

Twitter's capacity to forecast tourism demand: the case of way of Saint James

Twitter and
tourism
demand

Adrián Mendieta-Aragón

*Department of Economic Analysis, Faculty of Economics and Business,
National University of Distance Education, Madrid, Spain*

Julio Navío-Marco

*Department of Business Organization and Management,
Faculty of Economics and Business, National University of Distance Education,
Madrid, Spain, and*

Teresa Garín-Muñoz

*Department of Economic Analysis, Faculty of Economics and Business,
National University of Distance Education, Madrid, Spain*

Received 12 September 2023

Revised 9 January 2024

10 February 2024

20 February 2024

Accepted 20 February 2024

Abstract

Purpose – Radical changes in consumer habits induced by the coronavirus disease (COVID-19) pandemic suggest that the usual demand forecasting techniques based on historical series are questionable. This is particularly true for hospitality demand, which has been dramatically affected by the pandemic. Accordingly, we investigate the suitability of tourists' activity on Twitter as a predictor of hospitality demand in the Way of Saint James – an important pilgrimage tourism destination.

Design/methodology/approach – This study compares the predictive performance of the seasonal autoregressive integrated moving average (SARIMA) time-series model with that of the SARIMA with an exogenous variables (SARIMAX) model to forecast hotel tourism demand. For this, 110,456 tweets posted on Twitter between January 2018 and September 2022 are used as exogenous variables.

Findings – The results confirm that the predictions of traditional time-series models for tourist demand can be significantly improved by including tourist activity on Twitter. Twitter data could be an effective tool for improving the forecasting accuracy of tourism demand in real-time, which has relevant implications for tourism management. This study also provides a better understanding of tourists' digital footprints in pilgrimage tourism.

Originality/value – This study contributes to the scarce literature on the digitalisation of pilgrimage tourism and forecasting hotel demand using a new methodological framework based on Twitter user-generated content. This can enable hospitality industry practitioners to convert social media data into relevant information for hospitality management.

Keywords Hotel demand forecasting, Social media, Big data, Twitter, Pilgrimage tourism

Paper type Research paper

1. Introduction

Research on the use of social media and social networking sites in hospitality and tourism has proliferated in recent years (Buhalis *et al.*, 2017; Jamil *et al.*, 2023; Kozak *et al.*, 2018; Leung *et al.*, 2021; Sigala, 2015). Social media information enables the analysis of user behaviour

© Adrián Mendieta-Aragón, Julio Navío-Marco and Teresa Garín-Muñoz. Published in *European Journal of Management and Business Economics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Funding: This work was supported by the National University of Distance Education (Spain) under Grant [BICI N.3, October 21, 2019].



European Journal of Management
and Business Economics
Emerald Publishing Limited
e-ISSN: 2444-8494
p-ISSN: 2444-8451

DOI 10.1108/EJMBE-09-2023-0295

(Bigne *et al.*, 2016; Küster Boluda *et al.*, 2024; Navío-Marco *et al.*, 2018; Payntar *et al.*, 2021), accelerates the knowledge transfer process, provides a direct link between users and knowledge (Abdollahi *et al.*, 2023; Rita *et al.*, 2022) and helps analyse the relationship between brand equity and social media intensity (Stojanovic *et al.*, 2018).

Lately, it has been used as a data source for estimating tourism demand in a very incipient way. Li *et al.* (2021), in their review of tourism and hospitality forecasting research using Internet data, have identified only ten studies adopting social media data for forecasting. Since then, studies using social media data to improve predictions in the field of tourism have been increasing (e.g. Hu *et al.*, 2022; Li *et al.*, 2022; Sulong *et al.*, 2022). Regarding Twitter, Bigné *et al.* (2019) have extracted important relevant information from this application to determine how destination marketing organisation (DMO) activities on Twitter affect hotel occupancy forecasting.

Assaf *et al.* (2022), in their investigation to establish an expert-informed agenda for future research on tourism after COVID-19, have considered forecasting an area in which to progress, including the use of scenario forecasts using judgemental and econometric methods based on big data, tourism portals and social media. Several scholars have observed that during and after the pandemic, tourist demand was seriously impacted and the traditional methods of forecasting in these industries have become obsolete (Song and Li, 2021; Utkarsh and Sigala, 2021). Researchers have now begun to seek the best methods to predict the recovery of tourism from the devastating effects of COVID-19 (Polyzos *et al.*, 2020; Zhang *et al.*, 2021).

The relationship between tourism and pilgrimages has been studied in a fragmentary manner (Caber *et al.*, 2021), despite the growing economic importance of this kind of tourism [1]. While motivations and experiences have been analysed (Terzidou *et al.*, 2018), limited attention has been paid to behaviour on online platforms and digital devices (de Ascaniis *et al.*, 2019).

Accordingly, this study aims to fill the gap in the scarce literature on pilgrims' use of social networks and the suitability of user-generated data for accurately predicting hotel demand. The contribution of this research is threefold. First, it evaluates Twitter as a tool for predicting demand – in this case, for pilgrimage tourism to “the Way” – and provides insights into the time lag between tweets and demand manifestation. Second, it sheds light on the changes in hospitality demand and explores new forecasting approaches for estimating tourism demand during tumultuous times. Additionally, it provides new data on the digital footprint of pilgrimage tourism, an area where research is also very scarce.

As a research question, this study examines how hotel demand at a tourist destination can be accurately predicted using Twitter data. Particularly, this study analyses an international destination of special interest for pilgrimage tourism, namely, Santiago de Compostela, Spain. Accordingly, we assess the predictive performance of the seasonal autoregressive integrated moving average (SARIMA) time-series model with and without including the Twitter activity of pilgrims, considering the lagged effect of Twitter data and external factors, such as the Jubilee year in Santiago de Compostela. Accordingly, this study predicts tourism demand from January 2018 to September 2022 (using 110,456 tweets posted).

The remainder of this study is structured as follows: Section 2 briefly reviews the literature on techniques for forecasting tourism demand, use of social network data for forecasting and digital footprint of pilgrimage tourism. Section 3 presents the empirical analysis, including descriptions of the data and methodology. Section 4 presents and discusses the results. Finally, Section 5 presents the conclusion, major theoretical and managerial implications, study limitations and new avenues for future research.

2. Literature review

2.1 Tourism demand forecasting

Demand forecasting is essential for the hospitality and tourism sectors because of the transient nature of tourism. Therefore, growing interest in tourism demand forecasting is

reflected in the literature. Several studies have assessed the performance of different sources of big data generated on the internet for forecasting tourism demand (Li *et al.*, 2021; Mariani and Baggio, 2022; Stylos *et al.*, 2021).

Tourism demand forecasting studies have predominantly applied time-series and econometric models. The most popular time-series analysis methods are autoregressive with exogenous variables (ARX) (Choi and Varian, 2012; Li *et al.*, 2017), autoregressive integrated moving average (ARIMA) (Artola *et al.*, 2015; Li *et al.*, 2018), SARIMA (Qiu *et al.*, 2021; Wickramasinghe and Ratnasiri, 2021) and SARIMAX (Hu *et al.*, 2022; Park *et al.*, 2021) models. Moreover, autoregressive distributed lag (ARDL) (Husein and Kara, 2020; Li *et al.*, 2020), time-varying parameter (TVP) (Smeral and Song, 2015) and dynamic factor (DFM) (Camacho and Paccé, 2017) econometric models have also been widely employed in tourism demand forecasting.

