Revealing the value of social media data in forecasting tourism demand: evidence from Twitter

Yuanming Qiu

Doctor of Philosophy

The University of Edinburgh

2023

Conference papers adapted from thesis

Qiu, Y., Lacka, E., & Ansell, J. (2022). Exploring the Twitter myth: the value of Twitter-generated variables on forecasting tourist arrivals. Presented at the *European Marketing Academy (EMAC) Annual Conference 2022*, May 2022, Budapest, Hungary.

Qiu, Y., Lacka, E., & Ansell, J. (2022). Exploring the role of Twitter communication flow in tourism demand forecasts. Presented at the *AIRSI2022: Technologies 4.0 in Tourism, Services & Marketing Conference*, July 2022, online.

Qiu, Y., Lacka, E., & Ansell, J. (2022). Forecasting tourism demand for tourist attractions: the applicability of Twitter-generated variables. Presented at the *British Academy of Management (BAM) Annual Conference 2022*, September 2022, Manchester, United Kingdom

Qiu, Y., Lacka, E., & Ansell, J. (2023). Measuring the impact of social media boycotts on tourist arrivals: evidence from the British Museum. Presented at the *Academy of Marketing Science (AMS) Annual (48th) Conference 2023*, May 20223, New Orleans, United States

Abstract

As the digital era advances, tourists are increasingly navigating a wide range of internet-based resources to make travel decisions. Among these, social media platforms have emerged as a beacon, guiding tourists' decisions about where to visit with an enormous amount of publicly accessible travel-related information. This significant shift has catalysed a surge in research aimed at leveraging the potency of internet-based data (e.g., search engine query data and web traffic data) as predictors for generating precise tourism demand forecasts. Such enhanced forecasts emerge as vital components in implementing efficient crowd control and capacity management strategies within the tourism industry.

While this emergent field is witnessing significant expansion, researchers have primarily utilized search engine data in their forecasts, due to their structured, time-series nature. These structured data elucidate the evolving patterns of tourists' attention and interest, effectively signalling volumes of tourist arrivals. Conversely, the potential of unstructured social media data for forecasting tourism demand remains a largely untapped area of tourism research, still nascent and significantly sparse in its contributions. Although some initial research signals the potential of social media data in tourism demand forecasting, the arena of demand forecasting at the individual attraction level needs to be investigated more. Therefore, an overarching aim of this thesis is to ***reveal the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level*** and, by doing so, fill this void in the tourism literature.

As the leading micro-blogging platform worldwide, Twitter presents a fertile landscape for researchers across various disciplines seeking to capture intricate informative signals. In tourism literature, however, the value of Twitter data is particularly underexplored within the context of forecasting tourism demand for individual tourist attractions, which are at the heart of the tourism experience and easily constrained by physical capacity. Addressing the above research aim, this thesis aims to reveal the value of Twitter data for tourism demand forecasts for attractions. Specifically, this research aims to address the following objectives:

(1) Assess the role of Twitter data in tourism demand forecasting at the attraction level.

(2) *Examine the advantages of deciphering embedded communication flows on Twitter for generating improved tourism demand forecasts at the attraction level.*
(3) *Evaluate the value of commonly excluded Twitter "noise" (i.e., - tweets that neither stimulate visits nor signal tourists' intentions to visit) for supplementing tourism demand forecasts at the attraction level.*

Grounded in signalling theory, this thesis embodies three empirical studies, utilizing the British Museum as a case study.

Initially, Chapter 4 provides an overall evaluation of the value of unprocessed Twitter data for signalling tourism demand at the attraction level and encourages efforts to generate refined forecasts of tourist arrivals.

Following this, Chapter 5 clusters collected tweets into three categories based on the embedded communication flows, as identified through text analysis. Subsequent to the classification, several forecasting models are applied to scrutinize the signalling value of variables derived from each communication flow. The findings reveal the importance of mapping Twitter communication flows to decipher the information dissemination process within the Twittersphere. Specifically, the results demonstrate that direct communications between the attraction of interest (in this case, the British Museum) and tourists offer more valuable signals for tourism demand prediction than other communication flows.

Before proceeding to the final study, it is essential to note that tweets that failed to stimulate visiting intentions (i.e., negative signals: tweets criticizing the British Museum's possession of contested artefacts and corporate sponsorship) were systematically removed from the dataset prior to variable construction in the above-noted studies. In fact, the insights from Chapter 5 underscore the vital necessity for data pre-processing to counteract the inherent "noise" in Twitter data. Chapter 6, drawing its focus back to those tweets previously categorized as noise, examines the signalling value of these tweets regarding the volume of tourist arrivals. To precisely extract those tweets roughly detected in Chapter 5, Chapter 6 first proposes a three-staged framework that adopts sentiment analysis, judgemental screening, topic modelling, and in-context keyword searching. It is revealed that the noise consists of boycott tweets opposing the British Museum's contested artefacts and its sponsorship by fossil fuel companies. Then, variables generated from these boycott tweets are tested for their dynamic relationship with the volume of tourist arrivals. The findings of Chapter 6 underscore the short-to-medium-term negative impact of likes on tweets that oppose fossil fuel sponsorship, on seasonally adjusted tourist arrivals to the British Museum.

The thesis contributes to the corpus of literature on tourism demand forecasting in three distinct areas. First, it substantiates the importance of Twitter data in forecasting tourism demand at the attraction level, by providing both a solid theoretical underpinning and compelling empirical evidence. This strengthens the emerging recognition of social media data (i.e., Twitter data) as a valuable resource in tourism research, particularly tourism demand forecasts for attractions. Second, it formulates and appraises a communication flow mapping framework within the Twittersphere, assessing its efficacy in improving demand forecasts. Focusing on specific communication flows, such as direct interactions between Twitter users and attractions, demonstrates the importance of these flows for predicting tourism demand. Third, it explores the value of the noise within the Twittersphere, which is often overlooked in existing literature, for deciphering fluctuations in tourism demand. This analysis deepens our understanding of the complexities of unstructured social media data and highlights the potential usefulness of the noise for tourism demand forecasts.

In closing, the thesis discusses these findings' theoretical and practical implications. It also acknowledges the inherent limitations of the current research. It postulates potential research trajectories, thereby setting the stage for continued scholarly exploration in the field of tourism research utilizing Twitter data.

Lay summary

In today's digital age, internet-based resources, particularly social media platforms, are becoming go-to hubs for tourists planning their travels. The vast amount of readily available travel-related information on these platforms plays a significant role in shaping tourists' decisions. This has prompted researchers to harness internet data, such as search engine queries and web traffic, to aid tourism demand forecasting. Such data have been found to be critical in predicting the influx of tourists, which is crucial for efficient crowd control and capacity management within the tourism sector. Traditionally, researchers have relied on structured search engine data for their forecasting efforts due to their ability to reflect trends in tourist interests and predict future visitor numbers. However, the potential of unstructured social media data in this domain has yet to be explored, especially when it comes to forecasting demand for individual tourist attractions. This thesis aims to address this shortfall by placing a specific emphasis on exploiting Twitter data to predict tourist demand at tourist attractions.

Twitter, a micro-blogging platform, provides a wealth of informative signals for researchers across an array of disciplines. Nevertheless, within tourism research, the potential of Twitter data to predict tourism demand for individual attractions has been mostly overlooked. This research aims to cover this uncharted territory by first exploring *whether* and *how* Twitter data can improve attraction demand forecasts. It sets out to assess the role of Twitter data in demand forecasting, explore the benefits of analysing embedded communication flows on Twitter, and evaluate how the often-excluded Twitter "noise" might enhance demand forecasts.

This thesis, grounded in signalling theory, uses the British Museum as a case study. The research aim is to make several significant contributions to the body of literature on tourism demand forecasting. Firstly, the thesis provides both theoretical and empirical evidence affirming the value of Twitter data for predicting demand at the attraction level. Secondly, it maps communication flows on Twitter, revealing the importance of direct interactions between Twitter users and tourist attractions. Lastly, it ventures into the typically ignored "noise" on Twitter, emphasizing its potential for offering valuable insights into variations in tourism demand. In essence, this research goes beyond conventional methods, exploring the untapped

potential of social media data (i.e., Twitter data) for attraction-level tourism demand forecasting.

Acknowledgement

As the day of my PhD completion finally arrived, it was not excitement that filled me most, but a deep sense of thankfulness and reflection. As I stand at this pivotal milestone in my life, I am keenly aware of the numerous individuals who have significantly contributed to my PhD journey. I recognize the depth of my gratitude to each of them for their invaluable support.

Foremost among them are my wonderful supervisors, whose expertise and dedication have been instrumental in shaping this work. I would like to first extend my heartfelt appreciation to my principal supervisor, Dr Ewelina Lacka. Her scholarly achievements have set a high bar of excellence that has inspired me throughout this process. Despite this, her commitment to providing constructive and detailed feedback never wavered. Dr Ewelina Lacka not only spurred my involvement in academic research but also provided essential guidance on disseminating my work effectively, thereby enhancing its impact. Her readiness to share her resources, such as research funding and academic networks, has given an immense boost to my research. Her firm belief in my capabilities, along with her persistent enthusiasm, saw me through the most demanding stages of my PhD journey.

My sincere gratitude also goes to my secondary supervisor, Professor Jake Ansell. It was his belief in my potential and his decision to accept me into the marketing group that paved the way for my PhD journey. Without him, this path may never have been possible. His invaluable insights and continuous encouragement for my involvement in conferences and academic activities significantly broadened my research experience. During my period of medical leave, his genuine concern for my well-being served as a source of strength and motivation. Professor Jake Ansell's unwavering dedication and guidance have been pivotal to my academic pursuit, and I am eternally grateful for his support throughout this journey.

I wish to extend a special thank you to my friends, colleagues, and faculty members, particularly Jiaman Xu, Cheng Lu, Zeyu Zhang, Xiyuan Ma, Xiaobai Shen, Antonia Gieschen, Qian Wang, Victoria Rodner, and Euejung Hwang. Each of these individuals has played a unique role in my journey, offering assistance and inspiration at various stages. Every interaction has left a lasting impact on my growth as a scholar, and their camaraderie has been invaluable.

I also want to express my deep appreciation to the University of Edinburgh Business School, and specifically the Marketing Group. I am grateful for the opportunity they have provided me to conduct my research in Edinburgh and for equipping me with all the necessary resources and comprehensive training to do so.

My final words of thanks are dedicated to my parents, whose unshakeable faith in me has been a consistent source of strength. Their constant support has formed the foundation of all my academic pursuits.

To everyone who has contributed to my PhD journey, I extend my sincerest gratitude. Your joint efforts have made this thesis possible, and sharing this experience with each one of you has been a true privilege.

Table of contents

List of figures

List of tables

List of abbreviations

ADF	Augmented Dickey-Fuller
ANN	Artificial Neural Networks
AR	Autoregression
ARDL	Autoregressive Distributive Lag
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
DMO’s	Destination Marketing Organization’s
ECM	Error Correction Model
EG	Engle and Granger
FGC	Firm-Generated Content
FPE	Final Prediction Error
GARCH	Generalized Autoregressive Conditional Heteroscedasticity model
GDP	Gross Domestic Product
HAC	Heteroscedasticity and Autocorrelation Consistent
HQ	Hannan-Quinn Information Criterion
IMI	Improvised Marketing Interventions
IRFs	Impulse Response Functions
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
LightGBM	Light Gradient Boosting Machine
LR	Likelihood Ratio
MA	Moving Averages
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MIDAS	Mixed Data Sampling
MSE	Mean Squared Error
MWOM	Micro-blogging Word of Mouth
OLS	Ordinary Least Squares
VAR	Vector Autoregression
VECM	Vector Error Correction Model

Chapter 1: Introduction

Over the past few decades, the tourism industry has experienced a notable expansion, becoming a dominant force in both advanced and emerging economies (Brida et al., 2020; Dogru & Bulut, 2018). This tremendous growth is closely linked to a consistent surge in tourist arrivals worldwide (Kirilenko et al., 2023). The consequence of this phenomenon is a burgeoning concern called “overtourism”, where the sheer volume of tourists can outpace the infrastructural or ecological capacity of a destination or attraction (Capocchi, Vallone, Amaduzzi, & Pierotti, 2019; Capocchi, Vallone, Pierotti, & Amaduzzi, 2019; Goodwin, 2017; Ko Koens et al., 2018; Phi, 2019). Tourist attractions, which lie at the heart of the tourism experience, are especially susceptible to these pressures. They include a diverse range of sites and experiences, such as historical monuments, natural landmarks, museums, theme parks, and festivals. The popularity of these attractions is often central to the prosperity of the surrounding communities and economies (Lew, 1987). Overtourism, however, has a detrimental effect on attractions, local communities, and destinations (Cheung & Li, 2019).

The narrative of steady tourism growth experienced a significant disruption due to the COVID-19 pandemic, leading to a substantial downturn in global tourism (Buckley & Chauvenet, 2022). This interruption momentarily alleviated the overtourism issue (Zhang et al., 2021), as travel restrictions and health concerns drastically reduced tourist numbers at popular destinations. Despite the impact of the pandemic, the tourism industry has demonstrated remarkable resilience (Duro et al., 2022). The United Nations World Tourism Organization (UNWTO) reports that international tourism is gradually recovering and is expected to return to pre-pandemic levels. In the first quarter of 2023, international arrivals reached 80% of pre-pandemic levels, with an estimated 235 million tourists travelling internationally. This represents more than double the number travelling during the same period in 2022 (UNWTO, 2023b).

This projected revival, attributed to pent-up demand and the gradual easing of travel restrictions (UNWTO, 2023), underscores the likely return of the overtourism phenomenon in the near future. Its characteristics, however, may differ from those of the pre-COVID era (Assaf et al., 2022), especially in urban areas that have traditionally faced high tourist activity (Page & Duignan, 2023). While tourist attractions, particularly those in urban areas, are central to the tourism experience, they often face significant challenges (K. Koens et al., 2018). On the one hand, they play a crucial role in local economies, generating revenue, creating jobs, and contributing to cultural exchange. On the other hand, they need to manage their physical capacity effectively to maintain the quality of the visitor experience and preserve cultural and/or natural heritage (Lew, 1987). Overcrowding, environmental degradation (Ziegler et al., 2016), and local community backlash (Cheung & Li, 2019) are potential downsides of overtourism, making effective capacity management a pressing concern (Peeters et al., 2018). These expectations emphasize the need for precise tourism demand forecasting, especially at the level of individual tourist attractions. Timely and precise demand forecasting can provide valuable insights, helping attractions to formulate proactive and sustainable management strategies (H. Li et al., 2020). These strategies can help mitigate the negative impacts of overtourism, balance the needs of tourists and local communities, and enhance the overall visitor experience, thereby ensuring the sustainability of tourist attractions in the long term (Bi et al., 2020).

Despite the wealth of research confirming the value of tourism demand forecasting for effective management in the sector, more studies are needed conducting forecasts specifically at the individual attraction level. Data availability often limits such studies (Song & Li, 2008). A considerable amount of existing research has targeted broader geographic scopes, such as provinces, countries, or regions (Bangwayo-Skeete & Skeete, 2015; Dergiades et al., 2018; Gunter & Önder, 2016; Havranek & Zeynalov, 2021; S. Li et al., 2018; Yang et al., 2015) which typically do not face the same physical space constraints as individual attractions. Moreover, the data used in these studies are often sampled at relatively low frequencies (i.e., monthly, quarterly, or yearly tourist arrival data), yielding insights primarily useful for long-term crowd management strategies (Song et al., 2019). There has been limited exploration of high-frequency tourism demand forecasts (Bi et al., 2021; Bi et al., 2020; H. Li et al., 2020; Volchek et al., 2019). Such high-frequency forecasts could complement longer-term predictions by offering more granular, short-term insights for effective crowd management (H. Li et al., 2020).

In this context, the advent of the digital era presents an intriguing prospect. As tourists increasingly turn to internet-based resources, particularly social media platforms, to plan their travel, an enormous amount of publicly accessible travel-related information is becoming available (Leung et al., 2013; Önder et al., 2019; Sedera et al., 2017; Xiang & Gretzel, 2010). These data can be leveraged to provide precise insights into tourists' preferences and behaviours, including travel intentions (Li et al., 2021), guiding strategic planning for crowd control and capacity management in the tourism industry (Gretzel, 2019). While the integration of internet data, such as search engine queries, into tourism demand forecasting has been shown to be beneficial, the potential of social media data remains largely underexplored (Hu et al., 2022). Research in this area has been limited, as demonstrated by Önder et al. (2019), who find that likes given by users on Facebook to a destination marketing organization's (DMO's) homepage could serve as an indicator for forecasting tourism demand at the destination level. Consequently, there is a need for further research to explore the potential of data from other social media platforms for predicting tourism demand (Gunter, Önder, & Gindl, 2019).

In the tourism sector, DMOs, attractions, and individual tourists utilize Twitter as a key communication tool. Its use varies from broadcasting marketing content (Leung et al., 2015) and promoting events (Bigné et al., 2019) to sharing personal travel experiences (Alkhammash et al., 2019; Sotiriadis & Van Zyl, 2013), making Twitter a key communication channel in tourism. Additionally, potential tourists increasingly turn to Twitter to seek travel-related information (Nicolau et al., 2020) and to gain answers to their inquiries (Alkhammash et al., 2019). Therefore, Twitter, with its real-time feed, vast user base, and a wide array of content – spanning from personal experience and opinions to news updates – serves as a distinctive tool for identifying tourism trends and intentions, subsequently enabling the forecasting of tourism demand (Bigné et al., 2019). Despite the recognized importance of social media data in tourism research (Li et al., 2021), and research that used Twitter data for tourism demand forecasting purposes (Cano-Marin et al., 2023; Petropoulos et al., 2022), there remains a notable gap in studies leveraging Twitter data for forecasting tourism demand, particularly in predicting tourist arrivals at individual attractions. Current research has primarily focused on using Twitter data to forecast hotel occupancy rates (e.g., Bigné et al. (2019)) and analyse tourists' preferences (e.g., acceptance of recommendations: Chang et al. (2018)). However, the broader application of Twitter data for understanding tourist arrivals to attractions is still unexplored. This thesis aims to address this research gap and by doing so respond to a growing number of calls for research that effectively utilizes Twitter data, and the diverse behavioural and textual cues available on Twitter such as comments, likes, shares, sentiment, and textual analysis, to inform and improve business decision-making (Sheng et al., 2021). In response to these gaps and demands, an overarching aim of this thesis is to ***reveal the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level***, thereby filling this notable gap in the existing tourism literature.

This thesis, drawing upon the principles of signalling theory (Spence, 1973) and the many-to-many communication framework (Chaffey & Ellis-Chadwick, 2019), effectively broadens our understanding of the extraction of valuable signals from Twitter data for the purpose of tourism demand forecasting. Using the British Museum as a case study, this research not only explores the value of Twitter data but also probes the impact of mapping Twitter-mediated communication flows to forecast tourist arrivals. Importantly, it considers the latent signalling value encapsulated in the often-overlooked "noise" data in the domain of tourism demand modelling. Thus, this thesis firmly positions itself at the vanguard of research that investigates the signalling potential of Twitter data in relation to the influx of tourist arrivals at the attraction level.

The key findings of the thesis can be summarized in three main points. Firstly, it has been established that Twitter data is a valuable resource for generating predictors of tourism demand at the attraction level. Secondly, the precision of short-term forecasts for tourist arrivals at attractions can be enhanced by analyzing the communication flows in attraction-specific tweets. Specifically, results of various forecasting models have demonstrated that the volume of tweets reflecting direct communication between tourists and the attraction yields the most precise forecasts. Thirdly, the thesis has illuminated the often-ignored "Twitter noise", confirming the signalling value of boycott tweets in predicting tourist arrivals at the attraction level. Empirical evidence from the case study reveals that increasing public endorsement (as measured by tweet likes) of criticisms against the British Museum's fossil fuel sponsorship exhibits a negative correlation with tourist arrival volumes. In contrast, tweets criticizing the museum's exhibition of contested artefacts do not demonstrate a clear correlation with tourist arrivals.

The rest of the thesis is organized as follows. Chapter 2 provides a review of the literature in the field of tourism demand forecasting. Importantly, this chapter identifies the research gap and establishes the three research objectives of this study: (1) Assessing the role of Twitter data in tourism demand forecasting at the attraction level; (2) Examining the benefits of deciphering embedded communication flows on Twitter to enhance tourism demand forecasts at the attraction level; and (3) Evaluating the significance of commonly excluded Twitter "noise" data (i.e., tweets that neither stimulate visits nor signal tourists' intentions to visit) in supplementing tourism demand forecasts at the attraction level. Theoretical frameworks underpinning the research are also discussed in Chapter 2.

Chapter 3 details the research context and methodological choices made to achieve each research objective. Specifically, it starts with an introduction to the research philosophy and approach, explaining the selection of pragmatism and abductive reasoning. The narrative then moves to the use of quantitative methods as the main analytical framework, with qualitative methods playing a supportive role in the thesis. This is followed by a discussion on the research context, data collection strategies, and ethical considerations. Then, Chapter 3 further introduces the methodological choices tailored to each study within this thesis. This includes detailing the framework for conducting social media data-based tourism demand forecasting and outlining the strategies employed to accurately identify boycott tweets within the vast Twitter database. Finally, it considers the limitations associated with the chosen research methods and suggests potential directions for future research methodologies.

Subsequently, Chapters 4 to 6 detail the empirical findings from three distinct studies using The British Museum as the main case study, with each chapter dedicated to addressing a specific research objective outlined in this thesis.

Chapter 4 (Study 1) focuses on the initial research objective, which aims to evaluate the utility of Twitter data in forecasting tourism demand at the attraction level. The findings from Chapter 4 validate the effectiveness of Twitter data in creating predictors for tourism demand forecasts. It encourages further investigation into predictors sourced from Twitter data, emphasizing their potential to improve forecast accuracy for tourist attractions.

In Chapter 5 (Study 2), the focus shifts to the second research objective, exploring the benefits of analyzing embedded communication flows on Twitter to enhance tourism demand forecasts at the attraction level. Insights from Chapter 5 demonstrate the value of Twitter data in producing precise demand forecasts for attractions. Notably, the chapter highlights the importance of resampling potential predictors based on tripartite communication flows on Twitter, suggesting this approach as a means to refine and increase the precision of tourism demand forecasting.

Chapter 6 (Study 3) addresses the final research objective, investigating the utility of often-excluded Twitter "noise" (tweets that do not directly encourage visits or indicate tourists' intentions to visit) in augmenting tourism demand forecasts at the attraction level. Specifically, Chapter 6 explores the relationship between boycott tweets and tourist arrivals, uncovering significant short-to-medium-term effects of fluctuations (defined as changes by a standard deviation) in the volume of likes on tweets advocating for a boycott of The British Museum due to its fossil fuel sponsorship. The findings underscore the critical need to reassess the "noise removal fallacy" in the context of social media data for tourism demand forecasting.

The thesis concludes with Chapter 7, which synthesises the significant theoretical, methodological, and practical contributions of this research. This final chapter also discusses the limitations of the thesis and proposes directions for future research in this evolving field.

Chapter 2: Literature review

This chapter presents a review of research relevant to the central focus of this thesis. In Section 2.1, we underscore the significance of tourism demand forecasting. Section 2.2 delineates the challenges posed by current tourism demand forecasting research, revealing a gap in the literature, which predominantly concentrates on destination-based demand forecasting, thereby leaving attraction-level demand forecasting a largely unexplored research area. Section 2.3 conducts a critical examination of the existing body of tourism demand forecasting research, emphasizing prominent issues and limitations. Section 2.4 introduces the utility of Twitter data within the tourism domain, simultaneously proposing the overarching research aim of this thesis. Section 2.5 presents a review of milestone marketing research that capitalizes on Twitter data, aiming to obtain a nuanced understanding of the opportunities and challenges tied to the incorporation of Twitter data into demand forecasting research. Following this, Section 2.6 explicates the theoretical foundations of the thesis, specifically signalling theory and the many-to-many communication framework. Finally, Section 2.7 summarizes the three research objectives that guide this thesis and provides indications for the ensuing research design.

2.1 Importance of tourism demand forecasting

As posited by Petropoulos et al. (2022), the theory of forecasting is founded on the principle that previous and present information can be utilized to forecast future outcomes. This notion is especially encapsulated within time-series data, where it is believed that discernible patterns in the past can be utilized to precisely predict future values. The prospect of an uncertain future can be both thrilling and daunting, prompting individuals and organizations to strive for risk reduction and utility maximization. Therefore, predicting future outcomes has long been an essential component of making informed decisions and developing effective plans.

It has also been pointed out by Petropoulos et al. (2022) that forecasting procedures are best when they relate to a problem to be solved in practice. The forecasting theory can then be developed based on an understanding of the essential features of the problem. In turn, the theoretical results can lead to improved practice. Given the wide range of situations that require forecasting, including but not limited to tourism management, there is a need for a variety of forecasting approaches to address the complexities of real-world scenarios such as the demand to visit places of interest (e.g., attractions, destinations).

Tourism demand is predominantly measured by the number of tourist arrivals and the level of tourist expenditure in both aggregate and per capita terms (Song et al., 2010). In accordance with the fundamental principles of forecasting theory, forecasting tourism demand plays a significant role in the broader discipline of forecasting (Petropoulos et al., 2022). It merges theories particular to the tourism context, such as demand, economic, and consumer behaviour theories (Song et al., 2019). These theories facilitate the precise prediction of tourist arrivals and expenditure, which in turn enhance decision making for stakeholders in the tourism sector.

Frechtling (2001) highlights five primary reasons precise demand forecasts are necessary in the tourism industry:

- Firstly, the perishable nature of tourism products underscores the importance of effective demand management (Perdue, 2002). Once a tourism opportunity is missed - the closing of a tourism attraction (e.g., a museum or a theme park) for the day, an airline departure, or the end of the day of a hotel - unsold admission tickets, seats, or rooms are lost, along with the potential revenue they represent. This situation requires effective short-term demand management and long-term anticipation to avoid unfulfilled demand and unsold inventory.
- Secondly, tourism involves an inseparable process of production and consumption. This process frequently includes direct interaction between service providers and consumers (tourists), such as between on-site attraction staff who provide support services and attraction visitors (Calver & Page, 2013). This necessitates the availability of adequate service personnel when and where they are needed by tourists (Jensen et al., 2017).
- Thirdly, a tourist's satisfaction is dependent on complementary services. For example, Calver and Page (2013) suggest that, presenting heritage in a way that enhances visitors' latent interest in history, the arts, and the environment can lead to successful outcomes. Jensen et al. (2017) also suggest that presentation platforms and support services can significantly improve tourist satisfaction and experiences at tourist attractions.
- Fourthly, leisure tourism demand is highly sensitive to both natural and human-made disasters (e.g., the COVID-19 pandemic and the 2008 financial crisis) (Page et al., 2012; UNWTO, 2023a). Holidays often serve as an escape from everyday stress, and crises such as wars, terrorist attacks, disease outbreaks, crime, or extreme weather conditions

can deter potential leisure travellers. The ability to forecast such events and their potential impact on tourism demand can help mitigate the negative effects on tourism-related income, employment, and tax revenue (Assaf et al., 2022; Wu et al., 2023).

- Finally, the supply side of tourism requires substantial, long-lead-time investments in infrastructure, facilities, and equipment. Most tourism products are characterized by high fixed or sunk costs for infrastructure such as attraction venues and lodging units (Perdue, 2002). Precise forecasting of future demand is crucial to avoiding the financial implications of overcapacity or the opportunity costs of unmet demand.

Frechtling (2001) contends that, while other industries may face one or more of these decision-making constraints, the tourism sector experiences all five, uniquely. Expanding on Frechtling's comprehensive understanding, research on "overtourism" also highlights the essential role of demand forecasting. An excessive influx of tourists (K. Koens et al., 2018; Peeters et al., 2018) can have negative impacts on the tourism ecosystem at both the micro and macro level (Capocchi, Vallone, Amaduzzi, & Pierotti, 2019; Capocchi, Vallone, Pierotti, & Amaduzzi, 2019; Goodwin, 2017; Ko Koens et al., 2018; Phi, 2019). Precise tourism demand forecasts can help the supply side of the industry adjust capacity thresholds in a timely manner, preventing overtourism rather than just managing it (H. Li et al., 2020). Short-term or real-time forecasts can also support the creation of an early-warning system, detecting trends in tourist concentration and spreading them across various destinations and attractions, thereby reducing the negative impacts of overcrowding (Gunter & Onder, 2016).

More recently, the onset of the COVID-19 pandemic in early 2020 has dealt a substantial blow to the tourism industry worldwide (UNWTO, 2023a), causing unprecedented disruption and in turn easing overtourism (Zhang et al., 2021). As such, in accordance with Frechtling (2001), a renewed focus has been placed on the forecasting of tourism demand in disaster scenarios (Song et al., 2023). Accurate forecasts can equip stakeholders with tools to navigate and recover from such crises more effectively, thereby ensuring the industry's resilience in the face of future disaster (Boto-García & Baños-Pino, 2023; Cheng & Liu, 2022).

2.2 Challenges of tourism demand forecasting research

2.2.1 Challenges of current tourism demand forecasting research

Despite the importance of tourism demand forecasting, there are still several challenges that require further attention in future research. In a critical appraisal, Gunter, Önder and Smeral (2019) scrutinize the scientific merit inherent in econometric tourism demand studies. To this end, they analyse a multitude of articles[1] from 2007 to 2017. Several key findings emerge from their analysis. Firstly, a lack of distinction between economic and statistical significance was pervasive across many studies. This interchangeability of terms will potentially lead to misinterpretations of results and implications. Secondly, they highlight a prevalent tendency to avoid discussions pertaining to the scale and dependability of estimation results. The absence of such discourse can restrict the comprehension of the real-world impact of findings. Furthermore, Gunter, Önder and Smeral (2019) point out an inadequate treatment of limitations in many instances. This oversight could potentially inflate the perceived validity and generalizability of the results. Additionally, a lack of justification for the chosen methods is noted, a deficiency that can render the assessment of methodological suitability and resultant impact problematic.

Based on these observed shortcomings, Gunter, Önder and Smeral (2019) advocate a greater focus on the extent and significance of estimated effects within econometric tourism demand studies. Such a shift, the authors argue, would bolster the scientific value of these studies and yield more actionable insights for both scholarly and industry contexts. In response to Gunter, Önder and Smeral (2019), Song et al. (2023) survey 425 articles relevant to tourism demand analysis and find that 39.78% of the surveyed papers do not make explicit statements about the effect sizes of relationships identified between tourism demand measures and predictors. Some research also lacks descriptive statistics of variables, which will further hinder audiences looking to discern the marginal effects of the relationships (Song et al., 2023).

Another substantial issue pinpointed by Song et al. (2023) is the unsystematic inclusion of variables in tourism demand models without theoretical support. The authors acknowledge the practicality and technical feasibility of adding a variable to a regression and indicating its

[1] A total of 115 articles were published between 2007 and 2017 in the journals Annals of Tourism Research, Journal of Travel Research, Tourism Management, and Tourism Economics.

statistical significance, especially given the vast data availability. However, they insist on the necessity of a theoretical underpinning that establishes a causal relationship between the newly incorporated variables and the dependent variable. Song et al. (2023) argue that, if this lack of theoretical support is not cautiously and explicitly addressed in tourism demand forecasting research, it could lead to problems such as misuse of behavioural economic concepts and theoretical models. For instance, Song and Wu (2022) question the inclusion of tourism variables as inputs in the Solow-Swan model. According to their argument, strong empirical evidence can be derived from variables that are theoretically unrelated to tourism demand. Yet, any conclusions based on such evidence could result in undesirable practical outcomes.

Furthermore, it is worth noting that approximately one fifth of the articles surveyed by Song et al. (2023) are forecasting studies that focus on improving the accuracy of tourism demand forecasting models, while the remaining articles examine tourism demand. A lack of theoretical understanding of the distinction between determinants (influencers) and reflectors (indicators) of tourism in the analytical framework would bias the interpretation of the results of these two research fields. Distinguishing between determinants and indicators is crucial in tourism demand forecasting. Determinants explain demand fluctuations, while indicators measure and predict these changes. Failing to discern this difference can result in incorrect causality assumptions, deficient forecasting models, misdirected strategies, and undermined research validity (Song et al., 2023). For instance, although search queries are proven to be valuable for forecasting tourism demand in different contexts (Volchek et al., 2019; Yang et al., 2015), their efficacy for reflecting tourism demand remains ambiguous and cannot be directly generated from improved forecasting performance (Song et al., 2023). This issue extends beyond the realm of tourism research and permeates a broader range of social science studies leveraging internet data. For instance, Ledford (2020) acknowledges the disruptive effects of social media data on social science research while simultaneously noting that some computational scientists are dismissive of the importance of theory in their work. Such dismissiveness can lead to a superficial interpretation of their findings, as evidenced in the study conducted by Boto-García and Baños-Pino (2023). In their exploration of habit and travel resilience post-COVID-19, they incorrectly interpreted the statistical significance of an inverted U-shaped relationship between past travel intensity and future travel intentions. According to Song et al. (2023), the findings presented by Boto-García and Baños-Pino (2023) were misrepresentations of both theory and reality, which resulted in the withdrawal of their study.

In addition to the interpretation issue and lack of theoretical justification of variables, Song et al. (2023) raise further concerns about model specifications. First, both theoretical and statistical justification should be given to the order of lagged components incorporated in a model. However, 40% of the surveyed studies do not describe their lag selection process. From the perspective of data empirics, the optimal lag length can be determined using information criteria. From the theoretical perspective, for example, it has been found in previous studies that travel decisions are typically made at least two to three weeks in advance of visits (Liu et al., 2019; Yang et al., 2015). Therefore, judgemental adjustments could be made to abnormal lag lengths determined based on information criteria when modelling tourism demand.

Another concern relates to the stationarity of the data used in the estimation process (Song et al. (2023). Time-series analysis using non-stationary data can yield spurious regressions and lead to unwarranted conclusions (Hyndman & Koehler, 2006). Surprisingly, only half of the reviewed studies explicitly discuss the stationarity of their data. Nevertheless, while some studies apply transformations such as taking the natural logarithm or the first difference of their variables, whether these transformations successfully achieve stationarity in the time-series data warrants explicit discussion (Önder et al., 2019; Song et al., 2023).

Furthermore, given that indicators (e.g., internet-data-generated predictors) signal the tourism demand influenced by determinants, an indirect causal relationship may exist between the indicators and determinants. Consequently, a model incorporating both could potentially face issues related to multi-collinearity. Moreover, indicators tend to carry a degree of noise. For example, a tourism researcher may explore an attraction’s website without having any actual intention to visit. These variances in the data could render multi-collinearity imperceptible in certain regression analyses (Song et al., 2023). In contemporary analyses of tourism demand, numerous internet-related variables, such as search queries (Bangwayo-Skeete & Skeete, 2015; Dergiades et al., 2018; Gunter & Önder, 2016; Havranek & Zeynalov, 2021; Liu et al., 2019), web traffic data (Gunter & Önder, 2016; Yang et al., 2014), and the number of reviews (Hu et al., 2022), have been employed. These highlighted issues collectively underscore the need for a more meticulous approach in future tourism demand forecasting research.

2.2.2 Challenges of the tourism demand forecasting research context

The existing body of literature on tourism demand forecasting, often limited by data availability, has primarily focused on regional or destination trends (Song & Li, 2008). A recent observation also suggests a concentrated focus on international tourist flows (Song et al., 2019). Hu et al. (2022) venture into the potential of online review data for enhancing tourism demand forecasting models, albeit within the broader context of international arrivals, specifically inbound tourism into Sweden. Notably, search engine query data, commonly used in these studies, have been almost singularly applied to broader destination-level tourism demand forecasting (Bangwayo-Skeete & Skeete, 2015; Dergiades et al., 2018; Gunter & Önder, 2016; Havranek & Zeynalov, 2021; S. Li et al., 2018; Yang et al., 2015). There have been exceptions such as the studies conducted by Volchek et al. (2019) and Huang et al. (2017), which examine attraction-level tourist arrivals.

Attractions such as museums and theme parks, bounded by physical capacity (Khaleeli, 2015; Rawlinson, 2019), frequently face the repercussions of excessive tourist inflows. These repercussions not only affect infrastructure and tourist experiences but also have broader implications for the community at large (K. Koens et al., 2018). In this context, Huang et al. (2017) emphasize the vital role of precise forecasting in the planning, operation, and management of tourist attractions. Precise and timely forecasts enable operators to devise and adapt strategic marketing plans, achieve financial objectives via effective pricing strategies, optimize staff scheduling, and establish efficient policies (Pan & Yang, 2017). Expanding on this, Li and Jiao (2020) note that precise tourist arrival forecasts can assist operators with implementing strategic crowd control and improving tourist experience management for attractions. Furthermore, they help optimize the allocation of resources and facilities. Such optimization not only minimizes the adverse effects on local ecological and social capacities but also enhances the long-term competitiveness and sustainability of tourist attractions. In light of these considerations, and acknowledging the modelling and data challenges discussed above, the importance of dedicated research at the attraction level becomes clear, yet it is thus far unexplored. To address this void in the literature, this thesis seeks to contribute to the tourism literature and practice exploring tourism demand for attractions. To be able to address the aforementioned knowledge gap, and overcome the challenges related to Twitter data processing and modelling, an in-depth understanding of tourism demand forecasting research is necessary.

2.3 Progress of tourism demand forecasting

2.3.1 Rival tourism demand forecasting models

Current tourism demand modelling and forecasting approaches can be broadly classified into two categories: quantitative and qualitative methods (Song & Li, 2008; Witt & Witt, 1995). Throughout the history of research in this field, the advancement of quantitative research methods and econometrics has led to a significant predominance of quantitative methodologies in tourism demand studies. Within the field of quantitative tourism demand forecasting research, a further subdivision can be made into two primary methodological categories: non-causal time-series models (e.g., naïve, autoregressive (AR), single exponential smoothing (ES), moving average (MA), and historical average (HA) models) and causal econometric approaches (e.g., the distributed lag (DL) model, the autoregressive distributed lag model (ARDL), the error correction model (ECM), and the ARIMAX (X represents the exogenous variables) model) (Höpken et al., 2021). These categories differ in their treatment of the relationship between the tourism demand variable and its contributing factors. Non-causal time-series models, as the name suggests, do not incorporate the effect of influencing factors on tourism demand. Instead, they rely solely on historical data pertaining to the variable itself. On the other hand, causal econometric approaches establish a connection between tourism demand and its influencing factors. This allows the model to quantify the impact of changes in these factors on tourism demand (Song & Li, 2008), and more recently the involvement of Artificial Intelligence (AI)-based models (e.g., artificial neural networks (ANNs), support vector regression (SVR), and long short-term memory (LSTM)) have attracted extra attention (Song et al., 2019). Due to AI-based models' ability to explain non-linear data without a priori knowledge of the relationships between input and output variables, they are widely used in data-driven tourism demand forecasting research (e.g., Höpken et al. (2021) and Sun et al. (2019)).

It is important to highlight that existing literature emphasizes the absence of a universally superior model for tourism demand forecasting for all situations (Song et al., 2019). Instead, the literature suggests that the choice of model should be guided by the specific research context and the availability of data, advocating a tailored approach to forecasting rather than a "one size fits all" methodology (Petropoulos et al., 2022). For example, although dominating the field, neither non-causal pure time-series models nor causal econometric models can easily process time-series data with non-linear, non-stationary, and complex patterns (Bi et al., 2020).

AI-based models are suggested to be effective for processing data with these patterns, but are reliant on substantial amounts of training data being available (C. Li et al., 2020) and are criticized for producing results that are difficult to interpret (Song & Li, 2008).

Moreover, while qualitative approaches to tourism demand forecasting, such as judgemental forecasting, have been criticized for potentially introducing human bias (Petropoulos et al., 2022; Song et al., 2019), scenario-based forecasting has been found to be valuable for adjusting for the impact of the COVID-19 pandemic on quantitative methods (Zhang et al., 2021). This highlights an important limitation of current quantitative tourism demand forecasting practices in terms of how they account for unexpected events like pandemics (Song et al., 2023).

As no single model universally outperforms the others in tourism demand forecasting, researchers have recognized an emerging trend towards the use of combined and hybrid models (e.g., average-based methods, forecast-error-based weightings, and regression-based integrations) (Song et al., 2019). These innovative approaches integrate different quantitative methodologies or amalgamate quantitative and qualitative methods. The intent behind this synthesis is to harness the strengths of each component while mitigating their individual weaknesses, thereby enhancing overall forecast accuracy (Chu, 1998; Petropoulos et al., 2022; Song & Li, 2008).

The evolution of tourism demand forecasting models has also seen a concerted effort to identify the determinants of tourism demand. These determinants vary depending on whether a macro or micro perspective is taken. At the macro level, gross domestic product (GDP), exchange rates, and relative prices are key determinants (Adedoyin et al., 2021). Notably, the use of leading indicators, such as industrial production, which is a leading indicator for real GDP, has been proposed for short-term tourism demand forecasting (Petropoulos et al., 2022). Conversely, at the micro level, personal and travel-related characteristics, such as age, income, travel companions, previous visits, pricing, and seasonality, play significant roles (Pak, 2020; Wong et al., 2017; Yang et al., 2017). Yet no determinant is deemed to be superior over the others across different research contexts.

2.3.2 Data in tourism demand forecasting

As noted by H. Li et al. (2020), the accuracy of a forecasting model is intrinsically tied to the quality of the input data. Traditionally, the time-series data used for tourism demand forecasting are those that have already been discussed. A significant portion of research has focused on the economic determinants of tourism demand, understanding that, as a discretionary good, tourism is often directly affected by the economic climate. The majority of these studies have explored tourism demand at the region or destination level. For example, Eugenio-Martin et al. (2008) suggest that, for the worldwide case, economic development (measured by GDP) matters and makes a difference to tourists' decision making. Similarly, by analysing world tourism demand (in 218 countries), Martins et al. (2017) find that while the world's GDP per capita was a more relevant factor for explaining tourist arrivals, relative prices had a significant impact on explaining expenditure, with a unitary elasticity. By modelling the unemployment rate at the macro level, Wong et al. (2017) reveal a significant cross-level moderating effect on the relationship between the travel frequencies of domestic and international markets. More recently, Adedoyin et al. (2021) have modelled international tourist arrivals to the Maldives and discovered that tourism tax adversely influences inbound travel.

This preference for data stems from the lower costs associated with data collection and model estimation (Song & Li, 2008), as well as the proven capability of these variables to yield satisfactory results (Song et al., 2019). Nevertheless, the use of solely historical archival data has its critics, with some researchers arguing that the results might be inaccurate in the face of unexpected events (Yang et al., 2014). The late publication and low-frequency nature of such data present additional challenges, both of which can negatively affect the accuracy of forecasts (H. Li et al., 2020). Therefore, pure time-series models, such as autoregressive integrated moving average (ARIMA) model, generalized autoregressive conditional heteroscedasticity (GARCH) model, and ES model, which primarily incorporate historical tourist arrivals data, often serve merely as performance benchmarks in contemporary forecasting (Li et al., 2021; Song & Li, 2008). Following the advice of Lütkepohl (2005), forecasters should include relevant information in their datasets at the forecast origin, provided it enhances the accuracy of their predictions. This implies that indicators generated from other data sources, if proven to enhance forecast accuracy, should be integrated into the forecasting process (Petropoulos et al., 2022). With the increased accessibility of internet data in recent years, there has been a shift in the literature towards exploring this rich data source (J. Li et al., 2018).

2.3.3 Internet data in tourism demand forecasting

As suggested by Petropoulos et al. (2022), the value of internet interactions as indicators is underpinned by the fundamental premise that prospective tourists seek information about their intended travel destination before their actual trip. Due to the relatively low search costs, internet mediums such as websites and search engines enable potential tourists to collect information easily (Önder et al., 2019; Zipf, 2016). Therefore, as a way to depict tourists' online behaviour (e.g., destination/attraction selection and itinerary planning) and further infer their decision-making process (J. Li et al., 2018), new input indicators generated from internet data that (a) reveal preferences, (b) provide data at a relatively high frequency (e.g., daily or weekly), and (c) depict changes in a tourist's preferences have been employed in the pursuit of higher accuracy in tourism demand forecasting models. According to Li et al. (2021), frequently used internet data in the field of tourism demand forecasting include search query data, web traffic data, and social media data.

2.3.3.1 Search engine query data

Search engine query data obtained from search engines such as Google or Baidu provide search volumes for different keywords, based on which it is possible to identify tourists' interests and intentions, and further predict their behaviour, including where and how they travel (Sun et al., 2019). As search engines have long been suggested to be essential online sources of destination and/or attraction-related information for tourists, in the travel-planning process (Dergiades et al., 2018), it is reasonable to assume that such data can explain fluctuations in tourism demand (H. Li et al., 2020; Volchek et al., 2019).

A crucial advantage of query data obtained from search engines is that it is volume-based normalized data and is easily accessible (Volchek et al., 2019). As a free and publicly available online portal of Google Inc., Google Trends provides billions of pieces of daily search data, revealing the geospatial and temporal patterns of search volumes for user-specified keywords (Gunter, Önder, & Gindl, 2019).

Based on an extensive literature review, Li et al. (2021) attest that search engine query data are the data most commonly incorporated into tourism demand forecasting studies, as they are recognized for their ability to increase forecasting performance in both econometric (Dergiades et al., 2018; Hu & Song, 2019; Huang et al., 2017; Li et al., 2017) and AI-based models (Bi et al., 2020; Law et al., 2019; H. Li et al., 2020; Volchek et al., 2019). Li et al. (2017), for example, employed Google Trends data to forecast international tourist arrivals to Hong Kong. Their findings revealed improved efficiency using econometric models (i.e., ARIMA series models). Dergiades et al. (2018), Huang et al. (2017), Hu and Song (2019), and Law et al. (2019) all reached similar conclusions. Using adjusted search query data, Dergiades et al. (2018) achieved improved forecasts of international tourism arrivals to Cyprus. Employing search engine query data to predict tourist arrivals to the Forbidden City in Beijing, Huang et al. (2017) revealed that a causal econometric model (i.e., ARDL) with Baidu Index data outperformed a pure time-series model (i.e., ARIMA) based on historical data. More recently, Hu and Song (2019) proved the efficiency of search engine query data in their work predicting the arrivals of short-haul visitors from Hong Kong to Macau using AI models (i.e., ANNs). Law et al. (2019) used search engine query data from both Google and Baidu to forecast tourist inflows into Macau with AI models, reporting a significant improvement in forecasting performance. With the notable exceptions of Volchek et al. (2019) and Li et al. (2022), search query data have been used to examine tourism demand at destinations, leaving tourism demand forecasts at the attraction level an under-researched area of study.

Although search engine query data provide demonstrable improvements in tourism demand forecasting accuracy, such studies suffer from several limitations. First, too many keywords used in the search process may result in highly correlated data series, which can lead to biased forecasts (Li et al., 2017). Models incorporating search engine query data require sophisticated query keyword selection designs, such as dimensional reduction techniques (Li et al., 2017), and principal component analysis models (S. Li et al., 2018). Moreover, although easily accessible, search engine query data are generated by search engines' unpublished algorithms, which may limit the reliability of the data. Google Trends, for example, does not report the raw search volumes of a given query. Instead, it reports a normalized index that displays how frequently a keyword has been searched for relative to the total search volume (Yang et al., 2015). The timeframe for daily-frequency data generated by Google Trends is also restricted, to 90 days (Carrière-Swallow & Labbé, 2013; Hamid & Heiden, 2015). Data retrieved under multiple time windows cannot easily be merged into one meaningful time series without complex pre-processing (Hamid & Heiden, 2015). This restricts researchers' ability to obtain high-frequency data sampled over an extended timeframe, and limits them to low-frequency forecasts (Havranek & Zeynalov, 2021). Modifications to Google Trends' user interface and its capabilities are also not formally reported. This may result in variations in search query results and could lead to biased conclusions (Nuti et al., 2014). Additionally, the unavailability of raw data may prevent researchers from correcting any potential biases in the data (Geva et al., 2017). Lastly, existing literature posits the importance of distinguishing whether heightened tourist awareness stems from positive or negative circumstances (Dellarocas et al., 2007). Yet, search engine query data, being volume-based, only offers insights into tourists' overarching attention and awareness of tourism products and services (depending on the query used) (Hu et al., 2022). Such data, however, fall short of revealing tourists' preferences or sentiments towards tourism products and services they are searching for, thereby limiting the data's capacity to sieve out negative awareness or attention (Geva et al., 2017). As such, it underscores the necessity of leveraging more sophisticated data sources to capture the full spectrum of tourist awareness and attention and its implications on tourist behaviour (refer to the discussion in Section 2.5.2), that is, social media data.

2.3.3.2 Web traffic data

Another type of structured, volume-based internet data that can be used in tourism demand forecasting is web traffic data. Web traffic data represent numbers of visits to websites and indicate tourists' interests (Yang et al., 2014). Yang et al. (2014) suggest that web traffic data may provide greater predictive power than search engine query data, considering tourists' reliance on tourism directory websites when planning their trips (Gunter & Önder, 2016). Although potentially useful, unlike search engine query data, web traffic data can only be retrieved from Google Analytics after permission has been obtained from the website's owner, which makes such data difficult to access. Such restrictions limit scholars' attempts to conduct tourism demand forecasting research using web traffic data (Li et al., 2021).

Despite restrictions related to data access, there have been some notable research exceptions. Utilizing web traffic data, Yang et al. (2014) and Yang et al. (2015) forecasted hotel demand in South Carolina and Beijing respectively. Gunter and Önder (2016) used web traffic data to predict tourist arrivals to Vienna. Thus far, however, no studies have used web traffic data to examine tourism demand for attractions.

2.3.3.3 Social media data

Social media data have received considerable attention in tourism demand forecasting research in recent years (Li et al., 2021) (see Table 2.1). For example, Önder et al. (2019) and Gunter, Önder and Gindl (2019) examined the volume of likes on a DMO's Facebook homepage and confirmed that Facebook likes could be used as a supplement to search engine query data to forecast tourist arrivals to a destination in the short term. Their studies are underpinned by the least effort theory, which assumes that individuals use the most convenient search method, when trying to find information, and the most familiar tools available (Zipf, 2016). Considering the popularity of Facebook (i.e., as of the first quarter of 2023, there were 2.99 billion users worldwide: Statista (2023)), Önder et al. (2019) suggest it is easier for regular Facebook users to look for the destination and/or attraction's Facebook homepage rather than going to the actual website for event and other tourism-related information. As a marker of tourists' search behaviour on Facebook, likes on a DMO's Facebook homepage therefore serve as an efficient indicator of users' visiting intentions in the short run.

Table 2.1 Selected studies on tourism demand forecasting utilizing social media data

Study	Social media data			Resources	Inclusion of user online behaviour data from other sources	Data pre-processing procedures	Forecasting model	Forecasting objectives
	Categories	Numeric data	Textual data					
Bigné et al. (2019)	1. *User reaction data*: volume of retweets and replies posted by users 2. *DMO-related data*: volume of event tweets, tourist attraction tweets, and DMO retweets.	√	√	*Micro-blog platforms*: Twitter		1. *Content analysis*: judgemental classification via two experts 2. *Text mining*: clustering via QDA Miner	ANN	Hotel occupancy rate
Gunter, Önder and Gindl (2019)	*User reaction data*: volume of likes on DMO's Facebook homepage	√		*Social networking platforms*: Facebook	*Search engine query data*: Google Trends		ARDL and MIDAS	Tourist arrivals to destinations
Önder et al. (2019)	*User reaction data*: volume of likes on DMO's Facebook homepage	√		*Social networking platforms*: Facebook			ARDL and MIDAS	Tourist arrivals to destinations
H. Li et al. (2020)	*Online review data*: review volume and average review rating	√		*Online travel agencies*: Ctrip and Qunar	*Search engine query data*: Baidu Index		ARIMAX, SVM and RF	Tourist arrivals to attractions
Chen et al. (2021)	1. *User demographic data*: age, gender, province 2. *User online operational behaviour data*: browsing and purchasing 3. *Online review data*: review volume and average review rating	√	√	*Online travel agencies*: Haobaoche		1. The probabilistic linguistic term sets (PLTSs) model is used to describe user online operational data 2. *Text mining:* textual preference extraction model is used to obtain purchase-related variables	LR, RF, XGBoots, LightGBM, and Stacking	Customer purchasing behaviour
Hu et al. (2022)	*Online review data*: review volume, average review rating and average review sentiment	√		*Online travel forum*: TripAdvisor		*Text mining*: Vader is used to generate the sentiment polarity score for each review	MIDAS	Tourist arrivals to attractions, hotels, and shopping markets
Wu et al. (2022)	*Online review data*: sentiment index	√	√	*Online travel agencies*: Ctrip and Qunar		*Text mining*: LSTM and naïve Bayesian (NB) models are used for sentiment polarity determination	ARIMAX	Hotel occupancy measured by review count

In addition to Facebook, Twitter – the most popular micro-blogging platform – has received scholarly attention. Building on the framework of content co-creation among DMOs and users on social media, Bigné et al. (2019) examined the effect of DMOs' social media activity (i.e., retweets and replies to tweets created by users), the number of followers DMOs have, and users' actions (i.e., replies, retweets, and likes on tweets created by DMOs) on the hotel occupancy rate during the Easter holiday. Confirming the fact that DMOs' social media content plays an important role in shaping tourists' behaviour (Peters et al., 2013; Uşaklı et al., 2017), Bigné et al. (2019) further analysed the effects of tweets in different categories (i.e., event tweets, tourist attraction tweets, socialization tweets, commercial tweets), differentiated by text mining, on hotel occupancy. Overall, they confirmed that the number of retweets and replies by users, and the number of event tweets, tourist attraction tweets, and retweets by DMOs, can predict the hotel occupancy rate for a given destination. It is worth noting that Bigné et al. (2019) discover that event tweets had the greatest impact on hotel occupancy, with importance values greater than 80%, while socialization tweets had only a moderate effect of around 50%. The main findings of a meta-analysis conducted by Liadeli et al. (2023), which included 86 articles examining the relationship between social media engagement and sales, support Bigné et al. (2019) in showing that social media content should be more functional than emotional, in order to drive sales. Additionally, Liadeli et al. (2023) concluded that building a large social media following is not necessarily critical for increasing sales, a finding that aligns with Bigné et al. (2019)'s discovery that the effect of the number of followers of a DMO's Twitter account on the hotel occupancy rate is not significant, with an importance value below 40%.

Overall, existing studies on social media confirm its advantages over other types of internet data such as search engine query data or web traffic data in depicting tourist behaviour and contributing to tourism demand forecasting. Yet, as shown in Table 2.2, limited research has utilized this data type to examine tourism demand at the attraction level. With this in mind, the focus of this section now shifts to the value of social media data in the tourism domain.

2.4 The importance of social media data in the tourism domain

Academics increasingly acknowledge the value of social media data in the tourism domain (Gunter, Önder, & Gindl, 2019). Defined as "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0" (Kaplan & Haenlein, 2010, pp. 61), social media (e.g., Twitter and Facebook) are claimed to be spaces that enable tourists' generation, dissemination, and consumption of travel-related information (Magasic & Gretzel, 2020). Providing focused, up-to-date, and credible information, social media are used by tourists as an essential information source, especially in the planning stage of decision making (Leung et al., 2013; Önder et al., 2019; Sedera et al., 2017; Xiang & Gretzel, 2010). Social media also allow tourists to exchange information, consult, and discuss travel plans with others, especially with those who have already visited a destination or attraction (Munar & Jacobsen, 2014). This is particularly important as interpersonal communication through social media has become a way to reduce perceived risks related to travel decisions (Luo & Zhong, 2015).

Research investigating the value of social media data in tourism demand forecasting is growing (Gunter, Önder, & Gindl, 2019). The literature has affirmed that significant advantages of social media data over historical arrival data and economic indicators lie in (1) the revealing of preferences, (2) the provision of higher-frequency data (e.g., hourly and daily), and (3) the depiction of changes in consumers' preferences (Yang et al., 2014). However, limited studies provide theoretical explanations on the incorporation of internet data (Song et al., 2023).

Xiang et al. (2015) suggest that pre-trip planning can be considered a fundamental component of the trip experience, with tourists often needing to obtain a substantial amount of information to develop a travel plan. Information available to individual tourists has a significant impact on various aspects of the tourist's decision-making process, especially when choosing a destination or attraction to visit. In particular, while the internet has become the primary information source for travel planning, and search engines dominate the online information search environment, today's consumers continue to prioritize social media platforms to obtain rich and personally relevant information for upcoming travel plans (Xiang et al., 2015). Social media, with its numerous communicative modalities, is a growing force in shaping tourists' online behaviour (Dinhopl & Gretzel, 2016; Fogg, 2002; Gretzel, 2019; Jacobsen et al., 2019; Munar & Jacobsen, 2014), especially travel-planning activities (Xiang & Gretzel, 2010) (refer to Table 2.2 for detailed illustrations).

