Insights from Tweets: Analysing Destination Topics and Sentiments, and Predicting Tourist Arrivals

LI, YULEI

Insights from Tweets: Analysing Destination Topics and Sentiments, and Predicting Tourist Arrivals

Yulei Li

Abstract

Social media has gained great popularity among consumers and social media data offers enormous potential for marketing researchers to generate consumer insights. This thesis attempts to examine how tweets can be analysed and used to predict tourist arrivals. The research objectives are: 1) to identify the most important topics from tweets; 2) to quantify the sentiments of the topics extracted; 3) to examine how the topics can predict tourist arrivals.

Two large and popular destinations were used for empirical analysis. Study 1 focuses on Sydney, Australia and Study 2 focuses on London, the UK. First, tweets mentioning the destination were extracted and tourist arrivals data were retrieved from official sources. A topic modelling algorithm, BERTopic, was then used to extract important topics from the tweets data. The sentiment scores of each topic then were calculated. Finally, H2O AutoML was used to the forecast model of tourist arrivals using the sentiment scores. The best performing model for the destination was selected. The relationships between different topics and tourist arrivals are then identified.

The results indicate that people pay attention to different topics for the two destinations. For Sydney, the main topics are events and activities and food, while for London, the main topics are events and activities, and tourism facilities. The influential Twitter topics for predicting tourist arrivals are also different. For Sydney, these are events and activities, travel costs, food, and symbolic factors, while for London, food, symbolic factors, internal factors, and geographic factors are the important factors.

The thesis extends the marketing research literature by identifying influential topics for tourist arrivals from social media, and how the sentiments on different topics discussed on social media influence tourist arrivals. It also makes methodological contributions and has important implications for practice.

Insights from Tweets: Analysing Destination Topics and Sentiments, and Predicting Tourist Arrivals

YULEI LI

Doctor of Philosophy (PhD)

Durham University Business School

DURHAM UNIVERSITY

2022

Table of Contents

Declaration

I hereby declare that this dissertation is entirely unique and has never before been submittedto any institution for evaluation. Additionally, I have adopted the similar methods proposed in this thesis to publish 3 relevant articles on reputable journals (*Please see details below*).

- Filieri, R., Lin, Z., **Li, Y.**, Lu, X. and Yang, X. (2022). Customer Emotions in Service Robot Experiences: A Hybrid Machine Learning and Thematic Analytic Approach, Journal of Service Research (ABS: 4*, IF: 10.667, Q1)
- Liu, B., Yang, Wang., Katsumata, S., **Li, Y**., Gao, W., and Li, X. (2022). National Culture and Culinary Exploration: Japan Evidence of Heterogenous, *Frontiers in Psychology* (IF: 2.990, Q2)
- **Li, Y.**, Lin, Z. and Xiao, S., (2022). Using social media big data for tourist demand forecasting: A new machine learning analytical approach*, Journal of Digital Economy*

Statement of Copyright

Chapter 1 Introduction

1.1 Research background

1.1.1 The rise of social media: The role of tweets in marketing research

Social media has become immensely popular, with the number of users reaching 3.6 billion in 2020 and projected to grow to 4.41 billion by 2025 (Statista, 2020). It has become deeply integrated into people's daily lives, impacting various aspects of their existence (Alalwan et al., 2016; Sangwaan, 2019; Ten Wong et al., 2017; Usher et al., 2014). Consequently, the amount of time individuals spend on social media has increased, with adults averaging 25.1 hours per week on social media platforms (Ofcom, 2020)

Marketing scholars now have access to vast amounts of valuable data generated by social media users and their interactions. With numerous platforms available (e.g., Facebook, Twitter, Instagram, Pinterest, YouTube), each catering to specific needs or audiences, users generate diverse forms of data, including text, visuals, and verbal content (Okazaki & Taylor, 2013). The abundance and richness of user-generated content (UGC) offer valuable insights for marketing practice (Dhaoui & Webster, 2020; Fletcher-Brown et al., 2020; Heimbach & Hinz, 2016; Reich & Pittman, 2020; Toubia & Stephen, 2013).

Furthermore, social media not only generates data for understanding consumer behavior but also influences people's intentions and behaviors, subsequently impacting product demand and destination choices (Alaei et al., 2019). Social media serves as a platform for communication and learning, enabling tourists to gain knowledge and attitudes about destinations that can influence their choices. Even reading reviews on social media can shape consumers' intentions and behaviors (Zajonc, 1968). An illustrative example is the role of social media, such as Facebook, Twitter, and YouTube, in the 2008 and 2016 presidential elections, where they played a significant part in the victories of Barack Obama and Donald Trump (Allcott & Gentzkow, 2017; Bovet & Makse, 2019; Qualman, 2012). The spread of fake news through social media notably influenced the outcome of the 2016 US presidential election (Bovet & Makse, 2019).

Since the 2010s, many marketing scholars have empirically studied the effects of social media on consumer behavior. For instance, Goh et al. (2013) investigated the impact of social media content, including user-generated and marketer-generated content, on consumer purchase behavior, finding that both types of content can influence purchasing decisions through embedded information and persuasion. Varkaris and Neuhofer (2017) noted that social media can transform consumers' decision journeys by influencing their search and booking behaviors for hotels. Messages shared on social media can sway individuals to choose or avoid specific destinations, as they acquire skills, knowledge, and attitudes through communication with peers (Kozinets et al., 2010; Ward, 1974).

1.1.2 Tourism, tourist arrivals and tourism demand

The tourism sector plays a significant role in destinations, providing both economic and non-economic benefits. Economically, it has become the world's largest industry, contributing over ten percent of global GDP, employment, and investment (World Tourism Organisation, 2021). Many local economies rely heavily on tourism (Akinboade & Braimoh, 2010; Cárdenas-García et al., 2015; Schubert et al., 2011; Sinclair, 1998).

Tourism activities also bring non-economic advantages, such as cultural and social changes. This can occur through changes in people's demands and production related to culture, art, food, recreation, and infrastructure. For instance, advancements in flying technology have facilitated mass tourism. Additionally, tourism fosters cultural exchange through direct interactions between tourists and residents, leading to cultural changes in the destination (Pan & Li, 2004) and influencing tourists' own cultural behaviors (Ramkissoon & Uysal, 2011).

Given its importance, many marketing and tourism researchers have focused on studying the mechanism of tourist demand. Understanding and forecasting demand is crucial in the market-driven economic environment, where businesses gain a competitive advantage by better understanding their customers' demand for products and services. In the context of tourism, tourist demand refers to the quantity of tourism products (goods and services) that tourists are willing to purchase under specified conditions, such as budgets and time constraints (Lohmann & Netto, 2016; Song & Li, 2008)

Tourism demand research primarily consists of two streams: tourism demand forecasting and the exploration of factors driving tourist demand. Forecasting tourism demand aids in resource allocation and sustainable tourism practices, avoiding issues like overtourism or underutilisation (Peng et al., 2014; Song & Li, 2008; Witt et al., 1994). Accurate forecasting enables businesses and destination managers to make informed decisions regarding capacity planning, efficient resource management, and sustainable development (George & Ioana, 2007; Xu et al., 2020; Yoopetch et al., 2022). Moreover, understanding the drivers of tourism demand helps marketers and destination marketing organizations enhance their products and services (Song & Li, 2008)

With the knowledge of tourism demand, destination policy-makers can plan and develop infrastructure in a sustainable manner, incorporating environmentally friendly practices and minimising negative impaces on natural and cultural resources.

Tourist arrivals is one of the essential proxies for tourism demand in tourism literature. Tourist arrivals refer to the number of tourists visiting a destination and many scholars have attempted to forecast tourist arrivals by selecting appropriate variables in their forecasting models (Fourie & Santana-Gallego, 2011; Naudé & Saayman, 2005; Nepal et al., 2019; Song et al., 2008; Sun et al., 2019).

Mere tourism arrivals forecasting, however, is not sufficient for tourism firms and destination marketing organisations (DMOs) to make decisions. Marketers focus mainly on the overarching drivers of tourism demand rather than predicting it, to gain a deeper understanding of influential factors of tourism demand to develop and improve their products and services (Song & Li, 2008). Major transportation companies, for example, need to plan to deal with the very significant surge in travel experienced each annual holiday season. Without understanding the effects of how consumer behaviour can influence tourism arrivals in each season, it is impossible for them to prepare for the surge.

The uneven distribution of tourism resources also urges tourism shareholders and researchers to investigate the determinants of tourism arrivals. The rapid expansion of international tourism due to social, economic, political and technological changes has made tourism spread from developed countries to newly industrialised countries. Competition among these destinations may arise with the growth of the industry (Lim, 2004). Investigating influential factors of tourist arrivals can assist DMOs and tourism firms in gaining a competitive advantage.

To explore the influence of social media on tourist arrivals, the study focuses on Sydney and London, two popular global destinations with a vital tourism sector (Destination NSW, 2019; Visit Britain, 2020). The COVID-19 pandemic drastically reduced international visitor numbers to both cities. In 2020, Sydney experienced an 80.9% decrease in tourist arrivals, with only 0.8 million visitors (Destination NSW, 2021). Similarly, London's arrivals dropped from 21 million in 2019 to 4.6 million in 2020 (Visit Britain, 2020). Investigating the factors influencing tourist arrivals is essential for decision-makers in the tourism sector of these cities, enabling the development of short- and long-term strategies to drive the recovery of their tourism industries.

1.2 Research problems

Scholars have developed various models to understand the factors influencing tourism demand, often drawing from classic marketing theories. For instance, Van Raaij and Francken (1984) introduced a model based on a sequential decision-making process by Engel et al. (1968) , emphasizing the role of family members in tourists' destination choices, in agreement with Clawson and Knetsch (2013) that the process of tourists' destination choice is sequential. Specifically, a travel experience may include five sequential stages: anticipation, travelling to destination, on-site experience, travelling back and feedback of experience (Clawson & Knetsch, 2013) Observing all five stages can determine whether a tourist visits a destination and if they would come back. Another example is the model by Woodside and Lysonski (1989), who proposed that tourists choose destinations by categorising potential destinations, based on the model by Nicosia (1966).

The push-pull framework is a prevalent model in tourism demand literature (Chen & Chen, 2015; Dann, 1977; Dean & Suhartanto, 2019; Han, 2019; Nikjoo & Ketabi, 2015; Uysal et al., 2008; Whyte, 2017; Yousefi & Marzuki, 2015). It posits that psychological needs 'push' tourists towards a travel experience, while destination-related factors 'pull' them towards specific locations (Dann, 1977).

However, the push-pull model has limitations in providing a comprehensive understanding of the complexities behind tourism demand (Kirkwood, 2009; Pesonen et al., 2011; Whyte, 2017). It oversimplifies the process through which factors influence decision-making and overlooks important elements like destination image and economic considerations. Additionally, the push-pull factors are often destination-specific (Pesonen et al., 2011)

Many tourism scholars have attempted to uncover the psychological process of how push-pull factors influence tourist behavior. One essential concept that connects pull factors to destination choice is destination image. Assael (1984) defines destination image as the general perception or image of a destination that is formed over time through the processing of various information. Gartner (1994) further categorises perceptions into three components: cognitive image, affective image, and conative image. Baloglu and McCleary (1999) developed Gartner's model further by replacing the conative image with the overall image, which provides a more comprehensive synthesis of cognitive and affective images.

Numerous empirical studies have also examined the impacts of influential factors on tourism arrivals. Song and Li (2008) propose that influential factors of tourism demand should include both economic and non-economic factors. Regarding economic factors, variables such as income, population, price, exchange rate, and travel costs have been widely recognized for their significant impacts by marketing and tourism scholars (Cho, 2010; De Vita & Kyaw, 2013; Demiralay, 2020; Dogru et al., 2017; Lanouar & Goaied, 2019; Martin & Witt, 1988; Martins et al., 2017; Pham et al., 2017; Shafiullah et al., 2019; Song & Li, 2008; Tavares & Leitão, 2017; Y. Yang et al., 2019).

Non-economic factors also play a role in influencing tourist destination choice and, consequently, the demand for a destination. Uysal (1998) classifies non-economic factors into exogenous and socio-psychological factors. Exogenous factors refer to macro-level information about a destination, while socio-psychological factors encompass tourists' preferences, perceptions, and attitudes towards a destination (Uysal, 1998). For example, the overall safety of a destination is considered an exogenous factor, while individual perceptions and attitudes towards safety fall under socio-psychological factors.

Um and Crompton (1990) urther refine socio-psychological factors by categorizing them into internal and external inputs. Internal inputs involve tourists' values and motives, while external inputs encompass various stimuli, including *significant stimuli* (e.g., climate and attractions), *symbolic stimuli* (destination marketing and promotion), and *social stimuli* (e.g., friends' recommendations). The interplay between internal factors and external inputs can influence tourists' intentions. Numerous empirical studies have also confirmed the impact of socio-psychological factors on tourists' behavioral intentions, such as the intention to visit or re-visit a particular destination (Baloglu, 2001; Cho, 2010; Goh, 2012; Law et al., 2019; H. Li et al., 2018; Tasci et al., 2007; Woodside & Lysonski, 1989).

Despite abundant theoretical and empirical demand research, there still exist three main literature gaps that require research attention:

1.2.1 Explicating the Spectrum of Influential Factors

Tourism demand, being a complex domain, necessitates understanding an extensive gamut of factors that impact destination selection. The variance in these factors across individuals is evident, underscoring the role of aspects such as geographical origin (Z. Li et al., 2020; A. Liu, D. X. Fan, et al., 2021; H. Liu et al., 2021; McKercher & Mak, 2019). From this expansive, ostensibly unlimited pool, it is both critical and challenging to discern and focus on the most consequential determinants.

Empirical inquiries typically opt for proxies grounded in economic theory, antecedent research findings, experiential data, or data availability (Demiralay, 2020; Law et al., 2019; Martins et al., 2017; Morley, 1992). Illustratively, consumer demand or utility theory underpins numerous econometric models (Morley, 1992), while Demiralay (2020) probes into political and economic influences on US tourism index returns. The study includes six variables in the model: geopolitical risks, partisan conflict, economic policy uncertainty, exchange rate, crude oil prices, and Dow Jones US index; involving both political and economic factors (Demiralay, 2020).

However, it is not feasible to incorporate every potential factor. Firstly, not all determinants are equally influential; their impact can vary per destination (Yang et al., 2010). For example, the effect of exchange rate fluctuations may be negligible in less developed countries (Economist Intelligence Unit, 1975). Second, many determinants likely interrelate, posing challenges such as multicollinearity in econometric models (Lee et al., 1996; Martin & Witt, 1987). Also, Law et al. (2019) noted the inclusion of an excessive number of indicators can be problematic for traditional tourism demand forecasting models. Thus, the selection process requires a delicate balance, highlighting the intricacies of tourism demand modeling.

1.2.2 Exploring Fine-grained Dimensions of Social Media data for Tourism

Demand Analysis

Social media provides valuable insights into evolving tourist preferences. Previous research has utilised social media data as predictors to enhance tourism demand forecasting, but often overlooks the direct impact of social media data on tourism demand (Colladon et al., 2019).

Many existing studies related to social media and tourism demand primarily focus on volume- or rating-based data, such as the quantity of reviews or average ratings. While these data types can improve forecasting performance, their limitations hinder more comprehensive research. For example, volume-based social media data primarily reflect tourists' interest in a destination without revealing specific preferences or sentiments (Hu et al., 2022). Rating-based data may offer glimpses into tourists' attitudes, but they lack the granularity to explore different dimensions in detail. Studies like Khatibi et al. (2018) suggest that average ratings may not significantly correlate with tourist arrivals, as tourists may prioritize specific destination characteristics over average ratings. Goulias et al. (2015) emphasise that emotional and psychological aspects can have a greater impact on decision-making than cognitive measures like cost, time, and distance. This highlights the need to extract finer dimensions from social media data to uncover the specific elements that attract tourists.

1.2.3 Need for exploring the influence of social media sentiments on tourist arrivals

While numerous factors influence tourist arrivals, understanding the relative contribution of each factor and how it changes over time is crucial. This understanding is particularly important in comparison to solely improving the absolute accuracy of forecasting algorithms. Tourists' sentiments not only affect their intentions to revisit but also have the potential to sway potential tourists who read reviews. Social media platforms and online review platforms, such as TripAdvisor, have become increasingly influential in destination choice (Hu et al., 2022; Önder et al., 2020). Although some studies have incorporated sentiment analysis in tourism research, their focus has primarily been on improving tourism demand forecasting performance (Ma et al., 2018; Önder et al., 2020; Önder et al., 2019; Sun et al., 2019; Y. Zhang et al., 2021).

1.3 Aims and Objectives

This research aims to contribute to the marketing research literature by examining the influence of social media on tourism demand. The primary objective is to investigate how different topics discussed on social media affect tourism demand for a specific destination. To address the existing literature gaps, the research establishes three main objectives:

1. **Identify the most important tweet topics from unstructured textual tweets for tourist arrivals modelling.** Considering the vast number of influencing factors, it is crucial to limit and prioritize the most significant factors. Extracting important topics from tweets can provide insights into the aspects that tourism consumers prioritize when it comes to a particular destination. By focusing on topics discussed by social media users, irrelevant factors can be ignored, leading to a more streamlined and influential tourist arrivals model.

   In addition, many existing social media tourism demand studies rely on pre-set variables in their forecasting models, based on previous literature or personal experiences. This approach may overlook emerging aspects that are important to tourists. Therefore, incorporating methods that extract the most relevant aspects valued by tourists into forecasting models can enhance accuracy and improve understanding of the effects of emerging dimensions.

2. **Extract sentiment information from unstructured social media data for tourist arrivals analysis.** This objective involves extracting sentiment and calculating average sentiment scores for each topic and time period. By transforming unstructured social media data into structured tabular data, richer information, including topics and sentiments, can be obtained compared to relying solely on volume- or rating-based data. The resulting tabular data will be used for subsequent forecasting analyses..

3. **To investigate how social media sentiment for each topic influences tourist arrivals.** This objective aims to bridge the literature gaps between social media and tourism demand. By modeling tourism arrivals using the quantified information extracted from tweets, this research serves as a pioneering study that explores the impact of social media sentiment on tourist arrivals.

These objectives collectively contribute to advancing the understanding of the relationship between social media, sentiment analysis, and tourism demand, providing valuable insights for marketing research in the field.

1.4 Structure of the thesis

The remaining part of this thesis will be divided into seven chapters to present a comprehensive analysis. Here is an overview of each chapter:

Chapter 2 delves into the existing literature on social media analytics in the context of tourism. It begins by exploring the various applications of social media in tourism research. The focus then shifts to social media tourism demand forecasting, reviewing three theories that justify the use of social media in this area. The chapter further examines empirical studies on tourism demand forecasting using social media data, covering proxy selection, data processing, forecasting models, and performance evaluation. Additionally, the literature gaps in social media tourism demand forecasting are identified and discussed.

Chapter 3 reviews theories related to tourist decision making as it forms the basis for understanding tourist arrivals. The push-pull framework in travel decision making is introduced to explore influential factors that impact tourists' destination choices. It is emphasized that these factors indirectly affect travel decisions by shaping the destination image, which in turn influences the decision to visit a particular destination.

Chapter 4 focuses on specific factors that influence tourism demand, based on previous literature. It covers two broad categories: economic factors and non-economic factors. The five most common and influential economic factors are identified as income, population, price, exchange rate, and travel costs. Non-economic factors are classified into exogenous factors and socio-psychological factors, with a particular emphasis on the external inputs. These external inputs are further categorized as significant stimuli, symbolic stimuli, and social stimuli. The chapter concludes by highlighting the gaps in the literature regarding determinants of tourism demand.

Chapter 5 outlines the methodology employed in this research. It comprises two main sections: the philosophical approach and the specific research design. The philosophical approach discusses the differences between positivism and interpretivism, followed by an analysis of the pros and cons of quantitative and qualitative methods. The research design section describes the procedure, including data collection for Sydney and London, natural language processing, theme interpretation, sentiment analysis, forecasting models, baseline models for performance evaluation, and interpretation of the forecasting model.

Chapter 6 and Chapter 7 present the results of the two studies conducted in Sydney and London, respectively. Each study begins with the dimensions extracted from the collected tweets using the BERTopic algorithm. The results of the time series analysis are then presented, covering the impact of overall sentiment on tourism arrivals and the sentiments of individual topics on tourism arrivals. The performance of the models is evaluated by comparing them with state-of-the-art time series models. The interpretation of the effects of individual topics on tourism arrivals is supported by feature importance figures and Partial Dependence Plots.

Finally, provides a summary of the key results and findings, discussing the theoretical, methodological, and practical contributions of the thesis. The limitations of the research are acknowledged, and suggestions for future studies are presented.

Chapter 2 Social media data and analytics

2.1 Social media analytics in marketing research

Social media analytics in marketing represents a transformative approach that enables businesses to glean insights from the data-rich landscape of social media platforms. By capturing, analysing, and interpreting a wide range of user interactions, from likes and shares to comments and reviews, social media analytics offers a direct line to consumers' thoughts, feelings, and behaviours. These insights are increasingly important in shaping effective marketing strategies in the digital age.

Social media platforms yield various types of data, each carrying distinct implications for marketing research. User-generated content (UGC) such as posts, reviews, and comments, provides insights into consumer opinions, preferences, and sentiments. Social network structures reveal patterns of influence and the diffusion of information, key for identifying influencers and understanding viral phenomena. User demographics enable businesses to understand and target specific consumer groups.

The methods and techniques used in social media analytics include sentiment analysis, network analysis, and predictive modeling, amongst others. *Sentiment analysis*, for example, allows researchers to gauge public sentiment towards a product, brand or campaign (Bhuta et al., 2014; Liu, 2012; Vamshi et al., 2018). Sentiment analysis offers a powerful advantage by enabling real-time collection and analysis of online comments, an asset of increasing importance considering the incessant updates of user-generated content on social media platforms (Rambocas & Pacheco, 2018). Furthermore, it facilitates the automatic extraction of objective, measurable, and consistent data on emotional expressions, making it a valuable tool that has piqued the attention of scholars and industry practitioners alike (Rambocas & Pacheco, 2018).

The majority of these studies leveraged sentiment analysis within rigorous research designs, driven by pre-defined research frameworks and primarily applied in explanatory or causal contexts, rather than descriptive or exploratory ones. Key examples include the work of Sonnier et al. (2011), who find that positive feedback has the most significant impact on sales, followed by negative and neutral comments. Tang et al. (2014) expand on this, revealing that mixed-neutral comments amplify the effect of positive and negative feedback, due to the perceived credibility and honesty they imbue. Liang et al. (2015) demonstrated that customer feedback sentiments, particularly on service quality, could predict overall sales effectively in the context of mobile phone applications.

Research by Ludwig et al. (2013) and Hennig-Thurau et al. (2015) underline the powerful influence of affect and communication style on consumer behavior, revealing that this effect persists even when controlling for multiple factors. Further, Makarem and Jae (2016) and Tirunillai and Tellis (2012) broaden the scope beyond sales to examine the role of sentiment analysis in understanding the emotional intensity of boycott messages and the impact of negative product reviews on stock market volatility. Schweidel and Moe (2014) find that the influence of online sentiments on a brand's stock market performance could vary according to the social media platform.

The importance of relationships in marketing has further evolved with the advent of social media, with networks influencing customer behavior {Kar, 2022 #641}. Social media platforms have become instrumental in providing individuals with a platform to engage with a wider audience, facilitating the rise of social media influencers who can significantly impact marketing strategies {Kar, 2022 #641}.

Network analysis in the marketing research field has been employed in studying these phenomena, especially in B2C and C2C communications across various social media platforms {Risselada, 2014 #642}{Sharma, 2018 #643}. There has been a marked use of network analysis in understanding WOM and electronic WOM communications and the diffusion of innovation{Katona, 2011 #644}{Risselada, 2014 #642}. Additionally, network structures such as size and density have been found to have substantial effects on information sharing and diffusion of innovation {Vilpponen, 2006 #645}{Katona, 2011 #644}.

Network analysis has also proved instrumental in identifying social network characteristics of individuals with potential to make viral marketing successful {Zhang, 2008 #646}). The number of followers in social media, an indicator of network size, is commonly used to identify influencers {Yoganarasimhan, 2012 #647}. However, there seems to be a trade-off between the size of one's social connections and the individual's influence over the social network, suggesting that both network size and strength of connections are important determinants of influence on social media {Liu-Thompkins, 2012 #648}{Katona, 2011 #644}{Smith, 2007 #649}.

Having delved into the depth and breadth of social media analytics in marketing research, it becomes evident that these analytical methods have transformative potential across various sectors. One such sector that can profoundly benefit from these advances is tourism. The next section will pivot our focus towards understanding how social media analytics finds its utility in tourism research. It will explore how the tourism industry, intrinsically tied to the experiences and feedback of individuals, can leverage the wealth of user-generated content on social media platforms. In turn, this can facilitate a more nuanced understanding of tourist behaviour, destination preferences, and travel trends, offering invaluable insights for both researchers and practitioners in the field.

2.2 Social media analytics in tourism research

Social media data is becoming crucial in tourism research due to its extensive influence on tourist behaviour and its potential to enhance industry practices. Social media platforms are fundamentally transforming how travellers plan, select, and review their tourism experiences, forming a rich source of data for researchers (Dwivedi et al., 2007; Leung & Bai, 2013). User-generated content (UGC), in particular, provides unfiltered insights into the perspectives of tourists, shedding light on consumer trends, preferences, and behaviours (Tham et al., 2013).

Social media data also plays a critical role in tourism marketing and management strategies. It aids in understanding the effectiveness of marketing campaigns, allows for targeted promotions, and assists in improving products and services (Gonzalo, 2012; Matloka & Buhalis, 2010). Furthermore, it enables a deeper understanding of the local community's impacts on the tourism system, thus contributing to the development of more holistic and sustainable tourism strategies (Zeng, 2008).

Quantitative studies using social media data help ascertain the economic contributions of social media to the tourism industry, demonstrating its impact on the number and length of visits, visitor satisfaction, and the frequency of return visits (Lu, 2012; Milano et al., 2011). Hence, social media data serves as a powerful tool in the realm of tourism research, contributing to comprehensive insights and data-driven decision-making.

Tourism scholars have attempted to extract those insights from various types of user-generated data, such as texts, images, audio, videos, and search engine indices (Zhang et al., 2019). Among those types, textual data is the most prevalent in marketing and tourism research. Social media platforms enable customers to express their satisfaction and dissatisfaction toward products, or tourists to express their feelings toward the destinations that they have visited or plan to visit.

Social media textual data includes two major categories: *review* data and *blog* data (J. Li et al., 2018). *Review* data refers to those comments or reviews by tourists to express their satisfaction with tourism products, such as a certain hotel, destination, or restaurant. Tourism scholars collect review data mainly for exploring the prevalent topics that can influence tourists' satisfaction. For *Blog* data, on the other hand, refers to those comments posted on blog or micro-blog platforms. Its examples include blogs, tweets, Facebook, etc. Blog data, compared with *review* data, is usually more diverse as online users can express their feelings toward any topic on those social media platforms.Tourism research usually adopt *blog* data for tourism recommendation and tourist sentiment analysis (J. Li et al., 2018). Tourism recommendation focuses on designing a better experience by understanding tourists' behaviour.

Automated analytical techniques are required in tourism and marketing research to analyse social media data. Manual analysis of unstructured social media data, such as texts and images, at a large scale is not feasible (Edwards et al., 2020). As mentioned before, textual data is the most common type of data used in tourism literature and thus, various text mining techniques have been adopted in tourism and marketing research. According to a literature review by (Xie et al., 2020), text mining techniques account for over 88% of the total papers selected. *Text mining* refers to a series of analytical techniques that extract information from natural language. Among those techniques, topic modelling is a popular approach marketing researchers adopt.

Topic modelling involves extracting topical data from natural language posted on social-media. Various machine learning models, such as latent Dirichlet Allocation (LDA) and hierarchical LDA, can help researchers analyse topical information from large-scale or big data. For example, LDA was used to extract five topics based on 6,074 Airbnb listings and analyse how those topics influence revenue in the last year (García et al., 2019). Vermeer et al. (2019) utilise topic modelling to help firms identify relevant electronic word-of-mouth on social-media. Zhang et al. (2017)use the key themes identified by topic modelling to characterise consumers' interests and match content with the right individuals.

The process of topic modelling is an inductive approach. Most algorithms for topic modelling are unsupervised, which means an algorithm can automatically extract topics of a text without associated target variables given in advance. Take the LDA model as an example, the only required parameter before the algorithm starts is the number of topics. This feature makes topic modelling an ideal tool to generate concepts and thus, this research adopted topic modelling as the qualitative section in the exploratory sequential design. The output of this step is the most discussed general topic on social media.

The output of topic modelling is a set of topics containing important keywords. Researchers have to manually interpret the keywords to conceptualise the topics. This requirement makes the ultimate results rather subjective and plausible. Moreover, similar to traditional data, the accuracy and validity of topic modelling are another critical concern. The topic generation process is automatic without human input, whether the output is accurate and valid remains mysterious. Therefore, according to the exploratory sequential design as discussed above, a quantitative study should assess the predictive power of those topics generated and interpreted.

Sentiment analysis is another important application of text mining in marketing since it provides an easy-to-interpret metric of the valence of social media content (Vermeer et al., 2019). As explained before, the quantification of relevant concepts is necessary for further quantitative analysis in the following step. According to the previous literature review, attitudes or sentiments towards certain topics play an important role in consumer behaviours. The author thus decided to use sentiment analysis for those topics and quantify the consumers' attitudes towards the topics by averaging sentiment score.

Sentiment analysis is the process of classifying a piece of textual information with an associated sentiment level and exploring the authors' views on specific entities (Feldman, 2013; Liu, 2012). This makes sentiment analysis rather prevalent when exploring attitudinal information in marketing research. For example, sentiment analysis is used to identify those response-worthy content and to handle customer complaints (Liu et al., 2017; van Laer & de Ruyter, 2010). Sentiment analysis can also be used to explore whether content sentiment can influence other marketing indicators, such as virality (Heimbach & Hinz, 2016). Another application of sentiment analysis is to integrate sentimental value with other marketing theoretical frameworks. For instance, sentiment analysis, integrated with theories of consumer information process, is adopted by Colicev et al. (2019) to explore the relationship between the firm and user-generated content (UGC).

2.3 The rationale for using social media data for marketing research

The development of social media has enabled tourists to share on abounding channels, such as social network sites or consumer review sites. Utilizsing social media big data for demand forecasting has gained popularity due to its accessibility and information-rich nature (Geva et al., 2015). Researchers have incorporated social media data to predict TV ratings (Seiler et al., 2017) and explore tourists' information search behavior and decision-making (Gavilan et al., 2018).

Sharing travel information, experience or knowledge may influence potential travellers' travel decision making and destination choice (Chu et al., 2020; Hudson & Thal, 2013; Mariani et al., 2019). According to Xiang et al. (2015) and Colladon et al. (2019), customers tend to trust reviews or opinions from other tourists more than the message from the service providers. Thus, researchers have included various social media data in their demand forecasting. Furthermore, the inclusion of social media can usually improve forecasting accuracy. Colladon et al. (2019) crawl over two million posts and find information extracted from social media big data can improve the forecasting accuracy compared with those forecasting models with Google Trend Index.

It has become evident that elements such as *word of mouth*, *consumer socialisation*, and *mere exposure effects* play crucial roles in impacting tourism demand. However, the current literature review needs a tighter connection between these elements and their implications on model development, along with a less descriptive and more critical approach.

Words of mouth (WOM). Tourists are encouranged and also love to share their experiences or opinions about a tourism product or destination. WOM can influence consumers' behaviour in three ways: tie strength, homophily, and source credibility (Brown et al., 2007). Tie strength refers to a multidimensional model that reveals the dynamism of the dyadic interpersonal relationship in the context of social networks (Money et al., 1998). A strong tie between members can positively influence those members' decision making due to a strong tie relationship may urge more frequent information exchange (Brown & Reingen, 1987). Therefore, the change in information flows can alter consumers' behaviour. Kozinets et al. (2010) analyse 83 blogs and argue that online communal messages not only enhance marketing messages but modify the process of information received by consumers.

Homophily refers to the members' characteristics' similarity within a social network (Rogers, 2010). While traditional homophily focuses on social network members' characteristics, such as age and gender, shared group interests and group mind-set are more important in an online context (Brown et al., 2007). These similarities can contribute to interpersonal attraction and trust and further influence members' behaviour within a social network (Ruef et al., 2003).

Source credibility involves the evaluation of the credibility of an information sender. The elaboration likelihood model notes that a communicator's credibility plays an essential role in the persuasion process (Hass, 1981). Consumers tend to be influenced by the information with higher perceived credibility. Source credibility theory believes that the source credibility of a piece of information is influenced by perceived source expertise and perceived source bias (Birnbaum & Stegner, 1979).

In the social media context, evaluating the source expertise or source bias is different and more difficult than the process in traditional WOM. Without knowing sufficient knowledge of an information sender's characteristics, such as profession and education level, social media users have to evaluate the source's credibility using other methods. For example, users on TripAdvisor can assess the credibility of a review for a hotel by examining the past reviews posted by the reviewer. In addition, Brown et al. (2007) emphasise that in the online context, evaluations of source credibility are based on the credibility of the social media website and of the content itself.

WOM literature suggests three main stages of consumer influence: organic inter-consumer influence, linear marketer influence, and network co-production (Kozinets et al., 2010). The organic inter-consumer influence model argues that the exchange of information and influence between two consumers without the involvement of marketers. In other words, this model means the direct communication of a product or service between two consumers because of their desire to help each other, to warn about inadequate service, and/or to share their experiences and opinions (Arndt, 1967; Engel et al., 1968).

The linear marketer influence model involves the influence of opinion leaders on a herd of consumers (Feick & Price, 1987; King & Summers, 1970). People believe that those 'influencers' or 'opinion leaders' are more likely to provide accurate and authentic information (Engel et al., 1968). Marketers, thus, have been interested in identifying those 'influencers' or 'opinion leaders' and encouraging them to deliver marketing messages.

The network coproduction model refers to how a piece of information influences a consumer who may influence other consumers within the individual's social network. The development of the Internet and social media dragged scholars' attention from transaction-based to relationship-based (Vargo & Lusch, 2004). A consumer may share his/her opinion about a particular product. The shared message may influence another group of consumers who may influence each other. At this stage, consumers become co-producers of products' value and meaning and this can influence the psychology and behaviour of the consumers within the same network (Kozinets et al., 2010; Thompson & Sinha, 2008). This phenomenon makes the exploration of the influence of WOM more complex because the process involves more implicit factors, such as social or communal norms.

Those shared messages are more likely to influence their audience to choose or avoid a destination at all these three stages of the influence of WOM. As Kozinets et al. (2010) clarify, the three models in WOM may coexist, especially in the social media world. For the organic inter-consumer model, social media platforms allow consumers to share their thoughts on products or services via a direct message function. For the linear marketer influence model, opinion leaders on social media play an essential role in influencing their followers' consumption behaviour. The nature of social media may be the best environment for the network coproduction model. Social media users publicly share their experiences or opinions which may be seen by their followers and by non-followers. Those who have seen a particular message may share or discuss it with others. These models coexist in the social media world, where consumers share their experiences directly, opinion leaders influence consumption behavior, and shared messages have a cascading effect on others' attitudes and behaviour. Jalilvand and Heidari (2017) also confirm that electronic WOM has a more powerful effect on destination image and travel intention than face-to-face WOM.

Consumer socialisation theory. Consumer socialisation theory was proposed by Ward (1974) who claims that people tend to learn skills, knowledge and attitudes from others through various communication which may alter people's cognition and behaviours. The development of social media extends the effect of socialisation to a much larger community and the communication between social media users is expected to influence their knowledge and attitudes of them towards a certain destination or tourism product.

Social media platforms provide various functionalities that facilitate socialisation processes among consumers. First, most social media platforms provide fundamental communication functionality to promote socialisation processes. Social network sites enable users to share their thoughts or opinion through texts, images, or videos. Those audiences who saw a particular post can 'like', 'share' or 'reply' to the posts. The exchange of information and opinion can encourage the process of socialisation. For example, new members of a social network site can be socialised easily into virtual groups with similar interests or opinions (Ahuja & Galvin, 2003).

Second, peer interactions on social media play a significant role in shaping consumers' attitudes and behaviours (Bush et al., 1999). Consumers often rely on reviews and peer opinions on social media when making consumption decisions (BrightLocal, 2020). Furthermore, according to the report, two essential industries in tourism, restaurants, and hotels, are the top two industries in which consumers are most likely to read online reviews.

Third, The algorithmic display of relevant information on social media platforms enables easy access and evaluation of product information. For example, a blog by Twitter engineers Koumchatzky and Andryeyev (2017) reports that Twitter uses deep learning as the central model in its timeline ranking and achieves excellent performance in displaying relevant and recent tweets to its users.

Consumer socialization theory encompasses four components: Antecedents, Socialization Processes, Socialization Agents, and Outcomes (Moschis & Churchill Jr, 1978). Antecedents refer to a consumer's social structure or demographic information, such as age and social class (Moschis & Churchill Jr, 1978). Socialisation Processes involve different ways of how the learning process between agents and learners happens. Socialisation Agents means peer groups with which learner consumers interact (Moschis & Churchill Jr, 1978; Ward, 1974). Outcomes are those specific consumer learning properties and behaviour, such as attitudes, knowledge, and purchasing behaviour.

For travellers in a social media setting, the socialisation process and socialisation agents may be the most important components. Socialisation processes take place in three major ways: modelling, reinforcement, and social interaction (Moschis & Churchill Jr, 1978). As Moschis and Churchill Jr (1978) explain, a modelling process happens when the learner wishes to imitate a particular socialisation agent since the agent's behaviour is relevant or desirable to the learner. This aligns with influencer literature that people's purchase intention and behaviour may be significantly influenced by social media influencers (Audrezet et al., 2020; Hudders et al., 2021; Kay et al., 2020). In terms of the tourism industry, Xu and Pratt (2018) postulate social media influencers can play a role as endorsers to attract travellers to choose a particular destination.

Peers and organizations act as socialization agents in a social media setting. Peers transmit attitudinal and behavioral norms, which influence consumers within the same community (Bush et al., 1999; Moschis & Churchill Jr, 1978). Social media algorithms often connect users with like-minded individuals, shaping their perception of norms. For example, if a traveller shares a post complaining that the price of a luxury hotel is too high, the audience may learn that a high price is always bad. However, if another traveller says the price is high but totally worth it, the audience may learn that price is not the only standard but merely one dimension, to evaluate a hotel. Literature has confirmed this effect in various contexts (Colliander, 2019; De Bruyn & Lilien, 2008; Trusov et al., 2010).

Official accounts of Destination Marketing Organizations (DMOs) and non-DMO accounts, such as news media, also influence consumers' decision-making (Chinchanachokchai & de Gregorio, 2020). A DMO's official account is expected to share mostly positive attributes about the destination, while the attitudes of a non-DMO toward the same destination are not always positive depending on specific circumstances. More exposure time to the DMO's posts or non-DMO's posts may determine the audience's attitudes and behaviour.

Two models, the cognitive development model and social learning theory, explain how socialisation influences consumers' cognition and behavior at different age periods (Kim et al., 2009; Shim, 1996). Cognitive development model focus on the impact of socialisation on young consumers before adulthood (Kim et al., 2009). Most travel decisions are led by adults even if young family members and relatives can influence the decision-making process. The cognitive development model, thus, is not sufficient to explain the tourism demand for a destination.

Social learning theory, on the other hand, attempts to explain how socialisation process influences adult consumers' attitudes, cognition and behaviour. Social learning theory postulates that environmental resources or socialisation agents can transfer norms, attitudes, motivations and behaviours between peers within a social network (Köhler et al., 2011; Shim, 1996). The socialisation agents can be any person or organisations that interact with learners (Moschis & Churchill Jr, 1978). Social media platforms provide a platform for adult consumers to be influenced by non-family members and non-DMO accounts (Ahuja & Galvin, 2003; Chinchanachokchai & de Gregorio, 2020; De Gregorio & Sung, 2010; Taylor et al., 2011). For example, Chinchanachokchai and de Gregorio (2020) find that social media platform usage and the susceptibility to the opinion posed on social media can influence the attitude toward social media advertising.

Mere exposure effect. The mere exposure effect refers to a finding that the more people see something, the more they would like it (Zajonc, 1968). The mere exposure theory argues that people may show a positive attitude toward objects by merely exposing them to those objects without calling cognitive resources to actively process the information (Bojanic, 1991). Social media posts, according to the mere exposure effect, can still impact the cognition and affection of those readers, even though they do not directly interact with the post.

Many tourism studies have confirmed that exposing to social media posts about a destination may alter travellers' destination image and behaviour (Bigné et al., 2019; Bojanic, 1991; Colladon et al., 2019; Echtner & Ritchie, 1993; Jeong & Holland, 2012; Starosta et al., 2019; Um & Crompton, 1990). The increase in the exposure time and frequency to travel information sources may improve destination image (Choi et al., 2007; Fakeye & Crompton, 1991; Gartner, 1994). The improved image of a destination further increases the likelihood of choosing the destination (Afshardoost & Eshaghi, 2020; Ahmad et al., 2021; Crompton, 1979).

Exposure to destination advertisements may increase the familiarity with those destinations. Many studies have reported the effects of commercial advertisements on a destination (Butterfield et al., 1998; Kim et al., 2005; McWilliams & Crompton, 1997). Destination familiarity, as Kim et al. (2019) claim, can influence both travellers' cognitive and affective destination image which can further affect their intention to travel to the destination.

Tourists' virtual exposure to unknown others, however, as Latané (1981) claims, may lead to little social influence. This argument was made before the social media era and many empirical studies have confirmed the exposure (both continuous or incidental) can influence consumer behaviour. (Ferraro et al., 2009), as the pioneer of research on the impact of incidental exposure on brand choice, argue that incidentally exposing to a brand may also alter the audience's awareness of that brand.

The theories of *word of mouth*, *consumer socialisation*, and *mere exposure effects* provide a critical foundation for guiding the development of models in tourism demand forecasting (Geva et al., 2015; Moschis & Churchill Jr, 1978; Zajonc, 1968). These theories highlight the influential role of social media in shaping consumer behaviour and decision-making processes. However, to advance the field of tourism demand forecasting, it is essential to establish clear and definitive relationships between these theoretical underpinnings and the specific components of the forecasting model (Bigné et al., 2019; Colladon et al., 2019; Kim et al., 2009).

By establishing these relationships, researchers can identify precise proxies for unstructured social media data, which is crucial for accurate demand forecasting (Colladon et al., 2019). Proxies serve as indicators or measures that capture the relevant information contained within social media platforms, allowing for the integration of social media data into forecasting models (Geva et al., 2015). These proxies can include variables such as user-generated travel reviews, sentiment analysis of social media posts, engagement metrics, and network analysis of user interactions (Bigné et al., 2019; Geva et al., 2015; Moschis & Churchill Jr, 1978).

By incorporating precise proxies for unstructured social media data, forecasting models can capture the rich and diverse information shared by consumers on social media platforms. This enables a more comprehensive understanding of the factors influencing tourism demand, such as the impact of word of mouth, consumer socialisation, and the mere exposure effect (Geva et al., 2015; Moschis & Churchill Jr, 1978; Zajonc, 1968). Furthermore, the integration of social media data in forecasting models allows for a more nuanced and accurate prediction of consumer behavior, facilitating strategic decision-making for tourism industry stakeholders (Bigné et al., 2019; Colladon et al., 2019).

2.4 Analysing social media data for forecasting tourism demand

2.4.1 Tourism demand proxies

Various proxies have been selected to measure tourism demand in the literature. The number of tourist arrivals is the most prevalent dependent variable. For example, Starosta et al. (2019) extracted information from social media to forecast the tourist arrivals to seven countries, including Egypt, Greece, Italy, Spain, Portugal, Tunisia and Turkey. Colladon et al. (2019) adopted social media data in their forecasting model to predict tourist arrivals to seven European capitals.

Hotel occupancy is another popular proxy for tourism demand. Bigné et al. (2019) include tweets data to forecast the hotel occupancy in 10 Spanish tourist destinations. Another study by Ampountolas and Legg (2021) utilise social media data to predict the hotel occupancy of a resort in a major US market.

2.4.2 Social media data processing

Social media data, unlike other internet data, are usually unstructured data, such as texts and images. This characteristic makes the adoption of social media analysis in tourism demand forecasting more challenging than other data sources. According to a recent literature review by Li, Law, et al. (2021) on the use of Internet big data, in terms of social media data in tourism demand forecasting, textual data accounts for 60% while imagery data accounts for 40% in recent literature.

When dealing with these unstructured big data, researchers generally follow the five-step process: data selection, data extraction, processing and transformation, forecasting and evaluation (Li, Law, et al., 2021).

Data selection. This step involves the choice of social media platforms. According to the categorisation of Xiang and Gretzel (2010), four types of social media platforms are usually adopted in literature: social network sites (e.g. Twitter and Facebook) (Ampountolas & Legg, 2021); media/content communities (e.g. Flickr) (Miah et al., 2017); discussion forums (e.g. TripAdvisor travel forum) (Colladon et al., 2019); consumer review sites (e.g. TripAdvisor and Yelp) (Colladon et al., 2019; Eslami et al., 2018).

Data extraction. After deciding on the social media platform, as Li, Law, et al. (2021) summarise, researchers have to extract or crawl data from those social media platforms. Researchers usually adopted three main methods at this stage: manual, official API and own web crawlers. Some researchers manually collect social media data from those platforms. For example, Zhang et al. (2017) manually collected 14562 reviews by 451 tourists for 4820 restaurants on TripAdvisor. Official application programming interface (API) refers to the tools provided by platforms to enable users to scrape or interact with the contents on those platforms. This tool is widely adopted by researchers. Bigné et al. (2019) use the Twitter API to extract tweets by DMOs and other information, such as retweets and likes. Önder et al. (2020) adopted the Facebook Graph API to acquire relevant social media data to forecast the tourist arrivals to 4 cities of Graz, Innsbruck, Salzburg and Vienna. However, when the functions provided by official APIs cannot meet the requirement of researchers, they tend to take advantage of customised web crawlers. For example, Colladon et al. (2019) collect over two million posts from TripAdvisor by using a web crawler written in Java programming language.