Time-series models have maintained increasing acceptance in the literature on tourism demand forecasting studies (Huang and Zheng, 2023; Teixeira and Gunter, 2023; Wu *et al.*, 2023). This is mainly because of their ability to forecast future time series by identifying historical patterns and capturing seasonality and trends in time series (Ma *et al.*, 2023). However, in recent literature, a trend has emerged to incorporate exogenous explanatory variables into time-series models for predicting tourism demand (Hu *et al.*, 2023; Jiao and Chen, 2019; Li *et al.*, 2023a). Thus, SARIMAX models have gained importance among academics, especially after the COVID-19 pandemic. They improve the performance of pure time-series forecasting models during turbulent periods and allow the incorporation of exogenous variables with real-time information. For example, researchers have compared the performance of SARIMA models with exogenous variables using information collected from search engines (Li *et al.*, 2023b; Wickramasinghe and Ratnasiri, 2021), online news (Park *et al.*, 2021) and online reviews (Hu *et al.*, 2022; Li *et al.*, 2023a, b). The results confirm that the incorporation of this type of big data generated on the internet is useful for forecasting tourist demand for destinations or companies.

2.2 Social media as a source of prediction data

Studies have demonstrated that social media data measures people's attention and sentiments and provides real-time insights to predict consumer demand in different research areas, including economics and management. The main areas covered include the following: (a) stock market performance accurately predicted based on investors' opinions on social media (Guan *et al.*, 2022; Nofer and Hinz, 2015; Yang *et al.*, 2020), (b) transport and power demand predicted using real-time data from social media (Luna, Nunez-del-Prado, Talavera and Holguin, 2017; Punel and Ermagun, 2018; Roy *et al.*, 2021) and (c) crude oil prices predicted with social media data during periods of sharp fluctuations caused by conflicts or political instability (Elshendy *et al.*, 2017; Wu *et al.*, 2021).

Regarding Internet-structured data in tourism, search engine data (Bangwayo-Skeete and Skeete, 2015; Choi and Varian, 2012; Wu *et al.*, 2022) and web traffic data (Gunter and Önder, 2016) have been widely used to forecast tourism demand. Conversely, social media data are unstructured and require crawler tools to collect and apply big data techniques for extracting useful information from online textual data or images, thereby making them relatively less popular (Li *et al.*, 2021).

Focusing on Twitter, tourism studies have utilised this data source for sentiment analysis to identify tourist preferences and opinions on tourist services (Nadeau *et al.*, 2022; Philander and Zhong, 2016), geographic information (Chua *et al.*, 2016; Piramanayagam and Seal, 2022; Xin and MacEachren, 2020), promotion of tourist attractions (Bokunewicz and Shulman, 2017; Meehan *et al.*, 2016) and international trade show organisation (Geldres-Weiss *et al.*, 2023). However, only a few studies have analysed the usefulness of big data from Twitter to

analyse tourism demand (e.g. [Bigné et al., 2019](#); [Sulong et al., 2022](#); [Yang et al., 2022](#)) and define management approaches and business responses to the COVID-19 pandemic in real-time ([Chen et al., 2023](#); [Yang and Han, 2021](#)).

Previous literature has recognised Twitter's representativeness as a concern ([Beninger and Lepps, 2014](#)), but some authors recognise its interest if a contextual interpretation is made ([Tromble, 2019](#)). Twitter data differ in nature from data collected through traditional quantitative methods, such as surveys or experiments ([Chen et al., 2022](#)). Survey data are controlled and designed by researchers, while social media data can be considered organic data ([Groves, 2011](#)). The concept of organic data refers to data that are not collected following an explicit research design but documented using a technology that collects natural "digital footprints" of human activities, such as data from sensor devices, mobile applications or online social networks ([Xu et al., 2020](#)).

According to [Xu et al. \(2020\)](#), the advantages of these data coexist with challenges regarding data quality that researchers must consider because of their organic nature. First, data quality is more likely to be guaranteed in surveys and experiments because researchers have more control over which participants are recruited and what questions to ask. However, the emergent nature of social media discussions offers researchers opportunities to identify new perspectives and frameworks not previously identified ([Klašnja et al., 2018](#)). Although researchers have more control over the data generation process in surveys and experiments, it is expensive to collect surveys. Furthermore, organic data generated on social networks allows information to be extracted in real-time. Traditionally, hotel demand forecasts have been based solely on government statistical reports published annually or monthly ([Huang et al., 2017](#)). Nevertheless, hospitality industry professionals need up-to-date information to adjust to changes in tourism demand in real-time and achieve greater efficiency in the sector.

Newness is a strength of social media data, which is especially useful for studying emerging topics. The novelty of the data brings with it a data quality challenge that requires researchers to develop methods to indirectly assess user characteristics, such as user identity and motivations. Similarly, numerous authors have indicated that the pandemic has called into question traditional forecasting methods because data from official sources with guaranteed representativeness are not available in real-time, which makes it even more interesting to explore new data sources that are open and original, as done in this study.

2.3 Pilgrimage tourism's digital footprint

Literature on the digital aspects of pilgrimage tourism is scarce, recent and focused on human mobility ([Barnett et al., 2016](#)). [De Ascaniis et al. \(2019\)](#) have reviewed 13 academic papers and identified the following four themes: the adoption of information and communication technology (ICT) by religious travellers, usage and functionalities of mobile applications, online travel reviews to understand visitors' experiences at religious sites and online transmission of religious mass events. Research interest in religious tourists' behaviour on digital platforms, such as social media and social networking sites, remains incipient. [Caber et al. \(2021\)](#) have identified a few early works, such as [Haq and Jackson \(2009\)](#) investigating the impact of ICTs on religious tourists' perceptions and [Park et al. \(2015\)](#) surveying American participants to gauge their interest in visiting pilgrimage destinations and willingness to share their experiences on social networking sites.

"The Way" is a pilgrimage tourism destination that generates both religious and tourist interest worldwide ([López et al., 2017](#)). [Vila et al. \(2020\)](#) have indicated that religious or spiritual motivation is present but interlinked with other motivations, such as heritage, culture and experience. "The Way" is an international and multiconfessional space where pilgrims and tourists interact to co-create the route's postmodern identity and personality ([López and Lois González, 2020](#)). Pilgrims in "the Way" benefit from using mobile phones

while walking (Antunes and Amaro, 2016; Nickerson *et al.*, 2014). Fernández-Poyatos *et al.* (2012) have studied the presence of “the Way” on regional tourism websites in Spain, while Vázquez *et al.* (2020) have analysed the usage and effectiveness of Facebook fan pages of institutions in Spanish regions through which the French Way of Saint James passes for tourism promotion. No other research has been conducted on social media use pertaining to this topic.