Table 2.2 Examples of research investigating the impact of social media on tourists' behaviour

Research	Impact
Leung et al. (2013); Önder et al. (2019); Sedera et al. (2017); Xiang and Gretzel (2010)	Social media as an information source: social media is considered a convenient source for tourists to obtain others' views and accounts of their experiences when making travel-related decisions, due to its interactive feature and ability to provide easy access to information. Tourists perceive the content as more focused, experiential, up-to-date, and credible than the information available via other channels in the age of information overload.
Dinhopl and Gretzel (2016); Lo et al. (2011)	Social media as frames: social media provides new outlets for tourists to curate and share their travel photographs. Social media also changes how tourists take photographs and videos, and how they frame their visuals.
Dinhopl and Gretzel (2016); Jacobsen and Munar (2012)	Social media as panoptic forces: tourists who present themselves and their travel experiences on social media are subject to the gaze of social media audiences. These audiences not only praise but are also quick to shame and punish. Hence, this networked gaze becomes an omnipresent and powerful disciplinary force, shaping both social-media-related behaviour and real-life experiences, because tourists seek out experiences that will look particularly good on social media, or they rehearse these experiences so as to later receive the approval of others. Consequently, there is low "real-time" use of social media for holiday content sharing.
Fogg (2002)	Social media as persuasive technologies: there are social-media-designed business models which integrate techniques of persuasion to encourage particular behaviour, primarily to ensure that users feel the need to check their social media feeds and posts regularly. For instance, TripAdvisor awards travel reviewers with so-called "badges" to psychologically reward them and motivate them to produce more reviews, while Facebook uses notifications to condition users to stay engaged. Indeed, social media use can be so rewarding that it becomes addictive. The fear-of-missing-out (FOMO) spurred by social media further fuels its addictive potential.
Munar and Jacobsen (2014)	Social media as a social forum: social media facilitates sociality across time and space, allowing patrons of a hotel to share their experiences with potential guests, connecting locals and tourists in travel forums, and permitting the sharing of experiences with those who stay at home. Personal-related expectations, such as establishing and maintaining social capital, as well as community-oriented altruistic motivations, such as helping others avoid a wrong choice, have been confirmed as important considerations for tourists sharing their experiences on social media.
Gretzel (2017)	Social media as political means: in the tourism context, activism is commonly constrained because of the myriad of stakeholders involved and the substantial geographic distances between them. However, given a facelift by social media, which allow for more fluid membership and asynchronous participation in sustained movements beyond time and distance, a wide variety of activism types are applied to tourism through social media, ranging from individual actions against companies to large-scale consumer boycotts and social movements (e.g., "#antitourism").

2.5 The importance of Twitter data in the literature

2.5.1 Facebook vs Twitter

As is indicated by Table 2.3, a significant difference between social networks (e.g., Facebook) and micro-blogs (e.g., Twitter) relates to audience expectations and goals (Marchand et al., 2017). Social networks revolve around users' social ties, which creates trust and personal connections and relates to the goal of building relationships (Buzeta et al., 2020). On Facebook, a follower relationship is reciprocal (Toubia & Stephen, 2013). By contrast, micro-blogs (i.e., Twitter) are content-based and focus on broadcasting: micro-blogging platforms like Twitter are more publicly visible and allow wider access than other types of social media (Kaplan & Haenlein, 2011).

Table 2.3 Differences across social media channels

Channels	Social Community	Ties	Involvement	Interactivity	Private vs. Public	Persuasiveness
Social networks (e.g., Facebook)	√	Strong	Low	High	Private	Medium
Micro-blogs (e.g., Twitter)	√	Weak	Low	Medium	Public	Low
Blogs, forums, & online communities	√	Strong	High	High	Private	High

Source: Liadeli et al. (2023), web Appendix C

Kaplan and Haenlein (2011) also suggest that micro-blogging is a phenomenon that refers to the broadcasting of brief messages to some or all members of the sender's social network through a specific web-based service. Liadeli et al. (2023) argue that, due to the significant differences among various social media channels, the effects of owned social media[2] content on engagement and sales are likely to vary. They predict that content shared on social networks such as Facebook will have a greater impact than content shared on micro-blogs because social networks are characterized by stronger social ties and higher levels of trust. On the other hand,

[2] The term "owned social media" refers to the active social media presence of brands, which enables direct interaction with customers (Stephen and Galak, 2012).

micro-blogs are better suited for the wider dissemination of content and generally reach a larger number of consumers in a shorter amount of time.

De Oliveira Santini et al. (2020) discover that Twitter is twice as likely as other social media platforms to improve customer engagement by promoting satisfaction and positive emotions. Since the default setting for Twitter is that all tweets are public, user-generated content on Twitter is more likely to be seen by the general population (Lu et al., 2022). In contrast, discussion groups are private in Facebook by default[3] (Hewett et al., 2016). Hennig-Thurau et al. (2015) note that micro-blogging's ability to disseminate information quickly among networked individuals has been demonstrated during recent world events, such as the Arab Spring movement (Bruns et al., 2013) and the 2016 U.S. presidential election (Grover et al., 2019). This argument is reinforced by De Oliveira Santini et al. (2020) as discussed above.

In the literature, Bigné et al. (2019) and Önder et al. (2019) utilized distinct strategies in constructing indicators from Twitter and Facebook data, due to the unique characteristics of each platform, as identified by Liadeli et al. (2023), which necessitate different theoretical understandings and analytical approaches. An essential aspect of Twitter, crucial when analysing Twitter data, is that posting content serves as a means for users to attract new followers. This claim is supported by data provided by Toubia and Stephen (2013) and is consistent with Valsesia et al. (2020), who find that the number of people users choose to follow is a significant indicator of their online influence. It is also worth noting that, unlike on other social networks, reciprocity (i.e., user A follows user B, and vice versa) on Twitter is only moderate. According to Toubia and Stephen (2013), out of all pairs of users on Twitter who share at least one link, only 22.1% have a mutual relationship where both users follow each other. This means that a user's follower count is not solely determined by who they follow, and posting tweets can be an effective way for users to gain new followers. Building on that, Valsesia et al. (2020) further find that if a social media user has a significant number of followers, limiting the number of people they follow can enhance their perceived level of

[3] Twitter offers the option to unfollow or block other users and to set an account to "protected". Unprotected, or "public", accounts (the default setting) can be followed and accessed by any user. A protected account requires the user's approval of new followers and restricts tweet visibility to followers only. Yet, follower count, followed users count, and the total number of tweets are publicly accessible for both public and protected (or private) accounts.

influence. Moreover, a higher perceived influence leads to increased engagement with shared content (i.e., increased likes and retweets).

Despite the increasing acknowledgment of social media's effect on tourists' behaviour, the unique potential of social media data, and specifically that from Twitter, remains relatively unexplored, particularly at the attraction level. The platform's real-time feed, large user base, and diverse content (ranging from personal experiences and opinions to news updates) make it a unique instrument for discerning travel trends and intentions (Bigné et al., 2019). Furthermore, the multifaceted nature of Twitter data, encompassing elements such as textual content, timestamps, hashtags, retweets, likes, and geo-location tags (Hennig-Thurau et al., 2015; Lu et al., 2022), provides an additional wealth of information that can be harnessed to enhance forecasting models. Despite the potential value of Twitter data, however, only Bigné et al. (2019) have employed it, to examine hotel occupancy in holiday scenarios, while there has been sparse research on the facilitative role Twitter data might play in tourism demand forecasting, particularly at the attraction level. To contribute to the literature, therefore, the aim of this research is *to* ***reveal the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level*****.**

2.5.2 Existing marketing and management research on Twitter

While the value of Twitter data for forecasting tourism demand has yet to be fully explored, such data have been extensively used in the existing marketing and management literature. Within the literature, Twitter data have been analysed and broken down into multiple dimensions, which are outlined in the "key data dimensions" presented in Table 2.4.

Table 2.4 Selected marketing research utilizing Twitter data

Study	Theoretical Focus	Empirical Focus	Twitter Data	Key Data Dimensions	Main Findings
Herhausen et al. (2023)	Customer arousal in service failures and recoveries	Customer perception (gratitude)	FGC and UGC (i.e., 564 customer-initiated complaint interactions of a U.K. retailer)	Customer complaint (i.e., negative high arousal); firm response (i.e., active listening and empathy); and customer response (i.e., gratitude) identified via written style	1. By demonstrating active listening and empathy in their response, firms can elicit gratitude from customers in high-arousal states, regardless of whether the issue has been fully resolved yet. 2. When dealing with customers who are highly aroused, displaying empathy is more effective than simply engaging in active listening.
Villanova and Matherly (2023)	Unacceptable brand mentions and social media disengagement	User engagement (i.e., customer disengagement)	FGC and UGC (i.e., 10 MLB teams and 604,330 individuals)	Disengagement (i.e., unfollowing an MLB team on Twitter within 41 days), self-brand connection (i.e., measured by tweets, mentions, likes, retweets, and sentiment)	Individuals who have a strong connection to a brand are more likely to disengage when they observe socially unacceptable mentions of the brand.
Lacka et al. (2022)	FGC's impact on firm outcomes	Financial outcomes (i.e., stock price)	FGC (i.e., S&P 500 IT firms)	Sentiment (positive or negative) and subject matter (consumer or competitor orientation) of tweets	1. Tweets that focus solely on the valence or subject matter related to consumer or competitor orientation can cause short-lived price impacts, while tweets that incorporate both attributes can have a lasting effect on prices. 2. Permanent price impacts are most noticeable when negative-valence tweets are directed towards competitors.

Liaukonytė et al. (2022)	Adding empirical evidence to the literature on political consumerism	Financial outcomes (i.e., Food sales)	UGC: tweets mentioned “Goya” in July 2020; the focused-upon tweets are those posted using identified hashtags within the two weeks following the scandal (9 July 2020)	Tweets, mentions, and hashtag usage	1. Despite boycott-related tweets and media coverage outnumbering buycott-related ones, Goya experienced a temporary 22% increase in sales. However, this boost subsided entirely within three weeks. 2. The study revealed a considerable sales increase (56.4%) in counties with a predominantly Republican population, but no notable impact from the boycott was detected in counties with a predominantly Democratic population or among Latino consumers, who form the core customer base for Goya. 3. The findings may be explained by brand loyalty and switching costs.
Lu et al. (2022)	Joint effects of FGC and UGC	Financial outcomes (i.e., box office sales)	1. FGC and UGC (i.e., 145,502 FGC and 5.9 million UGC with 159 movies) 2. FGC according to its source (i.e., movie, studio, actor, and director) 3. Data obtained via Crimson Hexagon	Tweets, retweets, replies, and hashtag usage	The authors discovered a new positive indirect effect of FGC on sales that occurs through UGC, known as the ripple effect. They also demonstrated that this indirect effect is primarily responsible for the impact of FGC on sales.
Borah et al. (2020)	Improvised marketing interventions (IMI)	1. User engagement (i.e., virality) 2. Financial outcomes (i.e., firm value)	1. UGC: 99 hours of data of tweets mentioning @oreo 2. FGC: 462 IMI messages from 139 brands across 58 different industries over the six-year period between 2010 and 2015 3. FGC: 188 IMIs across eight unique firms 4. FGC: 470 tweets posted by 5% of the S&P 500 firms (randomly selected) for the month of April 2019	Tweets, retweets (virality), and specific text characteristics (i.e., humour and non-anticipation)	1. The inclusion of IMI messages results in increased virality and significantly amplifies virality compared to non-IMI messages. 2. IMIs that exhibit qualities such as humour, timeliness, or unexpectedness have the ability to amplify both virality and firm value.

Valsesia et al. (2020)	Two-step flow of communication and autonomy	User engagement (i.e., follows, likes, and retweets)	UGC: 1,581,522 tweets written in English on September 16, 2016, in the Los Angeles metropolitan area, from 784,170 distinct users	Tweets, follows, likes, retweets	1. Regardless of the number of followers a user has, the number of people they choose to follow is a significant indicator of their perceived influence online. 2. If a social media user has a significant number of followers, limiting the number of people they follow can enhance their perceived level of influence. 3. A higher perceived influence leads to increased engagement with shared content (i.e., increased likes and retweets).
Berman et al. (2019)	Adding empirical evidence to the literature on debates, political tweeting, and virality	User engagement (i.e., tweets and retweets)	UGC: 9.5 million tweets posted during and shortly after (from 60 minutes before to 90 minutes after) four key debates, obtained using selected hashtags	Tweets, retweets, linguistic style, and semantic characteristics (i.e., sentiment and topic)	1. During the course of the debates, the Twitter content gradually became disconnected from the live event. 2. The factors that contributed to the success of tweets during the debates were not the same as those that were observed to be successful after the debates.
Hennig-Thurau et al. (2015)	Micro-blogging word of mouth (MWOM)	Financial outcomes (i.e., movie adoption)	1. 4,045,350 tweets about 105 movies (obtained using a list of search terms) 2. These tweets are then classified into the following three groups via a self-built support vector machine (SVM) classifier: a. spam, non-English tweets, and tweets not related to the movie; b. movie-related tweets that contained no post-consumption quality assessment; c. review tweets	Tweets, subject matter (i.e., three groups), and sentiment	1. Negative MWOM reviews can impact the initial adoption of a movie. 2. Additionally, the impact of negative MWOM reviews is greater than that of positive MWOM reviews, indicating a bias towards negativity in MWOM.

Toubia and Stephen (2013)	Intrinsic vs image-related utility on Twitter	User engagement (i.e., tweets)	UGC: daily data collected for 2,493 non-commercial Twitter users accounts (randomly selected from a database of 3 million user accounts) for 160 days. 1,355 were further classified as active user accounts. 100 users were randomly chosen from this set of 1,355 active users as the treatment group.	Follows and tweets	1. Non-commercial Twitter users are driven to share content by two primary forms of utility: intrinsic utility and image-related utility. 2. Image-related utility is larger for most users.

An in-depth analysis of these dimensions elucidates the intricate dynamics of Twitter data and their potential utility for predicting tourism demand. The most prominent dimensions explored, as highlighted by Colicev (2019), include “volume” and “valence”. The dimension of “volume” encapsulates various measurable aspects, such as the quantity of tweets, retweets, replies (as explored by Lu et al. (2022) and Valsesia et al. (2020), and likes (Valsesia et al., 2020). On the other hand, “valence” relates to the sentiment of the tweet, as investigated by Hennig-Thurau et al. (2015) and Lacka et al. (2022).

Twitter data’s “volume” and “valence” have also frequently been examined in forecasting studies. For example, Behrendt and Schmidt (2018) investigated Twitter data’s attention effects (measured by the volume of tweets obtained from Bloomberg) and endorsement effects (measured by tweet sentiment obtained from Bloomberg) on retail customers’ investment decisions. Their results proved there was significant relevance of tweet volumes and sentiment for the return volatilities of individual-level stocks (i.e., IBM and WMT). Nevertheless, the results of an out-of-sample forecasting competition did not suggest significant improvements in forecasting accuracy for models incorporating tweet volumes and sentiment. Similarly, Audrino et al. (2020) discover the short-term effects of tweet volumes and sentiment on stock price volatility. However, again, no significant improvement in models’ forecasting accuracy, after incorporating tweet volumes and sentiment, was observed for either a single stock or the general stock market (based on 18 U.S. companies’ stocks). More recently, Twitter-generated variables have also been used to forecast newly emerging assets such as cryptocurrencies. For example, Shen et al. (2019) argue that investors’ attention measured by the volume of Bitcoin-related tweets could better capture the volatility of Bitcoin returns than could Google Trends data. Using a large dataset containing over 25 million tweets about the top 23 cryptocurrencies, Kraaijeveld and De Smedt (2020) also showed that increased investor attention (measured by volumes of tweets, retweets, and favourites) corresponded to greater price volatility. Building on Twitter data’s attention effects, their results further confirmed the relevance of Twitter sentiment (captured by the Loughran and McDonald sentiment classification dictionary) and cryptocurrency price volatility.

Apart from these primary dimensions, such as volume or valence, researchers have explored additional dimensions rooted in the textual content of tweets. These include “linguistic style” and “semantic characteristics”, as examined by Berman et al. (2019). Providing insights in the context of political debate, Berman et al. (2019) collected and analysed 9.5 million tweets posted during and shortly after four political debates. Their findings highlighted the evolving nature of the tweet content and the changing drivers of tweet success over time, with the narrative role of users shifting from that of narrator to interpreter, as they enriched their posts with detailed and emotionally compelling content.

Further, “text characteristics” such as the use of humour or the element of surprise have been considered by researchers such as Borah et al. (2020). Confirming that IMIs exhibiting specific qualities (e.g., humour, timeliness, or unexpectedness) have the ability to amplify firm value, Borah et al. (2020) also indicate that the inclusion of IMI messages will result in increased virality. More recently, the “subject matter” of a tweet, as discussed by Lacka et al. (2022), has also been shown to serve as a crucial dimension of Twitter signals. Lacka et al. (2022) analysed the subject matter of firm-generated tweets by S&P 500 IT firms and find that tweets reflecting either a consumer or a competitor orientation resulted in temporary price impacts. More recently, a unique dimension that has emerged as warranting further attention is the measure of users’ disengagement activities constructed by Villanova and Matherly (2023). This measure, deviating from traditional approaches (i.e., utilizing experimental settings to control for potential external factors) by recording real-world Twitter unfollowing activities, provides authentic and context-specific insights.

Although Twitter plays an important role in consumers’ lives, it is important to note that not all tweets are positive (Valsesia & Diehl, 2022). Prior literature acknowledges that there is a tendency towards negative tweets. In particular, Hennig-Thurau et al. (2015) argue that such “negative bias” significantly lowers customers’ initial acceptance of new experiential products. Examining 4,045,350 tweets about 105 movies (obtained using a list of search terms), Hennig-Thurau et al. (2015) proved that the effect of negative reviews dominates that of positive reviews, indicating a negative bias in the effect of Twitter reviews on customers’ initial acceptance of new experiential products (i.e., movies). Also demonstrating a negative bias in users’ tweets, Liaukonytė et al. (2022) shed extra light on the effect of Twitter boycott campaigns on sales. More recently, the negative effect of a brand’s socially unacceptable tweeting behaviour on Twitter was examined by Villanova and Matherly (2023). The authors find that a brand's socially unacceptable mentions on Twitter would lead to individuals' (individuals with stronger self-brand connections) unfollowing behaviour. Such research suggests that, although a negative bias exists on Twitter, negative tweets may also contain useful information.

As discussed in Section 2.3.2.3, Bigné et al. (2019) stands out as the singular study utilizing Twitter data to predict tourism demand, specifically hotel occupancy rates, during a specific holiday season. Beyond the field of tourism demand forecasting, Twitter data has garnered broader interest within tourism literature. As summarized in Table 2.5 below, the majority of studies have explored Twitter's geographical and textual data attributes. Studies such as Bordogna et al. (2016), Chua et al. (2016), and Salas-Olmedo et al. (2018) have primarily employed the geographical aspect of Twitter data to develop methodologies for mapping tourist movements. Alternatively, research utilizing the textual content of Twitter, as demonstrated by Philander and Zhong (2016) and Ainin et al. (2020), has been directed toward examining sentiment trends among individuals and within the tourism industry. Specifically, Lu and Zheng (2021) highlight the positive sentiment trends of tourists toward the cruise sector (i.e., more eager to travel) during the COVID-19 pandemic. From a macro perspective, Lee et al. (2021) leverage national-level Twitter happiness indices data to examine its influence on tourist entries and revenue, employing annual data for country-scale analysis. Additionally, at the destination level, Nicolau et al. (2020) analyse the impact of President Trump's Twitter engagements on the Dow Jones U.S. Travel and Tourism Index.

Collectively, studies such as those by Lee et al. (2021), which focus on tourist arrivals, are country-specific and have not provided predictive insights into tourist arrivals to attractions. Those that scrutinize individual-level Twitter data are limited to illustrating tourist flows (e.g., Bordogna et al. (2016), Chua et al. (2016), and Salas-Olmedo et al. (2018)) or sentiment trends (e.g., Philander and Zhong (2016), Ainin et al. (2020) and Lu and Zheng (2021)), without exploring the potential of this data in predicting tourist arrivals. Furthermore, compared to the wider management and marketing literature (refer to Table 2.4), the tourism field's application of Twitter data is relatively underdeveloped. The multifaceted nature of Twitter data in deducing individual tourists' intentions and behaviour remains underutilized, especially concerning long-term attraction-level trends. Consequently, the previously outlined objective of this research – ***to reveal the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level*** – has been further underscored.

Table 2.5 Selected tourism studies utilizing Twitter data

Study	Research focus	Attributes of Twitter data used	Data collection frequency/amount	Analytical techniques
Bordogna et al. (2016)	Developing a knowledge-based approach to identify, track, and cluster tourists' trips.	Geographic tags & timestamps	300,000 tweets in real-time	Clustering
Philander and Zhong (2016)	Demonstrating the use of Twitter sentiment analysis in developing cost-effective and real-time indicators of customer attitudes and perceptions within the hospitality industry.	Textual content & timestamps	34,315 tweets posted by 34 resorts within 5 weeks	Lexicon-based sentiment analysis
Chua et al. (2016)	Demonstrating the use of geotagged Twitter data in characterizing spatial, temporal, and demographic features of tourist flows at the regional level.	User profiles & geographic tags & timestamps	72,031 tweets posted by 3135 unique individuals	Flow analysis (i.e., trajectory mining & tourist identification) and visualization
Salas-Olmedo et al. (2018)	Demonstrating the use of geotagged Twitter data in analyzing the spatial behaviour of urban tourists	Geographic tags & timestamps	234,159 tweets posted by tourists visiting Madrid between 2012 and 2014	Density maps, OLS, spatial self-correlation and clustering
Ainin et al. (2020)	Exploring the trends in halal tourism by analyzing Twitter posts	Geographic tags & textual content & timestamps	85,259 tweets related to tourism in English and Bahasa Malaysia between 2008 and 2018	Lexicon-based sentiment analysis
Lee et al. (2021)	Exploring if and how the happiness indices of host countries, derived from Twitter data, influence tourism development (measured by tourist arrivals and tourism revenue).	Textual content & timestamps	Yearly frequent country happiness indices collected from 2008 to 2017 for 119 countries	/
Lu and Zheng (2021)	Understanding the public's sentiment trends to the cruise industry during the pandemic.	Hashtags & textual content & timestamps	53,546 tweets posted between 1 Feb and 18 Jun 2020	Lexicon-based sentiment analysis & LDA
Nicolau et al. (2020)	Analyzing the effect of President Trump's Twitter activities on the performance of the United States as a tourism destination (as reflected in the Dow Jones U.S. Travel and Tourism Index).	Textual content & timestamps	170 tweets posted by Donald J. Trump (after cleaning)	Lexicon-based sentiment analysis & judgemental topic classification

2.6 Theoretical underpinnings

2.6.1 Actor engagement theory

Actor engagement theory is characterized by an actor's willingness to engage and their active involvement in a collaborative process of investing resources within a service ecosystem (Storbacka et al., 2016). It goes beyond the traditional dyadic model of communication by embracing a network perspective, which highlights the dynamic interactions among various stakeholders (actors) within the service ecosystem (Brodie et al., 2019).

Yadav and Pavlou (2014) observe that various actors contribute to the service ecosystem not just by being present but through active engagement as content creators, critics, or disseminators of information. Within this interrelated network, actors – including firms, customers, and bystanders – play unique roles (Sheng et al., 2019). Firms, in particular, assume a variety of functions: monitoring online conversations and trends, guiding discussions to maintain community norms, facilitating interactions among different users, and directly participating by engaging in conversations, sharing content, and responding to customer queries (Godes et al., 2005). Notably, social media has evolved into an influential service ecosystem that accommodates a diverse array of actors (Alaimo et al., 2020). This transformation positions actor engagement theory as a valuable framework for examining the dynamic interactions and engagements occurring within this environment. Shedding light on the tourism and hospitality sector, Sheng et al. (2019) underscore the impact of firm interventions, such as managerial responses, on the network. They observe that these actions influence not only the customers currently engaging in reviews but also bystanders observing these interactions in the digital realm. Importantly, this continuous, interactive process markedly affects the overall engagement valence, namely customer ratings, over time.

Overall, actor engagement theory plays an important role in elucidating the roles of various actors within the service ecosystem, offering valuable insights for research focused on the dynamic and interactive communications within the Twittersphere. For the current research, understanding the network structure of different actor groups - such as tourists, bystanders (potential tourists), and attractions - within this ecosystem is crucial. This identification enables the mapping of complex networks among these actors, laying the groundwork for insights into tourists' visiting intentions and factors explaining their behavioural changes (Bi et al., 2021;

Hu et al., 2022), which are essential for forecasting tourist arrivals. Nonetheless, as suggested by Brodie et al. (2019), the present development of actor engagement theory does not adequately illuminate the key properties of actor engagement networks, particularly the extent to which relationships within the network affect its actors. In other words, it overlooks the strength (or credibility) of the signals within actors' interactions, a shortfall that diminishes its utility in generating predictors from the interaction traces of Twitter users to precisely forecast tourism demand. Addressing this gap, the joint application of actor engagement theory and signalling theory, as recommended by Sheng et al. (2019) and Tumasjan et al. (2021), positions each actor as both a sender and receiver of signals of varying strengths. This approach aids in a fuller understanding of actors' online engagement and, more importantly, how these interaction traces will impact actors' behaviour, such as visiting activities.

2.6.2 Signalling theory

2.6.2.1 Fundamentals of signalling theory: signaller, receiver and signal

Signalling theory enriches the framework of actor engagement theory by prioritizing the credibility of signals. Originally proposed by Spence (1973), signalling theory has been widely applied across various selection scenarios in numerous fields of the business and management literature (Connelly et al., 2011). The theory suggests that signallers transmit signals about their traits, attributes, or intentions to receivers, who then use these signals to make informed decisions (BliegeBird & Smith, 2005). The informative signal is transmitted from the signaller to the receiver via a signalling environment, such as social media. As acknowledged by Lacka et al. (2022), tweets serve as information signals. Further, Kirmani and Rao (2000) contend that signalling is most pertinent to products whose quality is unascertainable prior to purchase, for example, experiential goods such as tourism products. Therefore, the principles of signalling theory align well with the research context of this thesis.

As outlined by Connelly et al. (2011), signalling theory is composed of three primary elements: the signaller, the receiver, and the signal. Additionally, it encompasses two ancillary mechanisms, potential feedback to the signaller and the signalling environment, as depicted in Figure 2.1. In this subsection, the three primary elements will be introduced. The signalling environment and the feedback mechanism will be discussed in Section 2.6.3, where we also introduce the many-to-many communication framework as a tool for capturing those two ancillary mechanisms.

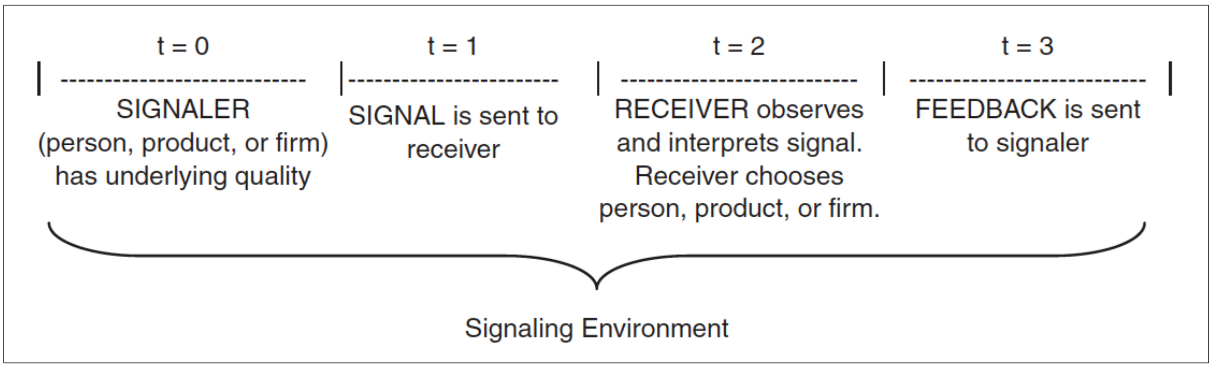

Figure 2.1 A timeline depicting the primary elements of signalling theory

Source: Connelly et al. (2011, pp. 44)

Signaller. Central to the signalling theory is the idea that the *signaller*, or sender, possesses information about an individual (e.g., Spence (1973) and Valsesia and Diehl (2022)), a product (e.g., Kirmani and Rao (2000)), an organization (e.g., Drover et al. (2018), Ross (1977), and Tumasjan et al. (2021)), or any other entity (e.g., DMOs and attractions). This information could benefit the receiver but is not readily available to them (Connelly et al., 2011). For example, a DMO or an attraction holds information that could be advantageous to tourists, such as details related to events and admission (Bigné et al., 2019).

This information offers the signaller a beneficial perspective regarding the underlying quality of the individual, product, or organization in question. Given this advantage, the signaller has to decide whether, how, and via which communication channel to convey this information. Conversely, the receiver, interested in acquiring such information, must decide how to interpret the signal and respond accordingly (Connelly et al., 2011). For example, a tourist who has visited a destination or attraction may opt to use social media such as Twitter as a channel to share their visiting experience (Hu et al., 2022; Ibrahim & Wang, 2019b), while prospective tourists may be interested in learning from that experience. Concurrently, prospective tourists have the task of interpreting this information and deciding whether to act upon it, such as making a decision to visit the destination or attraction.

Receiver. The recipient of the signal (i.e., *receiver*) represents the second component in the signalling process. According to the model, receivers are typically outsiders who are devoid of certain information but interested in acquiring it (Connelly et al., 2011). Empirical studies applying signalling theory as their theoretical underpinning often consider investors (Tumasjan et al., 2021), consumers (Kirmani & Rao, 2000), and visitors (Gao & Su, 2021) as receivers. In the field of marketing, customers are typically seen as receivers (Basuroy et al., 2006; Kirmani & Rao, 2000). A crucial aspect of the signalling is that the receivers stand to benefit (either directly or jointly with the signaller) from decisions (e.g., to invest, purchase, or visit) made based on the information gleaned from these signals. For example, a prospective visitor may be keen to learn about the quality of services at an attraction (Poria et al., 2013).

Signal. In the marketing area, advertising is regarded as a way to *signal* to customers regarding potential product information (quality-related), reduce consumer uncertainty, and foster purchase decisions (Kirmani & Rao, 2000; Nian & Sundararajan, 2022). Examples of marketing signals include quality certifications (Dewally & Ederington, 2006; Gao et al., 2010) and corporate social responsibility (CSR) (Saxton et al., 2019), which enhance buyers' trust and contribute to higher sales and consumer resonance. Tourism products are highly experiential; their quality cannot be perceived unless they are consumed by tourists (Sugathan & Ranjan, 2019). To address this challenge, tourism destinations and attractions often rely on third-party accreditations, such as quality ratings of tourist attractions (Gao et al., 2022) and destinations (Gao & Su, 2021), World Heritage Site (WHS) designation (Poria et al., 2013), and eco-labels (Capacci et al., 2015), as quality signals influence tourists' perceptions and purchase intentions in the increasingly competitive tourism market. Shedding light on tourism attractions, Gao et al. (2022) discover that attractions should invest more in self-initiated marketing communications to signal their quality, for example, travellers' reviews or social media posts.

Tourism organizations including DMOs and attractions increasingly rely, as a signalling strategy, on social media as a way to share information, with the hope of encouraging receivers (i.e., tourists) to act upon it (i.e., visit) (Mumi et al., 2019). In particular, the widespread use of social media platforms like Twitter has made experiential purchases, such as visits to tourist attractions, more visible (Valsesia & Diehl, 2022) and trustworthy (Hennig-Thurau et al., 2015).

As a result, the information shared through tweets can serve as “informative signals” due to its inherent characteristics and the nature of the Twitter platform, as detailed in Section 2.5.

To begin with, the brevity of tweets, which is a limitation imposed by Twitter itself, necessitates the creation of succinct and meaningful content (Lovejoy et al., 2012). This encourages users to concentrate their messages, encapsulating the most critical and informative elements of their thoughts or experiences. Moreover, Twitter’s public and far-reaching nature allows for the rapid and widespread dissemination of information (Valsesia & Diehl, 2022). When a user shares a tweet, it becomes instantly accessible to all their followers and, if the account is public, to any Twitter user around the globe (Toubia & Stephen, 2013). This feature amplifies the tweet’s potential impact as an “informative signal”, extending its reach beyond a person’s immediate social circle. Furthermore, various features associated with tweets add layers of context that can enhance their informative value. For example, the use of hashtags can signal the main topic or sentiment of a tweet (Kumar et al., 2022), while the act of retweeting can indicate endorsement or agreement with the original message (Saxton et al., 2019). These elements can provide additional clues to the receiver about the signal’s relevance and credibility. Finally, it is important to note that the real-time nature of Twitter allows tweets to serve as timely signals (Cano-Marin et al., 2023). This is particularly crucial in rapidly changing contexts, such as tourism demand forecasting, where up-to-date information can significantly influence decision-making processes (Peeters et al., 2018).

Overall, while tweets may be considered “informative signals”, they have not been fully explored in the field of tourism demand forecasting. To address this gap, the first research objective is ***to assess the role of Twitter data in tourism demand forecasting at the attraction level***.

2.6.2.2 Social-media-mediated signalling environment

Within the context of signalling theory, the signalling environment constitutes the conditions and circumstances that affect the transmission of signals from signallers to receivers and reception by the latter (Connelly et al., 2011). It holds an essential role in the signal transmission process, significantly impacting the creation, propagation, reception, and interpretation of signals (Gulati & Higgins, 2003). These effects contribute to the extent to which signalling alleviates information asymmetry (Park & Patel, 2015).

The emergence of social media platforms, notably Twitter, has incited a significant paradigm shift within the signalling environment, thereby garnering substantial academic interest (Harris et al., 2021). These platforms have redefined the contours of the signalling terrain, facilitating enduring and pervasive signal exposure that transcends the constraints inherent in conventional offline environments (Valsesia & Diehl, 2022). This transformation is characterized by its provision of widespread, dynamic channels for information transmission, thereby amplifying the proficiency and expediency of signalling processes (Saxton et al., 2019; Tumasjan et al., 2021). Consequently, it has unveiled a myriad of innovative prospects in both scholarly research and practical applications.

For tourist attractions, social media like Twitter provides a platform to broadcast signals about their offerings, events, and any other relevant information aimed at both potential (Bigné et al., 2019) and returning tourists (Sheng et al., 2019). For tourists, Twitter acts as a platform to communicate their experiences, opinions, and sentiments regarding these attractions (Sotiriadis & Van Zyl, 2013). Tourists can share positive experiences and satisfaction or voice their criticisms and complaints (Hu et al., 2022; Ibrahim & Wang, 2019b), thereby providing feedback to the attractions and information to other potential visitors (Alkhammash et al., 2019).

Moreover, Twitter's public, transparent nature allows for the tracking and examination of these interactions (Toubia & Stephen, 2013). Academics and industry professionals can analyse these signals and countersignals (refer to the discussion in Section 2.6.2.2) to gain insights into various aspects of tourist behaviour, attraction management strategies, and overall trends in the tourism industry. Specifically, Twitter facilitates the embedding of semantic layers within signals (Berman et al., 2019), allowing tourists to express their awareness or attention vividly. As a result, tourist attractions can discern whether the signals from tourists emerge from positive or negative contexts (Hu et al., 2022). This bifurcated understanding paves the way for the targeted utilization of these signals. For instance, positive signals can be leveraged for tourism demand forecasting (Dellarocas et al., 2007), while negative signals can be used to recalibrate the quality of their own communications (Liaukonytė et al., 2022; Villanova & Matherly, 2023). In this sense, Twitter is not only a conduit for communication between attractions and tourists but also a rich data source that reflects the dynamics of the tourism signalling environment.

2.6.2.3 The feedback mechanism

The capacity of social media as a signalling environment to facilitate countersignalling, or the feedback process, is notably significant (Harris et al., 2021). Numerous studies accentuate the importance of receivers' feedback to signallers about signal effectiveness (Saxton et al., 2019). To enhance the efficacy of signalling, receivers can respond with countersignals. The foundational idea here is that information asymmetry works in both directions: while receivers are in pursuit of information about signallers, signallers also need information from receivers. Such information enables signallers to discern which signals are deemed most credible, which ones attract the most attention from receivers, and how receivers are interpreting the signals (Connelly et al., 2011). Signallers who are attuned to such countersignals can refine their future signals in a timely way, thereby improving their trustworthiness and efficiency (Taj, 2016).

In a social-media-mediated signalling environment, the feedback process becomes public and directly connected to specific messages, as manifested in the form of likes, comments, and shares on social media platforms (Saxton & Waters, 2014). This visibility offers signallers the opportunity to promptly adjust their signals (Taj, 2016), and also presents researchers with the means to empirically investigate the effectiveness of signalling strategies via countersignals (e.g., public's retweeting behaviour: Saxton et al. (2019)). However, despite the significant role of the bidirectional signalling process in social media such as Twitter, there exists a paucity of examinations of communication flows in the current literature (Connelly et al., 2011; Taj, 2016). Therefore, in the following subsection, we employ the many-to-many communication framework as a method for discerning and understanding signals and countersignals within the signalling environment mediated by Twitter.

2.6.2.4 The many-to-many communication framework

The many-to-many communication framework characterizes communication flows on social media platforms like Twitter as a complex network of interconnected interactions, as depicted in Figure 2.2. This approach shifts away from the traditional paradigm of isolated one-to-one exchanges or one-to-many information broadcasts, instead accentuating the dynamic intermingling of numerous signals from a broad spectrum of signallers and receivers. The framework importantly incorporates a frequently overlooked feedback mechanism within the signalling process, indicated by the bidirectional arrow in black (see Figure 2.2). This framework accommodates the various constituents of the signalling process: the signallers, receivers, informative signals, and feedback mechanisms outlined above. It also enables the identification of distinct communication flows; that is, it sheds light on who communicates with whom. The complex interweaving of these communication flows within Twitter's signalling environment, therefore, contributes in-depth insights into the interactions among tourist attractions, tourists, and the general audience.

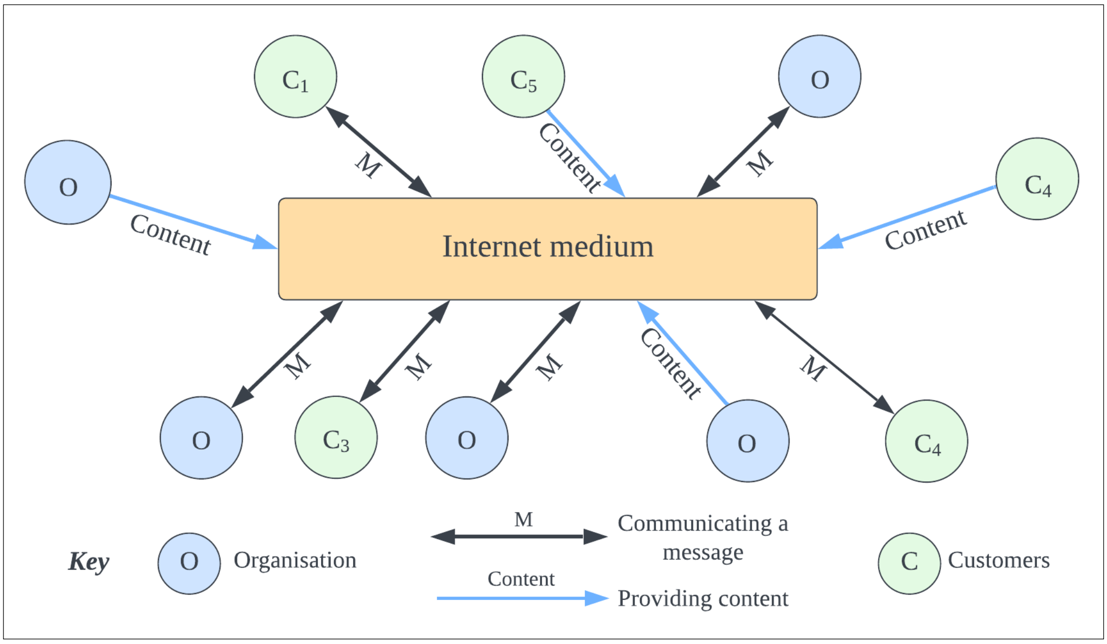

Figure 2.2 Many-to-many communication via internet medium

Source: (Chaffey & Ellis-Chadwick, 2019, pp. 426)

Referring to Figure 2.2, these communication flows can be categorized into three discrete pathways in the tourism context:

(1) *Direct communication with attractions*: This communication flow involves direct interaction between Twitter users, often potential tourists, and the attractions themselves. Typically, such exchanges include tourists seeking specific information or

providing feedback, while attractions address inquiries, comments, or grievances (Alkhammash et al., 2019). This dynamic fosters an interactive sphere of communication, serving as an immediate conduit between tourists and attractions, and thus enhancing customer service and engagement. Research such as that by Bigné et al. (2019) exemplifies the analysis of this communication stream on Twitter, while Önder et al. (2019) have considered it within the Facebook context.

(2) *Peer-to-peer conversations*: Here, Twitter users converse about an attraction, sharing their thoughts or experiences (Sotiriadis & Van Zyl, 2013). This communication flow involves Twitter users discussing attractions among themselves, sharing personal experiences, opinions, or even advice with fellow users. These discussions offer first-hand accounts and reviews of attractions, serving as valuable user-generated content for prospective visitors. The peer-to-peer sharing of experiences and thoughts amplifies the social aspect of Twitter, extending its role beyond that of a simple information-sharing platform to that of a virtual community of shared experiences (Ibrahim & Wang, 2019b). Due to the complexity of monitoring this communication stream (it does not necessarily involve the participation of the attraction), it remains relatively unexplored in existing tourism studies.

(3) *Communication from attractions to the broader audience*: This communication stream emanates from the attractions, who target a broad audience encompassing potential visitors, tourists, industry stakeholders, and the general public. Here, attractions leverage Twitter as an informational and promotional tool, disseminating updates about events, attractions, services, or news. These communications are aimed not only at engaging the existing audience but also at captivating potential tourists, thereby enhancing the attraction's visibility and appeal. This communication stream forms the primary focus of much existing research, including that of Bigné et al. (2019).

While the literature calls for the consideration of social media data generated by a wide array of tourism stakeholders (Bigné et al., 2019), current tourism demand forecasting research does not fully consider these communication flows. To address this gap, the second objective of this research is to ***examine the advantages of deciphering embedded communication flows on Twitter for generating improved tourism demand forecasts at the attraction level***.

2.6.2.5 The "noise" challenge presented by the social-media-mediated signalling environment

Considering the inherent characteristics of the social-media-mediated signalling environment, signals derived from it often encapsulate substantial "noise". In the process of social-media-mediated signalling, the transmission of multiple signals among signallers and receivers frequently contributes to the amplification of noise in the data (Xie et al., 2020). As Belch and Belch (2007) explain, during the communication process, messages can experience distortion or interference – known as "noise" – due to external factors. This interference presents a notable challenge to the identification of relevant "signals" amidst the noise within a group of tweets (O'Neal et al., 2018). For example, in the context of this research, while positive tweets can indeed encourage visits and hence aid tourism demand forecasting (Dellarocas et al., 2007), negative micro-blogging posts can be considered noise since they do not encourage visits (Su et al., 2022).

The literature notes that signallers transmit both positive and negative information, and they must decide whether to communicate this information to receivers. In the context of forecasting tourism demand, positive information can increase customers' purchase intentions (Dellarocas et al., 2007), but the interest an individual expresses in a destination or attraction on the internet is not always positive (Hu et al., 2022). Although signalling theory focuses primarily on the deliberate communication of positive information in an effort to convey positive attributes (Connelly et al., 2011), scholars have also examined actions taken by signallers that communicate negative information (Drover et al., 2018). For instance, accepting corporate sponsorship is generally considered a negative signal for non-profits because it reduces prospective donors' willingness to support them (Bennett et al., 2013). Specifically, low-fit sponsorship would produce a negative signal, affecting a non-profit's brand identity, brand meaning, brand response, and brand relationships (Becker-Olsen & Hill, 2006). Prior research demonstrates that corporate sponsorships from companies displaying non-altruistic intentions (e.g., fossil fuel sponsors: Adams (2019)) could be considered negative signals that would affect the authenticity of museums and ultimately reduce visiting intentions (Biraglia & Gerrath, 2020; Biraglia et al., 2018). As suggested by Hu et al. (2022), if a tourism destination is facing unfavourable conditions but tourist interest is still rising, relying solely on attention-based or internet traffic data (i.e., search query and web traffic data) for forecasting purposes may lead to biased conclusions. The signalling theory literature suggests that, as noise intensifies within the signalling environment, the efficacy of the signalling process may decrease, posing challenges for receivers in terms of discerning and interpreting quality signals sent by the signaller (Connelly et al., 2011).

The above arguments emphasize the importance of noise detection and elimination in social media data-based forecasting practice (Ibrahim & Wang, 2019a, 2019b). On the one hand, researchers have adopted various strategies for extracting pertinent elements of interest. For example, they have used specific Twitter handles to collect tweets between certain entities (Bigné et al., 2019; Borah et al., 2020) or utilized lists of hashtags or keywords to obtain tweets on specific topics (Liaukonytė et al., 2022). On the other hand, the social-media-mining literature underscores that effective data mining requires comprehensive data pre-processing and noise removal, encapsulated by the principle of "garbage in, garbage out".

In light of the substantial role "noise" plays in social media data research, Zafarani et al. (2014) emphasized the cautious consideration of the "noise removal fallacy" prior to the mining of social media data. The term "noise removal fallacy" is multifaceted. First, it accentuates the complexity and relativity involved in defining "noise", as this largely depends on the specific task in question (Zafarani et al., 2014). It is advised that the objective when analysing social media data should be to obtain relevant elements (the "signal"), while reducing and eliminating irrelevant components (the "noise") (Hollingshead et al., 2022). This suggests that the identification of "noise" is subjective, depending on the unique research context, and thus advocates a context-specific identification of noise. In the context of forecasting tourism demand, it has been widely recognized in the literature that digital traces reflecting tourism stakeholders' marketing efforts (Bigné et al., 2019), tourists' attention to such efforts (Önder et al., 2019), or tourists' visiting intention (Volchek et al., 2019) can serve as informative signals for predicting tourism demand. Consequently, we define tweets that ***neither stimulate visits nor signal tourists' intentions to visit*** as noise in this thesis. On the flip side, the "noise removal fallacy" cautions that indiscriminate noise elimination might compound problems by potentially deleting valuable information (Zafarani et al., 2014). For example, even though often classified as random noise, trolling content can provide valuable insights into the relational context of collected data (Nevin et al., 2022). Despite this, explorations into the signalling value of this "noise" have been limited.

To bridge the identified knowledge gap, this thesis's last research objective is to devise an effective strategy for noise detection. In achieving this, we aim to ***evaluate the value of commonly excluded Twitter "noise" (i.e., - tweets that neither stimulate visits nor signal tourists' intentions to visit) for supplementing tourism demand forecasts at the attraction level.***

2.7 Research aim and objectives

In summary, with the specific focus on Twitter, the aim of this study is to ***reveal the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level***. Thereby, the hope is to enrich the literature on tourism demand forecasting. Building upon the identified research gaps, the following research objectives are formulated:

(1) Assess the role of Twitter data in tourism demand forecasting at the attraction level.
(2) Examine the advantages of deciphering embedded communication flows on Twitter for generating improved tourism demand forecasts at the attraction level.
(3) Evaluate the value of commonly excluded Twitter "noise" (i.e., - tweets that neither stimulate visits nor signal tourists' intentions to visit) for supplementing tourism demand forecasts at the attraction level.

To thoroughly address these research objectives, we have devised a sequence of three studies, each probing a distinct facet of the overarching research aim. Chapter 3 will offer an exhaustive elucidation of the methodological approaches adopted for each of these studies.

Chapter 4, based on signalling theory and utilizing the British Museum as a case study, will initially investigate the predictive accuracy of different dimensions of raw Twitter data (such as tweet volume, likes, retweets, and replies) in forecasting tourist arrivals at individual attractions.

Chapter 5, underpinned by the framework of many-to-many communication, will generate an enhanced understanding of the information dissemination process within the Twittersphere. After clustering tweets based on their intrinsic communication flows as indicated by text mining, a diverse range of forecasting models will be used to assess the signalling value of the variables derived from each cluster. Of note, tweets that fail to reflect tourists' attention and visiting interests will be identified as “noise” and systematically excluded from the dataset prior to variable construction. The aim of Chapter 5 is to emphasize the importance of mapping Twitter communication flows and to identify signals that make the greatest contribution to generating improved forecasts of tourist arrivals to the British Museum.

Chapter 6 will re-evaluate the tweets previously classified as “noise”, in Chapter 5, to ascertain their signalling value for tourist arrivals. To fulfil the research needs, we will propose a three-stage framework involving sentiment analysis, judgemental screening, topic modelling, and in-context keyword searching for precisely identifying the “noise” consisting of tweets related to various boycott campaigns. The variables constructed from these boycott-related tweets will then be evaluated for their dynamic relationship with tourist arrivals. In this way, Chapter 6’s aim is to explore the value of the “noise” within the Twittersphere, which is often overlooked in the existing literature, for deciphering fluctuations in tourism demand.

Chapter 3: Research methodology

3.1 Overview

In academic research, the term "research methodology" refers to the systematic, principles-based approach used to guide the investigation of a research problem. This includes the research philosophy, research approach, and research method, as well as the selection of specific methods for data collection and analysis, and ethical considerations. The purpose of the research methodology is to provide a framework for addressing the research question systematically and transparently. In this chapter, the methodology used in the thesis is thoroughly explained. The chapter begins with an introduction to the research philosophy and research approach, including the reasoning behind the adoption of pragmatism and abductive reasoning. The focus then shifts to the adoption of quantitative methods and the supplementary role of qualitative methods in the thesis. Subsequently, the research context, data collection strategy, and ethical considerations are thoroughly discussed, followed by an illustration of the methodological choices made for each study conducted in this research. Last, limitations emanating from the chosen research methods and avenues for future research designs are discussed.

3.2 Research philosophy and research approach

According to Saunders et al. (2019), research philosophy is characterized by "the development of new knowledge and the nature of that knowledge". The research philosophy relates to a set of beliefs and assumptions that fundamentally underpin and guide a study. Therefore, it is essential to have a strong comprehension of the research philosophy in order to select an appropriate study design. In this section, the research philosophy of this thesis is drawn from three interlinked aspects: ontology, epistemology, and axiology. Ontology is the branch of philosophy that deals with the nature of reality and existence. It is concerned with questions such as what is the nature of the world and what exists in the world (Smith, 2012). Epistemology is the branch of philosophy that concerns knowledge and belief. It addresses questions such as how we know, what we know, and what is the nature of belief (Audi, 2010). Axiology is the branch of philosophy that refers to values and ethics within the research process. It is concerned with questions such as what are the most important values and how should we make moral decisions (Hartman, 2011). In an academic context, these three branches of philosophy are often studied together as they are all interconnected and can provide insight into how we understand and interact with the world around us (Saunders et al., 2019). Five research philosophies can be identified based on differences in research epistemology, ontology, and axiology: positivism, critical realism, interpretivism, postmodernism, and pragmatism (refer to Table 3.1).

Table 3.1 Comparison of five research philosophy positions in business and management research

Ontology (nature of reality or being)	Epistemology (what constitutes acceptable knowledge)	Axiology (role of values)	Typical methods
	Positivism		
• **Real, external, independent** • **One true reality (universalism)** • **Granular (things)** • **Ordered**	• Scientific method • Observable and measurable facts • Law-like generalizations • Numbers • Causal explanation and prediction as contribution	• Value-free research • Researcher is detached, neutral, and independent of what is researched • Researcher maintains objective stance	• Typically deductive highly structured, large samples, measurement, typically quantitative methods of analysis, but a range of data can be analysed
	Critical realism		
• **Stratified/layered (the empirical, the actual, and the real)** • **External, independent** • **Intransient** • **Objective structures** • **Causal mechanisms**	• Epistemological relativism • Knowledge historically situated and transient • Facts are social constructions • Historical causal explanation as contribution	• Value-laden research • Researcher acknowledges bias by world views, cultural experience, and upbringing • Researcher tries to minimize bias and errors • Researcher is as objective as possible	• Retroductive, in-depth historically situated analysis of pre-existing structures and emerging agency • Range of methods and data types to fit subject matter
	Interpretivism		
• **Complex, rich** • **Socially constructed through culture and language** • **Multiple meanings, interpretations, realities**	• Theories and concepts too simplistic • Focus on narratives, stories, perceptions and interpretations • New understandings and	• Value-bound research • Researchers are part of what is researched, subjective • Researcher interpretations key to contribution	• Typically inductive • Small samples, in-depth investigations, qualitative methods of analysis, but a range of data can be interpreted

• **Flux of processes, experiences, practices**	worldviews as contribution	• Researcher reflexive	
	Postmodernism		
• **Nominal** • **Complex, rich** • **Socially constructed through power relations** • **Some meanings, interpretations, realities are dominated and silenced by others** • **Flux of processes, experiences, practices**	• What counts as "truth" and "knowledge" is decided by dominant ideologies • Focus on absences, silences and oppressed/repressed meanings, interpretations, and voices • Exposure of power relations and challenge of dominant views as contribution	• Value-constituted research • Researcher and research embedded in power relations • Some research narratives are repressed and silenced at the expense of others • Researcher radically reflexive	• Typically deconstructive – reading texts and realities against themselves • In-depth investigations of anomalies, silences, and absences • Range of data types, typically qualitative methods of analysis
	Pragmatism		
• **Complex, rich, external** • **"Reality" is the practical consequence of ideas** • **Flux of processes, experiences, and practices**	• Practical meaning of knowledge in specific contexts • "True" theories and knowledge are those that enable successful action • Focus on problems, practices, and relevance • Problem solving and informed future practice as contribution	• Value-driven research • Research initiated and sustained by researcher's doubts and beliefs • Researcher reflexive	• Following research problem and research question • Range of methods: mixed, multiple, qualitative, quantitative, action • Research emphasis on practical solutions and outcomes

Source: adopted from Saunders et al. (2019, pp.144-145)

Positivist researchers believe that only observable phenomena and/or observable social reality can provide reliable data, and they tend to use highly structured methodologies in their research. This approach often involves the use of quantitative methods and seeks to generate “law-like” generalizations that can be applied to the broader population being studied (Crotty, 1998). Critical realism, another research philosophy, also assumes that objects have an independent existence beyond the human mind. Critical realism emphasizes the importance of understanding the underlying structures of reality to explain and make sense of observed events (Fleetwood, 2005). It often involves conducting historical analyses of societal and organizational structures, using a range of research methods, to understand how these structures shape and are shaped by the world around us. Critical realists seek to uncover the underlying causes and mechanisms that give rise to the phenomena we observe, to gain a deeper understanding of the world (Reed, 2009). However, interpretivism researchers critique the positivist and realist approaches for not allowing for the exploration of rich insights into the phenomena being investigated. Instead, interpretivism focuses on subjective evaluation and understanding the motivations, actions, and contexts behind cases and situations (Crotty, 1998). Postmodernist researchers go even further than interpretivists in their critique of positivism and objectivism, attributing even more importance to the role of language and power relations. Postmodernists critique dominant modes of thought and seek to amplify the voices of alternative worldviews that have been silenced by hegemonic viewpoints. They believe that any sense of order is provisional and lacking foundation, and can only be brought about through our language, with its categories and classifications (Reed, 2009). Pragmatism, on the other hand, shifts the focus from the research findings to the research questions themselves and seeks to identify the practical consequences of the phenomena being studied. Pragmatism allows for methodological flexibility and encourages the use of both quantitative and qualitative methods (Kelemen & Rumens, 2008). With the methodological freedom offered by a pragmatic approach, it is believed that we will be able to fully explore Twitter data’s potential for identifying tourists’ behavioural patterns and further inform the theory and practice with our research findings.

Directly connected to the research philosophy, the research approach reflects how theory and/or knowledge is developed. Deduction, induction, and abduction are three different research approaches, differentiated by their features of logic, generalizability, and use of data and theory (see Table 3.2) (Saunders et al., 2019). Deductive reasoning involves constructing a theory or hypothesis and designing a research strategy to test it, while inductive reasoning involves collecting data and developing a theory based on the analysis of those data. The deductive research approach has been seen as the logical progression of positivism. On the other hand, the inductive approach, characterized by its connection to the humanities and focus on subjective interpretations, aligns with interpretivism (Saunders et al., 2019). Combining both deductive and inductive reasoning (Suddaby, 2006), an abductive approach involves using data to explore a phenomenon, identify patterns, and generate or modify a theory (Van Maanen et al., 2007). Due to the flexibility of the abductive approach, it can be used by researchers within several different research philosophies (Saunders et al., 2019). The selection of the research approach is guided by practical considerations, such as the availability of literature with which to form a theoretical framework and the nature of the research topic (Creswell, 2002). Deductive reasoning is recommended if the theoretical framework and hypothesis can be derived from existing literature. If the literature is lacking, for instance in the case of novel phenomena, an inductive approach would be recommended. Otherwise, if there is a significant amount of literature in one research context but less in others, an abductive approach will be more suitable. As is discussed in the literature review, there are a handful of studies exploring Twitter data's capabilities when it comes to forecasting real-world outcomes such as asset prices (Behrendt & Schmidt, 2018), election results (Huberty, 2015), and public sentiment (Ridhwan & Hargreaves, 2021). However, tourism demand forecasting research using Twitter data is scarce (Li et al., 2021). Consequently, it is proposed that we use an abductive research approach as the foundation for this research, based on which we will be able to investigate Twitter data's role in reflecting and affecting tourists' decision-making process, and further generate practical implications for attractions' social media marketing and capacity management strategies.