The two major formats of social media data collected in tourism demand forecasting literature, according to the literature review of Li, Law, et al. (2021), are texts and images. Specifically, textual data accounts for 60% while imagery data accounts for 40% in previous literature. Li et al's review, however, ignores one essential type of data, which is the structured statistics generated by the interaction on social media platforms, such as the number of likes (Gunter et al., 2019) and the number of shares (Tian et al., 2021)). The statistical data generated by social media behaviour have been widely included in various forecasting models for tourism demand (Khatibi et al., 2018; Hailong Li et al., 2020; Önder et al., 2020). Including the statistical data is relatively simple in tourism demand forecasting because the data are already structured and can be converted to a longitudinal format easily.

Processing and transformation. The unstructured social media data, such as texts and images must be processed and transformed before conducting the analysis. The processing of unstructured social media data is much more complex than other types of data. The general purpose of this step is to generate time-stamped data representing individual experiences and expressions (Yang et al., 2015). Researchers adopt different processing approaches depending on the nature of the social media data collected. For textual data, such as tweets or reviews, counting the volume of the texts collected based on a certain criterion, is the most common approach (Bigné et al., 2019; Khatibi et al., 2018; Hailong Li et al., 2020; Önder et al., 2020; Qiu et al., 2021; Tian et al., 2021). Önder et al. (2020) use the number of Facebook likes to forecast the monthly tourist arrivals to four cities of Graz, Innsbruck, Salzburg and Vienna. Instead of merely counting the general texts, Bigné et al. (2019) classify the tweets they collect into four groups, namely events, tourist attractions, socialisation and commercials and count the tweets under each category for each destination to predict tourism demand and also explore the what types of tweets can influence tourism demand.

Extracting sentimental information from textual social media data is another useful and common technique in tourism demand forecasting literature (Afzaal et al., 2019; Ampountolas & Legg, 2021; Colladon et al., 2019). Tourists share their emotions and sentiments online and those emotions can be received by their friends and followers Empirical studies have proved that sentiments of user-generated content (UGC) can influence the initial acceptance of new experiential products or services (Hennig-Thurau et al., 2015). Therefore, sentiments and emotions in social media data are expected to be correlated with tourists' travel decision and should improve the forecasting accuracy of tourism demand. Colladon et al. (2019) include sentiments together with other features from the TripAdvisor travel forum in the forecasting model for tourist arrivals in seven European capitals. They note that those information does improve forecasting accuracy. Ampountolas and Legg (2021) extract sentiments from 3500 Tweets to forecast the occupancy of a resort in a major U.S. market, and also confirm the positive influence of sentimental information on the forecasting performance.

Recognising patterns and using those patterns to analyse tourism demand is another technique to process textual social media data. Toral et al. (2018) argue that unique attributes of tourist destinations from social media data are the best predictors of tourist arrivals for a destination.

Imagery data is also an important type of social media data used in tourism demand forecasting literature (Gunter & Önder, 2021; Li, Law, et al., 2021; Ma et al., 2018; Miah et al., 2017; Önder, 2017). Researchers aim to identify features from images posted on social media platforms to forecast tourism demand for a certain destination. Miah et al. (2017) identify the representative photos

Researchers often combine two or more processing techniques to extract valuable information from social media data (Khatibi et al., 2018; Hengyun Li et al., 2020; Tian et al., 2021). Extracting more than one dimension of textual data and including those dimensions into the forecasting model is a common method adopted by literature. Khatibi et al. (2018) combine the volume and ratings of reviews on TripAdvisor to predict monthly tourist arrivals to U.S. National Parks. Tian et al. (2021) include the number of new followers, the number of unfollows, the number of reads, the number of comments, the number of likes, and the number of posts by the DMO for three social media platforms, namely WeChat, Douyin (Chinese TikTok) and Sina Weibo.

Researchers also attempt to include the information extracted from both textual and imagery social media data (Gunter & Önder, 2021; Miah et al., 2017). Miah et al. (2017) identify the entities that interest tourists from textual data and extract the representative photos using Speeded-up Robust Features (SURF) and Kernel Density Estimation (KDE) for each entity. Gunter and Önder (2021) combine the metadata and volume, such as the number of likes, of photos from Instagram to forecast the tourism demand in Vienna.

Forecasting models. Researchers typically use one of two methods to forecast tourist arrivals. First, econometric models can be applied to the traditional data. Econometrics involves using statistical methods to predict future trends using past trends as a model for the data. Alternatively, machine learning and natural language processing (NLP) can be used to process the unstructured social media data. For unstructured text data, such as social media posts, they must be put into a longitudinal format so it can be used to make forecasts. Li, Law, et al. (2021) find that the econometrics models account for approximately 60% while machine learning models account for approximately 40% of the literature on tourism demand forecasting with social media data. Among the machine learning models, the artificial neural network (ANN) is the most prevalent architecture in literature (Bigné et al., 2019; Li, Law, et al., 2021; Ma et al., 2018; Starosta et al., 2019).

Performance evaluation. Developing a reliable evaluation approach is necessary to validate the performance and generalisability of proposed models. Literature usually adopts various error metrics so that they can compare their model with other benchmark models. The three most used error metrics are: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE)(Bi et al., 2021; Hengyun Li et al., 2020; Tian et al., 2021; Xie et al., 2020). These three metrics will be discussed in detail in the methodology section below.

Chapter 3 Relevant theories on tourist decision making

This chapter will review relevant theories on tourist decision making. It will first introduce the three grand models by Howard and Sheth (1969), Nicosia (1966) and Engel et al. (1968), in consumer behaviour due to travellers or tourists are essentially consumers. Travellers also have unique characteristics comparing with other consumers. For example, tourists are often on the go and have less time to comparison and make detailed decisions about purchases. This chapter, therefore, turn to relevant models and the push-pull framework in tourism to investigate what factors can influence travellers' decision making. Those factors, however, cannot directly influence tourists' decision making. Instead, there should be a psychological bridge linking those factors and desination choice decision. Destination image, therefore, as a psychological bridge between influential factors and actual decision, will be introduced at last.

3.1 Essential models and frameworks for travel decision making

The five-stage model presented by Clawson and Knetsch (2013) provides a holistic depiction of the outdoor recreation experience: anticipation, travelling to the site, on-site experience, the return journey, and feedback of the experience. Its utility is demonstrated in its predictive capacity for aggregate demand on a macro level (Sirakaya & Woodside, 2005). This model offers an overarching framework that encapsulates the entire tourism experience, providing a roadmap for our study on predicting tourist arrivals based on Twitter data. However, it lacks a granular analysis of individual decision-making processes across the stages, which necessitates the exploration of other models.

Supplementing this, the model proposed by Van Raaij and Francken (1984), evolving from Engel et al. (1968). delves deeper into the joint decision-making process and stresses the role of family and memories in shaping travel choices. This model's inclusion of collective decision-making and memory's role dovetails with our research's goal to forecast tourist arrivals. Analysing tweets can offer insights into shared decision-making processes and past experiences, hence augmenting our understanding of tourist behaviours.

Another salient model by Woodside and Lysonski (1989) views the tourist destination decision as a categorisation process for potential destinations. Echoing Nicosia (1966), hey underscore the impact of tourism suppliers' marketing activities, i.e., the 4Ps: product, price, place (channel), and promotion, on the decision process. Moreover, they assert that variables like previous experience and personal demography can shape destination awareness, and consequently, preference and intention to visit (Woodside & Lysonski, 1989). This model affirms the pertinence of our research, indicating how public sentiment towards certain topics (product, price, etc.) can influence tourist arrivals.

The "push-pull" framework is instrumental for understanding tourists' destination selection. It is fundamental to unravel what motivates visitors and allows better utilization of tourism resources. Hence, an examination of this framework was deemed crucial.

3.2 Push-pull framework for destination selection

The push-pull framework has gained wide acceptance for studying tourists' destination selection (Chen & Chen, 2015; Dann, 1977; Dean & Suhartanto, 2019; Han, 2019; Nikjoo & Ketabi, 2015; Uysal et al., 2008; Whyte, 2017; Yousefi & Marzuki, 2015).

3.2.1 Push factors

Push factors, the internal motives triggering a person to seek balance through tourism experience, play an important role in the thesis. Push factors refer to those psychological needs causing a person to feel a disequilibrium that can be corrected through tourism experience (Dann, 1977). In other words, push factors are those factors that motivate people to a destination for tourism purposes. Common push factors such as escape, relaxation, and exploration (Botha et al., 1999; Jensen, 2011) align with our research as they likely feature in social media conversations, thus influencing our sentiment analysis and prediction model.

The hierarchy of needs theory by Maslow (1981) and the psychographic theory by Plog (1974) are intimately related to these tourist-related factors (Kim & Lee, 2002). Maslow (1981) categorises human needs into five hierarchical levels namely *physiological needs, safety needs, love and belonging, esteem* and *self-actualisation*. He notes that a person would seek the higher level of needs if the lower ones are satisfied. Under this assumption, a person has the need for spending a holiday with family (*love and belonging*) or escaping from daily routine (*freedom* *needs in the esteem level*) once the person does not need to worry about his/her *physiological* and *safety needs.* In our study, tweets can act as indicators of these needs and personality types, enriching our understanding of tourist arrivals.

In the tourism context, Plog (1974) proposes a psychographic theory for travellers' decision making. The author uses the psychocentric-allocentric dimension to measure the different personalities of tourists. *Psychocentric* personality focuses more on one's self-thoughts and pays much attention to the small problem areas of one's life. A person with *allocentric* personality, on the other hand, concentrates more on seeing new things and doing new activities (Plog, 1974). The distribution of those personalities, as Plog (1974) assumes, generally follows the normal or bell-shape distribute. Most travellers locate in the mid-centric section while fewer locate at the two ends: complete psychocentric and allocentric.

Push factors connect to motivation and thus, the review of motivation theories in consumer psychology is necessary. In addition, the models by Engel et al. (1968) and Van Raaij and Francken (1984) propose that customer satisfaction can influence future decisions. Understanding motivation can help predict tourism demand, justifying the necessity of this theoretical review.

Numerous tourism researchers have paid attention to tourism motivation (Albayrak & Caber, 2018; Dann, 1981; Huang & Hsu, 2009; Iso-Ahola, 1982; Wolfe & Hsu, 2004; Yoon & Uysal, 2005). Gnoth (1997) adopts the push-pull framework to develop a motivation and expectation formation process; and he argues that both internal and external variables can turn felt needs and motives into motivations.

Four main models can explain tourism motivation: (a) need-based, (b) value-based, (c) benefits-based and (d) expectancy-based (Albayrak & Caber, 2018). (a) Need-based model, based on the theory of hierarchy of needs developed by Maslow (1954) and is one of the most significant motivation theories (Hsu & Huang, 2008). Because of the simplicity of Maslow's model, many scholars adopted this model in the field of tourism. For example, Pearce (1982) applied Maslow's theory to analyse travel motivation and Pearce (1988) developed the travel career ladder (TCL) model, including a five-hierarchy needs: relaxation, safety or security, relationship, self-esteem, and fulfilment. However, some scholars disagree Pearce's framework. Ryan (1998) notes that few empirical studies can support Pearce's concept.

(b) Value-based model attempts to explain how personal value can influence tourism motivations (Rokeach, 1968; Skidmore & Pyszka, 1987). The Rokeach Value Survey (RVS) developed by Rokeach (1968), the Value and Lifestyles (VAL) by Mitchell (1983) and the List of Values (LOV) by Kahle and Kennedy (1988) are the three mostly used measures for values. The value-based model has been used for tourism market segmentation (Madrigal & Kahle, 1994; Pitts & Woodside, 1986; Skidmore & Pyszka, 1987).

(c) Benefits-based model explores the impact of the benefits derived from tourism activities on traveller motivation (Pearce & Caltabiano, 1983). Albayrak and Caber (2018) claim that the benefits of tourism can be attribute-based (e.g. pull factors of destinations), psychological-based (e.g. emotional pleasure), or both. The benefits-based model usually links to the need-based model (Albayrak & Caber, 2018). The benefits-based model also aligns with the rationality assumption in the field of consumer behaviour that tourists try to maximise utility (Van Raaij & Francken, 1984; Wahab et al., 1976; Woodside & MacDonald, 1994).

(d) Expectancy model refers to the belief that travellers' motivation is greatly influenced by the expectancy of the outcomes of the intended trip (Hsu et al., 2010). Witt and Wright (1992) state that the expectancy of many related factors, such as cost can lead to the final destination decision. Even though Witt and Wright admit that it is difficult to apply the model in reality because of its complexity, the model depicts a fairly realistic view of tourist motivation (Sharpley, 2018; Witt & Wright, 1992). It is challenging for traditional data collection to extract those data at a large scale. Therefore, there are few empirical studies investigating the impacts of internal factors on tourism demand.

3.2.2 Pull factors

Pull factors pertain to supply-side or destination-associated elements (Albayrak & Caber, 2018; Dann, 1977). These are the factors that inspire individuals to select a particular destination to meet their psychological needs for a tourist experience (Swarbrooke & Horner, 2007). Those pull factors have a great impact on tourists' destination choice, destination image, and destination competitiveness (Klenosky et al., 1993; Sirakaya et al., 1996; Um & Crompton, 1990). Accordingly, these pull factors are central to this thesis as they may influence the topics discussed on social media that, in turn, affect the tourist arrivals at a destination. Understanding the importance of these pull factors and how they impact tourism demand is key to this research. Consequently, the upcoming sections will primarily discuss pull factors and how they impact a destination's tourism demand.

The push-pull framework can provide a theoretical perspective to interpret the results and understand the driving forces behind tourist arrivals. The differences in influential topics and sentiments between Sydney and London, for example, can be seen as reflections of different pull factors in these two destinations. This contributes to a deeper understanding of destination competitiveness in attracting tourists.

3.3 Destination Image (DI)

3.3.1 Conceptualisation Destination Image

The concept of Destination Image (DI) is a crucial determinant linking influential factors and tourists' travel decision (Assael, 1984; Baud Bovy & Lawson, 1977; Boulding, 1956; Zhang et al., 2014). Early studies like Boulding (1956) and Martineau (1958) proposes that human behaviour is dependent not on objective reality but on image perception. These studies were the precursors to the development of "image theory", which posits that an individual perceives the world as a psychological representation of objective reality existing in their mind (Myers, 1968). Drawing from "image theory", the mental representation of a destination, formed over time, can be shaped by the topics discussed in social media. (Assael, 1984; Fridgen, 1987).

The multifaceted nature of the term "destination image" has led to an array of definitions and dimensions. It comprises two interconnected components: cognitive and affective evaluations of a destination (Gartner, 1994). Cognitive evaluation involves what a person knows and thinks about a destination, while affective evaluation pertains to someone's feelings towards a destination (Baloglu & Brinberg, 1997). In marketing and consumer behaviour, these two components often fall under "beliefs" and "affect" (Holbrook, 1978). Numerous researchers have used this categorization to examine the mechanisms of these two components (Mak, 2017; Martín-Santana et al., 2017; Stylidis, 2020; Stylidis et al., 2017; Stylos et al., 2016). For example, Mak (2017) compared the projected and perceived online destination images for Taiwan by analysing tourist-generated and national tourism organisation generated content.. The author decomposes the destination into cognitive and affective evaluation together with overall image and conduct a visual content analysis and *mise en scene* analysis to identify the difference where the national tourism organisation of Taiwan can improve. Stylidis (2020) is another example of using the cognitive and affective evaluation for destination image. He uses a five-dimensional scale with 14 items to for the cognitive image and a 7-point sematic differential scale for affective image. These two components are vital to this study as they form the basis for assessing the impact of topics and sentiments extracted from tweets on tourism arrivals.

3.3.2 The Components of Destination Image

Various scholars have developed different models and frameworks to illustrate image formation. The influential and interrelated factors that impact this process mainly include stimulus factors and personal factors (Baloglu & McCleary, 1999; Beerli & Martin, 2004). These factors influence how the different components of image formation, such as cognitive and affective elements, are created (Baloglu & Brinberg, 1997; Gartner, 1996). Understanding how these components contribute to destination image and in turn, influence tourist arrivals is crucial for this study.

TDI is impacted by stimulus factors and personal factors it was argued that in this case, stimulus factors include the external stimulus and physical object, as well as previous experience of a traveller. Researchers pointed out that for a traveller to form an image of a destination, they are influenced by various information sources (amount and type); their previous experience and distribution as stimulus factors. On the other hand, personal factors were classified as either psychological or social. Under psychological factors, it was found that the destination image is impacted by travellers' values, motivations, and personalities.

For social factors, the researchers mentioned the age, education and marital status of a traveller as key in destination image formation (Baloglu & McCleary, 1999). Therefore, what is clear is that destination image can be impacted by either personal factors or stimulus factors. The actual visitation to a certain destination, a traveller is influenced by the type of information sources, age, education, and socio-psychological travel motivations in destination image formation. Jenkins (1999) went further to mention socio-economic characteristics (income, occupation and age, among others), perceptions, motivations, and psychological characteristics as crucial factors that influence a traveller's destination image formation.

More concretely, it is necessary to understand the construct of TDI within cogitation and affection. Destination image is made up of three hierarchical components: the cognitive image, the affective image and the conative image (Gartner, 1996). The function of conative image is considered to be behaviour-like and evolves from the affective and cognitive image, which motivates the tourist to choose a suitable destination (Gartner, 1996). Baloglu and McCleary (1999) fatherly developed Gartner's classification, which argues that the TDI consists of the cognitive image, the affective image, and the overall image. The overall image here is similar to the complex image proposed by Fakeye and Crompton (1991), which is a synthesis of cognitive and affective images. However, no matter how the concept format, cognitive and affective are the core component of constructing TDI, as the cognitive component is used for referring to factual knowledge, while affective components are through impressions and feelings of a destination (Alrawadieh et al., 2018; Baloglu & Brinberg, 1997).

3.3.3 Impacts of Destination Image on Travellers' Behaviour

Destination image is a crucial factor in tourism choice and marketing (Bramwell & Rawding, 1996; Chen & Phou, 2013; Chon, 1991; Dann, 1996; Echtner & Ritchie, 1993; Govers et al., 2007). It is a demand-side image formation of a tourism destination that begins with the knowledge of the destination, which can be either experiential or by searching for information (Baloglu & Brinberg, 1997; Chon, 1991; Chon, 1990).

The image of a destination may influence travellers' behaviours before, during and after the trip. Destination image can represent the destination and it can influence the decision of tourism consumers (Tasci & Gartner, 2007). Tourism products at destinations are mostly intangible. Goodrich (1978) postulates that the perception of a particular destination has a positive relationship with the preference for that destination. Chen and Kerstetter (1999) believe that travellers compare the positive image aspects of a destination with its negative image aspects before choosing a destination. More specifically, when the positive image exceeds its negative image, tourists would choose this particular destination.

Not only different types of image, but also different components of destination image can influence travellers' destination choices (Afshardoost & Eshaghi, 2020; Crompton, 1979; Tapachai & Waryszak, 2000; Tasci & Gartner, 2007; Walmsley & Young, 1998). Crompton (1979) notes that evaluative images can have more influence on tourists' destination choice than descriptive images do. A recent meta-analysis by Afshardoost and Eshaghi (2020)) finds destination image can be predictive of tourists' intentional behaviour. They also find that overall and affective images of a destination can mostly influence travellers' intentions while cognitive image may have less predictive power. The authors also conclude that destination image can significantly influence the intention to recommend which would further increase the tourism demand for a given destination.

Destination image can influence travellers' behaviour during a visit. Researchers postulate that destination image before-visit can influence enjoyment and satisfaction during a visit (Alhemoud & Armstrong, 1996; Britton, 1979; Ross, 1993; Tasci & Gartner, 2007). This statement is based on the expectancy theory (Vroom, 1964) that consumers are satisfied when their actual experience can meet their expectations. The image of a destination before visiting can form an expectation for that destination. During a visit, if the traveller's experience can live up to the expectation, he/she would feel enjoyment and satisfaction. Therefore, the level of the expectation derived from destination image may influence the enjoyment and satisfaction.

At the post-visit stage, destination image can have a positive effect on perceived quality and satisfaction, which may increase their intention to revisit or intention to recommend to other people (Crompton & Ankomah, 1993; Gartner, 1989). Overall destination image can influence the intention to revisit a destination (Ahmad et al., 2021; Baker & Crompton, 2000; Chew & Jahari, 2014; Dolnicar & Grün, 2013; Petrick, 2004; Qu et al., 2011; Ross, 1993). Ross (1993) is one of the pioneers who explore the influence of destination image on travellers' evaluation and revisit intention. He argues that visitors' positive image of a destination is more likely to attract them to revisit the destination. In a recent study, Ahmad et al. (2021) argue that the image of destinations can mediate the relationship between destination attributes and tourists' visit intention. Chew and Jahari (2014) also confirm the mediation role of destination image between perceived risks and intention to revisit.

The effect of destination image on the aggregated tourism demand has not yet been examined in detail empirically. Most studies focus on the effects of destination image at a micro-level, such as individuals' intention to revisit a destination. For an aggregated level, it is reasonable to predict the destination image can influence travellers' intention to visit or revisit a destination, and this intention can impact the tourism demand. However, the intention to revisit cannot represent an actual return and the distance between home and destination can have a negative effect on this relationship (McKercher & Tse, 2012).

Destination image (DI) is essentially a psychological construct representing how potential tourists perceive a destination. It is influenced by a multitude of factors, both economic and non-economic, and significantly impacts tourists' decision-making process (Ahmad et al., 2021; Beerli & Martin, 2004; Haarhoff, 2018; Kislali et al., 2020).

Chapter 4 Factors influencing tourism demand

The exploration of tourism demand forms a vital core of research within the broader tourism discipline. At its simplest, tourism demand refers to the quantity of individuals willing and able to travel and procure tourism services, considering a given price and specific time frame(Andersson, 2007). The comprehension of tourism demand is fundamentally crucial, serving as a bedrock for strategic planning and policymaking within the tourism sector, and ensuring sustainable progression of tourism destinations. The vast literature dedicated to understanding tourism demand reflects its significance, delving into themes spanning from socio-economic influences to digital impacts, amid the continually evolving trends and multifaceted external influences like global events, technological advancements, and climate change (Buhalis et al., 2019; Kaján & Saarinen, 2013; Poon, 1993; Van Nuenen & Scarles, 2021). For instance, Kaján and Saarinen (2013) argue that tourism contributes to climate change through various channels, including carbon emissions from transportation and energy-intensive accommodations, and it often exacerbates the vulnerabilities of local communities due to the sector's adaptation challenges to changing climate conditions.

The scholarship dedicated to tourism demand has undergone substantial evolution, with initial models and theoretical constructs offering foundations that have been built upon and nuanced over time. Early research adopted models like the Gravity Model, positing that tourism demand is influenced by factors such as the size of origin and destination populations and the distance between them (Khadaroo & Seetanah, 2008; Morley et al., 2014; María Santana-Gallego et al., 2016; Santeramo & Morelli, 2016). Classical economic theories, including utility theory, also found application in understanding tourism demand. Seminal works, such as those by Crouch (1994) and Witt and Martin (1987), offered significant contributions, applying utility theory to tourism and providing comprehensive overviews of tourism demand modelling and forecasting. Over time, the research lens widened, acknowledging the complexities of tourism demand beyond economic and geographical factors. Contemporary research themes have thus expanded to incorporate data-driven approaches, machine learning and AI in forecasting, political stability, and the impact of pandemics like COVID-19 on tourism demand (Chebli, 2020; Chen et al., 2012; Cho, 2010; Gössling et al., 2020; Law et al., 2019; Škare et al., 2021; Sun et al., 2019; Y. Zhang et al., 2021).

Having established an understanding of the evolution and importance of tourism demand, it is crucial to delve deeper into the specific factors that have been found to influence this demand. Conventional economists suggest that those variables representing price, income and tourists' preferences should be included in the demand model for any product (Barry & O'Hagan, 1972). Therefore, Song et al. (2008) provide a comprehensive function for the tourism demand in destination *i* by residents of origin *j* as below:

$$Q_{ij} = f\left(P_i, P_s, Y_j, T_j, A_{ij}, \varepsilon_{ij}\right)$$

where Q_{ij} is the quantity of the tourism demand in destination i by travellers from country j; P_i refers to the price of tourism in destination i; P_s refers to the price of tourism for substitute destinations; Y_j represents the level of income in origin country j; T_j is the tastes of the travellers from original country j; A_{ij} is advertising expenditure on tourism by destination i in origin country j; ε_{ij} represents all other influential factors for tourism demand of destination i from origin country j (Song et al., 2008). The model captures the most essential factors influencing tourism demand and generally meet the mutually exclusive and collectively exhaustive (MECE) principle. The three variables, $P_i, P_s \; and \; Y_j$ capture the economic elements whistle $T_j, A_{ij} \; and \; \varepsilon_{ij}$ capture all non-economic elements. Consequently, the section below will follow the economic/non-economic framework to explore different influential factors for tourism demand and their specific mechanism.

4.1 Economic Factors

Economic models of tourism demand studies are based on conventional economic theories, such as the demand-supply model, consumption behavioural theory and utility theory (Goh, 2012). Neoclassical economic theory believes that a consumer's behaviour is the process of utility maximisation constrained by budget (Clawson & Knetsch, 2013; Varian, 1983). These theories suggest that economic factors, such as income and price, can influence the decision making of potential tourists, which further influences a destination's tourism demand. Specifically, when making tourism decision, such as destination choice, a tourist must consider various constrains, such as budget, which is determined by his/her income level, and relative prices of a destination and its alternative choices. Therefore, the travel decision by a tourist depends on multiple economic factors.

The five most commonly examined economic factors are: *income* (Lanouar & Goaied, 2019; Martins et al., 2017; Shafiullah et al., 2019; Tavares & Leitão, 2017); *population* (Cho, 2010; Lanouar & Goaied, 2019; Pham et al., 2017); *price* (Dogru et al., 2017; Song & Li, 2008; Y. Yang et al., 2019); *exchange rate* (De Vita & Kyaw, 2013; Demiralay, 2020; Martins et al., 2017; Tavares & Leitão, 2017) and *travel costs* (De Vita & Kyaw, 2013; Demiralay, 2020; Martin & Witt, 1988).

Income

Income is one of the most examined factors which can influence tourism demand of a destination (Martins et al., 2017; Shafiullah et al., 2019; Song, Li, et al., 2010; Tavares & Leitão, 2017; Witt & Martin, 1987). Traditional economic theory argues that the income level can affect the purchasing power of origin tourists. The increase in incomes enables people to visitor other countries and purchase new types of tourism products (Barry & O'Hagan, 1972; Wakimin et al., 2018). As a result, when modelling tourism demand, it is common to include income or private consumption of the origin country as a key explanatory variable (Song et al., 2008).

The studies on the influence of income on tourism demand started before 1968. Keintz (1968) develops an aggregate demand model with the income level of tourists' origin countries as one of the explanatory variables. Barry and O'Hagan (1972) also include the UK disposable income as a proxy for income level in their demand model to explore its impact on British tourist expenditure in Ireland. They find that recreational tourists are more sensitive to income than tourists in general. Balli et al. (2016) examine the effect of the income level of 34 Organisation for Economic Co-operation and Development (OECD) countries influence their visits to 52 middle-to-low-income countries. They establish that if the income per capita in those OECD countries increase 1%, there will be a 0.53%-0.64% increase in tourism to those middle-to-low-income countries.

For the measurement of local income level, relevant measures usually enter the demand function in the form of per capita (Song et al., 2008). GDP per capita is one of the most used measures in tourism demand literature (Albalate & Bel, 2010; Gao & Su, 2019; Gao et al., 2019). For example, Martins et al. (2017) use the World's GDP per capita as the proxy of the world income level and report its positive impact on tourist arrivals. Wakimin et al. (2018) use income per capital as one of the indicators of tourist arrivals to the 5 Association of Southeast

Asia Nations (ASEAN-5) countries, including Indonesia, Malaysia, the Philippines, Singapore and Thailand.

Apart from using GDP per capita as the proxy, some scholars have attempted to seek for alternative representations of income level due to the data availability issue. For example, Dogru et al. (2017) use the industrial production index (IPI) because of the availability of monthly GDP data.

Population

Population refers to how many people live in the country, which does not include refugees not permanently settled in the country of asylum because they are generally considered part of the population of their country of origin (Cho, 2010). With more population, the cost per person in that area goes down. For example, more population urges the government in a destination to introduce a mass transport system. The efficiency of mass transport systems is higher while the costs of those systems per capita are lower. A lower price would attract more inbound tourists and thus increase the tourism demand. In addition, a bigger population encourages tourism markets to increase the quality and variety of tourism products which further leads to a rise in tourism demand (Cho, 2010).

Most analyses have provided empirical evidence that the population of origin can have a positive effect on the tourism demand of receiving countries. Eugenio-Martin et al. (2019) conduct a dynamic panel gravity model to examine the determinators of tourism demand in Africa between 1996 and 2016. They report that the increase in the population of origin countries will lead to an increase in tourist arrivals to Africa.

The measures of population vary across the literature. Most studies use the absolute population of origin (Adeola & Evans, 2020; Martins et al., 2017; Song, Li, et al., 2010). Song, Li, et al. (2010) collect population data from the IMF publications and use the data in the model of annual tourism demand for Hong Kong. Martins et al. (2017) also include population as an independent variable in their model to study the determinants of tourism expenditure and tourist arrivals in 218 countries. They obtain data for the population variable for both the model and the transforming of other variables, such as GDP, into per capita forms.

Apart from the absolute population, some scholars also use the density of the population as the proxy for the population. Population density refers to the amount of human activity in a given area and contributes to positive spillovers for locals and visitors. Thus, previous studies also confirmed the function of population density and controlled the variable in their model (Albalate et al., 2017; Cho, 2010; Gao & Su, 2019; Gao et al., 2019).

While most researchers use the population of the origin country as one of the independent variables, some scholars argue that the population of receiving countries also has an impact on tourism demand due to scale effects. Noval (1976) argues that with an increase in the population of receiving countries, the number of incoming trips will decrease. With regard to scale effects, the Poole et al. (1988) concluded that the elasticity of demand is likely to be greater when the share of travel to a particular destination is small.

Price

According to traditional economic theory, the price of products and services is a fundamental component of any demand modelling, including tourism demand. When tourists make their travel decisions, the price of a destination is an inevitable restriction. The price of a destination, therefore, should be an essential determinator for the tourism demand of the destination. When the general price of tourism products increase, the quantity of demand for the destination is expected to shrink based on the economic assumption.

Tourism demand models in literature usually use two types of price: relative price and substitute price (Song & Li, 2008). Relative price (RP) is the prices at a destination relative to the prices at tourists' origin country. First, the RP is used to measure how international prices compared to domestic prices. The RP captures how prices vary between countries, which can influence a traveller's decision on where to go on vacation. It is commonly added to the tourism demand model to see whether a person would rather travel internationally or domestically based on price differences (Dogru et al., 2017). Lim (1997) reviews 100 articles published in the 80s and report that most studies include the relative prices in their international tourism demand models.

Relative prices work well when tourists only need to make decisions on whether to travel to a single destination. However, when tourists evaluate and compare more than one destination, they need to consider another essential factor, substitute prices. Substitute price (SP) is the prices at a destination relative to the prices in competing or alternative destinations (Dogru et al., 2017; Martin & Witt, 1987; Uysal & Crompton, 1984). SP captures the impacts of price differences between the receiving country and substitutive destinations. Tourism demand models add the SP to explore how tourists' preferences change as the prices of substitutive destination change (Dogru et al., 2017). Traditional demand theory postulates that, for a good, the prices of its substitute goods can affect the demand for the good itself. Athanasopoulos et al. (2018) use the adjusted substitute price to forecast the tourism demand to Australia from six markets and confirm the reliability of the SP as a predictor.

Regarding the choice between RP and SP, some scholars prefer one to the other. For example, Martins et al. (2017) use the RP as the proxies for the "World CPI" because they believe the RP is appropriate from the perspectives of tourists. They report the negative relationship between prices and tourist arrivals which aligns with the economic assumption. Koo et al. (2017) also include the RP in their econometric model to examine the causality of both inbound and outbound tourist arrivals in Australia. They, however, claim that the correlation between relative prices and tourism demand is not statistically significant. This finding aligns with the conclusion of scholars such as Peng et al. (2015) and Lim (1999). In addition, some scholars attempt to integrate these two measures into their models. Uysal and Crompton (1984) assign 50% weights to each item to develop a weighted average price index. Martin and Witt (1988) extend the model by Uysal and Crompton (1984) and develop a similar index including RP and SP by using a different methodology for assigning weights.

When constructing prices, researchers have used different approaches. Two methods, consumer price index (CPI) and tourist services prices, are the most common ones. Consumer price index (CPI) The consumer price index measures the prices of a defined set of consumer goods (Martins et al., 2017). Uysal and Crompton (1984) adopt relative consumer price indices to examine their effect on tourist inflows into Turkey. Martins et al. (2017)collect CPI data from the World Bank and use the relative domestic prices as proxies for the World CPI. Martin and Witt (1987) sample several middle-category hotels in multiple countries and collect the average daily costs of those hotels to explore. However, they elucidate that the tourist services prices are less likely to capture the effect of price change on tourism demand than CPI.

Exchange rates

Exchange rate is the rate at which one currency is exchanged for another. It closely links to the real prices of a destination, and tourists' perceived prices of the destination. The exchange rate is simply a *price* of money (Copeland, 2008). For international tourism, exchange rates are essential when comparing prices between destination, origin countries and substitute/competing destinations. The change in exchange rates can lead to a change in prices, income, interest rates, and production levels (Isard, 1995), which will further impact tourism demand.

Although the exchange rate is a type of price, in tourism demand modelling, many authors, including Crouch (1994) and Webber (2001), argue that tourism demand models should sperate prices, such as CPI, and exchange rates. De Vita and Kyaw (2013) agree that if a country's currency devalues, international tourism will become cheaper to that country and this, as a result of decreased prices, should cause an increase in international visitors to that country. Witt and Martin (1987) also agree that foreign tourists are more aware of the fluctuation of exchange rates than the cost of living in a destination. Another reason why exchange rates should be included in the tourism demand model is they can influence tourists' perception of the price level. Tourists may not obtain knowledge of the CPI of a destination. They, however, are more likely to consider the exchange rate when they convert their domestic currencies into foreign currencies. Sharma and Pal (2020) argue that international tourists care more about currency rates than prices in the destination.

The two reasons above encourage scholars to include the exchange rates in their tourism demand models and most literature confirm the a significant relationship between exchange rates and tourism demand (De Vita, 2014; De Vita & Kyaw, 2013; Little, 1979; Santana-Gallego et al., 2010). Some scholars focus on the effects of real exchange rate on tourism demand. Little (1979) analyses annual and monthly data for the US and find that exchange rates can significantly influence tourist inflows. Lim and McAleer (2001b) examine the impacts of exchange rates between two origin countries and Australia and confirm a long-run relationship between the tourism demand in Australia and the real exchange rates of Hong Kong and Singapore. An analysis of tourist outflows from OECD nations during 1995–2004 shows that a less volatile exchange rate promotes tourism (Santana-Gallego et al., 2010). Using an annual data set, Schiff and Becken (2011) find that for a 10% increase in the exchange rate of the New Zealand dollar against the Japanese Yen, the inflow of tourists to New Zealand declined by 15.5% per year.

A number of tourism economists have argued that the exchange rate's fluctuation can have significant impacts on tourism demand. Saayman and Saayman (2013) analyse quarterly data for the period 2003 – 2010. They find that exchange rate volatility may have a significant adverse effect on foreign tourists' spending in South Africa. In another study, using a panel of 27 OECD nations from 1980 to 2011, De Vita (2014) finds that countries with a stable currency had greater demand for international tourism. Agiomirgianakis et al. (2014) report exchange rate volatility has a negative effect on demand for Turkey, by using a quarterly dataset spanning over 1994:Q4–2012:Q4.

Similar to other factors, tourists from different regions or countries may have different sensitivity to exchange rates. Quayson and Var (1982) develop a tourism demand function for the Okanagan, which is a region in Canada and confirm that residents in Washington state are more sensitive to changes in the exchange rates. Due to the easy accessibility of exchange rates data, researchers tend directly use the real exchange rates data collected. When integrating exchange rates in tourism demand models, scholars tend to combine the term with other economic factors, such as prices and travel costs.

Travel/transportation costs

Travel or transport costs involve the money and time spent in travelling from an origin to a destination (Uysal & Crompton, 1985). Different from the price factor mentioned above, the costs of travel occur during the movement of tourists from their origin to a destination. The monetary value of transport tickets, the monetary costs, including food, and the time spent during a trip can all impact the travel decision of tourists to a destination.

The impacts of international transportation costs on outbound tourism have been well examined. Lim (1997) reviewed 70 studies in the framework of microeconomic consumption theory and confirm the negative correlation between transportation costs and international demand. Dogru et al. (2017) also argue that the function of tourism demand should include the cost of transportation.

Since the travel costs can be seen as the price of travel, the mechanism of prices in tourism demand discussed above can also perform here. According to the price theory, together with the own prices of travel, the substitute cost of travel can also influence tourists' destination choice. The concept of substitute cost of travel is constructed by two dimensions. For a given destination, it can mean the travel costs to a substitute destination. It can also mean the travel costs via substitute vehicles. Martin and Witt (1988), therefore, mention the importance of substitute travel costs. They, then, disaggregate the substitute travel prices into four different variables by the two dimensions: the cost of travel by air from origin to destination; the cost of travel by air from origin to substitute destinations; the costs of travel by surface from origin to destination; the cost of travel by surface to substitute destinations. They report a composite result on the impacts of these four variables on UK aggregate outward tourism. For example, the cost of travel by *air* to substitute destinations significantly influences tourism flows to Austria while the costs of travel by *surface* to substitute destinations can impact the demand to Austria and the Federal Republic of Germany. Some scholars also find that transportation costs the most price-sensitive product that people want to cut from their budgets (Pyo et al., 1991).

Some scholars use exchange rates as the proxy for the costs of travel. Webber (2001) develops a model for the outbound tourism demand from Australia. The author adopts the exchange rates as the proxy for the travel costs variable because of the limited data availability for quarterly airfare data. He elucidates two reasons. The first one is the change in exchange rates can capture the change in costs of airline companies, and further influence the change in the prices. However, this assumption is generally against the fundamental economic theory. The economic theory claims that it is the demand instead of costs that influences the prices. Otherwise, airline companies would prefer to increase the costs as much as possible in order to increase their prices. The reality is just the opposite. The intersection of supply and demand determines the price of a good, and businesses must reduce their costs to make more profits. Compared with the first explanation, his second reason why using exchange rates as the proxy of travel costs is more convincible. Webber (2001) explains that the passengers of major airlines usually come from different countries of origin. Therefore, when airlines sell their tickets in tourists' origin countries, they have to price their tickets in the home currency. The change in exchange rates will influence airline ticket prices and further influence the tourism demand. Martin and Witt (1988) agree with the practice to use the exchange rates as the proxy for travel costs.

4.2 Non-Economic Factors

In addition to economic factors, non-economic factors, such as a tourist's own culture and tastes, can also influence the destination choice of a tourist. Due to the composite nature of those non-economic factors, scholars in tourism have attempted to categorise and structuralise those countless factors. Uysal (1998) is one of the pioneers and he split those non-economic factors into two categories: *exogenous* and *social-psychological* factors.

Exogenous factors

Exogenous factors, according to Uysal (1998), include some macro information regarding a destination, such as the general economic development; political stability; technology advancement; epidemics; terrorism, among others. The occurrence of terrorism attacks is one of the most studies exogenous factors (Araña & León, 2008; Lanouar & Goaied, 2019; Liu & Pratt, 2017; Neumayer & Plümper, 2016). Terrorism is a type of violent attack on non-combatants with the goal of spreading fear through intimidation, coercion, or instilling frustration (Schmid, 2011). Terrorism activities can lead to negative publicity for the victim destination, which may increase the global fear of travelling to the attacked destination. For example, Yaya (2009) reports, over the period from 1997 to 2006, terrorism activities in Turkey resulted in a reduction of six million foreign tourists. Krajňák (2020) conducts a systematic review and meta-analysis of a total of 45 journal articles. He concludes that generally, terrorism incidents in a destination attenuate its inbound tourism demand.

The negative effect of terrorism attacks on international tourism demand has been observed empirically (Lanouar & Goaied, 2019; Neumayer & Plümper, 2016; Samitas et al., 2018; Maria Santana-Gallego et al., 2016). Samitas et al. (2018) examine the influence of terrorism on tourist arrivals to Greece and confirm the negative impact of terrorism on tourism in the short and long term. In addition, Lanouar and Goaied (2019) explore how terrorist attacks and political violence influence tourist arrivals and overnight stays in Tunisia from January 2000 to September 2016. They find that two recent terrorist attacks on March 18, 2015, and on June 26 2015 significantly influence the tourism demand of the country. Neumayer and Plümper (2016), by analysing spatial spillover effects of those attacks in Islamic countries on Western countries' citizens, find that both tourism from the victims' countries and from other Western countries will decline.

Destination's general economic development is another essential exogenous factor influencing inbound tourism demand. While the development of the tourism industry can contribute to economic development, such as economic growth and employment (Akinboade & Braimoh, 2010; Cárdenas-García et al., 2015; Schubert et al., 2011; Sinclair, 1998), the development of the economy can also promote tourism development in a destination. Economic prosperity increases the disposable income of local people, which may lead to higher demand for recreation in the destination. The increase in the recreation demand can then increase recreation supply by attracting more recreation providers. More international tourists, therefore, may be attracted to the destination because of better recreation and tourism supplies. In addition, the economic development in the destination can increase the local population which encourages mass development of infrastructure and facilities to supply the extra demand. The improved infrastructure can attract more tourists to the destination. Previous empirical studies also confirm the positive relationship between general economic development and tourism demand (Eugenio-Martin et al., 2008)

Socio-psychological factors

Socio-psychological factors, on the other hand, include travel preferences, perceptions and attitudes about a destination, cultural similarities and tourists' demographic factors (Uysal, 1998). The social-psychological framework states that people have various and nearly unlimited wants. Certain stimuli, then, turn those wants into various motives which will further become tourism demand (Goh, 2012).

The social-psychological factors are classified by Um and Crompton (1990) into two types: *internal* and *external* inputs. *Internal* inputs are psychological factors, such as an individual's values and motives. This type of input determines how an individual processes the received information from external inputs. *External* inputs refer to the stimulus from social and marketing environments. They help people be aware of the existence of potential destinations. Those who have been aware of a destination will, then, evaluate their preferences and situational constraints to choose the ultimate decision (Um & Crompton, 1990). This process is similar to destination image formation which will be discussed below.

These *external* inputs, as Howard and Sheth (1969) summarise, can be further classified into three types of stimulus: ***significative stimuli; symbolic stimuli*** and ***social stimuli***. ***Significative stimuli***, as defined by Howard and Sheth (1969), are those attributes of a destination. This type includes both attractions (natural and human-made) and tourist activities, such as climate (natural), beaches (natural), museums (human-made); skiing (tourist activities).

Among those significative stimuli, *geographic factors* are heavily researched by tourism scholars. Tourists tend to enjoy the unique geographic resources and attractions, such as weather and beach, offered by their travel destination; also, they wish to appreciate the historical remains or cultural value that survived therein (Martín, 2005; Sánchez, 1985).

According to Martín (2005), people tend to settle in those spaces with the greatest comfort and thus, weather or climate should influence the location decision of tourists, and thus, c*limate/weather* is one of the most examined destination attributes in tourism demand literature (Cho, 2010; Goh, 2012; Law et al., 2019; H. Li et al., 2018; Li et al., 2017). For instance, Goh (2012) reports the essential role of climatic, as a non-economic variable can significantly influence the tourism demand in Hong Kong. The author constructs the climatic index by including the maximum daily temperature, the daily temperature, the precipitation, the duration of sunshine the wind speed data of Hong Kong. Becken (2013) agrees with this argument and after exploring the impact of weather on tourist arrivals in Westland, New Zealand, he finds out a significant and positive correlation between weather and tourist arrivals. Law et al. (2019) also propose that, based on an analysis of Google search volume data, tourists do pay abounding attention to the weather in Macau when they search Macau-related keywords.

However, the geographic relationship between origin and destination may affect the correlation between tourist demand and climate. If an origin and a destination country are in the same region so that the climates in those two countries are similar, tourists may be less sensitive to the weather in the destination country. Goh (2012)verifies this theory by examining the impact of climate on tourism demand to Hong Kong from four different countries. He states that those tourists from the US are more concerned by climate conditions than those from Mainland China.

Coast is another important reason why a tourist chooses a certain destination. As defined by Hall (2001), coastal tourism includes various activities that take place in the coastal zone. The quality of the coast is one of the most important reasons why tourists choose a certain destination. A survey by Gössling et al. (2012)suggests that 77% of tourists to Barbados and Bonaire would not return if the beaches in the region largely disappear. For many coastal nations, tourism is essential for their economy. For example, 26% overseas visitors visited a beach during their holiday in Scotland (Scotland Insight Department, 2016). Furthermore, tourism in Caribbean islands, as a coastal region, accounts for nearly 35% of its export earnings.

In fact, as reported by TripAdvisor (2016), beach holiday has become the most popular type of bookings (57%) on TripAdvisor.com globally. Therefore, it is reasonable to predict that how tourists perceive the coasts or beaches of Sydney can significantly influence its tourist arrivals.

Climate. To include the climatic factor into their forecasting models, authors adopted different measurements for the climate in a destination. Most of those measurements are based on objective secondary data. Mieczkowski (1985) is one of the pioneers in the field. He develops a tourism climate index (TCI) by combining several tourism-related climatic factors. The TCI has become one of the most cited climate indices to measure the climatic characteristics of a destination (Denstadli et al., 2011). For example, H. Li et al. (2018) develop a relative climate index based on the TCI developed by Mieczkowski (1985). They, then, use the index to predict a quarterly data set of tourist arrivals from Hong Kong to 13 major Chinese cities. A positive correlation between intra-annual relative climate and tourism demand is found. Nunes et al. (2013), propose a measurement for seasonal climate by combining maximum summer temperature and minimum winter temperature. They also include their measurements into their tourism demand model and find that the temperature plays different roles for domestic and international tourists.