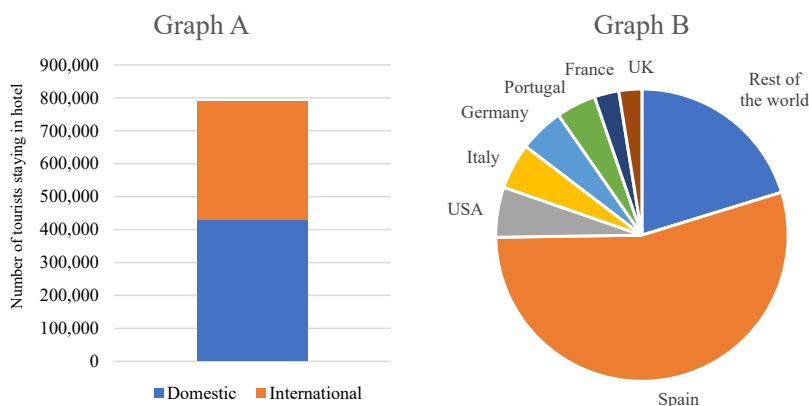
Pilgrimage tourism, gaining popularity since the COVID-19 outbreak, has demonstrated great resilience during the pandemic (Lin and Hsieh, 2022; Mittal and Sinha, 2021). As outdoor activities, pilgrimage routes can provide a safe environment and improve tourist well-being, offering an alternative to mass tourism (Lin *et al.*, 2022). Therefore, tourist destinations have used religious tourism as a key market segment to mitigate disruptions in tourism demand caused by the COVID-19 pandemic (Mittal and Sinha, 2021). In fact, pilgrimage tourism is positioned as a novel travel trend in tourism in the “new normal” (Campos *et al.*, 2022). This makes research that combines tourist demand, social media and pilgrimage tourism particularly interesting.

3. Empirical analysis

3.1 Data

Pilgrimage tourism is in a state of rejuvenation and is gaining importance among various tourism segments (Collins-Kreiner, 2020). This empirical analysis investigates the relationship between the digital footprint of pilgrims on “the Way” and hotel tourism demand for Santiago de Compostela. This is a major European pilgrimage itinerary recognised as the first European Cultural Route by the Council of Europe. Figure 1 presents the international dimensions of Santiago de Compostela as a tourist destination in 2019 (the year before the COVID-19 pandemic). Graph A reveals that foreign tourism represents 45.5% of the total hospitality demand, whereas Graph B reveals the distribution of international tourism demand by country of origin. The USA, Italy, Germany, Portugal, France and the UK generated 55.5% of international tourism demand.

Figure 2 depicts the framework used in this study to predict tourism demand in Santiago de Compostela based on big data generated on Twitter by pilgrims to the Saint James Way. It presents the data sources, data collection, model specifications and processes used in the empirical analysis.



Note(s): Data: Hotel Occupancy Survey (INE)
Source(s): Figure by authors

Figure 1.
Volume and
distribution of tourism
demand in Santiago de
Compostela in 2019
(pre-COVID-19)

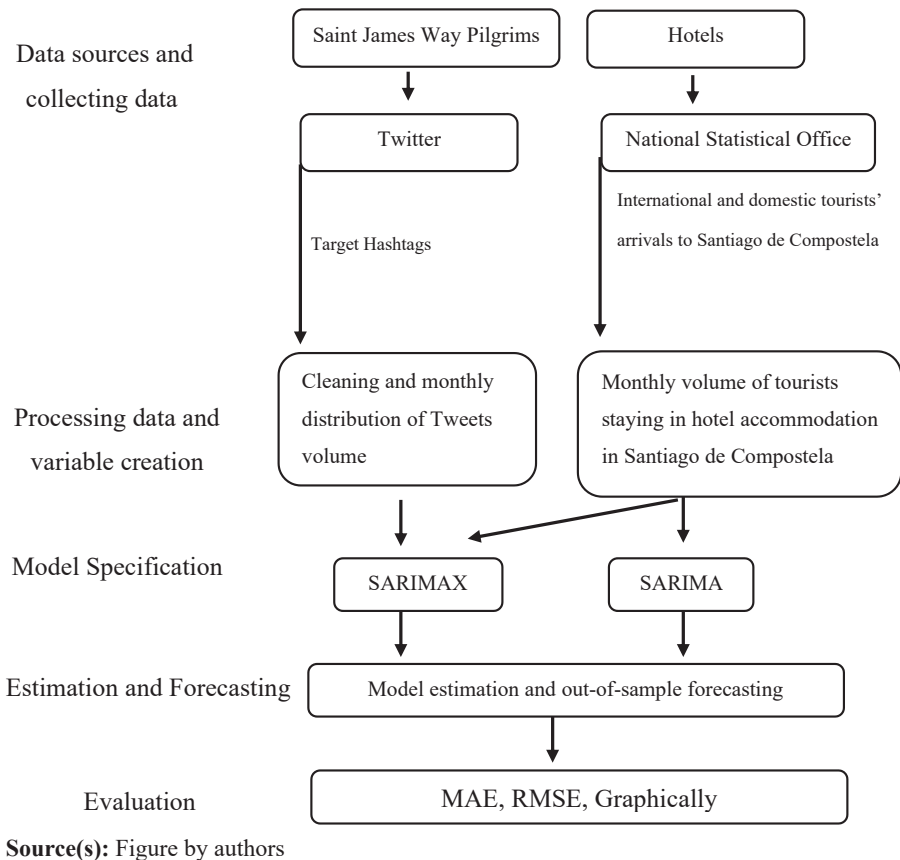


Figure 2.
 Framework for tourism demand predictions based on Twitter data

As shown in [Figure 2](#), the tourism demand for Santiago de Compostela is measured using the total number of tourists staying in hotel accommodations (TOUR). Monthly tourist arrivals are collected from the Hotel Occupancy Survey (HOS), published by the Spanish National Statistics Institute (INE) since 1996. It provides disaggregated information on travellers by country of origin and destination (regions, provinces and tourist sites). This measure includes the total number of travellers arriving by any means of transportation and staying in an establishment that provides hotel accommodation services (hotels, apart-hotels, motels, hostels, B&Bs, pensions and guesthouses).

[Figure 2](#) shows the digital footprint of tourists on Twitter as a secondary source of data. A crawler created with the programming language Python is used to extract the digital footprints of tourists on Twitter. Specifically, a script is designed to collate tweets posted with target hashtags using Twitter API V2. As Santiago de Compostela is an international pilgrimage destination, the decision to use hashtags was supported by an exhaustive search for hashtags related to tourism. Previous literature has supported the idea that the use of hashtags on Twitter is a powerful and helpful source of data ([Geldres-Weiss et al., 2023](#); [Wang et al., 2016](#)). According to [Carvache-Franco et al. \(2023\)](#), using hashtags to gather information is advantageous because it allows the concentration of users' opinions on a specific topic. Although the use of hashtags may exclude some data, it also helps avoid irrelevant data.

Twitter is a massive platform with a large amount of noisy and irrelevant data. Using hashtags helps categorise topics, making it easier to identify users who are talking about the same topic (Bruns and Burgess, 2011). Using hashtags also allows us to filter this noise and focus on the data most relevant to our study.