Table 3.2 Deduction, induction, and abduction: from reason to research

	Deduction	**Induction**	**Abduction**
Logic	In a deductive inference, when the premises are true, the conclusion must also be true	In an inductive inference, known premises are used to generate untested conclusions	In an abductive inference, known premises are used to generate testable conclusions
Generalizability	Generalizing from the general to the specific	Generalizing from the specific to the general	Generalizing from the interactions between the specific and the general
Use of data	Data collection is used to evaluate propositions or hypotheses related to an existing theory	Data collection is used to explore a phenomenon, identify themes and patterns, and create a conceptual framework	Data collection is used to explore a phenomenon, identify themes and patterns, locate them in a conceptual framework, and test this through subsequent data collection
Theory	Theory falsification or verification	Theory generating and building	Theory generation or modification; incorporating existing theory where appropriate, to build new theory or modify existing theory

Source: adopted from Saunders et al. (2019, pp.153)

3.3 Research method

The research method refers to the principles that guide the selection of the methods used in research (Huff, 2008). Within methodology, there are two primary approaches: quantitative and qualitative research. These methods are commonly used in empirical social science research and differ in the forms, media, and means of communication utilized. Quantitative research relies on numerical and quantifiable data (Saunders et al., 2019) while qualitative research utilizes unstructured and textual data (Bansal et al., 2018). A comparison of these two approaches is presented in Table 3.3, which highlights their different research goals, specific methods, and criticisms they have received respectively.

Table 3.3 The dichotomy between the qualitative and quantitative research methods

	Qualitative	Quantitative
Typical goals	• Explanation (how and why things happen) • Detail (providing detailed analysis about the phenomenon and causal relations) • Exploration (seeking unanticipated consequences and contingencies) • Theorization (theorizing based on practical experience to abstract idea frameworks)	• Inference (from sample to population) • Prediction (from past to future) • Hypothesis testing (testing and increasing confidence in a theoretical explanation) • Generalization (expanding the range of a theoretic explanation)
Methods	• Narrative research • Phenomenological research • Grounded theory research • Ethnographic research • Case study research	• Statistical analysis • Linear or non-linear regression • Principal component and factor analysis • Analysis of variance • Structural equation modelling • Game theory
Critique	• Subjectivity (unscientific, only exploratory, or entirely personal and full of bias)	• Oversimplification

Source: (Huff , 2008, pp.184-185)

The distinction between quantitative and qualitative methodologies can be traced back to their divergent philosophical positions. Quantitative research is typically associated with an objectivist ontology and positivist epistemology, and employs natural science research methods such as statistical and regression analyses to quantitatively describe features of a population, predict the future, test hypotheses, and demonstrate cause and effect relationships (Nuti et al., 2014). In the scenario of forecasting studies, the quantitative method is suitable for developing predictive models that explain behaviours and processes, and can serve as grounds for anticipating future results (Bertrand & Fransoo, 2002). In contrast, qualitative research aligns with a constructivist or critical realist ontology and interpretivist epistemology, and focuses on understanding social events by exploring the explicit and implicit mechanisms underlying these events, examining contingencies, developing hypotheses, and theorizing about practical experiences in order to generate abstract conceptual frameworks. To achieve these objectives, qualitative researchers often use narrative analysis, ethnographic research, and qualitative case study methods. In the field of management studies, the quantitative methodology has traditionally held a dominant position. However, over the past two decades, the advantages of the qualitative methodology have increasingly been recognized (Bluhm et al., 2011). The contribution of qualitative research in terms of advancing the theoretical development of the management and organizational literature has been widely acknowledged (Bansal et al., 2018).

In recent years, some researchers have advocated for the use of a mixed method that combines both quantitative and qualitative methods (Tashakkori & Creswell, 2007). They argue that social science research should follow a "pragmatist" approach, which prioritizes the research question and avoids debates about research ontology and epistemology (Parvaiz et al., 2016). While acknowledging that both qualitative and quantitative methodologies have their own strengths, researchers also note that caution should be exercised when attempting to use both methodologies in a single research project. This caution is due to two main reasons: first, as previously mentioned, quantitative and qualitative methodologies have different philosophical foundations, and using a mixed methodology can lead to a lack of philosophical coherence (Östlund et al., 2011). Second, from a practical perspective, it can be challenging for a single researcher to carry out a study that involves both quantitative and qualitative analyses concurrently (Johnson, 2018).

More specifically, in the field of tourism demand forecasting research, the primary methodological developments have been made in quantitative methods over the past five decades (Song et al., 2019). Although qualitative methods (i.e., judgmental methods such as the Delphi technique) have made significant contributions to the development of long-term tourism demand forecasting methods (Vanhove, 1980), their value is mainly reflected in their ability to detect uncertainty in tourism demand (Kaynak & Macaulay, 1984) and further revise quantitative-based forecasts (Edgell et al., 1980). Qualitative methods have also been criticized for their biased predictions, created by the involvement of subjective individual opinions

(Kaynak & Macaulay, 1984). Therefore, quantitative methods (i.e., econometric models and algorithm-based automatic natural language processing techniques) are selected as the basis for examining aggregated Twitter data's capabilities when it comes to generating destination management implications. Considering the fact that this research will incorporate Twitter data with a large amount of unstructured information (Bigné et al., 2019), qualitative methods (e.g., judgemental screening of stop words and seed words in the topic-modelling process) are selected as a supplement to interpret the semantic information and thereby enable the identification of communication flows, irrelevant information, and noise. Detailed illustration of the methodological choices made for each study conducted in this research will be given in the following subsections.

3.4 Research context and data collection

3.4.1 Research context

For this thesis, the British Museum has been selected as the primary research context. Established by an Act of Parliament in 1753, the British Museum stands as one of the world's most renowned institutions for the preservation and display of human history and culture. However, it continues to grapple with issues of overcrowding (Maddison & Foster, 2003), diminished visitor satisfaction (Khaleeli, 2015), and ethical issues (Michaelson, 2022; Siddique, 2018).

The British Museum originated from the collections of Sir Hans Sloane and the Harleian assemblage of manuscripts, and first opened its doors to the public in 1759. While predominantly dependent on government funding, the museum also garners income through sponsorships and various fundraising efforts. Its collections are categorized into ten divisions: Coins and Medals; Egyptian Antiquities; Ethnography; Greek and Roman Antiquities; Japanese Antiquities; Mediaeval and Later Antiquities; Oriental Antiquities; Prehistoric and Romano-British Antiquities; Prints and Drawings; and Western Asiatic Antiquities (Maddison & Foster, 2003, pp. 173). The institution generally offers free admission to visitors, with the exception of temporary exhibitions that may require an entrance fee (ThBritishMuseum, 2022). Consequently, it is ranked as the top free-entry museum in London (VisitLondon, 2022). In 2018, the museum attracted a total of 5.8 million visitors (see Figure 4.1 in Chapter 4). This high volume of visitors can adversely affect the museum experience, leading to overcrowding and difficulties accessing exhibits (Santana-Jiménez & Hernández, 2011). Even during less crowded periods, negative externalities may arise from visitor interactions. The possibility of increased attendance resulting in unfavourable outcomes in terms of overall enjoyment is a pressing concern for the museum’s management (Khaleeli, 2015; Maddison & Foster, 2003).

In recent years, the British Museum has encountered mounting pressure from protests against its affiliations with fossil fuel companies and the possession of stolen artefacts. Activist groups, such as “BP or not BP?”, have staged demonstrations and performance interventions within the museum, questioning its ethics in partnering with controversial sponsors such as BP (Addley, 2023; Michaelson, 2022). Additionally, the museum has faced calls for the repatriation of certain artefacts, such as the Parthenon Sculptures (previously known as Elgin Marbles) (Razzall, 2022b) and the Benin Bronzes (Razzall, 2022a), which were acquired through colonialist actions or looting (Siddique, 2018). These issues highlight the complex ethical landscape that the British Museum must navigate in order to maintain its reputation and secure funding, while also addressing the legitimate concerns of protestors and source communities.

The rationale for selecting the British Museum as the primary research context is threefold:

(1) Representativeness: As a globally renowned cultural landmark, the British Museum shares similarities with institutions such as the Louvre and the Smithsonian in terms of its significance to the host city (Macdonald, 2010). VisitLondon (2022) indicates that, among the top 10 free attractions in the UK, six are museums. Within this group, the British Museum was the most visited museum during the 2014-2019 financial year (see Table 3.4 for detailed information). Overcrowding challenges, frequently faced by such institutions (Khaleeli, 2015), result from the unique and irreplaceable nature of museum exhibits (Maddison & Foster, 2003). Numerous studies have employed the British Museum as a case study to investigate ways to address this concern, particularly focusing on improved tourism demand forecasting (Kim et al., 2022; Maddison & Foster, 2003; Volchek et al., 2019). Therefore, re-examining the British Museum’s case using novel research designs offers the opportunity to generate comparative insights and develop a generalizable modelling strategy for tackling common factors

contributing to overcrowding, such as limited capacity, seasonality, local events, and marketing efforts (Capocchi, Vallone, Amaduzzi, & Pierotti, 2019).

(2) Data availability: First of all, the UK government provides official and verified data, unlike unverified secondary sources, on tourist arrivals to the British Museum, which is essential for tourism demand forecasting. Second, the British Museum's Twitter account, boasting over 2 million followers (refer to Table 3.4), offers a wealth of engagement data (see Figure 3.4) that facilitate high-frequency, real-time, and information-rich analyses (Bigné et al., 2019; Ridhwan & Hargreaves, 2021). Furthermore, the British Museum's distinct name contributes to a reduction of irrelevant tweets during the data collection process on Twitter. By using a regular expression[4], we can efficiently collect tweets containing the phrase "British Museum" from the entire Twitter platform. For comparison, consider the Science Museum in Kensington. Given that numerous museums incorporate the phrase "Science Museum" in their names (e.g., Boston Museum of Science/Boston Science Museum, Science Museum of Virginia, and Science Museum Oklahoma), ensuring the relevance and coverage of data collected using the search keyword "Science Museum" for the Kensington location within the entire Twittersphere would present a significant challenge.

(3) Ethical complexity: The British Museum's complex ethical challenges, originating from its affiliations with fossil fuel corporations (as per Table 3.4, indicating that the museum is among the two principal London institutions with such partnerships) and the ownership of disputed artefacts (Siddique, 2018), provide an unparalleled context to explore the influence of noise - tweets that neither stimulate visits nor signal tourists' intentions to visit, as identified in Section 2.6.3 - on tourist arrivals. This context allows for the exploration of strategies aimed at enhancing visitor satisfaction while addressing complex ethical challenges. The findings of this study will not only contribute to the existing literature but also provide valuable insights for museum management into how they might improve tourists' visiting experiences and handle ethical concerns associated with corporate sponsorship and artefact ownership.

[4] str.contains(r'\british museum\b')

Table 3.4 Information on major museums in London

No	Museum	Twitter username	Followers	Tourist arrivals (2014-2019 financial year)	Fossil fuel sponsors
1	British Museum	britishmuseum	2 million	30,196,206	BP (Addley, 2023)
2	Tate Modern	tate	4.7 million	28,083,538	BP (Brown (2016): ended in 2017)
3	National Gallery	NationalGallery	982.9k	26,766,384	Shell (Vaughan (2018): ended in 2018)
4	Natural History Museum (South Kensington)	NHM_London	2.2 million	23,584,617	Shell (Parry (2009): ended in 2008) Ørsted (Ørsted (2023): divestment of oil and gas in 2017)
5	Science Museum (South Kensington)	sciencemuseum	664.66k	15,536,632	BP, Shell, Adani (Monbiot, 2021)
6	National Portrait Gallery	NPGLondon	411k		BP (BP (2022): no extension beyond 2022)

Table notes: 1. This table is sorted in descending order based on the tourist arrivals column. 2. Major museums in London are taken to be those listed as the most famous free attractions by VisitLondon (2022). 3. The V&A Museum is not included in this table as it has no history of collaboration with energy companies. 4. Additionally, the National Portrait Gallery, which has received controversial funding from BP, is included in the table to give further context.

3.4.2 Data collection strategy and ethical consideration

In this research, two types of data were collected: tourist arrivals data and Twitter data. Specifically, we first obtained data illustrating the volume of tourist arrivals to The British Museum from the UK government website,[5] which were used to generate an important variable y_t^j across the three studies. y_t^j denotes the number of tourist arrivals to museum j at time t, spans 1 January 2014 to 31 December 2019, and is sampled at a monthly frequency.

Then, we collected Twitter data via the Twitter application programming interface[6] (API). Notably, almost all studies listed in Table 2.4 collected data via Twitter's API except for Lu et al. (2022), which collected data on the daily number of non-follower impressions from a secondary source, Crimson Hexagon. Their measure represented the number of times the content was seen by users who did not follow any associated actor, director, studio, or movie

[5] Obtained from: https://www.gov.uk/government/statistical-data-sets/

[6] We accessed the Twitter API using an academic research account. More information about academic research access to the Twitter API can be found at the following link: https://developer.twitter.com/en/products/twitter-api/academic-research

on Twitter. With these data, the authors were able to differentiate between tweets posted by followers and those by non-followers of a selected Twitter account. By combining data collected from Twitter's API and from Crimson Hexagon, Lu et al. (2022) created an individual-level tweets dataset. They discovered that tweets volume is significantly associated with follower impressions but not with non-follower impressions. These studies highlight the potential of unorthodox methods in investigating the complexities of content created by customers in Twitter research. By incorporating novel approaches and pushing the envelope of data collection and measurement, they have expanded the possibilities for future research, encouraging fresh avenues of exploration in the study of Twitter communication and user behaviour. However, it is crucial to acknowledge that the methodologies employed in these studies can be resource-intensive and time-consuming. For purposes such as forecasting tourism demand, there is a need to strike a balance between data curation and efficiency. This balance ensures a cost-effective and efficient approach for tourism operators, facilitating informed decision-making without overburdening resources.

The first step in this research was to identify the appropriate keywords relevant to the research context. We followed Dergiades et al. (2018) in our keyword selection process, identifying generic and more specific search queries and assessing their relevance to the research before retrieving Twitter data. Based on the search query assessment, the term "British Museum" was selected as the query keyword. Considering that The British Museum is located in London and English is the language most used by tourists of different nationalities to exchange opinions (Dergiades et al., 2018), we used search queries written in English. Using the selected query keywords, tweets were continuously obtained and stored between 1 January 2014 and 30 September 2022.

Researchers have adopted various strategies to extract Twitter signals of interest. They have used specific Twitter handles to collect tweets between certain entities (Bigné et al., 2019; Borah et al., 2020) and utilized lists of hashtags or keywords to obtain tweets on specific topics (Liaukonytė et al., 2022). These approaches, while able to produce datasets primarily comprised of focused signals, might limit the breadth of the collected data. Furthermore, collecting data through specific Twitter handles, hashtags, or keywords might only reduce the proportion of noise in the dataset, without completely eradicating it. Irrelevant tweets sent by bots (Kraaijeveld & De Smedt, 2020) or spam tweets (Hennig-Thurau et al., 2015) can still constitute noise, potentially hampering the accuracy and effectiveness of data analysis. Therefore, the application of effective techniques to filter out noise from collected data, ensuring trustworthy insights, is essential (O'Neal et al., 2018). In this regard, Hennig-Thurau et al. (2015) provided a valuable example of managing noise problems using clustering techniques (i.e., SVM models). However, efforts to effectively differentiate signals from noise within the vast scope of social media data have been scant (Xie et al., 2020). Consequently, this thesis will also explore the potential of text analysis and clustering techniques for efficient social media noise detection. Detailed illustrations will be given in the methodology sections for Studies 2 and 3 respectively.

The Twitter dataset consists of several columns providing valuable information about each tweet and its respective engagement metrics. The columns included in the dataset are as follows:

- "created_at": The timestamp of when the tweet was posted.
- "date": The date on which the tweet was posted.
- "time": The time of day when the tweet was posted.
- "username": The username of the user who posted the tweet.
- "tweet": The textual content of the tweet.
- "name": The display name of the user who posted the tweet.
- "likes": The number of likes the tweet has received, reflecting the general approval of the message (Saxton et al., 2019).
- "retweets": The number of times the tweet has been retweeted. According to Saxton et al. (2019), retweets can be seen as a message-specific countersignal, and the number of retweets can effectively represent a user's ability to generate content with "pass-along value," drawing upon signaling theory.
- "replies": The number of replies to the tweet. Similar to retweets, they reflect dialogic engagement and the ability to generate conversation (Saxton et al., 2019).

This structured dataset allows for the analysis of various aspects of the tweets, especially those attributes that reflecting user engagement including tweeting, retweeting, liking and replying (commenting).

The COVID-19 pandemic significantly impacted tourist arrivals, as demonstrated by the dramatic drop in visitor numbers to The British Museum during the lockdown period (see Figure 6.1 in Chapter 6). Furthermore, this impact on the tourism industry persists (Seddon, 2023). As it has not been long since the easing of lockdown and travel restrictions (refer to Table 6.1 in Chapter 6 for details), there is limited data available for concurrently accounting for the seasonality feature of tourist arrivals and the pandemic's influence for the data spanning from March 2020 to September 2022. Consequently, directly incorporating such data into the analytical frameworks of Studies 1 and 2 could generate biased results. To ensure the stability and robustness of the research findings in Studies 1 and 2, we exclude the data from March 2020 to September 2022 from the subsequent analyses. Our investigation of the dynamics of tourist arrivals and Twitter data, therefore, focuses on the pre-lockdown period, from January 2014 to December 2019. With this approach, we aim to maintain the reliability of the research findings while paving the way for future studies to validate the present study's results, taking into account the impact of the COVID-19 pandemic on tourists' offline and online behaviour as more data become available. In Study 3, the time range for the Twitter data is extended to cover the period from 1 January 2014 to 30 September 2022, so that we can examine whether the intensity of noise (tweets that neither stimulate visits nor signal tourists' intentions to visit) changes during and after the lockdown period. This extension allows for a more comprehensive analysis of the evolving public sentiment in response to the pandemic and its subsequent effects on the museum sector.

3.5 Methodological choice

The research philosophy, approach, method, and context have already been discussed in detail. The focus of this section will now shift to the specific methodological choices that were made for each of the three studies conducted for this thesis. In order to comprehensively address each research objective proposed in Chapter 2, more than a single study was needed. Thus, an in-depth exploration of each research objective was undertaken through three studies, as shown in Figure 3.1 below. Detailed descriptions of the methodological choices made for each study are provided in the following subsections.

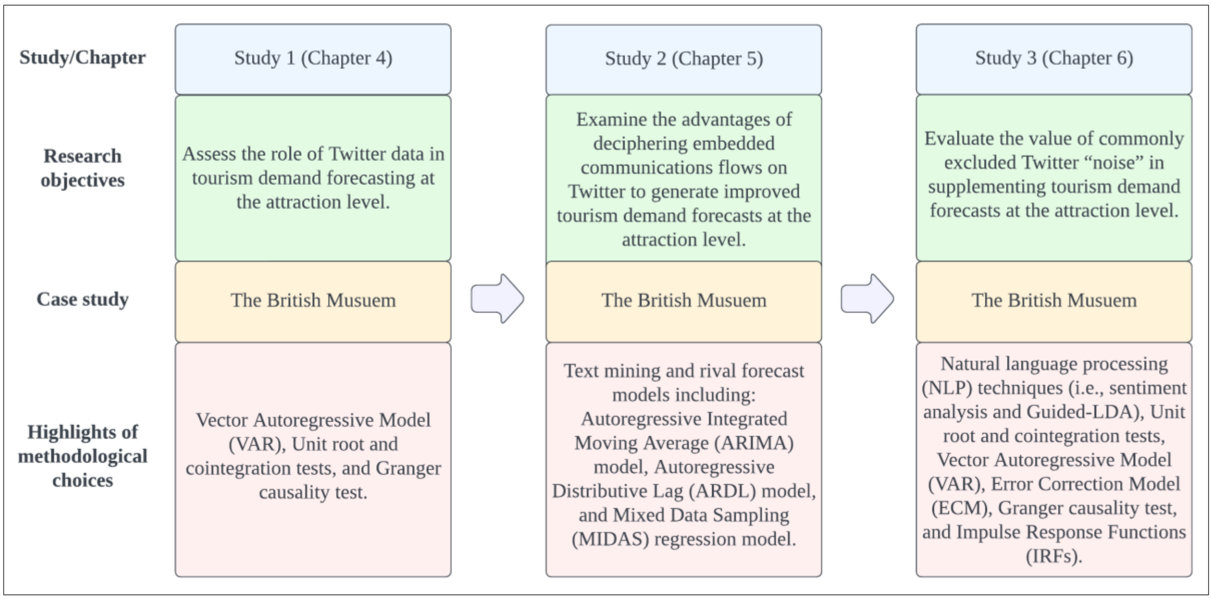

Figure 3.1 Research design

3.5.1 Study 1: Assess the role of Twitter data in tourism demand forecasting at the attraction level

As shown in Figure 3.1, Study 1 aims to explore the predictive value of Twitter data for tourist arrivals. To achieve this, as will be presented in Chapter 4, we employ the vector autoregressive (VAR) model as our estimation method, which offers a more comprehensive approach than traditional econometric models that often struggle to effectively capture the dynamic interrelationships between variables (Liu et al., 2019). The VAR model excels at characterizing dynamic variables using non-structural techniques, making it a highly valuable first step for conducting tourism demand forecasting studies, as exemplified by Huang et al. (2017) and Liu et al. (2019). One of the key advantages of the VAR model is its compatibility with the Granger causality test, which enables us to assess the predictive power of one variable over another. By employing the Granger causality test in our analysis, we can further strengthen the validity and reliability of our findings, providing a solid basis for future research. In the following sections, we will delve into the mathematical foundations of VAR models and the Granger causality test, elucidating their applicability and effectiveness for our study.

3.5.1.1 Vector autoregressive (VAR) model

Introduced in the 1980s, the VAR model is a non-structural equation model used for predicting multivariate time series and the dynamic effects of random disturbing variables in a system (Watson, 1994). Due to its advantages, including being less constrained by theory, avoiding default variables, and being more intuitive for analysing long-term dynamic effects between variables (Liu et al., 2019), the VAR model has frequently been applied in the tourism sector to analyse the dynamic relationships between tourist arrivals and various online predictors (Li et al., 2021; Song & Li, 2008). A standard $VAR(p)$ model is expressed as in equation (1):

$$Z_t^m = c^m + A_1 Z_{t-1}^m + A_2 Z_{t-2}^m + \cdots + A_p Z_{t-p}^m + \varepsilon_t^m, \tag{1}$$

where, for our research purposes, Z_t^m is a vector composed of two endogenous variables, that is, $Z_{m,t} = (y_t^m, Twitter - generated\ variables\ )$, y_t^m represents tourist arrivals to museum m at time t, and the vector $Twitter - generated\ variables$ includes $Volumes_t^m$, $Likes_t^m$, $Retweets_t^m$ and $Replies_t^m$, which have been developed in line with previous research (Chen et al., 2021; Hu et al., 2022; Valsesia et al., 2020; Villanova & Matherly, 2023; Wu et al., 2022); c^m is the constant term; ε_t^m is a two-dimensional random perturbation vector; p is the lag; and the 2×2 dimensional matrix $A_1 + A_2 + \cdots + A_p$ is the matrix to be estimated.

Following Guitart and Hervet (2017), we must make three crucial decisions to ensure the proper specification of the VAR model: (a) determining whether to use an error correction specification, (b) choosing between a linear or a log-linear specification, and (c) selecting the lag for the dependent and independent variables. In order to decide whether to use an error correction specification (decision (a)), it is essential to examine whether two or more series display a long-term equilibrium, indicative of cointegration. As cointegration can only occur if the series are non-stationary, unit root tests are employed to determine the nature of the series. To ensure the robustness of the conclusions, we utilize four well-established tests: the augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Each test offers unique advantages. The ADF method is widely used for detecting unit roots in time series, while the PP test builds upon the ADF by allowing more flexible assumptions regarding error terms. The KPSS test, with a null hypothesis of stationarity, complements other tests that assume non-stationarity. Applying these tests provides a comprehensive understanding of the stationarity of the series, which is crucial for determining the suitability of an error correction specification in the model. If the test results meet the premise (both variables are non-stationary) for building an error correction model (ECM), we would then examine the long-run equilibrium between the variables using the Engle and Granger (EG) cointegration test.

The EG cointegration test is used to determine if two time series are cointegrated (Asteriou & Hall, 2007; Enders, 2008). The test aims to establish a long-run equilibrium relationship between the time series (i.e., variables), while allowing for short-term adjustments. Specifically, we first estimate a linear regression model that includes all variables of interest (known as the cointegrating regression) using ordinary least squares (OLS). For example, if y and x are the two variables, the estimated equation will be $y_t = \alpha + \beta x_t + \varepsilon_t$. Then, we check the stationarity of the residual series ε_t from the OLS regression using the ADF test. If ε_t is stationary, y and x are cointegrated, indicating a long-run equilibrium relationship. If not, y and x are not cointegrated, and an ECM is not appropriate.

If ε_t is stationary, an ECM is then used to model the short-term dynamics and adjust for deviations from the long-run equilibrium (Asteriou & Hall, 2007; Enders, 2008). To estimate the short-run dynamics between y and x in the ECM, we calculate the first differences of both time series: $\Delta y_t = y_t - y_{t-1}$ and $\Delta x_t = x_t - x_{t-1}$. Next, we include the lagged residuals ε_{t-1} from the cointegrating equation as an additional independent variable, creating the error correction term for the ECM. This term captures deviations from the long-run equilibrium and how they adjust in the short run. Then, we can estimate the ECM by running an OLS regression of the first difference of the dependent variable Δy_t on the first difference of the independent variable Δx_t and the error correction term ε_{t-1}. The estimated equation will be $\Delta y_t = \alpha + \beta \Delta x_t + \gamma \varepsilon_{t-1} + u_t$, where γ represents the adjustment coefficient and u_t is the error term. Notably, γ represents the speed of adjustment back to the long-run equilibrium. A significant negative γ implies that deviations from the long-run equilibrium are corrected over time.

In terms of decision (b), three practical factors must be considered. First, it is essential to determine whether we need to use the differenced value of the variables or another approach (e.g., the Season-trend Decomposition (STL) approach, the moving average filter, or the X-12 process) to address seasonality issues. While a log-linear model permits the interpretation of coefficients as elasticities, incorporating differenced values can complicate results interpretation and hinder the comparison of findings with other studies. Second, if both specifications would account for a high percentage of data variance, we would need to ascertain whether the optimal number of lags in the log-linear model would be smaller than that in the linear specification. In such a case, we should opt for the specification that yielded a more parsimonious model. Lastly, models built with a log-linear specification often provide an inferior in-sample model fit (Önder et al., 2019). Consequently, unless the first two factors strongly suggest adopting a log-linear specification, we would tend to prefer the linear specification to ensure a more accurate and easily interpretable model.

For decision (c), the optimal lag length is determined by considering several criteria, including the sequentially modified likelihood ratio test statistic (LR), the final prediction error (FPE), the Schwarz Bayesian information criterion (SC), the Akaike information criterion (AIC), and the Hannan-Quinn information criterion (HQ). The optimal lag length is selected based on the majority consensus of these five criteria. Specifically, as our goal is to ensure a parsimonious model specification, we give preference to the SC, which is more sensitive to smaller lag lengths. Taken together, decisions (a), (b), and (c) are made by balancing the needs for model accuracy and simplicity, ultimately resulting in a more efficient and effective model.

3.5.1.2 Granger causality test and impulse response analysis

The Granger causality test is based on the concept of predictability in time series analysis, aiming to determine whether one variable contains useful information for predicting another variable, while not implying a cause-and-effect relationship. The test is named after Clive Granger, who developed the concept and was awarded the Nobel Prize in Economics for his work in 2003 (Granger, 1969). The mathematical foundation of Granger causality is based on the autoregressive models, where a time series is regressed on its own lagged values and those of other time series, and hence is often applied within the context of VAR models, which are multivariate time series models capturing the linear interdependencies among multiple variables. Therefore, the test meets the needs of the research aim of Study 1.

Consider a bivariate $VAR(p)$ model for two time series, X and Y. The model can be written as follows:

$$y_t = \delta_0 + \sum_{i=1}^{n} \delta_{1i} y_{t-i} + \sum_{j=1}^{m} \delta_{2j} x_{t-j} + \varepsilon_{2t} \quad (2)$$

$$x_t = \beta_0 + \sum_{i=1}^{n} \beta_{1i} x_{t-j} + \sum_{j=1}^{m} \beta_{2i} y_{t-i} + \varepsilon_{1t} \quad (3)$$

where m and n are lag lengths for x_t and y_t, respectively, β and δ are parameters for estimation, and ε_{1t} and ε_{2t} are error terms.

In this $VAR(p)$ model, each variable is regressed on the lagged values of itself and all other variables in the system. The Granger causality test can be applied to examine whether the coefficients associated with the lagged values of one variable are statistically significant in the equation of the other variable. More specifically, the Granger causality test involves comparing the fits of two nested models, as shown below:

$$y_t = \delta_0 + \sum_{i=1}^{n} \delta_{1i} y_{t-i} + \sum_{j=1}^{m} \delta_{2j} x_{t-j} + \varepsilon_{2t} \quad (4)$$

$$y_t = \delta_0 + \sum_{i=1}^{n} \delta_{1i} y_{t-i} + \varepsilon_{2t} \quad (5)$$

If the inclusion of the lagged values of x_t in equation (4) significantly improves the fit of equation (5), then it can be concluded that x_t Granger causes y_t. In other words, past values of X contain information that helps predict Y. Similarly, if the inclusion of the lagged values of y_t significantly improves the fit of a model that includes only the lagged values of x_t, then it can be concluded that y_t Granger causes x_t. To determine the statistical significance of the improvement in fit, we can use an F-test, comparing the residual sum of squares (RSS) of the two models expressed in equations (4) and (5):

$$F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k)} \quad (6)$$

In equation (6), RSS_r refers to the RSS of the restricted model (equation (5)), RSS_{ur} refers to the RSS of the unrestricted model (equation (4)), q is the number of restrictions (i.e., the number of excluded lags of X), n is the number of observations, and k is the number of estimated parameters in the unrestricted model. The F-statistic follows an F-distribution with $(q, n - k)$ degrees of freedom. If the calculated F-statistic is greater than the critical value from the F-distribution at a chosen significance level, we reject the null hypothesis that X does not Granger-cause Y. If the F-statistic is not greater than the critical value then we accept the null hypothesis.

Alternatively, the linear Granger causality test can be expressed as follows:

$$\Delta y_t = \delta_0 + \sum_{i=1}^{n} \delta_{1i} \Delta y_{t-i} + \sum_{j=1}^{m} \delta_{2j} \Delta x_{t-j} + \varepsilon_{2t} \quad (7)$$

$$\Delta x_t = \beta_0 + \sum_{i=1}^{n} \beta_{1i} \Delta x_{t-j} + \sum_{j=1}^{m} \beta_{2i} \Delta y_{t-i} + \varepsilon_{1t} \quad (8)$$

where Δ is the difference operator, m and n are lag lengths for x_t and y_t, respectively, β and δ are parameters for estimation, and ε_{1t} and ε_{2t} are error terms. To test whether the Granger causality runs from x_t to y_t, the null hypothesis (H_0) is:

$$H_0: \delta_{2j} = 0, j = 1, 2, \ldots\ldots\ldots q.$$

If at least one of the δ_{2j} is not equal to zero, then H_0 is rejected, and this suggests that x_t Granger causes y_t. This means that the past values of x_t have significant linear predictive power for the current value of y_t.

When adopting the Granger causality test, it is important to note that Granger causality should not be interpreted as implying true causality in the sense of a cause-and-effect relationship. Rather, it suggests that one time series can provide useful information for predicting another time series (Granger, 1969; Sims, 1980).

Building upon the findings from the VAR model and Granger causality test, we proceed to conduct impulse response analysis to further investigate the magnitude and dynamics of the identified Granger causality. Impulse response analysis enables us to examine how shocks to one variable affect other variables in the system over time within the context of our estimated VAR model (Liu et al., 2019). Only the Twitter-generated variables that have been tested and found to have Granger causality with tourist arrivals are analysed in this step. The following steps outline the impulse response analysis within the VAR framework:

- Structural shocks identification: As the error terms in the VAR model may be contemporaneously correlated, it is essential to identify the structural shocks, which are orthogonal and have economic interpretations. Techniques such as Cholesky decomposition can be applied to transform the reduced-form errors into structural shocks.
- Impulse response functions (IRFs) computation: Utilizing the estimated bivariate VAR(p) model and the identified structural shocks, we can calculate the IRFs. The IRFs describe each variable's response to a one-time shock to another variable, while holding all other shocks constant. Mathematically, the IRFs can be formulated as follows:

$$IRF_{i,j}^{s} = \frac{\partial z_i^{t+s}}{\partial \varepsilon_i^{t}}, \text{for } i, j = 1, 2 \quad (9)$$

where s denotes the time horizon after the shock to variable j at time t (s is typically assumed to be 1, as we are interested in the response of variable i to a one-time, one-unit shock to variable j), $\partial \varepsilon_i^t$ denotes the partial derivative of the one-unit shock to variable j at time t, ∂Z_i^{t+s} represents the partial derivative of variable i at time $t+s$ with respect to the shock to variable j at time t, and the indices $i, j = 1, 2$ indicate that the formula applies to both variables in the bivariate $VAR(p)$ model. Then, $IRF_{i,j}^s$ represents the response of variable i at time $t+s$ to a one-time shock to variable j at time t. It can be computed recursively using the estimated coefficients from the bivariate $VAR(p)$ model and the identified structural shocks.

- Results analysis: We can then assess the calculated IRFs to understand how a shock to the tested Twitter-generated variables influences tourist arrivals over time. We visualize the response dynamics by plotting the IRFs, which can provide insights into the magnitudes and durations of the effects.
- Confidence intervals computation: To evaluate the statistical significance of the impulse responses, confidence intervals can be computed using methods such as asymptotic or Monte Carlo simulations. In this study, we use the asymptotic approach to calculate the response standard error for our analysis.

By implementing impulse response analysis within the VAR system, we can gain a more in-depth understanding of the magnitudes and dynamics of the relationships between tourist arrivals and the Twitter-generated variables found to have Granger causality with arrivals, in response to shocks to the system.

3.5.2 Study 2: Examine the advantages of deciphering embedded communication flows on Twitter for generating improved tourism demand forecasts at the attraction level

Building on the findings from Study 1 (see Chapter 4), Study 2 aims to investigate the advantages of harnessing Twitter communication flows for generating more precise demand forecasts. To achieve this objective, we adopt the methodological framework proposed by Li et al. (2021). The framework is composed of five carefully designed steps: (1) Data selection, (2) Data extraction, (3) Data processing and transformation, (4) Model estimation, and (5) Performance evaluation, as depicted in Figure 3.2. The first two steps have been thoroughly discussed in previous sections (see Section 3.4), providing a strong foundation for our research. In this section, we will place emphasis on providing a detailed account of steps (3), (4), and (5), which are integral to the overall research process and ultimately contribute to the validity and reliability of our findings.

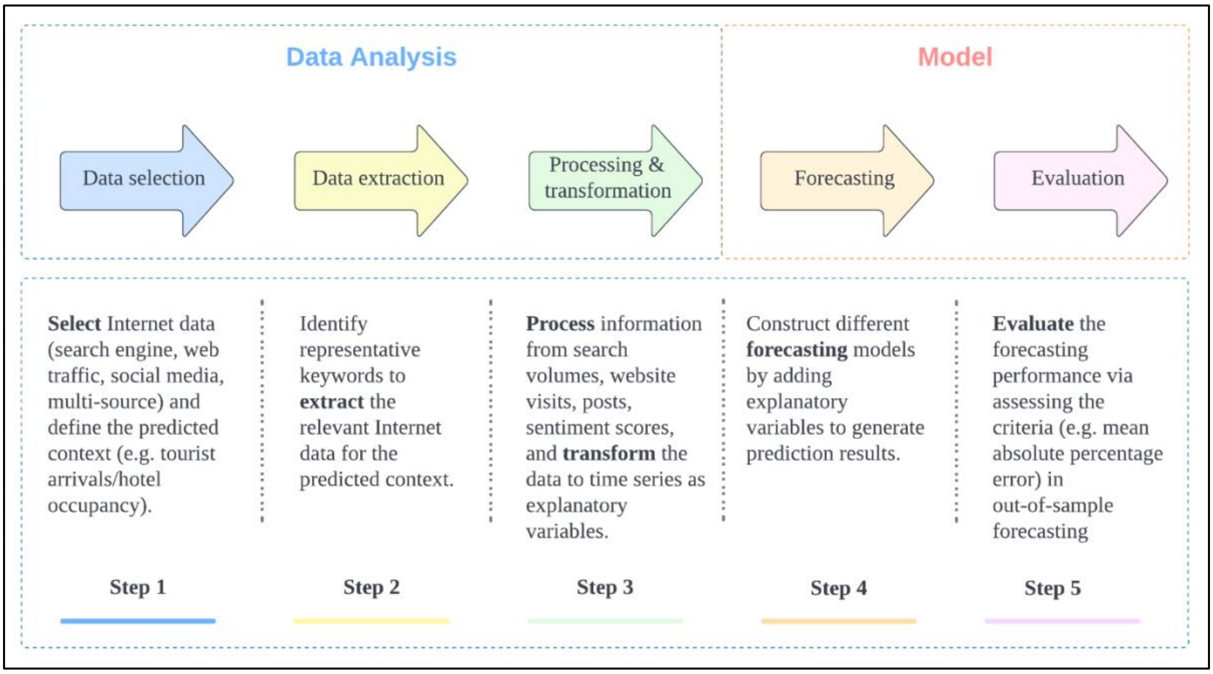

Figure 3.2 Process of tourism demand forecasting with social media data

Source: Li et al. (2021, pp. 5)

3.5.2.1 Twitter data processing and transformation

As discussed in the literature review, research using social media data tends to rely on temporally aggregated and volume-based data. The primary reason is that social media data are unstructured, with the majority being textual, and thus challenging to process (Berger et al., 2020). However, it is unreasonable to ignore the noise issue when attempting to capitalize on the value of social media data for signalling customer (tourist) expectations and experiences. The noise issue may stem from two factors. First, compromise in the data collection strategy may result in the inclusion of irrelevant data. For example, to collect as much relevant information as possible, we adopted a search-query-based approach to gather data across the entire Twitter platform (see Section 3.4.2). This approach resulted in a dataset with extensive coverage of diverse signallers and receivers, but also more irrelevant data (see the example given in Section 3.5.2.1.1). Second, while relevant to the research case, tweets that do not reflect the attraction's marketing efforts, or indicate tourist interest and intention to visit (e.g., negative tweets), are often considered "noise" in the context of tourism demand forecasting (refer to Section 2.6.3 for a detailed discussion on the definition of "noise"). In Study 2, since our goal is to use Twitter data to generate improved demand forecasts, all data not relevant to tourists' visiting intentions are considered noise. In line with signalling theory, as the context in which the signalling process occurs becomes increasingly noisy, the value of the signalling process is expected to diminish (Connelly et al., 2011). Taking these arguments together, Twitter data pre-processing is an essential step for Study 2, especially for fulfilling the research aim of generating improved forecasts of tourist arrivals, and cross-validating the research findings from Study 1 to ensure they are not biased by noise.

Various text-mining approaches can be employed to analyse the Twitter data and fulfil the pre-processing needs, such as entity extraction (e.g., sentiment analysis) and relation extraction (e.g., supervised machine-learning models) (Berger et al., 2020). The design of the pre-processing framework depends heavily on the research context and aims. In the context of Study 2, given that we have already made significant efforts regarding case selection and data collection strategies, our objective in Chapter 5 is to address the noise issues in the Twitter dataset from two perspectives: (1) filtering out irrelevant tweets and noise, and (2) charting the communication flow of the processed Twitter dataset.

3.5.2.1.1 Filtering out irrelevant tweets and noise

As previously mentioned, determining whether to filter out a tweet from the data is based on two factors. The first factor is whether the tweet is relevant to the case study. To achieve this, we use regular expressions to screen the textual content of each tweet. For example, with the Twitter dataset of the British Museum, we filter the "text" column to retain only those tweets containing the phrase "British Museum" (i.e., str.contains(r'\british museum\b')) and those tweets posted by the official account of the British Museum (which may not necessarily include the phrase "British Museum" but are still considered relevant data). This process allows us to exclude tweets containing the words "British" and "museum" separately, rather than the specific phrase "British Museum".

Secondly, while it is essential to identify relevant tweets for demand forecasting purposes, it is also crucial to determine whether a tweet should be classified as noise. In other words, we must evaluate whether the tweet indicates the tourist's intention to visit a specific destination. It is evident from press releases that the British Museum faces pressure from climate protests (both offline and online) and criticism regarding its contested artefacts (Adams, 2019; Addley, 2023; Michaelson, 2022; Siddique, 2018). Therefore, our goal is to filter out tweets related to these two topics because they are not relevant to the British Museum's marketing efforts, or indicate tourist interest and intention to visit. We employ the state-of-the-art keyword-in-context search approach to achieve this goal. To identify these in-context keywords, we analysed the top 50 tweets with the most likes for each case to gain insights. This decision is also supported by the surging trend of series "likes" for both cases (see Figure 5.4 in Chapter 5). By examining these popular tweets, we can better understand the context and sentiment surrounding the British Museum and develop a more precise keyword list so as to filter out noise by simply discarding tweets containing words from the formed list. Furthermore, to ensure the accuracy and stability of the above tweet-filtering process, all tweets underwent a strict cleaning process, including text cleaning, tokenization, stop words removal, and lemmatization, as suggested by Berger et al. (2020). This refined dataset provides a clearer picture of tourists' visiting intentions and allows for improved demand forecasting. Detailed results will be given in Chapter 5.

3.5.2.1.2 Communication flow mapping for Twitter data

As discussed in Section 2.6.2.2, the digital marketing communication framework proposed by DONNA and Novak (1997) complements the signalling theory and highlights the significant impact of internet-mediated environments on traditional marketing communication paradigms. This notion is particularly evident in the widespread popularity of social media (Chaffey & Ellis-Chadwick, 2019). As depicted in Figure 2.2, customers (tourists) and organizations (attractions) can now freely interact with each other through internet mediums, such as social media platforms, in particular Twitter, which further complicates the communication process. As a result, the literature emphasizes the necessity of clarifying signallers, receivers, the broader audience, and their communication flows in this complex signalling environment (Branzei et al., 2004). In the context of Study 2, it implies that a critical component of using Twitter data for modelling tourism demand is to identify and capture signals embedded in diverse communication flows that either represent tourists' attention and interest in an attraction or the attraction's marketing efforts.

Using the British Museum as an example, we can recognize the following communication flows. Communication flow one (C1): Direct communication with the British Museum; communication flow two (C2): Peer-to-peer conversations about the British Museum; and communication flow three (C3): Communication from the British Museum to the broader audience. Considering the structure of the tweets dataset (see the illustration in Section 3.4.2), we can assign each tweet to the communication flow that best matches the tweet's content through keyword matching, using the columns "username", "text", and "mentions" in the Twitter dataset. More specifically, tweets that mentioned the British Museum's official Twitter account (i.e., tweets that contain "britishmuseum" in the "mentions" column or include "@britishmuseum" in the "text" column) are assigned to C1. Then, tweets posted by individuals (excluding the British Museum's official Twitter account) that contain the term "British Museum" in their tweets' textual content (i.e., tweets that include "British Museum" in the "text" column) are assigned to C2. Finally, tweets posted by the British Museum's official Twitter account (i.e., tweets with the value "britishmuseum" in the "username" column) are assigned to C3.

3.5.2.2 Rival forecasting models

In the fourth step of the proposed framework, rival models are selected to generate forecasts of tourist arrivals to the British Museum. Specifically, all models are estimated and ex-post forecasts are generated, evaluated, and compared. All models are estimated based on training data (January 2014 to December 2016). Restricted by the low-frequency feature of $Arrivalsa_t$, we follow the strategy of generating pseudo-out-of-sample one-step-ahead static forecasts (Önder et al., 2019). Following Bangwayo-Skeete and Skeete (2015) and Hu et al. (2022), we use 36 months as the out-of-sample forecasting period. Specifically, 36 one-step-ahead static forecasts are generated recursively from January 2017 to December 2019. As discussed in the literature review, a wide range of models (i.e., time series models, econometric models, and AI-based models) have been examined by scholars for their ability to accommodate internet data (Li et al., 2021; Song et al., 2019). Although there is a trend to incorporate AI-based models in data-driven forecasting practice, these models are mainly used to accommodate non-linear data without prior knowledge about the relationships between the input and output variables (Hu et al., 2022). Considering that this research is exploratory in nature, the selection of forecasting models is therefore restricted to within the scope of time series models and econometric models.

3.5.2.2.1 The autoregressive integrated moving average (ARIMA) model

The autoregressive integrated moving average (ARIMA) is the most widely used time series forecasting model in the field of tourism demand forecasting as it performs well in diverse research contexts (Song et al., 2019). Considering the exploratory nature of this research, the ARIMA model class is first employed as the benchmark. ARIMA is a classical time series stochastic process model, which contains autoregression (AR), moving averages (MA), and a difference component (Cho, 2003). A general ARIMA model takes the form:

$$\varphi^*(L^j)y_t^j = a^j + \theta(L^j)\varepsilon_t^j \tag{10}$$

In equation (10), y_t^j for our purposes denotes the number of tourist arrivals in period t to attraction j (in this case, y_t^j denotes the monthly aggregated volume of tourist arrivals); a^j is the attraction-specific intercept; ε_t^j represents the attraction-specific error term in period t. Finally, $\varphi^*(L^j)$ and $\theta(L^j)$ are attraction-specific lag polynomials of finite orders p and q with d unit roots ($\varphi^*(L^j) = \varphi(L^j)(1 - L^j)^d$).

3.5.2.2.2 The autoregressive distributive lag (ARDL) model

A dynamic model class employed in this study is the autoregressive distributive lag (ARDL) model class, which incorporates lags of the dependent variable as additional explanatory variables (Narayan, 2004). The ARDL class has frequently been used in the tourism demand modelling literature (Gunter, Önder, & Gindl, 2019; Narayan, 2004; Önder et al., 2019). By allowing not only for a current but also for a lagged influence of the explanatory variable(s), this model class is much more flexible than the benchmark ARIMA model. More precisely, specific (or reduced) ARDL models are derived, through a general-to-specific approach with automatic lag selection through the application of SC values as criteria, from the following general ARDL model:

$$y_t^j = a^j + \sum_{i=1}^{12} b_i^j \cdot y_{t-i}^j + \sum_{i=1}^{12} c_i^j \cdot x_{t-i}^j + \varepsilon_t^j \tag{11}$$

In equation (11), y_t^j denotes the volume of tourist arrivals in period t to attraction j, a^j is the attraction-specific intercept, b_i^j and c_i^j are the attraction-specific regression coefficients on past realizations of tourists' arrivals ($y_{t-i}^j; i = 1,2, \ldots ,12$), and on current and past realizations of the Twitter-generated variables ($x_{t-i}^j; i = 1,2, \ldots ,12$), respectively. Finally, ε_t^j represents the attraction-specific error term in period t. Comparing equations (10) and (11), it can be found that the established ARIMA model (if it contains no MA terms) could be seen as a special case of an ARDL model.

Initially, the optimal lags for lagged forecast variables y_t^j and the predictor(s) x_t^j are determined using the AIC and an automatic model selection process with a maximum initial lag of 10. The ARDL models are then generated through a general-to-specific modelling approach. To avoid potential biases arising from autocorrelation of the error terms and heteroscedasticity in their variances, heteroscedasticity and autocorrelation consistent (HAC) standard errors are utilized.

3.5.2.2.3 The mixed data sampling (MIDAS) regression model

Finally, to avoid information loss when processing high-frequency data, which may lead to inefficient and biased model estimation, the MIDAS model, proposed by Ghysels et al. (2006) and Andreou et al. (2011), is employed. The MIDAS model incorporates high-frequency variables into a low-frequency process with a parsimonious weighting scheme. By doing so, it leads to a better in-sample model fit and may also lead to higher out-of-sample forecast accuracy. A growing number of tourism demand forecasting researchers have applied the MIDAS model (Bangwayo-Skeete & Skeete, 2015; Havranek & Zeynalov, 2021; Wen et al., 2021). Notably, Önder et al. (2019) and Hu et al. (2022) applied MIDAS models to social media data.

In this study, we employ a restricted MIDAS autoregression (R-MIDAS-AR) (Clements & Galvão, 2008). In the R-MIDAS-AR model, the AR component refers to the inclusion of lags of the low-frequency dependent variable, which accounts for habit persistence and tourist expectations (Song et al., 2003). To ensure consistency with the optimal number of lags for the dependent variable determined through the general-to-specific approach in equation (5), the exact optimal number of low-frequency lags is incorporated during R-MIDAS-AR estimation. Following the notation of Önder et al. (2019), a general R-MIDAS-AR is expressed as follows:

$$y_t^j = a^j + \sum_{i=1}^{12} b_i^j \cdot y_{t-i}^j + c^j \cdot \sum_{k=1}^{m} \omega(k^j, \theta) L_{HF}^{kj} x_t^j + \varepsilon_t^j \qquad (12)$$

In equation (12), x_t^j denotes daily-aggregated Twitter-generated variables, L_{HF}^{kj} represents the attraction-specific high-frequency lag operator with k^j daily lags, which are automatically selected in a general-to-specific approach from a maximum of m = 60, c^j denotes the attraction-specific vector of coefficients of the shape parameters for the high-frequency volume of tweets. The maximum number of daily lags is determined by the results of the Granger causality test conducted in Study 1, which implies that only Twitter activities dated close to an actual visit to museum j have explanatory power over y_t^j.

Furthermore, $\omega(k^j, \theta)$ in equation (12) denotes the weighting function of the high-frequency Twitter data. More specifically, it is a functional lag polynomial that determines the weights of the individual and museum-specific high-frequency lags for temporal aggregation (Clements & Galvão, 2008) and whose shape is determined by θ, the vector of shape parameters. Following Önder et al. (2019), we apply a non-exponential Almon function with four shapes for the specification of θ as it has been suggested as demonstrating the best overall in-sample model fit. Equation (12) can then be estimated by applying standard OLS. The remainder of the notation in equation (12) has the same interpretation as that in equation (11).

3.5.2.3 Evaluation strategy

3.5.2.3.1 In-sample model fit evaluation

In-sample model fit evaluation refers to the process of evaluating the performance of a statistical model using the same data that was used to train the model. This is also known as "training set evaluation" or "internal evaluation". The goal of in-sample model fit evaluation is to assess how well the model fits the data it was trained on, and to identify any potential issues with the model, such as overfitting or underfitting (Inoue & Kilian, 2005). For time series forecasting models, the AIC and the Bayesian Information Criterion (BIC) are the two most commonly used criteria (Qi & Zhang, 2001). Rissanen (1980) demonstrated that the BIC provides a consistent estimation of the order of an AR model. Consequently, the BIC is frequently favored over the AIC in practical applications since it is a more dependable criterion for model selection. Another commonly used criterion for selecting the number of regressors is Theil's adjusted R^2. While the standard coefficient of determination (R^2) measures the model's goodness of fit, it almost always increases and never decreases with the number of regressors. Consequently, if R^2 were used as a model selection criterion, it would always favour a larger number of lags. The adjusted R^2 corrects the problem with an adjustment based on the degrees of freedom (Qi & Zhang, 2001). Therefore, we chose the BIC and adjusted R^2 as the criteria for the in-sample model fit evaluation in this research. It is also important to note that in-sample model fit evaluation is only one aspect of model evaluation, and should be complemented by out-of-sample evaluation using holdout or test data to assess the model's ability to generalize to new, unseen data.

3.5.2.3.2 Out-of-sample forecasting performance evaluation

Out-of-sample forecasting performance evaluation refers to the process of evaluating the accuracy of a statistical model's predictions on data that were not used to train the model. This is also known as "holdout set evaluation" or "external evaluation". The goal of out-of-sample forecasting performance evaluation is to assess the model's ability to generalize to new, unseen data, and to identify any potential issues with the model, such as overfitting or underfitting (Inoue & Kilian, 2005). According to Petropoulos et al. (2022), there are several common methods for evaluating the out-of-sample forecasting performance of a model, including:

(1) Mean absolute error (MAE): MAE is the average absolute difference between the observed values and the predicted values. It is a measure of the model's overall forecasting accuracy. MAE is easy to interpret and compare across different models or datasets, as it is expressed in the same units as the original data. A smaller MAE indicates that the model's predictions are closer to the observed values, and vice versa.

(2) Mean squared error (MSE): MSE is the average squared difference between the observed values and the predicted values. It is a measure of the model's overall forecasting accuracy, with a higher weight given to larger errors. MSE is generally more sensitive to the magnitude of the errors than other measures such as MAE, which gives equal weight to all errors regardless of their size. MSE is expressed in the same units as the original data, but squared. A smaller MSE indicates that the model's predictions are closer to the observed values, and vice versa.

(3) Root mean squared error (RMSE): RMSE is the square root of MSE. It is a measure of the model's overall forecasting accuracy. RMSE is expressed in the same units as the original data, which makes it easy to interpret and compare across different models or datasets. A smaller RMSE indicates that the model's predictions are closer to the observed values, and vice versa. It is important to note that RMSE is sensitive to the scale of the data, and it may not be the most appropriate evaluation criterion in all cases.

(4) Mean absolute percentage error (MAPE): MAPE is the average absolute percentage difference between the observed values and the predicted values. It is a measure of the model's overall forecasting accuracy, with a higher weight given to larger errors. MAPE is especially useful when the scale of the data is important, as it allows you to compare the model's performance across different datasets with different scales. A MAPE of a smaller value indicates better performance. However, it is important to note that MAPE

can be undefined or misleading in cases where either the predicted values or the observed values are close to zero.

(5) Symmetric mean absolute percentage error (SMAPE): SMAPE is a measure of the average absolute percentage difference between the observed values and the predicted values, with the average being taken over the sum of the absolute percentage difference between the observed and predicted values. It is a symmetric measure that treats positive and negative errors equally, and is useful when the scale of the data is important and the researcher wants to give equal weight to all errors regardless of their size. However, as with MAPE, it can be undefined or misleading in cases where either the predicted values or observed values are close to zero.

It is important to note that different evaluation criteria may be more or less appropriate depending on the specific forecasting problem and the characteristics of the data. It is often helpful to compare the performance of a model using multiple evaluation criteria to get a more comprehensive understanding of its accuracy (Chai & Draxler, 2014; Willmott & Matsuura, 2005). In line with Song et al. (2019), four criteria, namely MAE, RMSE, MAPE, and SMAPE, are selected to measure the forecasting performance in this research. Specifically, these indicators are calculated using the following equations:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y_i}|}{n} \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y_i}|}{y_i} \tag{14}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y_i}|}{(y_i + \hat{y_i})/2} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{16}$$

Here, y_i denotes the actual number of tourist arrivals while $\hat{y_i}$ denotes the predicted number of tourist arrivals; n denotes the number of forecasts used for evaluation. The model with the lowest values of the four criteria can be considered the best forecasting model. To make the comparison of the values of the above evaluation criteria more straightforward, we further calculate the improvement of rival models over the benchmark model. Taking the improvement measured by RMSE as an example, the equation is as follows:

$$Improvement = \frac{RMSE(benchmark) - RMSE(rival\ model)}{RMSE(benchmark)} \quad (17)$$

The model with the highest values of improvement according to the above evaluation criteria can be considered the best forecasting model. Comparison groups can be constructed according to the employed model classes and incorporated variables to draw specific conclusions (see Figure 3.3 as an example).

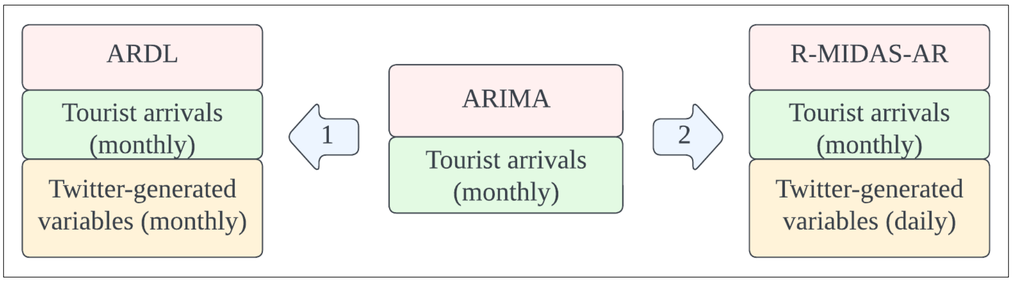

Figure 3.3 An example of the construction of a comparison group in the forecasting performance evaluation process

3.5.3 Study 3: Evaluate the value of commonly excluded Twitter "noise" for supplementing tourism demand forecasts at the attraction level

Drawing on the results of Studies 1 and 2, the main objective of Study 3 is to reevaluate the signalling value of Twitter "noise" - tweets that do not stimulate visits or signal tourists' intentions to visit - for tourist arrivals. Specifically, the study focuses on tweets related to boycotts aimed at the British Museum, building on our knowledge of the case and the findings of Study 2 (refer to Appendix 1).

3.5.3.1 Supplementary data

As was discussed in the data collection strategy section, in Study 3, the time range is extended to cover the period from 1 January 2014 to 30 September 2022, for tourist arrivals data and Twitter data, to analyse whether the intensity of noise (boycotts) changed during and after the lockdown period. To gain a more thorough understanding of the context and impact of noise, we decided to include the following supplementary data in Study 3.