Heritage and *attraction*. Factors regarding *heritage* (natural and cultural heritage) and *other attraction sites* are another type of variables in those tourism demand models (Cho, 2010; Gao & Su, 2019; Y. Yang et al., 2019). The list of World Heritage Sites (WHSs) can provide a “magnet for visitors” so tourism researchers have paid particular attention to it (Fyall & Rakic, 2006). Cho (2010) collects and analyses data from 135 countries to explore the non-economic determinants of tourism demand. He postulates that cultural heritage sites significantly attract tourists Asian and European tourists while natural heritages attract Asian tourists more than other regions. Yang et al. (2010) report that WHSs have a great tourism-enhancing effect and can have a significant and positive impact on international tourism demand. They also postulate that international tourists are more attracted to cultural heritages than to natural ones due to China’s long-standing historical and cultural assets.

Despite the positive effect identified by some quantitative research, other empirical studies report the heritage sites may not influence or even negatively affect tourism demand (Cuccia et al., 2016; Y. Yang et al., 2019). Y. Yang et al. (2019) conduct a meta-analysis of 43 studies with 344 econometric estimates regarding the effects of WHS on tourism demand. They report the effects of WHS listing on tourism are not as significant as expected. . Gao and Su (2019) conduct a heterogeneity analysis on panel data for 288 Chinese cities from 2000 to 2015. They report a negative effect of the World Heritage inscription on domestic tourism revenue in the developed eastern region of China, but this negative influence disappears for other regions.

Events and activities. The holding of various events and activities may attract tourists to visit a particular destination. Many tourists visit a destination to attend certain events at the destination. In addition, apart from directly attracting visitors, special tourism may also help to create an image for the host destination. A number of studies have empirically examined the effects of various events and activities on tourism demand and most of them report a positive relationship between the holding of events and the tourist arrivals to host countries (Brännäs & Nordström, 2006; Fourie & Santana-Gallego, 2010, 2011). Fourie and Santana-Gallego (2011) use a gravity model to explore whether mega-events can increase tourist arrivals to the host country. They report a generally positive relationship between mega-events and tourism demand but this relationship may various depending on various factors of the events, such as the type of mega-events or whether the event is during a peak or off-peak season (Fourie & Santana-Gallego, 2011). In addition, cultural festivals also can promote tourism demand at the host destination (Brännäs & Nordström, 2006).

Not all scholars agree that hosting mega-events, like Olympic Games, can always promote tourism demand. Vierhaus (2019) explore a country-level tourism effect for two mega-events, the Olympic Games and the FIFA World Cup and report that hosting the latter does not promote tourism before and after the event. Most literature, however, focuses on the effects of events and activities themselves on tourism demand. For example, several scholars study how the host of mega-events can influence tourists' intention to visit and spend (R. Baumann & V. Matheson, 2018; Fourie & Santana-Gallego, 2011; Gursoy & Kendall, 2006; Ritchie & Smith, 1991; Vierhaus, 2019).

Symbolic stimuli. Tourism marketing and promotion can influence the benefits and costs associated with traveling and provide a reason for making it in the first place. Symbolic stimuli refer to the promotional or marketing messages delivered by the travel industry (Howard & Sheth, 1969). Tourism marketing and promotion activities can influence the tourism demand for a particular destination from an economic perspective. According to Middleton et al. (2009), tourism marketing adapts to basic principles of traditional marketing developed and practiced for a long period. Tourist arrivals can be formulated as follows:

$$Tourist\ arrivals = leads * conversion\ rates * average\ number\ of\ repeated\ visits$$

Tourism marketing activities can influence the demand for a destination to visit by affecting all the variables in the equation above. The coverage of marketing can increase the leads. Destination marketing can increase destination familiarity, which may have positive effects on destination awareness (Chi et al., 2020). In addition, researchers postulate that destination marketing activities can increase the familiarity with a destination, and this familiarity can affect travellers' destination decision making (Baloglu, 2001; Tasci & Gartner, 2007; Woodside & Lysonski, 1989). Targeted marketing communication can increase the awareness and interest in a destination which can boost the desire of travelling to the destination and further increase its tourism demand (Court & Lupton, 1997). Chi et al. (2020) collect the opinion of 531 foreign travellers to a destination in Vietnam and report that destination familiarity moderates the relationship between destination awareness and perceived quality of travel intentions.

The quality of marketing activity can increase the conversion rates; and marketing can help tourists who have visited the destination recall the experience so that it can increase the average number of repeated visits and expenditure (Dogru et al., 2017; Song et al., 2008). Destination advertisement may influence the perceived quality of the tourism products at a destination. Tourists' perception of the quality of a destination may influence the tourists' satisfaction and further affect their intention to revisit the place (Ranjbarian & Pool, 2015; Wang et al., 2017). Tourists show a strong bias for visiting places that appear in the media. This phenomenon becomes more significant for out-of-the-way destinations, typically known by few tourists. It has been shown that word of mouth, advertisement campaigns and media coverage should be prioritized when marketing a destination to tourists, especially when trying to reach small or neglected regions.

Many empirical studies have confirmed the relationship between tourism marketing and tourism demand. Kulendran and Dwyer (2009) report that the return per dollar invested in marketing by the Australian government is 17:1 for Asia and 36:1 for New Zealand. Witt and

Martin (1987) explored the relationship between tourist arrivals to Greece and actual advertising budgets in different countries and regions; they also confirm that the increase in marketing expenditure can lead to an increase in tourist arrivals at different levels for different countries. A recent study by Tsui and Balli (2017) reports consistent results. The authors deploy three volatility models to model the arrivals of international passengers to eight Australian airports. They suggest that tourism marketing expenditure can significantly influence the tourist arrivals to the majority of those Australian airports.

Marketing communication, however, as some researchers argue, may not necessarily increase the tourism demand for a destination. For example, , Milman and Pizam (1995) imply that marketing activities may increase awareness or encourage potential travel consumers to research more about a destination. However, it is the positivity of the image of this destination that determines whether a traveller would choose this destination.

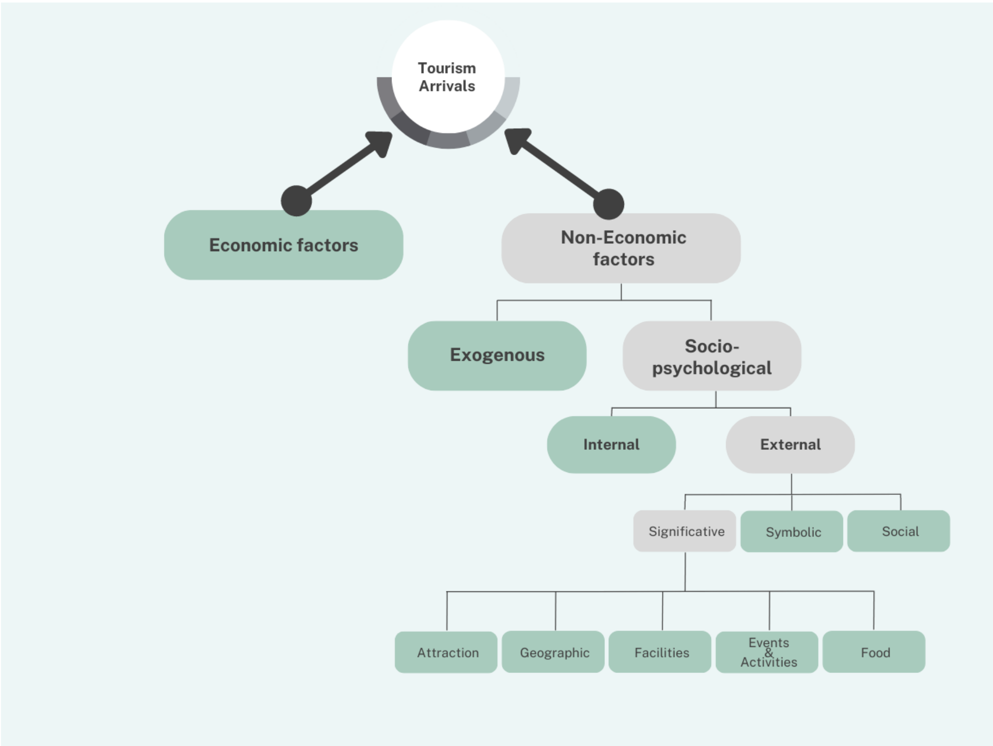
Social media has become a prevalent platform for destination marketing as it can enable destinations to offer visitors or potential visitors ample possibilities to interact with their destination. Sites like Facebook or Instagram can spread word of mouth among individuals who influence each other's decision to go to a destination, or it may just be used for entertainment such as YouTube videos of tourist attractions. Because of the fast expansion of the Internet and social media, it is now possible for destination marketers to approach potential tourists in a simple and cost-effective way (Katsikari et al., 2020).

The effect of destination marketing on social media has also been empirically proved (Hays et al., 2013; Hudson & Thal, 2013; Magno & Cassia, 2018; Minazzi, 2015; Önder et al., 2019). Social media marketing may influence tourists' satisfaction, positive word-of-mouth, behavioural intention, and thus tourism demand. Önder et al. (2019) report that the number of likes on a DMO's Facebook page can be a leading indicator of tourism demand. Hays et al. (2013) report that when customers arrived at one of the official Facebook pages of VisitBritain, visitors were 28% more likely to make a purchase and 58% more likely to spend compared to non-social-media-driven visitors. However, a criticism of social media argues, that although DMOs are investing money and time in social media marketing, they are not aware of the results which are not always measurable (Kumpu et al., 2021). Kumpu et al. (2021) also explain that non-measurable results because of social media marketing, such as tourist emotions, are as important as financial values.

Social stimuli:

4.3 Theoretical framework

Based on a comprehensive review of previous literature, the figure below presents the a structured categories of influential factors of tourist arrivals.



This thesis selects ten constructs (presented in green colour) by integrating the broad categorises outlined in the chart with the fine-grained, empirically identified factors of tourist attraction.

Economic factors: This category is directly from the literature's basic divide of tourism demand into economic and non-economic factors(Song & Li, 2008). It encompasses income, population, price, exchange rates, and travel costs.

Exogenous factors: This category is part of the non-economic factors defined byUysal (1998). They represent the macro information regarding a destination.

Internal factors: This category, a part of the ‘socio-psychological factors’, captures a tourist’s values and motives of visiting a particular destination (Um & Crompton, 1990)

Attraction: This construct is among the significative factors. According to the referenced literature, it's a common element driving tourist interest in a destination (Cho, 2010; Cuccia et al., 2016; Gao & Su, 2019; Um & Crompton, 1990; Y. Yang et al., 2019).

Geographic factors: Also identified as one of the significative factors, it deals with the geographical attributes of the destination that appeal to tourists (Cho, 2010; Goh, 2012; Law et al., 2019; H. Li et al., 2018; Li et al., 2017; Martín, 2005; Sánchez, 1985).

*Tourism facilit*ies: This factor, also a significative factor, relates to the availability and quality of facilities in a tourist destination (Khadaroo & Seetanah, 2008; Khoshnevis Yazdi & Khanalizadeh, 2017; Mathieson & Wall, 1982; Virkar & Mallya, 2018).

Events and Activities: Another significative factor, focusing on the various activities and events that a destination offers to engage tourists (R. Baumann & V. Matheson, 2018; Brännäs & Nordström, 2006; Fourie & Santana-Gallego, 2010, 2011; Gursoy & Kendall, 2006; Ritchie & Smith, 1991; Vierhaus, 2019).

Food: The last of the significative factors identified, representing the culinary aspect of the tourist experience, which often plays a major role in the choice of destination (Tommy D Andersson et al., 2017; Henderson, 2009; Rousta & Jamshidi, 2020).

Symbolic factors: This construct represents part of the 'external inputs' according to Um and Crompton (1990). It relates to the symbolic meanings or associations attached to a destination (Chi et al., 2020; Hays et al., 2013; Hudson & Thal, 2013; Magno & Cassia, 2018; Martin & Witt, 1987; Middleton et al., 2009; Minazzi, 2015; Önder et al., 2019; Tsui & Balli, 2017).

Social factors: The last construct comes from the external inputs identified by Howard and Sheth (1969). Social stimuli originate from in-person interactions with others. These encompass direct or indirect travel experiences shared by other individuals (Um & Crompton, 1990).

In summary, these ten constructs represent a fine-grained approach to understanding the dynamics of tourism demand. They integrate overarching constructs, like economic and exogenous factors, with more specific factors identified in the empirical literature (attraction, geographic factors, etc.) and represent a balanced and comprehensive view of the various inputs that may affect tourist arrivals.

Chapter 5 Methodology

5.1 Overview

This chapter explains the research philosophy; the rationale of the research design and the specific methods used.

5.2 Philosophy of research

Research philosophy is a belief about how to extract *known* knowledge from things *believe*d by data collection, data analysis and result interpretation and this is usually called *epistemological* considerations. An essential question in the social world is whether social reality can be studied as the natural sciences. Two distinctive assumptions to this question split social research into two epistemological positions: *positivism* and *interpretivism*.

5.2.1 Positivism vs interpretivism

Positivism, as defined by Levin (1988), is an epistemological stance that promotes to apply the methods of the natural sciences to obtain acceptable knowledge in social reality. Positivism researchers advocate five main principles: Knowledge is acceptable only if the senses can confirm them; positivists adopt deductive research by generating hypotheses from theory and testing those hypotheses; gathering facts is necessary; research must be conducted objectively; positivists believe that scientific statements are the true domain of the scientist while normative statements are not because they cannot be confirmed by the senses (Bryman, 2016).

Interpretivism, however, is at odds with the positivism assumption about whether the social world is identical to the natural scientific world. Unlike positivists, interpretivists believe that the subject of social science is different from that of natural science. Schultz (1962) argues that the world of nature explored by natural scientists does not mean anything to the subjects explored, such as atoms. The social reality, however, has a specific meaning for the subjects explored, such as human beings and organisations.

5.2.2 Quantitative, qualitative and mixed method

According to different epistemological positions, research methods in marketing are usually divided into two main types: quantitative and qualitative research. The distinguish between quantitative and qualitative research can help the author describe the rationale of the analytical methods proposed. As Bryman (2016) lists, *quantitative* research is a research strategy dealing with quantification in the process of data collection and analysis. He lists three major features of quantitative research.

1) Quantitative research usually entails a deductive approach. Bryman explains that quantitative researchers extract ideas from what is known and deduce relevant hypotheses followed by empirical scrutiny.
2) Quantitative research incorporated a positivism view.
3) Quantitative research represents a view that social reality is external and objective.

Bryman (2016) introduces general guidance containing eleven sequential steps, namely "theory; hypothesis development; research design; measure of concepts design; research sites selection; research subjects/respondents selection; data collection; data process; data analysis; findings/conclusions; writing up. Theory is the basis and the start of a quantitative study. Concepts are the building blocks of theory. Bulmer (1984) defines 'concepts' categorise ideas and observations in social reality. For example, the concept 'consumer behaviour', defined by Blackwell et al. (2001, p. 4), means "those activities directly involved in obtaining, consuming, and disposing of products and services, including the decision processes that precede and follow these actions". Without concepts, studying social reality is impossible. Measurements and/or indicators of a concept are necessary for quantitative research because of three primary reasons: to differ different individuals; to provide a consistent gauge for the differences; to precisely estimate the relationships between concepts (Bryman, 2016).

After measurements design, multiple data collection and analysis strategies are used in quantitative research. For data collection methods, Bryman (2016) mentions six popular ones: structured interviewing; self-administered questionnaires; asking questions; structured observation; content analysis and secondary data. For quantitative data analysis, three main approaches are: univariate analysis (analysing one variable at a time); bivariate analysis (examining the relationship between two variables; multivariate analysis (examining the relationship between three or more variables) (Bryman, 2016).

The development of quantitative data collection tactics and data analysis models has driven the popularity of quantitative research in the marketing discipline. For example, J*ournal of Consumer Research* is one of the leading journals in marketing. Over 90% of the articles in the journal have adopted quantitative research, mostly an experimental approach.

However, quantitative research has also received considerable criticism because of its epistemological and ontological assumptions. Schultz (1962) criticises that quantitative research and its positivistic view fail to distinguish the differences between the social and the natural world. This issue is common in marketing and tourism literature. For instance, Becken (2013) explores the relationship between weather and tourist arrivals by using actual climate data, such as maximum temperature. However, the direct driving force of tourist arrivals may not be the actual weather but the tourists' feelings towards the weather.

Another criticism of quantitative research is participants' answers may contrast with their actual behaviours. One reason is that research topics may be of different importance to participants' everyday life (Cicourel, 1982). Another reason can be Hawthorne effect which means people may behave differently when they are being watched (Adair, 1984; McCambridge et al., 2014). Take the study by Tezer and Bodur (2020) as an example, it explores the impacts of using green products on the consumption experience. One of their experiments tests whether using a green product can enhance their experience of headphones after listening to a piece of music. They do control the Hawthorne effect by not revealing the real purpose to the participants. However, they fail to recognise the differences in participants' music tastes, and the different levels of importance of green products to the participants.

The criticisms of quantitative research urge the development of another crucial research method, called *qualitative* research. Bryman (2016) notes qualitative research often focuses on the meanings of words instead of quantification in the data collection and analysis. He also summarises three primary features of qualitative research:

1) Qualitative research often entails an inductive approach, which is just the opposite of a deductive approach. Instead of developing and testing hypotheses, an inductive approach generates theory from observation or findings (Bryman, 2016).
2) Qualitative research entails an interpretivism epistemological position, instead of adopting a natural scientific model and positivism in quantitative research.
3) Qualitative research involves a constructionism ontological position.

> Constructionists hold a ground belief that social reality is dynamic rather than static because of both social interaction and constant revision by social members (Bryman, 2016).

Owing to the different epistemological and ontological positions, qualitative research clearly adopts a distinctive approach. Bryman (2016) introduces a six-step procedure for conducting qualitative research, namely research question; relevant sites and subjects' selection; data collection; data interpretation; conceptualisation; writing up. Qualitative research starts with research questions rather than relevant theories. Data will then be collected and interpreted followed by the induction of theory. For example, Mallat (2007) conducted a qualitative study to explore consumer adoption of mobile payment. Instead of adopting existing theory, he uses a qualitative focus group to interpret the meaning of participants' responses and finds variable factors, such as queue avoidance and transaction costs, that can influence consumers' adoption of mobile payment.

There are abundant data collection approaches and data sources: ethnography; qualitative interview; focus group; conversation and communication recording are the four most popular data collection approaches. Various documents, including personal documents, official documents, mass-media documents and virtual documents, are important data sources for qualitative researchers (Bryman, 2016).

There are two general approaches to qualitative data analysis: analytic induction and grounded theory. As defined by Johnson (2004), analytic induction requires researchers to formulate hypothetical explanations and collect data to test the proposed explanation. If there is a certain case that fails to confirm the explanation, Johnson adds, researchers need to repeat reformulating the explanation until all cases are consistent with the hypothetical explanation.

On the other hand, Bryman notes that the grounded theory approach interprets collected data to generate concepts and hypotheses. Once a hypothesis is developed, further data are collected to test the hypothesis. At last, a formal theory may be generated depending on whether the findings of new data can be consistent with the hypothesis (Bryman, 2016).

Qualitative research overcomes some criticisms received by quantitative studies. For example, focus group interviews can help obtain attitudinal information from consumers. However, qualitative research also receives similar if not more critiques. Its dissenters criticise qualitative research for subjectivity (Cavana et al., 2001; Povee & Roberts, 2014), lack of replicability (Aguinis & Solarino, 2019), problems of generalisation (Williams, 2000) and lack of transparency (Guenther & Falk, 2019).

Both quantitative and qualitative have their own advantages and limitations. Pure quantitative research does not enable researchers to extract the in-depth meaning of attitudes of consumers while a purely qualitative method cannot validate its generalisation and subjectivity. This study thus attempted to use a *mixed method*, combining quantitative and qualitative to break down the divide between these two approaches, and to overcome their shortfalls respectively (Cresswell & Plano Clark, 2011; Creswell, 1999).

According to the combination modes of quantitative and qualitative research, Cresswell and Plano Clark (2011) classify mixed methods into four major designs: 1) Convergent parallel design where quantitative and qualitative research weigh equally and simultaneously; 2) Exploratory sequential design, where qualitative data is collected and analysed before quantitative research; 3) Explanatory sequential design, where quantitative data is collected followed by qualitative data collection; 4) Embedded design, when the researcher feels qualitative (or quantitative) research alone is not sufficient and the other data should be used to compliment.

The nature of this study encouraged this study to adopt the exploratory sequential design. The author first conducted the qualitative research. In this step, social media big data was collected, and topical information was extracted and analysed. Further quantitative data analysis, including sentiment analysis and forecasting models, was applied to generate theoretical and practical implications.

5.3 Research Design

The key research question of the thesis is:

To what extent, the topics discussed in social media are predictive of tourist arrivals?

To answer the question, this thesis uses the exploratory sequential design proposed by Cresswell and Plano Clark (2011). For the qualitative section, I conducted topic modelling and sentiment analysis for social media big data (tweets) to extract themes and construct indicators of these induced themes. In the quantitative part, I applied multiple econometrics and machine learning models to forecast tourism demand using the indicators generated from the previous step. Finally, influential concepts and indicators were generated by the best-performing model.

This section describes 1) how the research collected data including social media data and tourism arrivals data; 2) how the research analysed the data using NLP techniques; 3) how the author built the optimal model for tourist arrivals forecasting.

5.3.1 Data collection

The study chose two cities for the two empirical studies: Sydney in Australia and London in the United Kingdom. Tourism is a significant contributor to the economy of these two cities.

Sydney. A report by Destination NSW (2021) states that tourism contributed over thirty billion Australian dollars to the state's gross state product. London is also among the world's leading tourism destinations. Each year, this city serves over twenty million visitors who spent over 15 million pounds (Visit Britain, 2020). In addition, there are significant indirect contributions such as tourism supporting jobs created by local businesses such as retailers, service providers and other tourism-related organisations. For example, according to Destination NSW (2021), one in every sixteen jobs was employed in the tourism sector in Sydney.

London. London is one of the most visited tourist destinations in the world (Maxim, 2019). This city has a large number of various attractions, such as historical buildings, events, restaurants and parks (Stevenson & Inskip, 2009). those attractions attracted over 21 million international tourists in 2019. Tourism is vital for London's economy, generating almost £600 million in Gross Value Added (GVA) in 2019 before Covid-19 (City of London, 2019). City of London (2019) also reports visitors to the city spent over £2 billion in 2019.

Due to the availability of data collection tools and data sources, the thesis utilised two different approaches to collect social media data and tourist arrivals data for these two cities. The specific approach used for each city will be described later in the individual study section.

5.3.2 Natural Language Processing (NLP)

Pre-processing

The data scraped contains multiple information, such as timestamps and the count of likes. Only the actual tweets were extracted due to the main objective of the study is to analyse the topics discussed and the general attitudes towards those topics. In other words, we got rid of other variables which do not help analyse topics and sentiments, such as likes, replies and retweets (Pak & Paroubek, 2010).

The author, then, used an easy-to-use python package called Preprocessor (Preprocessor, 2020). This tool can conduct a basic cleaning for tweets by removing URLs, hashtags, mentions, reserved words (e.g. RT, FAV), Emojis, Smileys, which have little effect on the main objective of this thesis. For example,

Topic modelling -BERTopic

In this step, the author extracted topics using a state-of-the-art topic modelling algorithm called BERTopic (Grootendorst, 2020). BERTopic is a topic modeling technique that uses phrase representations and context vectors to assign topics. Phrase representation refers to converting words or phrases into a mathematical form that a computer can understand and process. Context vectors are used to capture the 'context' of a word or phrase, that is, its meaning based on the words around it (Devlin et al., 2018). For example, the word 'bank' can mean different things based on the context - a place where people deposit money, or the edge of a river. Context vectors help BERTopic understand these differences.

Those topics are flexible and have multiple attributes, so they are easy to interpret (Grootendorst, 2020). Previous studies have compared BERTopic and other traditional topic modelling algorithms, such as LDA and confirm the efficiency and interpretability of BERTopic (Egger & Yu, 2022; Scarpino et al., 2022). For instance, Scarpino et al. (2022) attempt to extract insights from an adequate survey using BERTopic and LDA-based approach and their results show that BERTopic outperforms LDA, reaching an overall accuracy of 91.97%

In specific, BERTopic took three main stages to extract topics (Grootendorst, 2020).

1. **Converting textual documents to embeddings**. Computers cannot directly understand natural language. Thus, texts must be converted to 'numbers' that computers can interpret. There have been many embedding algorithms for this task, including traditional ones, such as Word2vec (Mikolov et al., 2013) and Glove (Pennington et al., 2014), and more

advanced Transformers ones, such as Bidirectional Encoder Representations from Transformers or BERT ((Devlin et al., 2018) and XLNet (Z. Yang et al., 2019). Empirical studies have proved that advanced Transformer embedding models can improve the performance of many natural language processing (NLP) tasks (Chaudhary et al., 2019; Dai et al., 2020; Devlin et al., 2018; Sun et al., 2019; Z. Yang et al., 2019). BERTopic uses those Transformers embedding models, such as BERT and other models to extract document embeddings from textual documents. BERTopic support multiple sentence transformer embedding models, such as BERT. For computation efficiency, this model chose the default embedding model "all-MiniLM-L6-v2" to convert texts into embeddings.

2. **Reducing dimensionality and clustering documents**. In this step, BERTopic uses UMAP and HDBSCAN algorithms to reduce embedding dimensionality and to cluster reduced embeddings. Document embeddings generated from the previous stage often cause the sparsity of the points and Clustering algorithms often have difficulty in high-dimensional space because of inherent sparsity (Aggarwal & Yu, 2002; Assent, 2012). The sparsity leads to the fact that interpoint distances become less informative (Shah & Koltun, 2018). Another reason why the dimensionality of embeddings should be reduced is its quadratic or near-quadratic time computational complexity because of the distance calculation of sparse data (Borodin et al., 1999; Wang et al., 2018).

   Therefore, BERTopic uses the Uniform Manifold Approximation and Projection or UMAP (McInnes et al., 2018) to reduce the dimensionality of embeddings. UMAP is a non-linear dimensionality reduction algorithm for discovering low-dimensional embeddings of high-dimensional data. UMAP, according to McInnes et al. (2018), is better with its visualisation quality and run-time performance than other dimensionality reduction algorithms, such as t-SNE.

   The reduced embeddings were clustered using the Hierarchical Density-Based Clustering or HDBSCAN (McInnes et al., 2017). HDBSCAN is an unsupervised clustering algorithm that can be used to find natural groupings in large sets of unlabelled data. It uses a novel distance function and a hierarchical tree-like layout allowing for hierarchical and overlapping clusters (Ester et al., 1996). The idea behind HDBSCAN, according to McInnes et al. (2017), is to first learn a graph-based representation of density via modularity optimization, then use the graph to guide a local search for denser clusters, and finally by growing clusters to also account for uncertainty regarding the true distance between points in the space.

3. **Creating topic representation.** This stage identifies what is different about one cluster from another based on its content (Grootendorst, 2020). The class-based TF-IDF or c-TF-ICF model was used to extract important words in each cluster and Grootendorst (2020) defined the importance score for term x within class c as

$$W_{x,c} = tf_{x,c} \times log\left(1 + \frac{A}{f_x}\right),$$

where $tf_{x,c}$ refers to the frequency of word x in class c. f_x refers to the frequency of word x across all classes. A is the average number of words per class. At the end of this stage, the important words in each cluster were extracted for interpretation.

BERTopic was selected to extract topics from reviews because of its two advantages. First, the topics generated by BERTopic are more interpretable than other traditional topic modeling methods, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003). LDA assumes that topics are distributions over words and does not take contextual information into account. In contrast, BERTopic leverages embedding models like BERT, which encode contextual information along with words. This enables BERTopic to capture the nuances and contextual meaning within the reviews, resulting in more interpretable topics. Additionally, BERTopic utilizes UMAP (Uniform Manifold Approximation and Projection) for document clustering, which helps preserve local structures within the data (Chong & Chen, 2021). This further enhances the quality of the topics generated by BERTopic.

The second advantage of BERTopic is that it eliminates the need for pre-setting the number of topics, unlike LDA (Ebeling et al., 2021; Grootendorst, 2020). Determining the optimal number of topics in LDA can be subjective and may introduce biases in the analysis. BERTopic overcomes this limitation by dynamically identifying the number of topics present in the data. This data-driven approach increases the objectivity of the results and ensures that the topics are derived based on the inherent patterns within the reviews, rather than relying on arbitrary assumptions (Grootendorst, 2020)

Since BERTopic was only proposed recently, there has been very limited research adopting this state-of-the-art algorithm. Sánchez‐Franco and Rey‐Moreno (2022) use BERTopic to extract 34 topics from 73K reviews of Airbnb guests. Ebeling et al. (2021), in their research, adopt BERTopic to discover people's concerns regarding COVID vaccines expressed on social media. As BERTopic enables enable researchers to use customised embedding models, non-English language can also be analysed using this algorithm by using non-English embedding models. Abuzayed and Al-Khalifa (2021) compared the performance of BERTopic and other topic modelling algorithms, such as LDA, based on reviews written in Arabic.

Sentiment analysis

This thesis used natural language processing (NLP) to analyse the sentiment of all tweets collected. The use of sentiment in this thesis can be justified from both a theoretical and methodological perspective. From a theoretical perspective, sentiment reflects the emotional state or attitude of individuals, which has been found to significantly impact decision-making processes, including decisions about travel (Alaei et al., 2019; Kirilenko et al., 2018; Zhang et al., 2020). A positive sentiment could indicate a pleasant travel experience or positive perception about a destination, leading to increased interest or actual visits. In addition, in tourism, personal recommendations and experiences shared (word-of-mouth) are crucial (Brown et al., 2007; Brown & Reingen, 1987; Money et al., 1998). Sentiment analysis can quantify these narratives to provide insights into collective attitudes towards a destination. If sentiments are mostly positive, it can drive up tourism demand and vice versa.

From a methodological perspective, sentiment provides a quantifiable measure to understand qualitative data (like tweets), enabling numerical analysis of otherwise subjective content. It allows the conversion of subjective feelings into a format that can be statistically analysed. Also, there are well-established machine learning tools for performing sentiment analysis, which can be applied at large scales (like on Twitter data) to give an aggregate measure of public sentiment.

The open-source CardiffNLP/Twitter-roBERTa-base model is a variant of the RoBERTa model trained by Cardiff NLP for sentiment analysis specifically on Twitter data (Barbieri et al., 2020). The model is fine-tuned on a large corpus of Twitter data to classify the sentiment of tweets into positive, negative, or neutral categories (Barbieri et al., 2020).

Twitter data presents unique challenges for sentiment analysis due to its informal language, character limitations, and the presence of emojis, hashtags, and user mentions (Bhuta et al., 2014). Models like CardiffNLP/Twitter-roBERTa-base are specifically designed to handle these challenges and perform well on sentiment analysis tasks for Twitter data (Barbieri et al., 2020). By leveraging pre-training on a large amount of text data and fine-tuning on sentiment-labeled Twitter datasets, models like CardiffNLP/Twitter-roBERTa-base can effectively capture the sentiment expressed in tweets.

The candidate used the CardiffNLP/Twitter-roBERTa-base model to determine the sentiment score for each tweet linked with a particular latent topic. Subsequently, the candidate computed the average sentiment scores per topic for each year, reflecting the general attitudes towards that topic. These average sentiment scores were then used as independent variables in the following steps.

5.3.3 Aggregation of sentiment

The sentimental data extracted was transformed into a longitudinal format at this stage. In specific, the quantification converts the sentiment score for each tweet into an aggregate sentiment score for each year (Sydney) or for each quarter (London). This aggregate scores for each topic in each period are calculated by the formula below:

$$s_{tp} = \frac{\sum s_{itp}}{N_{tp}},$$

where S_{tp} is the average sentiment score for topic t in period p. S_{itp} is the sentiment score for each tweet discussing topic t in period p. This score was calculated by TextBlob as discussed in the previous step. N_{tp} refers to the number of tweets about a certain topic t posted during the period p.

The quantification process converted the unstructured textual tweets collected into a longitudinal format, which is ready for the forecasting model. Previous studies usually measure destination image by asking tourists a series of questions on their a priori image regarding a destination via traditional surveys. This method, however, does not match the scale of tourism demand and therefore, this thesis aggregate the affective images (represented by average sentiments) for each cognitive image (topics).

5.3.4 Forecasting model: H2O AutoML

This section introduces the tourist arrivals forecasting approach by using the quantification information extracted from the steps above. Selecting relevant and significant factors or features is one of the most essential tasks for tourism demand forecasting. Tourism demand forecasting is concerned with predicting tourism demand – how many visitors will travel to a region – and tourism activity – where will they go and what will they do – based on the observations of current and historical data and information about external factors that could affect tourists' decisions to visit a place (Song & Turner, 2006).

Tourism demand forecasting is challenging and complex because of three major reasons. First, tourism market conditions and tourists' personal conditions that drive tourism demand fluctuate on a timely scale. Tourism market conditions involve economic and non-economic attributes of a destination, such as price, weather and political stability. Accurate tourism demand requires a deep understanding of the impacts of those attributes on tourism demand. Numerous attributes, however, make it difficult for researchers and practitioners to capture all different factors and understand their roles in tourism demand. For example, the reasons why a tourist chooses a destination in summer. As technologies advance, consumers' travel behaviour is becoming more difficult to predict as it is shaped by self-expression and other motivations outside of the normal "rational" actor description (Gross & Brown, 2008; Trauer, 2006).

Second, different tourism products in different destinations and different individuals can show the varying direction of the effects and sensitivities to tourism drivers over time. For example, as one of the most important determinants of tourism demand, the price plays a varying role in different situations (Crouch, 1995; Mello et al., 2002; Song, Kim, et al., 2010). In addition, the literature suggests that the impact of distance on destination choice depends on individual characteristics (McKercher & Mak, 2019; Nicolau, 2008).

Scholars have paid continuous attention to forecasting tourism demand since 1960, as accurately forecasting is essential for shareholders in the tourism industry (Artus, 1970; Gerakis, 1965; Gray, 1966; Witt & Martin, 1987). Song and Li (2008) broadly divide the tourism demand modelling and forecasting methods into quantitative and qualitative methods. There are only a few researchers adopted a qualitative approach to tourism forecasting. For example, Schwartz and Cohen (2004) invite fifty-seven experience hotel revenue managers to forecast their own daily occupancy by using simulated forecasting software. At the same time, the majority of research on forecasting tourism demand used a quantitative approach (Li et al., 2005; Song & Li, 2008).

There are three essential categories in quantitative modelling of tourist arrivals: *time series, econometric* and *machine learning* approaches. Based on whether the model examines any causal relationship between the determinants and the tourism demand, the non-causal time-series approach, and the causal econometrics approaches were mostly adopted by researchers before the rapid development of machine learning (Song & Li, 2008).

Time series models utilise historical time series data, such as historical tourist arrivals data, to predict future trends (Peng et al., 2014). Time series models have been frequently used in tourism demand forecasting studies over the past five decades due to their ease of implementation and their reasonable ability to capture historical patterns (Song et al., 2019a). Various models, including Naïve I (Chu, 1998; Martin & Witt, 1989; Witt et al., 1994), Naïve II (Chan et al., 1999; Chu, 1998; Wu et al., 2017), ARIMA (Cho, 2003; Hassani et al., 2017), Exponential Smoothing methods (Cho, 2003; Goh & Law, 2002; Lim & McAleer, 2001a), Holt-Winter (Chu, 1998; Lim & McAleer, 2001a) and many other ones have been adopted in tourism forecasting.

Among the various time series models, the autoregressive integrated moving average (ARIMA) model and its variants are the most prevalent type and account for over 60% of the papers using time series methods (Song et al., 2019a). Chu (2008) uses the ARIMA model to forecast the monthly international tourist arrivals to Singapore. The author reports that the model can achieve relatively lower mean absolute percentage errors (MAPEs) compared with other models.

The seasonal ARIMA (SARIMA) model has been adopted by a tremendous number of studies. Seasonality has been acknowledged as a key feature in tourism demand forecasting due to the nature of tourism activity (Song & Li, 2008). The SARIMA is one of the models accounting for seasonality. The study by (Geurts & Ibrahim, 1975) is one of the earliest attempts of using SARIMA to predict tourism demand. They evaluate the performance of the model by comparing it with the exponentially smoothed model in predicting tourist arrivals to Hawaii. Even though it is not evident the SARIMA model can outperform the other one.

The advantage of the non-causal approach is obvious: it can generate reasonable results without using more than one data series of historical tourism demand. However, due to the forecasting results are not based on economic theories, it can be difficult to analyse tourists' behaviour and measure the effectiveness of their strategies (Peng et al., 2014).

Unlike time series models which attempt to predict tourist arrivals only, *econometric approaches* focus on examining the relationships between tourist arrivals and potential independent variables. Classical regression models (Fourie & Santana-Gallego, 2010; Lee et al., 1996; Uysal & Crompton, 1984), are the most common approaches. Moreover, as a type of dynamic econometric models, the vector autoregressive (VAR) performs quite well for a medium to long term period (Song & Witt, 2006). Besides the two broad approaches, there still are many other important models, such as the Autoregressive Distributed Lag Model (ADLM) model (Maria Santana-Gallego et al., 2016; Song, Li, et al., 2010; Song et al., 2003) and the time varying parameter (TVP) model (Song & Wong, 2003; Witt et al., 2003).

Besides causal and non-causal approaches, the *machine learning* or *artificial intelligence (AI)* methods become more prevalent in the last decade because of the development of machine learning algorithms and the availability of big data. In tourism forecasting, Wang (2004) argues that AI forecasting methods are more likely to provide accurate results than traditional approaches. The artificial neural network (ANN) can deal with the nonlinear relationship between independent variables and tourist arrivals and this makes the ANN the most popular approach in tourism forecasting (Peng et al., 2014). A classic ANN is constructed by three layers: the input layer, the hidden layer and the output layer. Independent variables are put in the input layer; a certain number of neurons are put in the hidden layers and the results are produced in the output layer. Even if the ANN can generate relatively accurate forecasting results, similarly to time-series approaches, ANN cannot report the relationships between those independent variables and tourism demand (Wu, 2010).

Artificial intelligence-based techniques have been widely used to predict different phenomena in a variety of scientific disciplines, even though the 'black box' nature of ANN and other AI-based models limits the theoretical and interpretability of those models (Díaz & Mateu-Sbert, 2011). Apart from ANN, some other models, including the rough set approach, the support vector regression (Chen & Wang, 2007), the fuzzy time series method (Wang, 2004) and the grey theory (Chiang et al., 1997; Wang, 2004) are also adopted in tourism forecasting studies.

Tourism demand literature has paid tremendous attention to increasing the accuracy of the prediction of tourism demand and the forecasting of tourism demand has been well developed in the last decades. The tourism sector, however, does require insights more than just an accurate prediction. Instead of merely being *reactive* to the tourism demand, the tourism sector may be more *proactive* to it. In other words, those stakeholders in the tourism sector should proactively seek to impact the tourism demand. Therefore, a deeper insight into the sociological and psychological mechanisms should be explored.

As discussed above, scholars have attempted to adopt various econometric models to examine the causality between some determinants, such as economic factors, and tourism demand. The literature, however, fails to recognise the psychological links between external stimuli and internal inputs. According to Um and Crompton (1990), before external stimuli can influence tourists' destination decisions, cognitive constructs can integrate those external stimuli, such as destination attributes, with internal inputs, such as values and preferences, to generate attitudes. It is those attitudes, not stimuli or inputs themselves, influence whether or not they go to a certain destination. However, the majority of tourism demand studies merely examine the direct relationship between inputs and tourism demand. Without considering tourists' attitudes towards various determinants, the literature may fail to provide insights into what factors matter the most to tourists when selecting a destination and how those aggregated attitudes influence tourism demand.

The omission of the intrinsic links in tourism forecasting can be attributed to the lack of available data and the challenge of measurement. Most researchers collect attitudinal data by using qualitative methods, such as interviews. For example, Kock et al. (2016) conduct 50 semi-structured interviews to collect destination-specific and salient incorporation that tourists link to Germany and Spain. Moreover, predicting tourism arrivals demand requires aggregated data and thus, it is not common nor practical to collect qualitative data on such a large scale.

The author used an automated machine learning package called H2O AutoML to forecast the tourism demand for the two selected destinations. H2O AutoML was proposed by H2O.ai and based on an open source framework, H2O (LeDell & Poirier, 2020). This novel tool has been used for various research purposes, such as predicting COVID-19 cases (Han et al., 2021; Marques et al., 2021); predicting employee retention (Patro et al., 2021), and predicting customer value in real-time (Fröberg & Rosengren, 2020). The author chose the H2O AutoML package because of three main reasons: *ease of use, accuracy,* and *interpretability*.

Ease of use

H2O AutoML provides an easy-to-use API and this allows researchers in the non-machine learning field to take advantage of state-of-art models (LeDell & Poirier, 2020; Leist et al., 2021). In addition, H2O AutoML is a fully automated tool that provides a workflow from data pre-processing to forecasting models with performance evaluation (LeDell & Poirier, 2020). Researchers can also easily use the exposed parameters to further fine-tune their training models to achieve better results.

Accuracy and validity

H2O AutoML can often achieve better prediction accuracy than other forecasting models because it can operate a random grid search for various state-of-art algorithms with different parameter settings (LeDell & Poirier, 2020). The automated 10-fold cross-validation also ensures the accuracy and generalisability of this model (Cankurt, 2016; Cankurt & SUBAŞI, 2016; Cheng et al., 2018; Feng et al., 2019; Sugiartawan et al., 2018).

In a recent comparison study by Hanussek et al. (2020), the authors argue that most auto machine learning models (AutoML) outperform models developed by a human in primary metrics. H2O AutoML, as one of the AutoML models, performs better than other AutoML models, such as Auto-Keras, Auto-Sklearn, Ludwig, Darwin, and TOPT in binary classification and regression tasks (Truong et al., 2019).

Interpretability

Another important reason why the author adopted H2O AutoML is its great interpretability. As discussed above, tourism demand forecasting should not stop at the evaluation performance stage. The explanation of how endogenous variables influence tourism demand is essential for those stakeholders in the tourism industry. One of the literature gaps that this thesis aims to fill is the non-linear relationship between tourists' attitudes towards a certain aspect and the tourism demand. Therefore, a clear interpretation of what variables influence the forecasting of the proposed model is necessary. H2O AutoML can provide a detailed report of the models using feature importance and other different visualisation method, such as partial dependence plots (LeDell & Poirier, 2020). These reports can help the research explore the linear and non-linear relationship between endogenous variables and tourism demand.

5.3.5 Available models in H2O AutoML

Extreme Gradient Boosting (XGBoost)

Gradient tree boosting, also known as gradient boosting machine (GBM) or gradient boosted regression tree (GBRT), is a prevalent technique among machine learning algorithms (Friedman, 2001). Chen and Guestrin (2016) made improvements to the traditional gradient tree boosting to achieve faster speed and fewer resources requirement. Given a dataset with *n* examples and *m* features, a tree ensemble model using K additive functions is shown as,

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)\,,\ f_k \in F,$$

Where *F* is the space of the regression tree (CART). Unlike decision trees, each regression tree has a continuous value on each leaf. The main goal of this algorithm is to minimise the following function,

$$L(\phi) = \sum_i \{l(\hat{y}_i, y_i\}) + \sum_k \{\Omega(f_k\})$$

where l is the loss function measuring the difference between the predicted $\hat{y}_i$ and the actual y_i. The Ω is the penalisation function of the complexity of the model.

Compared with non-causal time-series and ANN models, XGBoost has an advantage that can generate the importance for each independent variable. This feature brings similar interpretability to traditional econometrics models.

Random Forest (RF)

Random Forest (RF) is a bagging procedure that adopts a combination of tree predictors to vote for the most popular class or calculate the mean prediction based on individual trees. These tree predictors are based on the value of a random vector with the same distribution for all trees in the forest (Breiman, 1996, 2001). The definition by Breiman (2001, p. 6) is as follows:

A random forest is a classifier consisting of a collection of tree-structured classifier $h(x, \theta_k)$ *where* θ_k *are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input* x.

Five major advantages of the random forest algorithm attract researchers. First, random forest overcomes the overfitting issue because of the Law of Large Numbers (Breiman, 2001; Zhou et al., 2021). Second, this algorithm can achieve accuracy for high dimensional data with a nonlinear relationship because it can learn complex interactions between input features (Friedman et al., 2001; Hartmann et al., 2019). Third, it performs well with noise and outliers (Breiman, 2001). Fourth, random forest has great scalability due to the increase in training time being linear (Hartmann et al., 2019). Finally, the tuning of model hyperparameters is easy (Breiman, 2001).

Because of the five advantages of the model, random forest classifier and regressor have been commonly used in various research fields, such as consumer behaviour (Lismont et al., 2018; Zhou et al., 2021); marketing effectiveness (Chen et al., 2020); retailing (Hirche et al., 2021); and finance (Krauss et al., 2017). Many researchers have also used this algorithm to forecast tourism demand (Brida et al., 2018; Hu et al., 2021; Li, Li, et al., 2021; A. Liu, L. Vici, et al., 2021).

Deep Neural Networks (DNN)

Deep neural networks (DNN) are neural networks composed of multiple non-linear hidden layers between input and output layers (Bengio, 2009). DNN adopts representation learning that automatically extracts representation from raw data for further classification or regression (LeCun et al., 2015). This approach has performed greatly in various fields, such as speech recognition and visual object recognition (LeCun et al., 2015).

It has the following three major advantages compared with other machine learning algorithms. DNN can identify intricate non-linear relationships and structures in high-dimensional data (Law et al., 2019; LeCun et al., 2015; Pouyanfar et al., 2018; Shi et al., 2017). Neural networks, as the universe approximation theorem claims, can accurately approximate any non-linear functions with reasonable network size (Hornik et al., 1989). Deep neural networks can learn non-linear relationships better than shallow neural networks because DNN can learn the non-linear function of features while shallow neural networks can only learn the relationship of the inputs (Shi et al., 2017).