All hashtags included in tweets published during the study period that contained the key search “Santiago de Compostela” were identified. By comparing the most repeated hashtags related to tourism for this destination, the following categories were identified:

- (1) “Saint James Way”,
- (2) “Pilgrims” and “Pilgrimage” and
- (3) “Xacobeo” and “Jacobeo”.

We excluded hashtags related to “Pilgrims” and “Pilgrimage” because they could include tweets pertaining to other pilgrimage destinations. However, tweets pertaining to St. James Way and Xacobeo were exclusive to tourism in Santiago de Compostela. Therefore, a combination of the 20 most published hashtags related to categories (1) and (3) in Spanish, English, German, French and Portuguese was selected (Table 1). These languages were selected because countries with these languages as their native languages represented 75% of hotel tourism demand in Santiago de Compostela in 2019.

After eliminating duplicate retweets, 110,456 tweets remained, based on which the monthly number of tweets was used to derive the explanatory variable – Twitter Data (TD). According to Guizzardi and Mazzocchi (2010), factors that occur at a specific moment in time, such as the Jubilee Year, can determine short- or long-term modifications in tourist flow. Therefore, a temporary dummy was created to control the effect of an extraordinary increase in tourism demand in 2021 and 2022, the Jubilee years in Santiago de Compostela (Compostela Holy Year, Xacobeo Year or Jacobeo Year). This variable takes the value of one for 2021 and 2022 and zero otherwise.

Language	Hashtag
Spanish	#CAMINODESANTIAGO #ELCAMINODESANTIAGO #BUENCAMINO #JACOBEO #XACOBEO
English	#WAYOFSTJAMES #THEWAYOFSAIN TJAMES #WAYOFSAIN TJAMES #SAIN TJAMESWAY #SANTIAGOWAY #WAYOFSANTIAGO #WALKCAMINO
German	#JAKOB SWEG #DERJAKOB WEB #DERWEGNACHSANTIAGO
French	#CHEMINDESAIN TJACQUES #LECHEMINDESAIN TJACQUES #SAIN TJACQUESCHEMIN
Portuguese	#OCAMINHODESANTIAGO #CAMINHODESANTIAGO

Source(s): Table by authors

Table 1.
Selected hashtags

3.2 Methodology

In this study, we compare two ARIMA-based forecasting models (SARIMA and SARIMAX models) to evaluate the appropriateness of using user-generated content on social media to improve the predictive capacity of time-series models in turmoil stages. In this exploratory case, we forecast monthly tourism demand for the internationally known destination of Santiago de Compostela.

The comparison of the SARIMA models in our time-series prediction methodology aligns with the goal of achieving accurate predictions, considering the specific characteristics of our dataset. We aim to capture the effects of exogenous shocks as part of the SARIMA model. To achieve this, we compare the predictive capacity of the SARIMA pure time-series forecasting and SARIMA models with exogenous variables (SARIMAX).

The SARIMA model was selected because of its various statistical advantages, supported by previous research on tourism demand forecasting (Qiu *et al.*, 2021; Song *et al.*, 2019). According to Song *et al.* (2019), the SARIMA model is the most commonly used model in tourism research because it considers the trends and/or seasonality components of a time series. Additionally, the parsimonious structure of the SARIMA models balances complexity and performance (Lama *et al.*, 2022; Saz, 2011).

The SARIMA (p,d,q) (P,D,Q) model is as follows:

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d y_t = \Theta(B^m)\theta(B) \varepsilon_t \quad (1)$$

where y_t expresses the tourism demand at time t ; the autoregressive (AR) and moving average (MA) components are represented by ϕ and θ of orders p and q , respectively; $\Phi(B^m)$ and $\Theta(B^m)$ denote the seasonal AR(P) and seasonal MA(Q) components, respectively; $(1-B)^d$ and $(1-B^m)^D$ represent the difference and seasonal difference indicators, respectively; ε_t expresses the white noise error term.

Using a linear regression, external variables can be added to the SARIMA model to create a SARIMAX model. Eq. (2) indicates that SARIMAX is a regression model with SARIMA errors where the regression is first conducted.

$$\Phi(B^m)\phi(B)(1-B^m)^D(1-B)^d y_t = \mu + \sum_{k=1}^n \beta^k \cdot X_t^k + \Theta(B^m)\theta(B) \varepsilon_t \quad (2)$$

where X_t^k is the exogenous variable at time t and β^k is the corresponding coefficient of the exogenous variable added to the parameters of the aforementioned SARIMA model described.

To validate the models and assess their respective predictive capacities, we fit the models with data from January 2018 to December 2021 and use those from January 2022 to September 2022 to test the accuracy of the predictions. To evaluate the forecast accuracy of the models, we use the following common evaluation measures from tourism and hospitality forecasting research: the mean absolute error (MAE) and root mean square error (RMSE), calculated using Eq. (3) and (4).

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2} \quad (4)$$

where \hat{y}_t and y_t are the predicted and actual values representing tourism demand in Santiago de Compostela, respectively.

4. Results and discussion

An exploratory analysis during the fitting period reveals that the variable TD displays the same trend as the variable TOUR, which denotes the volume of tourists staying in hotels in Santiago de Compostela; however, the peaks of the former occur one month earlier than those of the latter (see Figure 3). This indicates that tourists' Twitter activity is a good predictor of hotel demand.

Tourism demand has a high seasonal component, which is adjusted according to the model specifications. The augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) unit root tests confirm the presence of a unit root in the dependent and independent variables at the 1% significance level. Therefore, the first differences of all the variables are considered to ensure a stationary series. Correlograms and partial autocorrelation functions are examined to determine the appropriate order of the AR and MA components.

To analyse Twitter data's dynamic structure to forecast tourism demand, we use the Akaike information criterion (AIC) and Schwartz Bayesian information criterion (SBIC) to determine the monthly lagged distribution of the explanatory variable. The results indicate that the optimal lag length for the independent variable is two months. Additionally, the Granger causality test confirms a causal relationship between hotel demand and tourists' Twitter activity.

Table 2 presents the forecast errors of the in-sample estimation and improvement achieved in the final SARIMAX model compared to the SARIMA model [2]. The results indicate that including exogenous variables improves the SARIMA model's fit by 5.75 and 9.05% for the MAE and RMSE evaluation measures, respectively.