First, the dates of any physical protests related to each boycott campaign were collected. By comparing the dates of the protests with the timing of the Twitter boycott campaigns, we aimed to determine whether the social media boycott campaigns were reactive to these offline events. Specifically, we identified the dates of physical protests by searching newspaper articles via the Lexis-Nexis database. Lexis-Nexis is a well-established and widely used source of newspaper articles and other media content. It provides a comprehensive database of news articles and other publications from around the world, covering a wide range of topics and time periods. In particular, Lexis-Nexis's interface provides several filters that facilitate the identification of boycott-related news under a specific search query (in our case "protest British Museum"), including "negative news" and "practice area and topics". Several recent studies have used Lexis-Nexis to identify negative events (Gutierrez-Huerter O et al., 2023; Janney & Gove, 2017; Varma, 2021).

Second, according to Biraglia and Gerrath (2020), in times of crisis, such as the COVID-19 pandemic, tourists at museums may be more open to the idea of corporate sponsorship. Therefore, we consider the effects of COVID-19-related travel restrictions to examine whether the intensity of boycotts changes during and after the lockdown period. More specifically, we incorporate the dates when COVID-19 travel restrictions in the UK were imposed and later lifted. These dates were identified from official announcements made by the UK government. Descriptions of the collected supplementary data are presented in Chapter 6.

3.5.3.2 Framework of noise detection

For Study 2, we utilized a state-of-the-art keyword-in-context search approach to identify Twitter noise and discovered that the majority of these tweets are associated with boycott topics (refer to Appendix 1). The lists of keywords were constructed from an analysis of the most popular tweets with boycott themes. This methodological design ensured that the list covered the most representative keywords related to boycott intentions, subject to the most popular boycotting campaigns, although sacrificing the scope of coverage for negative sentiment triggered by other reasons or campaigns beyond our knowledge. At the same time, the simplicity of the keywords included in the list may have caused the issue of over-screening. While this approach is convenient and efficient for demand forecasting since tweets with boycott intentions account for a small proportion of the overall data (i.e., less than 5%), Study

3 focuses on tweets with boycott intentions (hereafter referred to as boycott tweets). Consequently, for this study, we needed to design a more sophisticated and robust framework for detecting boycott tweets and to further facilitate the precise capture of the influence of diverse boycott campaigns on tourist arrivals.

As discussed in Section 3.5.2.1, the challenge in mining Twitter data lies in the fact that social media data are unstructured, the majority textual (Berger et al., 2020). Various natural language-processing techniques can be applied to mining unstructured textual data (e.g., sentiment analysis and topic modelling: Ibrahim and Wang (2019a)). While sentiment analysis has been applied in research analysing public concerns (Ridhwan & Hargreaves, 2021; Signorini et al., 2011), it is not necessarily the most effective or reliable means of detecting public concerns. Typically, sentiment analysis depends heavily on the frequency of specific relevant words or phrases. Consequently, such techniques are inadequate for considering the genuine semantic connections among various parts of a document, individual sentences, or even subclauses (Bhatia et al., 2015). Therefore, although it can provide a broad indication of overall sentiment, sentiment analysis may struggle to capture the subtleties and nuances of context-specific public concerns, which can vary widely depending on the situation. This is particularly relevant in the case of boycotts, where sentiment may be highly context-specific. As such, alternative approaches, such as topic modelling, may offer more effective means of identifying and analysing tweets related to boycott campaigns.

Topic modelling, unlike sentiment analysis, aims to uncover latent semantic structures in large collections of text by identifying recurring patterns of co-occurring words across documents (Berger et al., 2020). By generating insights and interpretations of the text, rather than focusing solely on prediction, topic modelling can increase our understanding of document content and provide a more nuanced perspective on public concerns such as boycotts. Via topic modelling, we aim to uncover key themes and concerns within the text, allowing for a deeper analysis of the context-specific issues related to boycott campaigns and their potential impact on tourist arrivals.

Latent Dirichlet allocation (LDA) and guided latent Dirichlet allocation (guided LDA) are two generative probabilistic models utilized for topic modelling in document collections. While both approaches aim to discover latent topics within a corpus, guided LDA incorporates prior knowledge in the form of seed words, to guide the topic discovery process. LDA, introduced by Blei et al. (2003), is an unsupervised machine-learning technique that models documents as mixtures of latent topics, where each topic is a probability distribution over words. The model assumes that documents are generated through a generative process involving Dirichlet distributions for document-topic and topic-word probabilities. By employing inference algorithms such as Gibbs sampling or variational inference, LDA estimates the underlying topic structure within the document collection (Yin & Wang, 2014). However, LDA does not consider any domain knowledge or user input, which may result in topics that are less interpretable or relevant to the domain of interest. Guided LDA, on the other hand, addresses these limitations by incorporating prior knowledge in the form of seed words (Andrzejewski & Zhu, 2009). The seed words are user-defined and represent specific themes or topics within the document collection. Guided LDA integrates these seed words into the topic modelling process by guiding the initialization of the topic-word distribution (β), thereby directing the algorithm to discover more relevant and interpretable topics (Jagarlamudi et al., 2012). This approach is particularly beneficial when users possess domain expertise or prior knowledge about the topics in the corpus, as it can improve the quality of the discovered topics and enhance their utility in downstream analyses (Toubia et al., 2019). In light of the existing knowledge about the British Museum, guided LDA is chosen as the preferred topic-modelling approach.

Guided by Shankar and Parsana (2022), we propose a three-stage framework that incorporates sentiment analysis and text-mining techniques to enable the topic-modelling process, as depicted in Figure 3.4. According to the understanding of the theoretical and practical foundation of the guided LDA approach (Andrzejewski & Zhu, 2009; Jagarlamudi et al., 2012), it is essential to recognize that the quality of the topics discovered in guided LDA depends heavily on the appropriateness and representativeness of the chosen seed words, necessitating careful selection and validation. Therefore, various text-mining approaches are adopted in Stage 2 to build comprehensive and representative lists of seed words for the initialization of the guided LDA. By leveraging sentiment analysis and seed word detection, the designed framework is expected to enable the discovery of more interpretable and relevant topics, ultimately enhancing the utility of the topic-modelling results in downstream analyses.

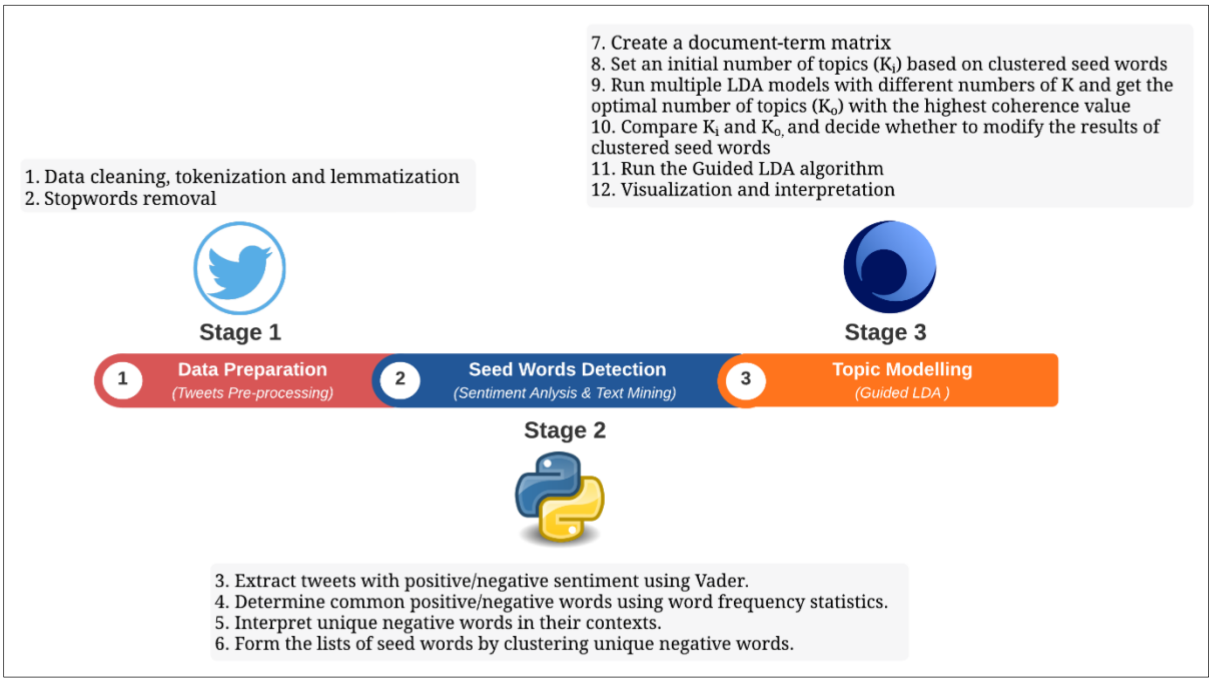

Figure 3.4 Framework of Twitter noise (boycotts) detection

Detailed descriptions of all the steps depicted in Figure 3.5 are provided below:

- Data cleaning, tokenization, and lemmatization: Clean the textual data by removing irrelevant content, tokenize the text into individual words, and lemmatize the words to reduce them to their root forms.
- Stopwords removal: Remove common stopwords (using Python packages including "SpaCy", "Gensim", and "Wordcloud") and customized stopwords (obtained from the previous step) that do not provide meaningful insights.
- Extract tweets with positive/negative sentiment using VADER: Analyse the sentiment of each tweet and identify those with positive, neutral, or negative sentiments using the VADER algorithm (Hutto & Gilbert, 2014). Tweets posted by the museum's official Twitter account are omitted from the analysis.
- Determine common positive/negative words using word frequency statistics: Calculate the frequency of words in positive and negative tweets and identify the most common ones (top 50).

- Interpret unique negative words in their contexts: Examine the contexts in which unique negative words appear, so as to better understand their meaning, and further decide whether to enter each keyword into the next step (Hannigan et al., 2019).
- Form the lists of seed words by clustering unique negative words: Group similar negative words together to create lists of seed words based on the understanding of the research context.
- Create a document-term matrix: Construct a matrix representing the frequency of words in each document (i.e., each negative tweet). To capture meaningful word combinations and phrases, the Python package "Gensim" is adopted to generate bigrams which are subsequently incorporated into the documents. Moreover, the term frequency-inverse document frequency (TF-IDF) transformation is applied to the corpus to emphasize the significance of specific words within the context of the documents, as suggested by Hong and Davison (2010). Lastly, any term that appears in more than half of the documents, or with a frequency lower than five, is removed.
- Set an initial number of topics (K_i) based on the clustered seed words constructed in the last step.
- Run multiple LDA models with different values of K and choose the optimal number of topics (K_o) based on the one with the highest coherence value (or the second peak after which the coherence value starts decreasing: Ridhwan and Hargreaves (2021)). For time-saving purposes, the Python package "tomotopy" is adopted at this step to run multiple LDA models and complete the tuning process of K_o.
- Compare K_i and K_o, and decide whether to modify the clustered seed words.
- Run the guided LDA algorithm: Implement the guided LDA model using the selected seed words and optimal number of topics.
- Visualization and interpretation: The Python library "pyLDAvis" is utilized to visualize and interpret the results of the guided LDA model and thereby help us understand the discovered topics and their relationships.

3.5.3.3 Model-free evidence

Following Sun et al. (2021) and Lin et al. (2021), prior to conducting econometric analysis, we first display the evolution of the key variables (i.e., tourist arrivals, boycott tweets, physical protests, and COVID-19 travel restrictions) over time to understand the correlations more intuitively. More specifically, assuming that Twitter boycotts, measured by the volume of tweets, likes, and retweets of or replies to pre-processed tweets, is negatively associated with tourist arrivals, we aim to capture the correspondence between the peak of Twitter boycotts and the lowest points of tourist arrivals in the pre-lockdown and post-lockdown periods. To avoid judgemental bias caused by visual measurement, we capture peaks of Twitter boycotts via the rolling-window z-score approach. In particular, peaks are identified based on their z-scores, which represent the number of standard deviations a data point is away from the rolling mean. The rolling mean and standard deviation are calculated using a window size of 12 months. Data points with a z-score greater than or equal to 1 are considered peaks, indicating that they are significantly different from the average behaviour in the dataset at the 16% significance level. To avoid selection bias and maintain the focus of the research on the pre- and post-lockdown periods, the rolling-window z-score is computed separately for two periods: firstly, the combination of the period from 31 January 2014 to 29 February 2020 and the period from 30 August 2021 to 30 September 2022, and secondly the period from 31 March 2020 to 31 July 2021. For the second of these, that is, the lockdown period, we intentionally use a window size of six months and extend the calculation window by including data for the six months prior to 31 March 2020.

3.5.3.4 Econometric evaluation

In Study 3, we also employ the VAR model as the estimation method, considering its advantages including being less constrained by theory, avoiding default variables, and being more intuitive for analysing long-term dynamic effects between variables (Liu et al., 2019). As discussed previously, the VAR model effectively characterizes these dynamic variables through non-structural techniques, making it particularly valuable in tourism demand forecasting studies. The standard VAR model used in this study is as follows:

$$Z_{m,t} = A_1 Z_{m,t-1} + A_2 Z_{m,t-2} + \cdots + A_l Z_{m,t-l} + \varepsilon_{m,t}, \tag{18}$$

where Z_t is a vector composed of two endogenous variables, that is, $Z_{m,t} = (SArrivals_{m,t}, Boycott_{m,t})$. $SArrivals_{m,t}$ presents the number of seasonally adjusted tourist arrivals to museum m at time t; $Boycott_{m,t}$ presents the intensity of Twitter boycotts, measured by the volume of tweets, likes, and retweets of or replies to pre-processed tweets, collected for museum m at time t; $\varepsilon_{m,t}$ is a two-dimensional random perturbation vector; l is the lag length; and the 2×2 dimensional matrix $A_1 + A_2 + \cdots + A_l$ is the matrix to be estimated.

Our study follows the modelling strategy implemented by Liu et al. (2019) and Huang et al. (2017), which consists of the following three main steps:

(1) First, given that our sample data are time series in nature, it is crucial to ensure stability and avoid spurious regression sequences by conducting unit root and cointegration tests before establishing econometric models. Consequently, we first employ the ADF test to evaluate stationarity. We then use the augmented EG two-step cointegration test to assess the cointegration relationship of the variables of interest. Following Huang et al. (2017), the first step is to establish the regression equation of our variables using OLS regression. In the second step, the estimated residuals from this regression are tested for stationarity. If these residuals are found to be stationary, it suggests that the variables under consideration are cointegrated, implying that they share a long-run equilibrium relationship.

(2) Subsequently, we apply the Granger causality test to investigate the causal relationships between the variables under consideration. The Granger causality test is crucial for determining the direction of causality between the variables and understanding their interactions over time. This test not only helps identify whether a variable can help predict another variable but also clarifies the temporal precedence of relationships between the variables, offering valuable insights into their causal structure (Diks & Panchenko, 2006).

(3) Finally, we utilize the IRF of the VAR model to reveal the dynamic interplay among these variables, providing an in-depth and precise understanding of their interactions. The IRF, derived from the VAR model, is used to examine the dynamic effect of a change in one variable on the entire system, capturing the system's response to the impact or disturbance of that variable (Götz et al., 2016). All subsequent analysis is performed using EViews version 11.

3.5 Ethical considerations

Although the data collection process for this research was not planned to include human participants, the collection and extensive analysis of a large amount of Twitter data obligated us to ensure that we obtained and utilized these data in a manner that upheld the highest ethical standards (Townsend & Wallace, 2016). Traditional ethical frameworks can provide some guidance for researchers working with social media data, but these approaches may not be sufficient to address the unique contextual challenges posed by this type of data. Therefore, we followed the guide constructed by Townsend and Wallace (2016) to address ethical challenges including anonymity and risk of harm in the process of collecting, processing, and publishing Twitter data for this research. Specifically, we only used publicly available content, collected via Twitter's official API. Our access to Twitter's API was enabled via the permitted academic developer account (case number: 0191389425). To ensure that no personal information was included in the research, only aggregate data was used in the forecasting and all account-related information, including username, user id, conversation id, and location, was omitted before the data were put into the text-mining process. Also, no direct quotes are included in this research nor will they be republished elsewhere. With regards to the tourist arrival data, since they are released by the Department for Culture, Media & Sport (DCMS) and are publicly available via the statistical dataset of museums and galleries on the UK government's website, no ethical hazards have been identified.

3.6 Limitations of research methods used

The limitations of the research methods used should be acknowledged to provide a comprehensive understanding of a study's scope and future research avenues. One of the primary concerns is the exclusion of demographic data on tourists and the focus on English-language tweets due to data availability constraints. This may reduce the explanatory power of the findings for tourists who use foreign languages (i.e., encounter the hazard of language bias: Dergiades et al. (2018)). Furthermore, ethical concerns led to the omission of any demographic data associated with the tweets, such as username and location. As a result, Studies 1 and 2 missed opportunities to explore spatial-temporal response patterns of tourist arrivals based on Twitter-generated variables. Second, the research methodology used in Study 2 does not account for the potential impact of boycott tweets on tourist arrivals or their effect on demand forecasting accuracy. The second limitation can be attributed to the interconnected nature of the three studies making up this thesis, as each subsequent study was informed by the findings of the previous one.

Chapter 4: Exploring Twitter data's value in forecasting tourism demand

4.1 Data

Following the data collection strategy discussed in Chapter 3, we obtained data depicting the volume of tourist arrivals to the British Museum from the UK government's website[7], which has been used to generate the dependent variable $Arrivals_t$. $Arrivals_t$ spans from 1 January 2014 to 31 December 2019 and is sampled on a monthly frequency. Then, we collected Twitter data relevant to the British Museum via the Twitter API using the term "British Museum" as the query keyword. Considering that the British Museum is located in London and English is the most used language for tourists of different nationalities to exchange opinions (Dergiades et al., 2018), this study is restricted to search queries and tweets written in English. Using the selected query keyword "British Museum", tweets are continuously obtained and stored between the periods of 1 January 2014 to 31 December 2019, resulting in a dataset with a total of *722,717* tweets. From this dataset, we generated variables, including $Volumes_t$, $Likes_t$, $Replies_t$ and $Retweets_t$. Descriptive statistics of $Arrivals_t$ and Twitter-generated variables are shown in Table 4.1 below.

[7] Tourist arrivals data can be accessed via the following link: https://www.gov.uk/government/statistical-data-sets/museums-and-galleries-monthly-visits

Table 4.1 Descriptive statistics of variables

Variable	Description	Period	Frequency	Max	Min	Mean	Std. Dev.
$\boldsymbol{Arrivals_t}$	Monthly volume of tourist arrivals to the British Museum at time t.	Jan 2014 – Dec 2019	Monthly	765877	363281	526110.1	94264.12
$\boldsymbol{Arrivalsa_t}$	Seasonally adjusted monthly volume of tourist arrivals to the British Museum at time t.	Jan 2014 – Dec 2019	Monthly				
$\boldsymbol{Volumes_t}$	Monthly volume of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	14998	5584	10037.74	2166.91
$\boldsymbol{Likes_t}$	Monthly volume of likes of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	221483	10485	49706.96	41139.15
$\boldsymbol{Retweets_t}$	Monthly volume of retweets of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	50176	7339	22724.68	10633.82
$\boldsymbol{Replies_t}$	Monthly volume of replies to tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	9383	1101	3232.528	1582.43

Source: UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Table notes: $Arrivalsa_t$ is generated by applying a moving average filter to $Arrivals_t$.

4.2 Stationarity test and cointegration test

Upon visual inspection of the time series plots for tourist arrivals (see Figure 4.1), it is evident that the variable $Arrivals_t$ exhibits a pronounced seasonal pattern. To address this, we apply a moving average seasonality filter[8] commonly used in forecasting studies to smooth data with pronounced seasonal patterns (Gunter & Önder, 2016; Kuzin et al., 2011). The resulting seasonally adjusted variable is denoted as $Arrivalsa_t$, which no longer exhibits a pronounced seasonal pattern, as shown in Figure 4.1. In addition, a visual inspection of the Twitter-generated variables (see Figure 4.2) indicates that both $Likes_t$ and $Replies_t$ present a clearly increasing trend.

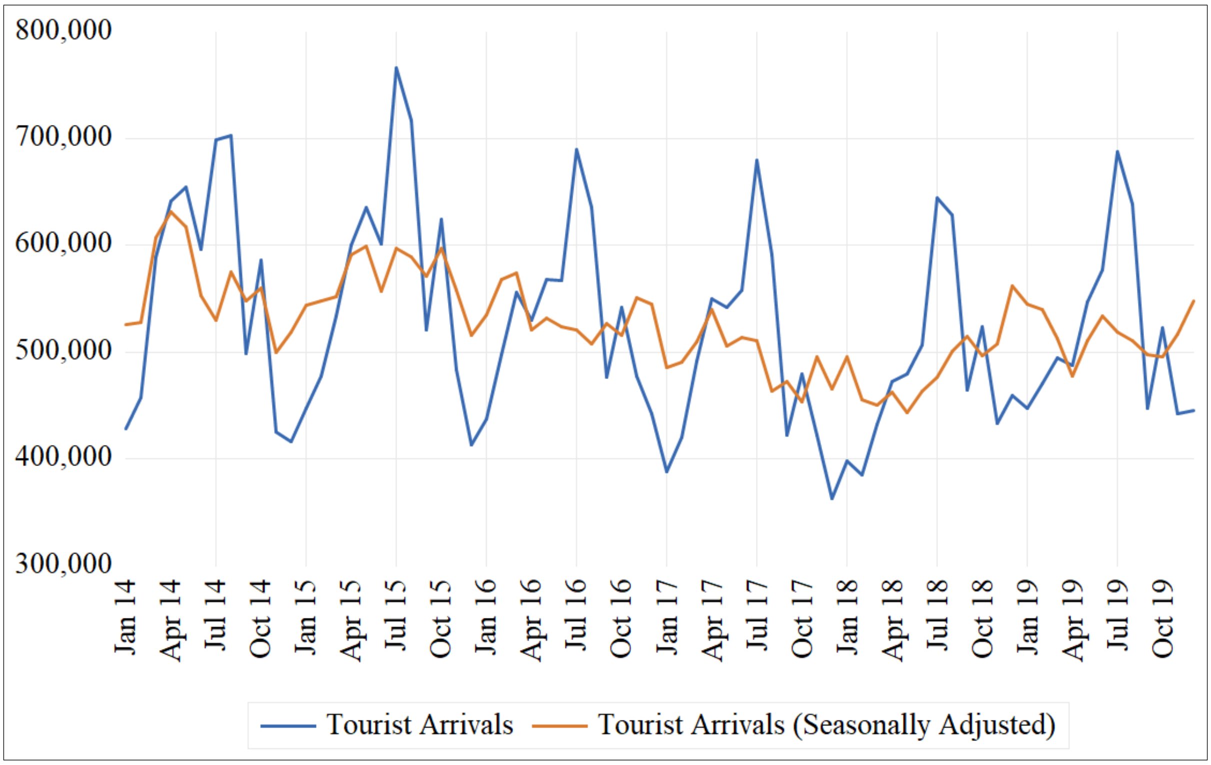

Figure 4.1 Visual inspection of tourist arrivals data

[8] These and all subsequent calculations in Study 2 are carried out with EViews Version 11.

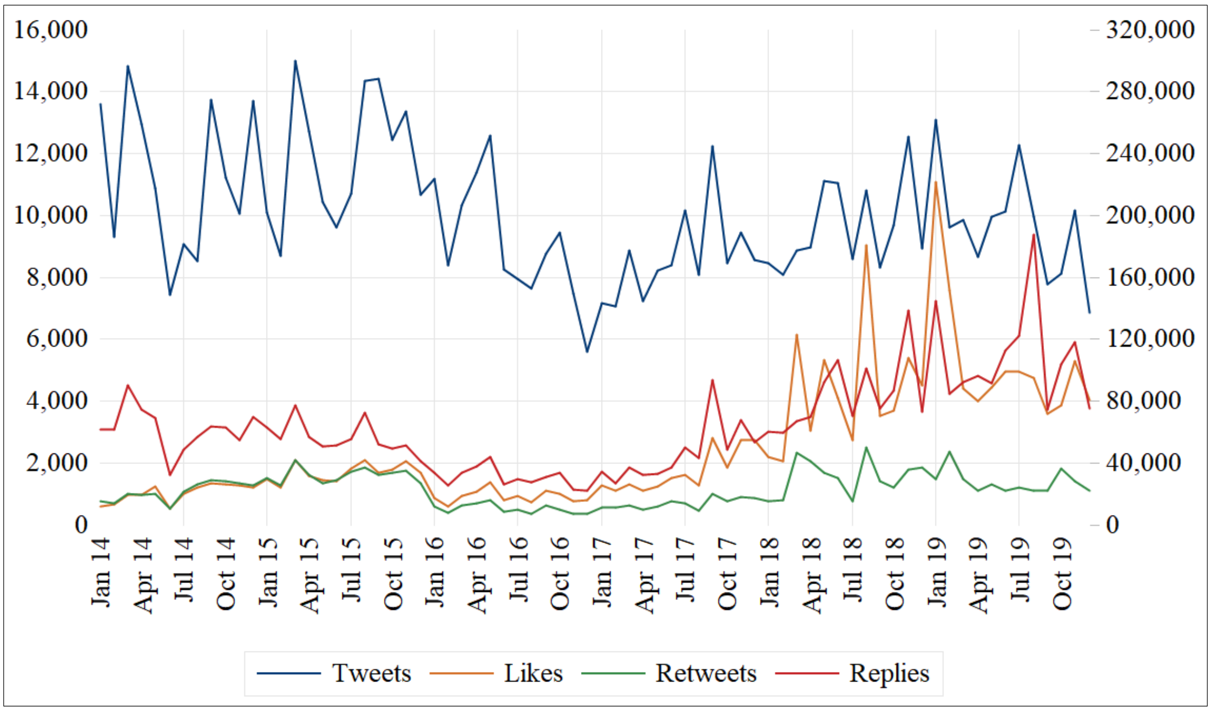

Figure 4.2 Visual inspection of Twitter-generated variables

Figure notes: The data of tweets and replies are scaled to the left axis, and the data of likes and retweets are scaled to the right axis.

Based on the above visual inspection, the stationary tests for these variables include the components of the intercept and trend, whereas, for the other variables, only the intercept is included. Table 4.2 presents the results of the ADF, Phillips-Perron, and KPSS tests for the various variables. To provide a more comprehensive understanding of the stationarity of these variables, we also include the first-order differenced value of each variable in the tests. We find that all three tests indicate that $Retweets_t$ is stationary, for $Arrivalsa_t$ and $Volumes_t$, the ADF and Phillips-Perron tests reject the null hypothesis of a unit root at the 5% significance level, while the KPSS test rejects the null hypothesis of stationarity. However, we conclude that both variables are stationary, based on the preponderance of evidence supporting their stationarity (i.e., the ADF and Phillips-Perron tests and visual examination). On the other hand, although the Phillips-Perron and KPSS tests support the stationarity of $Arrivals_t$, we conclude it to be non-stationary based on the strong evidence of seasonality obtained from visual inspection. More information on this judgement process can be found in Guitart and Hervet (2017). In addition, $Likes_t$ and $Replies_t$ are found to be non-stationary. However, their first-order differenced values are suggested by all three tests to be stationary at the 5% significance level. In summary, $Arrivalsa_t$, $Volumes_t$, and $Retweets_t$ are stationary at order $I(0)$, while $Arrivals_t, Likes_t$ and $Replies_t$ are stationary at order $I(1)$.

Table 4.2 Results of stationarity tests for Twitter-generated variables

Variables	ADF test (H_0: unit root)		Phillips-Perron test (H_0: unit root)		KPSS test (H_0: stationarity)		Terms included in the test equation
	t-Statistics	Probability	t-Statistics	Probability	LM-Statistics	5% Critical Value	
$\mathbf{Arrivals_t}$	-0.882	0.787	-4.376	0.001	0.256	0.463	Intercept
$\mathbf{\Delta Arrivals_t}$	-2.915	0.050	-9.766	0.000	0.045	0.463	Intercept
$\mathbf{Arrivalsa_t}$	-3.053	0.035	-2.936	0.046	0.667	0.463	Intercept
$\mathbf{Volumes_t}$	-5.736	0.000	-5.741	0.000	0.479	0.463	Intercept
$\mathbf{\Delta Volumes_t}$	-13.374	0.000	-45.616	0.000	0.322	0.463	Intercept
$\mathbf{Likes_t}$	-1.420	0.8461	-5.945	0.000	0.194	0.146	Intercept and trend
$\mathbf{\Delta Likes_t}$	-9.465	0.000	-43.188	0.000	0.093	0.146	Intercept and trend
$\mathbf{Retweets_t}$	-3.861	0.004	-3.722	0.006	0.176	0.463	Intercept
$\mathbf{\Delta Retweets_t}$	-8.802	0.000	-14.297	0.000	0.084	0.463	Intercept
$\mathbf{Replies_t}$	-2.223	0.4693	-5.084	0.001	0.259	0.146	Intercept and trend
$\mathbf{\Delta Replies_t}$	-6.100	0.000	-24.383	0.000	0.074	0.146	Intercept and trend

Source: the UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Table notes: "Δ" indicates the first-order differencing operation applied to the variable.

Based on the results of stationarity tests, we can conclude that $Arrivalsa_t$ and $Volumes_t$ meet the premise to fit a VAR model, while it is necessary to confirm whether there is a long-run equilibrium relationship among $Arrivals_t, Likes_t,$ and $Replies_t$. Following the test strategy discussed in Chapter 3, we apply the augmented Engle and Granger two-step cointegration test to assess the cointegration relationship of $Arrivals_t, Likes_t,$ and $Replies_t$. Specifically, we first establish the regression equation using the ordinary least squares (OLS) regression (refer to Table 4.3 for each test equation). Then, the estimated residuals (i.e., ε_t) from this regression are then tested for stationarity. If these residuals are found to be stationary, it suggests that the variables under consideration are co-integrated, implying that they share a long-run equilibrium relationship. Table 4.3 below presents the critical value for each test and test conclusions. According to Table 4.3, no long-run equilibrium relationship is detected among $Arrivals_t$, $Likes_t$, and $Replies_t$. Therefore, there is no need to fit error correction models for $Arrivals_t, Likes_t$ and $Replies_t$.

Table 4.3 Results of EG Cointegration Tests

No	Test equation	t-Statistics	Probability	Conclusion
1	$Arrivals_t = \alpha_1 + \beta_1 Likes_t + \varepsilon_{1,t}$	-0.772200	0.9363	Not co-integrated
2	$Arrivals_t = \alpha_2 + \beta_2 Retweets_t + \varepsilon_{2,t}$	-0.791809	0.9339	Not co-integrated

4.3 VAR modelling

Based on the evidence provided by the stationarity and cointegration tests, it is reasonable to fit bivariate $VAR(p)$ models for $Arrivalsa_t$ and Twitter-generated variables (i.e., $Volumes_t$, $Retweets_t$ $\Delta Likes_t$ and $\Delta Replies_t$). Following the modelling strategy proposed in Chapter 3, the following four $VAR(p)$ models are built (refer to Table 4.4). The optimal lag length for each model is determined by the majority consensus of these five criteria: LR, FPE, SC, AIC, and HQ. To determine if the selected model provides an adequate description of the data, we adopted the Lagrange Multiplier (LM) test for checking residual autocorrelation in each VAR model. The null hypothesis is that there is no residual autocorrelation in the VAR system. Test results are summarised in Table 4.4. As shown in Table 4.4, the null hypothesis is rejected at the 5% significance level for each model.

Table 4.4 Estimation and LM test results for VAR models

No	Variables	Model specification	LM Tests (H0: no serial correlation at lags 1 to h)		
			Lag	LRE-Statistics	Probability
1	$Volumes_t$ & $Arrivalsa_t$	$Arrivalsa_t = 0.687Arrivalsa_{t-1} + 3.563Volumes_{t-1} + 129049.4$ $Volumes_t = 0.01Volumes_{t-1} + 0.279Arrivalsa_{t-1} + 1946.311$	1	0.778	0.941
2	$Arrivalsa_t$ & $\Delta Likes_t$	$Arrivalsa_t = 0.593Arrivalsa_{t-1} + 0.026Arrivalsa_{t-2} + 0.001Arrivalsa_{t-3} + 0.167Arrivalsa_{t-4} + 0.07\Delta Likes_{t-1} - 0.16\Delta Likes_{t-2} - 0.278\Delta Likes_{t-3} - 0.037\Delta Likes_{t-4} + 109681.1$	1	1.769	0.778
			2	4.661	0.793
		$\Delta Likes_t = -0.705\Delta Likes_{t-1} - 0.607\Delta Likes_{t-2} - 0.458\Delta Likes_{t-3} - 0.584\Delta Likes_{t-4} - 0.09Arrivalsa_t + 0.147Arrivalsa_{t-1} + Arrivalsa_{t-2} - 0.226Arrivalsa_{t-3} + 0.038Arrivalsa_{t-4} + 72348.86$	3	7.085	0.852
			4	11.667	0.767
3	$Arrivalsa_t$ & $Retweets_t$	$Arrivalsa_t = 0.757Arrivalsa_{t-1} + 0.149Retweets_{t-1} + 124530$ $Retweets_t = 0.652Retweets_{t-1} - 0.014Arrivalsa_{t-1} + 15339.49$	1	4.565	0.335
4	$Arrivalsa_t$ & $\Delta Replies_t$	$Arrivalsa_t = 0.762Arrivalsa_{t-1} + 0.863\Delta Replies_{t-1} + 125483.5$ $\Delta Replies_t = -0.621\Delta Replies_{t-1} - 0.004Arrivalsa_{t-1} + 2354.788$	1	7.936	0.094

Table notes: "Δ" indicates the first-order differencing operation applied to the variable.

Furthermore, to assess the stability of each VAR system, we visually examined the roots of the characteristic polynomial. In Figure 4.3, these roots are represented by dots, and if none of them falls outside the unit circle, the VAR system meets the stability condition, indicating that the model is stable. As demonstrated in Figure 4.3, each model's stability is confirmed via visual examination.

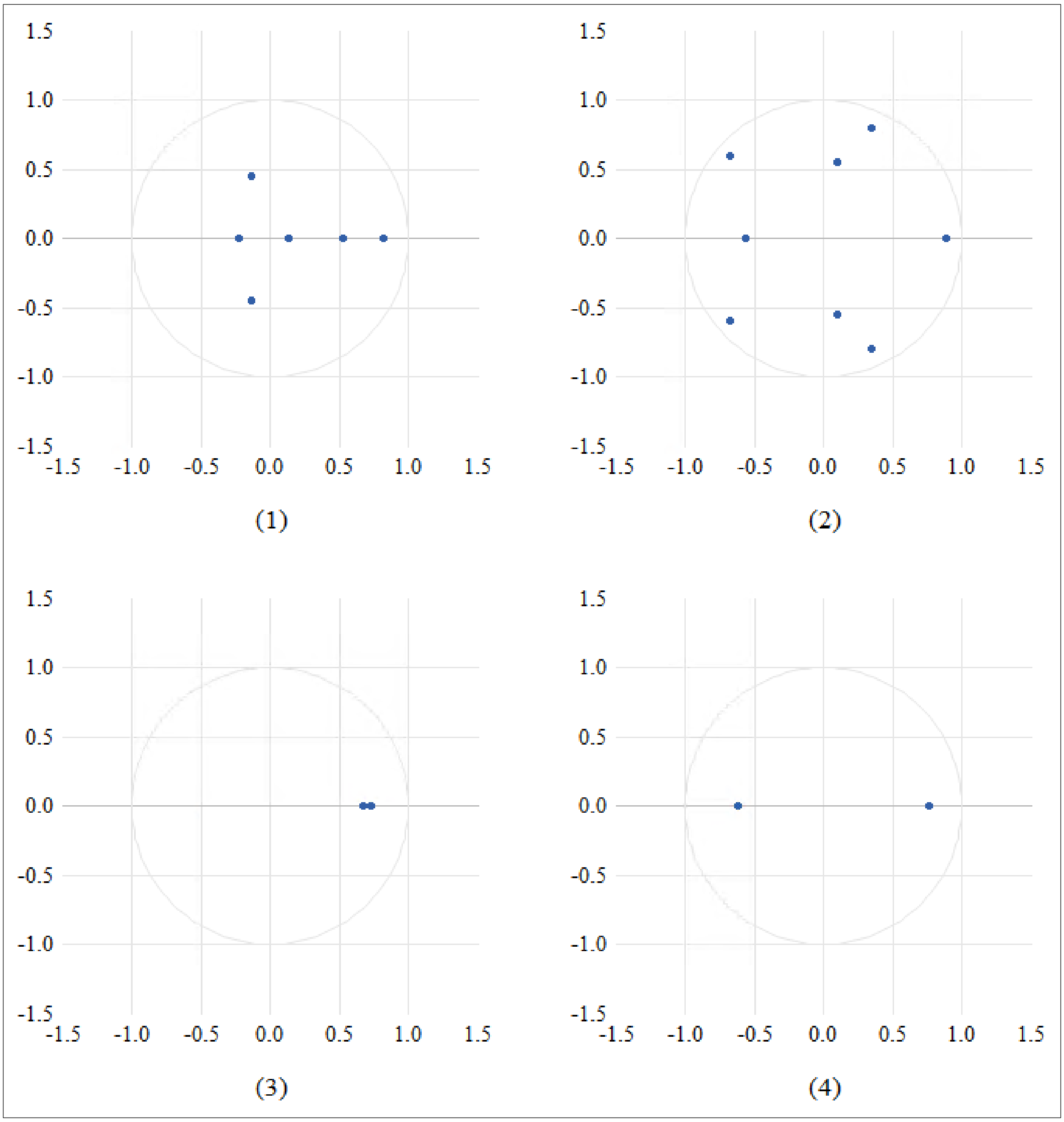

Figure 4.3 Stability test for VAR models

Figure notes: 1. The figure visualises the roots of the characteristic polynomial for each VAR model. 2. The numbering of figures (1) to (4) corresponds to the respective VAR models presented in Table 4.4.

4.4 Granger causality test and impulse response analysis

Although the cointegration test is capable of determining the existence of a long-term equilibrium relationship between $Arrivalsa_t$ and Twitter-generated variables, it does not reveal the causal relationship between them. To address this limitation, we employ the Granger causality test, as outlined in Chapter 3, to establish whether changes in one variable can be attributed to changes in another variable (Liu et al., 2019). The results of the Granger causality test are presented in Table 4.5. These results demonstrate that $Volumes_t$ Granger causes $Arrivalsa_t$, with a lag order of 1, at the 5% significance level. Additionally, it has been found that $\Delta Likes_t$ Granger causes $Arrivalsa_t$, with a lag order of 4, at the 10% significance level. This supports our expectation that including $Volumes_t$ and $\Delta Likes_t$ as relevant predictors for $Arrivalsa_t$ is appropriate. Meanwhile, no Granger causality has been detected between $Arrivalsa_t$ and $Rtweets_t$, or between $Arrivalsa_t$ and $\Delta Replies_t$. This suggests that these latter variables may not play a significant role in forecasting $Arrivalsa_{t.}$

Table 4.5 Results of the Granger causality test

Lag	Null hypothesis	F-Statistics	Probability	Conclusion
1	$Volumes_t$ does not Granger cause $Arrivalsa_t$	4.668	0.034	*Refuse*
1	$Arrivalsa_t$ does not Granger cause $Volumes_t$	2.621	0.110	Accept
4	$\Delta Likes_t$ does not Granger cause $Arrivalsa_t$	7.902	0.095	*Refuse*
4	$Arrivalsa_t$ does not Granger cause $\Delta Likes_t$	5.605	0.231	Accept
1	$Retweets_t$ does not Granger cause $Arrivalsa_t$	0.227	0.635	Accept
1	$Arrivalsa_t$ does not Granger cause $Retweets_t$	0.361	0.550	Accept
2	$\Delta Replies_t$ does not Granger cause $Arrivalsa_t$	0.116	0.735	Accept
2	$Arrivalsa_t$ does not Granger cause $\Delta Replies_t$	2.037	0.158	Accept

Table notes: "Δ" indicates the first-order differencing operation applied to the variable.

Subsequently, we employ the impulse response function in the VAR model to analyse the dynamic response process between $Arrivalsa_t$ and two Twitter-generated variables (i.e., $Volumes_t$ and $\Delta Likes_t$) based on the results of Granger causality and cointegration tests. By introducing a standard deviation shock to $Volumes_t$ and $\Delta Likes_t$, we can observe the resulting changes in $Arrivalsa_t$. The results are displayed in Figure 4.4, where the horizontal axis represents the number of impact response periods set to 10 months, and the vertical axis represents the impulse response function of seasonally adjusted tourist arrivals. The dotted lines on both sides represent the confidence bands of the function value, corresponding to two times the positive and negative standard deviations. If the confidence interval does not include zero, it suggests that the response is statistically significant at the chosen confidence level (($\pm$ 2 standard errors). The solid line represents the dynamic response process of the value, which reveals the change in $Arrivalsa_t$ under the impact of $Volumes_t$ and $\Delta Likes_t$, respectively. A value above 0 lines expresses the positive effect of $Volumes_t$ and $\Delta Likes_t$ to $Arrivalsa_t$, while a value below the 0 line signifies a negative effect.

As Figure 4.4 illustrates, $Arrivalsa_t$ exhibits a positive response to the impact of a unit change (i.e., a standard deviation) in $Volumes_t$. Throughout the response period, the relationship is positive, signifying that $Volumes_t$ has a positive correlation with $Arrivalsa_t$, albeit with a modest magnitude. Consequently, $Volumes_t$ can be considered a predictor of $Arrivalsa_t$,. Figure 4.4 also reveals that the duration of this positive response is relatively brief. More specifically, the positive response period is significant (at the 95% confidence level) within the first three months following the emergence of a shock to $Volumes_t$. Furthermore, the positive effect reaches its maximum impact in the second month after the shock to $Volumes_t$. This suggests that the influence of $Volumes_t$ on $Arrivalsa_t$ is most prominent during the short period, indicating a possible critical window for potential tourists to make decisions based on the information on Twitter. However, as time progresses beyond the second month, the positive effect gradually diminishes, indicating the influence of $Volumes_t$ on $Arrivalsa_t$weakens in the long term. In contrast, Figure 4.4 shows that $Arrivalsa_t$ has a generally negative response to the impact of a unit change in $\Delta Likes_t$. This negative response is only significant (at the 95% confidence level) in the fourth month following the emergence of a shock to $\Delta Likes_t$. The significance of the fourth month highlights that the effect of $\Delta Likes_t$ on $Arrivalsa_t$ occurs with a relatively short lag, suggesting that changes in the rate of $Likes_t$ can have an observable influence on tourist arrivals within a few months. Since $\Delta Likes_t$ denotes the first-order differenced value of $Likes_t$, the negative response process of $Arrivalsa_t$ to $\Delta Likes_t$ suggests that an increase in the rate of change of $Likes_t$ (i.e., the growth or decline in the number of Twitter likes) is associated with a decrease in tourist arrivals. In other words, when the change in the number of ‘likes’ becomes more volatile, the number of tourist arrivals tends to decrease.

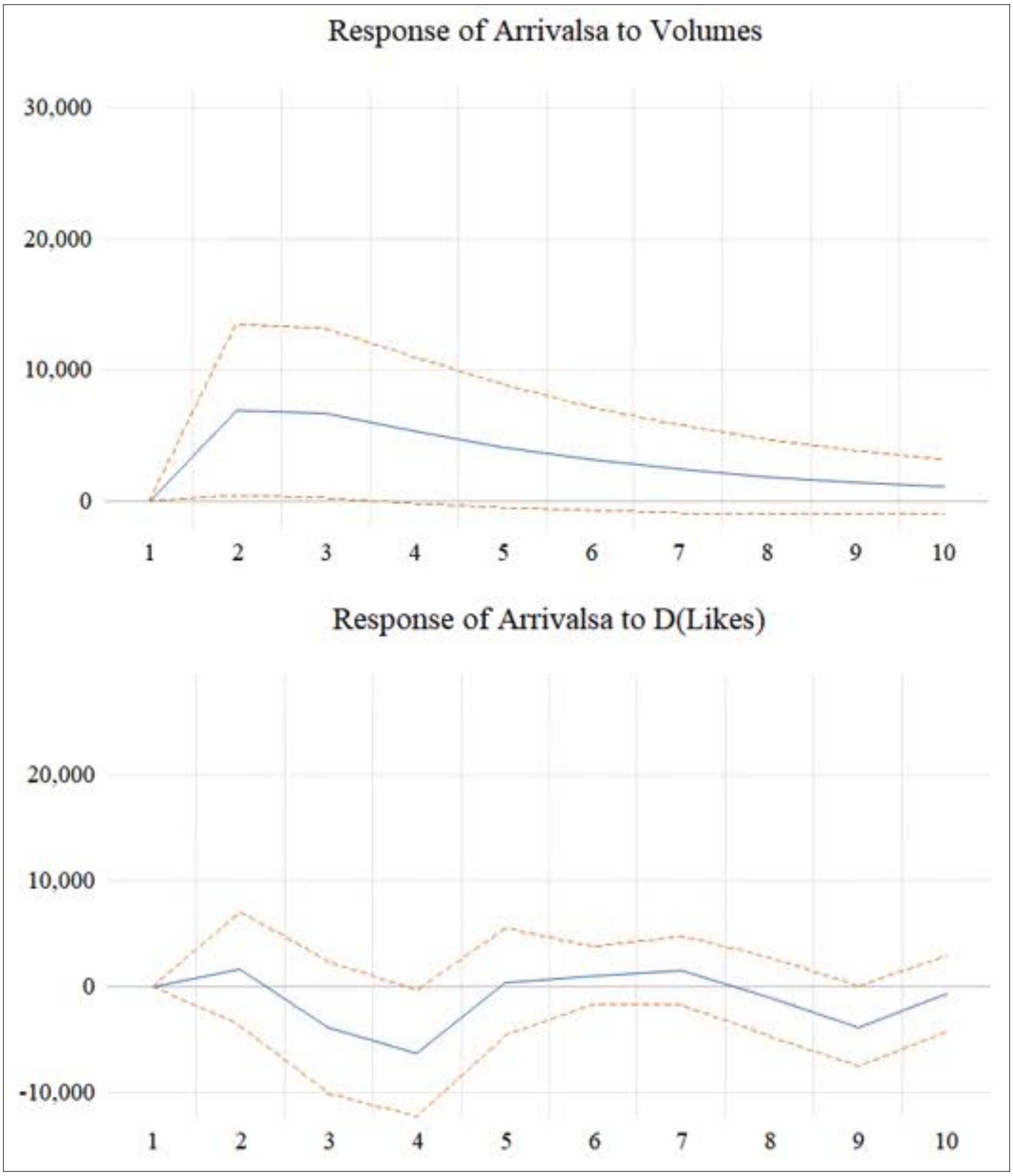

Figure 4.4 Results of impulse response analysis

Figure notes: 1. The figure depicts the response of $Arrivalsa_t$ to Cholesky One S.D. (d.f. adjusted) Innovations of $Volumes_t$ and $\Delta Likes_t$, respectively. 2. The two dotted lines represent the confidence interval (± 2 standard errors).

4.5 Findings

Although tourism demand forecasts are proven to be effective tools in developing crowd management strategies (Bi et al., 2020), existing studies mainly focus on the assessment of tourism demand for the destinations, overlooking the challenges faced by certain attractions (Bi et al., 2020; H. Li et al., 2020). Unlike destinations, tourist attractions are constrained by their physical capacity, yet the issue of managing attractions' physical capacity has been largely under-researched (Kim et al., 2022). As highlighted in the literature review, social media data presents several benefits to forecasting research, such as being information-rich, easily accessible, and generated in real-time. Consequently, an increasing number of studies have been utilizing social media data to forecast real-world outcomes (Asur & Huberman, 2010). However, studies employing social media data, particularly Twitter data, in tourism demand forecasting remain limited. Therefore, Study 1 aims to investigate the dynamics between tourist arrivals and Twitter-generated variables using the British Museum as a case study.

Applying the VAR model, Study 1 not only explores the Granger causal relationship between tourist arrivals and Twitter-generated variables but also examines the time horizon, direction, and magnitudes of these detected relationships. The main conclusions and implications of Study 1 are summarized as follows. First, it has been determined that the volume of tweets and the change in tweet likes Granger cause seasonally adjusted tourist numbers, suggesting that these Twitter-generated variables can serve as potential predictors for tourism demand forecasts. This finding shows that Twitter can be seen as a potential source for generating predictors at the attraction level. Second, the study reveals that the seasonally adjusted number of tourist arrivals exhibits a positive response to a unit change in the volume of tweets, with this positive response process being relatively brief. Study 1 also finds that when the change in the number of likes becomes more volatile, the number of tourist arrivals tends to decrease.

In summary, Chapter 4 (Study 1) tackles the first research objective of this thesis, which is to assess the role of Twitter data in tourism demand forecasting at the attraction level. Findings of Chapter 4 substantiate that Twitter data can be effectively utilized in generating predictors for tourism demand forecasts at the attraction level. This chapter further encourages ongoing exploration of predictors derived from Twitter data, with a focus on their capacity to enhance the accuracy of demand forecasts for tourist attractions.

Chapter 5: The role of Twitter communication flows in generating improved forecasts

5.1 Overview

Building on Study 1, Study 2 aims to address these gaps to generate improved tourism demand forecasts using Twitter-generated variables. Chapter 3.5.2 offers an in-depth description of the methodological design employed to achieve this research goal. To facilitate a comprehensive understanding of the research methodology, we present the framework for the empirical analysis in Study 2 through Figure 5.1.

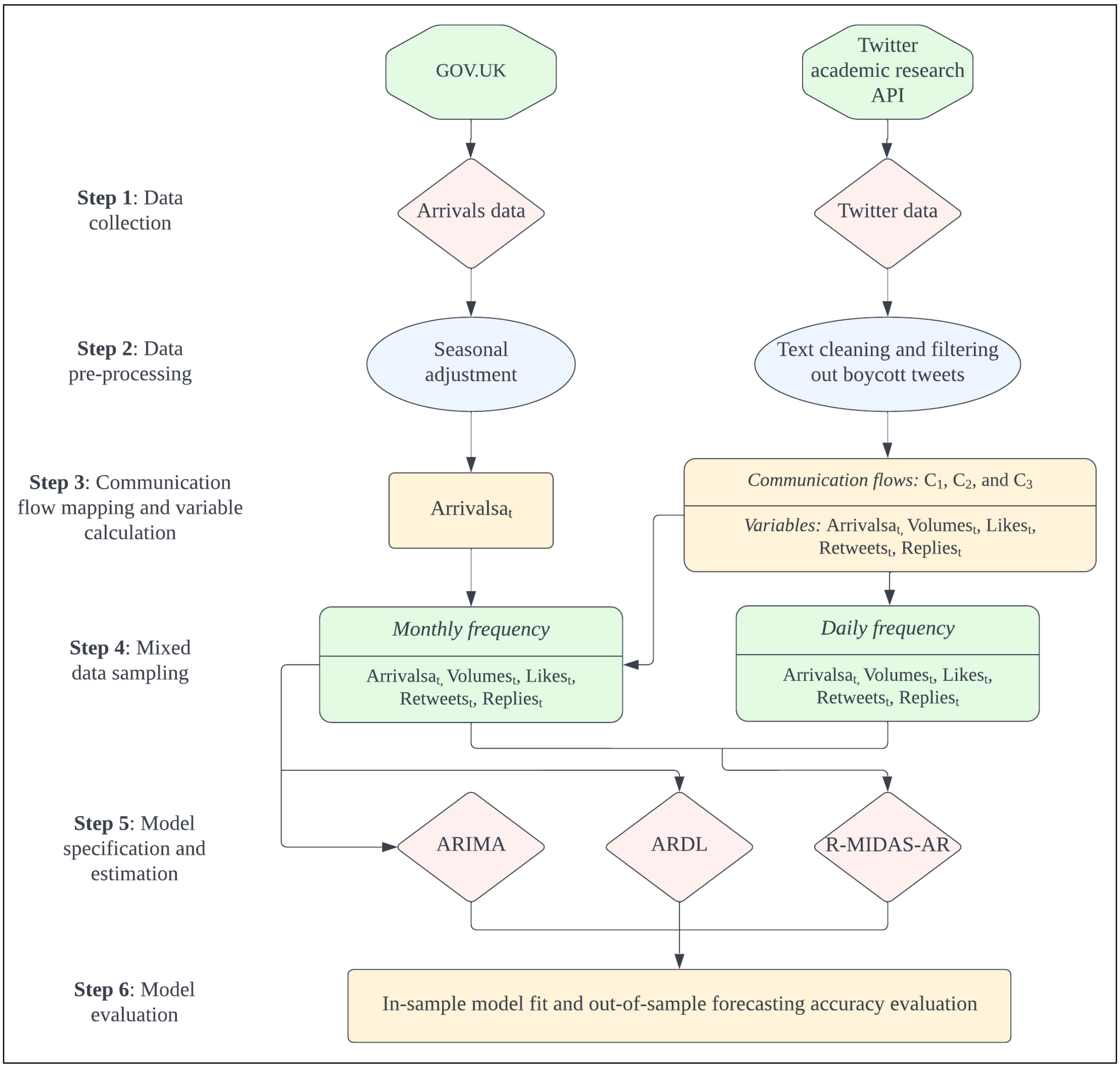

Figure 5.1 Empirical framework of Study 2

This framework comprises six steps: (1) data collection; (2) data pre-processing; (3) communication flow mapping and variable calculation; (4) mixed data sampling; (5) model specification and estimation; and (6) model evaluation. Steps 5 and 6 concentrate on the forecasting models that were estimated and evaluated in this study. We discuss the results and insights derived from each stage in detail below.

5.2 Data collection

Based on the identification of "irrelevant tweets" in Section 3.5.2.1.1, we first filter out tweets not related to the case of the British Museum. To achieve this, we use regular expressions to screen each text. Specifically, we filter the "text" column to retain only tweets containing the phrase "British Museum" (i.e., str.contains(r'\british museum\b')) and tweets posted by the official account of the British Museum. Considering the fact the following data pre-processing process with imply several text analysis procedures, all tweets written in languages other than English are also filtered out to avoid biased interpretation of textual content. This process filtered out $139,975$ tweets and left $582,742$ tweets in the dataset. The factor that causes "noise" will be addressed in Section 5.3.2.

5.3 Data processing and transformation

5.3.1 Tourist arrivals data

Visual inspection (see Figure 5.2**)** indicates that there is a strong seasonal variation in $Arrivals_t$. This is expected as the tourism sector is characterised by strong seasonal patterns. To preclude any potential distortions to the subsequent analyses stemming from the seasonal patterns, in line with Gunter and Önder (2016), $Arrivals_t$ is then seasonally adjusted through a moving average filter[9]. This process generates a new variable, $Arrivalsa_t$, which is defined as the seasonally adjusted series of monthly tourist arrivals to the British Museum.

Of note, we do not include differences in the data pre-processing procedure as we are interested in the short-term effects of the high-frequency Twitter data on tourism demand and not the long-term cointegration relationship. Following Önder et al. (2019), we also do not take natural logarithms as models built in natural logarithms provide an inferior in-sample model fit. To

[9] These and all subsequent calculations are carried out with EViews Version 11.

ensure stability and to avoid spurious regression sequence, we employed the ADF test with automatic optimal lag selection by applying the AIC to statistically evaluate if unit roots exist in the seasonally adjusted arrivals data. ADF tests allow for trends and intercepts to be calculated for $Arrivalsa_t$. The test results suggest that the null hypothesis of the ADF test is rejected at the 5% significance level with a t-statistic of -3.800 and a p-value of 0.022 (the 5% test critical value is -3.47). Consequently, it is confirmed that $Arrivalsa_t$ is stationary, and there is no need to further detrend the data by taking first differences or logarithms.