The capacity of learning non-linear relationships leads to the second advantage of DNN which is less dependency on domain expertise (Law et al., 2019). For a complex system, such as tourism, understanding the intricate relationship between countless components in the system is almost impossible. Thus, an algorithm with the ability to automatically construct suitable features does help researchers and practitioners make reasonable predictions and explore the non-linear relationship between variables.

The third advantage of DNN involves its competence to learn shared uncertainties (Shi et al., 2017). External influential factors may not evenly impact the samples. For example, the price of a destination may have a negative effect on the tourist demand for the destination. The magnitude of this effect varies from person to person as each person has a different sensitivity to price. Deep neural networks are good at capturing those uneven relationships because they can learn load features hierarchically (Shi et al., 2017).

DNN has been commonly adopted in tourism demand forecasting and other business research. For example, Law et al. (2019) use Recurrent Neural Network (RNN), which is a prevalent deep network architecture to predict monthly tourist arrivals to Macau. Long Short-Term Memory networks (LSTM) is a variant of RNN (Hochreiter & Schmidhuber, 1997) and has been used by Ma et al. (2018) to extract features from photos so they can explore how those photos can influence hotel review helpfulness.

Generalise Linear Models (GLM)

Generalised Linear models refer to an extension of the linear regression estimation. The GLM suit in H2O AutoML includes Gaussian regression, Poisson regression, Fractional binomial regression, Quasibinomial regression, Gamma regression, Ordinal regression, Negative Binomial regression and Tweedie distribution (LeDell & Poirier, 2020). H2O AutoML uses a list of alpha values to build a lambda_search enabled GLM model and returns the best performing model by maximising the log-likelihood. H2O's GLM can be estimated by maximising the likelihood optimisation below:

$$max_{\beta,\beta_0} = (GLMLog - likelihood - RegularisationPenalty)$$

where the regularisation penalty can reduce the variance in the predictions and increase the model interpretability (Nykodym et al., 2016). The penalty is defined as:

$$\lambda P_\alpha(\beta) = \lambda\left(\alpha\beta.|_1 + \frac{1}{2}(1-\alpha)\beta.|_2^2\right)$$

GLM has three major advantages over simple OLS linear regression. First, it is flexible because it relaxes the assumption of the OLS regression that the dependent variable has to be normally distributed (Nykodym et al., 2016). Thus, GLM can unify traditional regression models, such as linear regression and logistic regression and this attracts researchers (Santos, 2016). Another two advantages, according to Nykodym et al. (2016), are the availability of model-fitting software and GLM's scalability with large datasets.

GLM has been adopted in various literature, such as retailing, supply chain, marketing and tourism. Its most common function is to explore the relationship between independent variables and dependent variables. In tourism, the relationship covers two main aspects: the relationship between tourists/destination characteristics and their behaviour; the externalities of tourists' behaviour. Weigert et al. (2021)use GLM to explore how age, period and cohort can influence tourists' travel distances. In a study by Suni and Pesonen (2019), GLM is used to examine influential factors for hunting tourism frequency.

The externalities of tourists' behaviour refer to the impacts of tourism behaviour on the elements in the destination, such as the environment or animals. Argüelles et al. (2016) investigate how whale-watching tourism influences whales' behaviour by using GLM. Gabarda-Mallorquí et al. (2017) adopt GLM to explore the impacts of mass tourism on water efficiency at the destination.

The use of GLM in tourism demand forecasting is still rare in tourism literature (Asrin et al., 2015; Santos, 2016; Semeida, 2014). An example is a study by Santos (2016). He finds that GLM performs better than the simpler log-linear OLS regression when forecasting tourists' length of stay to 100 Brazilian destinations.

5.3.6 Baseline models for evaluation

Fast Fourier Transform

The Fast Fourier Transform (FFT) is an algorithm to compute the discrete Fourier transform (DFT) on a finite amount of data (Heideman et al., 1984). It was proposed by Cooley and Tukey (1965) and contains a comprehensive treatment of discrete Fourier transforms and their implementation. This algorithm is one of the most commonly used algorithms for signal processing (Rockmore, 2000). It is one of the most useful and widely used tools in applied mathematics, and some consider it to be among the greatest achievements of theoretical computer science, comparable to the invention of the wheel (Averbuch et al., 2006; Bell, 2012).

In tourism demand forecasting, however, the adoption of this method is not as common as other techniques, such as ARIMA. Tourism demand proxies, such as tourist arrivals, can be approximated by a function that can be approximated by a discrete Fourier series. Multiple scholars have adopted the Fourier transform in their tourism demand modelling (Hu, 2021; Kožić, 2014; Ridderstaat et al., 2014; C. Zhang et al., 2021). FFT can produce the function and the processing speed is faster than traditional DFT. In addition, seasonality usually exists in tourism demand time series and the FFT algorithm is able to find repeating cycles in time series data (Fernández-Morales & Cisneros-Martínez, 2019; Silva et al., 2019; Vatsa, 2021). Thus, it is reasonable to include FFT as one of the benchmarks in this project to forecast tourism demand.

Facebook Prophet

The Facebook Prophet algorithm is a state-of-the-art open-source time series forecasting algorithm developed by Facebook's Core Data Science team. Taylor and Letham (2018) argue that this algorithm is accurate, fast, fully automatic, and tunable. It uses an additive model in order to forecast time series data based on non-linear trends and seasonal effects. It's best used with time series that have strong seasonal effects, such as tourism demand (Zunic et al., 2020).

Temporal Convolutional Network

A Temporal Convolutional Network (TCN) is a "feed-forward"-type deep neural network that operates on time series with nonlinearity. Lea et al. (2016) propose this model for video-based action segmentation and then use it for time series forecasting. Comparing with other Recurrent and Convolutional Networks, Bai et al. (2018) outline three characteristics of TCN. First, the causal convolutions in the architecture ensure that there is no information “leakage” from the future to the past. Second, a sequence can be mapped to an output sequence with the same length, just as with an RNN.

TCNs have been shown to have strong performance for spatial domains and have recently been used to forecast time series in various disciplines. Wan et al. (2019)adopt TCN to forecast high-dimensional and diverse time series of Beijing PM2.5 and ISO-NE Datasets. TCN achieve over 25% improvement over benchmarking models. H. Zhang et al. (2021) propose a TCN-based hybrid model to predict short-term electricity prices and report a superior performance over other time series models. Wang et al. (2019) use the TCN to predict three high-dimensional datasets, including purchase and redemption of capital flow, Beijing PM2.5 and SML2010. They also claim that the TCN outperforms other models.

Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) is a special type of recurrent neural network (RNN) architecture that is able to learn long-term dependencies and representation over streams of data such as sentences (Hochreiter & Schmidhuber, 1997). This algorithm can distinguish long and short data inputs and state outputs by using several gates to capture the long states from the beginning unit and the shorter states from the last unit (Law et al., 2019).

LSTM and its variants have been widely adopted in tourism demand literature, beating the previous state-of-the-art models by a large margin with its simple, but insightful architecture. Kulshrestha et al. (2020) proposed a deep learning model based on LSTM to forecast the tourist arrivals to Singapore from five countries over 24 years. They report that their proposed model outperforms other benchmark models, including Support Vector Regression, Neural Network and Autoregressive Distributed Lag Model (ADLM). Polyzos et al. (2020) use an LSTM approach to explore the effects of the COVID-19 pandemic on tourist arrivals to China and conclude that the tourist arrivals will take one year and six months to return to the previous period. LSTM is also used for high-frequency tourism demand forecasting. Zheng et al. (2021) forecast hourly tourism demand for multiple attractions using the LSTM variant which significantly outperforms SARIMAX and Neural Network.

Performance metrics

We used two popular error metrics, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to compare the performance of the proposed model and the baseline models.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|$$

$$\frac{MAPE = \frac{1}{n}\sum_{i=1}^{n} \quad |y_i - \widehat{y_i}|}{y_i}$$

5.3.7 Interpretation

Variable Importance

One of the most important tasks in machine learning is to understand what features are important for making predictions. This thesis uses the variable importance plot to identify the variables which can influence the tourism demand the most. The H2O AutoML architecture uses different algorithms to calculate the variable importance for different models. H2O.AI (2020) explains the calculation as below.

For tree-based models, such as GBM and Distributed Random Forest, the variable importance is a native function. The H2O package scales it between 0 and 1. The calculation is dependent on what variable was selected to split on, and how much the squared error overall trees change as a result (LeDell & Poirier, 2020). The squared error for each individual node is the decrease in the variance of the response within that node, and it can be calculated by the formula below:

$$Squared\ Error = \ VAR \ \times N$$

Where VAR represents the variance calculated as:

$$VAR = \frac{1}{N}\sum_{i=0}^{N} \quad \left(y_i - \underline{y}\right)^2$$

Therefore, the Squared Error formular can be written as:

$$SquaredError = \left[\sum_{i=0}^{N} \frac{y_{i^2}}{N} - \underline{y}^2 \right] \times N$$

Apart from the single decision tree model, H2O AutoML also utilises tree-based ensemble methods to increase the architecture's performance. For this type of model, H2O compute the variable importance using the gains of the respective loss functions of each tree during their construction (H2O.AI, 2020).

For non-tree-based models, H2O adopts a different approach to computing the feature importance. There are two major non-tree-based models in the H2O architecture: deep learning and GLM. The feature importance of deep learning models adopts the approach proposed by Gedeon (1997). Gedeon (1997) uses sensitivity analysis to find the effect of input on the output of a neural network. Specifically, the importance of an input variable x_i on an output variable y_k depends on the calculation of the derivative as shown below:

$$\frac{dy_k}{dx_i} = f'(U_{k2}) \cdot f'\left(U_{j1}\right) \cdot \sum_{j} w_{jk} \cdot w_{ij}$$

Calculating this importance for each input and output variable will generate a weight matrix of the trained neural network.

For GLM models, the variable importance represents the normalised coefficient magnitudes by default (H2O.AI, 2020). The normalisation makes the comparison between variables easy and can be disabled by the user.

Variable importance is a common tool used by tourism research. Kuzey et al. (2019), by conducting a sensitivity analysis, report that the top three influential factors for the inbound tourist arrivals to Turkey are trade, the distance between origin countries and Turkey and whether origin countries and Turkey are in the same region. Brida et al. (2018) proposed a non-linear approximation of the total expenditure distribution of cruise tourists and compute the variable importance for each input variable. Researchers also use variable importance to select features. Shen et al. (2019) propose a hybrid model to forecast monthly tourist arrivals to Singapore. They use the variable importance tool to select the most significant variables and repeat the process until they obtain the required number of input variables.

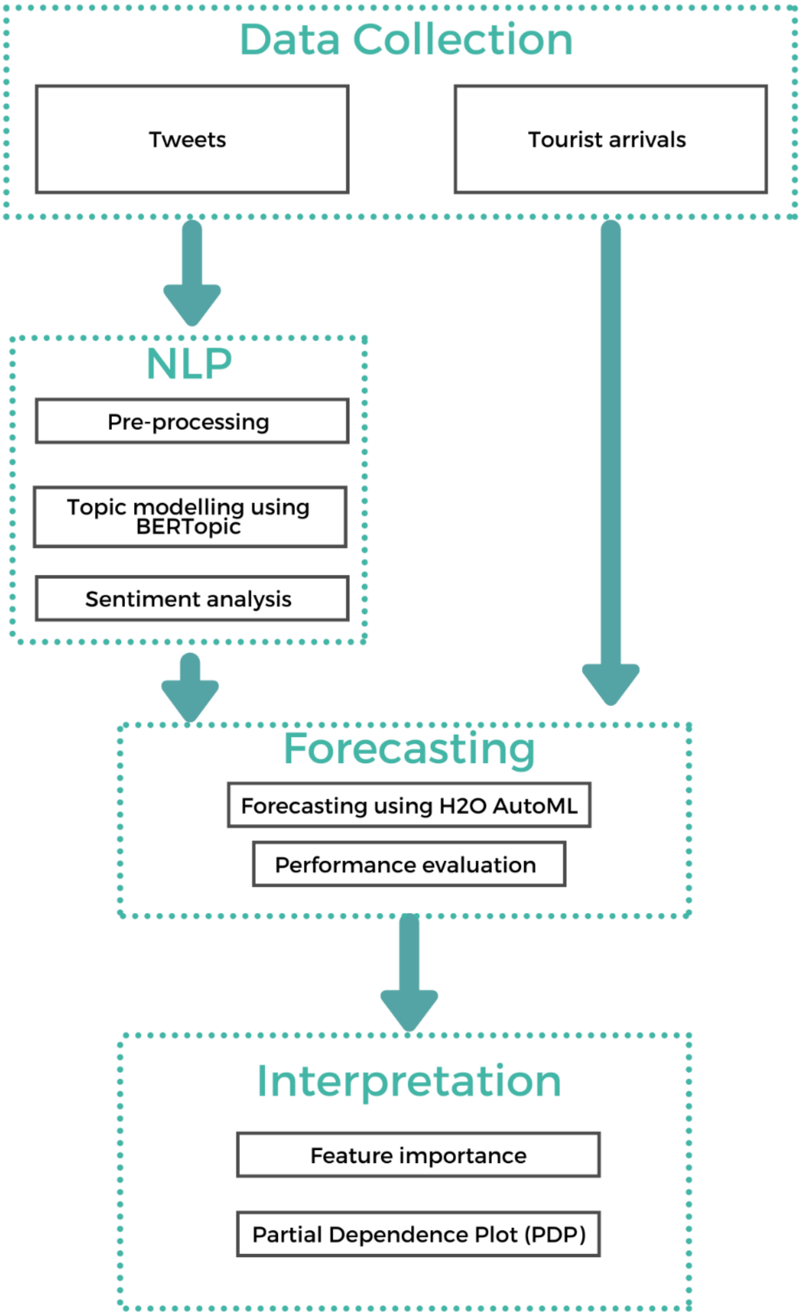
Partial Dependence Plot (PDP)

This thesis uses Partial Dependence Plot (PDP) to explore the specific relationship between input variables and tourist arrivals to Sydney and London. PDP is a method to derive a visualisation on the marginal effect of features on the forecasting result (H2O.AI, 2020). The partial dependence plot (PDP) is useful for visualizing pairwise interactions among variables in non-linear models that arise when using deep neural networks (Inglis et al., 2021). The algorithm is able to analyse linear, monotonous and other complex relationships between input and output variables, which can be used to interpret tourism demand (Chen et al., 2021). For example, it can be used to study how holidays and airline ticket prices influence tourist arrivals to a particular destination.

PDP is still a new tool but recent studies have started to use PDP to explore the fine-grained relationship between independent and dependent variables. By using PDP plot, Chen et al. (2021) figure the linear positive effect of the boutique order number on tourists' online purchases. They also report many non-linear relationships between behavioural variables and actual purchases. For example, the positive impacts of ratings on purchase behaviour stop when the ratings increase to a certain level. Alsaleh and Farooq (2021) adopt PDP to analyse how various factors can influence the trip production level in Belleville, Canada.

5.3.8 Methods summary

The methods used in this study is summarised in the figure below:


Data Collection
Tweets
Tourist arrivals
NLP
Pre-processing
Topic modelling using BERTopic
Sentiment analysis
Forecasting
Forecasting using H2O AutoML
Performance evaluation
Interpretation
Feature importance
Partial Dependence Plot (PDP)

Chapter 6 Study 1: Sydney

6.1 Data collection

The author chose Twitter as the platform to collect textual social media data for Sydney. Twitter, instead of other social media platform, was selected because Twitter has been widely used for marketing and tourism (Bigné et al., 2019). Companies and businesses also leverage Twitter to identify potential customers and understand their demands (Park et al., 2016). The widespread use of Twitter in these domains implies a rich source of data pertinent to this study's focus on tourism demand.

This study used a twitter scraping tool, called 'Twitterscraper', to crawl the tweets containing the name of the two cities. Twitterscraper uses a command prompt to crawl desired tweets by given keywords; the number of tweets; the start date; the end date; the language; and the output format (Taspinar & Schuirmann, 2017).

The candidate collect only tweets written in English because of three main reasons: First, English is one of the most widely spoken languages worldwide and serves as the de facto language of international communication. This is particularly true for social media platforms like Twitter, where a significant proportion of content is in English. By focusing on English, the study can tap into a broad, global discussion. Second, Natural Language Processing (NLP), the methodological approach of this study, has a heavy emphasis on English due to the language's global ubiquity. Most of the well-established, sophisticated algorithms and pre-trained models are designed and optimized for English, which ensures better accuracy and reliability in the study's analysis. Finally, By concentrating on English-language tweets, the study maintains consistency in the dataset, minimizing potential discrepancies or misinterpretations that could arise from language translation or cultural nuances.

After we collected all the data, we randomly sampled 10,000 tweets for each year and 100,000 tweets in total to speed up the followed NLP modelling and release the computation pressure. This is grounded in practical, computational and statistical considerations. Firstly, advanced NLP tasks, such as BERTopic, can be resource-intensive, requiring significant computational power and storage. By limiting the analysis to a manageable subset of tweets, the research ensures a faster and smoother processing, reducing the computational burden. This approach is crucial, especially when using complex models like BERTopic for topic modelling and sentiment analysis. Secondly, A well-conducted random sample of 10,000 tweets per year is often large enough to capture the diversity and variability in public sentiment and topics about the destinations of interest. Given the massive amount of tweets generated daily, this sizeable sample can adequately represent the broader conversation while providing a manageable dataset for analysis. Thirdly, the focus on a smaller dataset allows the research to prioritizse data quality. By managing a smaller dataset, researchers can devote more time to refining the data cleaning and preprocessing steps, which are critical for NLP applications. It ensures a high-quality dataset, leading to more accurate and reliable results.

The annual data of tourist arrivals in Sydney across ten years are collected from the report by Destination NSW, a lead Government agency for tourism and primary events sector (Destination NSW, 2018, 2019, 2021).

6.2 Topics Identified by BERTopic

The BERTopic generates eighty topics with keywords in total. The author categorises those topics according to the categorisation in Chapter 4.3. The table lists the keywords relating to each dimension. These words can describe the meaning and characteristics of each topic (Grootendorst, 2020). The author interpreted using those keywords. For example, based on the keywords for one of the topics are *flights; sale airfare; fly Sydney; fares,* this topic is more likely to be relevant to flight fare to Sydney. After interpreting all the topics based on their keywords, the author categorised those topics into ten groups, namely *economic, exogenous, internal, attraction, geographic factors, tourism facilities, events and activities,* food and *symbolic factors*.

The economic factors, as presented in Table 1, mainly refer to the sentiment of social media posts regarding flight ticket prices to Sydney. For example, "*888$ round trip flight to Sydney u down*" and "*Brisbane's Q3 2018 office sales worth $1.2b are now second only to Sydney. #brisbane #commercialproperty #commercialrealestate*". Travel costs that occur during the movement of tourists from their origin to a destination may influence the tourism demand for the destination. Economic theory postulates that when other conditions remain, as travel costs increase, the demand for visiting would decrease.

The type of exogenous factors is one of the fundamental components of non-economic influential factors for tourism demand and it involves the general perception of a destination (Uysal, 1998). For Sydney, results show that *general news, employment, real estate industry, traffic, architecture, pet-loving* and *politics* are the seven most discussed exogenous topics. For instance, "*Sydney man to face court in January accused of sexually assaulting girl at Tura Beach*". Exogenous factors may be one of the most studied types of drives of tourism demand.

The other component of non-economic influential factors of tourism demand is social-psychological factors. These factors consist of internal and external factors. Internal factors refer to the values, motives or feelings of tourists (Um & Crompton, 1990). BERTopic identifies one internal factor regarding people's sleeping in Sydney: "*Its my bestfriend birthdaym and im not with him*"; "*should i go to london or munich for my birthday this year?*". For external factors, significative and symbolic stimulus are the two most important types of topics discussed regarding Sydney. Significative stimuli are the attributes of a destination, such as attractions and geographic factors (Um & Crompton, 1990). This study identifies six most discussed attractions in Sydney. They are Sydney opera, Biennale of Sydney, Sydney bridge, Sydney Zoo, Royal Botanic Garden and Sydney fish market.

As a destination significantly impacted by weather and climate, it is important to explore the effects of geographic factors on Sydney's tourism demand. This topic modelling algorithm extracts two main topics: climate and beach. A tweet reporting weather or climate notes, "*Sydney ATIS H: 182308Z (10:08 AM AEST). Temperature: 20 QNH: 1017*." Another tweet mentioned the beach, "*Jaded by routine? This week, grab fish and chips and head to the beach*."

Another type of significant factors discussed on social media is tourism facilities in Sydney. Tourism facilities or infrastructure can support tourists' experience, leading to tourists' satisfaction. This category covers various areas of facilities tourists can enjoy during their holiday, including accommodation, airport, hospital, public transport, gym and Wi-Fi access. For example, "*I have been going INSANE WITH NO INTERNET Loving Sydney airport wifi Yayyyy*"; "*come to the gym with me*".

As another type of significative factor, the discussion on different events and activities in Sydney is one of the most popular topics. This category involves numerous events, such as concerts, conferences, festivals, shows, and sports events. For example, "*Ok here's the acer arena seating plan for justin biebers Sydney concert in 2011; I'm on the floor- section J*".

"*Fantastic shows Sat at The Ivy and Chinese Laundry in Sydney.*" Shows and events are important pull factors. As an essential part of significative factors, food experience was greatly discussed on social media. Two topics identified regards food and drinks, respectively.

Finally, BERTopic also identifies the discussion of multiple symbolic factors. Symbolic factors mainly refer to the marketing activities on social media (Um & Crompton, 1990). For example, the topic keywords "*wallaby; wallaby way; 42 wallaby; calle wallaby*" refer to the discussion on the address where, in the movie "*Finding Nemo*", Nemo was found by his father. This address has been one of the symbols of Sydney and many people visit Sydney to find this address. For example, "*I can't wait to hear our kids roast each other at the carne asadas (the name of a restaurant) every Friday at P. Sherman, 42 Wallaby Way, Sydney*".

Table 1. Categories with topics and keywords (Sydney)

Category	Topics with important keywords
Economic	*flights; sale airfare; fly Sydney; fares*
Exogenous	*sydney; sydney morning; sydney news; news aus*
	jobs Sydney; job Australia; sydney jobs; job jobs
	sydney property; real estate; house prices; property market
	sydney traffic; trafficnetwork; sydtraffic trafficnetwork; trafficnetwork sydneys
	Building; sydney architecture; building Sydney; house sydney
	sydney cat; dog Sydney; pets; dog lovers
	Iraq; Syria; iraq Sydney; syria world
	asian shares; sydney reuters; reuters Asian; betting china
	sydney Anglicans; archbishop Sydney; sydney Anglican; church sydney
	australian shares; australia shares; trading Sydney; australian stocks
	Cloud; aws; sydney cloud; cloudcentral
	aud asx; open aud; market open; sydney stock
	gay marriage; gay sydney; sydney gay; marriage sydney
Internal	*sleeps; sydney sleep; sleep sydney; rough sleepers*
	opera house; sydney opera; house sydney; opera sydney

Attraction	*biennale; biennale sydney; sydney art; art sydney*
	harbour bridge; bridge sydney; bridge climb; sydney bridge
	zoo; taronga zoo; zoo sydney; sydney zoo
	gardens; botanic gardens; botanical gardens; gardens sydney
	sydney fish; fish market; fish markets; sydney salmon
Geographic	*raining sydney; sydney rain; rain sydney; sydney raining*
	bondi beach; beach sydney; bondi sydney; sydney bondi
	kmh humidity; humidity; clouds wind; sydney weather
	atis; sydney atis; aest temperature; overhead sydney
	temp; temp increased; australia current; current temp
Tourism facilities	*hotels; hotel sydney; hotels sydney; sydney hotels*
	zayn sydney; sydney airport; airport sydney; zayn airport
	hospital sydney; medical; hospitals; health proposal
	qantas; london flights; direct sydney; qantas plan
	denpasarsydney delay; indonesia denpasarsydney; weather forecast; challenge bowling
	gym; fitness sydney; sydney gym; personal trainer
	telstra; free wifi; law twitter; telstra lays
Events and activities	*concert; sydney festival; sydney concert; concert sydney*
	conference sydney; workshop; sydney university; seminar
	new years; christmas; year sydney; sydney christmas
	fires; bushfire; reduction fires; bushfires
	shows; sydney shows; sydney; shows sydney
	fireworks; sydney fireworks; fireworks sydney; fireworks tonight
	sydney roosters; nrl; roosters vs; roosters sydney
	sydney fashion; fashion week; fashion sydney; fashion festival
	sydney writers; writers festival; books; sydney poetry
	film festival; sydney film; cinema sydney; film sydney
	swans; sydney swans; afl sydney; swans afl
	sydney surf; swimming sydney; city surf; pool sydney

	massage; salon; pamper package; clinic
	cycling sydney; sydney cyclist; bikelane skeptic; sydney cycling
	sydney wedding; wedding photographer; wedding sydney; wedding photography
	sydney dance; dancer sydney; dancing sydney; dance sydney
	justin bieber; meets justin; sydney meets; sydney bieber
	royal easter; sydney easter; easter sydney; easter happy
	ufc; sydney ufc; ufc sydney; ufc brisbane
	sydney olympic; olympic games; sydney olympics; 2000 olympics
	racing; jockey; sydney horse; race finish
	motorshow; motor motorshow; sydneymotorshow; motorshow sydneymotorshow
	party sydney; sydney party; party tonight; carry partying
Food	*sydney; sydneys; breakfast sydney; sydney food*
	coffee sydney; sydney coffee; morning coffee; cafe sydney
Symbolic	*wallaby; wallaby way; 42 wallaby; calle wallaby*
	tickets sydney; sydney tickets; tickets sale; ticket sydney
	liked video; video sydney; sydney liked; playlist sydney
	trending sydney; backstagepingpong trending; sydney torrential; sydney centrist
	tour sydney; touring sydney; sydney tour; australian tour
	social media; socialmedia; socialmedia crisis; blogthe social

6.3 Results for time series analysis

Figure 1 below shows the trend of the overall sentiments towards Sydney on Twitter from 2008 to 2019. The figure shows that the overall sentiment shown in tweets regarding Sydney kept decreasing until 2014 before it bounced back. The dip in 2014 can be associated with an event called Sydney Siege which took place on December 15th 2014 (Wendland et al., 2018). This siege lasted for 16 hours and ended with the death of three people (The Guardian, 2014). For example, "*Police did not act fast enough to deadly Sydney siege*".

Table 2 presents the descriptive summary of the tourist arrivals to Sydney and the sentiment change during the ten years. Results show that the average sentiment scores for all topics are positive. The minimum sentiment score is 1.09 in 2014 and the maximum sentiment, occurred in 2018, is 1.21.

Figure 1. Overall sentiment for Sydney from 2008 to 2019

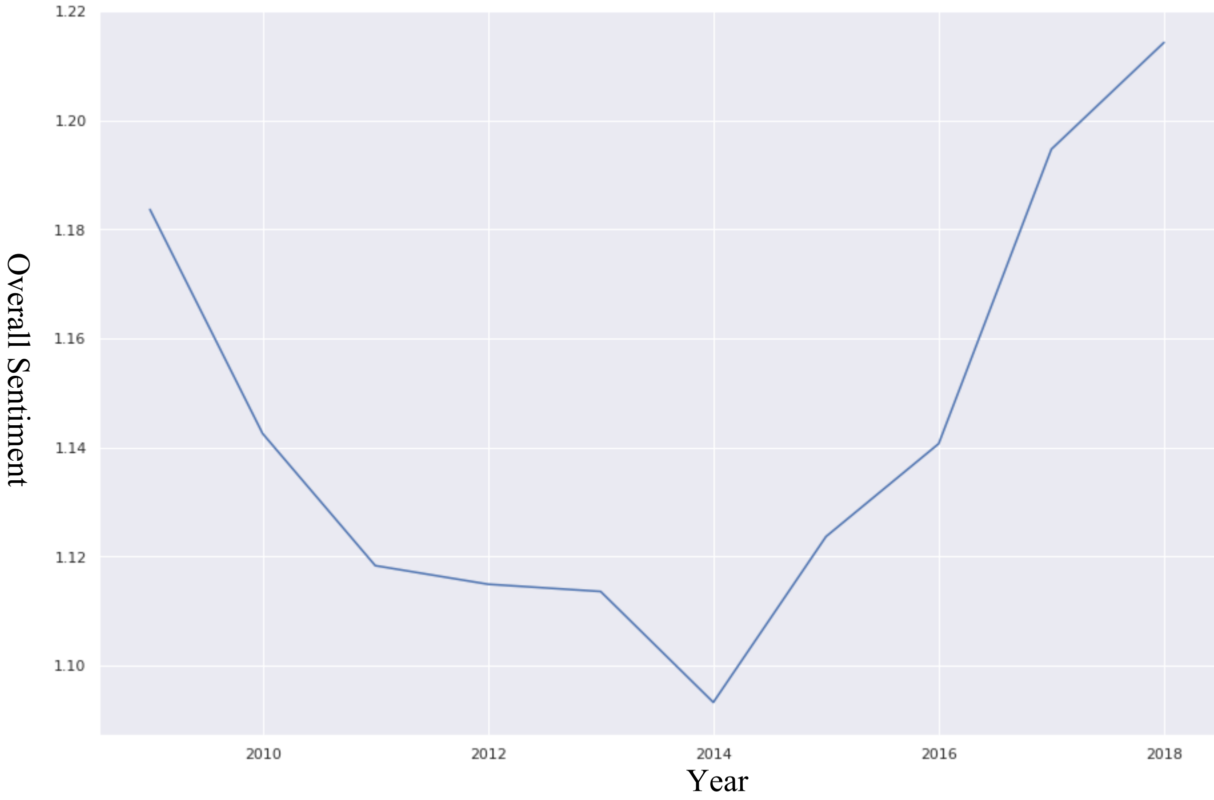


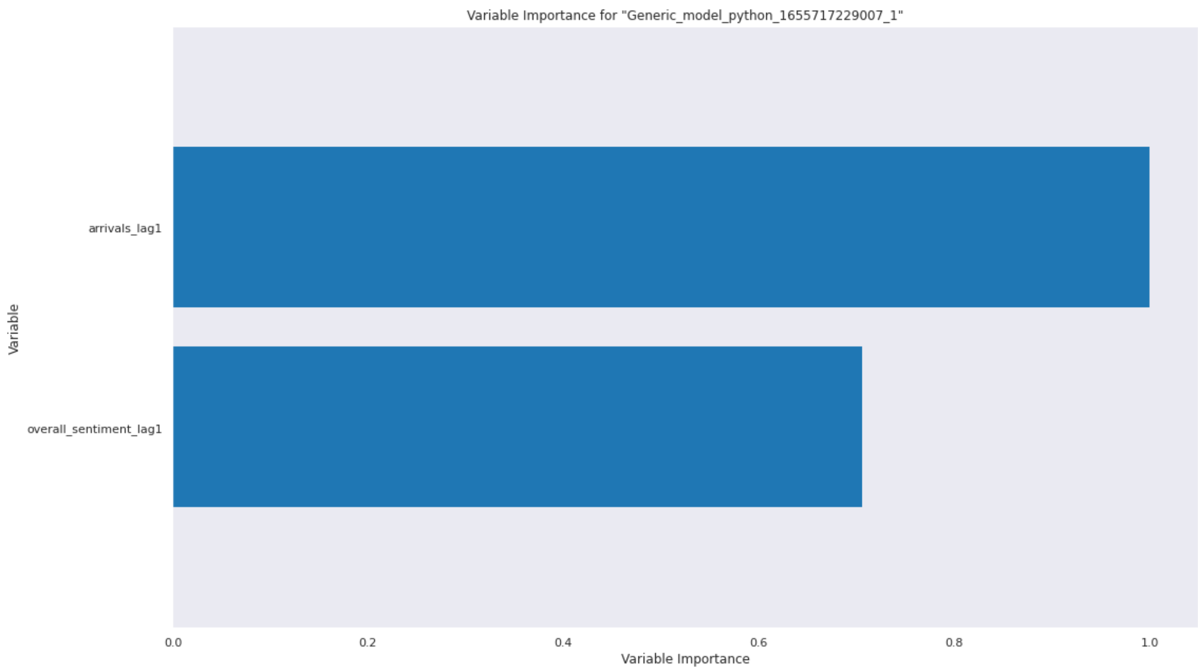
Table 2. Descriptive summary of Sydney arrivals and overall sentiment

	Arrivals	Overall Sentiment
Mean	30715.6	1.143
Std	3254.78	0.04
Min	25048.0	1.093
25%	29207.0	1.12
50%	30127.5	1.13
75%	32500.75	1.173
Max	36675.0	1.21

6.3.1 Preliminary study (overall sentiment vs tourist arrivals to Sydney)

Figure 2 below indicates the five most important variables in the forecasting model. It illustrates that the overall sentiment one year ago and the average sentiment in the recent two years are both influential for the tourist arrivals to Sydney.

Figure 2. Variable importance for Sydney (overall sentiments)



A Partial dependence plot (PDP) provides a visual interpretation of the influence of an explanatory variable on the dependent variable. The effect of a variable is measured in the change in the mean response. Figure 3 and Figure 4 demonstrate how the arrivals to Sydney one year before and the overall sentiment one year ago correlate with the dependent variable, tourist arrivals. Figure 3 shows a clear positive impacts of the arrivals one year ago on the tourist arrivals in the particular year. This rationalises the use of pure time-series methods for tourism demand forecasting (Cho, 2003; Chu, 1998; Goh & Law, 2002; Martin & Witt, 1989; Song et al., 2019b; Wu et al., 2017). The effects of the overall sentiments at lag (1), according to Figure 4, also confirm the positive effect of sentiments on tourist arrivals, even though the effect is limited when the sentiment is too positive.

This finding aligns with previous literature that positive sentiment before people visit a destination can mean a positive a priori image of the destination. According to destination image literature, destination image can positively influence tourists' perceived quality and satisfaction. Therefore, the increase in perceived quality and satisfaction can lead to more revisits or positive WOM, which further increases the tourism demand for the destination. Most empirical studies also confirm this effect (Lee, 2009; McKercher & Tse, 2012). Lee (2009) conducted an empirical study on the tourists to two destinations in Taiwan and confirms that both tourists' destination image and attitude can directly affect tourists' satisfaction and indirectly affects tourists' behaviour. Machado (2010) analyses the relationship between the destination image of Madeira Island and tourists' length of stay, and agrees that the increase in the destination image can increase the length of stay.

Figure 3. Partial Dependence plot for arrivals at lag (1)

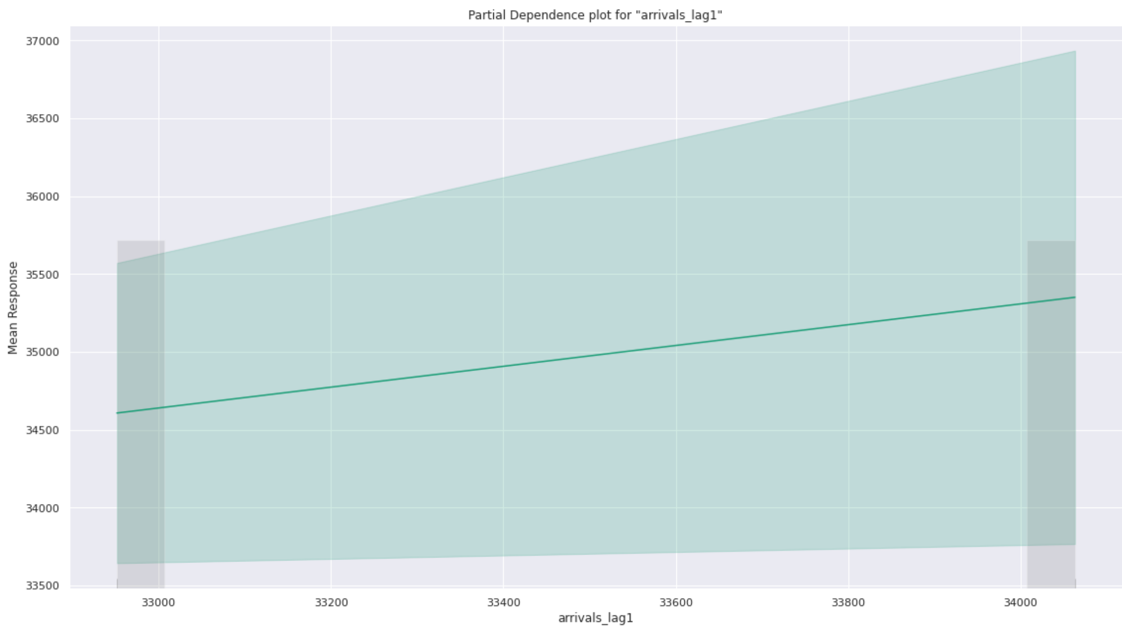
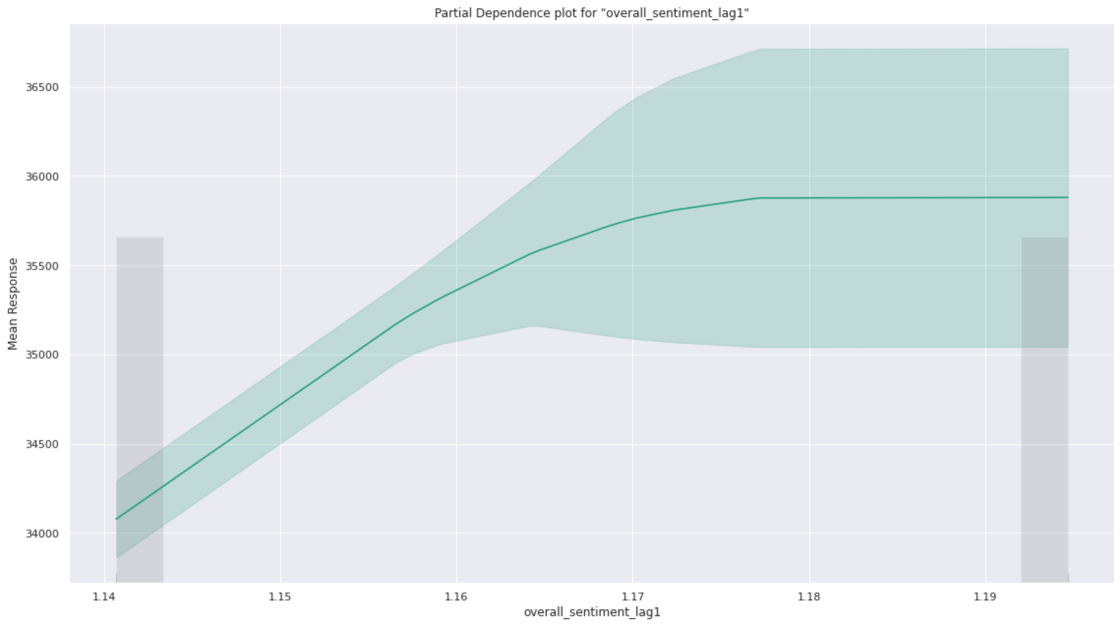


Figure 4. Partial Dependence plot for overall sentiment at lag (1)



6.3.2 Results (topical sentiments vs arrivals)

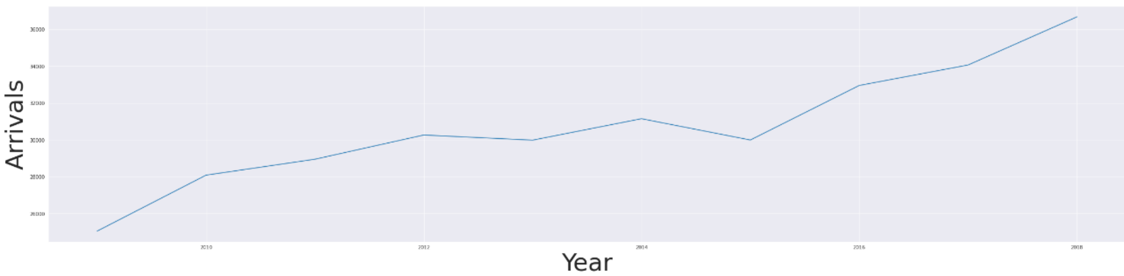
The sentiment scores for each topic from 2009 to 2018 were calculated by the formula discussed in the methodology section. Table 3 summarises the mean, standard deviation, minimum value, and maximum value for each category of topics. As per Table 3, social media posts displayed the highest positivity when discussing Sydney's attractions and the least positivity when referencing social factors.

Table 3. Descriptive summary of sentiment scores of the topics

	Attraction	**Events and activities**	**Exogenous**	**Food**	**Geographic**	**Social**	**Symbolic**	**Tourism facilities**	**Travel costs**
mean	1.23	1.27	1.07	1.3	1.08	0.93	1.22	1.15	1.01
std	0.08	0.07	0.04	0.07	0.06	0.13	0.05	0.08	0.12
min	1.03	1.16	1.03	1.16	0.99	0.68	1.14	1.05	0.75
25%	1.21	1.22	1.04	1.26	1.02	0.85	1.2	1.1	0.94
50%	1.25	1.26	1.06	1.3	1.09	0.91	1.23	1.13	1.05
75%	1.28	1.31	1.09	1.32	1.13	1.05	1.25	1.22	1.1
max	1.3	1.37	1.15	1.42	1.16	1.07	1.33	1.26	1.15

Figure 5 shows the annual tourist arrivals to Sydney from 2009 to 2018. The figure shows a clear increasing trend, except for a slight drop in the year 2015. The original time series was split into training (80%) and test sets (20%) for training and performance evaluation purposes.

Figure 5. Tourist arrivals (annually) to Sydney

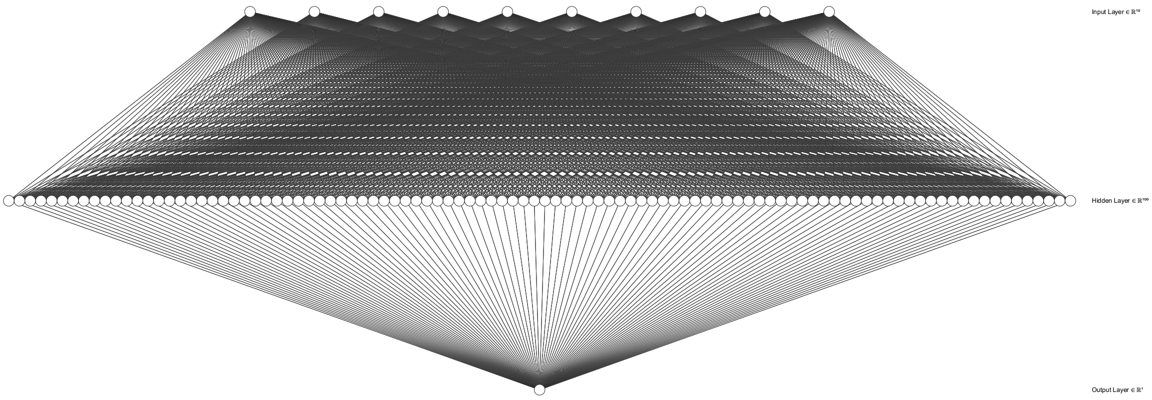


As described in the methodology section, H2O AutoML framework trains different models with a grid search of different parameters (LeDell & Poirier, 2020). After training all possible models and parameters, H2O AutoML reports the performance for a different combination of the models and parameters. Table 4 shows the leaderboard of the H2O AutoML models based on their prediction error on test data. The results show that the leading model with the name "DeepLearning_grid_1_AutoML_2_20220525_85253_model_114" outperforms other models by magnitudes based on their RMSE and MAE. Figure 6 shows the structure of the neural network for the leading model. This neural network is constructed with one input layer (10 input neurons), one hidden layer (100 hidden neurons) and one output layer (1 output neuron which is the variable tourist arrivals).

Table 4. Leader board of best performing models for Sydney

Model	RMSE	MAE
* DeepLearning_grid_1_AutoML_2_20220525_85253_model_114	2197.47	1593.71
DeepLearning_grid_3_AutoML_2_20220525_85253_model_25	2339.34	2235.71
DeepLearning_grid_2_AutoML_2_20220525_85253_model_11	2479.59	2038.13
DeepLearning_grid_1_AutoML_2_20220525_85253_model_341	2682.59	1967.63
DeepLearning_grid_1_AutoML_2_20220525_85253_model_63	2764.65	2144.91

Figure 6. Diagram for the leading model



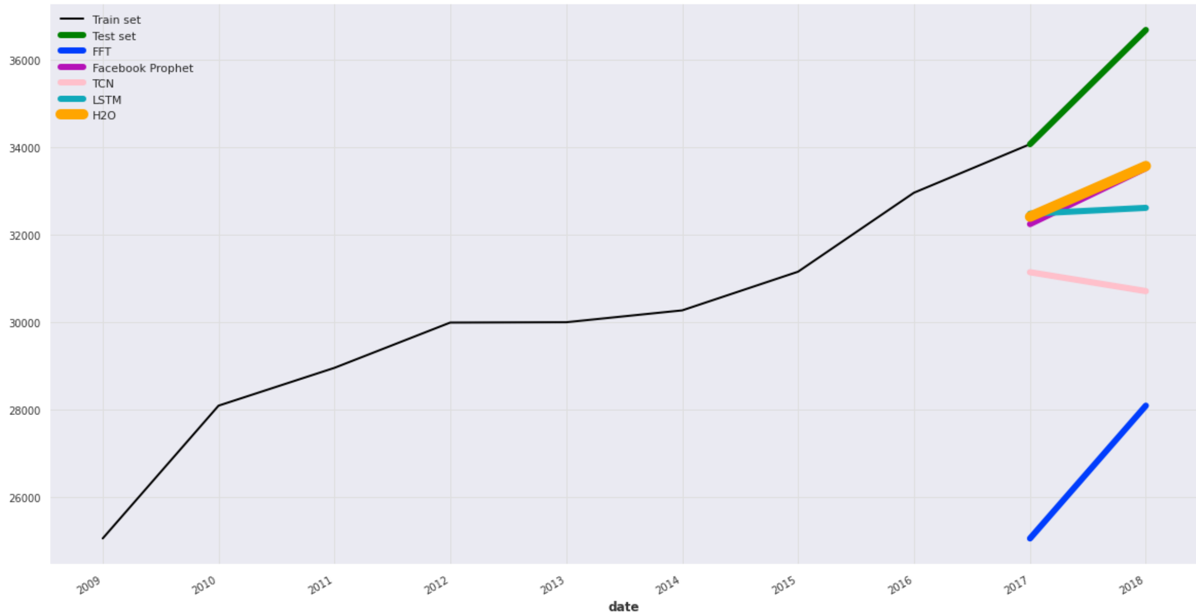
Accuracy of the models

To ensure the performance of the H2O AutoML framework, this study also compared the leader model with another four popular benchmark models. Table 5 reports the comparison between the prediction by the leading model and four other state-of-art models, including Fast Fourier Transformation, Facebook Prophet, Temporal Convolutional Network and LSTM. According to the results of MAE and MAPE, H2O AutoML outperforms other benchmark models in the prediction of tourist arrivals to Sydney with the lowest MAE and MAPE. Figure 7 visualises the actual arrivals and the prediction using different models. The black line represents the training data while the green line represents the test data. The orange line represents the prediction by the trained H2O model and those lines in other colours are the prediction from other models. The results show that the H2O model outperforms other models and thus, the leading model can be used to explain the impact of each topic on the response variable which is the tourist arrivals to Sydney.