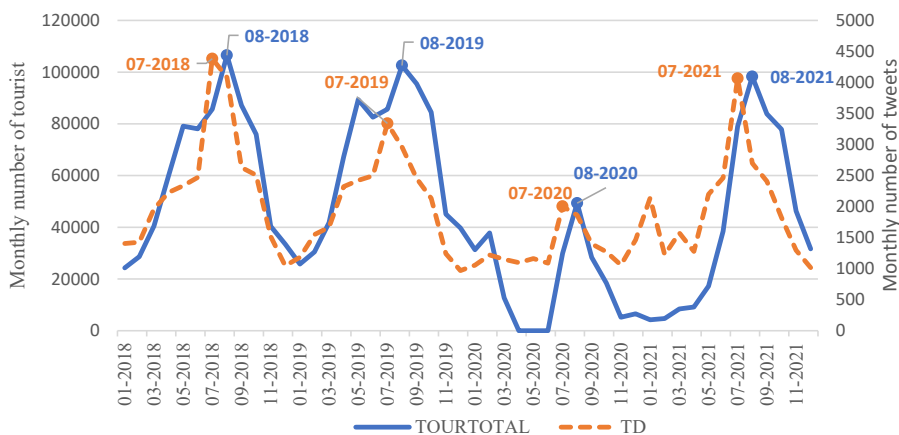


Figure 3. Volume and evolution of tourist arrivals and generated tweets in Santiago de Compostela (fitting period: Jan 2018–Dec 2021)

Note(s): Data: Hotel Occupancy Survey (INE) and Twitter

Source(s): Figure by authors

Evaluation metrics	SARIMA	SARIMAX	Improvement (%)
MAE	10466.18	<i>9864.05</i>	5.75
RMSE	13900.89	<i>12642.22</i>	9.05

Note(s): The values in italic indicate the model with the best evaluation metric

Source(s): Table by authors

Table 2. Estimation results for in-sample predictions of SARIMA and SARIMAX models (January 2018–December 2021)

The performance of the out-of-sample prediction summarised in Table 3 confirms a significant improvement in the SARIMAX model by 20.3 and 18.0% when using the MAE and RMSE evaluation measures, respectively. The robustness of the analysis is tested by modifying the fitting periods of the models and comparing their predictive performance after including Twitter data. This analysis confirmed the goodness of fit of the results.

Consistent with Yang and Han (2021), this study provides novel perspectives for practitioners to gain relevant hospitality business insights using social media data. Our results' alignment with those of previous studies verifies the utility of using Twitter to improve hotel demand forecasts, as in Bigné *et al.* (2019), and confirms a significant improvement in prediction accuracy, even during the pandemic, with the inclusion of new real-time data sources. Similarly, incorporating online review data improves the MAE forecast models by 2.97 and 6.19% and the RMSE between -3.41 and 7.98%, following Hu *et al.* (2022).

Moreover, our results confirm the importance of the lag structure of data sources in forecasting research, allowing tourism companies and policymakers to accurately anticipate future tourism demand. According to the results of our research, the Twitter activity of pilgrims from the previous two months can help hospitality companies predict the tourism demand for the Saint James Way.

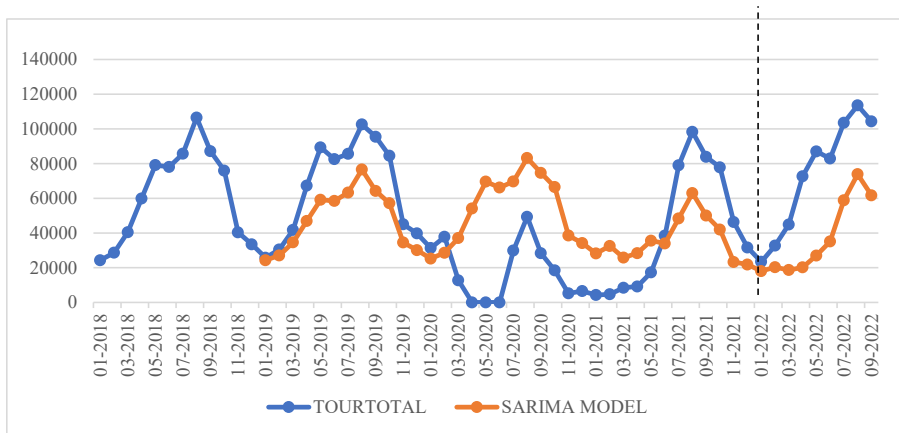
Figures 4 and 5 illustrate the actual and predicted tourism demand for Santiago de Compostela using the SARIMA and SARIMAX models, respectively. The evaluation measures of the SARIMA model and prediction accuracy shown in Figure 4 confirm that pure autoregressive models are inefficient in forecasting tourism demand during and after the pandemic. Therefore, we propose that researchers and stakeholders use Twitter activity data to accurately predict tourism demand (see Figure 5).

Table 3. Forecast performance of the out-of-sample predictions of SARIMA and SARIMAX models (January 2022–September 2022)

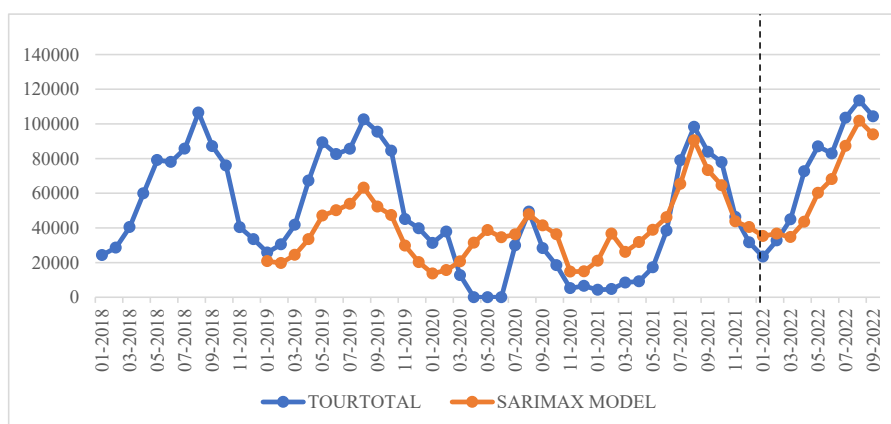
Evaluation metrics	SARIMA	SARIMAX	Improvement
MAE	9165.62	<i>7301.54</i>	20.3%
RMSE	11535.76	<i>9450.42</i>	18.0%

Note(s): The values in italic indicate the model with the best evaluation metric
Source(s): Table by authors

Figure 4. Forecast of tourist demand using SARIMA model



Source(s): Figure by authors



Source(s): Figure by authors

Figure 5.
Forecast of tourist
demand using
SARIMAX model with
Twitter data

Our findings answer the research questions and confirm our initial assumptions. With an improvement of between 18.0 and 20.3%, depending on the evaluation metric, pilgrim-generated digital content on social media can be used to improve the predictive capacity of time-series models. We agree with [Zhang et al. \(2021\)](#) in that hospitality companies' business planning, including budgeting, resource allocation and marketing, is based on demand forecasts. Consistent with [Li et al. \(2022\)](#), we avoid inaccurate predictions that could result in a supply-demand mismatch of tourism services, significantly affecting management, efficiency, productivity and the tourism sector's profitability. Therefore, this study makes a timely contribution to model development in tourism demand forecasting by proposing Twitter data as an exogenous variable to generate more accurate forecasts. Additionally, the results verify the lag-time structure of Twitter data, enabling the anticipation of changes in tourism demand during uncertain periods.

5. Conclusions

The pilgrim's footprint when walking "the Way" becomes a digital footprint in the 21st century. Our investigation contributes both to the scarce literature on digital pilgrimage tourism and research on forecasting hotel demand by proposing a new methodological framework based on user-generated content on Twitter for the case of the internationally known pilgrimage destination "the Way of Saint James".