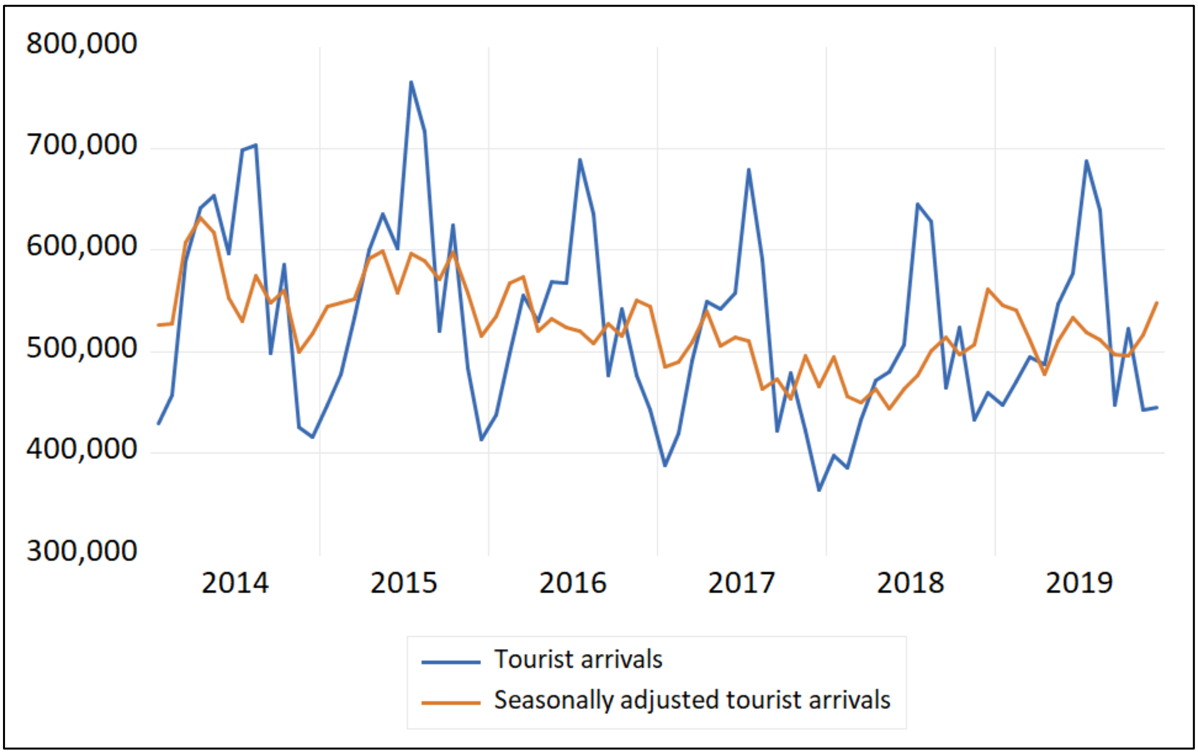

Figure 5.2 Monthly tourist arrivals to the British Museum: $Arrivals_t$

5.3.2 Twitter data and explanatory variables

As is discussed in the literature review, potential predictors we are interested in are $Volumes_t$, $Likes_t$, $Replies_t$ and $Retweets_t$. To meet the requirements of selected forecasting models (discussed in the next subsection), we sample these predictors in both daily and monthly frequencies by simple summation and then conducted an ADF test for each variable (refer to Table 5.1 for test results). We have observed that $Likes_t$ is not stationary at the 5% significance level. Further analysis shows that the majority of highly-liked tweets relate to negative attitudes towards the British Museum's exhibition of contested artefacts and BP sponsorships, which are not related to visitors' experiences or their intention to visit the museum (refer to Appendix 1 for a detailed description of the identification process). Therefore, these tweets are identified as noise, which may bias the forecasts and be discarded before the process of predictors construction. As a result, the adjusted dataset includes *563,301* tweets spanning 1 January 2014 to 31 December 2019. Figure 5.3 and Figure 5.4 visualise the adjusted variables $Volumes_t$, $Likes_t$, $Replies_t$ and $Retweets_t$. Although surges still exist, the series looks smoother than the unadjusted series. The unit root test employing the ADF examined method is used for adjusted variables. Based on the visualisation of each variable, we allow trend and intercept for monthly aggregated Likes_t, while allowing intercept only for monthly aggregated volumes and retweets. From Table 5.2, the test results show that almost all variables are stationary at the order $I(0)$ at the 1% significant level, except for monthly aggregated $Replies_t$, which is stationary around the 10% significant level. Table 5.3 provides descriptive statistics for Twitter-generated variables.

Table 5.1 Results of the ADF test for Twitter-generated variables sampled at different frequencies

Variables	Monthly frequency		Daily frequency	
	t-statistics	Probability	t-statistics	Probability
$\boldsymbol{Volumes_t}$	-5.732	0.000	-33.881	0.000
$\boldsymbol{Likes_t}$	-0.973	0.941	-11.024	0.000
$\boldsymbol{Retweets_t}$	-6.490	0.000	-44.761	0.000
$\boldsymbol{Replies_t}$	-2.616	0.095	-18.375	0.000

Source: the UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Note: Trend and intercept are included in the test equation for monthly aggregated Likes_t. For all other variables, only the intercept is included in the test equation.

Table 5.2 Results of the ADF test for Twitter-generated variables sampled at different frequencies (adjusted data)

Variables	Monthly frequency		Daily frequency	
	t-statistics	Probability	t-statistics	Probability
$\boldsymbol{Volumes_t}$	-5.383	0.000	-34.129	0.000
$\boldsymbol{Likes_t}$	-7.822	0.000	-19.160	0.000
$\boldsymbol{Retweets_t}$	-3.541	0.010	-20.450	0.000
$\boldsymbol{Replies_t}$	-2.715	0.077	-19.102	0.000

Source: the UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Note: Trend and intercept are included in the test equation for monthly aggregated $Likes_t$. For all other variables, only the intercept is included in the test equation.

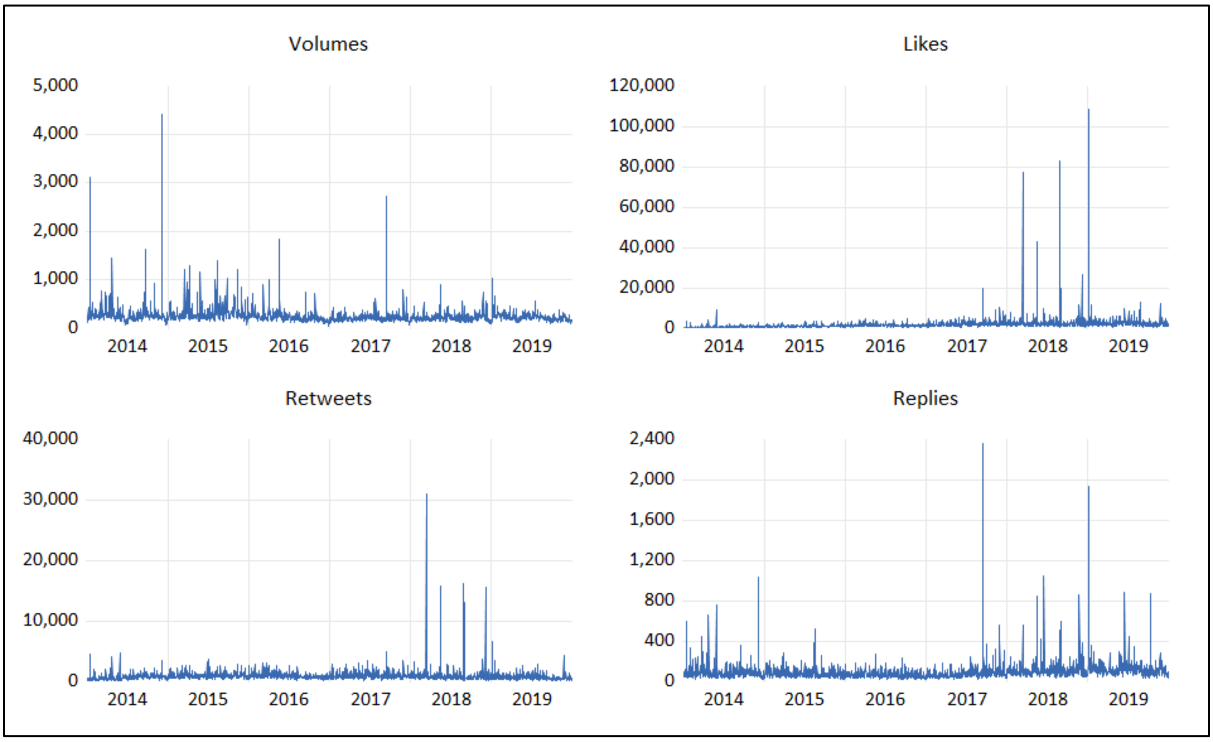

Figure 5.3 Visualisation of Twitter-generated variables sampled at a daily frequency (adjusted)

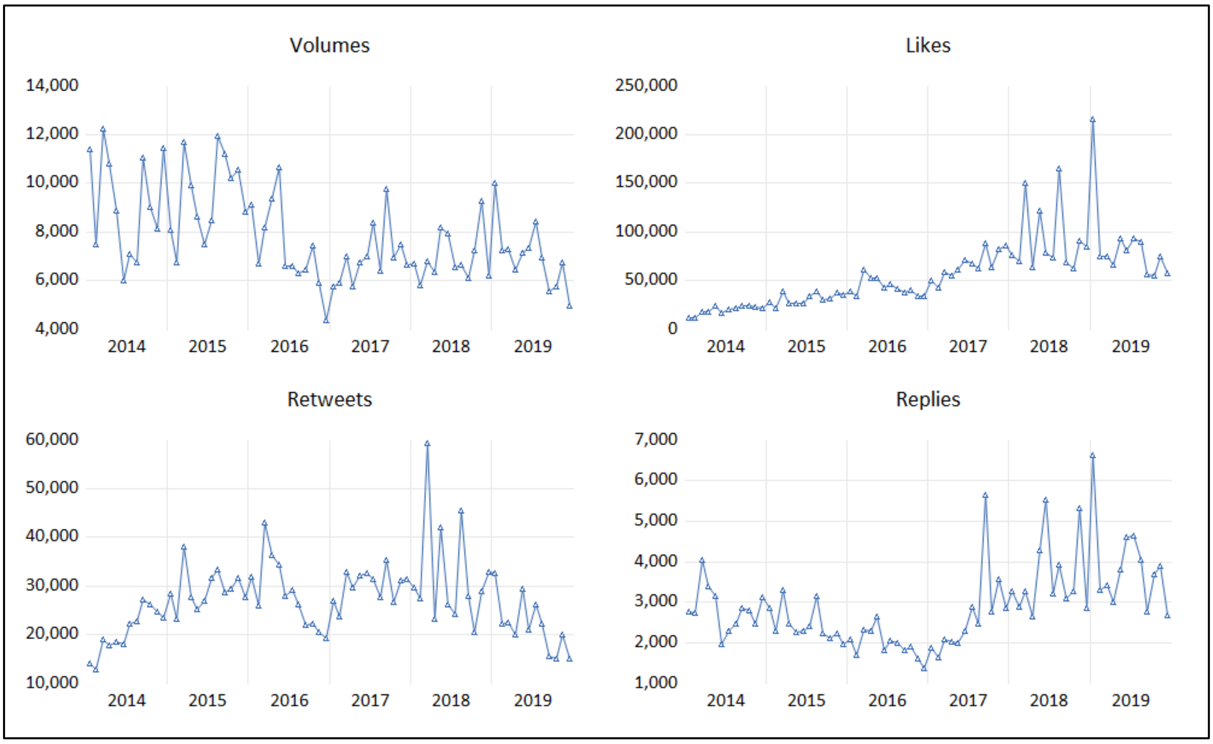

Figure 5.4 Visualisation of Twitter-generated variables sampled at a monthly frequency (adjusted)

Table 5.3 Descriptive statistics of variables

Variable	Description	Period	Frequency	Max	Min	Mean	Std. Dev.
$\boldsymbol{Arrivals_t}$	Monthly volume of tourist arrivals to the British Museum at time t.	Jan 2014 – Dec 2019	Monthly	765877	363281	526110.1	94264.12
$\boldsymbol{Arrivalsa_t}$	Seasonally-adjusted monthly volume of tourist arrivals to the British Museum at time t, adjusted through a moving average filter.	Jan 2014 – Dec 2019	Monthly	631862.6	443695.4	526110.1	42082.99
$\boldsymbol{Volumes_t}$	Monthly volume of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	12245	4379	7823.63	1851.79
$\boldsymbol{Likes_t}$	Monthly volume of likes of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	215544	11189	56369.38	35639.03
$\boldsymbol{Retweets_t}$	Monthly volume of retweets of tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	59359	12912	27081.83	7650.38
$\boldsymbol{Replies_t}$	Monthly volume of replies to tweets fetched from Twitter with the term "British Museum" at time t.	Jan 2014 – Dec 2019	Monthly	6620.00	1369.00	2908.99	1014.97
$\boldsymbol{Volumes_t}$	Daily volume of tweets fetched from Twitter with the term "British Museum" at time t.	1 Jan 2014 – 31 Dec 2019	Daily	4411	47	257.10	179.41
$\boldsymbol{Likes_t}$	Daily volume of likes of tweets fetched from Twitter with the term "British Museum" at time t.	1 Jan 2014 – 31 Dec 2019	Daily	108655	57	1852.394	3743.258
$\boldsymbol{Retweets_t}$	Daily volume of retweets of tweets fetched from Twitter with the term "British Museum" at time t.	1 Jan 2014 – 31 Dec 2019	Daily	30909	55	889.96	1044.34
$\boldsymbol{Replies_t}$	Daily volume of replies to tweets fetched from Twitter with the term "British Museum" at time t.	1 Jan 2014 – 31 Dec 2019	Daily	2353	14	95.59	98.23

Source: UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.

Before further data transformation, we first estimate vector autoregressive (VAR) models and conduct Granger causality tests to investigate the dynamics between $Arrivalsa_t$ and our explanatory variables, including the $Volumes_t$, $Likes_t$, $Retweets_t$ and $Replies_t$. The aim is to examine if findings from Study 1 still hold for these variables generated from the pre-processed (adjusted) Twitter dataset. We applied the same modelling and testing strategy as in Study 1, with the exception that we did not include differenced variables in the test, as all variables were found to be stationary. The test results are presented in Appendix 2 Table A2. The results indicate that there is bidirectional causality between $Volumes_t$ and $Arrivalsa_t$ with a lag order of 1 at the 5% significance level. Therefore, it aligns with Study 1 that it is reasonable to consider $Volumes_t$ as a potential predictor of $Arrivalsa_t$. Unlike Study 1, Study 2 finds evidence that $Retweets_t$ does Granger cause $Arrivalsa_t$ with a lag order of 2 at the 5% significance level. This suggests that $Retweets_t$ could potentially serve as an efficient predictor of $Arrivalsa_t$, and warrants further investigation. Granger causality is not detected between $Arrivalsa_t$ and $Likes_t$ or between $Arrivalsa_t$ and $Replies_t$. We note that all variables used in the Granger causality test are sampled at a monthly frequency. Temporal aggregation of Twitter-generated variables may cause the loss of high-frequency features in the data (Önder et al., 2019) and bias model estimation (Hu et al., 2022). To address this issue in our previous analysis, as well as in the analysis conducted in Study 1, we employ mixed-frequency data sampling to conduct a more robust examination of the results from the Granger causality test.

As discussed in Chapters 2 and 3, we further calculate these Twitter-generated variables based on three identified three communication flows, including direct communication with attractions: tourists (i.e., Twitter users) tweeting to the British Museum (C1), peer-to-peer conversations: tourists tweeting about the British Museum (C2), and communication from attractions to the broader audience: the British Museum tweeting to tourists (C3). Performing content analysis of the dataset, we placed each tweet in the most appropriate of these three communication flows based on the text content. Tweets mentioning the British Museum's official Twitter account were placed into the C1 category. Each tweet posted by individuals other than the British Museum's official Twitter account that contained the term 'British Museum' in its text content was placed into the C2 category. Finally, tweets posted by the British Museum's official Twitter account were placed into the C3 category. Table 5.4 shows the classification results and gives examples of tweets in each category. We then developed four Twitter-generated variables for each category and performed the ADF test for each variable. The test results are summarised in Table 5.5. From Table 5.5, all daily-aggregated variables are stationary at the 1% significant level. Almost all monthly-aggregated variables are stationary at the 5% significant level, with the following exceptions. Including trend and intercept in the test equation, the monthly-aggregated $Retweets_t$ for C1 and C3 are stationary at the 10% significant level. Notably, the monthly-aggregated $Likes_t$ is not stationary at the 10% significant level. We further took the first-order difference on the data, and the differenced variable is stationary at the 1% significant level.

Table 5.4 Classification of tweets based on communication flows

Tweet Category	Communication flow	Amount	Example (text content)
C1	Individuals talk to the British Museum	288,465	hey @britishmuseum give us your best duck
C2	Individuals talk about the British Museum	263,035	Remarkable Harry Kane exhibition at the British Museum
C3	The British Museum talks to individuals	11,677	We're excited to announce our next major show, opening in 2019! #MangaExhibition will explore the phenomenon of manga, and will be the largest exhibition of the art form ever to take place outside of Japan.

Note: 124 tweets are not categorised in any of the above three categories. This is due to errors that occur in the data encoding process. In these tweets, "British Museum" is encoded as "ritish Museum" and hence these tweets are not labelled by the clustering algorithm.

Table 5.5 Results of ADF test for Twitter-generated variables sampled for each category

Category	Frequency	Variable	t-statistics	Probability	5% Critical value	Note
C_1	Monthly	$Volumes_t$	-4.100	0.002	-2.904	/
C_1	Monthly	$Likes_t$	-3.370	0.015	-2.904	/
C_1	Monthly	$Retweets_t$	-3.359	0.066	-3.475	Intercept and trend are included in the test equation
C_1	Monthly	$Replies_t$	-3.516	0.045	-3.475	Intercept and trend are included in the test equation
C_1	Daily	$Volumes_t$	-22.022	0.000	-2.863	
C_1	Daily	$Likes_t$	-39.719	0.000	-2.863	
C_1	Daily	$Retweets_t$	-40.112	0.000	-2.863	
C_1	Daily	$Replies_t$	-36.367	0.000	-2.863	
C_2	Monthly	$Volumes_t$	-5.464	0.000	-2.903	
C_2	Monthly	$Likes_t$	-8.468	0.000	-3.474	Intercept and trend are included in the test equation
C_2	Monthly	$Retweets_t$	-8.509	0.000	-3.474	Intercept and trend are included in the test equation
C_2	Monthly	$Replies_t$	-3.240	0.085	-3.475	Intercept and trend are included in the test equation
C_2	Daily	$Volumes_t$	-11.112	0.000	-2.863	
C_2	Daily	$Likes_t$	-34.791	0.000	-2.863	
C_2	Daily	$Retweets_t$	-45.627	0.000	-2.863	
C_2	Daily	$Replies_t$	-45.971	0.000	-2.863	
C_3	Monthly	$Volumes_t$	-5.559	0.000	-3.475	Intercept and trend are included in the test equation
C_3	Monthly	$Likes_t$	-2.132	0.519	-3.475	Intercept and trend are included in the test equation
C_3	Monthly	$d(Likes_t)$	-12.72479	0.000	-3.475	Intercept and trend are included in the test equation
C_3	Monthly	$Retweets_t$	-3.167	0.100	-3.475	Intercept and trend are included in the test equation
C_3	Monthly	$Replies_t$	-6.618	0.000	-3.474	Intercept and trend are included in the test equation
C_3	Daily	$Volumes_t$	-10.023	0.000	-2.863	
C_3	Daily	$Likes_t$	-12.033	0.000	-2.863	
C_3	Daily	$Retweets_t$	-11.798	0.000	-2.863	
C_3	Daily	$Replies_t$	-13.831	0.000	-2.863	

Source: the UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.

5.4 Empirical results

In total, 21 models were estimated, including the benchmark ARIMA model, 4 ARDL models and 16 R-MIDAS-AR models. To evaluate the forecasting performance of these models, five groups of comparisons were conducted in two phases (see Figure 5.5).

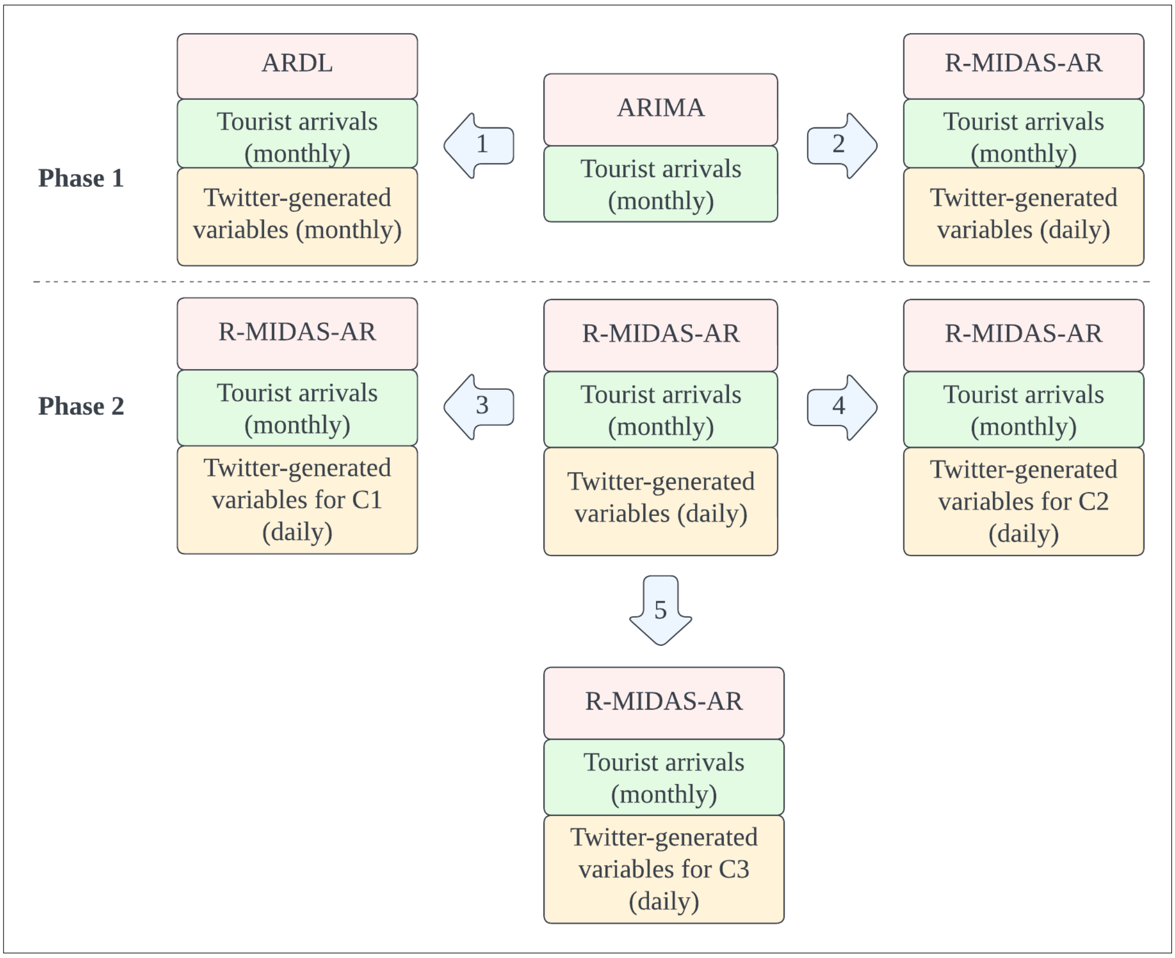

Figure 5.5 Evaluation framework

In phase 1, the benchmark ARIMA model based on historical tourist arrivals data sampled at monthly frequency is compared with four ARDL models (comparison group 1) and four R-MIDAS-AR models (comparison group 2) estimated with Twitter variables generated from the entire Twitter dataset. ARDL models are based on both historical tourist arrivals data and Twitter-generated variables sampled at monthly frequency. R-MIDAS-AR models are based on historical tourist arrivals data sampled at a monthly frequency and Twitter-generated variables sampled at a daily frequency. The aim of phase 1 is to investigate if Twitter-generated variables are relevant predictors of tourist arrivals in the short term and, if so, whether they can improve the accuracy of short-term tourist arrivals forecasts.

In-sample model fits and pseudo-out-of-sample forecasting accuracy for each model estimated in phase 1 are presented in Table 5.6. According to Table 5.6, estimates for the two dynamic model classes feature an improved in-sample model fit compared to the benchmark ARIMA model, as judged by high-adjusted R squared values and low SC values. Notably, four R-MIDAS-AR models present an improved in-sample model fit over the benchmark ARIMA model and four ARDL models. This meets our expectations as high-frequency Twitter-generated variables are assumed to have the ability to explain the fluctuation of tourist arrivals. Regarding the comparison of forecasting performance, the calculated RMSE, MAE, MAPE and RMSPE measures for the three comparison groups revealed that: (1) the ARDL model class does not generate more precise forecasts compared to the benchmark ARIMA model; and (2) forecasts generated by the R-MIDAS-AR model class significantly outperform the benchmark ARIMA model and the ARDL model class (refer to Table 5.8). Notably, the best pseudo-out-of-sample one-step-ahead static forecasts were produced by the R-MIDAS-AR (1, 50) model based on $Volumes_t$ generated from the entire Twitter dataset (see Figure 5.6). Overall, the comparison results in phase 1 suggest that: (1) Twitter-generated variables are relevant predictors of tourist arrivals in the short term; (2) mixed data sampling can improve the accuracy of short-term tourist arrivals forecasts. Therefore, we only incorporated the R-MIDAS-AR model class in phase 2 model estimation and comparison.

Phase 2 investigates whether mapping the communication flows for attraction-specific tweets can improve the accuracy of short-term tourist arrivals forecasts. In phase 2, 16 R-MIDAS-AR models were estimated, including four R-MIDAS-AR models for the entire Twitter dataset and four R-MIDAS-AR models for each tweet category (C1, C2 and C3). In-sample model fits and pseudo-out-of-sample forecasting accuracy for each model estimated in phase 2 are presented in Table 5.7. According to Table 5.7, all estimates feature a good in-sample model fit, as judged by improved high-adjusted R squared values and low SC values compared to the benchmark ARIMA model. Regarding the comparison of forecasting performance, the calculated MAE, RMSE, MAPE, and RMSPE measures for the three comparison groups reveal that predictors generated from the tweets belonging to the C1 subset perform best in comparison to C2 and

C3 (refer to Table 5.9). Notably, the best pseudo-out-of-sample one-step-ahead static forecasts were produced by the R-MIDAS-AR (1, 50) model based on $Volumes_t$ generated from tweets belonging to the C1 subset (see Figure 5.7). Meanwhile, the R-MIDAS-AR (1, 50) model based on $Likes_t$ generated from tweets belonging to the C1 subset also showed improved accuracy over the R-MIDAS-AR (1, 50) model based on $Volumes_t$ generated from the whole Twitter dataset. The results suggest that: (1) high-frequency Twitter-generated variables are relevant predictors of tourist arrivals in the short term; and (2) mapping the communication flow for attraction-specific tweets can improve the accuracy of short-term tourist arrival forecasts, particularly incorporating tweets that depict tourists' direct communication with the attraction (C1).

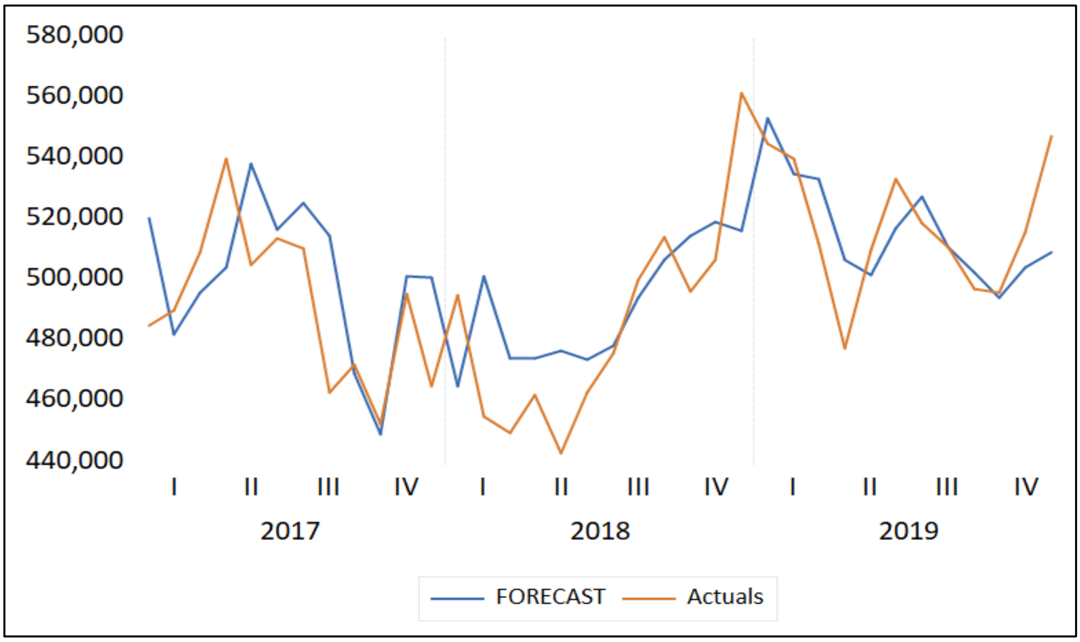

Figure 5.6 Pseudo-out-of-sample one-step-ahead static R-MIDAS-AR (1, 50) forecasts for tourist arrivals at the British Museum from Jan 2017 to Dec 2019 based on predictors generated from the whole Twitter dataset ($Volumes_t$)

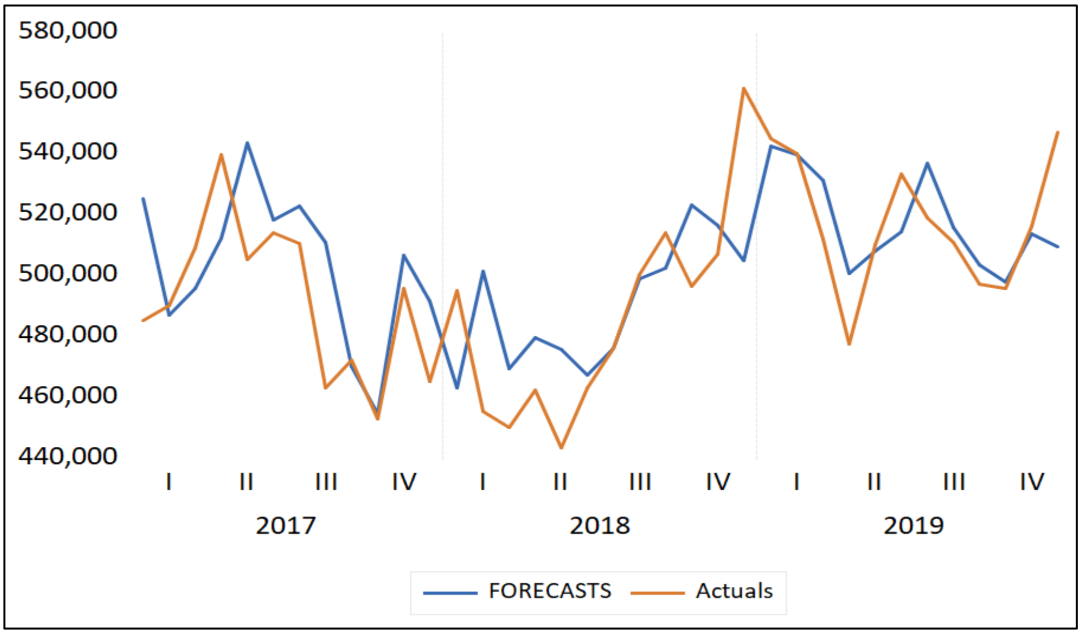

Figure 5.7 Pseudo-out-of-sample one-step-ahead static R-MIDAS-AR (1, 50) forecasts for tourist arrivals at the British Museum from Jan 2017 to Dec 2019 based on predictors generated from the tweets belonging to the C1 subset ($C1\_Volumes_t$)

Table 5.6 In-sample model fit and pseudo-out-of-sample one-step-ahead forecast evaluation (*phase 1*)

No.	Model	Exogenous predictors	P-values	Adjusted R^2	SC	RMSE	MAE	MAPE	SMAPE
1	ARIMA (1, 0, 1)	/	/	0.56	23.51	25047.60	19877.01	4.04	3.97
2	ARDL (1, 0)	$Volumes_t$	0.0524	0.59	***23.38***	24841.56	20242.78	4.10	4.05
3	ARDL (1, 0)	$Replies_t$	0.5670	0.56	23.45	/	/	/	/
4	ARDL (1, 0)	$Rewteets_t$	0.3681	0.57	23.45	/	/	/	/
5	ARDL (1, 0)	$Likes_t$	0.0697	0.58	23.41	25291.12	20430.52	4.13	4.08
6	R-MIDAS-AR (1, 50)	$Volumes_t$	All < 0.05	***0.64***	23.41	***23030.97***	***17891.14***	***3.62***	***3.58***
7	R-MIDAS-AR (1, 14)	$Replies_t$	All >0.05	0.62	23.47	/	/	/	/
8	R-MIDAS-AR (1, 27)	$Rewteest_t$	All > 0.1	0.59	23.56	/	/	/	/
9	R-MIDAS-AR (1, 49)	$Likes_t$	All < 0.05	0.61	23.49	23767.38	18664.63	3.76	3.72

Source: the UK government's official website, Twitter Inc., and own calculations using EViews Version 11.

Table notes: 1. ARDL models with insignificant (i.e., below the 10% significant level) exogenous predictors did not enter the pseudo-out-of-sample forecasting process. 2. R-MIDAS-AR models with insignificant exogenous predictors (i.e., all PDL lags are below the 10% significant level) did not enter the pseudo-out-of-sample forecasting process. 3. The bold and italic numbers are the best performance compared with other models. 4. Estimation results for each model can be provided on request.

Table 5.7 In-sample model fit and pseudo-out-of-sample one-step-ahead forecast evaluation (*phase 2*)

No.	Model	Exogenous predictors	P-values	Adjusted R^2	SC	RMSE	MAE	MAPE	SMAPE
1	R-MIDAS-AR (1, 50)	$Volumes_t$	All < 0.05	0.64	23.41	23030.97	17891.14	3.62	3.58
2	R-MIDAS-AR (1, 14)	$Replies_t$	All >0.05	0.62	23.47	/	/	/	/
3	R-MIDAS-AR (1, 27)	$Rewteest_t$	All > 0.1	0.59	23.56	/	/	/	/
4	R-MIDAS-AR (1, 49)	$Likes_t$	All < 0.05	0.61	23.49	23767.38	18664.63	3.76	3.72
5	R-MIDAS-AR (1, 50)	$C1\_Volumes_t$	All < 0.05	***0.66***	***23.34***	***23264.45***	***17329.86***	***3.50***	***3.46***
6	R-MIDAS-AR (1, 16)	$C1\_Replies_t$	All > 0.1	0.61	23.50	/	/	/	/
7	R-MIDAS-AR (1, 51)	$C1\_Rewteets_t$	All < 0.05	0.62	23.48	22773.50	18356.12	3.71	3.66
8	R-MIDAS-AR (1, 50)	$C1\_Likes_t$	All < 0.05	0.61	23.49	22379.41	17487.84	3.56	3.51
9	R-MIDAS-AR (1, 39)	$C2\_Volumes_t$	All > 0.1	0.61	23.49	/	/	/	/
10	R-MIDAS-AR (1, 40)	$C2\_Replies_t$	All > 0.1	0.58	23.56	/	/	/	/
11	R-MIDAS-AR (1, 23)	$C2\_Retweets_t$	All > 0.1	0.58	23.58	/	/	/	/
12	R-MIDAS-AR (1, 27)	$C2\_Likes_t$	All > 0.1	0.57	23.58	/	/	/	/
13	R-MIDAS-AR (1, 50)	$C3\_Volumes_t$	All < 0.05	0.64	23.42	24068.15	19174.25	3.88	3.83
14	R-MIDAS-AR (1, 19)	$C3\_Replies_t$	All < 0.05	0.64	23.41	22994.42	18577.86	3.78	3.73

15	R-MIDAS-AR (1, 27)	$C3\_Rewteets_t$	$PDL_{2 \text{ to } 4}$ < 0.1	0.60	23.52	22557.55	18309.59	3.73	3.67
16	R-MIDAS-AR (1, 26)	$C3\_Likes_t$	All > 0.1	0.63	23.45	/	/	/	/

Source: the UK government's official website, Twitter Inc., and own calculations using EViews Version 11.
Table notes: 1. ARDL models with insignificant (i.e., below the 10% significant level) exogenous predictors did not enter the pseudo-out-of-sample forecasting process. 2. R-MIDAS-AR models with insignificant exogenous predictors (i.e., all PDL lags are below the 10% significant level) did not enter the pseudo-out-of-sample forecasting process. 3. The bold and italic numbers are the best performance compared with other models. 4. Estimation results for each model can be provided on request.

Table 5.8 Pseudo-out-of-sample one-step-ahead forecast comparison (*phase 1*)

Indicator	*Benchmark*	*Comparison*							
	ARIMA	ARDL (1, 0) with $Volumes_t$	Improvement	ARDL (1, 0) with $Likes_t$	Improvement	R-MIDAS-AR (1, 50) with $Volumes_t$	Improvement	R-MIDAS-AR (1, 50) with $Likes_t$	Improvement
RMSE	25048	24841.56	0.82%	25291.12	-0.97%	23030.97	*8.05%*	23767.38	5.11%
MAE	19877	20242.78	-1.84%	20430.52	-2.78%	17891.14	*9.99%*	18664.63	6.10%
MAPE	4.04	4.1	-1.49%	4.13	-2.23%	3.62	*10.40%*	3.76	6.93%
SMAPE	3.97	4.05	-2.02%	4.08	-2.77%	3.58	*9.82%*	3.72	6.30%

Table notes: 1. The bold and italic numbers are the best performance compared with other models. 2. A negative improvement value implies that the former model provides a better forecast than the latter model in each comparison.

Table 5.9 Pseudo-out-of-sample one-step-ahead forecast comparison (*phase 2*)

Indicator	Benchmark	Comparison							
	R-MIDAS-AR (1, 50) with $Volumes_t$	R-MIDAS-AR (1, 49) with $Likes_t$	Improvement	R-MIDAS-AR (1, 50) with $C1\_Volumes_t$	Improvement	R-MIDAS-AR (1, 51) with $C1\_Rewteets_t$	Improvement	R-MIDAS-AR (1, 50) with $C1\_Likes_t$	Improvement
MAE	23030.97	23767.38	-3.20%	23264.45	-1.01%	22773.5	1.12%	22379.41	***2.83%***
RMSE	17891.14	18664.63	-4.32%	17329.86	***3.14%***	18356.12	-2.60%	17487.84	2.25%
MAPE	3.62	3.76	-3.87%	3.5	***3.31%***	3.71	-2.49%	3.56	1.66%
SMAPE	3.58	3.72	-3.91%	3.46	***3.35%***	3.66	-2.23%	3.51	1.96%

Indicator	Benchmark	Comparison					
	R-MIDAS-AR (1, 50) with $Volumes_t$	R-MIDAS-AR (1, 50) with $C3\_Volumes_t$	Improvement	R-MIDAS-AR (1, 19) with $C3\_Replies_t$	Improvement	R-MIDAS-AR (1, 27) with $C3\_Rewteets_t$	Improvement
MAE	23030.97	24068.15	-4.50%	22994.42	0.16%	22557.55	2.06%
RMSE	17891.14	19174.25	-7.17%	18577.86	-3.84%	18309.59	-2.34%
MAPE	3.62	3.88	-7.18%	3.78	-4.42%	3.73	-3.04%
SMAPE	3.58	3.83	-6.98%	3.73	-4.19%	3.67	-2.51%

Table notes: 1. The bold and italic numbers are the best performance compared with other models. 2. A negative improvement value implies that the former model provides a better forecast than the latter model in each comparison

5.5 Findings

While tourism demand forecasting studies have recognised the importance of internet data for enhancing forecasting model performance (J. Li et al., 2018), contemporary research has primarily investigated the capacity of low-frequency social media data to yield improved tourism demand forecasts (Li et al., 2021). Only a select few (Havranek & Zeynalov, 2021; Hu et al., 2022; Önder et al., 2019) have leveraged the MIDAS approach to delve into the potential of high-frequency social media data for the progression of tourism demand forecasting research. Moreover, although it has been proved in Study 1 that Twitter contains valuable informative signals that informs tourists' preferences and intentions (Bigné et al., 2019; J. Li et al., 2018), Twitter communication flow's value in signalling tourist arrivals has not been fully recognised in the tourism context.

Addressing these research gaps and exploring the case study of the British Museum, Study 2 reveals the value of Twitter data in generating enhanced tourist arrivals forecasts for attractions. Using the Granger causality test, we show that the volume of tweets and retweets reflect tourist arrivals to an attraction. Additionally, findings of Study 2 indicate that Twitter-generated variables can serve as predictors of tourist arrivals to an attraction in the short term. Notably, prior to entering rival forecasting models, we utilized a state-of-the-art keyword-in-context search approach to detect and eliminate the noise embedded in the input Twitter data. It has been identified that the majority of these tweets are associated with boycott topics (refer to Appendix 1).

Moreover, Study 2 establishes that the interval between individuals' travel-planning behaviour and subsequent visiting behaviour is short, typically spanning one to two months. Capitalising on this discovery, we employed a mixed data sampling approach to resample potential predictors. The outcome demonstrates that the R-MIDAS-AR model is capable of incorporating high-frequency Twitter-derived variables to enhance forecast accuracy.

Lastly, Study 2 demonstrates that mapping the communication flows for attraction-specific tweets can further improve the accuracy of short-term tourist arrivals forecasts. In particular, the estimating results of various R-MIDAS-AR models show that the volume of tweets that reflecting direct communications between the tourists and the attraction (C1) generates the best forecast. Additionally, it is worth noting that "Likes" of these tweets also perform well when they depict tourists' direct communications with the attraction.

Taken together, Chapter 5 (Study 2) addresses the second research objective of this thesis, which is to examine the advantages of deciphering embedded communication flows on Twitter for generating improved tourism demand forecasts at the attraction level. The insights derived from Chapter 5 highlight the efficacy of Twitter data in crafting precise attraction-level demand forecasts. Specifically, the results underscore the significance of resampling potential predictors informed by the tripartite communication flows on Twitter as a strategy to further refine and bolster the accuracy of tourism demand forecasting.

Chapter 6: Re-examining Twitter "noise": the signalling value of boycott tweets on tourist arrivals

6.1 Data

As outlined in Section 3.5.3.2, we collected four types of data for the period between 1 January 2014 and 30 September 2022, including tourist arrivals data, Twitter data, important dates of COVID-19 travel restrictions, and dates of physical protests. The subsequent parts of this study will provide a detailed illustration of this data.

6.1.1 Tourist arrivals data

In alignment with Study 1 and Study 2, we obtained data illustrating the volume of tourist arrivals to the British Museum from the official website[10] of Department for Digital, Culture, Media and Sport of the United Kingdom (DCMS), which were used to generate the dependent variable $Arrivals_t$. $Arrivals_t$ denotes the number of tourist arrivals at time t. $Arrivals_t$ spans 1 January 2014 to 30 September 2022 and is sampled at a monthly frequency. Visual inspection (see Figure 6.1) shows a clear seasonal pattern, indicating the need for seasonal adjustment. Furthermore, there has been a noticeable drop in tourist arrivals since 2020, which can be attributed to travel restrictions imposed as a response to the COVID-19 pandemic.

[10] Tourist arrivals data can be accessed via the following link: https://www.gov.uk/government/statistical-data-sets/museums-and-galleries-monthly-visits

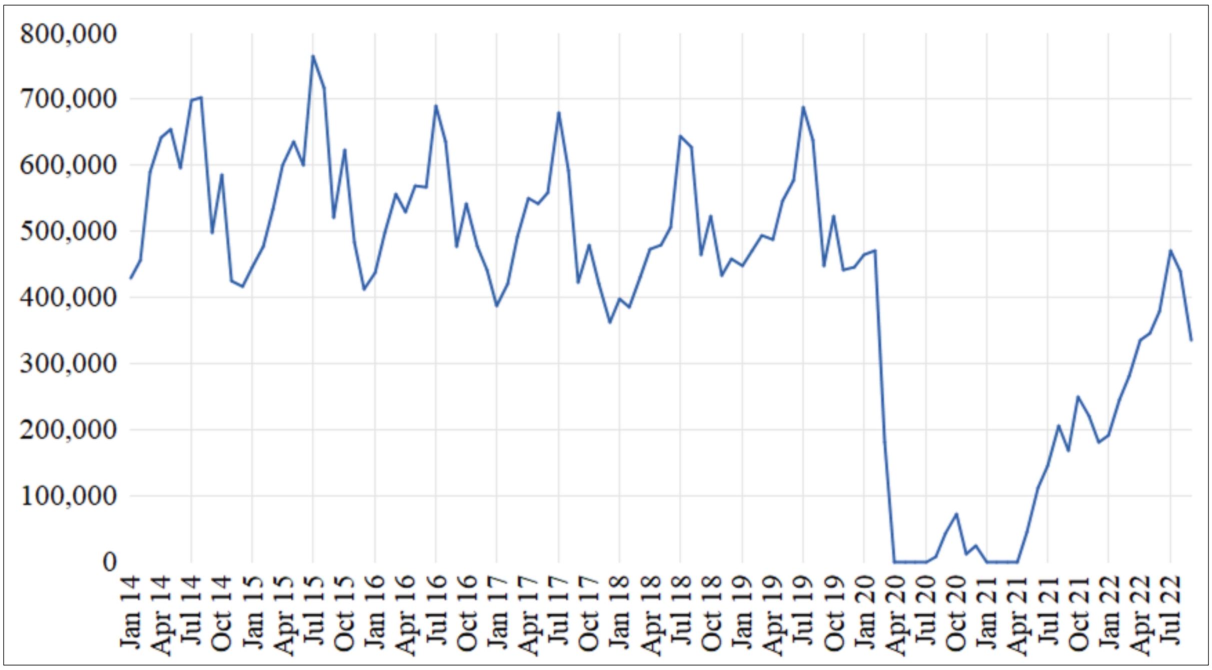

Figure 6.1 Monthly tourist arrivals to the British Museum: $Arrivals_t$

6.1.2 Twitter data

We collected Twitter data for the British Museum via the Twitter API. Considering that the British Museum is located in London and English is the most used language by tourists (Dergiades et al., 2018), this study is restricted to search queries and tweets written in English. Using the selected query keyword “British Museum”, tweets are continuously obtained and stored between 1 January 2014 and 31 December 2022, resulting in a dataset with a total of 837,630 tweets.

6.1.3 Key dates of UK government coronavirus lockdowns and measures

According to Biraglia and Gerrath (2020), in times of crisis, such as the COVID-19 pandemic, tourists to museums may be more open to the idea of corporate sponsorship. Therefore, we consider the effects of COVID-19 related travel restrictions in this research. More specifically, we have incorporated the dates when COVID-19 travel restrictions in the UK were imposed and later lifted, as summarised in Table 6.1. These dates have been identified from official announcements[11] made by the UK government.

[11] Government announcements regarding coronavirus lockdowns and measures can be accessed via the following link: https://www.gov.uk/coronavirus

Given the impact of the COVID-19 pandemic on tourist arrivals, as demonstrated by the steep decline in visitor numbers at the British Museum during the lockdown period (see Figure 6.1), we have deliberately omitted data from this period (March 2020 to June 2021) in the analyses, which includes variable computation and econometric evaluation. Our econometric exploration is predominantly focused on the pre-lockdown period, which extends from January 2014 to February 2020. On the other hand, the post-lockdown period, which stretches from August 2021 to September 2022, is considerably shorter. This period also exhibits a distinct trend, characterised by a clear pattern of recovery or growth. Due to these factors, implementing seasonal smoothing techniques on this limited dataset poses a significant challenge. Merging this data with pre-lockdown data may distort the outcomes of time-series modelling, as the post-lockdown period presents a unique and distinct trend. Therefore, it would be more judicious to analyse these two periods independently to ensure the robustness and validity of our results. This approach to data segmentation is designed to provide an exhaustive view of the variables of interest through econometric evaluation (Section 6.3) while concurrently acknowledging the exceptional circumstances of the pandemic (i.e., lockdown) through model-free evidence (Section 6.4).

Table 6.1 Key dates of UK government coronavirus lockdowns and measures

No	Date	Category	Description
1	2020/03/23	Lockdown/restrictions introduced	PM announces the first lockdown in the UK, ordering people to "stay at home".
2	2020/11/05	Lockdown/restrictions introduced	Second national lockdown comes into force in England
3	2021/01/06	Lockdown/restrictions introduced	England enters third national lockdown
4	2021/07/19	Lockdown/restrictions eased	Most legal limits on social contact were removed in England, and the final closed sectors of the economy reopened.

6.1.4 Dates of physical protests

To assess the relationship between physical protests and online activity of the British Museum, we took into account the dates of protests against these institutions. As discussed in Section 3.5.3.1, we identified the dates of protests against fossil fuel companies' sponsorship using newspaper articles accessed via the Lexis-Nexis database. The identified dates of protests against the British Museum were 2014/06/15, 2015/03/29, 2015/09/03, 2016/05/01, 2019/02/01, 2020/02/01, 2021/05/01, and 2022/03/21.

6.2. Model-free evidence

6.2.1 Revised "noise" identification

The process of identifying boycott tweets followed the methodological design discussed in Section 3.5.3.2. Here, we utilized tweets that mentioned the British Museum's official Twitter account (i.e., *382,619* tweets containing "@britishmuseum" in the column of "tweet") as input for the Guided LDA algorithm. This approach was motivated by two main considerations. Firstly, findings from Study 2 accentuated the importance of tweets depicting direct interactions between Twitter users and the tourist attraction. Secondly, a comparative analysis of Studies 1 and 2 highlighted the presence of noise in the unprocessed dataset, which could potentially skew the topic modelling process. By focusing on tweets directly engaging with the British Museum, we aimed to minimise this noise and enhance the accuracy of the topic modelling process. Due to the interpretability challenges that arise with the Guided LDA algorithm, we then adopted the keyword-in-context search approach to extract boycott-related tweets from the entire dataset (837,630 tweets). This process involved leveraging our contextual and domain knowledge of the research background to identify keywords and n-grams (refer to Tables 6.2 and 6.3) derived from the Guided LDA process, which was closely related to boycotting issues concerning the British Museum. Subsequently, we extracted all tweets from the entire dataset that contained these identified keywords and n-grams. This methodological adjustment is grounded in a body of social science research utilizing topic modelling techniques. These studies have highlighted the value of interactive topic modelling and constraint-based approaches for researchers well-versed in their document collections (Andrzejewski et al., 2011; Chang et al., 2009; Ramage et al., 2009; Roberts et al., 2016; Shankar & Parsana, 2022). Consequently, our strategy of using keyword-in-context searches, enriched by the insights from the Guided LDA model, intends to effectively address the limitations inherent in pure keyword tracking methods (Hannigan et al., 2019; Houghton et al., 2019) and probability-based topic classification models (Berger et al., 2020). Below, we provide a detailed illustration of the results obtained from the adjusted identification process for boycott-related tweets.

First, we employed the VADER algorithm to categorise the sentiment of the tweets. The majority were classified as positive, with a total of *206,728* tweets. Neutral sentiments followed, with *130,028* tweets, while negative sentiments were the least prevalent, with a total of *45,863* tweets. Subsequent to this sentiment classification, we identified unique negative words from the tweets with a negative sentiment, as detailed in Table 6.2. This process enabled us to form initial lists of seed topics and corresponding seed keywords, which are delineated in Table 6.3. In order to ensure the validity and reliability of our seed topics and keywords, we undertook additional steps for verification and refinement. Firstly, we performed a cross-validation of the number of seed topics (K_i) by estimating the statistically optimal number of topics (K_o); the results of this process can be found in Appendix 3, Figure A3.1. Secondly, the preliminary list of seed words from Table 6.2 was refined through multiple iterations of the seeded LDA model. This iterative process led to the final lists of seed topics and keywords, as presented in Table 6.3. This meticulous process aims to ensure the accuracy and relevance of the topic modelling, thus strengthening the robustness of the research findings. Expanding upon the results shown in Table 6.3, we used the Guided LDA algorithm to further investigate the distribution and semantic characteristics of each seed topic. The seed topics from Table 6.3 informed the LDA model, initially shaping the eta matrix, which represents the probabilities of topics associated with each word. This process also boosted the likelihood of seed words appearing in each respective topic. To ensure the model's effectiveness, we set the alpha parameter to "auto", which allowed for the automatic tuning of hyper parameters. Training the model over 20 iterations provided a more precise distribution of topics, enriching our understanding of the conversations around the seed topics.

Table 6.2 Most frequently used words in negative tweets

No	Words	Frequency	No	Words	Frequency	No	Words	Frequency
1	stolen	3144	10	sponsorship	1290	19	country	861
2	bp	2493	11	dropbp	1222	20	need	844
3	climate	1609	12	deal	1221	21	sorry	840
4	return	1524	13	death	1076	22	onthisday	823

5	right	1523	14	stop	1024	23	crime	800
6	war	1454	15	give	939	24	dont	781
7	died	1402	16	crisis	879	25	bad	777
8	miss	1356	17	dead	871	26	wrong	777
9	oil	1309	18	lost	868	27	marble	759

Table notes: 1. Words existing in both negative tweets and non-negative tweets are omitted from the table. 2. All words are converted to lowercase.

Table 6.3 Seed topics and keywords

No	Seed topics	Seed keywords
1	Contested artefacts	"stolen", "stolen artefact", "steal", "elgin", "greece", "loan", "russia", "sculpture", "return", "return_stolen", "back", "artefact", "object", "marble", "thief", "theft", "stone", "statue"
2	Fossil fuel sponsorship	"bp", "protest", "sponsorship", "activist", "oil", "drop", "sponsored", "deal", "stop"
3	Other tweets with negative valence	"war", "lost", "dead", "died", "attack", "onthisday", "city"

Table notes: 1. Seed topics and corresponding seed keywords are identified from Table 6.2 based on our prior knowledge of the research context. 2. For words with ambiguous meanings, we examined them within their respective contexts by analysing the top 50 most frequently retweeted tweets associated with each word. This approach enabled us to elucidate the intended meanings and precisely assign them to their appropriate seed topics. 3. All words have been converted to lowercase. 4. To ensure the validity and reliability of the number of seed topics (K_i) judgementally identified, we conducted cross-validation by running multiple Latent Dirichlet Allocation (LDA) models and estimating the statistically optimal number of topics (K_o). For the dataset, K_o was determined to be 3 or 4 (see Figure A3.1 in Appendix 3), which supports the reliability of K_i. 5. The initial list of seed words was adjusted and refined by running the seeded LDA model multiple times to ensure the most relevant and coherent topics.

Figures 6.2, 6.3, and A3.2 provide a comprehensive visualisation of the topic modelling results, each figure representing a distinct topic. Drawing upon the methodology suggested by Sievert and Shirley (2014), we focused on the 30 most relevant terms for each cluster. These terms are illustrated as bar charts on the right side of each figure, ranked from top to bottom based on their relevance. Grey bars indicate the overall frequency of each term across the entire corpus, while red bars reflect their specific frequency within the cluster. In the analysis, we adjusted the λ value to 1 to rank the terms based solely on their probability within the cluster. This adjustment did not significantly alter the alignment of the red bars with the grey bars' distribution pattern. To further ensure the robustness of the findings, we incrementally varied the λ value from 0 to 1 in steps of 0.2. Even with these adjustments, the distribution of the most relevant terms, especially the top 10, remained consistent across each cluster (exhibited minimal alterations). This observation suggests that the algorithm effectively identified the inherent clusters and precisely represented their term distributions. We also noted the simultaneous high relevance (frequency) and high probability (likelihood of appearing within a cluster) of the keywords present in each cluster, reinforcing the effectiveness of the approach. Importantly, we identified minimal overlap among the 30 most frequent terms defining these clusters, indicating distinct and separate cluster categories. In summary, the findings attest to the stability and robustness of the model, and hence we can move to the interpretation of the semantic feature of each cluster.

Each bubble's size in the figures corresponds to the prevalence of the associated cluster within the corpus. Cluster 3 is the most prevalent, representing 46.3% of the corpus, followed by Cluster 2 at 30.2%, and finally, Cluster 1 at 23.5%. Upon interpreting the top 30 most frequent terms for each cluster while varying the λ value, we observed that Clusters 1 and 2 maintain consistent semantic meanings that align with the contextually informed seed topics. Specifically, Cluster 1 is intimately linked with the topic of "contested artefacts," and Cluster 2 correlates closely with "fossil fuel sponsorships". This finding further affirms the relevance of the most frequent terms in Clusters 1 and 2 in guiding the keyword-in-context search process. On the other hand, Cluster 3 exhibits a diverse range of terms, suggesting the presence of multiple topics within this cluster. Considering that our focus here is the "noise", particularly boycott tweets that does not contribute to improved tourism demand forecasts, as identified in Appendix 2, Cluster 3 is not included in the subsequent analysis.