Table 5. Comparison between H2O model and benchmark models

Model	**MAE**	**MAPE**
Fast Fourier Transformation	8803.0	24.95%
Facebook Prophet	2494.0	6.99%
Temporal Convolutional Network	4231.5	11.82%
LSTM	3394.5	9.47%
H2O AutoML	**2385.0**	**6.68%**

Figure 7. Comparison between the leading model and other baseline models



Explainability of the model

Figure 8 demonstrates the importance of the variables for the leading model. The figure shows that, for the tourist arrivals in a year, the most powerful predictor is the number of tourists to Sydney one year ago. Apart from the lagged arrivals, Figure 8 shows that, the five most significant predictors for tourist arrivals to Sydney are events and activities; travel costs; exogenous; food and symbolic factors. According to the categorisation in the literature review section, these five factors belong to significative factors (events and activities; food); economic factors (travel costs); exogenous factors and symbolic factors.

Figure 8. Variable importance

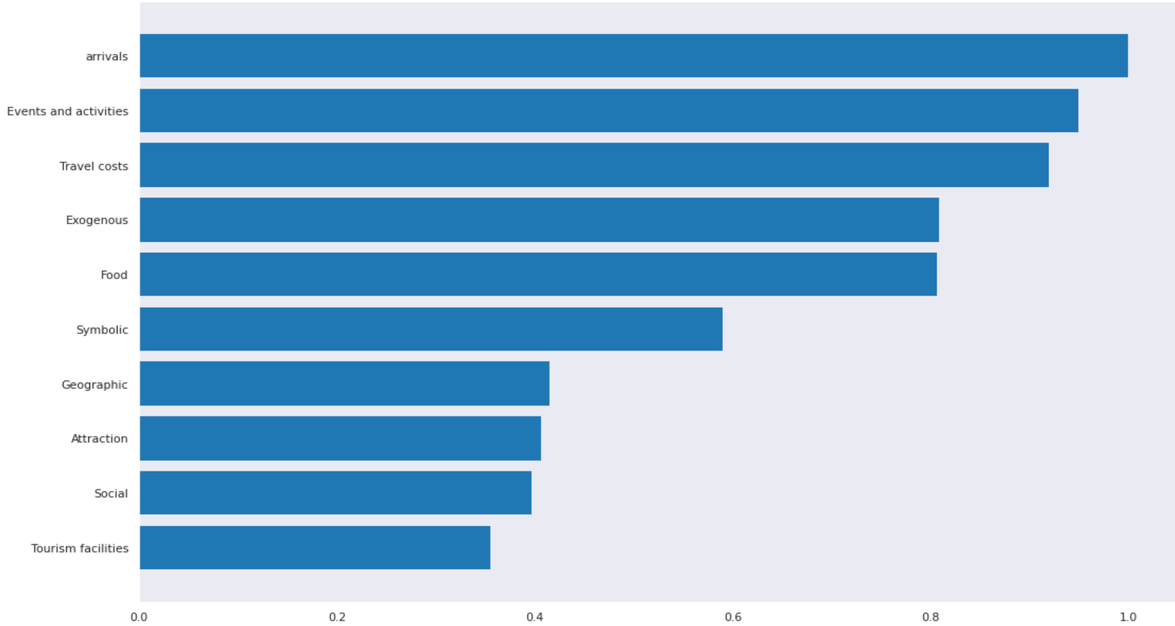
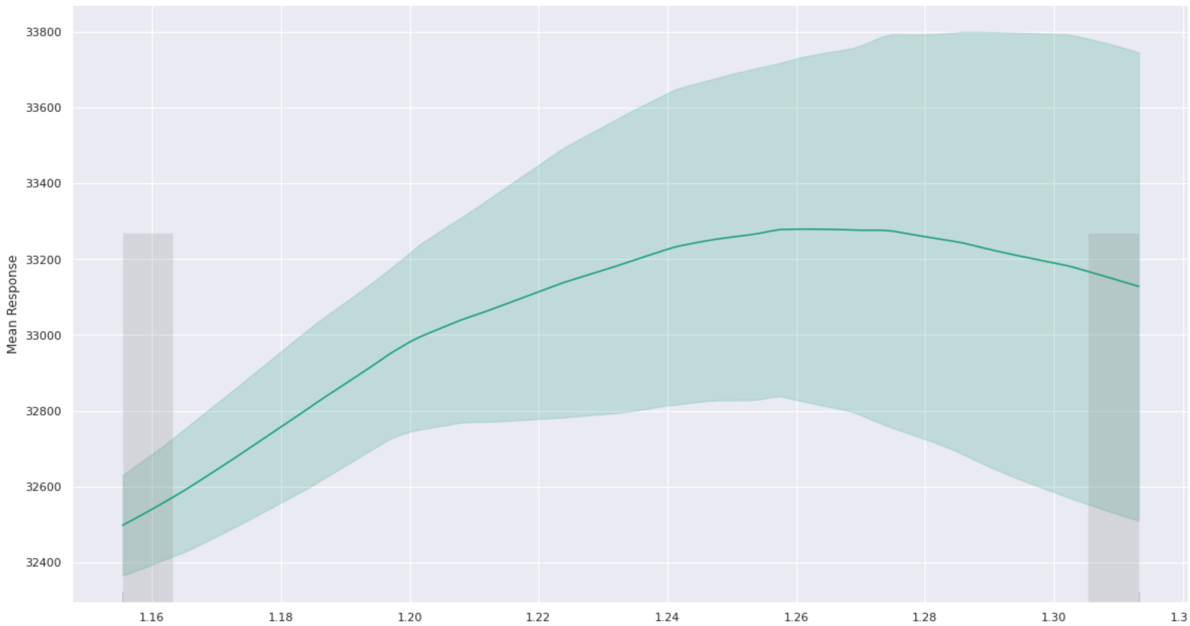


Figure 9 shows how the tourist arrivals respond to the change in the sentiment of the events and activities in Sydney. The results show an inverted-U shape of the relationship between the sentiments towards "events and activities" and tourist arrivals. As people show more positive sentiment when discussing events and activities in Sydney, tourist arrivals are more likely to increase in the next year. However, this positive effect starts to diminish when the average sentiment is around 1.26. After the threshold, if the sentiment continues to increase, the tourist arrivals next year might decrease. The results suggest that, while positive sentiment can be positively associated with tourist arrivals, too much excitement may hinder the number of visitors next year.

Figure 9. Partial dependence plot for the events and activities



According to Figure 10 below, which is the PDP for the relationship between how people perceive “travel costs” one year ago and the tourist arrivals. The results show that the increase in the sentiment shown when people discuss “travel costs” may decrease the number of tourists in the next year but this effect is not as obvious as the last predictor “events and activities”.

Figure 10. PDP for the sentiment of travel costs

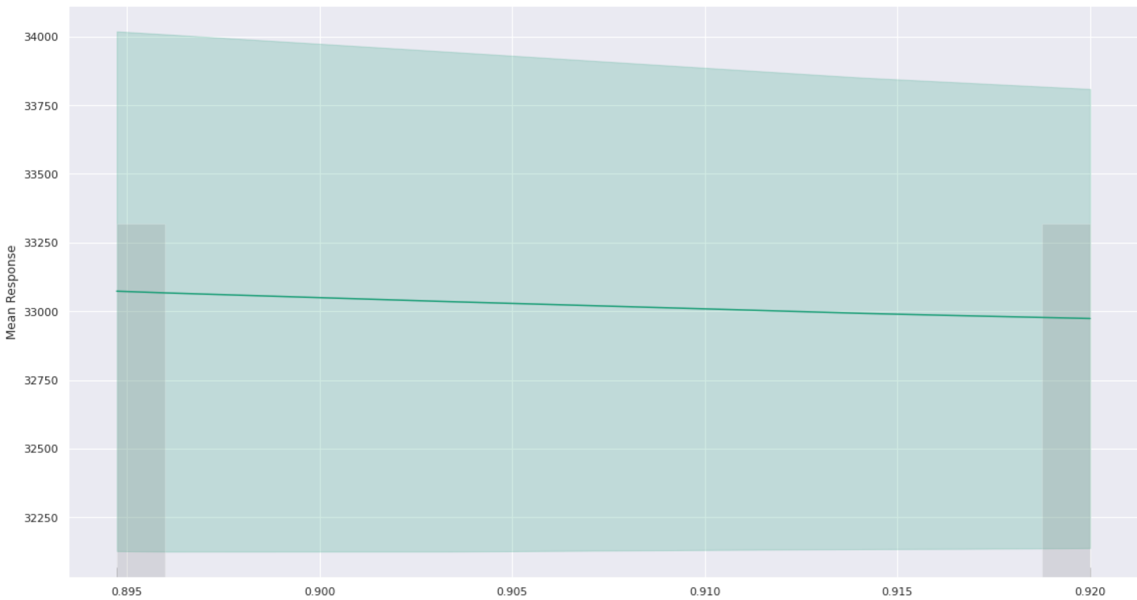


Figure 11 below presents how tourist arrivals may respond to the change of sentiment of “exogenous factors”. Exogenous factors, as described in the literature review section, refer to non-tourism-related macro information of a destination, such as political stability, general safety and technology advancement (Uysal, 1998). The figure shows that the change of sentiment towards those exogenous factors may slightly decrease the tourist arrivals to Sydney in the next year.

Figure 11. PDP for the sentiment of exogenous factors

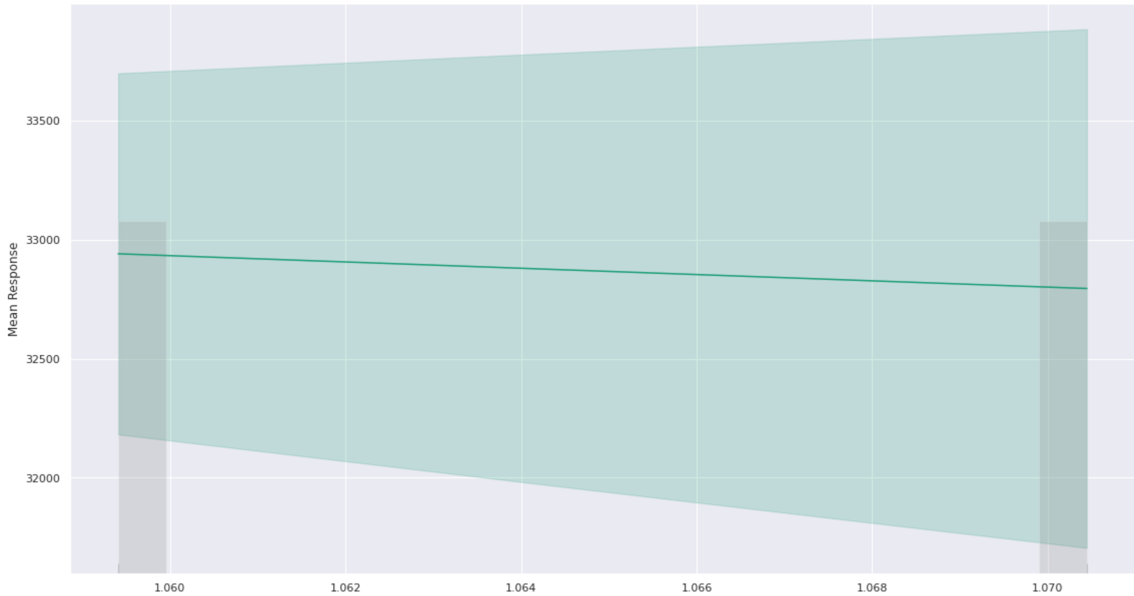


Figure 12 shows the effect of sentiment towards food in Sydney on the tourist arrivals to the city. According to the figure, the tourist arrivals to Sydney barely respond to the change of sentiment toward food in Sydney. In other words, tourists are less likely to be affected by the food when deciding to visit Sydney or not.

Figure 12. PDP for the sentiment of food

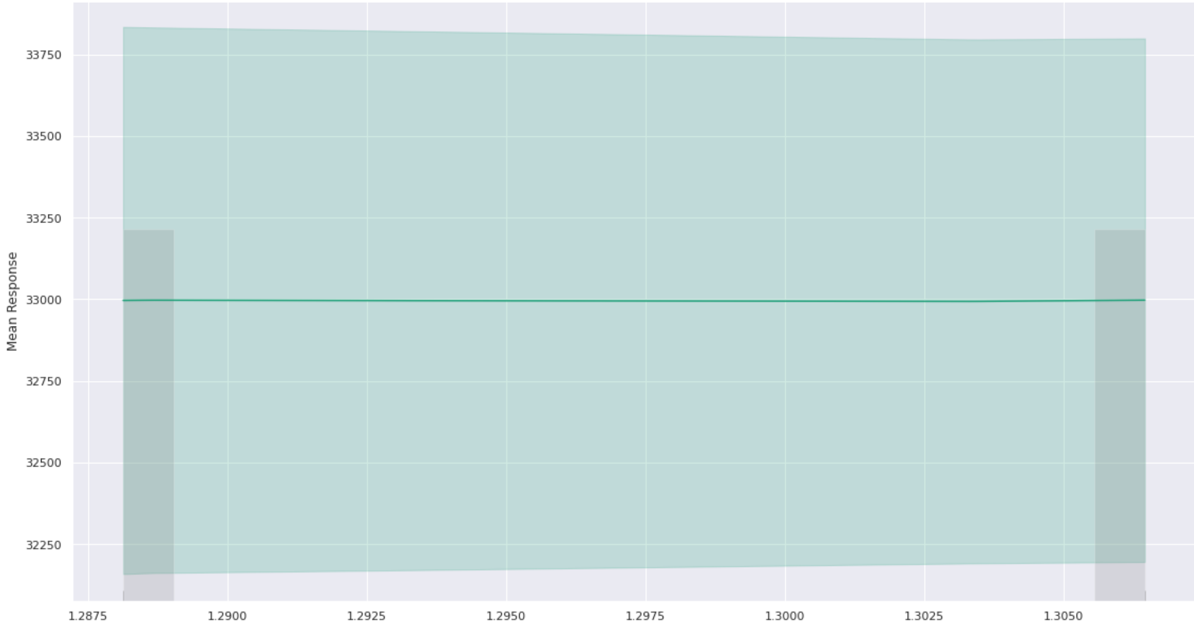


Figure 13 below shows, averagely, how tourists respond to the sentiment of symbolic factors, i.e., marketing and promotion information. The result shows that the sentiment shown in the marketing and promotion messages regarding Sydney may positively influence the number of tourists to the destination, but the effect is rather limited.

Figure 13. PDP for symbolic factors

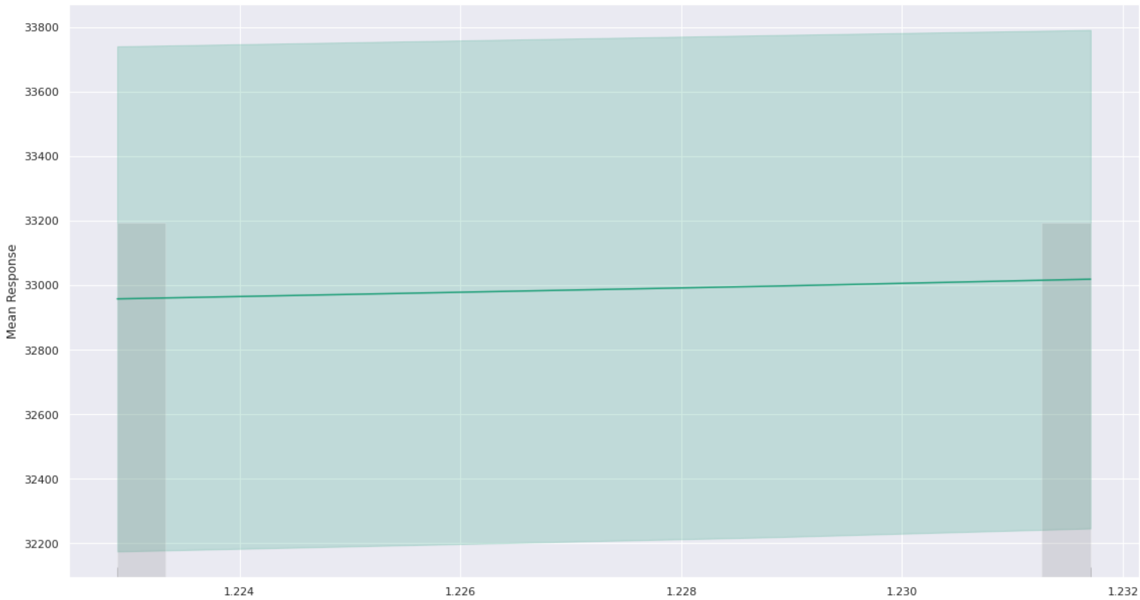
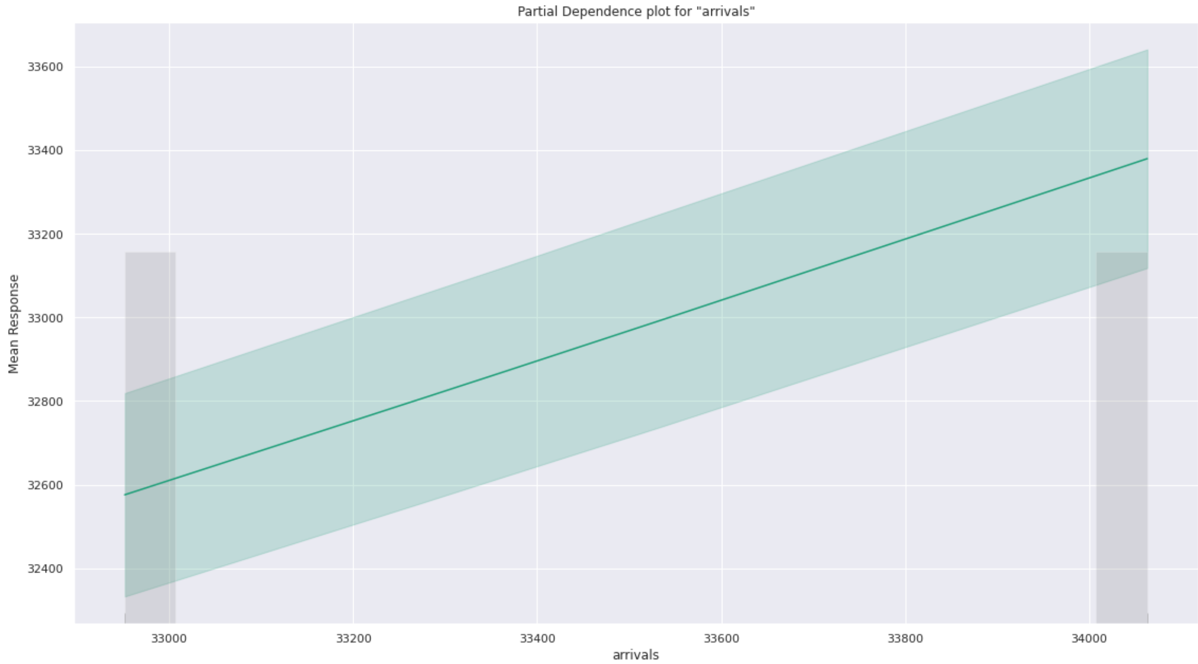


Figure 14 below shows a clear positive correlation between tourist arrivals and itself after one year. In other words, tourist arrivals to Sydney are autoregressive. Therefore, pure time-series models should perform well in tourist arrival forecasting (Apergis et al., 2017; Chu, 1998;

Goh & Law, 2002; Hassani et al., 2017; Kim et al., 2011; Martin & Witt, 1989; Song & Li, 2008; Song et al., 2019b; Thushara et al., 2019; Tsui & Balli, 2017; Witt & Martin, 1987; Wu et al., 2017).

Figure 14. PDP for arrivals at lag (1)



Chapter 7 **Study 2: London**

7.1 Data

For the social media data regarding London, the author adopted a different tool from the one used in Study 1, because the software Twitterscraper stopped working at that time. The author applied for an official academic research account for Twitter API which allows researchers to collect the full archive of tweets posted since 2006. A python script was written by the author to automate the process of collecting all tweets mentioning London for ten years from 2010 to 2019. At last, 1,048,574 tweets in total were collected.

While conducting the studies on Sydney and London, the timeframe for the data collection of tweets varied; the data collection period for Sydney extended from 2008 to 2019, while for London, the period was from 2010 to 2020. Despite the inconsistency in the timeframes, it is essential to appreciate that the period difference does not inherently invalidate or undermine the findings of the two studies. Three factors justify this discrepancy.

Firstly, this study aims to assess the impact of Twitter sentiment on tourism demand, necessitating data across a substantial duration to capture meaningful trends and patterns. The selected periods provide ample timeframes to observe these trends and establish the predictive power of the sentiment analysis..

Secondly, the aim of the two studies is not to make direct comparative evaluations between Sydney and London. Instead, each study explores the impact of social media sentiment on local tourism demand. Therefore, a direct overlap in the periods of study isn't strictly necessary.

Lastly, the machine learning algorithm used in this study, H2O's AutoML, is robust in handling data across different periods. It is designed to work effectively on diverse and non-uniform datasets (LeDell & Poirier, 2020). By employing cross-validation and other rigorous validation techniques, the algorithm ensures that the forecasting model generated is valid and accurate, regardless of minor discrepancies in the data collection periods.

The quarterly data of tourist arrivals to London are collected from a dataset prepared by Visit Britain (2020), a non-departmental national tourism agency.

7.2 Results for topic modelling

The BERTopic model identified 88 topics. After filtering all outlier topics (labelled as -1 by BERTopic), the author attempts to group the rest into the ten categories according to Chapter 4.3. Table 6 shows the topics with keywords under each category. The main categories of topics are exactly the same as those topics extracted in Study 1 including economic factors; exogenous factors; internal factors; attraction; geographic factors; tourism facilities; events and activities; food; symbolic factors. However, the specific topics in these two studies are not the same. For economic factors, people discussed the price of products and various deals available. An example is "*Just need to count my pennies now to check I can afford the weekends in London*". Exogenous factors discussed regarding London on social media are rather similar to Study 1, including various factors, such as employment market, politics and business environment. For instance, "*Product Management Director: London-London, Product Management Director - London £80-100k Basic, plus Bonus, Car* ". *Internal factors* for London are different from those for Sydney. For example, the discussion on personal occasions, such as birthdays, is one of the popular topics regarding London, such as "*not bad! i live in london now! it's my birthday tomorrow! i got married on saturday!*"

Sydney and London are two different cities and thus, discussion on different *attractions* is expected. Results of the topics show that London bridge, London zoo, London Eye, London theatre and different museums are the most discussed attractions. For example, *"#Somebodyshould go 2 the eye in london & sky dive off of it."* *Geographic factors*, including weather and pollution in London, are another category discussed on Twitter. "*Landed weather in London is nearly as cold as the dolomites. Thai dinner, early to bed. Prep day tomorrow. Maybe Sherlock Homes in the PM.*"

Similar to Study 1, the *tourism facility* in London is well discussed by Twitter users. Tourism facilities are infrastructure that can support tourists' tourism activities during being at a destination. The discussion involves airport, public transportation (taxi; tube; busses; trains), accommodation, shopping and beauty facilities. For example, "*Did you even go to London if you didn't go on the tube? That was me today. London, but didn't see any of it*". People pay attention to different *events and activities* in London and Sydney as expected. For London, Twitter users are interested in music, art, fashion, sports, film, books and others. For instance, "*A Great Big World - Live gig at , 5.12.2019. Watch the full video on my channel*" and "*A v* *special film* 🎥💥*born in the east end I constantly heard that title growing up/reminds me of my late Mum n dad*".

Unlike Study 1 where two separate *food experience* topics were extracted, the BERTopic algorithm extracted one unified topic covering the discussion on both food and drink. For instance, "*Report: UK Drinking Culture Strains Health System: By AP LONDON (AP)*" and "*Organic beef marrow bone from Planet Organic" in #London #UnitedKingdom #foodwaste #free*". This means when discussing London, people tend to discuss food and drinks at the same time while, when discussing Sydney, people tend to discuss food or drinks separately. *Symbolic factors* for the two cities are rather similar, including photos, blogs and videos of London. For example, "*someone on my curious cat asked me to write a blog post on being a student and being able to afford to go to shows in London*". Another example is "*Bumped into at the Pokemon London store today!* 💻 *His video is already out*"

Table 6. Ten categories with keywords (London)

Category	Topics with important keywords
Economic	*expensive; money; pay; cash*
	deals; groupon; off at; daily
Exogenous	*jobs; job; manager; united*
	mayor; boris; brexit; labour
	nhs; health; nurse; hospital
	iran; syria; pakistan; palestine
	fire; fireworks; firefighters; grenfell
	paris; london paris; to paris; paris london
	muslim; muslims; islamic; islam
	gay; women; lgbt; womens
	buhari; nigeria; nigerian; president
	riots; protest; london riots; march
	stock; exchange; london stock; stock exchange
	startup; tech; startups; business
	khan; sadiq; sadiq khan; mayor

	homeless; rough; homelessness; the homeless
	accent; london accent; accents; cockney
	migrants; merkel; migrant; eu
	russian; russia; steele; the russian
	dublin; irish; ireland; belfast
	awards; award; awards in; congratulations
	war; churchill; memorial; ww2
	missing; missing from; help find; help
	kate; prince; duchess; duchess of
	charity; donate; fundraising; fundraiser
Internal	*birthday; happy birthday; happy; my birthday*
	packing; pack; packing for; packed
Attraction	*bridge; tower; london bridge; tower bridge*
	zoo; london zoo; wildlife; animals
	eye; london eye; eyes; eye london
	theatre; london theatre; theatre london; theatre in
	museum; museums; natural history; history museum
Geographic	*weather; cold; rain; the weather*
	sunset; sky; lights; light
	snow; snowing; snow in; the snow
	pollution; air; low; smog
	climate; climate change; extinction; change
Tourism facilities	*flight; airport; flights; heathrow*
	uber; taxi; cab; london taxi
	flat; bedroom; rent; bed
	hotel; hotels; hotel london; london hotel
	underground; tube; london underground; map
	shopping; shop; store; store in
	fitness; gym; workout; trainer

	bus; london bus; buses; transport
	train; trains; the train; train to
	hair; salon; barber; beauty
Events and activities	*tickets; going to; tomorrow; cant*
	music; gig; tour; radio
	event; teacher; conference; school'
	fashion; fashion week; london fashion; lfw
	marathon; london marathon; running; running the
	art; exhibition; gallery; painting
	cycling; bike; ride; cycle
	show; london show; shows; show in
	film; film festival; festival; london film
	book; books; read; london book
	christmas; xmas; merry; merry christmas
	escort; sexy; massage; services
	harry; potter; harry potter; london today
	niall; liam; louis; zayn
	church; prayer; pray; worship
	boxing; fight; vs; conor
	wimbledon; tennis; murray; finals
	party; party in; to party; the party
	wedding; royal wedding; london wedding; weddinghour
	justin; performing; he; beliebers
	orchestra; symphony; london symphony; symphony orchestra
	dance; dancing; ballet; dancers
	comic; comic con; con; london comic
	knights; london knights; game; score
	american; an american; warren; horror
	play; playing; playing in; to play

Food	*food; restaurant; tea; coffee*
Symbolic	*photo; photography; photos; photographer*
	video; liked; liked video; video from
	blog; blog post; new blog; bloggers
Social	*None*

7.3 Results for time series analysis

Figure 15 shows the change in people's sentiments when they discuss London on Twitter. The plot shows that the sentiments kept increasing until the middle of the year 2013, followed by a consistent decrease until 2020. Table 7 reports the descriptive summary for the tourist arrivals and the overall sentiment scores during the 10 years from 2010 to 2019. The mean of the overall sentiment toward London is similar to the data towards Sydney (1.18 vs 1.14). Furthermore, like in Sydney, the minimum sentiment score is larger than 1 (1.07) meaning people generally showed positive attitudes toward the city.

Figure 15. Overall sentiment for London from 2010 to 2019

Table 7. Descriptive summary of London arrivals and overall sentiment

	sentiment	arrivals
mean	1.19	4447.25
std	0.05	751.28
min	1.07	2975.0
25%	1.16	3982.25
50%	1.20	4504.0
75%	1.22	4988.5
max	1.29	6122.0

7.3.1 Preliminary study (overall sentiment vs tourist arrivals to London)

According to Figure 16, the effect of overall sentiment four quarters/one year ago on tourist arrivals to London is positive until a threshold where tourists start to avoid visiting this city. This means that, generally, the increase in the overall sentiment can lead to an increase in the number of tourists. However, too positive sentiment may attenuate tourists' intention to visit London. The results confirm that sentiment can influence tourist arrivals. However, the specific relationship between sentiments on social media and tourists arrivals is contrary to previous literature arguing that positive sentiment or image may lead to positive expectations and a positive image of a destination (Afshardoost & Eshaghi, 2020; Ahmad et al., 2021; Baker & Crompton, 2000; Crompton & Ankomah, 1993; Gartner, 1989; Önder et al., 2019).

The Partial Dependence Plot shown in Figure 17 confirms the positive relationship between arrivals at lags (4) and tourist arrivals. This finding aligns with the results for Sydney and is consistent with previous research using a pure time-series approach to forecast tourism demand (Cho, 2003; Chu, 1998; Goh & Law, 2002; Martin & Witt, 1989; Song et al., 2019b; Wu et al., 2017).

Figure 16. Partial Dependence plot for overall sentiment at lags (4)

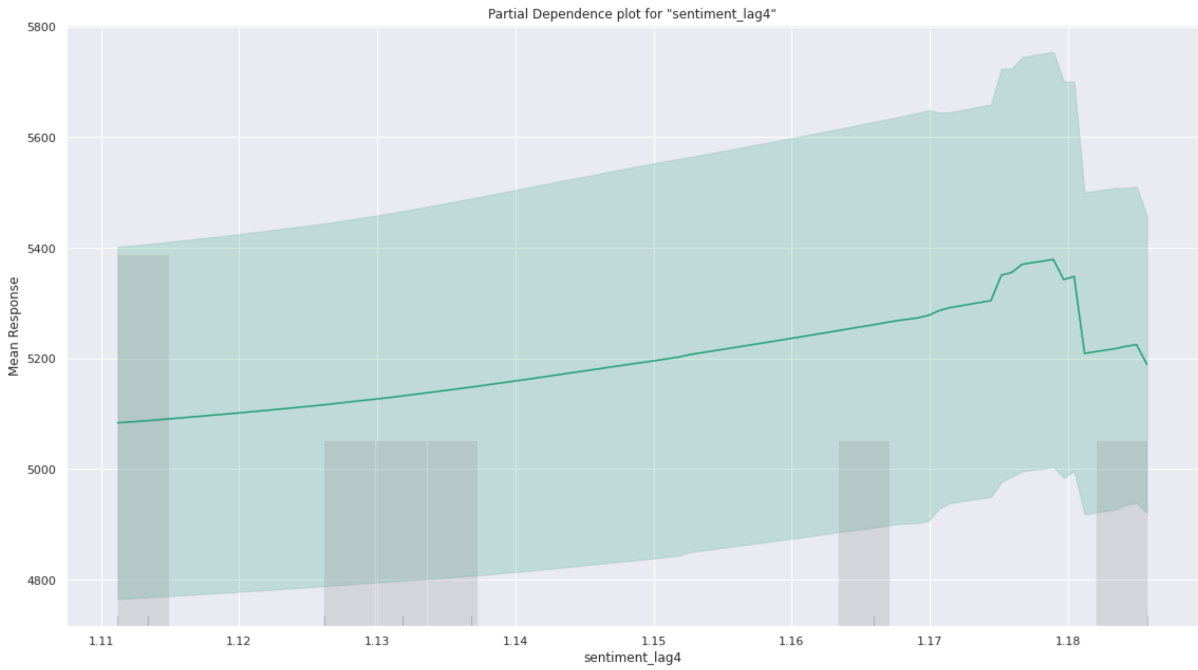
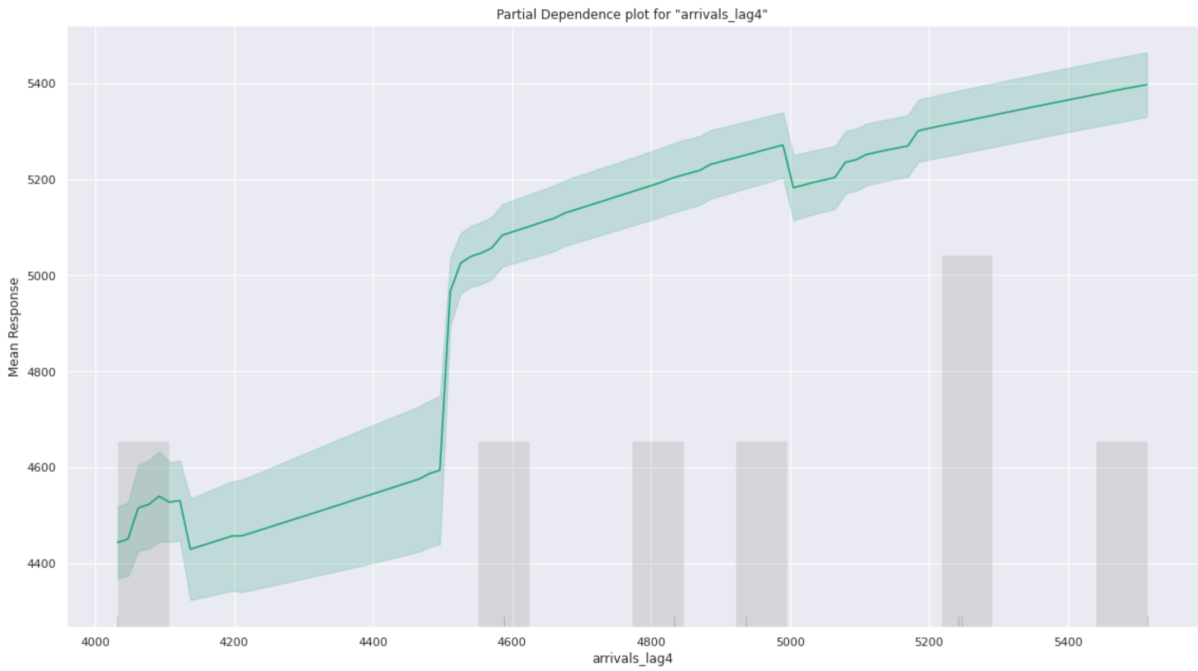


Figure 17. Partial Dependence plot for arrivals at lags (4)



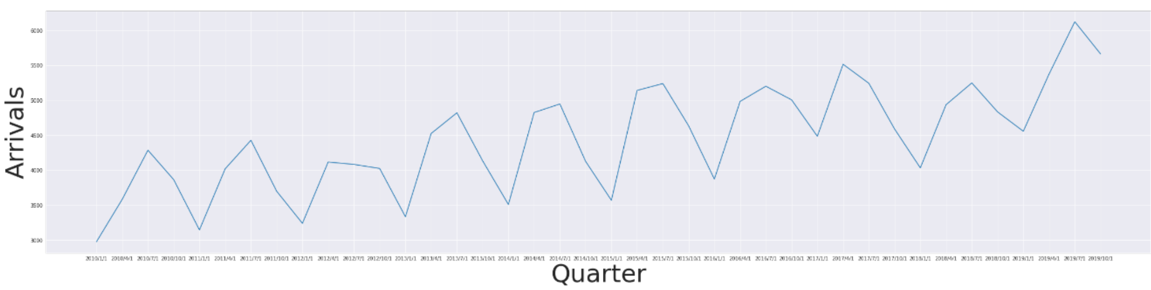
7.3.2 Results (topical sentiment vs arrivals)

The sentiment scores for each topic at each quarter from 2010 to 2019 were calculated according to the equation discussed in the methodology section. Table 8 and Figure 18 presents the descriptive summary of the sentiments towards each topic and visualise the change of sentiments during the ten years. According to Table 8, on average, people show the most positive sentiment (1.54) towards internal factors, and the least positive sentiment (0.95) towards exogenous factors. In other words, when talking about the internal motivation or feelings of visiting London, people tend to express positive sentiments. However, people are more likely to show negative sentiment when discussing economic factors.

Table 8. Descriptive summary of sentiment scores of the topics

	Attraction	**Economic**	**Events and activities**	**Exogenous**	**Food**	**Geographic factors**	**Internal**	**Symbolic**	**Tourism facilities**	**arrivals**
mean	1.19	0.96	1.34	0.95	1.41	1.06	1.54	1.36	1.1	4447.55
std	0.06	0.12	0.03	0.07	0.06	0.11	0.11	0.05	0.07	751.28
min	1.04	0.64	1.3	0.84	1.26	0.84	1.3	1.24	0.95	2975.0
25%	1.13	0.87	1.32	0.89	1.38	0.99	1.46	1.33	1.06	3982.25
50%	1.19	0.97	1.34	0.95	1.41	1.05	1.55	1.35	1.11	4504.0
75%	1.23	1.06	1.36	1.01	1.45	1.15	1.61	1.39	1.14	4988.5
max	1.32	1.15	1.4	1.07	1.54	1.25	1.82	1.53	1.29	6122.0

Figure 18. Tourist arrivals (quarterly) to London



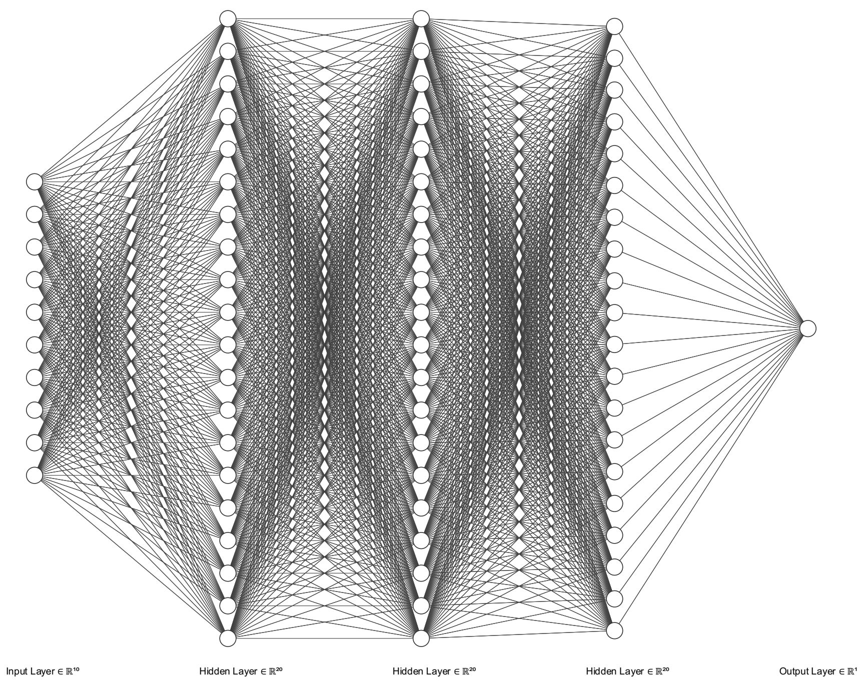
Accuracy of the models

As discussed in the methodology chapter, the time series was divided into training (80%) and test sets (20%). The training set was used to train the model and the test set was to evaluate the performance of the trained models. The H2O Auto Machine learning generated a report for the leaderboard of the best-performing models. Table 9 demonstrates the RMSE and MAE metrics of the five leading models on the test dataset. According to Table 9, the best performing model is a variant of the deep learning neural network model due to its lowest RMSE (504.02) and MAE (350.253). The structure for this neutral network is shown in Figure 19. The neural network consists of one input layer (10 input neurons), three hidden layers (each has 20 neurons) and one output layer (1 output neuron).

Table 9. Leaderboard of best performing models.

Model	RMSE	MAE
* DeepLearning_grid_3_AutoML_1_20220701_93140_model_76	504.02	350.253
DeepLearning_grid_3_AutoML_1_20220701_93140_model_28	509.169	389.023
XGBoost_grid_1_AutoML_1_20220701_93140_model_9	519.929	406.477
XGBoost_grid_1_AutoML_1_20220701_93140_model_4	524.847	394.713
DeepLearning_grid_3_AutoML_1_20220701_93140_model_569	525.416	387.353

Figure 19. Diagram for the leading neural network model

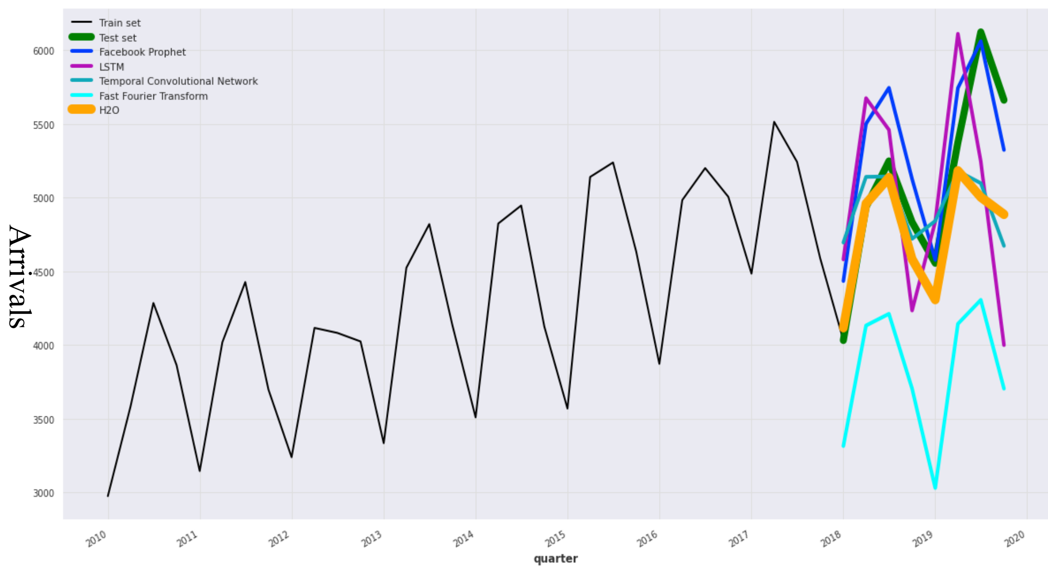


The leader model was, then, applied to the test set to predict the tourist arrivals to London based on the sentiments towards the 10 topics and the date. Figure 20 illustrates the comparison between the leading model and four other state-of-art baseline models, including Facebook Prophet; LSTM; Temporal Convolutional Network; and Fast Fourier Transformation. The figure shows that the leading model outperforms other baseline models due to its MAPE value (6.36%) being the lowest among other models.

Table 10. Comparison between H2O model and benchmark models

Model	MAE	MAPE
Fast Fourier Transform	1277.5	24.74%
Facebook Prophet	318.125	6.4%
Temporal Convolutional Network	447.375	8.63%
LSTM	706.75	13.57%
H2O	**350.5**	**6.36%**

Figure 20. Comparison between the leading model and baseline models



Explainability of the model

Figure 21 below lists the important variables according to how much they can contribute to the forecasting by the model. The top-5 influential factors, apart from the arrivals at lags (4), are *food*, *symbolic factors*, *internal factors*, *exogenous factors* and *geographic factors*. The following section will present the Partial Dependence Plots for each topic to discuss the specific effects of each topic on tourist arrivals to London.

Figure 21. Variable importance

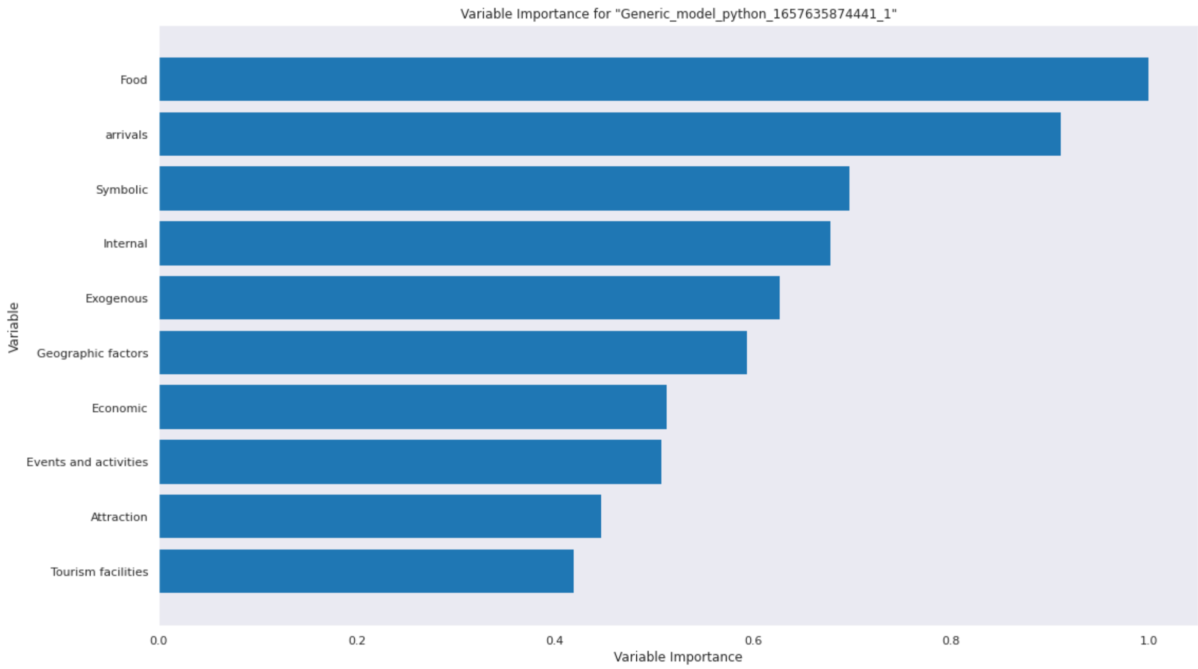


Figure 22 shows how the change in social media sentiments towards food can influence tourist arrivals to London. The figure shows an upslope line as the sentiment score increase, which means that the more positive the sentiment towards food is, the more tourists would visit London. In other words, London tourists are influenced by the discussion of food in the city and their intentional and actual behaviour can be affected positively. These results also show the importance of food for tourists visiting London which is similar to the Sydney study. However, unlike the Sydney study, the forecasting model identifies a positive relationship between the sentiment for food and tourist arrivals.