This study demonstrates the importance of regularly refining forecasting methods using new data sources available in the digital world for effective forecasting. Thus, some theoretical implications are derived from this study. First, it improves our understanding of the usefulness of social networks, particularly Twitter, in forecasting tourism models. Second, it identifies the time lag between user information generated on Twitter and consumer demand. Third, it connects the digitalisation of pilgrimage tourists with the use of social networks and digital footprints.

We agree with [Gunter and Önder \(2015\)](#), suggesting that an accurate prediction of the number of tourists visiting a destination has implications for tourism management, such as sustaining tourism demand and efficient planning to accommodate tourists. This study has three primary managerial implications. First, the possibility of accurately predicting tourism demand from publicly shared information by pilgrims can improve hotel management efficiency at tourist destinations and prevent hotel oversupply or undersupply. Second, our

findings indicate that content published on Twitter during the previous two months is significant for forecasting hotel demand in Santiago de Compostela. Consistent with [Huang et al. \(2017\)](#) and [Liu et al. \(2018\)](#), the lag time structure of the data enables a better prediction of the demand and management of tourist destinations. This is because it allows the number of visitors to a destination to be known before they arrive. Finally, the COVID-19 pandemic has generated instability in tourism demand, induced by perceived health risks and government-imposed mobility restrictions, forcing managers to modify demand predictions frequently. Therefore, this study provides stakeholders with a methodological framework to accurately forecast real-time tourism demand and anticipate changes during times of crisis and instability.

In summary, Twitter offers two primary practical advantages for tourism management in Santiago de Compostela. First, it provides real-time information, which is particularly important during periods of uncertainty and volatility, such as those caused by the COVID-19 pandemic. Second, it helps accurately predict tourism demand, which can improve the tourism industry's efficiency. Therefore, this study recommends that stakeholders and decision-makers use Twitter as a new source of big data because it can serve as a leading indicator of changes in tourism demand.

This study has some limitations, the main one being its exploratory nature because it is limited to a single destination. One limitation of sampling our data using hashtags is that tweets related to elections without a hashtag would be ignored. However, the results obtained make it advisable to replicate the study in other tourism environments to observe the feasibility of using Twitter as a source for forecasting tourism demand, especially considering some of the trends found in this study are promising.

Nevertheless, the exploratory nature of this study does not detract from the relevance of its results, in which we are able to identify opportunities for Santiago de Compostela hotel demand planning. Furthermore, this study is limited to the pilgrimage destination of the Saint James Way and the results for other destinations should be cross-checked in future studies. Thus, the application of the Twitter-based forecasting method to other destinations is a clear avenue for future research.

In any case, we consider that our findings represent a step forward in the search for new forecasting methods that work even in the event of strong demand shocks, such as those caused by the COVID-19 pandemic and in understanding the relationship between social media data and pilgrimage tourism demand.

Notes

1. The United Nations World Tourism Organization estimates that 330 m people travel for religious reasons each year (<https://www.unwto.org>). Additionally, it is estimated that global income from religious tourism will increase from a total of \$15.1 bn in 2023 to approximately \$41 bn in 2033, according to the market analysis firm FutureMarket Insights (<https://www.futuremarketinsights.com>).
2. The improvement achieved using Twitter data are measured as follows:

$$\text{Improvement} = \frac{\text{Evaluation Metric (SARIMA)} - \text{Evaluation Metric (SARIMAX)}}{\text{Evaluation Metric (SARIMA)}}$$

References

- Abdollahi, A., Ghaderi, Z., Béal, L. and Cooper, C. (2023), "The intersection between knowledge management and organizational learning in tourism and hospitality: a bibliometric analysis", *Journal of Hospitality and Tourism Management*, Vol. 55, pp. 11-28, doi: [10.1016/J.JHTM.2023.02.014](https://doi.org/10.1016/J.JHTM.2023.02.014).
- Antunes, A. and Amaro, S. (2016), "Pilgrims' Acceptance of a Mobile App for the Camino de Santiago", in Inversini, A. and Schegg, R. (Eds), *Information and Communication Technologies in Tourism 2016*, Springer, Cham, pp. 509-521, doi: [10.1007/978-3-319-28231-2_37](https://doi.org/10.1007/978-3-319-28231-2_37).

-
- Artola, C., Pinto, F. and de Pedraza, P. (2015), "Can internet searches forecast tourism inflows?", *International Journal of Manpower*, Vol. 36 No. 1, pp. 103-116, doi: [10.1108/IJM-12-2014-0259](https://doi.org/10.1108/IJM-12-2014-0259).
- Assaf, A.G., Kock, F. and Tsionas, M. (2022), "Tourism during and after COVID-19: an expert-informed agenda for future research", *Journal of Travel Research*, Vol. 61 No. 2, pp. 454-457, doi: [10.1177/0047287521101723](https://doi.org/10.1177/0047287521101723).
- Bangwayo-Skeete, P.F. and Skeete, R.W. (2015), "Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach", *Tourism Management*, Vol. 46, pp. 454-464, doi: [10.1016/j.TOURMAN.2014.07.014](https://doi.org/10.1016/j.TOURMAN.2014.07.014).
- Barnett, I., Khanna, T. and Onnela, J.P. (2016), "Social and spatial clustering of people at humanity's largest gathering", *Plos One*, Vol. 11 No. 6, e0156794, doi: [10.1371/JOURNAL.PONE.0156794](https://doi.org/10.1371/JOURNAL.PONE.0156794).
- Beninger, K. and Lepps, H. (2014), *Research Using Social Media; Users' Views*, NatCen Social Research, London, Vol. 20.
- Bigné, E., Oltra, E. and Andreu, L. (2019), "Harnessing stakeholder input on Twitter: a case study of short breaks in Spanish tourist cities", *Tourism Management*, Vol. 71, pp. 490-503, doi: [10.1016/j.tourman.2018.10.013](https://doi.org/10.1016/j.tourman.2018.10.013).
- Bigne, E., Andreu, L., Hernandez, B. and Ruiz, C. (2016), "The impact of social media and offline influences on consumer behaviour. An analysis of the low-cost airline industry", *Current Issues in Tourism*, Vol. 21 No. 9, pp. 1014-1032, doi: [10.1080/13683500.2015.1126236](https://doi.org/10.1080/13683500.2015.1126236).
- Bokunewicz, J.F. and Shulman, J. (2017), "Influencer identification in Twitter networks of destination marketing organizations", *Journal of Hospitality and Tourism Technology*, Vol. 8 No. 2, pp. 205-219, doi: [10.1108/JHTT-09-2016-0057/FULL/XML](https://doi.org/10.1108/JHTT-09-2016-0057/FULL/XML).
- Bruns, A. and Burgess, J. (2011), "How twitter covered the 2010 Australian federal election", *Communication, Politics & Culture*, Vol. 44 No. 2, pp. 37-56, doi: [10.3316/IELAPA.627330171744964](https://doi.org/10.3316/IELAPA.627330171744964).
- Buhalis, D., Kavoura, A. and Cooper, C. (2017), "Social media and user-generated content for marketing tourism experiences", *Tourismos*, Vol. 12 No. 3, pp. x-xvi.
- Caber, M., Drori, N., Albayrak, T. and Herstein, R. (2021), "Social media usage behaviours of religious tourists: the cases of the Vatican, Mecca, and Jerusalem", *International Journal of Tourism Research*, Vol. 23 No. 5, pp. 816-831, doi: [10.1002/JTR.2444](https://doi.org/10.1002/JTR.2444).
- Camacho, M. and Pacce, M.J. (2017), "Forecasting travellers in Spain with Google's search volume indices", *Tourism Economics*, Vol. 24 No. 4, pp. 434-448, doi: [10.1177/1354816617737227](https://doi.org/10.1177/1354816617737227).
- Campos, C., Laso, J., Cristóbal, J., Alberti, J., Bala, A., Fullana, M., Aldaco, R. and Margallo, M. (2022), "Towards more sustainable tourism under a carbon footprint approach: the Camino Lebaniego case study", *Journal of Cleaner Production*, Vol. 369, 133222, doi: [10.1016/j.jclepro.2022.133222](https://doi.org/10.1016/j.jclepro.2022.133222).
- Carvache-Franco, O., Carvache-Franco, M., Carvache-Franco, W. and Iturralde, K. (2023), "Topic and sentiment analysis of crisis communications about the COVID-19 pandemic in Twitter's tourism hashtags", *Tourism and Hospitality Research*, Vol. 23 No. 1, pp. 44-59, doi: [10.1177/14673584221085470/ASSET/IMAGES/LARGE/10.1177_14673584221085470-FIG3.JPG](https://doi.org/10.1177/14673584221085470/ASSET/IMAGES/LARGE/10.1177_14673584221085470-FIG3.JPG).
- Chen, K., Duan, Z. and Yang, S. (2022), "Twitter as research data", *Politics and the Life Sciences*, Vol. 41 No. 1, pp. 114-130, doi: [10.1017/PLS.2021.19](https://doi.org/10.1017/PLS.2021.19).
- Chen, J., Becken, S. and Stantic, B. (2023), "Travel bubbles to maintain safe space for international travel during crisis - emotions reflected in Twitter posts", *Current Issues in Tourism*, Vol. 26 No. 15, pp. 2479-2493, doi: [10.1080/13683500.2022.2089546](https://doi.org/10.1080/13683500.2022.2089546).
- Choi, H. and Varian, H. (2012), "Predicting the present with google trends", *Economic Record*, Vol. 88 No. 1, pp. 2-9, doi: [10.1111/j.1475-4932.2012.00809.x](https://doi.org/10.1111/j.1475-4932.2012.00809.x).
- Chua, A., Servillo, L., Marcheggiani, E. and Moere, A.V. (2016), "Mapping Cilento: using geotagged social media data to characterize tourist flows in southern Italy", *Tourism Management*, Vol. 57, pp. 295-310, doi: [10.1016/j.TOURMAN.2016.06.013](https://doi.org/10.1016/j.TOURMAN.2016.06.013).