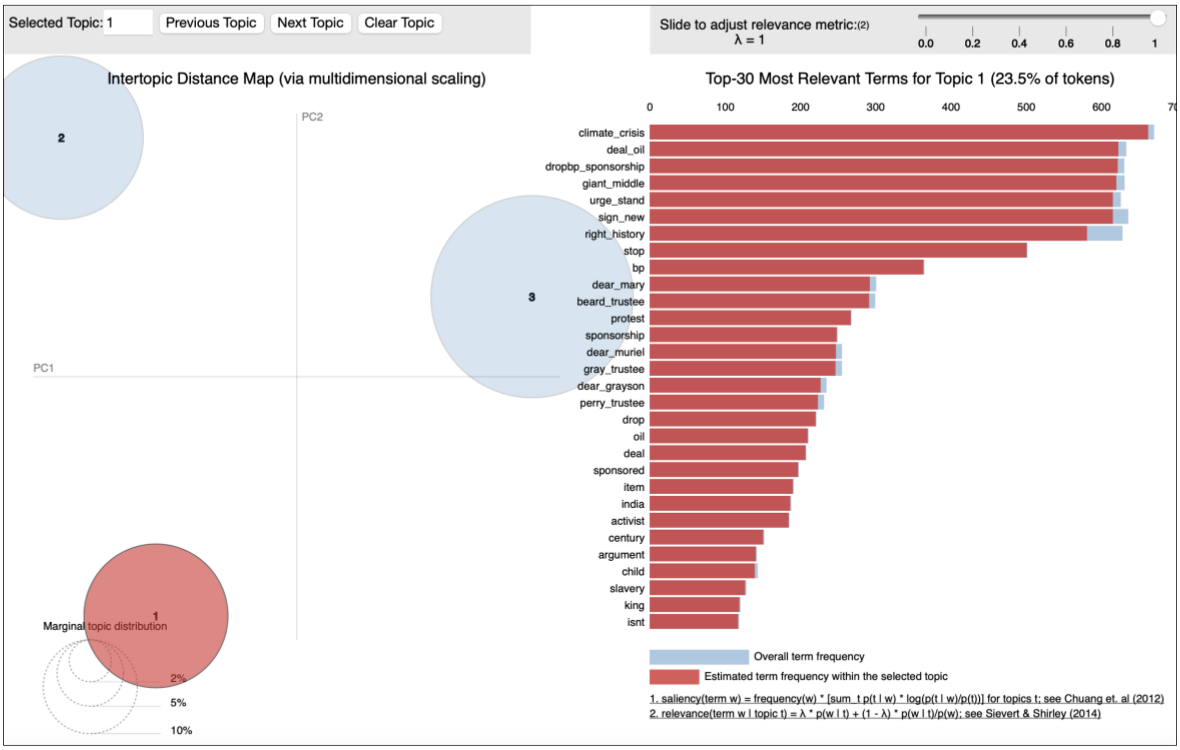

Figure 6.2 Topic distribution of negative tweets for the British Museum (part 1)

Figure notes: The red bar displays relevant terms for seed topic 1 (i.e., fossil fuel sponsorship).

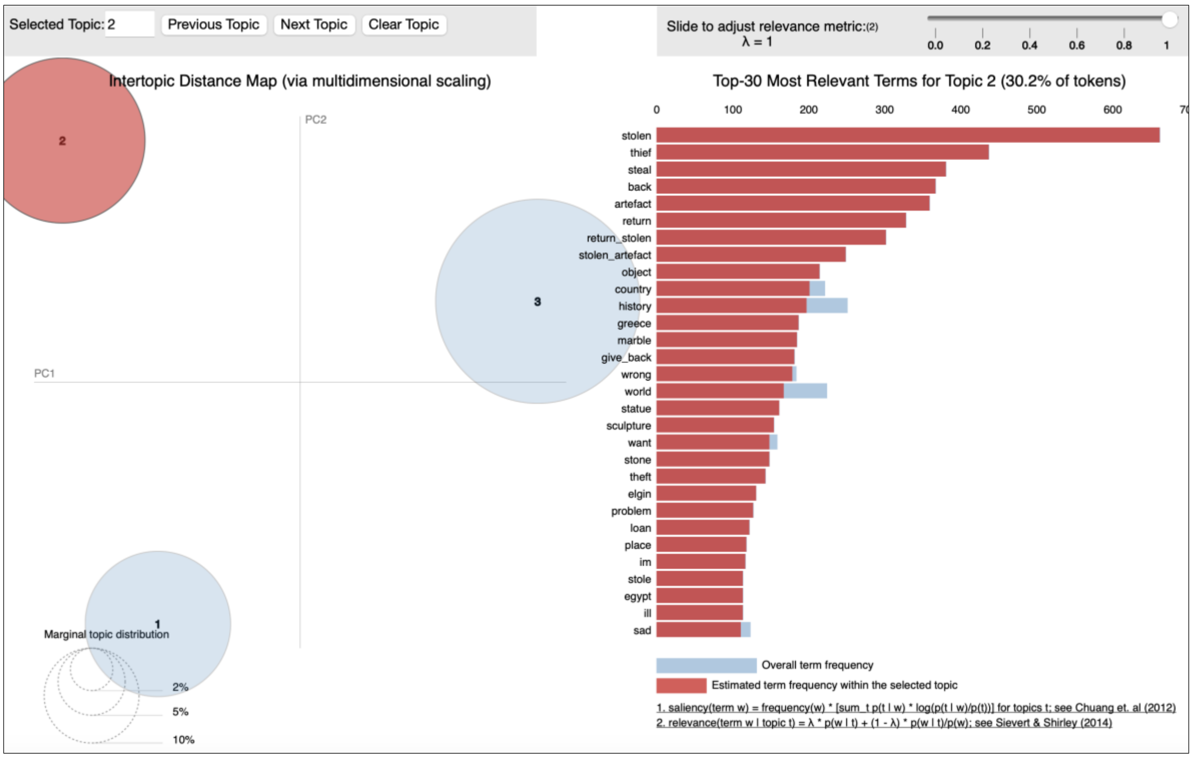

Figure 6.3 Topic distribution of negative tweets for the British Museum (part 2)

Figure notes: The red bar displays relevant terms for seed topic 2 (i.e., contested artefacts).

It is important to note that the findings are contextually dependent on the 45,863 negative tweets labelled by the VADER algorithm out of the 382,619 tweets mentioning "@britishmuseum". As such, they cannot be directly applied to the entire dataset and require further refinement to facilitate the keyword-in-context search process. The objective is to pinpoint those terms with high frequency and probability within each topic and that are unlikely to appear in overlapping topics in the entire dataset. Table 6.4 below summarises the results of this approach. Notably, despite the utilization of fewer keywords, the process yielded a substantially larger number of tweets related to "contested artefacts" as compared to "fossil fuel sponsorship". The reliability of this categorising process is further corroborated by the word cloud presented in Figure 6.4. This visual representation affirms the relevance of the categorised tweets to the seed topics, which were determined with contextual knowledge.

Table 6.4 Summary of boycott tweets extracted from the entire dataset

Category	Topic	Words	Amount	Example (tweet text)
T_1	Fossil fuel sponsorship	"climate crisis", "dropbp sponsorship", "deal oil", "giant middle", "urge stand", "sign new", "stop", "bp", "dear mary", "beard trustee", "protest", "sponsorship", "dear muriel", "gray trustee", "dear grayson", "perry trustee", "activist."	19868	We're outside the BP-sponsored exhibition in the British Museum. We have a simple message. Drop BP.
T_2	Contested artefacts	"stolen", "thief", "steal", "artefact", "return", "return stolen", "stolen artefact", "marble", "give back", "theft", "loan", "stole"	65762	@britishmuseum return what you've stolen! egyptians can open their own museums anywhere in europe. do not justificate theft! :)

Table notes: To maintain consistency with the time frame of the tourist arrivals data, only tweets posted between 1 January, 2014, and 30 September, 2022, are categorised.

Figure 6.4 Word cloud of tweets in topic 1 and topic 2

Figure notes: 1. Each word cloud includes the 200 most frequently occurring words from the input text. 2. The figure on the left side is the word cloud of tweets in topic 1 (fossil fuel sponsorship), while the figure on the right side is the word cloud of tweets in topic 2 (contested artefacts).

Informed by the insights gained from Studies 1 and 2, we delved into the effects of categorised tweets from four perspectives: (1) the volume of boycott tweets, (2) the number of likes these tweets received, (3) the count of retweets, and (4) the volume of replies to these tweets. Table 6.5 delineates the definitions and descriptive statistics for these variables. Concurrently, Figures 6.5 to 6.8 visualise these key variables, highlighting their varying frequencies (i.e., monthly and weekly). These figures exhibit a discernible escalation in the reaction-based variables (i.e., $Likes_t$, $Retweets_t$ and $Replies_t$) of tweets for both topics 1 and 2. The upward trend for topic 2 appears to have commenced around mid-2018 (see Figures 6.6 and 6.8), whereas for topic 1, the surge seems to have started later, around early 2019 (see Figures 6.5 and 6.7). In contrast, the pattern for $Volumes_t$ is different, with the peak occurring before the rise in reaction-based variables for both topics. Moreover, the increasing trend of $Volumes_t$ emerges prior to the start of the increasing trend of reaction-based variables. The distinct pattern of $Volumes_t$ and the reaction-based variables calls for a more thorough investigation. It also underscores the importance of including these variables in the evaluation, as they may reflect different signals (i.e., attitude and intention to boycott). Another noteworthy observation is the consistently higher frequency and variability of $Likes_t$ compared to the other two reaction-based variables, $Retweets_t$ and $Replies_t$, as evident in both the monthly and weekly data representations. Particularly, $Likes_t$ reached its zenith towards the end of the observation period, showing a significant deviation from $Retweets_t$ and $Replies_t$ for both topics. This suggests a virality in the appeal of certain tweets during that period (i.e., December 2022). These emerging trends and their potential implications necessitate further scrutiny in the subsequent analyses.

The rationale for including tweets-generated variables (hereafter referred to as boycott variables) at a weekly frequency in Study 3, rather than a higher frequency such as daily, is twofold. Firstly, Study 2's findings suggest an optimal mixed-frequency model lag of 50 days. This implies that the impact of tweets on tourists' decision-making processes does not occur instantaneously. This is consistent with previous research, such as Liu et al. (2019) and Yang et al. (2015), which found that after consulting online resources, travel decisions are typically made two to three weeks in advance. Thus, capturing data at a daily level may not yield additional valuable insights. Secondly, the size of the dataset consisting of boycott tweets is considerably smaller than the one used in Study 2. Consequently, the data points are sparser and contain null values for certain dates. Collating data on a weekly basis effectively addresses this issue, particularly for the variable $Volumes_t$, as it amalgamates data points and thus minimises the occurrence of null values. Therefore, using weekly frequency data strikes a balance between maintaining the integrity of the data and the ability to capture the signalling value of boycott variables relevant to the research context. This frequency allows us to glean meaningful insights without the noise that might be associated with daily data.

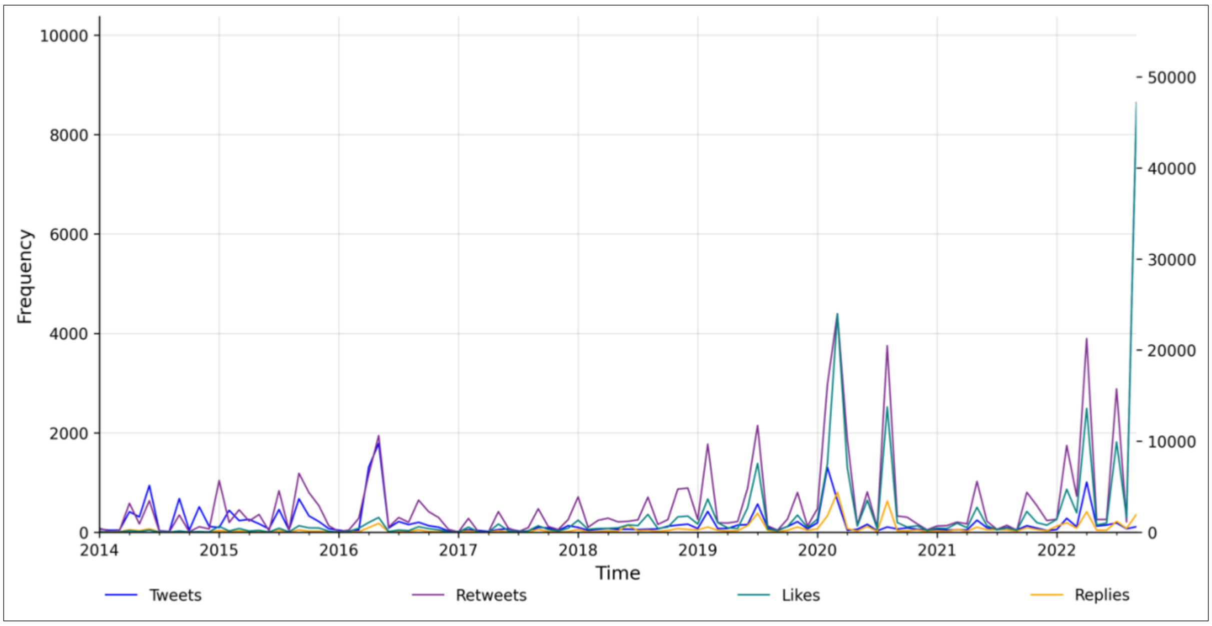

Figure 6.5 Monthly frequency boycott variables generated from tweets in category T1

Figure notes: Tweets, retweets, and replies are scaled to the left axis, while likes are scaled to the right axis, for the visualisation of the corresponding variable trends.

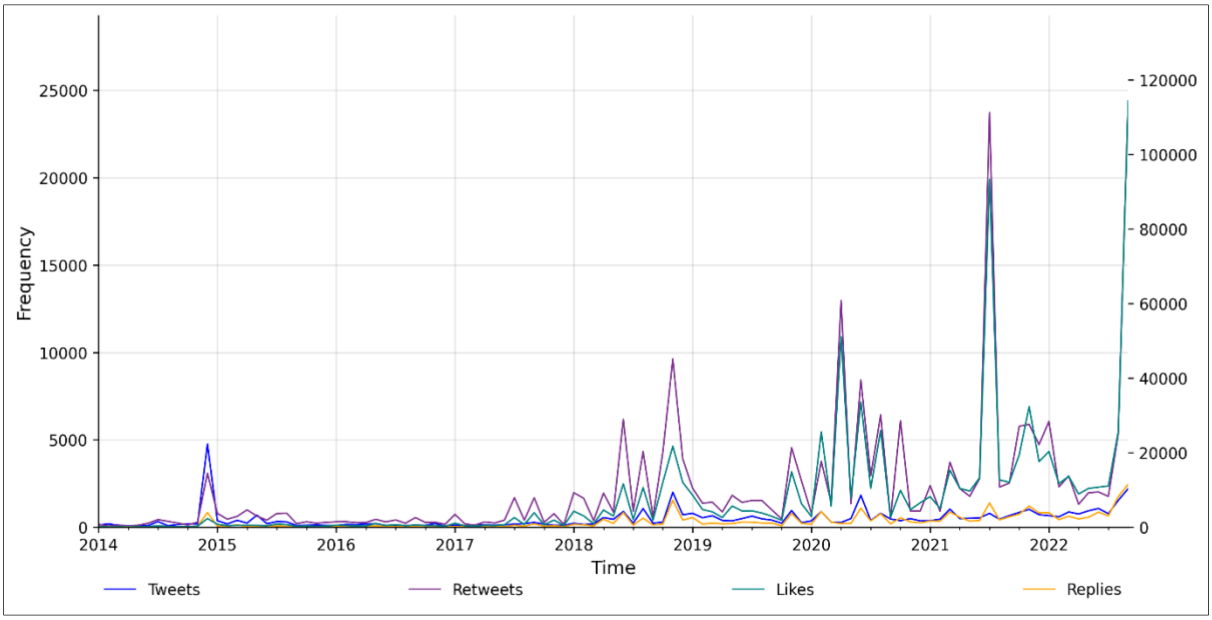

Figure 6.6 Monthly frequency boycott variables generated from tweets in category T2

Figure notes: Tweets, retweets, and replies are scaled to the left axis, while likes are scaled to the right axis, for the visualisation of the corresponding variable trends.

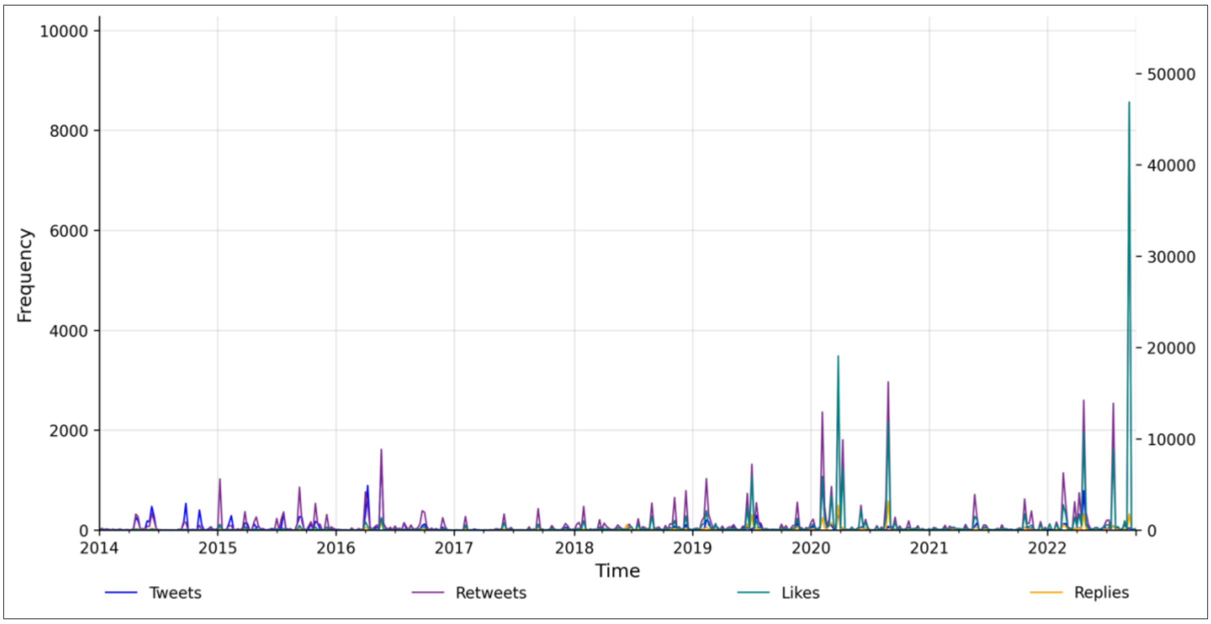

Figure 6.7 Weekly frequency boycott variables generated from tweets in category T_1

Figure notes: Tweets, retweets, and replies are scaled to the left axis, while likes are scaled to the right axis, for the visualisation of the corresponding variable trends.

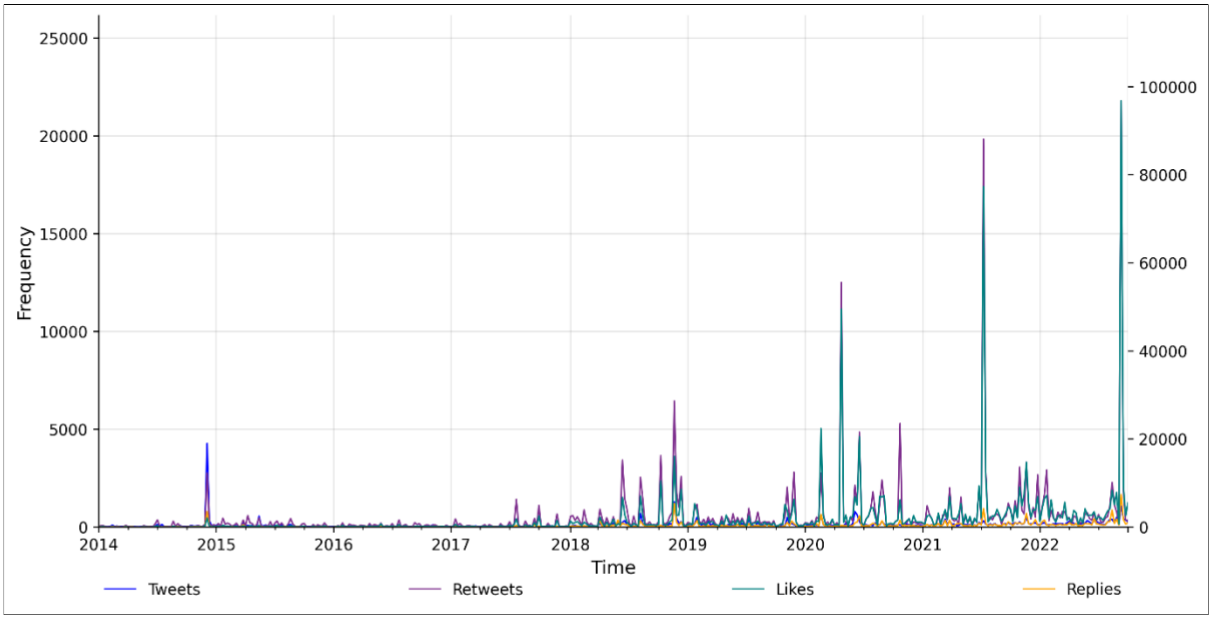

Figure 6.8 Weekly frequency boycott variables generated from tweets in category T_2

Figure notes: Tweets, retweets, and replies are scaled to the left axis, while likes are scaled to the right axis, for the visualisation of the corresponding variable trends.

Table 6.5 Descriptive statistics of variables

Variable	Description	Period	Observations	Max	Min	Mean	Std. Dev.
$\boldsymbol{Arrivals_t}$	Monthly volume of tourist arrivals to the British Museum at time t.	Jan 2014 – Feb 2020	74	765877	363281	524532.400	93453.300
$\boldsymbol{Arrivalsa_t}$	Seasonally adjusted monthly volume of tourist arrivals to the British Museum at time t, adjusted through the moving average process.	Jan 2014 – Feb 2020	74	631801.3	443634.1	526797.700	41739.700
$\boldsymbol{T_1\_Volumes_t}$	Monthly volume of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	74	1789	15	209.919	318.445
$\boldsymbol{T_1\_Likes_t}$	Monthly volume of likes of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	74	7562	4	721.095	1320.534
$\boldsymbol{T_1\_Retweets_t}$	Monthly volume of retweets of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	74	2968	2	411.784	535.886
$\boldsymbol{T_1\_Replies_t}$	Monthly volume of replies of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	74	380	1	40.730	62.958
$\boldsymbol{T_2\_Volumes_t}$	Monthly volume of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	74	4832	86	491.932	614.502
$\boldsymbol{T_2\_Likes_t}$	Monthly volume of likes of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	74	57395	79	4466.959	8420.344
$\boldsymbol{T_2\_Retweets_t}$	Monthly volume of retweets of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	74	24393	77	1796.324	3230.554
$\boldsymbol{T_2\_Replies_t}$	Monthly volume of replies of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	74	1680	20	220.378	291.854

$\boldsymbol{T_1\_Volumes_t}$	Weekly volume of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	322	1544	1	48.48137	130.1264
$\boldsymbol{T_1\_Likes_t}$	Weekly volume of likes of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	322	5962	0	166.8851	539.2083
$\boldsymbol{T_1\_Retweets_t}$	Weekly volume of retweets of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	322	2363	0	95.18634	227.908
$\boldsymbol{T_1\_Replies_t}$	Weekly volume of replies of boycott tweets in category T_1 at time t.	Jan 2014 – Feb 2020	322	304	0	9.596273	26.91474
$\boldsymbol{T_2\_Volumes_t}$	Weekly volume of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	322	4300	8	113.177	265.8644
$\boldsymbol{T_2\_Likes_t}$	Weekly volume of likes of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	322	52693	5	1032.674	3617.887
$\boldsymbol{T_2\_Retweets_t}$	Weekly volume of retweets of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	322	23109	3	414.4845	1430.201
$\boldsymbol{T_2\_Replies_t}$	Weekly volume of replies of boycott tweets in category T_2 at time t.	Jan 2014 – Feb 2020	322	1256	0	50.91304	108.4146

Source: UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Table notes: As discussed in Section 6.2.3, data from March 2020 to September 2022 are not included in the calculation for each variable as the investigation focuses on the pre-lockdown period.

6.2.2 Interactions among variables

Prior to embarking on econometric analyses, we put forth model-free evidence derived from the data to shed light on the interrelationships between boycott variables and supplementary variables identified in Section 3.5.3.1. These variables encompass tourist arrivals, dates of physical protests, dates of various rounds of COVID-19 pandemic lockdowns, and the date of COVID-19 lockdown easing. Given that the volume of tourist arrivals data is available at a monthly frequency, we have incorporated boycott variables aggregated at a similar monthly frequency to facilitate the evaluation of interactions. Furthermore, since we identified a similar trend for the three reaction-based variables throughout almost the entire observation period, the evaluation primarily centres on $Volumes_t$ and $Likes_t$ for each topic. The results of these evaluations are exhibited in Figures 6.9 to 6.12. In anticipation of a potential negative impact of boycott tweets on tourist arrivals, our primary focus is on discerning the correlation between the peaks in boycott variables and the troughs in tourist arrivals, both pre-lockdown and post-lockdown. To mitigate any bias that may arise from visual interpretation, peaks in boycott variables are determined via the rolling window z-score method, as elaborated in Section 3.5.3.3. The same method is used to dertermine the bottom points of adjusted tourist arrivals with the threshold of z-score setting to -1.

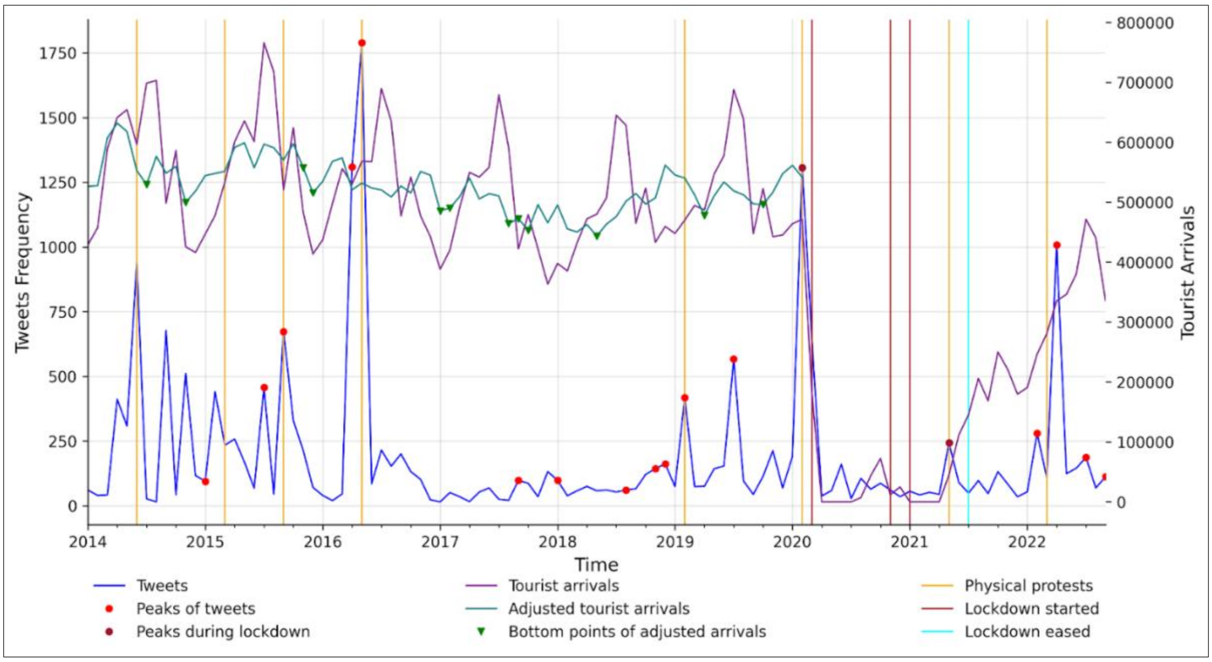

Figure 6.9 Interactions among $T_1\_Volumes_t$, tourist arrivals, travel restrictions and physical protests

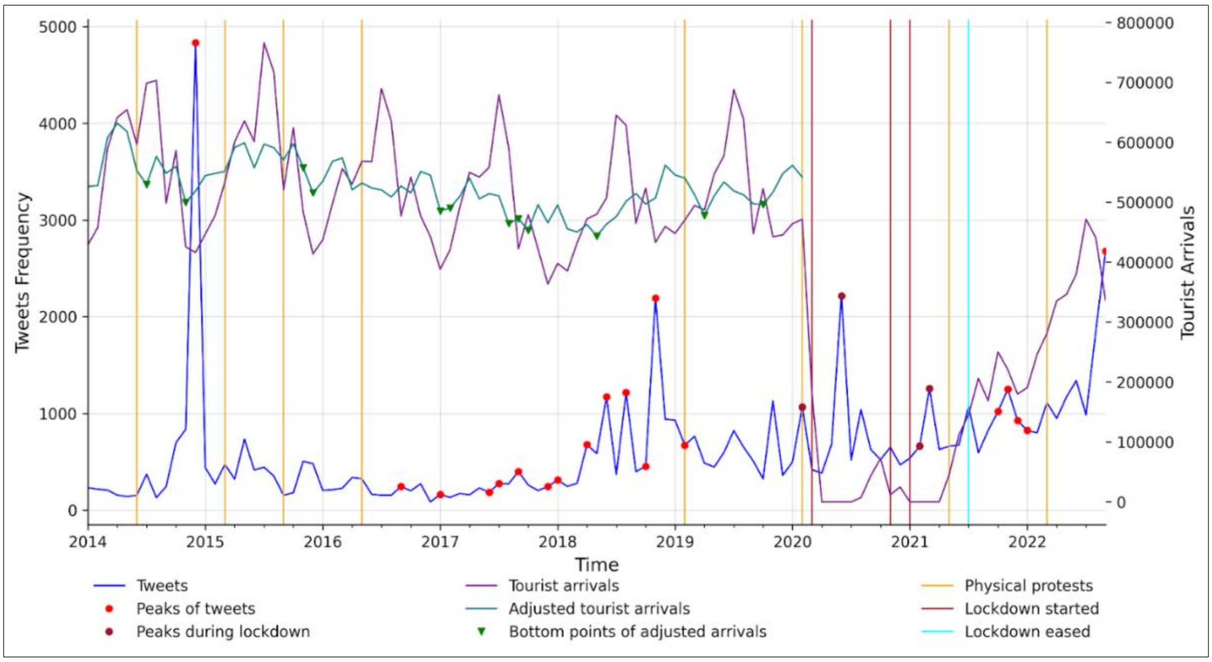

Figure 6.10 Interactions among $T_2\_Volumes_t$, tourist arrivals, travel restrictions and physical protests

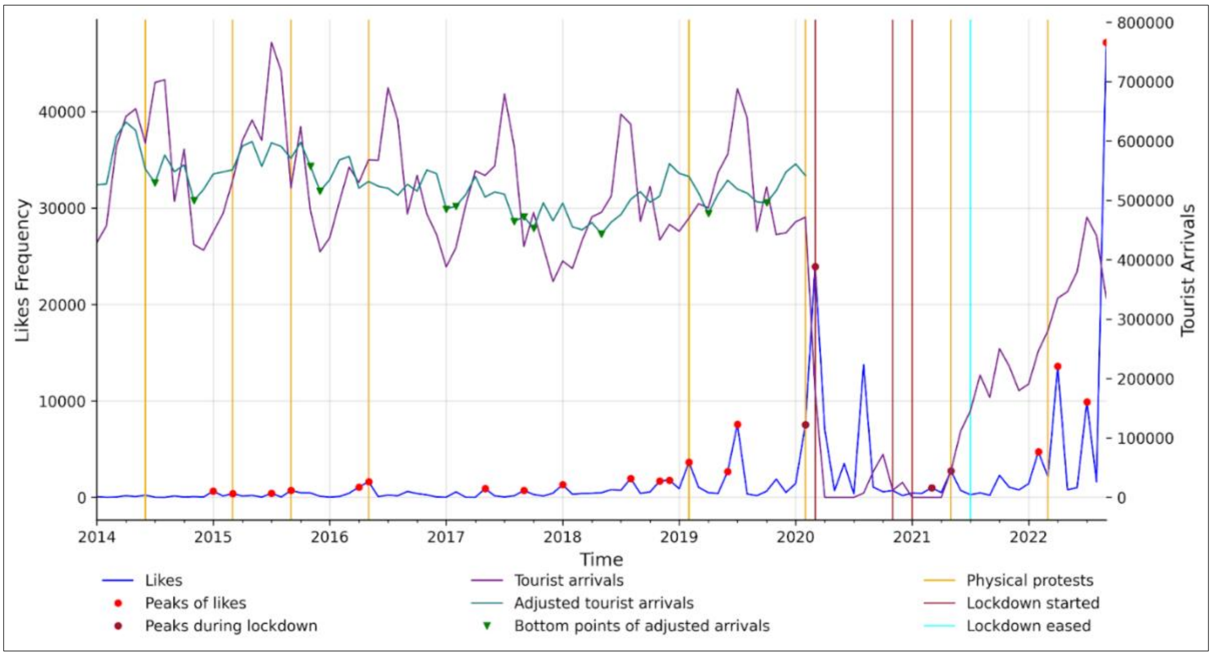

Figure 6.11 Interactions among $T_1\_Likes_t$, tourist arrivals, travel restrictions and physical protests

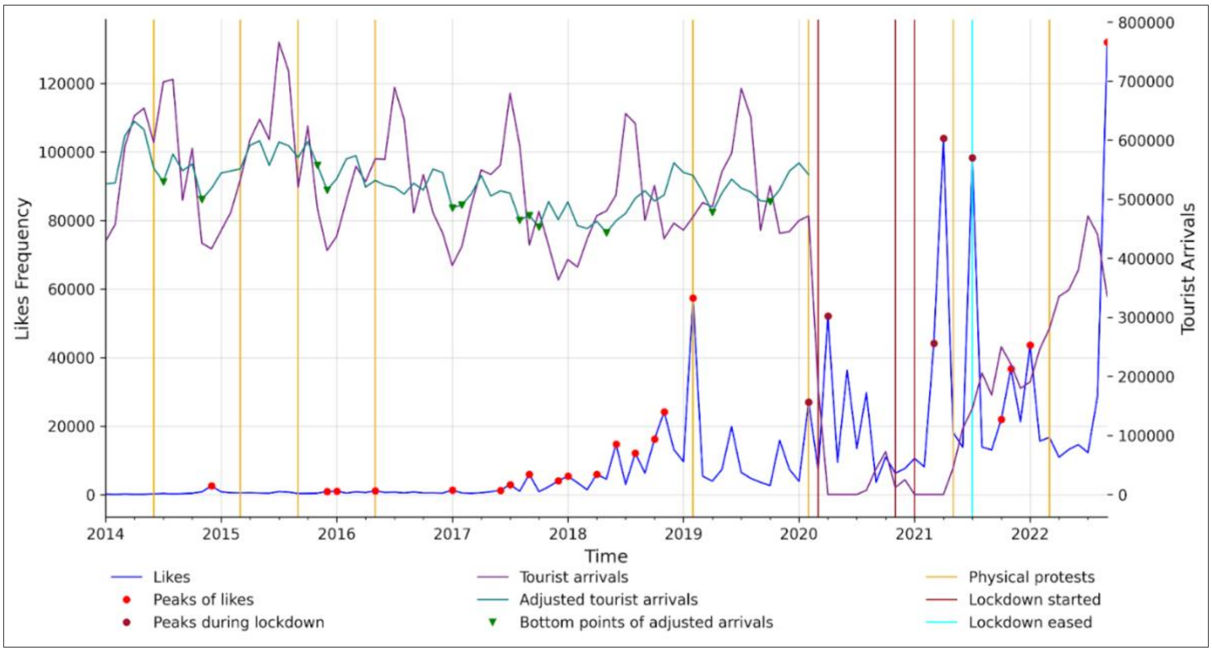

Figure 6.12 Interactions among $T_2\_Likes_t$, tourist arrivals, travel restrictions and physical protests

Figures 6.11 and 6.12 demonstrate an ascending trajectory for reaction-based variables-namely $Likes_t$ for both topics 1 and 2 - commencing with the onset of the COVID-19 lockdown. This escalating trajectory reaches its zenith for both topics by the end of the observation period. Contrastingly, the volumes of tweets addressing the two distinct topics (i.e., $T_1\_Volumes_t$ and $T_2\_Volumes_t$,) peak well before the introduction of the lockdown measures (see Figures 6.9 and 6.10 and discussions in Section 6.3.1). Given that the scale of $T_1\_Likes_t$ and $T_2\_Likes_t$ substantially outstrips the volume of tweets for both topics, we infer that the attention and engagement (approval in particular, as suggested by Saxton et al. (2019)) towards these subjects intensified during and subsequent to the lockdown period. This surge in social media usage during the lockdown, as affirmed by the increased number of daily active Twitter users and engagement growth since the onset of the pandemic (Twitter, 2021), possibly underpins this trend. An interesting pattern emerges when examining Topic 1. We observe a correlation between the peaks of volumes and likes and each physical protest occurrence. However, this congruence does not extend to Topic 2, where a matching pattern is less discernible and limited. In synthesising these findings and observed data trends, we anticipate a high probability of a forthcoming physical protest against the British Museum (Topic 1) post our observation window. This prediction is corroborated by the evidence from the Lexis-Nexis database, wherein we identified newspaper articles discussing climate protests against a BP-sponsored exhibition in the museum in October 2022 (Michaelson, 2022). Regarding the interactions between the peaks of variables generated by boycott tweets and adjusted tourist arrivals, no significant patterns were observed. These observations highlight the signalling value of boycott tweets, particularly those associated with Topic 1, to Twitter users' increasing attention and approval to the boycott context. Furthermore, the dates of physical boycotts can serve as a useful tool to confirm the time frame for the subsequent econometric analysis, which can precisely assess the impact of boycott tweets on tourist arrivals.

6.3 Econometric evaluation

6.3.1 Stationarity and cointegration testing

The ADF test results for all variables included in the upcoming time series modelling process are displayed in Table 6.6. In line with the testing strategy used in Study 1, the decision on whether to incorporate a linear trend term in the test equation is based on the visual inspection of each test variable. In the context of Study 3, this decision was also cross-checked by comparing information criteria, including AIC and SC, assessing the model's performance with and without a linear trend. If a model that includes a linear trend shows a lower information criterion value, it could be suitable to include a linear trend in the ADF test.

As per Table 6.6, the variable $Arrivals_t$ did not exhibit stationarity at the 5% significance level. However, its first-order differenced value $\Delta Arrivals_t$ and seasonally adjusted value $Arrivalsa_t$ were found to be stationary. With respect to the boycott variables at various frequencies, the majority were determined to be stationary at the 1% significance level, with the exception of the monthly aggregated $T_1\_Likes_t$. The first-order difference of $T_1\_Likes_t$, denoted as $\Delta T_1\_Likes_t$, was confirmed to be stationary at the 1% significance level. As such, it is worth examining whether $Arrivals_t$ and $\Delta T_1\_Likes_t$ share a cointegration relationship of order $I(0)$, which could help determine the most appropriate modelling approach.

Table 6.6 Results of ADF test for key variables sampled at different frequencies

Variable	Frequency	t-statistics	Probability	5% Critical value	Note
$Arrivals_t$	Monthly	-0.034	0.995	-3.484	Intercept and trend are included in the test equation
$\Delta Arrivals_t$	Monthly	-3.052	0.036	-2.910	Intercept is included in the test equation
$Arrivalsa_t$	Monthly	-3.694	0.029	-3.473	Intercept and trend are included in the test equation
$T_1\_Volumes_t$	Monthly	-5.889	0.000	-2.902	Intercept is included in the test equation
$T_1\_Likes_t$	Monthly	-1.365	0.862	-3.478	Intercept and trend are included in the test equation
$\Delta T_1\_Likes_t$	Monthly	-10.396	0.000	-2.906	Intercept is included in the test equation
$T_1\_Retweets_t$	Monthly	-6.603	0.000	-3.473	Intercept and trend are included in the test equation
$T_1\_Replies_t$	Monthly	-6.063	0.000	-3.473	Intercept and trend are included in the test equation
$T_2\_Volumes_t$	Monthly	-7.893	0.000	-2.902	Intercept is included in the test equation
$T_2\_Likes_t$	Monthly	-5.366	0.000	-3.473	Intercept and trend are included in the test equation
$T_2\_Retweets_t$	Monthly	-6.735	0.000	-3.473	Intercept and trend are included in the test equation
$T_2\_Replies_t$	Monthly	-7.922	0.000	-3.473	Intercept and trend are included in the test equation
$T_1\_Volumes_t$	Weekly	-4.658	0.000	-2.870	Intercept is included in the test equation
$T_1\_Likes_t$	Weekly	-10.355	0.000	-3.424	Intercept and trend are included in the test equation
$T_1\_Retweets_t$	Weekly	-16.300	0.000	-3.424	Intercept and trend are included in the test equation
$T_1\_Replies_t$	Weekly	-9.923	0.000	-3.424	Intercept and trend are included in the test equation
$T_2\_Volumes_t$	Weekly	-16.217	0.000	-2.870	Intercept is included in the test equation
$T_2\_Likes_t$	Weekly	-17.228	0.000	-3.424	Intercept and trend are included in the test equation
$T_2\_Retweets_t$	Weekly	-17.164	0.000	-3.424	Intercept and trend are included in the test equation
$T_2\_Replies_t$	Weekly	-15.637	0.000	-3.424	Intercept and trend are included in the test equation

Source: the UK government's official website, Twitter Inc., and author's estimation using EViews Version 11.
Table notes: 1. "Δ" indicates the first-order differencing operation applied to the variable. 2. The null hypothesis of the ADF test is that there is a unit root.

Consistent with the approach in Study 1, we implemented the augmented Engle and Granger two-step cointegration test to examine the cointegration relationship between $Arrivals_t$ and

$T_1\_Likes_t$. In the first step, an ordinary least squares (OLS) regression model was established: $Arrivals_t = \alpha + \beta T_1\_Likes_t + \varepsilon_t$. Given the characteristics of $Arrivals_t$ and $T_1\_Likes_t$, a linear trend variable was included as an additional regressor in the cointegration model. In the second step, the residuals from this regression (ε_t) were tested for stationarity. If these residuals were found to be stationary, it would suggest that the variables in question are cointegrated, indicating a long-run equilibrium relationship. However, the calculations revealed an Engle-Granger t-statistic of -0.872 with a probability of 0.952, suggesting that the null hypothesis of no cointegration could not be rejected. Consequently, we concluded that no long-run equilibrium relationship was detected between $Arrivals_t$ and $T_1\_Likes_t$. Thus, there was no requirement to fit an error correction model for $Arrivals_t$ and $T_1\_Likes_t$.

6.3.2 VAR and VECM modelling

Based on the findings from the stationarity and cointegration assessments, there is a solid rationale for fitting bivariate $VAR(p)$ models for $Arrivalsa_t$ and boycott variables. Adhering to the modelling approach outlined in Section 3.5.3.4, we constructed eight distinct bivariate VAR models (refer to Table 6.7 for details). The optimal lag length for each model is determined by a consensus from the following five criteria: LR, FPE, SC, AIC, and HQ. As evidenced in Table 6.7, for all the bivariate VAR models, the null hypothesis of the Lagrange Multiplier (LM) test is not rejected at the 5% significance level. This result suggests the absence of autocorrelation in the residuals, a prerequisite for a correctly specified model. Additionally, this indicates that the selected lag order for each model is satisfactory, affirming that the models are well specified. The absence of autocorrelation signifies that the models capture all relevant system dynamics effectively. To further ascertain the stability of each VAR system, we visually inspected the roots of the characteristic polynomial. As illustrated in Figure 6.13, all models, barring model (2), demonstrate confirmed stability. For models (1) and (3) to (8), none of the roots of the characteristic polynomial lie outside the unit circle. In contrast, model (2), which includes $Arrivalsa_t$ (stationary at $I(0)$) and $T_1\_Likes_t$ (stationary at $I(1)$), deviates from this pattern. Half of the roots for this model either lie on or outside the unit circle, indicating instability in the VAR system, which could result in unreliable estimates.

Table 6.7 Estimation and LM test results for VAR models

No	Variables	Model specification	LM Tests (H0: no serial correlation at lags 1 to h)		
			Lag	LRE-Statistics	Probability
1	$Arrivalsa_t$ & $T_1\_Volumes_t$	$Arrivalsa_t = 0.755 Arrivalsa_{t-1} + 3.169\, T_1\_Volumes_{t-1} + 128472.7$ $T_1\_Volumes_t = 0.003\, T_1\_Volumes_{t-1} + 0.183 Arrivalsa_{t-1} - 1144.616$	1	2.503	0.644
2	$Arrivalsa_t$ & $\Delta T_1\_Likes_t$	$Arrivalsa_t = 0.575 Arrivalsa_{t-1} + 0.061 Arrivalsa_{t-2} + 0.121 Arrivalsa_{t-3} + 0.118 Arrivalsa_{t-4} + 0.067 Arrivalsa_{t-5} - 0.116 Arrivalsa_{t-6} - 1.802 \Delta T_{1_{Likes_{t-1}}} - 7.067 \Delta T_{1_{Likes_{t-2}}} - 10.946 \Delta T_{1_{Likes_{t-3}}} - 9.39 \Delta T_{1_{Likes_{t-4}}} - 8.772 \Delta T_{1_{Likes_{t-5}}} - 1.673 \Delta T_{1_{Likes_{t-6}}} + 91368.53$	1	3.496	0.479
			2	7.178	0.518
			3	10.057	0.611
		$\Delta T_1\_Likes_t = -0.0017 Arrivalsa_{t-1} + 0.012 Arrivalsa_{t-2} - 0.013 Arrivalsa_{t-3} + 0.0012 Arrivalsa_{t-4} + 0.0029 Arrivalsa_{t-5} - 0.0031 Arrivalsa_{t-6} - 0.854 \Delta T_{1_{Likes_{t-1}}} - 1.191 \Delta T_{1_{Likes_{t-2}}} - 1.287 \Delta T_{1_{Likes_{t-3}}} - 1.083 \Delta T_{1_{Likes_{t-4}}} - 0.846 \Delta T_{1_{Likes_{t-5}}} - 0.96 \Delta T_{1_{Likes_{t-6}}} + 1256.426$	4	14.424	0.567
			5	15.034	0.775
			6	18.371	0.785
3	$Arrivalsa_t$ & $T_1\_Retweets_t$	$Arrivalsa_t = 0.763 Arrivalsa_{t-1} - 2.488 T_1\_Retweets_{t-1} + 125724.9$ $T_1\_Retweets_t = 0.0017 Arrivalsa_{t-1} + 0.087 T_1\_Retweets_{t-1} - 525.812$	1	1.285	0.864
4	$Arrivalsa_t$ & $T_1\_Replies_t$	$Arrivalsa_t = 0.758 Arrivalsa_{t-1} - 28.144 T_1\_Replies_{t-1} + 128475.6$ $T_1\_Replies_t = 0.000035 Arrivalsa_{t-1} + 0.305 T_1\_Replies_{t-1} + 11.407$	1	0.643	0.958

5	$Arrivalsa_t$ & $T_2\_Volumes_t$	$Arrivalsa_t = 0.769Arrivalsa_{t-1} + 6.227\ T_2\_Volumes_{t-1} + 118744.3$ $T_2\_Volumes_t = 0.002\ T_2\_Volumes_{t-1} + 0.157Arrivalsa_{t-1} + 1499.055$	1	4.013	0.404
6	$Arrivalsa_t$ & $T_2\_Likes_t$	$Arrivalsa_t = 0.659Arrivalsa_{t-1} + 0.005Arrivalsa_{t-2} - 0.037Arrivalsa_{t-3} + 0.18Arrivalsa_{t-4} + 0.075T_2\_Likes_{t-1} - 0.391T_2\_Likes_{t-2} + 0.161T_2\_Likes_{t-3} + 0.214T_2\_Likes_{t-4} + 99413.33$	1	3.443	0.487
			2	7.721	0.461
		$T_2\_Likes_t = 0.0017Arrivalsa_{t-1} + 0.055Arrivalsa_{t-2} - 0.048Arrivalsa_{t-3} - 0.045Arrivalsa_{t-4} - 0.03T_2\_Likes_{t-1} - 0.059T_2\_Likes_{t-2} + 0.323T_2\_Likes_{t-3} + 0.37T_2\_Likes_{t-4} + 21050.07$	3	9.444	0.665
			4	9.827	0.876
7	$Arrivalsa_t$ & $T_2\_Retweets_t$	$Arrivalsa_t = 0.757Arrivalsa_{t-1} + 0.149T_2\_Retweets_{t-1} + 124530$ $T_2\_Retweets_t = -0.014Arrivalsa_{t-1}0.652T_2\_Retweets_{t-1} + 15339.49$	1	3.308	0.508
8	$Arrivalsa_t$ & $T_2\_Replies_t$	$Arrivalsa_t = 0.751Arrivalsa_{t-1} - 0.033Arrivalsa_{t-2} + 0.11Arrivalsa_{t-3} + 18.421T_{2_{Replies_{t-1}}} + 3.895T_{2_{Replies_{t-2}}} - 9.328T_2\_Replies_{t-3} + 87173.04$	1	5.390	0.250
			2	10.275	0.246
		$T_2\_Replies_t = -0.0021Arrivalsa_{t-1} + 0.0013Arrivalsa_{t-2} - 0.000097Arrivalsa_{t-3} + 0.0791T_{2_{Replies_{t-1}}} + 0.294T_2\_Replies_{t-2} + 0.304T_2\_Replies_{t-3} + 976.859$	3	13.731	0.318

Table note: "Δ" indicates the first-order differencing operation applied to the variable

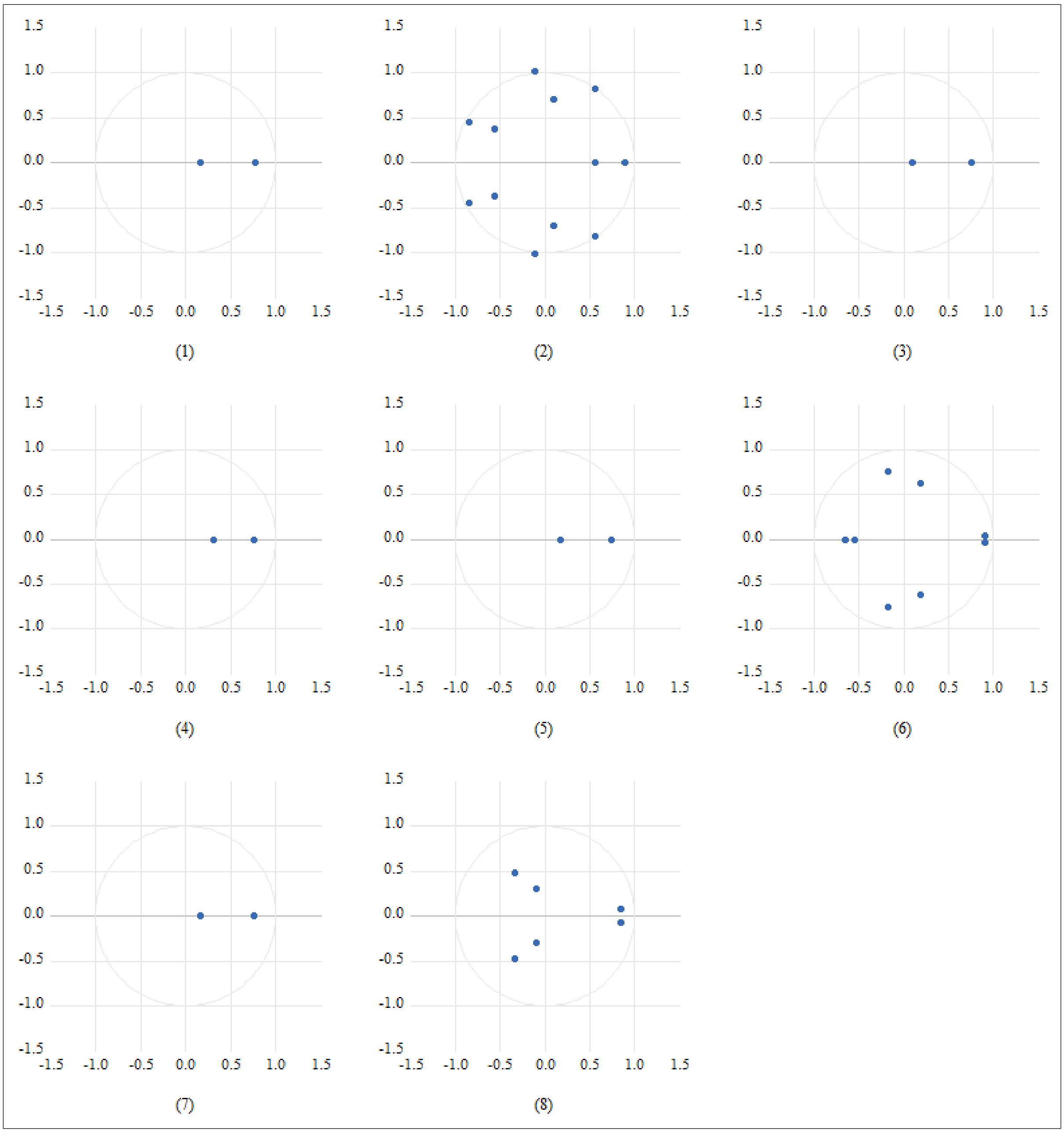

Figure 6.13 Stability test for VAR models

Figure notes: 1. The figure visualises the roots of the characteristic polynomial for each VAR model. 2. The numbering of figures (1) to (8) corresponds to the respective VAR models presented in Table 6.7.

Despite the confirmed stationarity of $Arrivalsa_t$, the instability observed in the model (2) prompts the need for additional cointegration testing. Given the different orders of integration for $Arrivalsa_t$ and $T_1\_Likes_t$, the existence of a long-run equilibrium relationship between these variables is plausible. A subsequent cointegration test was conducted to validate this potential relationship. In performing the Engle-Granger (EG) cointegration test on $Arrivalsa_t$ and $T_1\_Likes_t$, the t-statistic value is recorded as -4.557878 with a significance level of 0.01. This result substantiates the existence of cointegration between these variables. Consequently, the next logical step is to specify a Vector Error Correction Model (VECM) that precisely captures both the short-term fluctuations and long-term dynamics between the $Arrivalsa_t$ and $T_1\_Likes_t$. The lag order (i.e., 5) for the VECM model is determined by subtracting 1 from the optimal lag order selected for model (2) (i.e., 6). Based on the calculation, the VECM (5) model is constructed as follows:

$$\begin{aligned}\Delta Arrivalsa_t = &-0.156\left(Arrivalsa_{t-1} - 26.342 * T_{1_{Likes_{t-1}}} - 504897.07\right) - 0.224\Delta Arrivalsa_{t-1} \\ &- 0.207\Delta Arrivalsa_{t-2} - 0.109\Delta Arrivalsa_{t-3} - 0.026\Delta Arrivalsa_{t-4} \\ &+ 0.083\Delta Arrivalsa_{t-5} - 5.091\Delta T_{1_{Likes_{t-1}}} - 8.075\Delta T_{1_{Likes_{t-2}}} - 11.005\Delta T_{1_{Likes_{t-3}}} \\ &- 8.438\Delta T_{1_{Likes_{t-4}}} - 7.527\Delta T_{1_{Likes_{t-5}}} - 191.949\end{aligned}$$

$$\begin{aligned}\Delta T_{1_{Likes_t}} = &\ 0.0025\left(Arrivalsa_{t-1} - 26.342 * T_{1_{Likes_{t-1}}} - 504897.07\right) - 0.0062\Delta Arrivalsa_{t-1} \\ &+ 0.015\Delta Arrivalsa_{t-2} + 0.0027\Delta Arrivalsa_{t-3} - 0.00057\Delta Arrivalsa_{t-4} \\ &- 0.00057\Delta Arrivalsa_{t-5} - 0.6981\Delta T_{1_{Likes_{t-1}}} - 0.794\Delta T_{1_{Likes_{t-2}}} - 0.568\Delta T_{1_{Likes_{t-3}}} \\ &- 0.255\Delta T_{1_{Likes_{t-4}}} - 0.053\Delta T_{1_{Likes_{t-5}}} + 169.719\end{aligned}$$

Figure 6.14 below confirms the stability of the model, as none of the roots is located outside the unit circle (i.e., none has a modulus larger than 1). The bivariate VECM consists of two equations. The first equation models the change in $Arrivalsa_t$, while the second equation models the change in $T_1\_Likes_t$. Both equations incorporate the error correction term (the cointegration equation:$Arrivalsa_{t-1} - 26.342 * T_{1_{Likes_{t-1}}} - 504897.07$), encapsulating the long-term relationship between seasonally adjusted tourist arrivals and likes on boycott tweets regarding the British Museum's association with fossil fuel companies (Topic 1). In the first equation, the coefficient of the error correction term is -0.156, statistically significant at the 5% level (with a t-statistic of -2.068 and a probability of 0.041). This implies that when $Arrivalsa_t$ and $T_1\_Likes_t$ deviate from their long-term equilibrium relationship (as captured by the cointegrating equation), $Arrivalsa_t$ adjusts in the opposite direction by 15.6% of the deviation in the next period to restore the equilibrium. Essentially, one standard deviation in $T_1\_Likes_t$ will result in an approximately -15.6% adjustment in $Arrivalsa_t$ to re-establish its long-run equilibrium with $T_1\_Likes_t$. The significance of this adjustment highlights the robustness of the long-term relationship between $Arrivalsa_t$ and $T_1\_Likes_t$, and the responsive adjustment of tourist arrivals to deviations from this equilibrium.