Figure 22. Partial Dependence plot for food

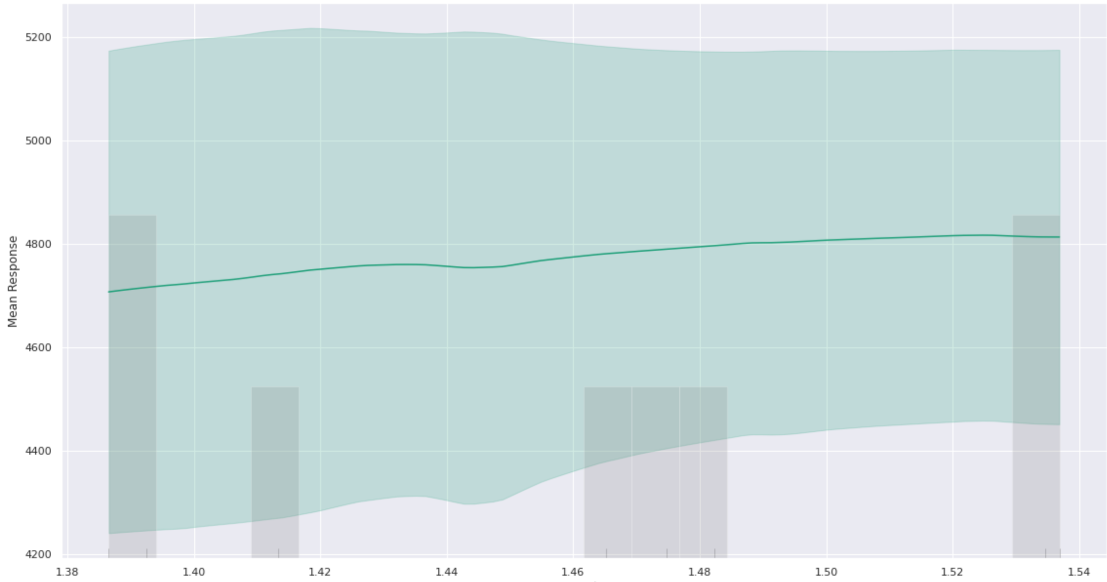


Figure 23 shows how the sentiment of symbolic posts, such as marketing and promotion messages, can influence tourist arrivals. The results show that people are less likely to be influenced by the sentiments of advertisements posted. In addition, when the sentiments are too positive (above 1.38), tourists start to slightly avoid visiting the city. This means that, when promoting a destination or activities in the destination, too much excitement may hinder tourist arrivals. This finding is similar to the Sydney case regarding the limited effects of sentiments in commercial advertisements on attracting tourists.

Figure 23. Partial Dependence plot for symbolic factors

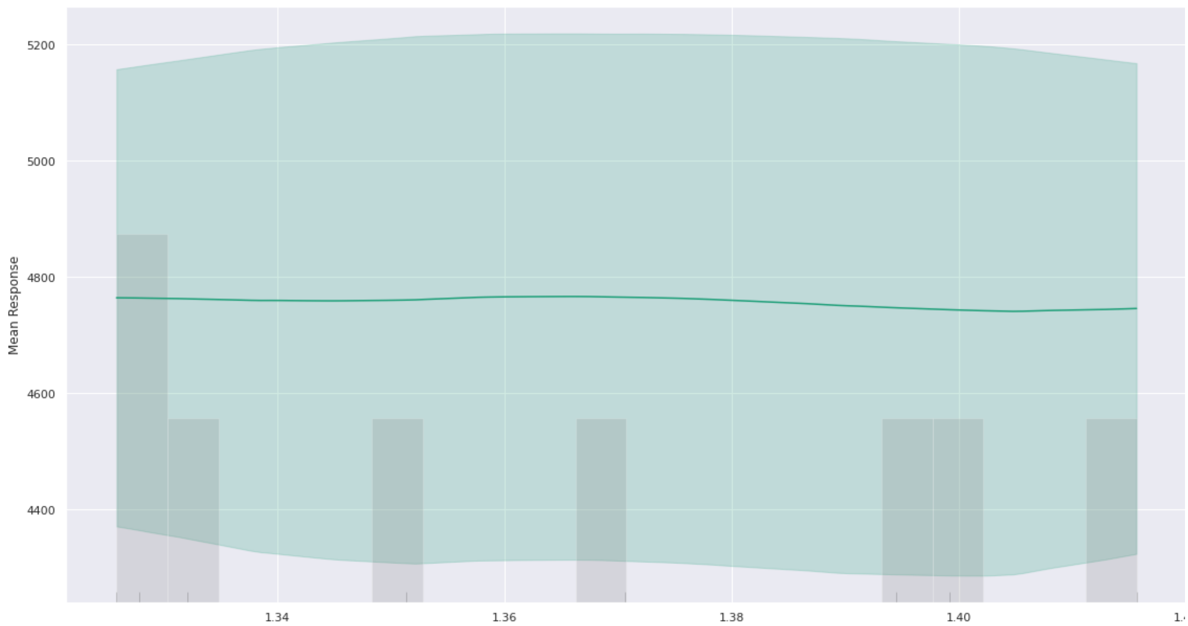


Figure 24 shows the partial dependence plot for the effects of the sentiment of internal factors on tourist arrivals after one year. The effect is not obvious when sentiments are close to neutral but as the sentiments increase, the number of tourists in the next year starts to decrease. Internal factors refer to psychological “push” factors which motivate tourism to a particular destination (Dann, 1977; Um & Crompton, 1990). These factors may include escaping from routine daily life, relaxation, exploration, social interaction, self-esteem, and novelty (Botha et al., 1999; Jensen, 2011).

Figure 24. Partial Dependence plot for internal factors

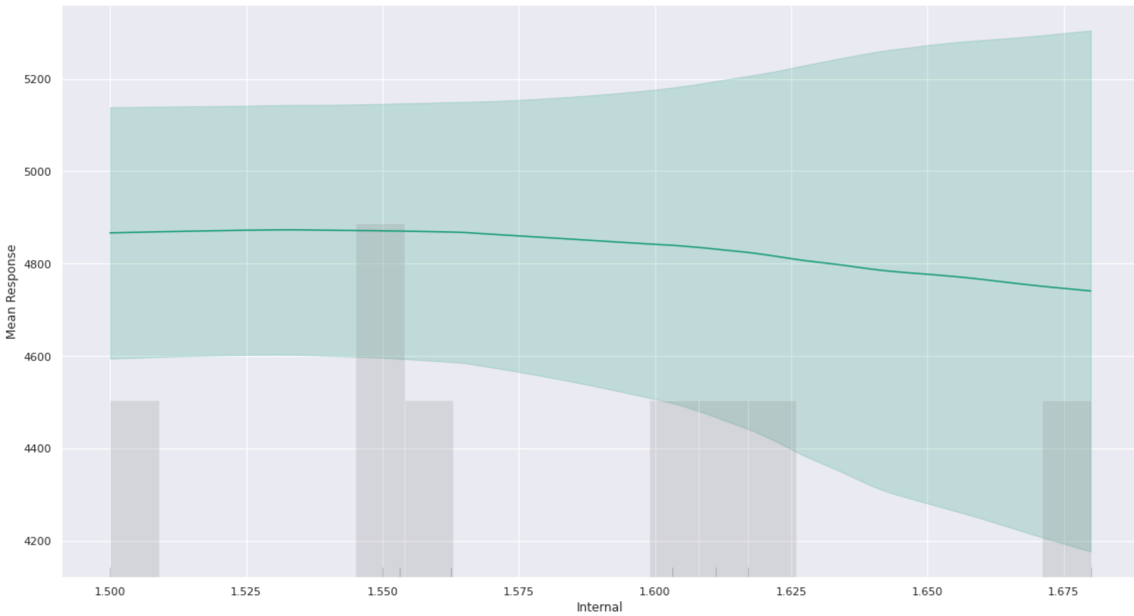


Figure 25 is the illustration of the partial dependence plot for the average sentiment of exogenous factors on tourist arrivals to London. As discussed in the literature review, exogenous factors include those non-tourism-related factors, such as general economic growth and technological advancement. The results show that, generally, people show negative sentiments when discussing those exogenous factors of London. Furthermore, the closer to neutral the sentiment is, the less likely people decide to visit this city after a year.

Figure 25. Partial Dependence plot for exogenous factors

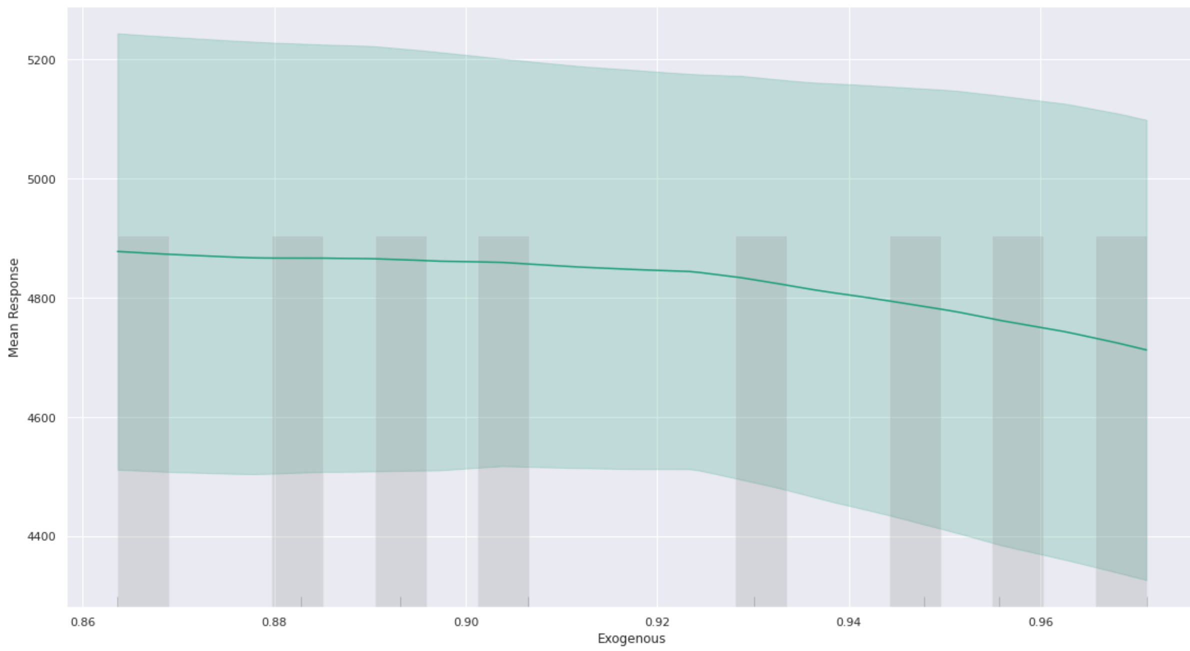
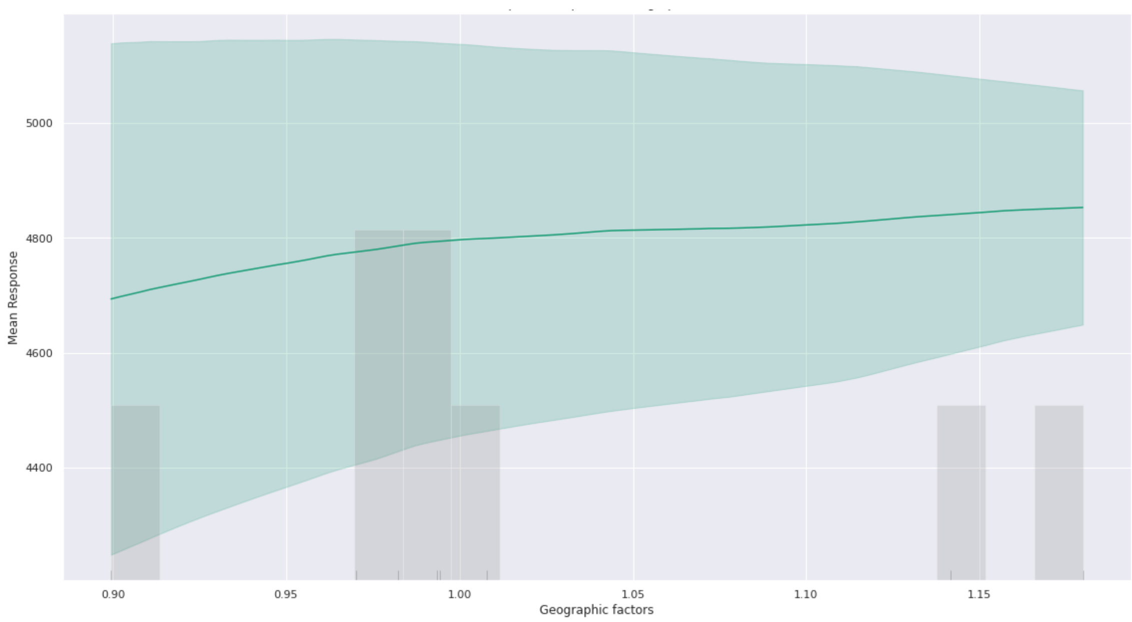


Figure 26 shows the sentiment of geographic factors of London can positively influence its tourist arrivals. The results mean that when the discussion of geographic factors of London, such as weather, is more positive, more people are likely to visit this city. Unlike the Sydney study, the forecasting model for the London study recognises the importance of geographic factors in the top-5 influential factors for tourism demand for London, even though the importance of geographic factors ranks the 6th for the Sydney study, as Figure 8 above shows.

Figure 26. Partial Dependence plot for geographic factors



Chapter 8 Discussion and Conclusions

The primary aim of this research is to examine how topics discussed on Twitter influence the number of tourist arrivals. Destination image can influence tourists' behavioural intention at the stages of pre-visit, during-visit and after-visit, therefore tourists' sentiment towards a destination could play an important role for tourists when selecting a destination to visit. It has been commonly studied as a predictor of tourists' behavioural intention (Afshardoost & Eshaghi, 2020; Souiden et al., 2017). However, the effect of tweets on the aggregated tourism demand has not yet been examined empirically.

This thesis attempted to address the literature gap by examining the effects of destination image on the actual number of tourists. The author proposes a framework for accomplishing this task. Specifically, this framework extracts topics discussed from a number of tweets mentioning the two destinations, namely Sydney and London. It, then, categorises the topics according to the previous literature. The state-of-the-art AutoML framework was then adopted to predict the tourist arrivals to those two destination cities. The forecasting performance of the leading model was evaluated by comparing it with other traditional and state-of-the-art forecasting models. At last, the author adopted the leading model according to the ranking generated from the H2O AutoML framework and interpreted the relationship between different images and tourist arrivals using partial dependence plots (PDPs).

8.1 Discussion of the Main Findings

This research reveals the major topics of a destination extracted from tweets, the sentiments of those topics, and their significance in predicting tourist arrivals.

8.1.1 **Tweets topics**

One of the purposes of this study is to identify the important topics discussed based on the textual tweets collected. The BERTopic algorithm extracted 80 different topics for Sydney, while extracting 88 dimensions for London. The author encapsulated those topics into nine categories, which are economic factors, exogenous factors, internal factors, attraction, geographic factors, tourism facilities, events and activities, food and drink, and symbolic factors.

The results show, for the two destinations, the topics contained in each category are distinct. In other words, people expressed different interests when discussing Sydney and London. When people discuss attractions, for example, they mention completely different topics. For Sydney, this category involves six topics (Sydney Opera House, Biemale Sydney art gallery, Sydney arbour bridge, Sydney zoo, Botanic gardens and Sydney fish market). For London, however, this category covers five different attractions (London Bridge, London zoo, London Eye, London theatre, and museums). Another example is geographic factors. For Sydney, people tend to discuss rain, beach, wind and temperature. For London, rain, sunset, snow, and air pollution are common topics.

8.1.2 **Forecasting performance evaluation of different models**

Tourism demand forecasting is essential in tourism research. Due to the characteristics of tourism products and increased competition, accurate tourism demand forecasting is necessary. Both service providers and destination marketing organisations (DMOs) need to make marketing and operational decisions based on the prediction of tourism demand. According to the results, both the two H2O AutoML forecasting models using sentiments of tweets for Sydney (MAE: 2385, MAPE: 6.68%) and London (MAE: 350.5, MAPE: 6.36%) outperform the other benchmark models (Fast Fourier Transform; Facebook Prophet; Temporal Convolutional Network; LSTM) in the two studies. In addition, the structure of the forecasting models for these two studies is different. For the tourist arrivals to Sydney, an artificial neural network with a hidden layer containing 100 neurons works the best. For the London study, on the other hand, a deep learning neural network with three hidden layers (each has 20 neurons) outperforms other benchmark models. Better performance means the forecasting models capture the ‘effect’ of the independent variables on the dependent variable which is tourist arrivals. This laid a foundation for the next step that investigates the specific effect captured by these two forecasting models.

8.1.3 The effects of the dimensions on tourism demand

The variable importance fiture and partial dependence plot (PDPs) generated from the forecasting models present the importance of each category and the direction of their effects on tourist arrivals to Sydney and London. The two models for the two cities captured different important categories. For Sydney, the most important five categories are events and activities, travel costs, exogenous factors, food, and symbolic factors. For London, the top 5 important categories are food, symbolic factors, internal factors, exogenous factors and geographic factors.

In addition, the partital dependence plot (PDPs) shows that the effect of each category on tourist arrivals varies. For Sydney, the sentiment towards food, symbolic factors are positively associated with tourist arrivals to the city. When people discussed ‘travel costs’, however, high sentiment score seems to hinder the tourist arrivals to Sydney. The events and activities, on the other hand, have a non-linear effect on tourist arrivals. In specific, as the sentiment scores for events and activities increase, the tourist arrivals to Sydney one year later are most likely to increase until the sentiment reaches a threshold, when the tourist arrivals start to decrease. The different effects also occur in the London study. The positive roles of food and geographic factors are revealed by the results. When the sentiment scores of internal factors and exogenous factors increase, on the other hand, people tend to avoid visiting the city.

8.2 Theoretical contributions

In specific, the theoretical contribution of this study is two-fold. First, this thesis extends the marketing research literature by identifying influential topics for tourist arrivals from social media. The research further reveals how the sentiments on different topics discussed on social media influence tourist arrivals.

8.2.1 Twitter topics that influence tourist arrivals.

Traditional forecasting models for predicting arrivals fail because of a large amount of influential factors. These factors sometimes are emerging or uncommon, or because the historical data from which models were derived no longer reflect current conditions. This challenges tourism scholars to include variables that matter the most from unlimited possible variables. For example, economic factors, such as price, have a relatively stable effect on demand (Cho, 2010; De Vita & Kyaw, 2013; Demiralay, 2020; Dogru et al., 2017; Lanouar & Goaied, 2019; Martin & Witt, 1988; Martins et al., 2017; Pham et al., 2017; Shafiullah et al., 2019; Song & Li, 2008; Sun et al., 2019; Tavares & Leitão, 2017). Nevertheless, non-economic factors, especially social-psychological factors, such as exogenous factors (Akinboade & Braimoh, 2010; Cárdenas-García et al., 2015; Schubert et al., 2011), are difficult to be identified.

This research extracts important topics using the BERTopic algorithm and identifies those topics that are uncommon in previous tourism demand literature. The results show that tourists pay attention to different topics regarding a destination. Both studies identify more than 80 specific topics which then be categoriesed into nine groups: *economic factors, exogenous factors, internal factors, attraction, geographic factors, tourism facilities, events and activities, culinary experiences and symbolic factors.* The results slightly differ from the proposed framework as neither of the studies explicitly mentioned 'social factors' as significant topics.

Economic factors refer to the sentiment of tweets discussing economic terms. For Sydney, people tend to discuss transportation costs, such as flight fares. Travel costs that occur during the movement of tourists from their origin to a destination may influence the tourism demand for the destination. Economic theory postulates that when other conditions remain, as travel costs increase, the demand for visiting would decrease. Many empirical studies have verified this effect (Dogru et al., 2017; Lim, 1997; Weis & Axhausen, 2009). Social media posts show different economic topics for London. It seems people are more likely to discuss general economic topics, such as whether a product or service is expensive and available deals in London.

Exogenous factors refer to the general perception of a destination (Uysal, 1998). This type is usually non-tourism related. For Sydney, the topic modelling algorithm extracts seven most discussed topics: general news, employment, real estate industry, traffic, architecture, pet-loving and politics. The exogenous factors for London are similar to Sydey, including employment, politics and business environment. Exogenous factors may be one of the most studied types of drives of tourism demand. Take the sentiment of general news as an example, positive sentiment of the news in a destination may create a more positive expectation of the general economy and business environment, leading people to spend holidays in the destination. Shapiro et al. (2022) conduct a text analysis of newspaper articles and report a positive correlation between the sentiment of general news and economic cycle indicators. Huang et al. (2018) also report that news sentiment can be an indicator to forecast an economic recession. In addition, various dimensions of politics, such as political instability and political ideology, can also influence the tourism demand for a certain destination (Clements & Georgiou, 1998; Hall, 1999; Hanon & Wang, 2020; Saha et al., 2017; Seddighi et al., 2001; Webster & Ivanov, 2016).

Internal factors involves the values, motives and feelings of tourists (Um & Crompton, 1990). Both studies identify internal factors, such as motives of visiting (birthday) and feelings in the city (sleeping). Tourism literature have vastly study the effect of internal factors, such as personal values on tourists' behaviour. According to a study of 60 understraduate, between 41 and 81% of travel behavior can be predicted by combined general values and object-specific values (Pizam & Calantone, 1987). Through a study of 408 young Australian tourists, their emotional, social and functional value can impact the perception of the beneficial image of a tourism product (Phau et al., 2014). A study in 2018 also confirm that people with different attitudes towards risk and uncertainty in travel decision-making greatly differ in their choice of destination (Karl, 2018).

Attractions discussed on Twitter are expected to differ across the two studies and the results confirm this. Results of the topics show that London bridge, London zoo, London Eye, London theatre and different museums are the most discussed attractions. The variety of tourist attractions and attractions within each attraction set that satisfies tourists' needs becomes increasingly important in determining tourist destination choice (McKercher, 2017; Mckercher & Koh, 2017; Navarro-Ruiz & McKercher, 2020). Some scholars even claim that tourism attractions are the core determinants of the attractiveness of a destination (Morozov & Morozova, 2016; Vengesayi et al., 2009)

The literature on the impacts of a specific attraction on tourism, however, is rather limited. Take zoos as an example, zoos have played an important role in Ecotourism (Puan & Zakaria, 2007; Ryan & Saward, 2004). Lee (2015) suggests that improving the zoo environment can enhance visitors' satisfaction levels.

Geographic factors are another type of significative stimulus. Due to the different geographical conditions of the two cities, BERTopic algorithm identified similar and different geographic topics discussed on Twitter. Climate and weather is a common topic for both Sydney and London. This finding aligns with previous studies arguing that climate or weather can influence tourists' intention of visiting a particular destination (Cho, 2010; Goh, 2012; Law et al., 2019; Lise & Tol, 2001; Martín, 2005). Moore (2010) uses a dynamic panel data analysis and confirms that climate change can influence tourism demand in the Caribbean.

Apart from the common topics, there are also distinct topics when Twitter users discuss the two cities. People tend to discuss Sydney's beach but this topic did not occur in the tweets for

London. Previous tourism demand literature postulates that the coast or beach can be an essential motive for attracting visitors (Gössling et al., 2012; Hall, 2001; Scotland Insight Department, 2016). Another unique topic discussed in the London study but not in the Sydney one is air pollution. Air pollution may negatively influence the image of a destination and then hinder visitors to choose the destination. A number of tourism studies have also investigated and confirmed the effects of air pollution on inbound tourist arrivals (Dong et al., 2019; Tang et al., 2019; Xu & Reed, 2017)

Tourism facilities are another type of topic discussed when people mentioned Sydney and London. The topic modelling algorithm identifies five main types of tourism facilities discussed for Sydney: accommodation, transportation, hospitals, gyms and Wi-Fi access. For London, accommodation, transportation, gyms, shopping and beauty facilities are identified. Previous literature has argued the importance of tourism facilities in attracting tourist arrivals and therefore has included this variable in their tourism demand equations (Seetanah et al., 2011). Empirical studies also confirmed the effects of tourism facilities on tourism demand to various destinations. Khoshnevis Yazdi and Khanalizadeh (2017), through a panel data analysis on the international tourist arrivals to the USA from 1995 to 2014, proved that tourism transport infrastructure is a significant determinant of tourism demand. Virkar and Mallya (2018) also agree with the effects of tourism transport facilities on tourists' overall satisfaction. Little research has paid attention to the impacts of Wi-Fi access or Internet connectivity on tourism demand or tourists' satisfaction. Tanti and Buhalis (2017) argue that connectivity can enhance the tourism experience and researchers should gain a better understanding of the role of connectivity to increase tourism demand. Eriksson and Fagerstrøm (2018) investigate the impact of the price and reviews of Wi-Fi access on tourists' hotel bookings. They figure both Wi-Fi price and Wi-Fi reviews can affect booking, but the effect is limited.

Events and activies are another general topic discussed for Sydney and London. Different events and activities discussed are expected for these two cities. The results of the Sydney study identify numerous events, such as concerts, festivals, shows, and mega sports events. For London, however, people are more likely to discuss music, art, and fashion. A number of scholars have investigated the impacts of specific events on tourism demand (Fourie & Santana-Gallego, 2011; Ghalia et al., 2019). Mega-events, such as sports or the Olympics, can have a huge impact on tourist arrivals. Mega-events are events that draw massive international attention and interest. Hosting mega-events can have a positive effect on tourism demand due to the increase in the tourists' awareness, through promotional campaigns of those mega-events. Empirical studies indicate that such events typically lead to an increase in tourism expenditure and a temporary increase in visitors (R. Baumann & V. A. Matheson, 2018; Fourie & Santana-Gallego, 2011; Vierhaus, 2018). R. Baumann and V. A. Matheson (2018) note the 2014 FIFA World Cup increased foreign tourism by roughly one million visitors to Brazil. Vierhaus (2018) also report that hosting the Summer Olympic Games increases international tourist arrivals significantly in the 8 years before, during, and 20 years after the event.

Apart from mega-events, other events, such as festivals, shows and concerts, are expected to affect tourist arrivals to host destinations. Festivals are held to celebrate, commemorate, or promote some phenomenon or locally made product, and many studies have confirmed the positive effects of festivals in increasing tourism demand to a destination (Kendall et al., 2020; O'sullivan & Jackson, 2002; Saleh & Ryan, 1993; Tonga Uriarte et al., 2019). However, some scholars may argue that short-duration cultural festivals may not be effective in attracting tourists (Mckercher et al., 2006). Other events, such as concerts, are also able to drive tourists to visit a destination. Music festivals, such as concerts, according to the study by Frey (1994) , almost perfectly combine culture and holidays. A study by Duarte et al. (2018), using PLS-SEM, empirically investigate the impact of music festivals on destination image. They report that music festivals can contribute to creating a positive image of a destination, which may increase the intention to visit.

Gastronomy or culinary experience is another important topic discussed for the two destination cities on Twitter. Both studies have identified similar topics regarding the discussion on food and drinks in these two cities. The motivation to travel for gastronomy reasons has been a valid construct of tourism demand modelling (Ab Karim & Chi, 2010; Kivela & Crotts, 2006; Sánchez‐Cañizares & López-Guzmán, 2012). Early research by Rimmington and Yüksel (1998) postulates that food experience is a critical factor determining tourists' satisfaction levels. Folgado-Fernández et al. (2017) conduct an analysis of the influence of tourists' food experience and confirm its positive effect on destination image and loyalty. In addition, food tourism is a growing sector that stimulates interest in visiting a destination (Tommy D. Andersson et al., 2017). Food experience can also influence tourism demand by adapting to meet tourist needs, creating new tourism products, stimulating innovation and supporting marketing and branding (Richards, 2012).

Finally, *symbolic factors* are the final general category of topics that involves marketing and promotion messages. Marketing activities are important to deliver essential messages to attract tourists. Two studies extract rather similar topics regarding symbolics, including tickets of various tourism products, interaction with the audience, photographs, videos and blogs. Previous literature has confirmed the effects of the quality of marketing activities (Baloglu, 2001; Chi et al., 2020; Tasci & Gartner, 2007; Woodside & Lysonski, 1989) and symbolic factors, such as characters of film and television regarding a destination (Albu, 2013; Bolan & Williams, 2008; Iwashita, 2008; Wen et al., 2018) on tourism demand to the destination. Iwashita (2008) studies the roles of films and television dramas in attracting Japanese tourists to the UK and reports that films and television dramas can attract international visits by creating destination awareness, consciousness and images. A study by Connell (2005) argues that television-induced tourism can increase the tourism demand to the Isle of Mull, Scotland. However, not all scholars agree with the effect of dramas on tourism demand. Kim and Kim (2018) empirically investigate whether Korean television dramas or K-dramas can attract a Chinese audience, but they did not find a direct effect.

Furthermore, the three essential marketing media, photos, blogs and videos of a destination can show people both cognitive (e.g. fun facts about a destination, specific activities) and affective (e.g. people's feelings when participating in a particular activity) information (Latorre-Martínez et al., 2014). The information can increase people's awareness of the destination and, form or alter their image of a destination. Hem et al. (2003) report that photos of nature-based tourism attractions can increase the awareness of a destination called 'Fjord Norway'. Many potential tourists may also desire to be part of similar experiences and activities. This can lead to an increase in the tourism demand for the destination. In addition, shared videos online can entertain viewers by activating their fantasy and imagination (Tussyadiah & Fesenmaier, 2009). A recent study by Briciu and Briciu (2020) also shows that video-sharing platforms, such as YouTube, are important for the development of the tourist industry as it provides travellers with new practices for choosing their destination.

8.2.2 The way how the sentiments of social media topics influence tourist arrivals

Social media data, such as UGC, has been commonly used in marketing and tourism literature (Colladon et al., 2019; Dhaoui & Webster, 2020; Fletcher-Brown et al., 2020; Reich & Pittman, 2020). Most of these studies pay attention to the effects of social media data on individual behavioural intention or behaviour (Rapp et al., 2013; Varkaris & Neuhofer, 2017). For example, Varkaris and Neuhofer (2017) interviewed 12 individuals to explore how social media influence their decision making when they search, decide and book hotels. Tourism demand literature has attempted to use social media data to improve the performance of their modelling (Bigné et al., 2019; Colladon et al., 2019; Gunter & Önder, 2021; Khatibi et al., 2018; Hailong Li et al., 2020; Miah et al., 2017; Starosta et al., 2019; Tian et al., 2021). Little literature, however, has investigated how individual behavioural intention or behaviour at a micro level can influence tourist arrivals at a macro level. Without understanding the specific effects of social media on tourism demand, tourism decision-makers cannot make better decisions.

This research adopts an AutoML framework and partial dependence plots (PDPs) to investigate the effects of different topics' sentiments on tourist arrivals. It reports that social media sentiments can have different effects on different destinations. First, some of the top-5 influential factors Sydney and London are similar but some factors are distinct. The results show the sentiment scores of exogenous factors, food experience and symbolic factors play important roles in attracting tourist arrivals to both cities. Nevertheless, not all the topics matter the same for the two cities. According to the results, tourists visiting Sydney are more sensitive to various events and economic factors than tourists visiting London. On the contrary, internal and geographic factors play a more important role in attracting visitors to London than to those to Sydney.

Second, the specific influence of sentiments about a particular topic seems to fluctuate across various destinations. Both studies establish a negative relationship between exogenous factors and tourist arrivals, deviating from much of the existing literature (Akinboade & Braimoh, 2010; Al-Mulali et al., 2021; Cárdenas-García et al., 2015; Schubert et al., 2011). *Exogenous factors* refer to those non-tourism-related factors, such as general economic development and technological advancement. In theory, better exogenous factors, such as fast economic and technological development of a destination may help to increase tourists' experience and further increase tourist arrivals to the destination (Al-Mulali et al., 2021). The results of both studies disagree with this theory. One explanation is that people may overestimate the degree of exogenous factors. The too positive sentiment means that people have a high general expectation towards a destination (Garau-Vadell et al., 2018). When they actually arrive at the destination, if the actual experience does not meet the high expectation, people may have lower satisfaction (Vroom, 1964).

The results present that *food experience* has a positive effect on tourists to London while having little effect on tourists to Sydney. Tourists need to eat and drink during their visit to a destination. Food can provide benefits for tourists physiologically and psychologically. It can also help tourists understand the difference between a destination and its origins (Bell & Valentine, 2013). The results in the London study agree with previous literature confirming the importance of food experience when tourists choose a destination (Bell & Valentine, 2013; Cohen & Avieli, 2004; Ellis et al., 2018; Henderson, 2009; Tikkanen, 2007). The results for Sydney, however, does not align with literature, meaning tourists to Sydney care less about food when they make travel decision to the city.

Symbolic factors have limited or negative influence on tourist arrivals to the two cities. A possible reason is that the sentiments can significantly influence tourists' decision on destination choice only when the sentiments are from peers rather than official accounts. In other words, people tend to put more trust in other individuals than in commercial organisations (Schmidt & Iyer, 2015). The impact of trust on purchase intention and behaviour has also been verified by theoretical and empirical studies (Hajli et al., 2017; Hong & Cha, 2013; McCole et al., 2010; Sannassee & Seetanah, 2015). Due to the moderation effects of trust, the sentiment of DMOs or companies marketing messages on social media may not be considered when tourists choose their destinations.

Except for the three influential topics for both two citiies (*exogenous, food experience,* and *symbolic factors*), this thesis also investigated the effects of other influential topics on the toursits arrivals to the two cities. For Sydney, the PDP plot finds that the relationship between events and activities and tourist arrivals is not linear. The results partially align with previous literature. Most previous empirical studies confirm the holding of events and activities may promote tourism demand (Brännäs & Nordström, 2006; Fourie & Santana-Gallego, 2010, 2011; Ritchie & Smith, 1991; Solberg & Preuss, 2007; Vierhaus, 2019). For example, Fourie and Santana-Gallego (2011) report that mega-events may promote tourism demand but this promotion effect may vary depending on different types of events. Brännäs and Nordström (2006) report that festivals can increase tourist receipts for hotels in cities by 2% and 6%. However, the results do not agree with the unidirectional relationship between events and tourism demand as argued by previous literature. The inverted-U shape in Figure 9 postulates that the relationships between events and activities are not necessarily linear or unidirectional. In specific, when the sentiments of events and activities in a particular year are too positive, the tourist arrivals may be reduced in the later year. As to the best knowledge of the author, few studies have reported this non-linearity relationship between events and tourism demand. This finding is rather similar to the results of Vierhaus (2019), who claims that the types of events hosted in a destination can decide the magnitude of their effects on tourism demand. The author report that, though hosting the summer Olympic Games can promote tourism before, during and after the effects, the FIFA World Cup can only promote the international tourist arrivals to the destination during the event. In other words, hosting the FIFA World Cup cannot promote tourism before and after the event.

The results also show that the sentiment scores of *economic topics* are negatively linked to the tourist arrivals to Sydney. Economic theory and previous tourism literature propose that costs of travel are one of the most influential factors for purchase decision-making. Travel costs refer to how much money or time visitors need to spend from their origin to a destination (Uysal & Crompton, 1985). Previous empirical studies often report a negative relationship between travel costs and tourism demand (Lim, 1997; Sunday, 1978). For instance, Sunday (1978) conducted a panel regression analysis and claims that higher airfares may attenuate the number of tourists but increase tourism expenditure per tourist visit. Previous literature also argues that, transportation is one of the most price-sensitive products and it can be the first category when tourists consider cutting their budget (Pyo et al., 1991).

The results shown in this study may be contrary to common sense. The figure below shows the effects of the sentiments of "travel costs", instead of "travel costs" itself, on tourist arrivals. In theory, when people are happy with the travel costs of a destination, they are more likely to travel to the destination. However, Figure 10 shows the opposite. In other words, the increase in the sentiments of transportation costs may slightly lead to a reduction in tourist arrivals. One explanation is that the sentiments shown in the posts cannot represent the actual sentiments of tourists. Through analysing the tweets under this category, the author finds that the tweets were mostly posted by airline companies and other tourism businesses, instead of tourists. For example, "Los Angeles to Sydney, Australia for only $310 USD one-way (& vice versa for $476 AUD)" and "888$ round trip flight to Sydney u down". The results show that the sentiments shown in commercial advertisements can barely exert effects on tourists' destination choices. Practically, airline companies should not show too much excitement when marketing their airlines. Instead, transportation companies, such as airline operators, should be more neutral when promoting their products.

The London study identifies two influential topics which are less significant in the Sydney study: *internal factors* and *geographic factors*. *Internal factors* refer to psychological “push” factors which motivate tourism to a particular destination (Dann, 1977; Um & Crompton, 1990). These factors may include escaping from routine daily life, relaxation, exploration, social interaction, self-esteem, and novelty (Botha et al., 1999; Jensen, 2011).

One explanation of the results can be from expectancy-based model. Albayrak and Caber (2018) identify four main models explaining tourism motivation: need-based; value-based; benefits-based and expectancy-based model. Because of the four seasons gap between the sentiment and tourist arrivals, the expectancy-based model is an appropriate model to explain the negative relationship between sentiments of internal psychological factors and tourist arrivals. The expectancy model proposes that the expectancy of the outcomes of the intended trip can affect travellers’ motivation to visit a destination (Hsu et al., 2010). This model allows a more realistic view of tourists’ motivation (Sharpley, 2018; Witt & Wright, 1992). Several tourism theorists agree with the effectiveness of the push-pull framework However, this model is difficult to apply in the real world. The so-called ‘expectancy’ can be rather complex because, as Witt and Wright (1992) note, includes various factors, such as costs and the quality of accommodation. The expectancy-based model and expectancy confirmation theory can explain the non-linear negative relationship (Alhemoud & Armstrong, 1996; Britton, 1979; Witt & Wright, 1992). Positive sentiments can lead to high expectation of tourists, which may reduce the probability of satisfaction during and after the visit (Alhemoud & Armstrong, 1996; Britton, 1979; Lee et al., 2011; Tasci et al., 2007)

Geographic factors also play a crucial role in attracting London visitors, according to the results. As an important component of significative stimuli (Um & Crompton, 1990), geographic factors, such as weather, climate and beaches, have been focused on by a number of tourism scholars (Cho, 2010; Goh, 2012; Gössling et al., 2012; Hall, 2001; Law et al., 2019; H. Li et al., 2018; Li et al., 2017; Martín, 2005). Most agree with the positive relationship between geographic factors and tourism demand. For example, Martín (2005) argues that those spaces with the greatest comfort can attract the destination choice decision by tourists. Law et al. (2019) also report that weather is one of the important factors that tourists consider on deciding whether to visit Macau. Li et al. (2017) conducted a dynamic panel data analysis on the impacts of climate on tourism demand. They postulate that the different climates of the home city and destination city, and their comparison can influence the tourists' decision on visiting the destination. The positive impacts of the people's sentiment toward geographic factors also align with previous literature (Bae & Nam, 2020; Becken, 2013; Law et al., 2019; Mieczkowski, 1985).

Third, the relationship between overall sentiments and tourist arrivals is not simply linear. Both studies confirm that higher overall sentiments score may lead to more visitors to those two destinations. However, when overall sentiments become too high, the positive impacts start to become limited or even negative. Overall sentiments on social media can partially represent the overall image of a destination. Most previous literature on destination image has identified the positive relationship between overall image of a destination and their visitors, but they fail to recognise the threshold where the tourist arrivals would be negatively influenced by sentiments (Afshardoost & Eshaghi, 2020; Bigné et al., 2019; Lee, 2009; Machado, 2010; McKercher & Tse, 2012). For instance, Afshardoost and Eshaghi (2020) conduct a meta-analysis and confirm that overall and affective images have the greatest impact on behavioural intention. Bigné et al. (2019) argue that tourism image is a direct antecedent of intention to return and willingness to recommend the destination. However, this positive impact does not last forever as overall sentiments grow. Once the sentiments reach a threshold that is similar for both sample cities (between 1.17 and 1.18), the number of tourists would barely change or start to decrease. This "joy begets sorrow" effect has not been commonly identified by literature, though some literature does recognise the existence of moderator for the relationship (Assaker & Hallak, 2013; Assaker et al., 2011; Papadimitriou et al., 2018). For example, Papadimitriou et al. (2018) compare three distinct groups, namely residents, past tourists, and prospective tourists, in terms of the effects of cognitive, affective, and overall destination image on future behaviour. They find tourists make a decision depending on overall image perception, while residents rely more on cognitive and affective image components. Assaker and Hallak (2013) also report that tourists' novelty-seeking tendencies can moderate the relationship between destination image and tourists' revisit intentions.

8.3 Methodological contributions /implications

This thesis shed light on the empirical studies in marketing methodologically. First, this research proposes a novel approach to extracting deeper information from large textual data and exploring the effects of different images on aggregated tourism demand. Most previous social media tourism demand literature simply uses volume- or rating-based data (Bigné et al., 2019; Khatibi et al., 2018; Hengyun Li et al., 2020; Önder et al., 2020; Qiu et al., 2021; Tian et al., 2021). Some scholars have attempted to extract deeper information, such as sentiment, in their tourism demand models (Afzaal et al., 2019; Ampountolas & Legg, 2021; Colladon et al., 2019; Hennig-Thurau et al., 2007). These studies, however, fail to realise that consumers discuss different topics for different destinations and the sentiments towards different topics also have different effects, as this research suggests. This thesis thus proposes an exploratory sequential approach using both qualitative and quantitative methods to extract important topics on a certain destination and to quantify the sentimental information about each topic. This novel approach can help researchers extract and quantify deeper information from unstructured textual data. Researchers can explore the quantified information or use them in further modelling, such as investigating the effects of that information on consumers' behaviour on a large scale.

Second, this study contributes to the tourism marketing literature by introducing a method of selecting appropriate variables from nearly unlimited influential factors of tourist arrivals. Feature selection is a process for choosing a subset of relevant predictors out of the set of all available predictors. Traditional methods of modelling micro consumer behaviour or macro tourism rely on previous experience, knowledge and a bit of luck to select relevant features. Song and Li (2008) propose a tourism demand model with factors including price, income, travellers' tastes, marketing expenditure on tourism by destination and all other influential factors. Most research has confirmed the effects of economic factors, such as price and income (Dogru et al., 2017; Lanouar & Goaied, 2019; Martins et al., 2017; Shafiullah et al., 2019; Song & Li, 2008; Tavares & Leitão, 2017; Y. Yang et al., 2019). However, extracting other influential factors is still challenging for researchers. Tourists are likely to be more likely to travel to places with natural features, such as mountains and lakes. This is especially true if the tourist has specific hobbies or interests. The outcome is that modelling how a potential tourist will decide where to travel is extremely difficult. By extracting the topics and the sentimental information from unstructured online reviews, researchers are able to identify important or emerging factors that are not common or have not been examined before.

8.4 Managerial implications

The proposed analytical framework and results of the current research have practical implications for both researchers and stakeholders in the tourism industry. First, it suggests that tweet data can capture the change in consumers' preferences and the emergence of important topics or issues. Black swan events in recent years, such as COVID-19, have led to a great fluctuation in tourism demand (Duro et al., 2021; Foo et al., 2021; Sharma et al., 2021). Those great fluctuation further challenges researchers to study the tourism demand mechanism and forecast tourism demand. This research shows that social media data can capture the timely fluctuation of tourists' attention and interests towards a particular destination. Social media data, thus, can be used to model tourism demand and improve demand forecasting (Colladon et al., 2019; Dhaoui & Webster, 2020; Fletcher-Brown et al., 2020; Heimbach & Hinz, 2016; Reich & Pittman, 2020; Toubia & Stephen, 2013).

Second, tourism businesses and DMOs can monitor what themes social media users are discussing and their sentiments towards those themes as a social listening approach (Reid & Duffy, 2018; Schweidel & Moe, 2014). Social listening has been widely used by marketing practitioners to understand their product/service position in consumers' mindsets, and to identify a trend or potential crisis (Reid & Duffy, 2018). Westermann and Forthmann (2020) also add that social listening can lower costs and deliver results in a timely manner. Listening to customers is important for DMOs but it is challenging to monitor big data social media conversations manually (Edwards et al., 2020). By combining topical modelling and sentiment analysis, tourism firms and DMOs are able to monitor and allocate the most resources to the most important topics discussed and ignore less important aspects or noises (García et al., 2019; Lim & Buntine, 2014; Vamshi et al., 2018; Vermeer et al., 2019; Yuan et al., 2018; Zhang et al., 2017).

Third, public and private sectors in the tourism industry can make use of this social media approach to make decisions based on how specifically each aspect influence tourism demand. Dynamic, volatile, and time-sensitive industries, such as tourism, travel and hospitality require agility and market intelligence to create value and achieve competitive advantage (Stylos et al., 2021). In general, or those themes whose sentiments are positively (negatively) correlated to tourism demand, decision-makers can increase (decrease) the allocation of the resources to those aspects. For marketers, communicating those positive themes and avoiding/directing those negative themes are expected to increase the tourism demand for a particular destination. For instance, this research reports that the sentiments of the discussion on the activities in Sydney have positive effects on the tourism demand for Sydney. DMOs, thus, can increase the quality and frequency of those posts regarding what people can do in this city. In addition, tourism marketers can also utilise the results in their marketing planning. For example, this research reports that the overall sentiment one year and two years ago can positively influence the tourism demand for Sydney. Therefore, Sydney marketers should make a long-term plan (at least two years ahead) for their marketing campaign on social media.