- Collins-Kreiner, N. (2020), "Pilgrimage tourism-past, present and future rejuvenation: a perspective article", *Tourism Review*, Vol. 75 No. 1, pp. 145-148, doi: [10.1108/TR-04-2019-0130](https://doi.org/10.1108/TR-04-2019-0130).
- de Ascaniis, S., Mutangala, M.M. and Cantoni, L. (2019), "ICTs in the tourism experience at religious heritage sites: a review of the literature and an investigation of pilgrims' experiences at the sanctuary of Loreto (Italy)", *Church, Communication and Culture*, Vol. 3 No. 3, pp. 310-334, doi: [10.1080/23753234.2018.1544835](https://doi.org/10.1080/23753234.2018.1544835).
- Elshendy, M., Fronzetti Colladon, A., Battistoni, E. and Gloor, P.A. (2017), "Using four different online media sources to forecast the crude oil price", *Journal of Information Science*, Vol. 44 No. 3, pp. 408-421, doi: [10.1177/0165551517698298](https://doi.org/10.1177/0165551517698298).
- Fernández-Poyatos, M.D., Aguirregoitia-Martínez, A. and Boix-Martínez, B. (2012), "The way of Saint James and the Xacobeo 2010 in the tourism websites of the Spanish autonomous communities", *Revista Latina de Comunicación Social*, Vol. 67, pp. 23-46, doi: [10.4185/RLCS-067-946-023-046](https://doi.org/10.4185/RLCS-067-946-023-046).
- Geldres-Weiss, S., Küster-Boluda, I. and Vila-López, N. (2023), "B2B value co-creation influence on engagement: twitter analysis at international trade show organizer", *European Journal of Management and Business Economics*, Vol. 32 No. 3, pp. 257-275, doi: [10.1108/EJMBE-04-2022-0121/FULL/PDF](https://doi.org/10.1108/EJMBE-04-2022-0121/FULL/PDF).
- Groves, R.M. (2011), "Three eras of survey research", *Public Opinion Quarterly*, Vol. 75 No. 5, pp. 861-871, doi: [10.1093/POQ/NFR057](https://doi.org/10.1093/POQ/NFR057).
- Guan, C., Liu, W. and Cheng, J.Y.C. (2022), "Using social media to predict the stock market crash and rebound amid the pandemic: the digital 'haves' and 'have-mores.'", *Annals of Data Science*, Vol. 9 No. 1, pp. 5-31, doi: [10.1007/S40745-021-00353-W/TABLES/9](https://doi.org/10.1007/S40745-021-00353-W/TABLES/9).
- Guizzardi, A. and Mazzocchi, M. (2010), "Tourism demand for Italy and the business cycle", *Tourism Management*, Vol. 31 No. 3, pp. 367-377, doi: [10.1016/J.TOURMAN.2009.03.017](https://doi.org/10.1016/J.TOURMAN.2009.03.017).
- Gunter, U. and Önder, I. (2015), "Forecasting international city tourism demand for Paris: accuracy of uni- and multivariate models employing monthly data", *Tourism Management*, Vol. 46, pp. 123-135, doi: [10.1016/J.TOURMAN.2014.06.017](https://doi.org/10.1016/J.TOURMAN.2014.06.017).
- Gunter, U. and Önder, I. (2016), "Forecasting city arrivals with google analytics", *Annals of Tourism Research*, Vol. 61, pp. 199-212, doi: [10.1016/j.annals.2016.10.007](https://doi.org/10.1016/j.annals.2016.10.007).
- Haq, F. and Jackson, J. (2009), "Spiritual journey to Hajj: australian and Pakistani experience and expectations", *Journal of Management, Spirituality and Religion*, Vol. 6 No. 2, pp. 141-156, doi: [10.1080/14766080902815155](https://doi.org/10.1080/14766080902815155).
- Hu, M., Li, H., Song, H., Li, X. and Law, R. (2022), "Tourism demand forecasting using tourist-generated online review data", *Tourism Management*, Vol. 90, 104490, doi: [10.1016/J.TOURMAN.2022.104490](https://doi.org/10.1016/J.TOURMAN.2022.104490).
- Hu, T., Wang, H., Law, R. and Geng, J. (2023), "Diverse feature extraction techniques in internet search query to forecast tourism demand: an in-depth comparison", *Tourism Management Perspectives*, Vol. 47, 101116, doi: [10.1016/J.TMP.2023.101116](https://doi.org/10.1016/J.TMP.2023.101116).
- Huang, L. and Zheng, W. (2023), "Hotel demand forecasting: a comprehensive literature review", *Tourism Review*, Vol. 78 No. 1, pp. 218-244, doi: [10.1108/TR-07-2022-0367/FULL/XML](https://doi.org/10.1108/TR-07-2022-0367/FULL/XML).
- Huang, X., Zhang, L. and Ding, Y. (2017), "The Baidu Index: uses in predicting tourism flows -A case study of the Forbidden City", *Tourism Management*, Vol. 58, pp. 301-306, doi: [10.1016/j.tourman.2016.03.015](https://doi.org/10.1016/j.tourman.2016.03.015).
- Husein, J. and Kara, S.M. (2020), "Nonlinear ARDL estimation of tourism demand for Puerto Rico from the USA", *Tourism Management*, Vol. 77, 103998, doi: [10.1016/J.TOURMAN.2019.103998](https://doi.org/10.1016/J.TOURMAN.2019.103998).
- Jamil, R.A., Qayyum, U., ul Hassan, S.R. and Khan, T.I. (2023), "Impact of social media influencers on consumers' well-being and purchase intention: a TikTok perspective", *European Journal of Management and Business Economics*, ahead-of-print(ahead-of-print), doi: [10.1108/EJMBE-08-2022-0270/FULL/PDF](https://doi.org/10.1108/EJMBE-08-2022-0270/FULL/PDF).
- Jiao, E.X. and Chen, J.L. (2019), "Tourism forecasting: a review of methodological developments over the last decade", *Tourism Economics*, Vol. 25 No. 3, pp. 469-492, doi: [10.1177/1354816618812588](https://doi.org/10.1177/1354816618812588).