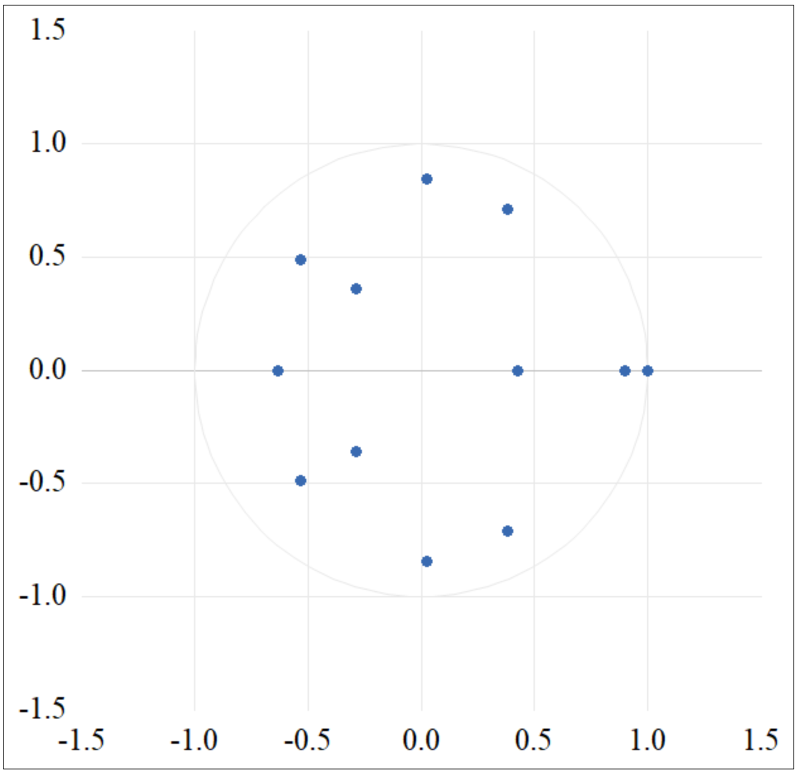

Figure 6.14 Stability test for the VECM (5) model

Figure notes: The figure visualise the roots of the characteristic polynomial for VECM (5) model for $Arrivalsa_t$ and $T_1\_Likes_t$.

The second equation models the change in $T_1\_Likes_t$. Here, the coefficient of the error correction term is 0.0025, but it is not statistically significant (with a t-statistic of 0.642 and a probability of 0.523). This suggests that, unlike $Arrivalsa_t$, $T_1\_Likes_t$ does not adjust in the subsequent period in response to any deviations from the long-term equilibrium with $Arrivalsa_t$. In other words, while there is a long-term relationship between $T_1\_Likes_t$ and $Arrivalsa_t$, the likes on boycott tweets with topic 1 (i.e., $T_1\_Likes_t$) do not necessarily respond to offset any disturbances that might push these variables away from their long-term equilibrium relationship.

Given the significant error correction term in the first equation and its non-significance in the second, it is important to conduct further analysis to understand the nature of interactions between $T_1\_Likes_t$ on $Arrivalsa_t$. In particular, it would be beneficial to compute the accumulated response as opposed to the instantaneous response. Calculating the accumulated response could provide more insights into the cumulative effect of a shock (i.e., a change by one standard deviation) to $T_1\_Likes_t$ on $Arrivalsa_t$ over time, especially considering the confirmed long-term relationship between these variables (see Ito and Sato (2008) and Liu et al. (2019) for the exemplary use of accumulated impulse response). Figure 6.15 displays the accumulated response of $Arrivalsa_t$ to $T_1\_Likes_t$ shock (i.e., a change by one standard deviation). The curve commences a rapid descent from period 1 (i.e., the first month), indicating a substantial initial negative response of seasonally adjusted tourist arrivals to an unexpected shock to the number of likes on boycott tweets regarding the British Museum's association with fossil fuel companies (Topic 1). This could be interpreted as a substantial decrease in tourist arrivals as a reaction to a sudden, unexpected increase in likes on boycott tweets (Topic 1). The curve reaches its nadir at the midpoint of the observation period, signifying that the adverse effect of an increase in boycott tweet likes on tourist arrivals is most pronounced at this juncture. It suggests that, on average, it takes about half of the total observation period (i.e., six months) for the impact of a shock to boycott tweets (Topic 1) to fully manifest in the form of a decrease in tourist arrivals. Subsequently, the accumulated response curve starts to ascend gradually and finally converges to nearly zero towards the end of the observation period. This upward shift signifies a recovery in tourist arrivals, suggesting that the initial negative impact of the shock to boycott tweet likes (Topic 1) gradually wanes over time. By the end of the observation period, the effect of the initial shock is almost absorbed completely, and the tourist arrivals return to their long-run equilibrium level. In other words, while an unanticipated increase in likes on boycott tweets (Topic 1) may lead to a temporary decrease in tourist arrivals, this effect does not persist in the long run. This dynamic response pattern implies that while shocks to $T_1\_Likes_t$ have a significant short-to-medium-term impact on $Arrivalsa_t$, these effects are transitory and diminish over time.

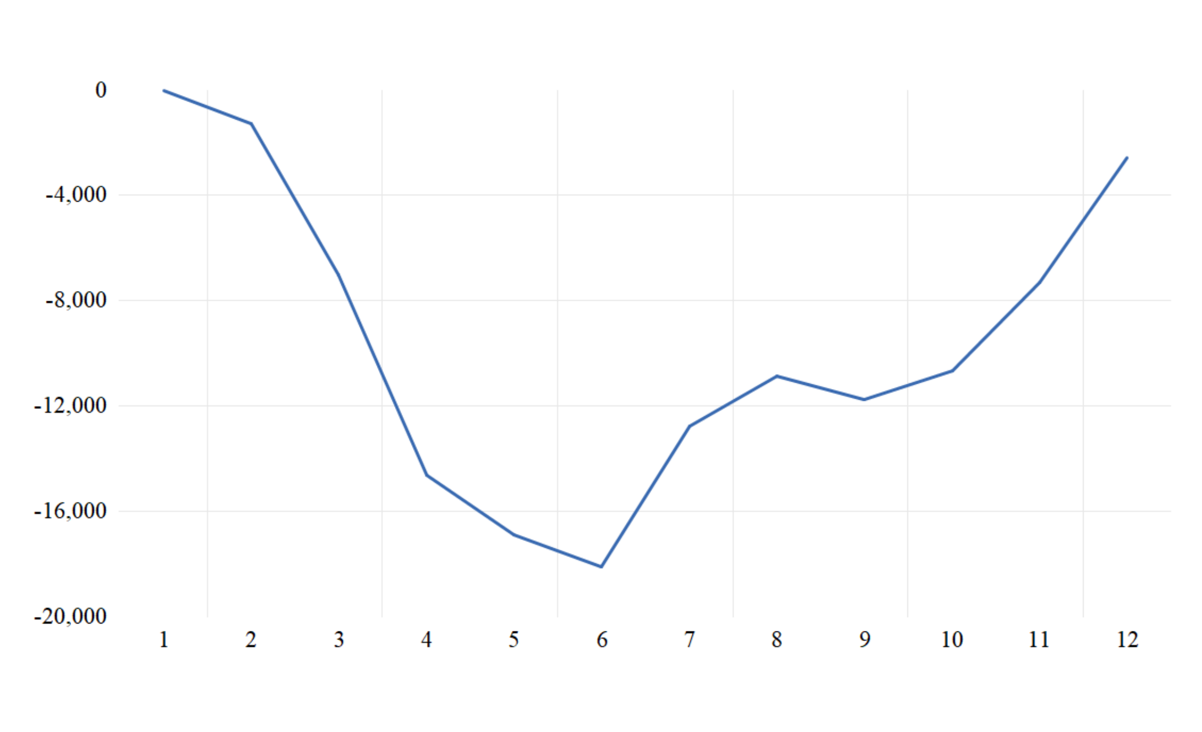

Figure 6.15 Accumulated response of $Arrivalsa_t$ to $T_1\_Likes_t$ shock

Figure notes: The Cholesky (d.f. adjusted) factors are used to generalise Impulses.

6.3.3 Granger causality test

After investigating the long-run relationships between co-integrated variables ($Arrivalsa_t$ to $T_1\_Likes_t$) using a VECM model, the focus now shifts to examining the short-run dynamics. Specifically, the Granger causality test is utilized on the VECM(5) model to analyse the short-term relationship between $\Delta T_1\_Likes_t$ and $\Delta Arrivalsa_t$. It aims to provide more insights into how deviations from the long-term equilibrium are corrected in the short term. Furthermore, the dynamic relationships between $Arrivalsa_t$ and seven other variables derived from boycott tweets will also be examined by applying the Granger causality test. The test results are summarized in Tables 6.8 and 6.9 below.

Table 6.8 Results of the Granger causality test to the VECM (5) model

Lag	Null hypothesis	Chi-sq	Probability	Conclusion
5	$\Delta T_1\_Likes_t$ does not Granger cause $\Delta Arrivalsa_t$	7.495	0.186	Accept
5	$\Delta Arrivalsa_t$ does not Granger cause $\Delta T_1\_Likes_t$	6.466	0.264	Accept

Table notes: "Δ" indicates the first-order differencing operation applied to the variable.

Table 6.9 Results of the Granger causality test to models (1) and (3) to (8)

Lag	Null hypothesis	F-Statistics	Probability	Conclusion
1	$T_1\_Volumes_t$ does not Granger cause $Arrivalsa_t$	0.077	0.782	Accept
1	$Arrivalsa_t$ does not Granger cause $T_1\_Volumes_t$	8.618	0.005	*Reject*
1	$T_1\_Retweets_t$ does not Granger cause $Arrivalsa_t$	0.113	0.737	Accept
1	$Arrivalsa_t$ does not Granger cause $T_1\_Retweets_t$	1.281	0.262	Accept
1	$T_1\_Replies_t$ does not Granger cause $Arrivalsa_t$	0.211	0.648	Accept
1	$Arrivalsa_t$ does not Granger cause $T_1\_Replies_t$	0.041	0.840	Accept
1	$T_2\_Volumes_t$ does not Granger cause $Arrivalsa_t$	1.385	0.243	Accept
1	$Arrivalsa_t$ does not Granger cause $T_2\_Volumes_t$	1.413	0.239	Accept
4	$T_2\_Likes_t$ does not Granger cause $Arrivalsa_t$	0.263	0.901	Accept
4	$Arrivalsa_t$ does not Granger cause $T_2\_Likes_t$	2.272	0.072	*Reject*
1	$T_2\_Retweets_t$ does not Granger cause $Arrivalsa_t$	0.197	0.659	Accept
1	$Arrivalsa_t$ does not Granger cause $T_2\_Retweets_t$	1.582	0.213	Accept
3	$T_2\_Replies_t$ does not Granger cause $Arrivalsa_t$	0.684	0.565	Accept
3	$Arrivalsa_t$ does not Granger cause $T_2\_Replies_t$	2.200	0.097	*Reject*

Table notes: "Δ" indicates the first-order differencing operation applied to the variable.

Referring to Table 6.8, the Granger causality test assesses whether $\Delta T_1\_Likes_t$ Granger cause $\Delta Arrivalsa_t$. The test yields a Chi-square statistic of 7.495 with a corresponding p-value of

0.186. Given that the p-value surpasses the 0.05 threshold, we lack sufficient evidence to reject the null hypothesis. Consequently, we uphold the null hypothesis, inferring that $\Delta T_1\_Likes_t$ do not Granger cause variations in $\Delta Arrivalsa_t$. In parallel, when investigating whether $\Delta Arrivalsa_t$ Granger cause $\Delta T_1\_Likes_t$, the Chi-square statistic is 6.466 with a corresponding p-value of 0.264. As this p-value is also greater than 0.05, we accept the null hypothesis, indicating that $\Delta Arrivalsa_t$ do not Granger cause fluctuations in $\Delta T_1\_Likes_t$. These outcomes underscore that, in terms of Granger causality, neither $\Delta Arrivalsa_t$ nor $\Delta T_1\_Likes_t$ can be employed to predict each other in the short run.

Turning to Table 6.9, when testing whether $Arrivalsa_t$ Granger causes $T_1\_Volumes_t$, the F-statistic is 8.618 with a corresponding p-value of 0.005. This p-value is less than 0.05. Thus, we reject the null hypothesis, indicating that $Arrivalsa_t$ does Granger cause changes in $T_1\_Volumes_t$. In the case of $Arrivalsa_t$ Granger causing $T_2\_Likes_t$, the test yielded an F-statistic of 2.272 with a p-value of 0.072. While this p-value is greater than 0.05, it is still relatively low, suggesting potential short-run predictability from $Arrivalsa_t$ to $T_2\_Likes_t$, but it does not pass the strict 5% significance level. Similarly, when testing whether $Arrivalsa_t$ Granger causes $T_2\_Replies_t$, the F-statistic is 2.200 with a p-value of 0.097. This p-value, although greater than 0.05, is relatively low, indicating some potential for short-run predictability from $Arrivalsa_t$ to $T_2\_Replies_t$. Once more, it falls short of the stringent 5% significance level. These outcomes collectively propose that $Arrivalsa_t$ might play a role in predicting the short-run changes in $T_1\_Volumes_t$, $T_2\_Likes_t$, and $T_2\_Replies_t$, although the latter two do not satisfy the strict significance threshold.

The cumulative findings from our Granger causality tests suggest a predictive role for $Arrivalsa_t$ in determining short-term variations in $T_1\_Volumes_t$, $T_2\_Likes_t$, and $T_2\_Replies_t$. However, it is important to note that the reciprocal relationships - those of the boycott tweets-generated variables Granger causing $Arrivalsa_t$ - were not established by the tests. Given the core objective of Study 3 - to evaluate the value of commonly excluded Twitter “noise” (i.e., - tweets that neither stimulate visits nor signal tourists’ intentions to visit) in supplementing tourism demand forecasts at the attraction level - these outcomes do not provide sufficient evidence to warrant further exploration of the impulse response of $T_1\_Volumes_t$, $T_2\_Likes_t$, and $T_2\_Replies_t$ to $Arrivalsa_t$.

6.4 Findings

In Study 2, we discovered that the majority of the "noise" are tweets associated with boycott topics (refer to Appendix 1) utilizing a state-of-the-art keyword-in-context search approach. Building on the insights gained from Study 2, Study 3 embarks on an exploration of the signalling value of boycott tweets for tourist arrivals at the British Museum. Utilizing a set of analytical tools that encompass sentiment analysis, text mining, discerning screening, and the guided LDA algorithm, we initially constructed a thorough three-stage framework. This approach efficiently recognises boycott tweets and successfully unravels their related topics. Moreover, by harnessing information from archival sources such as newspaper articles and press releases, this study probes the dynamic interactions between visitor numbers and variables derived from boycott tweets. Lastly, through the application of the VAR system and VECM modelling, Study 3 uncovers the significant short-to-medium term effects of shocks to $T_1\_Likes_t$ (i.e., change by a standard deviation) on the seasonally adjusted tourist arrivals $Arrivalsa_t$ with an opposite direction.

The main findings of Study 3 can be summarised as follows. Firstly, the investigation highlights that the contentious subjects of "fossil fuel sponsorship" and "contested artefacts" have generated significant discussion on Twitter. Furthermore, the model-free evidence we present illustrates an upward trajectory in the reactions (i.e., likes, retweets, and replies) to boycott tweets concerning these two areas. In particular, "likes" display the most pronounced engagement compared to other reaction types (i.e., retweets and replies). This heightened engagement noticeably increased during the UK's initial lockdown period, reaching its zenith following the relaxation of travel constraints. The discrepancy between the timing of peak in boycott tweet volume (i.e., $T_1\_Volumes_t$ and $T_2\_Volumes_t$) and the audience's approval (represented by $T_1\_Likes_t$ and $T_2\_Likes_t$) might be partly due to the surge in Twitter usage during the COVID-19 lockdown (Twitter, 2021). Additionally, boycott tweets corresponding to Topic 1 displayed a moderate but immediate uptick following each physical protest.

Second, employing the vector auto-regressive model and the vector error correction model, this study uncovers the dynamic interplay among seasonally adjusted tourist arrivals (i.e., $Arrivalsa_t$) and boycott tweets-generated variables., A long-run equilibrium relationship has been identified between tourist arrivals and the popularity of boycott tweets related to the

British Museum's acceptance of fossil fuel sponsorship (i.e., $T_1\_Likes_t$). Through impulse response analysis, we further assess the magnitudes and time horizons of the response pattern $Arrivalsa_t$ to $T_1\_Likes_t$. The aggregated results imply that when $Arrivalsa_t$ and $T_1\_Likes_t$ deviate from their long-term equilibrium relationship, $Arrivalsa_t$ adjusts in the opposite direction by about 15.6% of the deviation in the next period to restore the equilibrium. Essentially, one standard deviation in $T_1\_Likes_t$ will result in an approximately -15.6% adjustment in $Arrivalsa_t$ to re-establish its long-run equilibrium with $T_1\_Likes_t$ within a short-to-medium term (i.e., 6 months). This implies the importance for policymakers and tourism management to closely observe short-to-medium term fluctuations in likes for boycott tweets related to the British Museum's ties with fossil fuel companies and to devise responsive strategies to lessen their negative impact on tourist arrivals.

As for the variables generated from boycott tweets related to contested artefacts (Topic 2), despite their larger volume, significant dynamics between these variables and tourist arrivals have not been detected. This disparate impact of different boycott themes against the British Museum may be attributed to two potential factors. Firstly, escalating concerns about climate change could spur consumers to shun products and services supported by "dirty energy" providers like BP (Kennedy, 2017), thereby making the boycott theme of "fossil fuel sponsorship" negatively impact tourist arrivals to the British Museum. Secondly, while there have been calls to return artefacts to their places of origin, the British Museum is legally prohibited from doing so (TheBritishMuseum, 2018). Klein et al. (2004) suggest that consumers need to believe that boycotts will effect real change before they translate intention into action. Hence, Twitter boycotts against the British Museum on the theme of "contested artefacts" may not yield a substantial response other than attracting public attention. This conjecture is further reinforced by the fact that the identified physical protests (refer to Section 6.3.4) predominantly pertain to "fossil fuel sponsorship". This insight underlines the necessity of understanding the nuanced impacts of different boycott themes and, accordingly, adopting appropriate response strategies.

Collectively, Chapter 6 (Study 3) fulfils the third research objective of this thesis, which is to evaluate the value of commonly excluded Twitter "noise" (i.e., - tweets that neither stimulate visits nor signal tourists' intentions to visit) for supplementing tourism demand forecasts at the attraction level. The discoveries from Chapter 6 emphasize the importance of confronting the “noise removal fallacy” when dealing with social media data, especially when used for tourism demand forecasting practice. Thus, rather than merely disregarding the noise, efforts should be made to precisely identify and understand it, illuminating its potential contributions to more precise forecasts.

Chapter 7: Discussion and conclusion

7.1 Discussion

This thesis is primarily aimed at ***revealing the value of social media data, especially Twitter data, in forecasting tourism demand at the attraction level***. Addressing this research aim, the thesis adds to the existing body of literature on tourism demand forecasting, while offering practical insights for tourist attractions.

Thus far, researchers studying tourism demand have widely employed search engine queries and website traffic data (Dergiades et al., 2018; Havranek & Zeynalov, 2021; Volchek et al., 2019; Yang et al., 2014), which are only available at aggregate levels and do not distinguish between positive and negative attention or interest (Hu et al., 2022). This prevents researchers from identifying, analysing, and correcting possible biases and limitations in the raw data (Geva et al., 2017). Although Twitter's ability to reveal preferences, supply high-frequency data, and capture shifts in consumer inclinations (Yang et al., 2014) may be helpful when assessing tourism demand, thus far, this rich data source has not been fully explored in the tourism demand forecasting literature. This is surprising, as tourism organizations regularly utilize Twitter as a marketing communication tool (e.g., DMOs' commercial tweets, event tweets, socializing tweets, and tourist attraction tweets: Bigné et al. (2019)), and tourists use Twitter to search for travel-related information (Nicolau et al., 2020), make enquiries (Alkhammash et al., 2019), or share their experiences (Sotiriadis & Van Zyl, 2013). The emergence of studies using social media data (e.g., Facebook likes: Önder et al. (2019), DMOs' tweets: Bigné et al. (2019), and review data: H. Li et al. (2020)) to forecast tourism demand, as well as a call for more studies exploring the value of Twitter data for forecasting tourism demand (Bigné et al., 2019), and the apparent lack of studies examining tourism demand for attractions, further reinforces this research's motivation.

Previous research has investigated the utility of Twitter data for understanding what drives user engagement (Berman et al., 2019; Borah et al., 2020; Toubia & Stephen, 2013) and disengagement activities (Villanova & Matherly, 2023), assessing its impact on users' perceived influence (Valsesia et al., 2020), deciphering customer perceptions (Herhausen et al., 2023), influencing financial outcomes (Borah et al., 2020; Hennig-Thurau et al., 2015; Lacka et al., 2022; Lu et al., 2022), and managing service recoveries (Herhausen et al., 2023). While the usefulness of Twitter data has been widely studied in diverse contexts, including movie sales (Hennig-Thurau et al., 2015; Lu et al., 2022), political debate (Berman et al., 2019), the stock market (Borah et al., 2020; Lacka et al., 2022), service failure (Herhausen et al., 2023), boycott campaigns (Liaukonytė et al., 2022), and sports (Villanova & Matherly, 2023), its application in the tourism context, and specifically to tourism demand forecasting research, remains largely unexplored. This is particularly true in the context of tourist attractions, for which the literature on tourism demand is sparse.

This thesis, grounded on the principles of signalling theory (Spence, 1973) and the many-to-many communication framework (Chaffey & Ellis-Chadwick, 2019), expands on the understanding of how to extract valuable signals from Twitter data to forecast tourism demand for attractions such as the British Museum. It also explores the signalling value of multiple communication flows, for tourist arrivals. Specifically, it looks at three types of communication flow: direct communication with attractions, peer-to-peer conversations, and communication from attractions to the broader audience. Finally, it explores the potential signalling value that is often ignored in "noise" data (tweets that neither stimulate visits nor signal tourists' intentions to visit) in the context of tourism demand forecasting. This situates this thesis at the forefront of research investigating the signalling value of Twitter data for the volume of tourist arrivals at the attraction level.

Filling the aforementioned knowledge gap, Chapter 4 of this research discovers that the volume of tweets and the change in the number of likes of tweets Granger-cause seasonally adjusted tourist arrivals to the British Museum, which serves as a case study in this research. These findings address the first objective of this research by showing that, for the purpose of attraction-based tourism demand forecasting, Twitter can indeed be seen as a valuable supplement to other data sources (e.g., search engines, online travel forums, and Facebook) that have been examined by previous research. The findings of Chapter 4 reinforce the necessity for investigating further the potential of Twitter-generated variables to improve demand forecasts at the attraction level.

In their research, Bigné et al. (2019) assessed tweets disseminated by DMOs, but they ignored the large volume of tweets disseminated by tourists. Despite the call by Bigné et al. (2019) to consider the diverse array of tourism stakeholders generating social media data, existing literature often overlooks the role of communication flows nestled within the Twitter-mediated signalling environment. These communication flows depict the process of information dissemination between tourists and other tourists or attractions and, as this research shows, offers valuable insights for tourism demand forecasting research, such as the importance of including the commonly neglected countersignalling process in the framework. Consequently, building on the insights revealed in Chapter 4, Chapter 5 goes beyond the variables such as tweet volumes, likes, retweets, and replies previously examined in Chapter 4, incorporating the three communication flows identified in the literature review through text analysis. The results from competing forecasting models suggest that tracing the communication flow for attraction-specific tweets can bolster the accuracy of short-term tourist arrival forecasts. This improvement is particularly notable when incorporating tweets reflecting tourists' direct interaction with the attraction.

As discussed in Section 2.6, signalling theory emphasises the importance of the credibility and authority of signals (Connelly et al., 2011; Tumasjan et al., 2021), a concept particularly relevant in the context of direct interactions between attractions and tourists on Twitter. Communications from attractions and local experts are often perceived as credible due to their authoritative source (Edwards et al., 2017; Xiang & Gretzel, 2010), providing specific and targeted information that constitutes strong signals. This type of information can directly influence tourists' decisions to visit, aligning with signalling theory's assertion that clear and relevant signals from a trusted source have a significant impact on the receiver's decision-making process (e.g., Bigné et al. (2019); Önder et al. (2019)). In contrast, the credibility of signals in tourist-to-tourist interactions on Twitter is more variable. Personal experiences and opinions shared by tourists may lack the reliability and relevance of information from attractions (Munar & Jacobsen, 2014). These signals are often indirect and may not strongly impact immediate decision-making, instead providing emotional support (Munar & Jacobsen, 2014) and influencing long-term perceptions (e.g., trust and expectations: Narangajavana et al. (2017)) or general awareness (Leung et al., 2013), thereby explaining their limited capability of generating short-term tourist arrival forecasts. The findings further highlight the significance of tweet volumes and likes, within direct communications between tourists and attractions, for forecasting tourist arrivals. In essence, Chapter 5 addresses the second objective of this thesis and, by doing so, enhances both the literature on tourism demand forecasting and that on social media signalling by emphasizing the value of deciphering Twitter communication flows for generating improved demand forecasts for tourist attractions.

Finally, this thesis revolves around the critical role of the "noise removal fallacy" in Twitter-based tourism demand forecasting research. The identified "noise" could potentially contain valuable information (Zafarani et al., 2014). In Chapter 6, we illustrate a method for extracting tweets expressing the public's negative concerns, primarily centred on two boycott campaigns (i.e., fossil fuel sponsorship and contested artefacts) for the British Museum. Excluding these tweets facilitates the cleansing of the data, particularly the "like" attribute of the collected tweets, and this, in turn, generates more precise forecasts of tourist arrivals, as discussed in Chapter 5. However, as we contend in Chapter 6, these negative tweets, often dismissed as "noise" in the existing tourism demand forecasting literature, can indeed serve as insightful information sources for tourist attractions.

In particular, we investigate the dynamics between boycott tweets and tourist arrivals and find significant short-to-medium-term effects of shocks (i.e., changes by a standard deviation) in the volume of likes associated with tweets about boycotting the British Museum due to its fossil fuel sponsorship. Specifically, the findings suggest that the seasonally adjusted tourist arrivals negatively responded to a shock in the volume of likes of such tweets. However, the boycott campaigns criticizing the British Museum's contested artefacts did not yield a significant impact on tourist arrivals. This discrepancy warrants further investigation. One potential avenue worth exploring is whether tourists believe that the "stolen artefacts" boycott campaign will lead to any tangible change. Chapter 6 notes the legal constraints faced by the British Museum regarding the return of artefacts to their places of origin (TheBritishMuseum, 2018), presenting a significant barrier to the effectiveness of boycotts. Klein et al. (2004) emphasize that for customers to translate intention into action, such as participating in a boycott, they need to believe that their actions will effect tangible change. In the case of the British Museum, these legal restrictions may lead the public to perceive the boycotts as unlikely to effect change. Consequently, Twitter campaigns against the British Museum's exhibition of "stolen artefacts" may not have significantly impacted tourist arrivals, despite raising public awareness. The public might sympathize with the cause but still choose to visit the museum, especially if they believe their actions will not lead to a tangible outcome.

In the literature review, we noted the inherently unstructured and complex nature of Twitter data (Li et al., 2021). This complexity calls for sophisticated pre-processing methods to extract valuable information (Berger et al., 2020). To tackle this challenge, this thesis employs several text-mining techniques. These include textual analysis, sentiment analysis, and both unsupervised and semi-supervised topic-modelling algorithms. These methods facilitate the identification of multiple signallers and receivers within the Twittersphere, along with the classification of Twitter signals, irrelevant information, and noise. Competing forecasting models such as the VAR model, the ARIMA model, the ARDL model, and the MIDAS-AR model are deployed in Chapters 4 and 5. These models aim to explore the potential of Twitter signals for predicting tourist arrivals to the British Museum. Chapter 6 explicitly targets the investigation of the signalling value of "noise" with respect to tourist arrivals. To achieve this, we use Granger causality tests, the vector error correction model (VECM), and impulse response analysis. Through these sophisticated techniques, we have made strides towards understanding the complexities of Twitter data and how to process and utilize it for predicting outcomes in the tourism sector.

7.2 Contributions

This thesis makes significant contributions to the literature and practice, as discussed below, in line with the research aim and objectives presented in the literature review section. These contributions provide novel approaches for utilizing Twitter data in tourism demand forecasting. Due to the complexity of the signalling environment facilitated by Twitter, the contributions converge towards deciphering the signalling process among tourist attractions and tourists on Twitter. This allows for the extraction of embedded signals from various communication flows and the assessment of their predictive power for forecasting tourist arrivals. Moreover, this thesis contributes by evaluating the signalling value of typically excluded social media "noise" for demand forecasting research. These theoretical advancements rest upon a novel methodological approach designed to analyse the semantic features and underlying topics of the gathered tweets. This approach reinforces the necessity of considering the multifaceted attributes of Twitter signals when predicting tourism demand.

7.2.1 Theoretical contributions

As highlighted in the literature review, while the signalling value of Twitter data has been examined in various marketing and management studies (refer to Section 2.5.2), there has been a lack of research thoroughly investigating the potential of Twitter data for forecasting tourist arrivals, particularly at the attraction level. This thesis bridges this gap in the literature and makes five key contributions, as detailed below.

First, this thesis addresses an underexplored area in the tourism demand forecasting literature by focusing on tourist attractions. Although a substantial amount of research highlights the importance of tourism demand forecasting for improving management within the tourism sector (Frechtling, 2001; H. Li et al., 2020), there is a notable lack of studies focusing on forecasting at the level of individual attractions. Existing research concentrates on larger geographic scopes, such as provinces, countries, or regions (Bangwayo-Skeete & Skeete, 2015; Dergiades et al., 2018; Gunter & Önder, 2016; Havranek & Zeynalov, 2021; S. Li et al., 2018; Yang et al., 2015). These larger areas are typically not as restricted by physical space limitations as are individual attractions (Canestrelli & Costa, 1991; Manning et al., 2002). Therefore, by presenting a comprehensive case study of the British Museum, this thesis contributes to the literature on tourism demand forecasting, with a focus on individual tourist attractions.

Second, the data employed in such studies are often gathered at relatively low frequencies (i.e., monthly, quarterly, or yearly aggregated data for tourist arrivals and indicators), providing insights primarily conducive to formulating long-term crowd management strategies (Song et al., 2019). Exploration of high-frequency tourism demand forecasts remains minimal (Bi et al., 2021; Bi et al., 2020; Volchek et al., 2019), yet these high-frequency forecasts have the potential to enrich longer-term predictions by delivering more detailed, short-term perspectives beneficial for effective crowd management (H. Li et al., 2020). Innovatively, this thesis is one of the first few to adopt mixed-frequency data sampling in tourism demand forecasting (other examples: Havranek and Zeynalov (2021); Hu et al. (2022); Önder et al. (2019)). The MIDAS technique is an effective way to avoid information loss from higher-frequency data (Kuzin et al., 2011). The empirical results of this study indicate that the MIDAS-AR model (including high-frequency Twitter data) generated more precise forecasts than competing models using aggregated Twitter data as explanatory variables (i.e., the ARDL and ARIMAX models) and benchmark ARIMA models. Therefore, this work contributes to the model development literature in tourism demand forecasting.

Third, this research stands out as one of the pioneering works that integrate Twitter data generated by a diverse array of tourism stakeholders. This scope includes the attractions themselves, experienced tourists, and a wider audience exhibiting an interest in the attraction. In contrast to studies focused on specific holiday seasons, this research encompasses an expansive timeframe of six years, thereby offering a long-term viewpoint of Twitter data's value for forecasting tourist arrivals. In doing so, it responds to the call by Bigné et al. (2019), which encourages research in Twitter-based tourism demand forecasting to cover a broader range of tourism stakeholders and be extended over a lengthy observation period.

Fourth, addressing the second research objective, this thesis is pioneering in providing evidence supporting the advantages of deciphering Twitter communication flows. As highlighted in the literature review, prior research has validated the use of various attributes of Twitter data as insightful signals from which to infer consumer interest (Lu et al., 2022), attention (Liaukonytė et al., 2022), perceptions (Herhausen et al., 2023), and behaviour (e.g., engagement: Berman et al. (2019) and disengagement: Villanova and Matherly (2023)). However, research that contemplates the intercommunication dynamics among numerous signallers and receivers in the Twitter-mediated signalling environment remains limited. The prevailing focus of existing research is on communication initiated by DMOs and targeted at potential tourists (Bigné et al., 2019). To extend the research landscape in this area, this thesis delineates the interaction patterns among attractions and tourists within the Twittersphere. Specifically, three distinct communication flows are identified, namely direct communication with attractions, peer-to-peer conversations, and communication from attractions to the broader audience. In doing so, this thesis illustrates that a concentrated focus on direct information exchange between tourists and attractions can significantly enhance the precision of tourist arrival forecasts for attractions. This finding reinforces the value of mapping communication flows in tourism demand forecasting.

Lastly, this thesis brings attention to an often-overlooked aspect of tourism demand forecasting research: the propensity to disregard negative signals, which include boycott tweets. It must be acknowledged that not all social media conversations concerning a tourist attraction or destination are positive (Su et al., 2022). Nevertheless, negative tweets are often excluded from analysis or classified as "noise", a practice this research aims to reassess. By recognizing the potential influence of negative attention and sentiment, this thesis takes a significant initial step towards investigating the signalling value of boycott tweets for tourist arrivals at the attraction level. The empirical findings suggest that the public's growing approval (endorsement) of tweets (measured by tweet likes: Saxton et al. (2019)) criticizing the British Museum's fossil fuel sponsorship correlates negatively with the volume of tourist arrivals. On the other hand, the effects generated by signals from tweets criticizing the British Museum's exhibition of contested artefacts do not exactly correlate with tourist arrivals. The insights gleaned from this thesis, therefore, enrich our understanding of a broad spectrum of elements that can be used to forecast tourist arrivals, encompassing both positive and negative signals. It emphasizes the importance of exploring the underlying semantic characteristics of social media signals, including sentiment and topic, while highlighting that negative tweets – often disregarded as noise – may indeed contain meaningful information. These insights not only augment the body of literature on tourism demand forecasting but also contribute to the discourse surrounding consumer/tourist boycotts.

7.2.2 Methodological contributions

In addition to the five theoretical contributions outlined above, this thesis offers invaluable insights into methodological considerations for informative signal extraction, the tuning of tourism demand forecasting models, and noise identification tailored to diverse research objectives.

First of all, in the process of extracting informative signals from collected tweets, careful attention must be paid not only to the detection of noise but also to the consideration and removal of irrelevant tweets. As explored in the literature review, Twitter data, unlike search engine query data which is pre-processed via categorization, may contain irrelevant information when gathered through specific queries (Ibrahim & Wang, 2019a, 2019b). For example, in an effort to capture the broadest range of relevant data, we employed a search-query-based approach to collect information across the entire Twitter platform. While this strategy yielded a dataset encompassing a wide range of signallers and receivers, it also inevitably led to the inclusion of more irrelevant data. For example, a tweet like “British Motor Museum gears up for 30th birthday celebration” would be correctly excluded as irrelevant, even though it contains both of the sought-for words of “British” and “Museum”. However, the current literature devotes limited attention to this challenge. In the context of our Twitter dataset pertaining to the British Museum, we used a regular expression[12] to filter the “text” column to retain only those tweets that contained the phrase “British Museum”. This filtration approach enabled us to exclude tweets containing the words “British” and “Museum” separately, as opposed to the specific phrase “British Museum”.

Second, to effectively detect boycott tweets amid the vast Twitter database, we propose a boycott tweet detection approach. This framework combines VADER sentiment analysis, judgmental screening, and two probability-based topic-modelling algorithms: LDA and guided LDA. The goal of this combination is to mitigate the inefficiencies of a purely judgemental approach (Berger et al., 2020), the accuracy issues encountered with dictionary-based sentiment analysis models like VADER (Bhatia et al., 2015), and the interpretational constraints associated with pure probability-based topic-modelling algorithms like LDA (Berger et al., 2020). Our framework comprises three stages, with the first two serving as preparatory steps for the guided LDA algorithm. Importantly, we recommend feeding only tweets that directly engage with the attraction into the guided LDA algorithm. This decision stems from two primary considerations. First, our findings from Chapter 5 accentuated the importance of tweets that were direct interactions between Twitter users and the tourist attraction, in our case, the British Museum. These direct communication tweets demonstrated higher signalling value than tweets from other communication flows, making them an essential focus for data analysis. Second, comparing the findings from Chapters 5 and 6 unveiled the presence of irrelevant tweets in the unprocessed dataset. These irrelevant tweets could disrupt the topic-modelling process by skewing its output. By prioritizing tweets directly engaging with the British Museum, we could reduce the impact of this bias, thus enhancing the precision and accuracy of the topic-modelling process.

[12] str.contains(r'\british museum\b')

The approach proposed here offers a balanced method, amalgamating distinct techniques to address the intricacies of social media data. In particular, it addresses the intrinsic semantic features of each tweet and the general topic distributions. This contributes significantly to the research methodology in terms of the detection of noise, mainly when centred on measuring the impact of social media noise in the tourism literature. Additionally, Chapter 5 presents a cutting-edge keyword-in-context search approach for identifying tweets associated with boycott topics, specifically for demand forecasting. The list of keywords was methodically derived from an analysis of the most popular tweets centred on boycott themes. This methodological design ensured that the list encompassed the most representative keywords related to boycotts, reflecting the most prevalent boycotting campaigns. In this process, we accept the limitation of potentially sacrificing the scope of coverage of negative sentiment instigated by other reasons or campaigns beyond our current knowledge. Despite the risk of over-screening due to the simplicity of the keywords included in the list, this approach is convenient and efficient for demand forecasting. The rationale behind this is that tweets exhibiting boycott intentions constitute only a minor proportion of the overall data (that is, less than 5%). This method, despite its potential limitations, significantly streamlines the process and facilitates the generation of more precise demand forecasts.

Third, Chapters 4 and 5 uncover a compelling correlation between the volume of tweets and the number of seasonally adjusted tourist arrivals. Specifically, the research indicates that a unit change in the volume of tweets incites a favourable increase in tourist arrivals. However, it is worth noting that this advantageous reaction does not extend over a prolonged period – it is rather fleeting in nature. This finding aligns with Önder et al. (2019) and Liu et al. (2019), suggesting that the lag between individuals' travel planning and visiting behaviour is short.

7.3 Practical recommendations

In addition to the theoretical and methodological contributions, this thesis unveil several practical implications that could help tourist attractions optimize their Twitter marketing strategies and monitor tourist flows. First, tourist attractions like the British Museum should pay attention to communication on Twitter; they should monitor the volume of tweets and reply to tourists’ enquiries as they can provide insights into tourists’ arrivals. Moreover, they should pay close attention to the timing and frequency of tweet volumes (i.e., including short lag variables derived from tweet volumes) to leverage their forecasting potential optimally.

According to the findings from Chapters 4 and 5, since it might be overwhelming to monitor all Twitter activity, instead of monitoring Twitter communications on a broad level, attractions should pay attention to direct conversations they have with Twitter users. The captured data based on communication flow could help attractions to generate improved forecasts of tourist arrivals and optimize on-site tourist flow management strategies. Specifically, for demand forecasting and crowd management purposes, attractions such as the British Museum should utilize the volume of tweets and likes generated from tweets posted by Twitter users that mention the British Museum, rather than indicators generated from all tweets related to the British Museum. In addition, it is crucial to monitor and respond to changes in users' tweets (Sheng et al., 2019) within the critical window (i.e., around two months) to capitalize on the positive relationship between tweet volumes, likes, and tourist arrivals.

In light of the increasing trend of museums becoming focal points for boycott campaigns, particularly climate protests, it has become imperative for tourist attractions to diligently monitor those signals identified as "noise" for demand forecasting purposes, particularly on platforms like Twitter. The escalating propensity for these institutions, especially those associated with fossil fuel companies, to be targeted for boycotts (Adams, 2019), as brought to light in Chapter 6, proffers critical insights for the tourism sector. First, it needs to be vigilant about the prevalence of boycott-related tweets, especially those linked to climate protests, and respond in a proactive manner. This could involve directly addressing criticisms, dispelling misconceptions, and prominently displaying institutions' commitment towards sustainability and environmental responsibility (Su et al., 2022). To elucidate, tourism attractions could enhance their communication strategies, amplifying transparency about their partnerships, funding sources (Motion, 2019), and particularly challenges associated with funding in crisis scenarios (Yu et al., 2020). Second, attractions like the British Museum should be prepared for potential physical protests. This study underscores the connection between Twitter-based boycott campaigns and real-world demonstrations, implying that these institutions should anticipate potential protests and develop comprehensive strategies for managing such occurrences (Yu et al., 2020). By contemplating these practical implications, tourist attractions can better navigate the complexities presented by online boycott campaigns triggered by climate activism. Concurrently, they can work towards evolving into more sustainable and environmentally responsible organizations that retain the trust and garner the continued support of the public.

7.4 Limitations and future research

Despite its exploratory nature, this study has some limitations that should be acknowledged and considered for future research.

First, one significant constraint is the limited generalizability of the findings. This study primarily draws its empirical evidence from case studies on the British Museum. While the British Museum offers a unique context for research, its scale and prominence may not mirror the circumstances of smaller, less known attractions. Therefore, the results may only partially encapsulate the full range of tourism dynamics. This limitation underscores the need for a broader scope of inquiry in future research. It would be beneficial to apply the methodologies and analyses used in this study across a more diverse selection of tourist attractions, mainly focusing on those smaller or less prominent than the case studies selected here. Doing so could help reinforce the relevance and applicability of Twitter-generated indicators for predicting tourist demand across various scales and types of attractions. Furthermore, it could provide a richer understanding of how different attractions interact with their audiences on social media, adding nuance and depth to the current understanding of digital signalling in the tourism industry.

Second, the limitations concerning data frequency present a significant challenge. While the thesis navigates around the issue of low-frequency tourist arrival data by implementing mixed-frequency data sampling, it only partially utilizes the high-frequency nature of Twitter-generated indicators. This aspect of the research could have benefitted significantly from the continuous, real-time data that Twitter provides, potentially capturing more granular variations and trends related to tourism demand. Future research avenues could be focused on leveraging this high-frequency Twitter data in a more effective way. If more frequent data on tourist arrivals (for instance, on a monthly or daily basis) could be accessed, this could allow for more nuanced and real-time analysis of tourism trends. Moreover, it could lead to the discovery of intricate patterns that might be lost when observing data at a lower frequency (Lacka et al., 2022). In tandem with this, the integration of more advanced modelling techniques, such as machine learning algorithms, could help fully utilize this rich, high-frequency dataset. Machine learning's inherent ability to handle complex, multidimensional data and recognize subtle patterns could help shed light on the dynamic relationship between social media discourse and tourism demand (Song et al., 2019). This could enhance the validity of the results obtained in this research, making it more robust and thereby adding a more valuable contribution to the tourism demand forecasting literature.

While Chapter 6 establishes that boycott tweets represent informative signals that cannot be overlooked, this thesis does not explicitly focus on boycott issues within the tourism context. As per a study by Yu et al. (2020), political boycotts can result in substantial declines in tourist arrivals. However, there is a conspicuous gap in the current literature regarding the study of ethical-consumerism-driven boycotts in the context of tourism, particularly at the attraction level. Differing from political boycotts, ethical-consumerism-inspired boycotts refer to situations where consumers deliberately avoid products or services produced by entities (e.g., companies, organizations, or countries) that disregard human rights or engage in environmentally detrimental practices (Koku, 2022). Instances of the public boycotting or endorsing boycotts in response to a museum's fossil fuel sponsorship or exhibition of contested artefacts exemplify this form of consumer activism. Future research could further explore the discrepancy between the public's expressed intentions and their actual behaviour in the context of consumer activism (Lasarov et al., 2021). Such research might examine if the legal constraints faced by the British Museum indeed account for the varying impact of different Twitter boycott campaigns on tourist arrivals. Moreover, due to data limitations, the existing design of this thesis did not factor in the dynamic interaction between individuals' boycott-related behaviour on Twitter and tourist arrivals amid the COVID-19 pandemic and in the period following it. There have been suggestions that the public tends to be more tolerant of museum sponsorship during crisis scenarios such as the COVID-19 pandemic (Biraglia & Gerrath, 2020). Accordingly, future studies could delve into this aspect, potentially generating valuable insights into the interplay between public sentiment, sponsorship, and tourism demand during and after crisis periods.

Lastly, the methodology proposed in this thesis for detecting boycott tweets is based on a semi-supervised topic-modelling algorithm, specifically, the guided LDA algorithm. This choice of method was aimed at providing a more efficient way for researchers and attraction managers to monitor and analyse boycott tweets. However, the method's efficacy could be improved with the availability of more resources. If future research is equipped with more extensive human-coding capabilities and computing power, creating a fully labelled dataset could be a viable approach. This labelled dataset would allow researchers to test and compare the performance of various topic-modelling algorithms, including but not limited to unsupervised models such as auto-encoder-based algorithms and supervised models such as ANN, back-propagation neural networks (BPNN) and LSTM networks. By comparing these different methods, researchers could evaluate the strengths and weaknesses of each approach in the context of topic modelling in Twitter data, leading to more effective and precise methodologies for future studies. Such an enhanced methodology would not only further the academic understanding of boycott phenomena but also provide practical tools for stakeholders to use to manage and respond to such events more effectively.

References

Adams, G. K. (2019). Museums increasingly targeted as sites of protest. *Museums Association*. https://www.museumsassociation.org/museums-journal/news/2022/12/british-museum-seeks-to-no-longer-be-a-destination-for-climate-protest/

Addley, E. (2023). Fresh doubts raised over future BP funding of British Museum. *The Guardian*. https://www.theguardian.com/culture/2023/feb/28/fresh-doubts-raised-over-future-bp-funding-of-british-museum

Adedoyin, F. F., Seetaram, N., Disegna, M., & Filis, G. (2021). The Effect of Tourism Taxation on International Arrivals to a Small Tourism-Dependent Economy. *Journal of Travel Research*, *62*(1), 135-153.

Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, *34*, 100658.

Alaimo, C., Kallinikos, J., & Valderrama, E. (2020). Platforms as service ecosystems: Lessons from social media. *Journal of Information Technology*, *35*(1), 25-48.

Alkhammash, E. H., Jussila, J., Lytras, M. D., & Visvizi, A. (2019). Annotation of smart cities Twitter micro-contents for enhanced citizen's engagement. *IEEE Access*, *7*, 116267-116276.

Andreou, E., Ghysels, E., & Kourtellos, A. (2011). Forecasting with mixed-frequency data. In *The Oxford handbook of economic forecasting*.

Andrzejewski, D., & Zhu, X. (2009). Latent dirichlet allocation with topic-in-set knowledge. Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing,

Andrzejewski, D., Zhu, X., Craven, M., & Recht, B. (2011). *A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic*.

Assaf, A. G., Kock, F., & Tsionas, M. (2022). Tourism during and after COVID-19: An expert-informed agenda for future research. *Journal of Travel Research*, *61*(2), 454-457.

Asteriou, D., & Hall, S. G. (2007). *Applied econometrics : a modern approach using EViews and Microfit / Dimitrios Asteriou and Stephen G. Hall* (Revised edition. ed.). Palgrave Macmillan.

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01,

Audi, R. (2010). *Epistemology: A contemporary introduction to the theory of knowledge*. Routledge.

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, *36*(2), 334-357.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, *46*, 454-464.

Bansal, P., Smith, W. K., & Vaara, E. (2018). New ways of seeing through qualitative research. In (Vol. 61, pp. 1189-1195): Academy of Management Briarcliff Manor, NY.

Basuroy, S., Desai, K. K., & Talukdar, D. (2006). An empirical investigation of signaling in the motion picture industry. *Journal of marketing research*, *43*(2), 287-295.

BBC. (2022). National Portrait Gallery's BP sponsorship to end. *BBC*. https://www.bbc.co.uk/news/entertainment-arts-60476760

Becker-Olsen, K. L., & Hill, R. P. (2006). The impact of sponsor fit on brand equity: The case of nonprofit service providers. *Journal of service research*, *9*(1), 73-83.

Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, *96*, 355-367.

Bennett, C. M., Kim, H., & Loken, B. (2013). Corporate sponsorships may hurt nonprofits: Understanding their effects on charitable giving. *Journal of Consumer Psychology*, *23*(3), 288-300.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1-25.

Berman, R., Melumad, S., Humphrey, C., & Meyer, R. (2019). A tale of two Twitterspheres: Political microblogging during and after the 2016 primary and presidential debates. *Journal of marketing research*, *56*(6), 895-917.

Bertrand, J. W. M., & Fransoo, J. C. (2002). Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*.

Bhatia, P., Ji, Y., & Eisenstein, J. (2015). Better document-level sentiment analysis from rst discourse parsing. *arXiv preprint arXiv:1509.01599*.

Bi, J.-W., Li, C., Xu, H., & Li, H. (2021). Forecasting Daily Tourism Demand for Tourist Attractions with Big Data: An Ensemble Deep Learning Method. *Journal of Travel Research*, 00472875211040569.

Bi, J.-W., Liu, Y., & Li, H. (2020). Daily tourism volume forecasting for tourist attractions. *Annals of tourism research*, *83*, 102923.

Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. *Tourism Management*, *71*, 490-503.

Biraglia, A., & Gerrath, M. H. (2020). Corporate sponsorship for museums in times of crisis. *Annals of tourism research*.

Biraglia, A., Gerrath, M. H., & Usrey, B. (2018). Examining how companies' support of tourist attractions affects visiting intentions: The mediating role of perceived authenticity. *Journal of Travel Research*, *57*(6), 811-823.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

BliegeBird, R., & Smith, E. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current anthropology*, *46*(2), 221-248.

Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies*, *48*(8), 1866-1891.

Borah, A., Banerjee, S., Lin, Y.-T., Jain, A., & Eisingerich, A. B. (2020). Improvised marketing interventions in social media. *Journal of Marketing*, *84*(2), 69-91.

Bordogna, G., Frigerio, L., Cuzzocrea, A., & Psaila, G. (2016). Clustering geo-tagged tweets for advanced big data analytics. 2016 IEEE international congress on big data (BigData Congress),

Boto-García, D., & Baños-Pino, J. F. (2023). Deep habits and travel resilience after COVID-19. *Tourism Economics*, *29*(2), 543-550.

BP. (2022). *bp and the National Portrait Gallery announce end of partnership* https://www.bp.com/en_gb/united-kingdom/home/news/press-releases/bp-and-the-national-portrait-gallery-announce-end-of-partnership.html

Branzei, O., Ursacki-Bryant, T. J., Vertinsky, I., & Zhang, W. (2004). The formation of green strategies in Chinese firms: Matching corporate environmental responses and individual principles. *Strategic management journal*, *25*(11), 1075-1095.
Brida, J. G., Gomez, D. M., & Segarra, V. (2020). On the empirical relationship between tourism and economic growth. *Tourism Management*, *81*, 104131.
Brodie, R. J., Fehrer, J. A., Jaakkola, E., & Conduit, J. (2019). Actor engagement in networks: Defining the conceptual domain. *Journal of service research*, *22*(2), 173-188.
Brown, J. M. (2016). BP ends 27-year sponsorship of Tate as falling oil price takes toll. *Financial Times*. https://www.ft.com/content/fc89bc54-e78f-11e5-bc31-138df2ae9ee6
Bruns, A., Highfield, T., & Burgess, J. (2013). The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. *American behavioral scientist*, *57*(7), 871-898.
Buckley, R., & Chauvenet, A. L. (2022). Hidden economic counterflows in post-COVID international wildlife tourism. *Tourism Management*, *93*, 104624.
Buzeta, C., De Pelsmacker, P., & Dens, N. (2020). Motivations to use different social media types and their impact on consumers' online brand-related activities (COBRAs). *Journal of Interactive marketing*, *52*(1), 79-98.
Calver, S. J., & Page, S. J. (2013). Enlightened hedonism: Exploring the relationship of service value, visitor knowledge and interest, to visitor enjoyment at heritage attractions. *Tourism Management*, *39*, 23-36.
Canestrelli, E., & Costa, P. (1991). Tourist carrying capacity: A fuzzy approach. *Annals of tourism research*, *18*(2), 295-311.
Cano-Marin, E., Mora-Cantallops, M., & Sánchez-Alonso, S. (2023). Twitter as a predictive system: A systematic literature review. *Journal of Business Research*, *157*, 113561.
Capacci, S., Scorcu, A. E., & Vici, L. (2015). Seaside tourism and eco-labels: The economic impact of Blue Flags. *Tourism Management*, *47*, 88-96.
Capocchi, A., Vallone, C., Amaduzzi, A., & Pierotti, M. (2019). Is 'overtourism'a new issue in tourism development or just a new term for an already known phenomenon? *Current Issues in Tourism*, 1-5.
Capocchi, A., Vallone, C., Pierotti, M., & Amaduzzi, A. (2019). Overtourism: A literature review to assess implications and future perspectives. *Sustainability*, *11*(12), 3303.
Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.
Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing: strategy, implementation & practice*. Pearson uk.
Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3), 1247-1250.
Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, *22*.
Chang, J.-H., Tsai, C.-E., & Chiang, J.-H. (2018). Using heterogeneous social media as auxiliary information to improve hotel recommendation performance. *IEEE Access*, *6*, 42647-42660.
Chen, S.-x., Wang, X.-k., Zhang, H.-y., Wang, J.-q., & Peng, J.-j. (2021). Customer purchase forecasting for online tourism: A data-driven method with multiplex behavior data. *Tourism Management*, *87*, 104357.

Cheng, L., & Liu, L. (2022). Exploring posttraumatic growth after the COVID-19 pandemic. *Tourism Management*, *90*, 104474.

Cheung, K. S., & Li, L.-H. (2019). Understanding visitor–resident relations in overtourism: developing resilience for sustainable tourism. *Journal of Sustainable Tourism*, 1-20.

Chu, F.-L. (1998). Forecasting tourism: a combined approach. *Tourism Management*, *19*(6), 515-520.

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, *57*, 295-310.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, *26*(4), 546-554.

Colicev, A. K., Ashish ; O'Connor, Peter. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, *36*(1), 100-116.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of management*, *37*(1), 39-67.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative* (Vol. 7). Prentice Hall Upper Saddle River, NJ.

Crotty, M. J. (1998). The foundations of social research: Meaning and perspective in the research process. *The foundations of social research*, 1-256.

De Oliveira Santini, F., Ladeira, W. J., Pinto, D. C., Herter, M. M., Sampaio, C. H., & Babin, B. J. (2020). Customer engagement in social media: a framework and meta-analysis. *Journal of the Academy of Marketing Science*, *48*, 1211-1228.

Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, *21*(4), 23-45.

Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, *66*, 108-120.

Dewally, M., & Ederington, L. (2006). Reputation, certification, warranties, and information as remedies for seller‐buyer information asymmetries: Lessons from the online comic book market. *The Journal of Business*, *79*(2), 693-729.

Diks, C., & Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, *30*(9-10), 1647-1669.

Dinhopl, A., & Gretzel, U. (2016). Selfie-taking as touristic looking. *Annals of tourism research*, *57*, 126-139.

Dogru, T., & Bulut, U. (2018). Is tourism an engine for economic recovery? Theory and empirical evidence. *Tourism Management*, *67*, 425-434.

DONNA, L., & Novak, H. T. P. (1997). A new marketing paradigm for electronic commerce. *The information society*, *13*(1), 43-54.

Drover, W., Wood, M. S., & Corbett, A. C. (2018). Toward a cognitive view of signalling theory: Individual attention and signal set interpretation. *Journal of Management Studies*, *55*(2), 209-231.

Duro, J., Perez-Laborda, A., & Fernandez, M. (2022). Territorial tourism resilience in the COVID-19 summer. *Annals of Tourism Research Empirical Insights*, *3*(1), 100039.

Edgell, D. L., Seely, R. L., & Iglarsh, H. J. (1980). Forecasts of international tourism to the USA. *International Journal of Tourism Management*, *1*(2), 109-113.

Edwards, D., Cheng, M., Wong, I. A., Zhang, J., & Wu, Q. (2017). Ambassadors of knowledge sharing: Co-produced travel information through tourist-local social media

exchange. *International Journal of Contemporary Hospitality Management*, *29*(2), 690-708.
Enders, W. (2008). *Applied econometric time series*. John Wiley & Sons.
Eugenio-Martin, J. L., Martín-Morales, N., & Sinclair, M. T. (2008). The role of economic development in tourism demand. *Tourism Economics*, *14*(4), 673-690.
Fleetwood, S. (2005). Ontology in organization and management studies: A critical realist perspective. *Organization*, *12*(2), 197-222.
Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, *2002*(December), 5.
Frechtling, D. (2001). *Forecasting tourism demand*. Oxford, UK: Butterworth Heinemann
Gao, G., Gopal, A., & Agarwal, R. (2010). Contingent effects of quality signaling: evidence from the Indian offshore IT services industry. *Management Science*, *56*(6), 1012-1029.
Gao, Y., & Su, W. (2021). The disclosure of quality on tourism performance: evidence from top tourist cities in China. *Journal of Travel Research*, *60*(7), 1492-1509.
Gao, Y., Su, W., & Zang, L. (2022). Does regional tourism benefit from the official quality rating of tourist attractions? Evidence from China's top-grade tourist attraction accreditations. *Journal of China Tourism Research*, *18*(2), 268-293.
Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using forum and search data for sales prediction of high-involvement projects. *Mis Quarterly, 41*(1), 65-82.
Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1-2), 59-95.
Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., & Verlegh, P. (2005). The firm's management of social interactions. *Marketing letters*, *16*, 415-428.
Goodwin, H. (2017). The challenge of overtourism. *Responsible Tourism Partnership*.
Götz, T. B., Hecq, A., & Smeekes, S. (2016). Testing for Granger causality in large mixed-frequency VARs. *Journal of Econometrics*, *193*(2), 418-432.
Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.
Gretzel, U. (2017). Social media activism in tourism. *Journal of Hospitality and Tourism, 15*(2), 1-14.
Gretzel, U. (2019). The role of social media in creating and addressing overtourism. In *Overtourism: Issues, realities and solutions* (pp. 62-75). De Gruyter.
Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in US Election 2016 outcomes–Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*, *145*, 438-460.
Guitart, I. A., & Hervet, G. (2017). The impact of contextual television ads on online conversions: An application in the insurance industry. *International Journal of Research in Marketing*, *34*(2), 480-498.
Gulati, R., & Higgins, M. C. (2003). Which ties matter when? The contingent effects of interorganizational partnerships on IPO success. *Strategic management journal*, *24*(2), 127-144.
Gunter, U., & Onder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of tourism research*, *61*, 199-212. https://doi.org/10.1016/j.annals.2016.10.007
Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of tourism research*, *61*, 199-212.