8.5 Limitations and Future Research

This thesis presents several key limitations that should be considered for future research. First, the study primarily relied on BERTopic (Grootendorst, 2020) While BERTopic is a powerful tool, it has inherent constraints, primarily its assumption that a single document such as a tweet or review contains only one topic. This assumption often falls short in reality where documents contain multiple topics (Grootendorst, 2020). As a result, future research could benefit from the application of multiple topic modelling algorithms to generate a variety of topic sets. By comparing and integrating these topic sets, the validity of the research results could be improved.

Second, the approach to sentiment quantification could be refined. In this study, sentiment was quantified by averaging sentiment scores for each category. This simplification can overlook the complexity and nuances of human emotions which extend beyond a binary positive and negative scale. Emotions such as sadness or anxiety expressed in tweets may have different impacts on travel decision-making processes (Chiou & Wan, 2006; Yang et al., 2018). Future research should consider adopting methods to capture these fine-grained emotions, such as creating training datasets labeled with different emotions, and utilising these datasets to train forecasting models using advanced econometric or machine learning algorithms.

Finally, the study incorporated all extracted categories into one model without considering potential canceling effects between topics. The assumption that categories derived from previous literature are mutually exclusive may not hold true in certain cases. For example, the study found a correlation of 0.82 between 'food' and 'exogenous factors' for Sydney, suggesting a potential canceling effect.

Furthermore, the study underscores an increasing trend of using social media data for tourism demand forecasting, and highlights a gap in studies connecting insights from social media data to other aggregate measures, such as tourist arrivals. Consequently, future research can consider developing or adopting advanced machine learning or econometric algorithms to forge links between unstructured social media data and aggregated variables. Such an approach could yield novel insights that are absent in traditional macro psychological and aggregated economic research.In addition, future research could consider focusing on specific sub-domains, such as attraction or geographic factors. The examination of more fine-grained dimensions may provide deeper insights into how various factors can influence tourism.

References

Ab Karim, S., & Chi, C. G.-q. (2010). Culinary Tourism as a Destination Attraction: An Empirical Examination of Destinations' Food Image. *Journal of Hospitality Marketing & Management*, *19*, 531 - 555.

Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. *Procedia Computer Science*, *189*, 191-194.

Adair, J. G. (1984). The Hawthorne effect: a reconsideration of the methodological artifact. *Journal of applied psychology*, *69*(2), 334.

Adeola, O., & Evans, O. (2020). ICT, infrastructure, and tourism development in Africa. *Tourism Economics*, *26*(1), 97-114.

Afshardoost, M., & Eshaghi, M. S. (2020). Destination image and tourist behavioural intentions: A meta-analysis. *Tourism Management*, *81*, 104154.

Afzaal, M., Usman, M., & Fong, A. (2019). Predictive aspect-based sentiment classification of online tourist reviews. *Journal of Information Science*, *45*(3), 341-363.

Aggarwal, C. C., & Yu, P. S. (2002). Redefining clustering for high-dimensional applications. *IEEE transactions on Knowledge and Data Engineering*, *14*(2), 210-225.

Agiomirgianakis, G., Serenis, D., & Tsounis, N. (2014). Exchange rate volatility and tourist flows into Turkey. *Journal of Economic Integration*, 700-725.

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, *40*(8), 1291-1315.

Ahmad, A., Jamaludin, A., Zuraimi, N. S. M., & Valeri, M. (2021). Visit intention and destination image in post-Covid-19 crisis recovery. *Current Issues in Tourism*, *24*(17), 2392-2397.

Ahuja, M. K., & Galvin, J. E. (2003). Socialization in virtual groups. *Journal of Management*, *29*(2), 161-185.

Akinboade, O. A., & Braimoh, L. A. (2010). International tourism and economic development in South Africa: A Granger causality test. *International Journal of Tourism Research*, *12*(2), 149-163.

Al-Mulali, U., Solarin, S. A., Andargoli, A. E., & Gholipour, H. F. (2021). Digital adoption and its impact on tourism arrivals and receipts. *Anatolia*, *32*(2), 337-339.

Alaei, A., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*, *58*, 175 - 191.

Alalwan, A. A., Rana, N. P., Algharabat, R., & Tarhini, A. (2016). A systematic review of extant literature in social media in the marketing perspective. Conference on e-Business, e-Services and e-Society,

Albalate, D., & Bel, G. (2010). Tourism and urban public transport: Holding demand pressure under supply constraints. *Tourism Management*, *31*(3), 425-433.

Albalate, D., Campos, J., & Jiménez, J. L. (2017). Tourism and high speed rail in Spain: Does the AVE increase local visitors? *Annals of Tourism Research*, *65*, 71-82.

Albayrak, T., & Caber, M. (2018). Examining the relationship between tourist motivation and satisfaction by two competing methods. *Tourism Management*, *69*, 201-213. https://doi.org/https://doi.org/10.1016/j.tourman.2018.06.015

Albu, C. E. (2013). STEREOTYPICAL FACTORS IN TOURISM -Literature review.

Alhemoud, A. M., & Armstrong, E. G. (1996). Image of tourism attractions in Kuwait. *Journal of Travel Research*, *34*(4), 76-80.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, *31*(2), 211-236.

Alrawadieh, Z., Dincer, M. Z., Dincer, F. I., & Mammadova, P. (2018). Understanding destination image from the perspective of Western travel bloggers: the case of Istanbul. *International Journal of Culture, Tourism and Hospitality Research*.

Alsaleh, N., & Farooq, B. (2021). Interpretable data-driven demand modelling for on-demand transit services. *Transportation Research Part A: Policy and Practice*, *154*, 1-22.

Ampountolas, A., & Legg, M. P. (2021). A segmented machine learning modeling approach of social media for predicting occupancy. *International Journal of Contemporary Hospitality Management*.

Andersson, T. D. (2007). The tourist in the experience economy. *Scandinavian Journal of Hospitality and Tourism*, *7*(1), 46-58.

Andersson, T. D., Mossberg, L., & Therkelsen, A. (2017). Food and tourism synergies: Perspectives on consumption, production and destination development. In (Vol. 17, pp. 1-8): Taylor & Francis.

Andersson, T. D., Mossberg, L. L., & Therkelsen, A. (2017). Food and tourism synergies: perspectives on consumption, production and destination development. *Scandinavian Journal of Hospitality and Tourism*, *17*, 1 - 8.

Apergis, N., Mervar, A., & Payne, J. E. (2017). Forecasting disaggregated tourist arrivals in Croatia:Evidence from seasonal univariate time series models. *Tourism Economics*, *23*(1), 78-98. https://doi.org/10.5367/te.2015.0499

Araña, J. E., & León, C. J. (2008). The impact of terrorism on tourism demand. *Annals of Tourism Research*, *35*(2), 299-315. https://doi.org/https://doi.org/10.1016/j.annals.2007.08.003

Argüelles, M. B., Coscarella, M., Fazio, A., & Bertellotti, M. (2016). Impact of whale-watching on the short-term behavior of Southern right whales (Eubalaena australis) in Patagonia, Argentina. *Tourism Management Perspectives*, *18*, 118-124.

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research*, *4*(3), 291-295.

Artus, J. R. (1970). The effect of revaluation on the foreign travel balance of Germany. *Staff Papers*, *17*(3), 602-619.

Asrin, K., Pouya, F., & Khalid, A. R. (2015). Modeling and forecasting of international tourism demand in ASEAN countries. *American Journal of Applied Sciences*, *12*(7), 479-486.

Assael, H. (1984). *Consumer behavior and marketing action*. Kent Pub. Co.

Assaker, G., & Hallak, R. (2013). Moderating effects of tourists' novelty-seeking tendencies on destination image, visitor satisfaction, and short-and long-term revisit intentions. *Journal of Travel Research*, *52*(5), 600-613.

Assaker, G., Vinzi, V. E., & O'Connor, P. (2011). Examining the effect of novelty seeking, satisfaction, and destination image on tourists' return pattern: A two factor, non-linear latent growth model. *Tourism Management*, *32*(4), 890-901.

Assent, I. (2012). Clustering high dimensional data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(4), 340-350.

Athanasopoulos, G., Song, H., & Sun, J. A. (2018). Bagging in tourism demand modeling and forecasting. *Journal of Travel Research*, *57*(1), 52-68.

Audrezet, A., de Kerviler, G., & Moulard, J. G. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of business research*, *117*, 557-569.

Averbuch, A., Coifman, R. R., Donoho, D. L., Elad, M., & Israeli, M. (2006). Fast and accurate polar Fourier transform. *Applied and computational harmonic analysis*, *21*(2), 145-167.

Bae, J., & Nam, S. (2020). An analysis of the effect of climate indicators on tourism demand: A case study of Jeju Island. *Journal of Policy Research in Tourism, Leisure and Events*, *12*(2), 185-196.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Baker, D. A., & Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of Tourism Research*, *27*(3), 785-804.

Balli, F., Balli, H. O., & Jean Louis, R. (2016). The impacts of immigrants and institutions on bilateral tourism flows. *Tourism Management*, *52*, 221-229. https://doi.org/https://doi.org/10.1016/j.tourman.2015.06.021

Baloglu, S. (2001). Image variations of Turkey by familiarity index: Informational and experiential dimensions. *Tourism Management*, *22*(2), 127-133.

Baloglu, S., & Brinberg, D. (1997). Affective images of tourism destinations. *Journal of Travel Research*, *35*(4), 11-15.

Baloglu, S., & McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research*, *26*(4), 868-897.

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. T. (2020). Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics*.

Barry, K., & O'Hagan, J. (1972). An econometric study of British tourist expenditure in Ireland. *Economic and Social Review*, *3*(2), 143-161.

Baud Bovy, M., & Lawson, F. (1977). *Tourism and recreation development*.

Baumann, R., & Matheson, V. (2018). Mega‐events and tourism: The case of Brazil. *Contemporary economic policy*, *36*(2), 292-301.

Baumann, R., & Matheson, V. A. (2018). Mega‐Events and Tourism: The Case of Brazil. *Political Economy: Structure & Scope of Government eJournal*.

Becken, S. (2013). Measuring the effect of weather on tourism: A destination-and activity-based analysis. *Journal of Travel Research*, *52*(2), 156-167.

Beerli, A., & Martin, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research*, *31*(3), 657-681.

Bell, D., & Valentine, G. (2013). *Consuming geographies: We are where we eat*. Routledge.

Bell, R. (2012). *Introductory Fourier transform spectroscopy*. Elsevier.

Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc.

Bhuta, S., Doshi, A., Doshi, U., & Narvekar, M. (2014). A review of techniques for sentiment analysis of twitter data. 2014 International conference on issues and challenges in intelligent computing techniques (ICICT), Bi, J.-W., Li, H., & Fan, Z.-P. (2021). Tourism demand forecasting with time series imaging: A deep learning model. *Annals of Tourism Research*, *90*, 103255.

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. https://doi.org/https://doi.org/10.1016/j.tourman.2018.10.013

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*(1), 48.

Blackwell, R. D., Miniard, P. W., & Engel, J. F. (2001). Consumer behavior 9th. *South-Western Thomas Learning. Mason, OH*.

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

Bojanic, D. C. (1991). The use of advertising in managing destination image. *Tourism Management*, *12*(4), 352-355.

Bolan, P., & Williams, L. (2008). The role of image in service promotion: focusing on the influence of film on consumer choice within tourism. *International Journal of Consumer Studies*, *32*, 382-390.

Borodin, A., Ostrovsky, R., & Rabani, Y. (1999). Subquadratic approximation algorithms for clustering problems in high dimensional spaces. Proceedings of the thirty-first annual ACM symposium on Theory of computing,

Botha, C., Crompton, J. L., & Kim, S.-S. (1999). Developing a revised competitive position for Sun/Lost city, South Africa. *Journal of Travel Research*, *37*(4), 341-352.

Boulding, K. E. (1956). *The image: Knowledge in life and society* (Vol. 47). University of Michigan press.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications*, *10*(1), 7.

Bramwell, B., & Rawding, L. (1996). Tourism marketing images of industrial cities. *Annals of Tourism Research*, *23*(1), 201-221.

Brännäs, K., & Nordström, J. (2006). Tourist Accommodation Effects of Festivals. *Tourism Economics*, *12*(2), 291-302. https://doi.org/10.5367/000000006777637458

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123-140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Briciu, A., & Briciu, V.-A. (2020). Participatory Culture and Tourist Experience: Promoting Destinations Through YouTube.

Brida, J. G., Lanzilotta, B., Moreno, L., & Santiñaque, F. (2018). A non-linear approximation to the distribution of total expenditure distribution of cruise tourists in Uruguay. *Tourism Management*, *69*, 62-68.

BrightLocal. (2020). *Local Consumer Review Survey 2020*. https://www.brightlocal.com/research/local-consumer-review-survey/

Britton, R. A. (1979). The image of the Third World in tourism marketing. *Annals of Tourism Research*, *6*(3), 318-329.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, *21*(3), 2-20.

Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of consumer research*, *14*(3), 350-362.

Bryman, A. (2016). *Social research methods*. Oxford university press.

Buhalis, D., Harwood, T., Bogicevic, V., Viglia, G., Beldona, S., & Hofacker, C. (2019). Technological disruptions in services: lessons from tourism and hospitality. *Journal of Service Management*, *30*(4), 484-506.

Bulmer, M. (1984). Facts, concepts, theories and problems. In *Sociological research methods* (pp. 37-50). Springer.

Bush, A. J., Smith, R., & Martin, C. (1999). The influence of consumer socialization variables on attitude toward advertising: A comparison of African-Americans and Caucasians. *Journal of Advertising*, *28*(3), 13-24.

Butterfield, D. W., Deal, K. R., & Kubursi, A. A. (1998). Measuring the returns to tourism advertising. *Journal of Travel Research*, *37*(1), 12-20.

Cankurt, S. (2016). Tourism demand forecasting using ensembles of regression trees. 2016 IEEE 8th International Conference on Intelligent Systems (IS),

Cankurt, S., & SUBAŞI, A. (2016). Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey. *Turkish Journal of Electrical Engineering & Computer Sciences*, *24*(5), 3388-3404.

Cárdenas-García, P. J., Sánchez-Rivero, M., & Pulido-Fernández, J. I. (2015). Does tourism growth influence economic development? *Journal of Travel Research*, *54*(2), 206-221.

Cavana, R., Delahaye, B., & Sekaran, U. (2001). Applied business research: Qualitative and quantitative methods. new york: John willey & sons. In: Inc.

Chan, Y.-M., Hui, T.-K., & Yuen, E. (1999). Modeling the impact of sudden environmental changes on visitor arrival forecasts: The case of the Gulf War. *Journal of Travel Research*, *37*(4), 391-394.

Chaudhary, R., Jindal, A., Aujla, G. S., Aggarwal, S., Kumar, N., & Choo, K.-K. R. (2019). BEST: Blockchain-based secure energy trading in SDN-enabled intelligent transportation system. *Computers & Security*, *85*, 288-299.

Chebli, A. (2020). The impact of Covis-19 on tourist consumption behaviour: a perspective article.

Chen, C.-F., Lai, M.-C., & Yeh, C.-C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, *26*, 281-287.

Chen, C.-F., & Phou, S. (2013). A closer look at destination: Image, personality, relationship and loyalty. *Tourism Management*, *36*, 269-278.

Chen, K.-Y., & Wang, C.-H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, *28*(1), 215-226.

Chen, L.-J., & Chen, W.-P. (2015). Push–pull factors in international birders' travel. *Tourism Management*, *48*, 416-425.

Chen, P.-J., & Kerstetter, D. L. (1999). International students’ image of rural Pennsylvania as a travel destination. *Journal of Travel Research*, *37*(3), 256-266.

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.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining,

Chen, Y., Lee, J.-Y., Sridhar, S., Mittal, V., McCallister, K., & Singal, A. G. (2020). Improving cancer outreach effectiveness through targeting and economic assessments: Insights from a randomized field experiment. *Journal of marketing*, *84*(3), 1-27.

Cheng, Z., Jian, S., Maghrebi, M., Rashidi, T. H., & Waller, S. T. (2018). Is social media an appropriate data source to improve travel demand estimation models? Transportation Research Board,

Chew, E. Y. T., & Jahari, S. A. (2014). Destination image as a mediator between perceived risks and revisit intention: A case of post-disaster Japan. *Tourism Management*, *40*, 382-393.

Chi, H.-K., Huang, K.-C., & Nguyen, H. M. (2020). Elements of destination brand equity and destination familiarity regarding travel intention. *Journal of Retailing and Consumer Services*, *52*, 101728.

Chiang, K., Wu, P., Chiang, S., Chang, T., Chang, S., & Wen, K. (1997). The introduction of grey theory. *Taipei: Gauli Publishing Co*.

Chinchanachokchai, S., & de Gregorio, F. (2020). A consumer socialization approach to understanding advertising avoidance on social media. *Journal of business research*, *110*, 474-483.

Chiou, W.-B., & Wan, C.-S. (2006). The effects of anxiety and sadness on travelers' decisions and perceived risk: Mood management as an active process of affect-adjustment. *ACR Asia-Pacific Advances*.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, *24*(3), 323-330.

Cho, V. (2010). A study of the non-economic determinants in tourism demand. *International Journal of Tourism Research*, *12*(4), 307-320. https://doi.org/10.1002/jtr.749

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, *28*(1), 118-129.

Chon, K.-S. (1991). Tourism destination image modification process: Marketing implications. *Tourism Management*, *12*(1), 68-72.

Chon, K. S. (1990). The role of destination image in tourism: A review and discussion. *The tourist review*.

Chu, F.-L. (1998). Forecasting tourism demand in Asian-Pacific countries. *Annals of Tourism Research*, *25*(3), 597-615.

Chu, F.-L. (2008). A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management*, *29*(1), 79-88.

Chu, S.-C., Deng, T., & Cheng, H. (2020). The role of social media advertising in hospitality, tourism and travel: a literature review and research agenda. *International Journal of Contemporary Hospitality Management*.

Cicourel, A. V. (1982). Interviews, surveys, and the problem of ecological validity. *The American Sociologist*, 11-20.

City of London. (2019). *Facts of Tourism 2019*. https://www.cityoflondon.gov.uk/assets/Things-to-do/facts-of-tourism-2019.pdf

Clawson, M., & Knetsch, J. L. (2013). *Economics of outdoor recreation*. RFF Press.

Clements, M. A., & Georgiou, A. (1998). The impact of political instability on a fragile tourism product. *Tourism Management*, *19*, 283-288.

Cohen, E., & Avieli, N. (2004). Food in tourism: Attraction and impediment. *Annals of Tourism Research*, *31*(4), 755-778.

Colicev, A., Kumar, A., & O'Connor, P. (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.

Colladon, A. F., Guardabascio, B., & Innarella, R. (2019). Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems*, *123*, 113075.

Colliander, J. (2019). “This is fake news”: Investigating the role of conformity to other users’ views when commenting on and spreading disinformation in social media. *Computers in Human Behavior*, *97*, 202-215.

Connell, J. (2005). Toddlers, tourism and Tobermory: Destination marketing issues and television-induced tourism. *Tourism Management*, *26*, 763-776.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297-301.

Copeland, L. S. (2008). *Exchange rates and international finance*. Pearson Education.

Court, B., & Lupton, R. A. (1997). Customer portfolio development: Modeling destination adopters, inactives, and rejecters. *Journal of Travel Research*, *36*(1), 35-43.

Cresswell, J., & Plano Clark, V. (2011). Designing and conducting mixed method research. 2nd Sage. *Thousand Oaks, CA*, *201*.

Creswell, J. W. (1999). Mixed-method research: Introduction and application. In *Handbook of educational policy* (pp. 455-472). Elsevier.

Crompton, J. L. (1979). An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image. *Journal of Travel Research*, *17*(4), 18-23.

Crompton, J. L., & Ankomah, P. K. (1993). Choice set propositions in destination decisions. *Annals of Tourism Research*, *20*(3), 461-476.

Crouch, G. I. (1994). The study of international tourism demand: A survey of practice. *Journal of Travel Research*, *32*(4), 41-55.

Crouch, G. I. (1995). A meta-analysis of tourism demand. *Annals of Tourism Research*, *22*(1), 103-118.

Cuccia, T., Guccio, C., & Rizzo, I. (2016). The effects of UNESCO World Heritage List inscription on tourism destinations performance in Italian regions. *Economic Modelling*, *53*, 494-508.

Dai, X., Karimi, S., Hachey, B., & Paris, C. (2020). Cost-effective selection of pretraining data: A case study of pretraining BERT on social media. *arXiv preprint arXiv:2010.01150*.

Dann, G. M. (1977). Anomie, ego-enhancement and tourism. *Annals of Tourism Research*, *4*(4), 184-194.

Dann, G. M. (1996). Tourists' images of a destination-an alternative analysis. *Journal of travel & tourism marketing*, *5*(1-2), 41-55.

Dann, G. M. S. (1981). Tourist motivation an appraisal. *Annals of Tourism Research*, *8*(2), 187-219. https://doi.org/https://doi.org/10.1016/0160-7383(81)90082-7

De Bruyn, A., & Lilien, G. L. (2008). A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing*, *25*(3), 151-163.

De Gregorio, F., & Sung, Y. (2010). Understanding attitudes toward and behaviors in response to product placement. *Journal of Advertising*, *39*(1), 83-96.

De Vita, G. (2014). The long-run impact of exchange rate regimes on international tourism flows. *Tourism Management*, *45*, 226-233.

De Vita, G., & Kyaw, K. S. (2013). Role of the exchange rate in tourism demand. *Annals of Tourism Research*, *43*, 624-627. https://doi.org/https://doi.org/10.1016/j.annals.2013.07.011

Dean, D., & Suhartanto, D. (2019). The formation of visitor behavioral intention to creative tourism: the role of push–Pull motivation. *Asia Pacific Journal of Tourism Research*, *24*(5), 393-403

Demiralay, S. (2020). Political uncertainty and the us tourism index returns. *Annals of Tourism Research*, *84*, 102875. https://doi.org/https://doi.org/10.1016/j.annals.2020.102875

Denstadli, J. M., Jacobsen, J. K. S., & Lohmann, M. (2011). Tourist perceptions of summer weather in Scandinavia. *Annals of Tourism Research*, *38*(3), 920-940.

Destination NSW. (2018). *Travel to Sydney Tourism Region*.

Destination NSW. (2019). *About Us*. Retrieved July from https://www.destinationnsw.com.au/about-us

Destination NSW. (2021). *State Tourism Satellite Account 2019-20*. https://www.destinationnsw.com.au/wp-content/uploads/2021/09/economic-contribution-of-tourism-to-nsw-2019-2020.pdf

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dhaoui, C., & Webster, C. M. (2020). Brand and consumer engagement behaviors on Facebook brand pages: Let's have a (positive) conversation. *International Journal of Research in Marketing*.

Díaz, M. á., & Mateu-Sbert, J. (2011). Forecasting daily air arrivals in Mallorca Island using nearest neighbour methods. *Tourism Economics*, *17*(1), 191-208.

Dogru, T., Sirakaya-Turk, E., & Crouch, G. I. (2017). Remodeling international tourism demand: Old theory and new evidence. *Tourism Management*, *60*, 47-55. https://doi.org/https://doi.org/10.1016/j.tourman.2016.11.010

Dolnicar, S., & Grün, B. (2013). Validly measuring destination image in survey studies. *Journal of Travel Research*, *52*(1), 3-14.

Dong, D., Xu, X., & Wong, Y. F. (2019). Estimating the Impact of Air Pollution on Inbound Tourism in China: An Analysis Based on Regression Discontinuity Design. *Sustainability*.

Duarte, P., Folgado-Fernández, J. A., & Hernández-Mogollón, J. M. (2018). Measurement of the Impact of Music Festivals on Destination Image: The Case of a Womad Festival. *Event Management*.

Duro, J. A., Perez-Laborda, A., Turrion-Prats, J., & Fernández-Fernández, M. (2021). Covid-19 and tourism vulnerability. *Tourism Management Perspectives*, *38*, 100819. https://doi.org/https://doi.org/10.1016/j.tmp.2021.100819

Dwivedi, M., Shibu, T., & Venkatesh, U. (2007). Social software practices on the Internet: Implications for the hotel industry. *International Journal of Contemporary Hospitality Management*.

Ebeling, R., Sáenz, C. A. C., Nobre, J., & Becker, K. (2021). The effect of political polarization on social distance stances in the brazilian covid-19 scenario. *Journal of Information and Data Management*, *12*(1).

Echtner, C. M., & Ritchie, J. R. B. (1993). The Measurement of Destination Image: An Empirical Assessment. *Journal of Travel Research*, *31*(4), 3-13. https://doi.org/10.1177/004728759303100402

Economist Intelligence Unit. (1975). Currency changes, exchange rates and their effects on tourism. *International Tourism Quarterly*, *1*, 46-52.

Edwards, M., Mitchell, L., Tuke, S. J., & Roughan, M. (2020). The one comparing narrative social network extraction techniques. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 905-913.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*.

Ellis, A., Park, E., Kim, S., & Yeoman, I. (2018). What is food tourism? *Tourism Management*, *68*, 250-263.

Engel, J. F., Kollat, D. J., & Blackwell, R. D. (1968). *Consumer behaviour*. Holt, Rinehart and Winston.

Eriksson, N., & Fagerstrøm, A. (2018). The Relative Impact of Wi-Fi Service on Young Consumers' Hotel Booking Online. *Journal of Hospitality & Tourism Research*, *42*, 1152 - 1169.

Eslami, S. P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? A multi-method investigation. *Decision Support Systems*, *113*, 32-42.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. kdd,

Eugenio-Martin, J. L., Cazorla-Artiles, J. M., & González-Martel, C. (2019). On the determinants of Airbnb location and its spatial distribution. *Tourism Economics*, *25*(8), 1224-1244.

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.

Fakeye, P. C., & Crompton, J. L. (1991). Image differences between prospective, first-time, and repeat visitors to the Lower Rio Grande Valley. *Journal of Travel Research*, *30*(2), 10-16.

Feick, L. F., & Price, L. L. (1987). The market maven: A diffuser of marketplace information. *Journal of marketing*, *51*(1), 83-97.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82-89.

Feng, Y., Li, G., Sun, X., & Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. *Procedia Computer Science*, *162*, 628-633.

Fernández-Morales, A., & Cisneros-Martínez, J. D. (2019). Seasonal concentration decomposition of cruise tourism demand in southern Europe. *Journal of Travel Research*, *58*(8), 1389-1407.

Ferraro, R., Bettman, J. R., & Chartrand, T. L. (2009). The power of strangers: The effect of incidental consumer brand encounters on brand choice. *Journal of consumer research*, *35*(5), 729-741.

Fletcher-Brown, J., Turnbull, S., Viglia, G., Chen, T., & Pereira, V. (2020). Vulnerable consumer engagement: how corporate social media can facilitate the replenishment of depleted resources. *International Journal of Research in Marketing*.

Folgado-Fernández, J. A., Hernández-Mogollón, J. M., & Duarte, P. (2017). Destination image and loyalty development: the impact of tourists' food experiences at gastronomic events. *Scandinavian Journal of Hospitality and Tourism*, *17*, 110 - 192.

Foo, L.-P., Chin, M.-Y., Tan, K.-L., & Phuah, K.-T. (2021). The impact of COVID-19 on tourism industry in Malaysia. *Current Issues in Tourism*, *24*(19), 2735-2739. https://doi.org/10.1080/13683500.2020.1777951

Fourie, J., & Santana-Gallego, M. (2010). The impact of mega-events on tourist arrivals.

Fourie, J., & Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management*, *32*(6), 1364-1370.

Frey, B. S. (1994). The economics of music festivals. *Journal of Cultural Economics*, *18*, 29-39.

Fridgen, J. D. (1987). Use of cognitive maps to determine perceived tourism regions. *Leisure Sciences*, *9*(2), 101-117.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics New York.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

Fröberg, E., & Rosengren, S. (2020). Man (ager Heuristics) vs. Machine (Learning): Automation for Prediction of Customer Value for Brick-and-Mortar Retailers. *Machine (Learning): Automation for Prediction of Customer Value for Brick-and-Mortar Retailers (October 27, 2020)*.

Fyall, A., & Rakic, T. (2006). The future market for World Heritage sites. *Managing world heritage sites*, 159-175.

Gabarda-Mallorquí, A., Garcia, X., & Ribas, A. (2017). Mass tourism and water efficiency in the hotel industry: A case study. *International Journal of Hospitality Management*, *61*, 82-93.

Gao, Y., & Su, W. (2019). Is the World Heritage just a title for tourism? *Annals of Tourism Research*, *78*, 102748. https://doi.org/https://doi.org/10.1016/j.annals.2019.102748

Gao, Y., Su, W., & Wang, K. (2019). Does high-speed rail boost tourism growth? New evidence from China. *Tourism Management*, *72*, 220-231.

Garau-Vadell, J. B., Gutierrez-Taño, D., & Diaz-Armas, R. (2018). Economic crisis and residents' perception of the impacts of tourism in mass tourism destinations. *Journal of destination marketing & management*, *7*, 68-75.

García, M. N., Munoz-Gallego, P. A., Viglia, G., & Gonzalez-Benito, O. (2019). Be social! The impact of self-presentation on peer-to-peer accommodation revenue. *Journal of Travel Research*.

Gartner, W. C. (1989). Tourism image: Attribute measurement of state tourism products using multidimensional scaling techniques. *Journal of Travel Research*, *28*(2), 16-20.

Gartner, W. C. (1994). Image formation process. *Journal of travel & tourism marketing*, *2*(2-3), 191-216.

Gartner, W. C. (1996). *Tourism development: Principles, processes, and policies*. Wiley.

Gavilan, D., Avello, M., & Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, *66*, 53-61.

Gedeon, T. D. (1997). Data mining of inputs: analysing magnitude and functional measures. *International Journal of Neural Systems*, *8*(02), 209-218.

George, A., & Ioana, U. C. (2007). Forecasting tourism demand using ANFIS for assuaring successful strategies in the view of sustainable development in the tourism sector.

Gerakis, A. S. (1965). Effects of exchange-rate devaluations and revaluations on receipts from tourism. *Staff Papers*, *12*(3), 365-384.

Geurts, M. D., & Ibrahim, I. (1975). Comparing the Box-Jenkins approach with the exponentially smoothed forecasting model application to Hawaii tourists. *Journal of Marketing Research*, *12*(2), 182-188.

Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2015). Using forum and search data for sales prediction of high-involvement products. Tomer Geva, Gal Oestreicher-Singer, Niv Efron, Yair Shimshoni." Using Forum and Search Data for Sales Prediction of High-Involvement Products"-MIS Quarterly, Forthcoming,

Ghalia, T., Fidrmuc, J., Samargandi, N., & Sohag, K. (2019). Institutional quality, political risk and tourism. *Tourism Management Perspectives*, *32*, 100576. https://doi.org/https://doi.org/10.1016/j.tmp.2019.100576

Gnoth, J. (1997). Tourism motivation and expectation formation. *Annals of Tourism Research*, *24*(2), 283-304.

Goh, C. (2012). Exploring impact of climate on tourism demand. *Annals of Tourism Research*, *39*(4), 1859-1883.

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, *23*(5), 499-510.

Goh, K.-Y., Heng, C.-S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information systems research*, *24*(1), 88-107.

Gonzalo, F. (2012). The five challenges of social media management in tourism. *Téléchargeable à https://fredericgonzalo. com/en/2012/05/01/the-five-challengesof-social-media-management-in-tourism*.

Goodrich, J. N. (1978). The relationship between preferences for and perceptions of vacation destinations: Application of a choice model. *Journal of Travel Research*, *17*(2), 8-13.

Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*, *29*(1), 1-20.

Gössling, S., Scott, D., Hall, C. M., Ceron, J.-P., & Dubois, G. (2012). Consumer behaviour and demand response of tourists to climate change. *Annals of Tourism Research*, *39*(1), 36-58.

Goulias, K., Lee, J. H., Deutsch-Burgner, K., Davis, A. W., & Wu, F. (2015). Detection and Measurement of Latent and Manifest Heterogeneity of Familiarity, Perceptions, and Attractiveness of Places Using Multilevel Analysis. International Choice Modelling Conference 2015,

Govers, R., Go, F. M., & Kumar, K. (2007). Promoting tourism destination image. *Journal of Travel Research*, *46*(1), 15-23.

Gray, H. P. (1966). The demand for international travel by the United States and Canada. *International Economic Review*, *7*(1), 83-92.

Grootendorst, M. (2020). *BERTopic: leveraging BERT and c-TF-IDF to create easily interpretable topics* In (Version v0.9.2) Zenodo. https://doi.org/10.5281/zenodo.4381785

Gross, M. J., & Brown, G. (2008). An empirical structural model of tourists and places: Progressing involvement and place attachment into tourism. *Tourism Management*, *29*(6), 1141-1151.

Guenther, J., & Falk, I. (2019). Generalising from Qualitative Research (GQR): A new old approach. *The Qualitative Report*, *24*(5), 1012-1033.

Gunter, U., & Önder, I. (2021). An exploratory analysis of geotagged photos from Instagram for residents of and visitors to Vienna. *Journal of Hospitality & Tourism Research*, *45*(2), 373-398.

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. https://doi.org/10.1177/1354816618793765

Gursoy, D., & Kendall, K. (2006). Hosting mega events: Modeling locals' support. *Annals of Tourism Research*, *33*(3), 603-623.

H2O.AI. (2020). *Variable Importance*. Retrieved MAY 2021 from https://docs.h2o.ai/h2o/latest-stable/h2o-docs/variable-importance.html

Haarhoff, R. (2018). Tourist perceptions of factors influencing destination image: a case study of selected Kimberley resorts. *African Journal of Hospitality, Tourism and Leisure*, *7*(4), 1-21.

Hajli, N., Sims, J., Zadeh, A. H., & Richard, M.-O. (2017). A social commerce investigation of the role of trust in a social networking site on purchase intentions. *Journal of business research*, *71*, 133-141.

Hall, C. M. (1999). Tourism And Politics: Policy, Power And Place.

Hall, C. M. (2001). Trends in ocean and coastal tourism: the end of the last frontier? *Ocean & coastal management*, *44*(9-10), 601-618.

Han, J. (2019). Vacationers in the countryside: Traveling for tranquility? *Tourism Management*, *70*, 299-310. https://doi.org/https://doi.org/10.1016/j.tourman.2018.09.001

Han, T., Gois, F. N. B., Oliveira, R., Prates, L. R., & de Almeida Porto, M. M. (2021). Modeling the progression of COVID-19 deaths using Kalman Filter and AutoML. *Soft Computing*, 1-16.

Hanon, W., & Wang, E. (2020). Comparing the impact of political instability and terrorism on inbound tourism demand in Syria before and after the political crisis in 2011. *Asia Pacific Journal of Tourism Research*, *25*(6), 651-661. https://doi.org/10.1080/10941665.2020.1752750

Hanussek, M., Blohm, M., & Kintz, M. (2020). Can AutoML outperform humans? An evaluation on popular OpenML datasets using AutoML benchmark. *arXiv preprint arXiv:2009.01564*.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*(1), 20-38.

Hass, R. G. (1981). Effects of source characteristics on cognitive responses in persuasion. *Cognitive responses in persuasion*, 141-172.

Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, *63*, 112-127.

Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: its use by national tourism organisations. *Current Issues in Tourism*, *16*(3), 211-239.

Heideman, M., Johnson, D., & Burrus, C. (1984). Gauss and the history of the fast Fourier transform. *IEEE ASSP Magazine*, *1*(4), 14-21.

Heimbach, I., & Hinz, O. (2016). The impact of content sentiment and emotionality on content virality. *International Journal of Research in Marketing*, *33*(3), 695-701.

Hem, L. E., Iversen, N. M., & Grønhaug, K. (2003). Advertising Effects of Photos Used to Portray Nature-Based Tourism Attractions. *Scandinavian Journal of Hospitality and Tourism*, *3*, 48 - 70.

Henderson, J. C. (2009). Food tourism reviewed. *British food journal*.

Hennig-Thurau, T., Henning, V., & Sattler, H. (2007). Consumer file sharing of motion pictures. *Journal of marketing*, *71*(4), 1-18.

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*(3), 375-394.

Hirche, M., Farris, P. W., Greenacre, L., Quan, Y., & Wei, S. (2021). Predicting Under-and Overperforming SKUs within the Distribution–Market Share Relationship. *Journal of Retailing*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Hong, I. B., & Cha, H. S. (2013). The mediating role of consumer trust in an online merchant in predicting purchase intention. *International Journal of Information Management*, *33*(6), 927-939.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, *2*(5), 359-366.

Howard, J. A., & Sheth, J. N. (1969). The theory of buyer behavior. *New York*, *63*.

Hsu, C. H., Cai, L. A., & Li, M. (2010). Expectation, motivation, and attitude: A tourist behavioral model. *Journal of Travel Research*, *49*(3), 282-296.

Hsu, C. H., & Huang, S. (2008). Travel motivation: A critical review of the concept's development. *Tourism management: Analysis, behaviour and strategy*, 14-27.

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., Xiao, M., & Li, H. (2021). Which search queries are more powerful in tourism demand forecasting: searches via mobile device or PC? *International Journal of Contemporary Hospitality Management*.

Hu, Y.-C. (2021). Forecasting tourism demand using fractional grey prediction models with Fourier series. *Annals of Operations Research*, *300*, 467-491.

Huang, M. Y., Rojas, R. R., & Convery, P. D. (2018). News Sentiment as Leading Indicators for Recessions. *arXiv: Applications*.

Huang, S., & Hsu, C. H. (2009). Effects of travel motivation, past experience, perceived constraint, and attitude on revisit intention. *Journal of Travel Research*, *48*(1), 29-44.

Hudders, L., De Jans, S., & De Veirman, M. (2021). The commercialization of social media stars: a literature review and conceptual framework on the strategic use of social media influencers. *International Journal of Advertising*, *40*(3), 327-375.

Hudson, S., & Thal, K. (2013). The impact of social media on the consumer decision process: Implications for tourism marketing. *Journal of travel & tourism marketing*, *30*(1-2), 156-160.

Inglis, A., Parnell, A., & Hurley, C. (2021). Visualizing Variable Importance and Variable Interaction Effects in Machine Learning Models. *arXiv preprint arXiv:2108.04310*.

Isard, P. (1995). *Exchange rate economics*. Cambridge University Press.

Iso-Ahola, S. E. (1982). Toward a social psychological theory of tourism motivation: A rejoinder. *Annals of Tourism Research*, *9*(2), 256-262. https://doi.org/https://doi.org/10.1016/0160-7383(82)90049-4

Iwashita, C. (2008). Roles of Films and Television Dramas in International Tourism: The Case of Japanese Tourists to the UK. *Journal of travel & tourism marketing*, *24*, 139 - 151.

Jalilvand, M. R., & Heidari, A. (2017). Comparing face-to-face and electronic word-of-mouth in destination image formation. *Information Technology & People*.

Jenkins, O. H. (1999). Understanding and measuring tourist destination images. *International Journal of Tourism Research*, *1*(1), 1-15.

Jensen, J. M. (2011). The relationships between socio-demographic variables, travel motivations and subsequent choice of vacation. 2nd International Conference on Economics, Business and Management,

Jeong, C., & Holland, S. (2012). Destination image saturation. *Journal of travel & tourism marketing*, *29*(6), 501-519.

Johnson, P. (2004). Analytic induction. *Essential guide to qualitative methods in organizational research*, *165*, 179.

Kahle, L. R., & Kennedy, P. (1988). Using the list of values (LOV) to understand consumers. *Journal of Services Marketing*, *2*(4), 49-56.

Kaján, E., & Saarinen, J. (2013). Tourism, climate change and adaptation: A review. *Current Issues in Tourism*, *16*(2), 167-195.

Karl, M. (2018). Risk and uncertainty in travel decision-making: Tourist and destination perspective. *Journal of Travel Research*, *57*(1), 129-146.

Katsikari, C., Hatzithomas, L., Fotiadis, T., & Folinas, D. (2020). Push and pull travel motivation: segmentation of the Greek market for social media marketing in tourism. *Sustainability*, *12*(11), 4770.

Kay, S., Mulcahy, R., & Parkinson, J. (2020). When less is more: the impact of macro and micro social media influencers' disclosure. *Journal of Marketing Management*, *36*(3-4), 248-278.

Keintz, R. M. (1968). A study of the demand for international travel to and from the United States. *Travel Research Bulletin*, *7*(1), 6-10.

Kendall, G., Chan, J. H. T., Yeung, M. C. H., & Law, K. K. (2020). Do film festivals attract tourists? *Current Issues in Tourism*, *24*, 1482 - 1486.

Khadaroo, J., & Seetanah, B. (2008). The role of transport infrastructure in international tourism development: A gravity model approach. *Tourism Management*, *29*(5), 831-840.

Khatibi, A., Belem, F., Silva, A. P., Shasha, D., & Goncalves, M. A. (2018). Improving Tourism Prediction Models Using Climate and Social Media Data: A Fine-Grained Approach. Twelfth International AAAI Conference on Web and Social Media,

Khoshnevis Yazdi, S., & Khanalizadeh, B. (2017). Tourism demand: a panel data approach. *Current Issues in Tourism*, *20*(8), 787-800. https://doi.org/10.1080/13683500.2016.1170772

Kim, C., Lee, H., & Tomiuk, M. A. (2009). Adolescents' perceptions of family communication patterns and some aspects of their consumer socialization. *Psychology & Marketing*, *26*(10), 888-907.

Kim, D.-Y., Hwang, Y.-H., & Fesenmaier, D. R. (2005). Modeling tourism advertising effectiveness. *Journal of Travel Research*, *44*(1), 42-49.

Kim, J. H., Wong, K., Athanasopoulos, G., & Liu, S. (2011). Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. *International Journal of forecasting*, *27*(3), 887-901.

Kim, S., & Kim, S. (2018). Perceived values of TV drama, audience involvement, and behavioral intention in film tourism. *Journal of travel & tourism marketing*, *35*, 259 - 272.

Kim, S., Lehto, X., & Kandampully, J. (2019). The role of familiarity in consumer destination image formation. *Tourism Review*.

King, C. W., & Summers, J. O. (1970). Overlap of opinion leadership across consumer product categories. *Journal of Marketing Research*, *7*(1), 43-50.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, *57*(8), 1012-1025.

Kirkwood, J. (2009). Motivational factors in a push ‐ pull theory of entrepreneurship. *Gender in Management: An International Journal*.

Kislali, H., Kavaratzis, M., & Saren, M. (2020). Destination image formation: Towards a holistic approach. *International Journal of Tourism Research*, *22*(2), 266-276.

Kivela, J., & Crotts, J. C. (2006). Tourism and Gastronomy: Gastronomy's Influence on How Tourists Experience a Destination. *Journal of Hospitality & Tourism Research*, *30*, 354 - 377.

Klenosky, D. B., Gengler, C. E., & Mulvey, M. S. (1993). Understanding the factors influencing ski destination choice: A means-end analytic approach. *Journal of leisure research*, *25*(4), 362-379.

Kock, F., Josiassen, A., & Assaf, A. G. (2016). Advancing destination image: The destination content model. *Annals of Tourism Research*, *61*, 28-44. https://doi.org/https://doi.org/10.1016/j.annals.2016.07.003

Köhler, C. F., Rohm, A. J., de Ruyter, K., & Wetzels, M. (2011). Return on interactivity: The impact of online agents on newcomer adjustment. *Journal of marketing*, *75*(2), 93-108.

Koo, T. T., Lim, C., & Dobruszkes, F. (2017). Causality in direct air services and tourism demand. *Annals of Tourism Research*, *67*, 67-77.

Koumchatzky, N., & Andryeyev, A. (2017). Using Deep Learning at Scale in Twitter’s Timelines. https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines

Kožić, I. (2014). Detecting international tourism demand growth cycles. *Current Issues in Tourism*, *17*(5), 397-403.

Kozinets, R. V., De Valck, K., Wojnicki, A. C., & Wilner, S. J. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of marketing*, *74*(2), 71-89.

Krajňák, T. (2020). The effects of terrorism on tourism demand: A systematic review. *Tourism Economics*, 1354816620938900.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689-702.

Kulendran, N., & Dwyer, L. (2009). Measuring the return from Australian tourism marketing expenditure. *Journal of Travel Research*, *47*(3), 275-284.

Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research*, *83*, 102925.

Kumpu, J., Pesonen, J., & Heinonen, J. (2021). Measuring the value of social media marketing from a destination marketing organization perspective. In *Information and Communication Technologies in Tourism 2021* (pp. 365-377). Springer.

Kuzey, C., Karaman, A. S., & Akman, E. (2019). Elucidating the impact of visa regimes: A decision tree analysis. *Tourism Management Perspectives*, *29*, 148-156.

Lanouar, C., & Goaied, M. (2019). Tourism, terrorism and political violence in Tunisia: Evidence from Markov-switching models. *Tourism Management*, *70*, 404-418. https://doi.org/https://doi.org/10.1016/j.tourman.2018.09.002

Latané, B. (1981). The psychology of social impact. *American psychologist*, *36*(4), 343.

Latorre-Martínez, M. P., Iñíguez-Berrozpe, T., & Plumed-Lasarte, M. (2014). Image-focused social media for a market analysis of tourism consumption. *Int. J. Technol. Manag.*, *64*, 17-30.

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.

Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. European Conference on Computer Vision,

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

LeDell, E., & Poirier, S. (2020). H2o automl: Scalable automatic machine learning. Proceedings of the AutoML Workshop at ICML,

Lee, C.-K., Var, T., & Blaine, T. W. (1996). Determinants of inbound tourist expenditures. *Annals of Tourism Research*, *23*(3), 527-542.

Lee, H.-S. (2015). Measurement of visitors' satisfaction with public zoos in Korea using importance-performance analysis. *Tourism Management*, *47*, 251-260.

Lee, S., Jeon, S., & Kim, D. (2011). The impact of tour quality and tourist satisfaction on tourist loyalty: The case of Chinese tourists in Korea. *Tourism Management*, *32*(5), 1115-1124.