Gunter, U., Önder, I., & Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. *Tourism Economics*, *25*(3), 375-401.
Gunter, U., Önder, I., & Smeral, E. (2019). Scientific value of econometric tourism demand studies. *Annals of tourism research*, *78*, 102738.
Gutierrez-Huerter O, G., Gold, S., & Trautrims, A. (2023). Change in rhetoric but not in action? Framing of the ethical issue of modern slavery in a UK sector at high risk of labor exploitation. *Journal of business ethics*, *182*(1), 35-58.
Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, *117*, 62-81.
Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, *13*(2), 586-632.
Harris, E. E., Neely, D. G., & Saxton, G. D. (2021). Social media, signaling, and donations: Testing the financial returns on nonprofits' social media investment. *Review of Accounting Studies*, 1-31.
Hartman, R. S. (2011). *The structure of value: Foundations of scientific axiology*. Wipf and Stock Publishers.
Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics*, *27*(1), 129-148.
Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the Academy of Marketing Science*, *43*, 375-394.
Herhausen, D., Grewal, L., Cummings, K. H., Roggeveen, A. L., Villarroel Ordenes, F., & Grewal, D. (2023). Complaint De-Escalation Strategies on Social Media. *Journal of Marketing*, *87*(2), 210-231.
Hewett, K., Rand, W., Rust, R. T., & Van Heerde, H. J. (2016). Brand buzz in the echoverse. *Journal of Marketing*, *80*(3), 1-24.
Hollingshead, W., Quan-Haase, A., & Blank, G. (2022). Representativeness and Bias in Social Media Research: Quantitative and Qualitative Approaches to Sampling. *The SAGE Handbook of Social Media Research Methods*, 79.
Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2021). Improving tourist arrival prediction: a big data and artificial neural network approach. *Journal of Travel Research*, *60*(5), 998-1017.
Houghton, J. P., Siegel, M., Madnick, S., Tounaka, N., Nakamura, K., Sugiyama, T., Nakagawa, D., & Shirnen, B. (2019). Beyond Keywords: Tracking the evolution of conversational clusters in social media. *Sociological Methods & Research*, *48*(3), 588-607.
Hu, M., Li, H., Song, H., Li, X., & Law, R. (2022). Tourism demand forecasting using tourist-generated online review data. *Tourism Management*, *90*, 104490.
Hu, M., & Song, H. (2019). Data source combination for tourism demand forecasting. *Tourism Economics*, 1354816619872592.
Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu Index: Uses in predicting tourism flows–A case study of the Forbidden City. *Tourism Management*, *58*, 301-306.
Huberty, M. (2015). Can we vote with our tweet? On the perennial difficulty of election forecasting with social media. *International Journal of Forecasting*, *31*(3), 992-1007.
Huff, A. S. (2008). *Designing research for publication*. Sage.
Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the international AAAI conference on web and social media,

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688.

Ibrahim, N. F., & Wang, X. (2019a). Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. *Computers in Human Behavior*, *96*, 32-45.

Ibrahim, N. F., & Wang, X. (2019b). A text analytics approach for online retailing service improvement: Evidence from Twitter. *Decision Support Systems*, *121*, 37-50.

Inoue, A., & Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, *23*(4), 371-402.

Ito, T., & Sato, K. (2008). Exchange rate changes and inflation in post-crisis Asian Economies: Vector Autoregression Analysis of the exchange rate pass-through. *Journal of Money, Credit and Banking*, *40*(7), 1407-1438.

Jacobsen, J. K. S., Iversen, N. M., & Hem, L. E. (2019). Hotspot crowding and over-tourism: Antecedents of destination attractiveness. *Annals of tourism research*, *76*, 53-66.

Jacobsen, J. K. S., & Munar, A. M. (2012). Tourist information search and destination choice in a digital age. *Tourism Management Perspectives*, *1*, 39-47.

Jagarlamudi, J., Daumé III, H., & Udupa, R. (2012). Incorporating lexical priors into topic models. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics,

Janney, J. J., & Gove, S. (2017). Firm linkages to scandals via directors and professional service firms: Insights from the backdating scandal. *Journal of business ethics*, *140*, 65-79.

Jensen, Ø., Li, Y., & Uysal, M. (2017). Visitors' satisfaction at managed tourist attractions in Northern Norway: Do on-site factors matter? *Tourism Management*, *63*, 277-286.

Johnson, Y. R. R. (2018). *New developer requirements to protect our platform*. https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), 59-68.

Kaplan, A. M., & Haenlein, M. (2011). Two hearts in three-quarter time: How to waltz the social media/viral marketing dance. *Business horizons*, *54*(3), 253-263.

Kaynak, E., & Macaulay, J. A. (1984). The Delphi technique in the measurement of tourism market potential: the case of Nova Scotia. *Tourism Management*, *5*(2), 87-101.

Kelemen, M. L., & Rumens, N. (2008). *An introduction to critical management research*. Sage.

Kennedy, C. (2017). Boycott products from states with dirty energy. In: Nature Publishing Group.

Khaleeli, H. (2015). 'We're going to need a bigger door': how museums manage blockbuster crowds. *The Guardian*. https://www.theguardian.com/culture/shortcuts/2015/nov/03/british-museum-overcrowding-entrance-sistine-chapel

Kim, Y. R., Liu, A., Stienmetz, J., & Chen, Y. (2022). Visitor flow spillover effects on attraction demand: A spatial econometric model with multisource data. *Tourism Management*, *88*, 104432.

Kirilenko, A. P., Ma, S., Stepchenkova, S. O., Su, L., & Waddell, T. F. (2023). Detecting early signs of overtourism: bringing together indicators of tourism development with data fusion. *Journal of Travel Research*, *62*(2), 382-398.

Kirmani, A., & Rao, A. R. (2000). No pain, no gain: A critical review of the literature on signaling unobservable product quality. *Journal of Marketing*, *64*(2), 66-79.

Klein, J. G., Smith, N. C., & John, A. (2004). Why we boycott: Consumer motivations for boycott participation. *Journal of Marketing*, *68*(3), 92-109.

Koens, K., Postma, A., & Papp, B. (2018). Is overtourism overused? Understanding the impact of tourism in a city context. *Sustainability*, *10*(12), 4384.

Koens, K., Turisme, O. M. d., Organization, W. T., Postma, A., Centre of Expertise Leisure, T., Hospitality, Papp, B., Sciences, B. U. o. A., Yeoman, I., Sciences, S. U. o. A., & CELTH. (2018). *'Overtourism'? - Understanding and Managing Urban Tourism Growth Beyond Perceptions*. World Tourism Organization (UNWTO). https://books.google.co.uk/books?id=viRmwAEACAAJ

Koku, P. S. (2022). Are Ethical Boycotts Merely Signaling Value? The Financial Effect of Ethical Boycotts. Academy of Marketing Science Annual Conference-World Marketing Congress,

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188.

Kumar, N., Qiu, L., & Kumar, S. (2022). A hashtag is worth a thousand words: An empirical investigation of social media strategies in trademarking hashtags. *Information Systems Research*, *33*(4), 1403-1427.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, *27*(2), 529-542.

Lacka, E., Boyd, D. E., Ibikunle, G., & Kannan, P. (2022). Measuring the Real-Time Stock Market Impact of Firm-Generated Content. *Journal of Marketing*, *86*(5), 58-78.

Lasarov, W., Hoffmann, S., & Orth, U. (2021). Vanishing Boycott Impetus: Why and How Consumer Participation in a Boycott Decreases Over Time. *Journal of business ethics*, *182*, 1129–1154 (2023).

Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of tourism research*, *75*, 410-423.

Ledford, H. (2020). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, *582*(7812), 328-331.

Lee, C.-C., Chen, M.-P., & Peng, Y.-T. (2021). Tourism development and happiness: International evidence. *Tourism Economics*, *27*(5), 1101-1136.

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of travel tourism marketing*, *30*(1-2), 3-22.

Leung, X. Y., Bai, B., & Stahura, K. A. (2015). The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter. *Journal of Hospitality & Tourism Research*, *39*(2), 147-169.

Lew, A. A. (1987). A framework of tourist attraction research. *Annals of tourism research*, *14*(4), 553-575.

Li, C., Ge, P., Liu, Z., & Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. *Annals of tourism research*, *83*, 102943.

Li, G., & Jiao, X. (2020). Tourism forecasting research: a perspective article. *Tourism Review*.

Li, H., Hu, M., & Li, G. (2020). Forecasting tourism demand with multisource big data. *Annals of tourism research*, *83*, 102912.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301-323.

Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, *68*, 116-126.

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, *83*, 104245.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, *59*, 57-66.

Liadeli, G., Sotgiu, F., & Verlegh, P. W. (2023). A meta-analysis of the effects of brands' owned social media on social media engagement and sales. *Journal of Marketing*, *87*(3), 406-427.

Liaukonytė, J., Tuchman, A., & Zhu, X. (2022). Frontiers: Spilling the Beans on Political Consumerism: Do Social Media Boycotts and Buycotts Translate to Real Sales Impact? *Marketing Science*.

Lin, Y., Yao, D., & Chen, X. (2021). Happiness begets money: Emotion and engagement in live streaming. *Journal of marketing research*, *58*(3), 417-438.

Liu, P., Zhang, H., Zhang, J., Sun, Y., & Qiu, M. (2019). Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. *Tourism Management*, *73*, 105-114.

Lo, I. S., McKercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism Management*, *32*(4), 725-731.

Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public relations review*, *38*(2), 313-318.

Lu, S., Dinner, I. M., & Grewal, R. (2022). The Ripple Effect of Firm-Generated Content on New Movie Releases. *Journal of Marketing Research, forthcoming*.

Lu, Y., & Zheng, Q. (2021). Twitter public sentiment dynamics on cruise tourism during the COVID-19 pandemic. *Current Issues in Tourism*, *24*(7), 892-898.

Luo, Q., & Zhong, D. (2015). Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management*, *46*, 274-282.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Macdonald, S. (2010). *Difficult heritage: Negotiating the Nazi past in Nuremberg and beyond*. Routledge.

Maddison, D., & Foster, T. (2003). Valuing congestion costs in the British Museum. *Oxford Economic Papers*, *55*(1), 173-190.

Magasic, M., & Gretzel, U. (2020). Travel connectivity. *Tourist Studies*, 1468797619899343.

Manning, R., Wang, B., Valliere, W., Lawson, S., & Newman, P. (2002). Research to estimate and manage carrying capacity of a tourist attraction: a study of Alcatraz Island. *Journal of Sustainable Tourism*, *10*(5), 388-404.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing*, *34*(2), 336-354.

Martins, L. F., Gan, Y., & Ferreira-Lopes, A. (2017). An empirical analysis of the influence of macroeconomic determinants on World tourism demand. *Tourism Management*, *61*, 248-260.

Michaelson, R. (2022). British Museum urged to drop BP sponsor deal for Egypt exhibition. *The Guardian*. https://www.theguardian.com/culture/2022/oct/11/british-museum-urged-to-drop-bp-sponsor-deal-for-egypt-exhibition

Monbiot, G. (2021). Why is the Science Museum still being contaminated by Shell's dirty money? *The Guardian*.

https://www.theguardian.com/commentisfree/2021/apr/21/science-museum-shell-money-exhibition-climate
Motion, J. (2019). Undoing art and oil: An environmental tale of sponsorship, cultural justice and climate change controversy. *Environmental Politics*, *28*(4), 727-746.
Mumi, A., Obal, M., & Yang, Y. (2019). Investigating social media as a firm's signaling strategy through an IPO. *Small Business Economics*, *53*, 631-645.
Munar, A. M., & Jacobsen, J. K. S. (2014). Motivations for sharing tourism experiences through social media. *Tourism Management*, *43*, 46-54.
Narangajavana, Y., Fiol, L. J. C., Tena, M. Á. M., Artola, R. M. R., & García, J. S. (2017). The influence of social media in creating expectations. An empirical study for a tourist destination. *Annals of tourism research*, *65*, 60-70.
Narayan, P. K. (2004). Fiji's tourism demand: the ARDL approach to cointegration. *Tourism Economics*, *10*(2), 193-206.
Nevin, A. D., Chen, Y., Yang, S., & Quan-Haase, A. (2022). Key considerations in the interpretation of digital trace data. *The SAGE Handbook of Social Media Research Methods*, 54-66.
Nian, T., & Sundararajan, A. (2022). Social media marketing, quality signaling, and the Goldilocks principle. *Information Systems Research*, *33*(2), 540-556.
Nicolau, J. L., Sharma, A., & Shin, S. (2020). The tourism effect of President Trump's participation on Twitter. *Tourism Management*, *81*, 104133.
Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of google trends in health care research: a systematic review. *PloS one*, *9*(10), e109583.
O'Neal, A., Rodgers, B., Segler, J., Murthy, D., Lakuduva, N., Johnson, M., & Stephens, K. (2018). Training an emergency-response image classifier on signal data. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA),
Önder, I., Gunter, U., & Gindl, S. (2019). Utilizing Facebook Statistics in Tourism Demand Modeling and Destination Marketing. *Journal of Travel Research*(59(2)), 195-208.
Ørsted. (2023). *About our name: From magnetism to offshore wind turbines*. Retrieved April from https://orsted.co.uk/about-us/our-company/our-name-change
Östlund, U., Kidd, L., Wengström, Y., & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: a methodological review. *International journal of nursing studies*, *48*(3), 369-383.
Page, S., Song, H., & Wu, D. C. (2012). Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the United Kingdom. *Journal of Travel Research*, *51*(2), 142-153.
Page, S. J., & Duignan, M. (2023). Progress in Tourism Management: Is urban tourism a paradoxical research domain? Progress since 2011 and prospects for the future. *Tourism Management*, *98*, 104737.
Pak, T.-Y. (2020). Old-age income security and tourism demand: A quasi-experimental study. *Journal of Travel Research*, *59*(7), 1298-1315.
Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, *56*(7), 957-970.
Park, H. D., & Patel, P. C. (2015). How does ambiguity influence IPO underpricing? The role of the signalling environment. *Journal of Management Studies*, *52*(6), 796-818.
Parry, P. (2009). *Freedom of Info, The Natural History Museum* https://www.whatdotheyknow.com/request/funding_from_shell
Parvaiz, G. S., Mufti, O., & Wahab, M. (2016). Pragmatism for mixed method research at higher education level. *Business & Economic Review*, *8*(2), 67-79.

Peeters, P., Gössling, S., Klijs, J., Milano, C., Novelli, M., Dijkmans, C., Eijgelaar, E., Hartman, S., Heslinga, J., & Isaac, R. (2018). *Research for TRAN Committee-Overtourism: impact and possible policy responses*. European Parliament, Directorate General for Internal Policies, Policy Department B: Structural and Cohesion Policies, Transport and Tourism.

Perdue, R. R. (2002). Perishability, yield management, and cross-product elasticity: A case study of deep discount season passes in the Colorado ski industry. *Journal of Travel Research*, *41*(1), 15-22.

Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K. (2013). Social media metrics—A framework and guidelines for managing social media. *Journal of Interactive marketing*, *27*(4), 281-298.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., Bergmeir, C., Bessa, R. J., Bijak, J., & Boylan, J. E. (2022). Forecasting: theory and practice. *International Journal of Forecasting*.

Phi, G. T. (2019). Framing overtourism: a critical news media analysis. *Current Issues in Tourism*, 1-5.

Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, *55*, 16-24.

Poria, Y., Reichel, A., & Cohen, R. (2013). Tourists perceptions of World Heritage Site and its designation. *Tourism Management*, *35*, 272-274.

Qi, M., & Zhang, G. P. (2001). An investigation of model selection criteria for neural network time series forecasting. *European journal of operational research*, *132*(3), 666-680.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. Proceedings of the 2009 conference on empirical methods in natural language processing,

Rawlinson, K. (2019). Congestion on Everest leads to backlog of climbers in 'death zone'. *Guardian*. https://www.theguardian.com/world/2019/may/25/british-climber-latest-to-die-on-everest-amid-overcrowding-robin-fisher

Razzall, K. (2022a). Benin Bronzes: Nigeria hails 'great day' as London museum signs over looted objects. *BBC*. https://www.bbc.co.uk/news/entertainment-arts-63783561

Razzall, K. (2022b). Elgin Marbles: New body aims to return sculptures to Greece. *BBC*. https://www.bbc.co.uk/news/entertainment-arts-63231885

Reed, M. I. (2009). Critical realism: Philosophy, method, or philosophy in search of a method. *The SAGE handbook of organizational research methods*, 430-448.

Ridhwan, K. M., & Hargreaves, C. A. (2021). Leveraging Twitter data to understand public sentiment for the COVID‐19 outbreak in Singapore. *International Journal of Information Management Data Insights*, *1*(2), 100021.

Rissanen, J. (1980). Consistent order estimates of autoregressive processes by shortest description of data. *Analysis and Optimisation of Stochastic Systems. Academic Press, New York*.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988-1003.

Ross, S. A. (1977). The determination of financial structure: the incentive-signalling approach. *The bell journal of economics*, 23-40.

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, *66*, 13-25.

Santana-Jiménez, Y., & Hernández, J. M. (2011). Estimating the effect of overcrowding on tourist attraction: The case of Canary Islands. *Tourism Management*, *32*(2), 415-425.
Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students.
Saxton, G. D., Gómez, L., Ngoh, Z., Lin, Y.-P., & Dietrich, S. (2019). Do CSR messages resonate? Examining public reactions to firms' CSR efforts on social media. *Journal of business ethics*, *155*, 359-377.
Saxton, G. D., & Waters, R. D. (2014). What do stakeholders like on Facebook? Examining public reactions to nonprofit organizations' informational, promotional, and community-building messages. *Journal of public relations research*, *26*(3), 280-299.
Seddon, S. (2023). UK tourism: Visitor attractions still suffering from pandemic effects. *BBC*. https://www.bbc.co.uk/news/uk-64976013
Sedera, D., Lokuge, S., Atapattu, M., & Gretzel, U. (2017). Likes—the key to my happiness: The moderating effect of social influence on travel experience. *Information and Management*, *54*(6), 825-836.
Shankar, V., & Parsana, S. (2022). An overview and empirical comparison of natural language processing (NLP) models and an introduction to and empirical application of autoencoder models in marketing. *Journal of the Academy of Marketing Science*, *50*(6), 1324-1350.
Shen, D., Urquhart, A., & Wang, P. (2019). Does twitter predict Bitcoin? *Economics Letters*, *174*, 118-122.
Sheng, J., Amankwah‐Amoah, J., Wang, X., & Khan, Z. (2019). Managerial responses to online reviews: A text analytics approach. *British Journal of Management*, *30*(2), 315-327.
Sheng, J., Wang, X., & Amankwah-Amoah, J. (2021). The value of firm engagement: How do ratings benefit from managerial responses? *Decision Support Systems*, *147*, 113578.
Siddique, H. (2018). 'Not everything was looted': British Museum to fight critics. *The Guardian*. https://www.theguardian.com/culture/2018/oct/12/collected-histories-not-everything-was-looted-british-museum-defends-collections
Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. Proceedings of the workshop on interactive language learning, visualization, and interfaces,
Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS one*, *6*(5), e19467.
Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1-48.
Smith, B. (2012). Ontology. In *The furniture of the world* (pp. 47-68). Brill.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, *29*(2), 203-220.
Song, H., Li, G., Witt, S. F., & Fei, B. (2010). Tourism demand modelling and forecasting: how should demand be measured? *Tourism Economics*, *16*(1), 63-81.
Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of tourism research*, *75*, 338-362.
Song, H., Qiu, R. T., & Park, J. (2023). Progress in tourism demand research: Theory and empirics. *Tourism Management*, *94*, 104655.
Song, H., Wong, K. K., & Chon, K. K. (2003). Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management*, *22*(4), 435-451.

Song, H., & Wu, D. C. (2022). A critique of tourism-led economic growth studies. *Journal of Travel Research*, *61*(4), 719-729.
Sotiriadis, M. D., & Van Zyl, C. (2013). Electronic word-of-mouth and online reviews in tourism services: the use of twitter by tourists. *Electronic Commerce Research*, *13*, 103-124.
Spence, A. M. (1973). Time and communication in economic and social interaction. *The Quarterly Journal of Economics*, *87*(4), 651-660.
Statista. (2023). *Number of monthly active Facebook users worldwide as of 1st quarter 2023*. Retrieved May from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/
Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of marketing research*, *49*(5), 624-639.
Storbacka, K., Brodie, R. J., Böhmann, T., Maglio, P. P., & Nenonen, S. (2016). Actor engagement as a microfoundation for value co-creation. *Journal of Business Research*, *69*(8), 3008-3017.
Su, L., Jia, B., & Huang, Y. (2022). How do destination negative events trigger tourists' perceived betrayal and boycott? The moderating role of relationship quality. *Tourism Management*, *92*, 104536.
Suddaby, R. (2006). From the editors: What grounded theory is not. In (Vol. 49, pp. 633-642): Academy of Management Briarcliff Manor, NY 10510.
Sugathan, P., & Ranjan, K. R. (2019). Co-creating the tourism experience. *Journal of Business Research*, *100*, 207-217.
Sun, Q., Wu, F., Li, S., & Grewal, R. (2021). Consumer boycotts, country of origin, and product competition: Evidence from China's automobile market. *Management Science*, *67*(9), 5857-5877.
Sun, S., Wei, Y., Tsui, K.-L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, *70*, 1-10.
Taj, S. A. (2016). Application of signaling theory in management research: Addressing major gaps in theory. *European Management Journal*, *34*(4), 338-348.
Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. In (Vol. 1, pp. 3-7): Sage Publications.
ThBritishMuseum. (2022). *The British Museum story*. https://www.britishmuseum.org/about-us/british-museum-story
TheBritishMuseum. (2018). *British Museum policy: De-accession of objects from the collection* https://www.britishmuseum.org/sites/default/files/2019-10/De-accession_Policy_Nov2018.pdf
Toubia, O., Iyengar, G., Bunnell, R., & Lemaire, A. (2019). Extracting features of entertainment products: A guided latent dirichlet allocation approach informed by the psychology of media consumption. *Journal of marketing research*, *56*(1), 18-36.
Toubia, O., & Stephen, A. T. (2013). Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science*, *32*(3), 368-392.
Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics. *University of Aberdeen*, *1*, 16.
Tumasjan, A., Braun, R., & Stolz, B. (2021). Twitter sentiment as a weak signal in venture capital financing. *Journal of Business Venturing*, *36*(2), 106062.
Twitter, I. (2021). *Fiscal Year 2020 Annual Report*. https://investor.twitterinc.com/financial-information/annual-reports/default.aspx

UNWTO. (2023a). *IMPACT ASSESSMENT OF THE COVID-19 OUTBREAK ON INTERNATIONAL TOURISM*. https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism

UNWTO. (2023b). *TOURISM ON TRACK FOR FULL RECOVERY AS NEW DATA SHOWS STRONG START TO 2023*. Retrieved 10 July from https://www.unwto.org/news/tourism-on-track-for-full-recovery-as-new-data-shows-strong-start-to-2023

Uşaklı, A., Koç, B., & Sönmez, S. (2017). How'social'are destinations? Examining European DMO social media usage. *Journal of destination marketing & management*, *6*(2), 136-149.

Valsesia, F., & Diehl, K. (2022). Let me show you what I did versus what I have: Sharing experiential versus material purchases alters authenticity and liking of social media users. *Journal of Consumer Research*, *49*(3), 430-449.

Valsesia, F., Proserpio, D., & Nunes, J. C. (2020). The positive effect of not following others on social media. *Journal of marketing research*, *57*(6), 1152-1168.

Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. (2007). The interplay between theory and method. *Academy of management review*, *32*(4), 1145-1154.

Vanhove, N. (1980). Forecasting in tourism. *Tourist Review*, *35*(3), 2-7.

Varma, T. M. (2021). Responsible leadership and reputation management during a crisis: The cases of Delta and United Airlines. *Journal of business ethics*, *173*, 29-45.

Vaughan, A. (2018). Shell ends National Gallery sponsorship – to delight of campaigners. *The Guardian*. https://www.theguardian.com/business/2018/oct/19/shells-ends-national-gallery-sponsorship-to-delight-of-campaigners

Villanova, D., & Matherly, T. (2023). EXPRESS: For Shame! Socially Unacceptable Brand Mentions on Social Media Motivate Consumer Disengagement. *Journal of Marketing*, 00222429231179942.

VisitLondon. (2022). *Best 10 London attractions 2022*. Retrieved 20 September from https://www.visitlondon.com/things-to-do/sightseeing/london-attraction/top-ten-attractions

Volchek, K., Liu, A., Song, H., & Buhalis, D. (2019). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics*, *25*(3), 425-447.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of econometrics*, *4*, 2843-2915.

Wen, L., Liu, C., Song, H., & Liu, H. (2021). Forecasting tourism demand with an improved mixed data sampling model. *Journal of Travel Research*, *60*(2), 336-353.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*(1), 79-82.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, *11*(3), 447-475.

Wong, I. A., Law, R., & Zhao, X. (2017). When and where to travel? A longitudinal multilevel investigation on destination choice and demand. *Journal of Travel Research*, *56*(7), 868-880.

Wu, D. C., Zhong, S., Qiu, R. T., & Wu, J. (2022). Are customer reviews just reviews? Hotel forecasting using sentiment analysis. *Tourism Economics*, *28*(3), 795-816.

Wu, J., Li, M., Zhao, E., Sun, S., & Wang, S. (2023). Can multi-source heterogeneous data improve the forecasting performance of tourist arrivals amid COVID-19? Mixed-data sampling approach. *Tourism Management*, *98*, 104759.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179-188.
Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing Consumer Services*, *22*, 244-249.
Xie, P., Chen, H., & Hu, Y. J. (2020). Signal or noise in social media discussions: the role of network cohesion in predicting the Bitcoin market. *Journal of Management Information Systems*, *37*(4), 933-956.
Yadav, M. S., & Pavlou, P. A. (2014). Marketing in computer-mediated environments: Research synthesis and new directions. *Journal of Marketing*.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, *46*, 386-397.
Yang, Y., Fik, T. J., & Zhang, H.-l. (2017). Designing a tourism spillover index based on multidestination travel: A two-stage distance-based modeling approach. *Journal of Travel Research*, *56*(3), 317-333.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization’s web traffic data. *Journal of Travel Research*, *53*(4), 433-447.
Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining,
Yu, Q., McManus, R., Yen, D. A., & Li, X. R. (2020). Tourism boycotts and animosity: A study of seven events. *Annals of tourism research*, *80*, 102792.
Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press.
Zhang, H., Song, H., Wen, L., & Liu, C. (2021). Forecasting tourism recovery amid COVID-19. *Annals of tourism research*, *87*, 103149.
Ziegler, J. A., Dearden, P., & Rollins, R. (2016). Participant crowding and physical contact rates of whale shark tours on Isla Holbox, Mexico. *Journal of Sustainable Tourism*, *24*(4), 616-636.
Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.

Appendices

Appendix 1

Table A1 Top 50 tweets with most likes count since 2017

No	Date	Tweet text	Likes count	Boycott tweets
1	26/08/2018	Overheard at British Museum -Young boy to Dad - ‘when you die, can I use your skull to strike fear into the hearts of my enemies?’Dad - ‘...no.’	78317	
2	12/03/2018	Happy Commonwealth Day to 2.435 billion people bound together by a dream of getting their stuff back from the British Museum.	73604	
3	04/02/2019	Should British museums have more British artifacts?	54230	
4	17/08/2018	My mom called The British Museum "The Evidence Room" because its just "things they've stolen from colonies over the years"	34184	√
5	28/02/2019	My brother was asked to make a donation at the British Museum today and he politely told the lady,"my donation is the Sudan section"	32547	√
6	05/12/2018	Held at the British Museum in 2019, "Golden Kamuy" will participate in the manga-themed exhibition! Official account of the museum @britishmuseum Ashi (Ri) Pa-san is in the header of ...! Please check it.	21251	
7	07/06/2019	Salah for the next James Bond. The entire film is him returning British Museum artefacts to their countries of origin, and then he goes home to his wife and children.	15914	√
8	29/01/2019	Someone send Marie Kondo to the British Museum, so some of us can get our shit back.	13783	√
9	11/04/2019	Fully endorse the demand that British empire must apologize to the nations of Pakistan, India and Bangladesh on Jallianwala Massacre and Bengal famine .. these tragedies are the scar on the face of Britain, also KohENoor must be returned to Lahore museum where it belongs	13525	√
10	04/01/2019	hey @britishmuseum give us your best duck	13311	

11	17/05/2018	Mohamed Salah's boots have been added to the Egyptian collection at the British Museum	11456	
12	13/10/2018	Excited to see the new British Museum slogan, We Didnt Steal *All* Of It.	10584	√
13	28/03/2019	RM went to the solo exhibition of British artist, “David Hockney," at the Seoul Museum of Art (SeMA) features 133 artworks (3/22-8/4) @BTS_twt *Personally I am a big fan of David Hockney, I’ve seen his works at Tate London before. I’m so happy Namjoon enjoys his art.	10274	
14	23/11/2019	SO NOBODYS GONNA FUCKIN TELL ME THAT TOUHOU IS REPRESENTED IN A DISPLAY CASE AT THE BRITISH MUSEUM???????????????????? https://t.co/X9lxd53Mki	9798	
15	25/08/2019	2 very different heroes: Ben Stokes switch-hitting 6s while Jack Leach cleans his glasses and digs in 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1 Get those glasses in the British museum #Ashes2019 #TheAshes #ENGAUS https://t.co/kJEi5yGJ8X	9492	
16	03/09/2018	Reports from Rio are of the total destruction of the museum and the loss of upwards of 20 million items. This is the Brazilian equivalent of the British Museum or the Louvre. There are simply no words for a tragedy like this. https://t.co/dXNp4Kv379	9453	
17	15/06/2019	Pssst! Guess who's got a cameo role in the Manga Exhibition at the British Museum? Fun fact: this is the largest exhibition of manga ever to take place outside of Japan. https://t.co/IcOMEt44aL	6719	
18	23/08/2019	Remarkable Harry Kane exhibition at the British Museum https://t.co/uAjZJS4RFz	6273	
19	03/07/2019	Thank you for inviting me @britishmuseum ! #britishmuseum. Next stop @japanexpo in Paris. l play #redswan, the theme from #attackontitan #. https://t.co/PgwHYSQuiT https://t.co/CvSLKWeejl	6101	
20	02/03/2019	["Golden Kamuy" jumps out to London !!] The British Museum (Golden Kamuy) also participates @britishmuseum An exhibition on the theme of "manga" at) will be held from 23rd May (* local time) this year! A large signboard featuring Ashi (Ri) pa is being posted at the main entrance of the museum! Please look forward to the exhibition!	5975	
21	12/08/2019	Acknowledging Britain‘s role in the slave trade is key to challenging racist ideology and deepening our understanding of the past. That's why I’m backing @thefabians proposal for a British slavery museum in London.	4872	√
22	02/11/2019	@JeremyDBoreing I took middle school students to the British Museum in London and their overall observation was that nothing in there is Britain, it's just stuff they stole from somewhere else Out of the mouths of babes...truth	4665	√
23	08/04/2018	Footprints of an ancient dog on a brick. It remains in good shape! C. 2115-2095 BC, Third Dynasty of Ur, Mesopotamia, British Museum	4651	
24	04/01/2019	@GettyMuseum @britishmuseum guys have you even seen a duck	4577	

25	28/05/2019	What! !! !! I was invited to a special event of the manga exhibition being held at the British Museum! I am chewing on the happiness of drinking wine at the Great Court of the British Museum. You can take a picture with Pikachu and it's incredible. .. This is a tremendous exhibition. .. #MangaExhibition #BritishMuseum #London Art	4557	
26	14/12/2017	@VancityReynolds @metmuseum @GettyMuseum @MuseeLouvre @deadpoolmovie @Guggenheim @britishmuseum @bucadibeppo You'e a well hung piece of art!!	4515	
27	01/12/2019	Sorry, but we are still looking at it.JamesAcastor on British museums and colonial theft https://t.co/J29S3rKXFX	4209	√
28	22/11/2018	Easter Island delegation lobbies for the British Museum to return one of its ancient 'Moai' statues https://t.co/2n83PcMxkJ	4208	√
29	27/01/2019	There's a London crime museum that I am going to visit. It's full of objects used in historical crimes, and not to be confused with The British Museum which is a museum full of stolen things.	4121	√
30	14/11/2018	Incredible use of light in the Ashurbanipal exhibition at the British Museum, bringing the ancient Assyrian carvings to life. https://t.co/xaI3XJLDZ2	4071	
31	26/12/2017	ThrowBack 2017. Art Los Angeles Contemporary / British museum in London / Subway in where? I am https://t.co/5SphBTo3XF	4051	
32	01/05/2018	This week I fulfilled a life ambition - I did a stint as a Gallery Attendant @britishmuseum . Blimey it is really hard work. But it gives a whole new perspective & you learn a huge amount about museum visiting, which we will discuss on @BBCFrontRow Late bbc2 Friday 11.05 https://t.co/9Tf7mRPZLc	3904	
33	04/01/2019	@SpadinaMuseum @britishmuseum Look at that kind dog taking their friend for a walk	3825	
34	04/01/2019	@TheMERL @britishmuseum Big duck energy	3813	
35	10/12/2018	I went to the British Museum in London and then someone asked me how did you enjoy seeing literally all stolen artifacts and I still cannot get over the accuracy of that	3706	√
36	04/01/2019	@TheMERL @britishmuseum You want ducks? We've got ducks. But they're all, er...resting. https://t.co/O9p21KuULY	3675	
37	12/06/2018	Argh. As the Mum of the house, why is it always MY problem when people can't find stuff??!! Honestly, one of these days I'll have the British Museum ringing me up wanting to know where the Ark of the Covenant is.	3668	
38	12/08/2019	Acknowledging Britain,s role in the slave trade is key to challenging racist ideology and deepening our understanding of the past. That's why I'm backing @thefabians proposal for a British slavery museum in London. https://t.co/pk4oqmDdBZ	3578	√

39	21/03/2019	The British Museum denies Munch's best-known theory of 'The Scream'	3558	
40	16/06/2018	The British Museum is packed to the rafters with stolen treasures and artefacts from India, Egypt, China, Sudan, Ghana, Nigeria, Ireland and Greece. What's British about the British Museum? The looting. https://t.co/OHm22MRP7e	3419	√
41	14/09/2017	The British Museum is getting dragged for saying Asian names can be "confusing." https://t.co/q9qJ2r47FN	3146	√
42	16/12/2019	Roman crocodile-skin armour; this astonishing helmet and cuirass was likely worn by a distinguished soldier during parades and cult processions in Roman Egypt. Perhaps a personal trophy from a successful hunt on the Nile? From Manfalout, Egypt, 3rd century AD. British Museum https://t.co/iWLbdvvDOx	3135	
43	15/12/2018	The British Museum Visitors Guide is like a catalogue of the world's stolen shit.	3131	√
44	16/10/2017	The British Museum is famously full of removed trade barriers https://t.co/afUa3UsO8a	3084	
45	17/05/2018	AMAZING Mohamed Salah's boots have been added to the British Museum's Egyptian collection to celebrate his record-breaking season. https://t.co/jfk99ucPuU	3044	
46	03/09/2018	With the total destruction of Museu Nacional of Rio de Janeiro, a 200-year institution, 20 million artifacts and historic building were lost, Museu Nacional is the equivalent of the British Museum in Brazil. Losing it is devastating, still a disgrace that could have been avoided	2966	
47	14/07/2019	This world cup trophy deserves a place in the British Museum. Would fit perfectly well with all the other things England has no business owning.	2917	
48	16/08/2018	Tonight someone who works @britishmuseum told me they once had a mummified cat in the collection but then they x-rayed it and it turns out it wasn't a cat at all, just loads and loads of snakes twisted into a cat like shape and then mummified. So, there you have it.	2884	
49	12/07/2019	Tomorrow, we are launching a petition for the return of the Lion of Knidos, which is exhibited at the British Museum in Datça Marketplace.	2879	√
50	21/10/2019	Mummy Papyrus paper with real hieroglyphs. Pharaoh's throne made of gold bag from 3-5 thousand years ago At Egypt museum, it's very good, even I've been to the British Museum a lot, the feeling is not so good, it's so good.	2867	

Source: Twitter Inc., and author's own judgement.
Table notes: No 6, 20, 23, 25, 39, 49, 50 are not written in English. These non-English tweets are translated via the Google Translator.

As discussed in Chapter 3, prior to performing a statistical adjustment to the series, we first took a step back to our tweets dataset to see what happened to the surging trends. Table A1 in Appendix 1 lists the top 50 tweets with the most likes since 2017. From Table 1A, we can observe that 17 tweets (34%) accuse the British Museum of its stolen riches in the UK's colonial history. While among the top 10 tweets of the most likes, the proportion rises to 50% (5 out of 10). The word cloud (see Figure 1A) shows words frequently used in these 17 tweets. Words such as "stolen", "artefact", "slave", and "racist" rank the most frequently used negative words, along with some neutral words like "donation" and "return". Tourism boycotts are a key risk factor in destination management and can significantly decrease the number of tourist arrivals (Yu et al., 2020). Hence, it can be assumed that these boycott tweets do not express individuals' visiting intentions or encourage others to visit. For the purpose of forecasting tourist arrivals, we removed these tweets from the dataset identified based on the keywords depicted in the word cloud.

Figure 1A Word cloud of 17 tweets criticizing the British Museum for its stolen artefacts

More specifically, we exclude *19,424* tweets from the dataset. These tweets are detected through a snowballing approach. We first identified two boycott themes for the British Museum - "contested artefacts" and "BP sponsorship" - based on tweets fetched with the query keyword "boycott British Museum" from Twitter. Then we identified keywords (e.g., "loot", "steal", "racist" and "colony") and phrases (i.e., keywords connected by a regular expression such as "artefact.*return" and "sponsor.*bp") frequently used in tweets belonging to these two themes.

Last, we removed tweets containing these keywords or phrases from the dataset. To ensure the accuracy and stability of the above tweet-filtering process, all tweets underwent a strict pre-processing process, including text cleaning, tokenisation, stop words removal and lemmatisation.

Appendix 2

Table A2 Results of the Granger causality test for adjusted Twitter-generated variables

Lag	Null hypothesis	F-Statistics	Probability	Conclusion
1	$Volumes_t$ does not Granger cause $Arrivalsa_t$	4.241	0.043	*Refuse*
1	$Arrivalsa_t$ does not Granger cause $Volumes_t$	4.395	0.040	*Refuse*
5	$Likes_t$ does not Granger cause $Arrivalsa_t$	1.604	0.174	Accept
5	$Arrivalsa_t$ does not Granger cause $Likes_t$	1.462	0.217	Accept
2	$Retweets_t$ does not Granger cause $Arrivalsa_t$	3.217	0.047	*Refuse*
2	$Arrivalsa_t$ does not Granger cause $Retweets_t$	0.429	0.653	Accept
2	$Replies_t$ does not Granger cause $Arrivalsa_t$	0.049	0.952	Accept
2	$Arrivalsa_t$ does not Granger cause $Replies_t$	2.539	0.087	Accept

Appendix 3

Table A3.1 Details of other museums' "cutting ties" activities

No	Date	Museum	Sponsor	Newspaper article/Press release
1	2016/03/13	Tate Modern	BP	"BP ends 27-year sponsorship of Tate as falling oil price takes toll" (Brown, 2016)
2	2018/10/19	National Gallery	Shell	"Shell ends National Gallery sponsorship – to delight of campaigners" (Vaughan, 2018)
3	2017/10/31	Natural History Museum	Ørsted	NHM's sponsor Dong Energy divested its upstream oil and gas production and officially announced to change its name to Ørsted on October 31, 2017 (Ørsted, 2023).
4	2022/02/22	National Portrait Gallery	BP	"National Portrait Gallery's BP sponsorship to end" (BBC, 2022)

Table notes: 1. I utilized news articles accessed via the Lexis-Nexis database to identify the dates of "cutting ties" activities of the other four major museums. 2. Our analysis focused on examining the pressure arising from public attention on other museums' decisions to sever ties with specific organizations. To achieve this, we identified the dates when major museums terminated their relationships by pinpointing the dates with a high volume of newspaper articles discussing the issue, rather than relying on the official dates of the "cutting ties" events. 3. Although I discovered a press release regarding Ørsted's name change activities, I could not locate any relevant newspaper articles using the keywords "national gallery" and "Ørsted" or "Dong Energy". Consequently, the date "2017/10/31" is not recognized as the date of the "cutting ties" event in this particular instance.

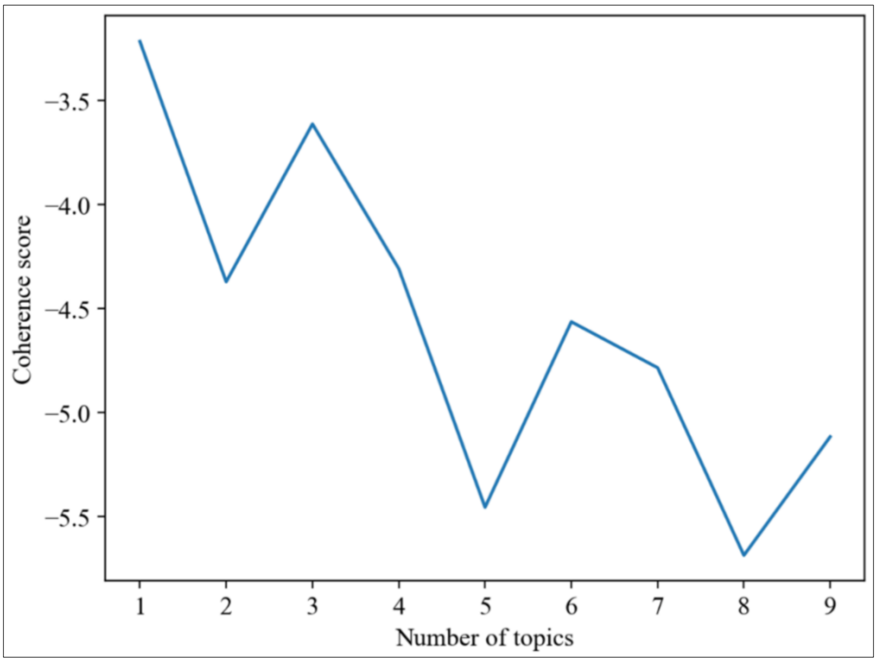

Figure A3.1 Coherence Scores of LDA Topic Models with Varying Number of Topics

Figure notes: 1. The figure presents the coherence scores of LDA topic models generated using the "tomotopy" library, with varying numbers of topics (k) for the given dataset of lemmatized and stopword-removed tweet tokens. 2. The x-axis represents the number of topics, ranging from 1 to 10, and the y-axis displays the corresponding coherence scores. 3. The LDA models were built using key parameters such as min_df set to 2 (minimum document frequency) and a fixed random seed of 555 for reproducibility. By analysing the plot, I can identify the optimal number of topics (k) for the LDA topic model that yields the highest coherence score (k=3).

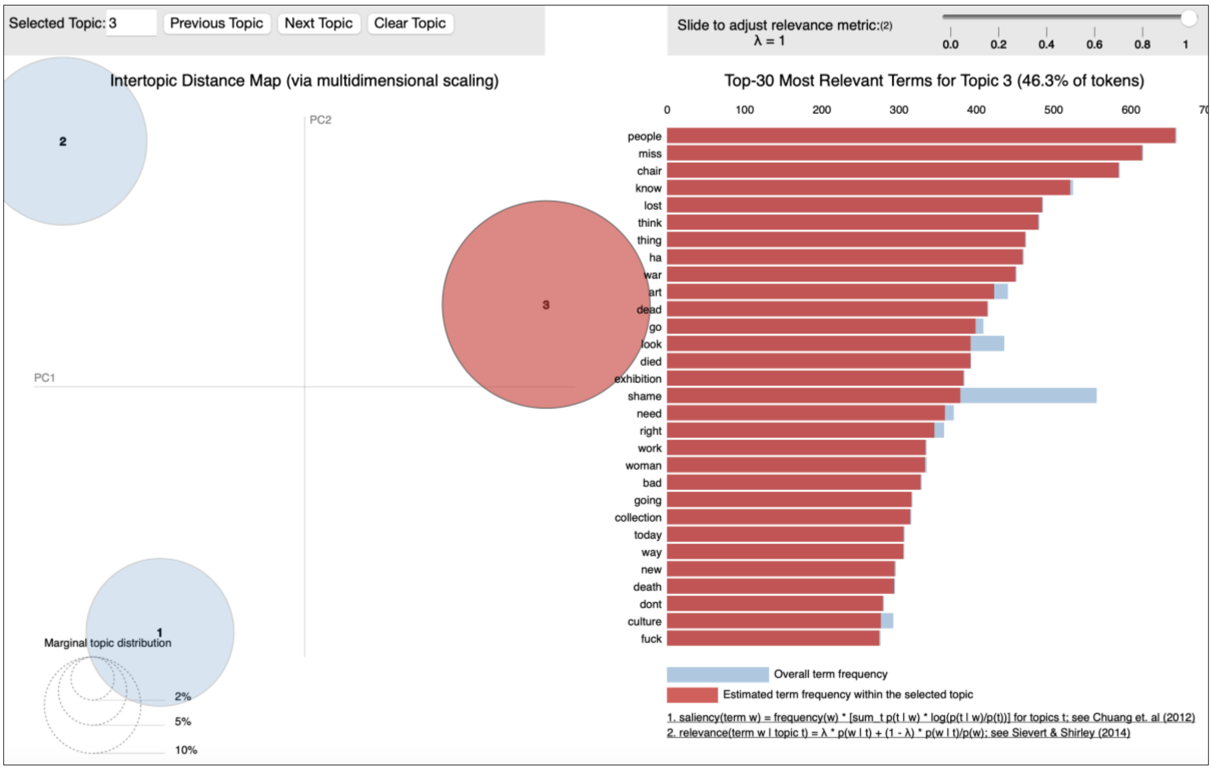

Figure A3.2 Topic distribution of negative tweets for the British Museum (part 3)

Figure notes: The red bar displays relevant terms for seed topic 3 (i.e., other tweets with negative valence).

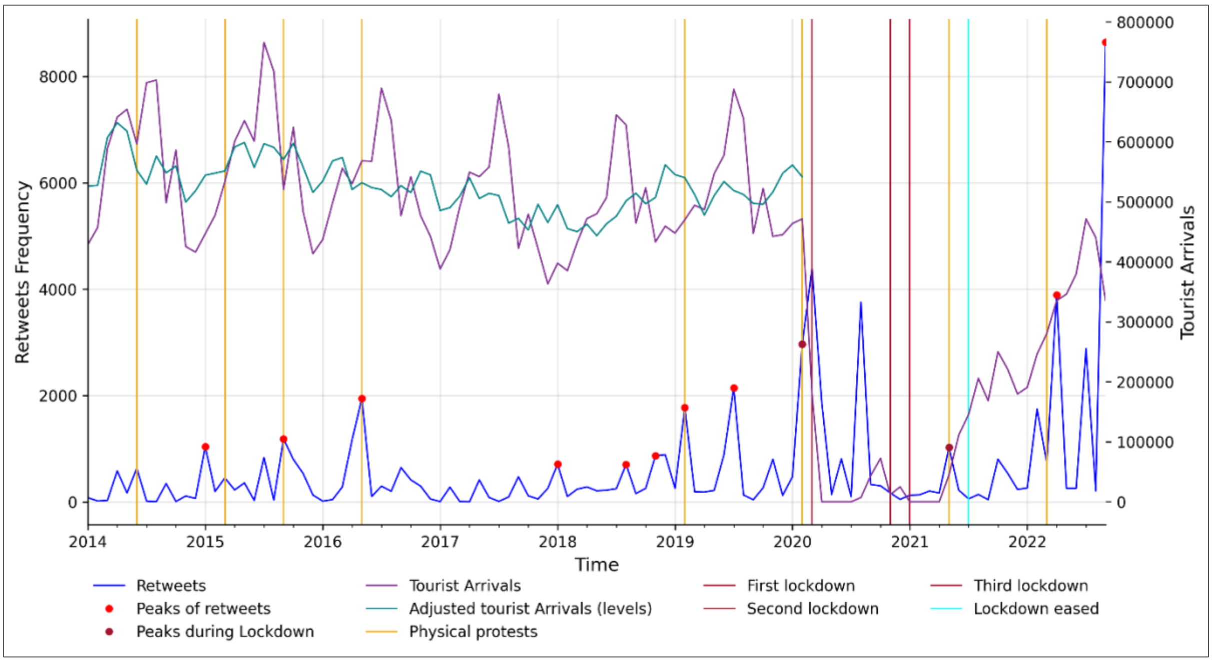

Figure A3.3 Interactions among $T_1\_Retweets_t$, tourist arrivals, travel restrictions and physical protests

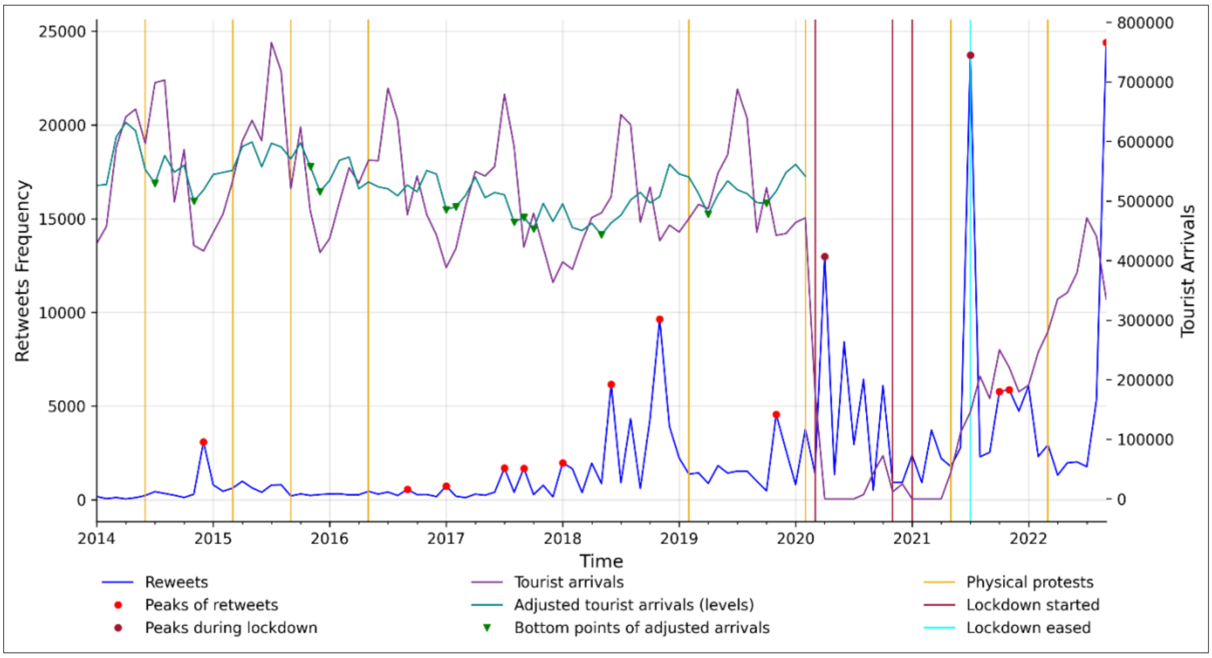

Figure A3.4 Interactions among $T_2\_Retweets_t$, tourist arrivals, travel restrictions and physical protests

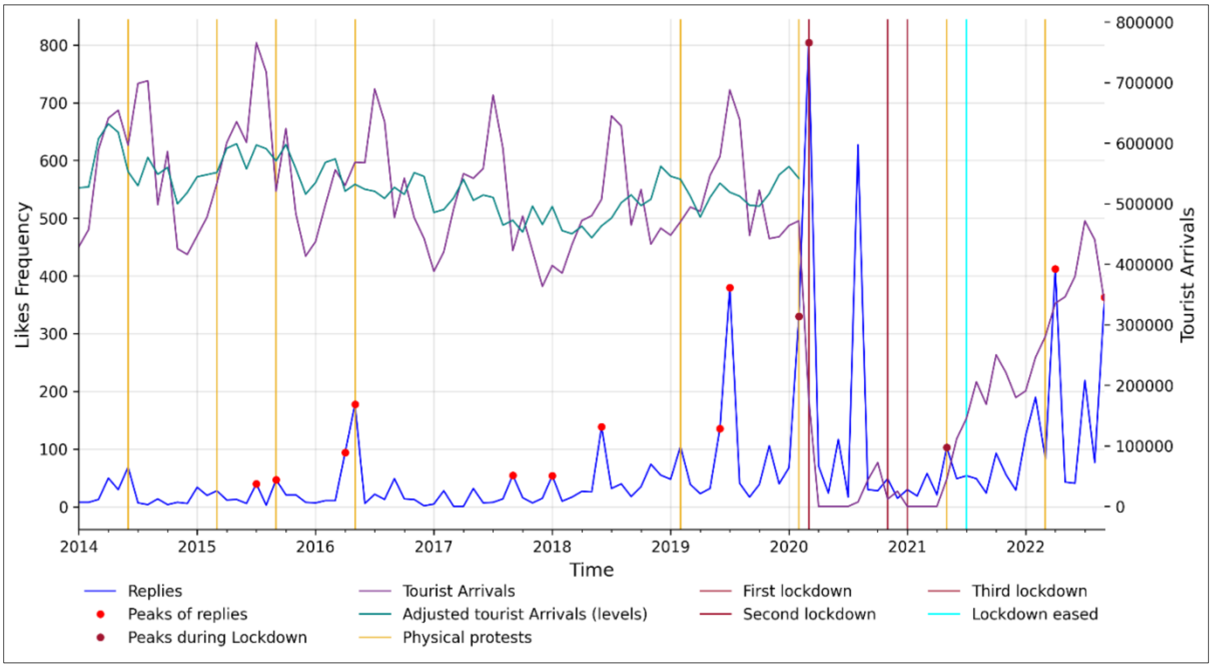

Figure A3.5 Interactions among $T_1\_Replies_t$, tourist arrivals, travel restrictions and physical protests

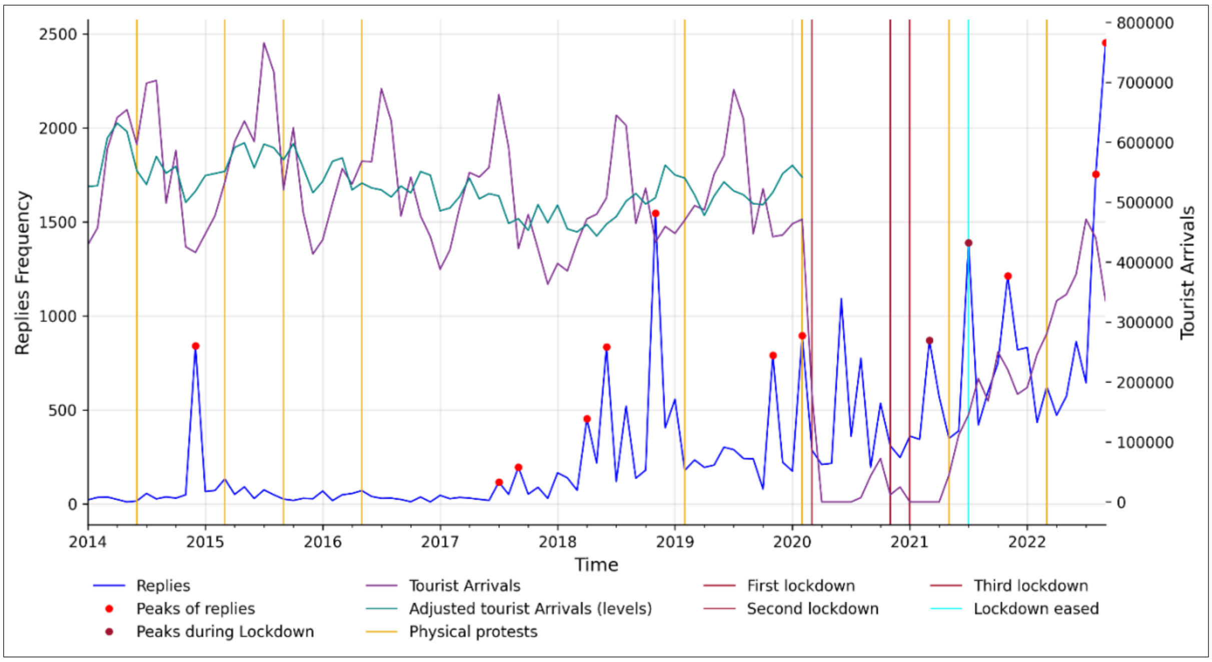

Figure A3.6 Interactions among $T_2\_Replies_t$, tourist arrivals, travel restrictions and physical protests