Lee, T. H. (2009). A structural model to examine how destination image, attitude, and motivation affect the future behavior of tourists. *Leisure Sciences*, *31*(3), 215-236.

Leist, A. K., Klee, M., Kim, J. H., Rehkopf, D. H., Bordas, S., Muniz-Terrera, G., & Wade, S. (2021). Machine learning in the social and health sciences. *arXiv preprint arXiv:2106.10716*.

Leung, X. Y., & Bai, B. (2013). How motivation, opportunity, and ability impact travelers' social media involvement and revisit intention. *Journal of travel & tourism marketing*, *30*(1-2), 58-77.

Levin, C. (1988). Art and the Sociological Ego: Value from a Psychoanalytic Point of View. In *Life after Postmodernism* (pp. 22-63). Springer.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, *44*(1), 82-99.

Li, H., Choi, J., Lee, S., & Ahn, J. H. (2020). Comparing BERT and XLNet from the Perspective of Computational Characteristics. *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, 1-4.

Li, H., Goh, C., Hung, K., & Chen, J. L. (2018). Relative Climate Index and Its Effect on Seasonal Tourism Demand. *Journal of Travel Research*, *57*(2), 178-192. https://doi.org/10.1177/0047287516687409

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

Li, H., Song, H., & Li, L. (2017). A dynamic panel data analysis of climate and tourism demand: Additional evidence. *Journal of Travel Research*, *56*(2), 158-171.

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, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, *83*, 104245. https://doi.org/https://doi.org/10.1016/j.tourman.2020.104245

Li, X., Li, H., Pan, B., & Law, R. (2021). Machine learning in Internet search query selection for tourism forecasting. *Journal of Travel Research*, *60*(6), 1213-1231.

Li, Z., Shu, H., Tan, T., Huang, S., & Zha, J. (2020). Does the demographic structure affect outbound tourism demand? A panel smooth transition regression approach. *Journal of Travel Research*, *59*(5), 893-908.

Liang, T.-P., Li, X., Yang, C.-T., & Wang, M. (2015). What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach. *International Journal of Electronic Commerce*, *20*(2), 236-260.

Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research*, *24*(4), 835-849.

Lim, C. (1999). A meta-analytic review of international tourism demand. *Journal of Travel Research*, *37*(3), 273-284.

Lim, C. (2004). The major determinants of Korean outbound travel to Australia. *Mathematics and Computers in Simulation*, *64*(3-4), 477-485.

Lim, C., & McAleer, M. (2001a). Forecasting tourist arrivals. *Annals of Tourism Research*, *28*(4), 965-977.

Lim, C., & McAleer, M. (2001b). Monthly seasonal variations: Asian tourism to Australia. *Annals of Tourism Research*, *28*(1), 68-82.

Lim, K. W., & Buntine, W. L. (2014). Twitter Opinion Topic Model: Extracting Product Opinions from Tweets by Leveraging Hashtags and Sentiment Lexicon.

Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management.

Lise, W., & Tol, R. S. J. (2001). Impact of Climate on Tourist Demand. *Climatic Change*, *55*, 429-449.

Lismont, J., Ram, S., Vanthienen, J., Lemahieu, W., & Baesens, B. (2018). Predicting interpurchase time in a retail environment using customer-product networks: An empirical study and evaluation. *Expert systems with applications*, *104*, 22-32.

Little, J. S. (1979). *International travel in the US balance of payments*. Research Department, Federal Reserve Bank of Boston.

Liu, A., Fan, D. X., & Qiu, R. T. (2021). Does culture affect tourism demand? A global perspective. *Journal of Hospitality & Tourism Research*, *45*(1), 192-214.

Liu, A., & Pratt, S. (2017). Tourism's vulnerability and resilience to terrorism. *Tourism Management*, *60*, 404-417.

Liu, A., Vici, L., Ramos, V., Giannoni, S., & Blake, A. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Europe team. *Annals of Tourism Research*, *88*, 103182.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Liu, H., Liu, W., & Wang, Y. (2021). A study on the influencing factors of tourism demand from mainland China to Hong Kong. *Journal of Hospitality & Tourism Research*, *45*(1), 171-191.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, *59*, 554-563.

Lohmann, G., & Netto, A. P. (2016). *Tourism theory: Concepts, models and systems*. Cabi.

Lu, R. (2012). Study on tourist attraction micro-blog marketing model—A case of China Zhejiang Wuzheng. The 13th International Joint World Cultural Tourism Conference Thailand,

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of marketing*, *77*(1), 87-103.

Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management*, *71*, 120-131.

Machado, L. P. (2010). Does destination image influence the length of stay in a tourism destination? *Tourism Economics*, *16*(2), 443-456.

Madrigal, R., & Kahle, L. R. (1994). Predicting vacation activity preferences on the basis of value-system segmentation. *Journal of Travel Research*, *32*(3), 22-28.

Magno, F., & Cassia, F. (2018). The impact of social media influencers in tourism. *Anatolia*, *29*(2), 288-290.

Mak, A. H. (2017). Online destination image: Comparing national tourism organisation's and tourists' perspectives. *Tourism Management*, *60*, 280-297.

Makarem, S. C., & Jae, H. (2016). Consumer boycott behavior: An exploratory analysis of twitter feeds. *Journal of consumer affairs*, *50*(1), 193-223.

Mallat, N. (2007). Exploring consumer adoption of mobile payments–A qualitative study. *The Journal of Strategic Information Systems*, *16*(4), 413-432.

Mariani, M., Styven, M. E., & Ayeh, J. K. (2019). Using Facebook for travel decision-making: an international study of antecedents. *International Journal of Contemporary Hospitality Management*.

Marques, J. A. L., Gois, F. N. B., Xavier-Neto, J., & Fong, S. J. (2021). Artificial Intelligence Prediction for the COVID-19 Data Based on LSTM Neural Networks and H2O AutoML. In *Predictive Models for Decision Support in the COVID-19 Crisis* (pp. 69-87). Springer.

Martin, C. A., & Witt, S. F. (1987). Tourism demand forecasting models: Choice of appropriate variable to represent tourists' cost of living. *Tourism Management*, *8*(3), 233-246. https://doi.org/https://doi.org/10.1016/0261-5177(87)90055-0

Martin, C. A., & Witt, S. F. (1988). Substitute prices in models of tourism demand. *Annals of Tourism Research*, *15*(2), 255-268. https://doi.org/https://doi.org/10.1016/0160-7383(88)90086-2

Martin, C. A., & Witt, S. F. (1989). Forecasting tourism demand: A comparison of the accuracy of several quantitative methods. *International Journal of forecasting*, *5*(1), 7-19.

Martín, M. B. G. (2005). Weather, climate and tourism a geographical perspective. *Annals of Tourism Research*, *32*(3), 571-591.

Martín-Santana, J. D., Beerli-Palacio, A., & Nazzareno, P. A. (2017). Antecedents and consequences of destination image gap. *Annals of Tourism Research*, *62*, 13-25.

Martineau, P. (1958). The personality of the retail store.

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.

Maslow, A. H. (1954). Motivation and personality. In: New York: Harper & Row.

Maslow, A. H. (1981). *Motivation and personality*. Prabhat Prakashan.

Mathieson, A., & Wall, G. (1982). *Tourism, economic, physical and social impacts*. Longman.

Matloka, J., & Buhalis, D. (2010). Destination marketing through user personalised content (UPC). In *Information and communication technologies in tourism 2010* (pp. 519-530). Springer.

Maxim, C. (2019). Challenges faced by world tourism cities–London's perspective. *Current Issues in Tourism*, *22*(9), 1006-1024.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology*, *67*(3), 267-277.

McCole, P., Ramsey, E., & Williams, J. (2010). Trust considerations on attitudes towards online purchasing: The moderating effect of privacy and security concerns. *Journal of business research*, *63*(9-10), 1018-1024.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McKercher, B. (2017). Do attractions attract tourists? A framework to assess the importance of attractions in driving demand. *International Journal of Tourism Research*, *19*(1), 120-125.

Mckercher, B., & Koh, E. (2017). Do attractions "attract" tourists? The case of Singapore. *International Journal of Tourism Research*, *19*, 661-671.

McKercher, B., & Mak, B. (2019). The impact of distance on international tourism demand. *Tourism Management Perspectives*, *31*, 340-347.

Mckercher, B., Mei, W. S., & Tse, T. S. M. (2006). Are Short Duration Cultural Festivals Tourist Attractions? *Journal of Sustainable Tourism*, *14*, 55 - 66.

McKercher, B., & Tse, T. S. (2012). Is intention to return a valid proxy for actual repeat visitation? *Journal of Travel Research*, *51*(6), 671-686.

McWilliams, E. G., & Crompton, J. L. (1997). An expanded framework for measuring the effectiveness of destination advertising. *Tourism Management*, *18*(3), 127-137.

Mello, M. d., Pack, A., & Sinclair, M. T. (2002). A system of equations model of UK tourism demand in neighbouring countries. *Applied Economics*, *34*(4), 509-521.

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. *Information & Management*, *54*(6), 771-785.

Middleton, V. T., Fyall, A., Morgan, M., Morgan, M., & Ranchhod, A. (2009). *Marketing in travel and tourism*. Routledge.

Mieczkowski, Z. (1985). The tourism climatic index: a method of evaluating world climates for tourism. *Canadian Geographer/Le Géographe Canadien*, *29*(3), 220-233.

Mikolov, T., Yih, W.-t., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies,

Milano, R., Baggio, R., & Piattelli, R. (2011). The effects of online social media on tourism websites. In *Information and communication technologies in tourism 2011* (pp. 471-483). Springer.

Milman, A., & Pizam, A. (1995). The role of awareness and familiarity with a destination: The central Florida case. *Journal of Travel Research*, *33*(3), 21-27.

Minazzi, R. (2015). *Social media marketing in tourism and hospitality*. Springer.

Mitchell, A. (1983). *The nine American lifestyles: Who we are and where we're going*. Scribner Book Company.

Money, R. B., Gilly, M. C., & Graham, J. L. (1998). Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan. *Journal of marketing*, *62*(4), 76-87.

Moore, W. (2010). The impact of climate change on Caribbean tourism demand. *Current Issues in Tourism*, *13*, 495 - 505.

Morley, C., Rosselló, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, *48*, 1-10.

Morley, C. L. (1992). A microeconomic theory of international tourism demand. *Annals of Tourism Research*, *19*(2), 250-267.

Morozov, M. A., & Morozova, N. S. (2016). Attractive tourist destinations as a factor of its development. *Journal of Environmental Management & Tourism*, *7*(1 (13)), 105.

Moschis, G. P., & Churchill Jr, G. A. (1978). Consumer socialization: A theoretical and empirical analysis. *Journal of Marketing Research*, *15*(4), 599-609.

Myers, J. G. (1968). *Consumer image and attitude*. Institute of Business and Economic Research.

Naudé, W. A., & Saayman, A. (2005). Determinants of tourist arrivals in Africa: a panel data regression analysis. *Tourism Economics*, *11*(3), 365-391.

Navarro-Ruiz, S., & McKercher, B. (2020). The usability of visitor attractions: State-of-the-art. *Tourism Review*.

Nepal, R., Al Irsyad, M. I., & Nepal, S. K. (2019). Tourist arrivals, energy consumption and pollutant emissions in a developing economy–implications for sustainable tourism. *Tourism Management*, *72*, 145-154.

Neumayer, E., & Plümper, T. (2016). Spatial spill-overs from terrorism on tourism: Western victims in Islamic destination countries. *Public Choice*, *169*(3), 195-206. https://doi.org/10.1007/s11127-016-0359-y

Nicolau, J. L. (2008). Characterizing tourist sensitivity to distance. *Journal of Travel Research*, *47*(1), 43-52.

Nicosia, F. M. (1966). CONSUMER DECISION PROCESSES; MARKETING AND ADVERTISING IMPLICATIONS.

Nikjoo, A. H., & Ketabi, M. (2015). The role of push and pull factors in the way tourists choose their destination. *Anatolia*, *26*(4), 588-597.

Noval, S. (1976). THE DEMAND FOR INTERNATIONAL TOURISM AND TRAVEL: THEORY AND MEASUREMENT.

Nunes, P., Cai, M., Ferrise, R., Moriondo, M., & Marco, B. (2013). An econometric analysis of climate change impacts on tourism flows: An empirical evidence from the region of Tuscany, Italy. *International Journal of Ecological Economics and Statistics*, *31*(4), 1-20.

Nykodym, T., Kraljevic, T., Wang, A., & Wong, W. (2016). Generalized linear modeling with h2o. *Published by H2O. ai Inc*.

O'sullivan, D. B., & Jackson, M. J. (2002). Festival Tourism: A Contributor to Sustainable Local Economic Development? *Journal of Sustainable Tourism*, *10*, 325 - 342.

Ofcom. (2020). *Adults' Media Use & Attitudes Report 2020*. https://www.ofcom.org.uk/__data/assets/pdf_file/0031/196375/adults-media-use-and-attitudes-2020-report.pdf

Okazaki, S., & Taylor, C. R. (2013). Social media and international advertising: theoretical challenges and future directions. *International marketing review*.

Önder, I. (2017). Classifying multi-destination trips in Austria with big data. *Tourism Management Perspectives*, *21*, 54-58.

Önder, I., Gunter, U., & Gindl, S. (2020). Utilizing Facebook statistics in tourism demand modeling and destination marketing. *Journal of Travel Research*, *59*(2), 195-208.

Önder, I., Gunter, U., & Scharl, A. (2019). Forecasting tourist arrivals with the help of web sentiment: A mixed-frequency modeling approach for big data. *Tourism Analysis*, *24*(4), 437-452.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. LREc,

Pan, Q., & Li, W. (2004). Review on the socio-cultural impacts of tourism on destination in China in recent years. *Economic Geography*, *24*(3), 412-415.

Papadimitriou, D., Kaplanidou, K., & Apostolopoulou, A. (2018). Destination image components and word-of-mouth intentions in urban tourism: A multigroup approach. *Journal of Hospitality & Tourism Research*, *42*(4), 503-527.

Park, S. B., Ok, C. M., & Chae, B. K. (2016). Using Twitter data for cruise tourism marketing and research. *Journal of travel & tourism marketing*, *33*(6), 885-898.

Patro, A. C., Zaidi, S. A., Dixit, A., & Dixit, M. (2021). A Novel Approach to Improve Employee Retention Using Machine Learning. 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT),

Pearce, P. (1982). The social psychology of tourist behaviour. *The social psychology of tourist behaviour.*

Pearce, P. L. (1988). The Ulysses factor: Evaluating visitors in tourist settings.

Pearce, P. L., & Caltabiano, M. L. (1983). Inferring travel motivation from travelers' experiences. *Journal of Travel Research*, *22*(2), 16-20.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, 181-193.

Peng, B., Song, H., Crouch, G. I., & Witt, S. F. (2015). A meta-analysis of international tourism demand elasticities. *Journal of Travel Research*, *54*(5), 611-633.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),

Pesonen, J., Komppula, R., Kronenberg, C., & Peters, M. (2011). Understanding the relationship between push and pull motivations in rural tourism. *Tourism Review*, *66*(3), 32-49.

Petrick, J. F. (2004). The roles of quality, value, and satisfaction in predicting cruise passengers' behavioral intentions. *Journal of Travel Research*, *42*(4), 397-407.

Pham, T. D., Nghiem, S., & Dwyer, L. (2017). The determinants of Chinese visitors to Australia: A dynamic demand analysis. *Tourism Management*, *63*, 268-276. https://doi.org/https://doi.org/10.1016/j.tourman.2017.06.015

Phau, I., Quintal, V., & Shanka, T. (2014). Examining a consumption values theory approach of young tourists toward destination choice intentions. *International Journal of Culture, Tourism and Hospitality Research*.

Pitts, R. E., & Woodside, A. G. (1986). Personal values and travel decisions. *Journal of Travel Research*, *25*(1), 20-25.

Pizam, A., & Calantone, R. (1987). Beyond psychographics—values as determinants of tourist behavior. *International Journal of Hospitality Management*, *6*(3), 177-181.

Plog, S. C. (1974). Why destination areas rise and fall in popularity. *Cornell hotel and restaurant administration quarterly*, *14*(4), 55-58.

Polyzos, S., Samitas, A., & Spyridou, A. E. (2020). Tourism demand and the COVID-19 pandemic: An LSTM approach. *Tourism Recreation Research*, 1-13.

Poole, M., Davis, G., & James, S. (1988). Trends and prospects for Australian international air transport. *OCCASIONAL PAPER, 88, ISSN: 0157-7085;-UNTRACED*(88).

Poon, A. (1993). *Tourism, technology and competitive strategies*. CAB international.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, *51*(5), 1-36.

Povee, K., & Roberts, L. D. (2014). Qualitative research in psychology: Attitudes of psychology students and academic staff. *Australian Journal of Psychology*, *66*(1), 28-37.

Preprocessor. (2020). *Preprocessor*. In https://github.com/s/preprocessor

Puan, C. L., & Zakaria, M. (2007). Perception of visitors towards the role of zoos: a Malaysian perspective. *International Zoo Yearbook*, *41*, 226-232.

Pyo, S. S., Uysal, M., & McLellan, R. W. (1991). A linear expenditure model for tourism demand. *Annals of Tourism Research*, *18*(3), 443-454.

Qiu, R. T., Liu, A., Stienmetz, J. L., & Yu, Y. (2021). Timing matters: crisis severity and occupancy rate forecasts in social unrest periods. *International Journal of Contemporary Hospitality Management*.

Qu, H., Kim, L. H., & Im, H. H. (2011). A model of destination branding: Integrating the concepts of the branding and destination image. *Tourism Management*, *32*(3), 465-476.

Qualman, E. (2012). *Socialnomics: How social media transforms the way we live and do business*. John Wiley & Sons.

Quayson, J., & Var, T. (1982). A tourism demand function for the Okanagan, BC. *Tourism Management*, *3*(2), 108-115.

Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: a review. *Journal of Research in Interactive Marketing*, *12*(2), 146-163.

Ramkissoon, H., & Uysal, M. S. (2011). The effects of perceived authenticity, information search behaviour, motivation and destination imagery on cultural behavioural intentions of tourists. *Current Issues in Tourism*, *14*(6), 537-562.

Ranjbarian, B., & Pool, J. K. (2015). The impact of perceived quality and value on tourists' satisfaction and intention to revisit Nowshahr city of Iran. *Journal of Quality Assurance in Hospitality & Tourism*, *16*(1), 103-117.

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, *41*, 547-566.

Reich, B. J., & Pittman, M. (2020). An Appeal to Intimacy: Consumer Response to Platform ‐ Appeal Fit on Social Media. *Journal of Consumer Psychology*.

Reid, E., & Duffy, K. (2018). A netnographic sensibility: Developing the netnographic/social listening boundaries. *Journal of Marketing Management*, *34*(3-4), 263-286.

Richards, G. (2012). An overview of food and tourism trends and policies.

Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P., & Martens, P. (2014). Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: Evidence from Aruba. *Tourism Management*, *41*, 245-256.

Rimmington, M., & Yüksel, A. (1998). Tourist satisfaction and food service experience: results and implications of an empirical investigation.

Ritchie, J. B., & Smith, B. H. (1991). The impact of a mega-event on host region awareness: A longitudinal study. *Journal of Travel Research*, *30*(1), 3-10.

Rockmore, D. N. (2000). The FFT: an algorithm the whole family can use. *Computing in Science & Engineering*, *2*(1), 60-64.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Rokeach, M. (1968). *Beliefs, attitudes and values; a theory of organization and change*.

Ross, G. F. (1993). Ideal and actual images of backpacker visitors to Northern Australia. *Journal of Travel Research*, *32*(2), 54-57.

Rousta, A., & Jamshidi, D. (2020). Food tourism value: Investigating the factors that influence tourists to revisit. *Journal of Vacation Marketing*, *26*(1), 73-95.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. *American sociological review*, 195-222.

Ryan, C. (1998). The travel career ladder an appraisal. *Annals of Tourism Research*, *25*(4), 936-957.

Ryan, C., & Saward, J. (2004). The Zoo as Ecotourism Attraction – Visitor Reactions, Perceptions and Management Implications: The Case of Hamilton Zoo, New Zealand. *Journal of Sustainable Tourism*, *12*, 245 - 266.

Saayman, A., & Saayman, M. (2013). Exchange rate volatility and tourism-revisiting the nature of the relationship. *European journal of tourism research*, *6*(2), 104-121.

Saha, S., Su, J.-J., & Campbell, N. (2017). Does Political and Economic Freedom Matter for Inbound Tourism? A Cross-National Panel Data Estimation. *Journal of Travel Research*, *56*(2), 221-234. https://doi.org/10.1177/0047287515627028

Saleh, F. A., & Ryan, C. (1993). Jazz and knitwear: Factors that attract tourists to festivals. *Tourism Management*, *14*, 289-297.

Samitas, A., Asteriou, D., Polyzos, S., & Kenourgios, D. (2018). Terrorist incidents and tourism demand: Evidence from Greece. *Tourism Management Perspectives*, *25*, 23-28. https://doi.org/https://doi.org/10.1016/j.tmp.2017.10.005

Sánchez, J. E. (1985). Por una geografía del turismo de litoral. Una aproximación metodológica. *Estudios territoriales*, *17*, 103-122.

Sánchez ‐ Cañizares, S. M., & López-Guzmán, T. (2012). Gastronomy as a tourism resource: profile of the culinary tourist. *Current Issues in Tourism*, *15*, 229 - 245.

Sánchez ‐ Franco, M. J., & Rey ‐ Moreno, M. (2022). Do travelers' reviews depend on the destination? An analysis in coastal and urban peer ‐ to ‐ peer lodgings. *Psychology & Marketing*, *39*(2), 441-459.

Sangwaan, R. (2019). USE OF SOCIAL MEDIA IN EDUCATION: POSITIVE AND NEGATIVE IMPACT ON THE STUDENTS.

Sannassee, R. V., & Seetanah, B. (2015). The influence of trust on repeat tourism: The Mauritian case study. *Journal of Hospitality Marketing & Management*, *24*(7), 770-789.

Santana-Gallego, M., Ledesma-Rodríguez, F. J., & Pérez-Rodríguez, J. V. (2010). Exchange rate regimes and tourism. *Tourism Economics*, *16*(1), 25-43.

Santana-Gallego, M., Ledesma-Rodríguez, F. J., & Pérez-Rodríguez, J. V. (2016). International trade and tourism flows: An extension of the gravity model. *Economic Modelling*, *52*, 1026-1033.

Santana-Gallego, M., Rosselló-Nadal, J., & Fourie, J. (2016). The effects of terrorism, crime and corruption on tourism. *Economic Research Southern Africa (ERSA)*, *595*, 1-28.

Santeramo, F. G., & Morelli, M. (2016). Modelling tourism flows through gravity models: A quantile regression approach. *Current Issues in Tourism*, *19*(11), 1077-1083.

Santos, G. E. d. O. (2016). An efficient method for modelling tourists' length of stay. *Tourism Economics*, *22*(6), 1367-1379.

Scarpino, I., Zucco, C., Vallelunga, R., Luzza, F., & Cannataro, M. (2022). Investigating topic modeling techniques to extract meaningful insights in Italian long COVID narration. *BioTech*, *11*(3), 41.

Schiff, A., & Becken, S. (2011). Demand elasticity estimates for New Zealand tourism. *Tourism Management*, *32*(3), 564-575.

Schmid, A. P. (2011). *The definition of terrorism*. Routledge.

Schmidt, K. N., & Iyer, M. K. S. (2015). Online behaviour of social media participants' and perception of trust, comparing social media brand community groups and associated organized marketing strategies. *Procedia-Social and Behavioral Sciences*, *177*, 432-439.

Schubert, S. F., Brida, J. G., & Risso, W. A. (2011). The impacts of international tourism demand on economic growth of small economies dependent on tourism. *Tourism Management*, *32*(2), 377-385.

Schultz, A. (1962). Collected parers, vol. 1: The problem of social reality (M. Natanson, Ed.). *Nijhoff: The Hague*.

Schwartz, Z., & Cohen, E. (2004). Hotel revenue-management forecasting: Evidence of expert-judgment bias. *Cornell hotel and restaurant administration quarterly*, *45*(1), 85-98.

Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, *51*(4), 387-402.

Scotland Insight Department. (2016). *Coastal Tourism in Scotland* (Topic Paper, Issue.

Seddighi, H. R., Nuttall, M. W., & Theocharous, A. L. (2001). Does cultural background of tourists influence the destination choice? an empirical study with special reference to political instability. *Tourism Management*, *22*(2), 181-191. https://doi.org/https://doi.org/10.1016/S0261-5177(00)00046-7

Seetanah, B., Juwaheer, T., Lamport, M. J., Rojid, S., Sannassee, R. V., & Subadar, A. (2011). Does Infrastructure Matter In Tourism Development. *University of Mauritius Research Journal*, *17*, 89-108.

Seiler, S., Yao, S., & Wang, W. (2017). Does online word of mouth increase demand?(and how?) evidence from a natural experiment. *Marketing Science*, *36*(6), 838-861.

Semeida, A. M. (2014). Derivation of travel demand forecasting models for low population areas: the case of Port Said Governorate, North East Egypt. *Journal of Traffic and Transportation Engineering (English Edition)*, *1*(3), 196-208.

Shafiullah, M., Okafor, L. E., & Khalid, U. (2019). Determinants of international tourism demand: Evidence from Australian states and territories. *Tourism Economics*, *25*(2), 274-296. https://doi.org/10.1177/1354816618800642

Shah, S. A., & Koltun, V. (2018). Deep continuous clustering. *arXiv preprint arXiv:1803.01449*.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, *228*(2), 221-243. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.07.053

Sharma, C., & Pal, D. (2020). Exchange rate volatility and tourism demand in India: unraveling the asymmetric relationship. *Journal of Travel Research*, *59*(7), 1282-1297.

Sharma, G. D., Thomas, A., & Paul, J. (2021). Reviving tourism industry post-COVID-19: A resilience-based framework. *Tourism Management Perspectives*, *37*, 100786. https://doi.org/https://doi.org/10.1016/j.tmp.2020.100786

Sharpley, R. (2018). *Tourism, tourists and society*. Routledge.

Shen, M.-L., Liu, H.-H., Lien, Y.-H., Lee, C.-F., & Yang, C.-H. (2019). Hybrid Approach for Forecasting Tourist Arrivals. Proceedings of the 2019 8th International Conference on Software and Computer Applications,

Shi, H., Xu, M., & Li, R. (2017). Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid*, *9*(5), 5271-5280.

Shim, S. (1996). Adolescent consumer decision ‐ making styles: The consumer socialization perspective. *Psychology & Marketing*, *13*(6), 547-569.

Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research*, *74*, 134-154.

Sinclair, M. T. (1998). Tourism and economic development: A survey. *The journal of development studies*, *34*(5), 1-51.

Sirakaya, E., McLellan, R. W., & Uysal, M. (1996). Modeling vacation destination decisions: A behavioral approach. *Journal of travel & tourism marketing*, *5*(1-2), 57-75.

Sirakaya, E., & Woodside, A. G. (2005). Building and testing theories of decision making by travellers. *Tourism Management*, *26*(6), 815-832.

Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the travel and tourism industry. *Technological Forecasting and Social Change*, *163*, 120469.

Skidmore, S., & Pyszka, R. (1987). US international pleasure travel market—a values perspective. *Tourism Management*, *8*(2), 121-122.

Solberg, H. A., & Preuss, H. (2007). Major sport events and long-term tourism impacts. *Journal of sport Management*, *21*(2), 213-234.

Song, H., Kim, J. H., & Yang, S. (2010). Confidence intervals for tourism demand elasticity. *Annals of Tourism Research*, *37*(2), 377-396.

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. (2019a). A review of research on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338-362.

Song, H., Qiu, R. T. R., & Park, J. (2019b). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, *75*, 338-362. https://doi.org/https://doi.org/10.1016/j.annals.2018.12.001

Song, H., & Turner, L. (2006). Tourism demand forecasting. *International handbook on the economics of tourism*, 89-114.

Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management*, *27*(2), 214-224.

Song, H., Witt, S. F., & Li, G. (2008). *The advanced econometrics of tourism demand*. Routledge.

Song, H., & Wong, K. K. (2003). Tourism demand modeling: A time-varying parameter approach. *Journal of Travel Research*, *42*(1), 57-64.

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.

Sonnier, G. P., McAlister, L., & Rutz, O. J. (2011). A dynamic model of the effect of online communications on firm sales. *Marketing Science*, *30*(4), 702-716.

Souiden, N., Ladhari, R., & Chiadmi, N. E. (2017). Destination personality and destination image. *Journal of Hospitality and Tourism Management*, *32*, 54-70.

Starosta, K., Budz, S., & Krutwig, M. (2019). The impact of German-speaking online media on tourist arrivals in popular tourist destinations for Europeans. *Applied Economics*, *51*(14), 1558-1573. https://doi.org/10.1080/00036846.2018.1527463

Statista. (2020). *Social media - Statistics & Facts*. Statista. Retrieved 23 MAY from https://www.statista.com/topics/1164/social-networks/#topicOverview

Stevenson, N., & Inskip, C. (2009). Seeing the Sites: Perceptions ofLondon. *City tourism: National capital perspectives*, 94.

Stylidis, D. (2020). Exploring Resident–Tourist Interaction and its Impact on Tourists' Destination Image. *Journal of Travel Research*, 0047287520969861.

Stylidis, D., Shani, A., & Belhassen, Y. (2017). Testing an integrated destination image model across residents and tourists. *Tourism Management*, *58*, 184-195.

Stylos, N., Vassiliadis, C. A., Bellou, V., & Andronikidis, A. (2016). Destination images, holistic images and personal normative beliefs: Predictors of intention to revisit a destination. *Tourism Management*, *53*, 40-60.

Stylos, N., Zwiegelaar, J. B., & Buhalis, D. (2021). Big data empowered agility for dynamic, volatile, and time-sensitive service industries: the case of tourism sector. *International Journal of Contemporary Hospitality Management*, *33*, 1015-1036.

Sugiartawan, P., Hartati, S., & Musdholifah, A. (2018). Tourist Visits Prediction with Fully Recurrent Neural Network. International Conference on Information Technology, Engineering, Science & its Applications,

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.

Sunday, A. A. (1978). Foreign travel and tourism prices and demand. *Annals of Tourism Research*, *5*(2), 268-273.

Suni, J., & Pesonen, J. (2019). Hunters as tourists–an exploratory study of push–pull motivations. *Scandinavian Journal of Hospitality and Tourism*, *19*(2), 175-191.

Swarbrooke, J., & Horner, S. (2007). *Consumer behaviour in tourism*. Routledge.

Tang, J., Yuan, X., Ramos, V., & Sriboonchitta, S. (2019). Does air pollution decrease inbound tourist arrivals? The case of Beijing. *Asia Pacific Journal of Tourism Research*, *24*, 597 - 605.

Tang, T., Fang, E., & Wang, F. (2014). Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of marketing*, *78*(4), 41-58.

Tanti, A., & Buhalis, D. (2017). The influences and consequences of being digitally connected and/or disconnected to travellers. *Information Technology & Tourism*, *17*, 121-141.

Tapachai, N., & Waryszak, R. (2000). An examination of the role of beneficial image in tourist destination selection. *Journal of Travel Research*, *39*(1), 37-44.

Tasci, A. D., & Gartner, W. C. (2007). Destination image and its functional relationships. *Journal of Travel Research*, *45*(4), 413-425.

Tasci, A. D. A., Gartner, W. C., & Tamer Cavusgil, S. (2007). Conceptualization and Operationalization of Destination Image. *Journal of Hospitality & Tourism Research*, *31*(2), 194-223. https://doi.org/10.1177/1096348006297290

Taspinar, A., & Schuirmann, L. (2017). Twitterscraper 0.2. 7: Python Package Index. In.

Tavares, J. M., & Leitão, N. C. (2017). The determinants of international tourism demand for Brazil. *Tourism Economics*, *23*(4), 834-845. https://doi.org/10.5367/te.2016.0540

Taylor, D. G., Lewin, J. E., & Strutton, D. (2011). Friends, fans, and followers: do ads work on social networks?: how gender and age shape receptivity. *Journal of advertising research*, *51*(1), 258-275.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

Ten Wong, D. H., Phang, C. S., Maarop, N., Samy, G. N., Ibrahim, R., Yusoff, R. C. M., Magalingam, P., & Azmi, N. F. M. (2017). Effect of social media on human interpersonal communication: A Review. *Open International Journal of Informatics (OIJI)*, *5*(2), 1-6.

Tezer, A., & Bodur, H. O. (2020). The Greenconsumption Effect: How Using Green Products Improves Consumption Experience. *Journal of consumer research*, *47*(1), 25-39.

Tham, A., Croy, G., & Mair, J. (2013). Social media in destination choice: Distinctive electronic word-of-mouth dimensions. *Journal of travel & tourism marketing*, *30*(1-2), 144-155.

The Guardian. (2014). *Sydney cafe siege: a timeline of events*. The Guardian,. Retrieved 12 Mar from https://www.theguardian.com/australia-news/2014/dec/15/sydney-cafe-siege-timeline-of-events

Thompson, S. A., & Sinha, R. K. (2008). Brand communities and new product adoption: The influence and limits of oppositional loyalty. *Journal of marketing*, *72*(6), 65-80.

Thushara, S., Su, J.-J., & Bandara, J. S. (2019). Forecasting international tourist arrivals in formulating tourism strategies and planning: The case of Sri Lanka. *Cogent Economics & Finance*, *7*(1), 1699884.

Tian, F., Yang, Y., Mao, Z., & Tang, W. (2021). Forecasting daily attraction demand using big data from search engines and social media. *International Journal of Contemporary Hospitality Management*.

Tikkanen, I. (2007). Maslow's hierarchy and food tourism in Finland: five cases. *British food journal*.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198-215.

Tonga Uriarte, Y. i., Antognozzi, T., & Catoni, M. L. (2019). Investigating Tourism Impacts of Festivals: An Exploratory Case Study of a Big-Scale Comic-Con. *Event Management*, *23*, 817-833.

Toral, S., Martínez-Torres, M., & Gonzalez-Rodriguez, M. (2018). Identification of the unique attributes of tourist destinations from online reviews. *Journal of Travel Research*, *57*(7), 908-919.

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.

Trauer, B. (2006). Conceptualizing special interest tourism—frameworks for analysis. *Tourism Management*, *27*(2), 183-200.

TripAdvisor. (2016). *TripBarometer 2016*.

Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C. B., & Farivar, R. (2019). Towards automated machine learning: Evaluation and comparison of AutoML approaches and tools. 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI),

Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. *Journal of Marketing Research*, *47*(4), 643-658.

Tsui, W. H. K., & Balli, F. (2017). International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. *Tourism Economics*, *23*(2), 403-428.

Tussyadiah, I. P., & Fesenmaier, D. R. (2009). Mediating tourist experiences: access to places via shared videos. *Annals of Tourism Research*, *36*, 24-40.

Um, S., & Crompton, J. L. (1990). Attitude determinants in tourism destination choice. *Annals of Tourism Research*, *17*(3), 432-448.

Usher, K., Woods, C., Casella, E., Glass, N., Wilson, R., Mayner, L., Jackson, D., Brown, J., Duffy, E., & Mather, C. (2014). Australian health professions student use of social media. *Collegian*, *21*(2), 95-101.

Uysal, M. (1998). The determinants of tourism demand. *The economic geography of the tourist industry: A supply-side analysis*, *79*.

Uysal, M., & Crompton, J. L. (1984). Determinants of demand for international tourist flows to Turkey. *Tourism Management*, *5*(4), 288-297.

Uysal, M., & Crompton, J. L. (1985). Deriving a relative price index for inclusion in international tourism demand estimation models. *Journal of Travel Research*, *24*(1), 32-34.

Uysal, M., Li, X., & Sirakaya-Turk, E. (2008). Push-pull dynamics in travel decisions. *Handbook of hospitality marketing management*, *2009*, 412-439.

Vamshi, K. B., Pandey, A. K., & Siva, K. A. P. (2018). Topic Model Based Opinion Mining and Sentiment Analysis. *2018 International Conference on Computer Communication and Informatics (ICCCI)*, 1-4.

van Laer, T., & de Ruyter, K. (2010). In stories we trust: How narrative apologies provide cover for competitive vulnerability after integrity-violating blog posts. *International Journal of Research in Marketing*, *27*(2), 164-174. https://doi.org/https://doi.org/10.1016/j.ijresmar.2009.12.010

Van Nuenen, T., & Scarles, C. (2021). Advancements in technology and digital media in tourism. *Tourist Studies*, *21*(1), 119-132.

Van Raaij, W. F., & Francken, D. A. (1984). Vacation decisions, activities, and satisfactions. *Annals of Tourism Research*, *11*(1), 101-112.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of marketing*, *68*(1), 1-17.

Varian, H. R. (1983). Non-parametric tests of consumer behaviour. *The review of economic studies*, *50*(1), 99-110.

Varkaris, E., & Neuhofer, B. (2017). The influence of social media on the consumers' hotel decision journey. *Journal of Hospitality and Tourism Technology*.

Vatsa, P. (2021). Seasonality and cycles in tourism demand—redux. *Annals of Tourism Research*, *90*, 103105.

Vengesayi, S., Mavondo, F., & Reisinger, Y. (2009). Tourism destination attractiveness: attractions, facilities, and people as predictors. *Tourism Analysis*, *14*, 621-636.

Vermeer, S. A., Araujo, T., Bernritter, S. F., & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, *36*(3), 492-508.

Vierhaus, C. (2018). The international tourism effect of hosting the Olympic Games and the FIFA World Cup. *Tourism Economics*, *25*, 1009 - 1028.

Vierhaus, C. (2019). The international tourism effect of hosting the Olympic Games and the FIFA World Cup. *Tourism Economics*, *25*(7), 1009-1028.

Virkar, A. R., & Mallya, P. D. (2018). A Review of Dimensions of Tourism Transport Affecting Tourist Satisfaction.

Visit Britain. (2020). *International visitors to London*. https://www.visitbritain.org/latest-quarterly-data-area

Vroom, V. H. (1964). Work and motivation.

Wahab, S., Crampon, L. J., & Rothfield, L. M. (1976). *Tourism marketing: a destination-orientated programme for the marketing of international tourism*. Tourism International Press.

Wakimin, N., Azlinaa, A., & Hazman, S. (2018). Tourism demand in Asean-5 countries: Evidence from panel data analysis. *Management Science Letters*, *8*(6), 677-690.

Walmsley, D. J., & Young, M. (1998). Evaluative images and tourism: The use of personal constructs to describe the structure of destination images. *Journal of Travel Research*, *36*(3), 65-69.

Wan, R., Mei, S., Wang, J., Liu, M., & Yang, F. (2019). Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics*, *8*(8), 876.

Wang, C.-H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management*, *25*(3), 367-374.

Wang, S., Liu, J., & Shroff, N. (2018). Coded sparse matrix multiplication. International Conference on Machine Learning,

Wang, T.-L., Tran, P. T. K., & Tran, V. T. (2017). Destination perceived quality, tourist satisfaction and word-of-mouth. *Tourism Review*.

Wang, Y., Liu, Z., Hu, D., & Zhang, M. (2019). Multivariate time series prediction based on optimized temporal convolutional networks with stacked auto-encoders. Asian Conference on Machine Learning,

Ward, S. (1974). Consumer socialization. *Journal of consumer research*, *1*(2), 1-14.

Webber, A. G. (2001). Exchange rate volatility and cointegration in tourism demand. *Journal of Travel Research*, *39*(4), 398-405.

Webster, C., & Ivanov, S. (2016). Political Ideologies as Shapers of Future Tourism Development. *PSN: Political Institutions (Topic)*.

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., & Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*, 1354816620987198.

Weis, C., & Axhausen, K. W. (2009). Induced travel demand: Evidence from a pseudo panel data based structural equations model. *Research in Transportation Economics*, *25*(1), 8-18.

Wen, H., Josiam, B. M., Spears, D. L., & Yang, Y. (2018). Influence of movies and television on Chinese Tourists perception toward international tourism destinations. *Tourism Management Perspectives*.

Wendland, J., Ehnis, C., Clarke, R. J., & Bunker, D. (2018). Sydney siege, December 2014: A visualisation of a semantic social media sentiment analysis.

Westermann, A., & Forthmann, J. (2020). Social listening: a potential game changer in reputation management How big data analysis can contribute to understanding stakeholders' views on organisations. *Corporate Communications: An International Journal*.

Whyte, L. J. (2017). Understanding the relationship between push and pull motivational factors in cruise tourism: A canonical correlation analysis. *International Journal of Tourism Research*, *19*(5), 557-568.

Williams, M. (2000). Interpretivism and generalisation. *Sociology*, *34*(2), 209-224.

Witt, C. A., Witt, S. F., & Wilson, N. (1994). Forecasting international tourist flows. *Annals of Tourism Research*, *21*(3), 612-628.

Witt, C. A., & Wright, P. L. (1992). Tourist motivation: life after Maslow. *Tourist motivation: life after Maslow.*, 33-35.

Witt, S. F., & Martin, C. A. (1987). International tourism-demand models—inclusion of marketing variables. *Tourism Management*, *8*(1), 33-40.

Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection. *Journal of Travel Research*, *42*(2), 151-158.

Wolfe, K., & Hsu, C. H. (2004). An application of the social psychological model of tourism motivation. *International Journal of Hospitality & Tourism Administration*, *5*(1), 29-47.

Woodside, A. G., & Lysonski, S. (1989). A general model of traveler destination choice. *Journal of Travel Research*, *27*(4), 8-14.

Woodside, A. G., & MacDonald, R. (1994). General system framework of customer choice processes of tourism services. *Spoilt for choice*, *30*.

World Tourism Organisation. (2021). *Yearbook of Tourism Statistics, Data 2015 – 2019* (2021 Edition ed.). UNWTO.

Wu, C. (2010). *Econometric analysis of tourist expenditures* The Hong Kong Polytechnic University].

Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*, *29*(1), 507-529.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179-188.

Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120-130.

Xie, G., Li, X., Qian, Y., & Wang, S. (2020). Forecasting tourism demand with KPCA-based web search indexes. *Tourism Economics*, 1354816619898576.

Xu, F., Nash, N., & Whitmarsh, L. (2020). Big data or small data? A methodological review of sustainable tourism. *Journal of Sustainable Tourism*, *28*(2), 144-163.

Xu, X., & Pratt, S. (2018). Social media influencers as endorsers to promote travel destinations: an application of self-congruence theory to the Chinese Generation Y. *Journal of travel & tourism marketing*, *35*(7), 958-972.

Xu, X., & Reed, M. (2017). Perceived pollution and inbound tourism in China. *Tourism Management Perspectives*, *21*, 109-112.

Yang, C.-H., Lin, H.-L., & Han, C.-C. (2010). Analysis of international tourist arrivals in China: The role of World Heritage Sites. *Tourism Management*, *31*(6), 827-837.

Yang, Q., Zhao, D., Wu, Y., Tang, P., Gu, R., & Luo, Y.-j. (2018). Differentiating the influence of incidental anger and fear on risk decision-making. *Physiology & behavior*, *184*, 179-188.

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., Xue, L., & Jones, T. E. (2019). Tourism-enhancing effect of World Heritage Sites: Panacea or placebo? A meta-analysis. *Annals of Tourism Research*, *75*, 29-41. https://doi.org/https://doi.org/10.1016/j.annals.2018.12.007

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, *32*.

Yaya, M. E. (2009). Terrorism and tourism: The case of Turkey. *Defence and Peace Economics*, *20*(6), 477-497.

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management*, *26*(1), 45-56.

Yoopetch, C., Kongarchapatara, B., & Nimsai, S. (2022). Tourism Forecasting Using the Delphi Method and Implications for Sustainable Tourism Development. *Sustainability*, *15*(1), 126.

Yousefi, M., & Marzuki, A. (2015). An analysis of push and pull motivational factors of international tourists to Penang, Malaysia. *International Journal of Hospitality & Tourism Administration*, *16*(1), 40-56.

Yuan, H., Xu, W., Li, Q., & Lau, R. Y. K. (2018). Topic sentiment mining for sales performance prediction in e-commerce. *Annals of Operations Research*, *270*, 553-576.

Zajonc, R. (1968). ``Attitudinal effects of mere exposure"Journal of Personality and Social Psychology 9 1^ 27 Conditions of use. *This article may be downloaded from the E&P website for personal research by members of subscribing organisations. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher*.

Zeng, B. (2008). *Tourism development and local poverty: A case study of Qinling Mountain Region, Shaanxi Province, China*. VDM Verlag Dr. Mueller eK.

Zhang, C., Li, M., Sun, S., Tang, L., & Wang, S. (2021). Decomposition Methods for Tourism Demand Forecasting: A Comparative Study. *Journal of Travel Research*, 00472875211036194.

Zhang, H., Fu, X., Cai, L. A., & Lu, L. (2014). Destination image and tourist loyalty: A meta-analysis. *Tourism Management*, *40*, 213-223.

Zhang, H., Hu, W., Cao, D., Huang, Q., Chen, Z., & Blaabjerg, F. (2021). A temporal convolutional network based hybrid model of short-term electricity price forecasting. *CSEE Journal of Power and Energy Systems*.

Zhang, K., Chen, Y., & Li, C. (2019). Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tourism Management*, *75*, 595-608.

Zhang, X., Yang, Y., Zhang, Y., & Zhang, Z. (2020). Designing tourist experiences amidst air pollution: A spatial analytical approach using social media. *Annals of Tourism Research*, *84*, 102999.

Zhang, Y., Li, G., Muskat, B., & Law, R. (2021). Tourism Demand Forecasting: A Decomposed Deep Learning Approach. *Journal of Travel Research*, *60*(5), 981-997. https://doi.org/10.1177/0047287520919522

Zhang, Y., Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, *34*(1), 100-119.

Zheng, W., Huang, L., & Lin, Z. (2021). Multi-attraction, hourly tourism demand forecasting. *Annals of Tourism Research*, *90*, 103271.

Zhou, M., Chen, G., Ferreira, P., & Smith, M. D. (2021). Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware. *Available at SSRN 3897111*.

Zunic, E., Korjenic, K., Hodzic, K., & Donko, D. (2020). Application of facebook's prophet algorithm for successful sales forecasting based on real-world data. *arXiv preprint arXiv:2005.07575*.