-
- Klašnja, M., Barberá, P., Beauchamp, N., Nagler, J. and Tucker, J.A. (2018), "Measuring public opinion with social media data", in Atkeson, L.R. and Alvarez, R.M. (Eds), *The Oxford Handbook of Polling and Polling Methods*, Oxford University Press, pp. 555-582, doi: [10.1093/OXFORDHB/9780190213299.013.3](https://doi.org/10.1093/OXFORDHB/9780190213299.013.3).
- Kozak, M., Rita, P. and Bigné, E. (2018), "New frontiers in tourism: destinations, resources, and managerial perspectives", *European Journal of Management and Business Economics*, Vol. 27 No. 1, pp. 2-5, doi: [10.1108/EJMBE-03-2018-066/FULL/PDF](https://doi.org/10.1108/EJMBE-03-2018-066/FULL/PDF).
- Küster Boluda, I., Vila-Lopez, N., Mora, E. and Casanoves-Boix, J. (2024), "Social media impact on international sports events related to the brand Spain: a comparison between inner versus outside events", *European Journal of Management and Business Economics*, Vol. ahead-of-print No. ahead-of-print, doi: [10.1108/EJMBE-06-2023-0171](https://doi.org/10.1108/EJMBE-06-2023-0171).
- Lama, A., Singh, K.N., Singh, H., Shekhawat, R., Mishra, P. and Gurung, B. (2022), "Forecasting monthly rainfall of Sub-Himalayan region of India using parametric and non-parametric modelling approaches", *Modeling Earth Systems and Environment*, Vol. 8 No. 1, pp. 837-845, doi: [10.1007/S40808-021-01124-5/TABLES/3](https://doi.org/10.1007/S40808-021-01124-5/TABLES/3).
- Leung, X.Y., Sun, J. and Bai, B. (2021), "Social media research in hospitality and tourism: a causal chain framework of literature review", *Tourism and Hospitality Management*, Vol. 27 No. 3, pp. 455-477, doi: [10.20867/THM.27.3.1](https://doi.org/10.20867/THM.27.3.1).
- Li, X., Pan, B., Law, R. and Huang, X. (2017), "Forecasting tourism demand with composite search index", *Tourism Management*, Vol. 59, pp. 57-66, doi: [10.1016/J.TOURMAN.2016.07.005](https://doi.org/10.1016/J.TOURMAN.2016.07.005).
- Li, S., Chen, T., Wang, L. and Ming, C. (2018), "Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index", *Tourism Management*, Vol. 68, pp. 116-126, doi: [10.1016/J.TOURMAN.2018.03.006](https://doi.org/10.1016/J.TOURMAN.2018.03.006).
- Li, H., Hu, M. and Li, G. (2020), "Forecasting tourism demand with multisource big data", *Annals of Tourism Research*, Vol. 83, 102912, doi: [10.1016/j.annals.2020.102912](https://doi.org/10.1016/j.annals.2020.102912).
- Li, X., Law, R., Xie, G. and Wang, S. (2021), "Review of tourism forecasting research with internet data", *Tourism Management*, Vol. 83, 104245, doi: [10.1016/j.tourman.2020.104245](https://doi.org/10.1016/j.tourman.2020.104245).
- 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*, Vol. 1 No. 1, pp. 32-43, doi: [10.1016/J.JDEC.2022.08.006](https://doi.org/10.1016/J.JDEC.2022.08.006).
- Li, H., Gao, H. and Song, H. (2023a), "Tourism forecasting with granular sentiment analysis", *Annals of Tourism Research*, Vol. 103, 103667, doi: [10.1016/J.ANNALS.2023.103667](https://doi.org/10.1016/J.ANNALS.2023.103667).
- Li, M., Zhang, C., Sun, S. and Wang, S. (2023b), "A novel deep learning approach for tourism volume forecasting with tourist search data", *International Journal of Tourism Research*, Vol. 25 No. 2, pp. 183-197, doi: [10.1002/JTR.2558](https://doi.org/10.1002/JTR.2558).
- Lin, L.P. and Hsieh, W.K. (2022), "Exploring how perceived resilience and restoration affected the wellbeing of Matsu pilgrims during COVID-19", *Tourism Management*, Vol. 90, 104473, doi: [10.1016/J.TOURMAN.2021.104473](https://doi.org/10.1016/J.TOURMAN.2021.104473).
- Lin, H.H., Lin, T.Y., Hsu, C.W., Chen, C.H., Li, Q.Y. and Wu, P.H. (2022), "Moderating effects of religious tourism activities on environmental risk, leisure satisfaction, physical and mental health and well-being among the elderly in the context of COVID-19", *International Journal of Environmental Research and Public Health*, Vol. 19 No. 21, 14419, doi: [10.3390/IJERPH192114419](https://doi.org/10.3390/IJERPH192114419).
- Liu, Y.Y., Tseng, F.M. and Tseng, Y.H. (2018), "Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model", *Technological Forecasting and Social Change*, Vol. 130, pp. 123-134, doi: [10.1016/J.TECHFORE.2018.01.018](https://doi.org/10.1016/J.TECHFORE.2018.01.018).
- López, L. and Lois González, R.C. (2020), "New tourism dynamics along the way of st. James. From undertourism and overtourism to the post-COVID-19 era", in Pons, G.X., Blanco-Romero, A., Navalón-García, R., Troitiño-Torrallba, L. and y Blázquez-Salom, M. (Eds), *Sostenibilidad Turística: overtourism vs undertourism*. Societat d'Història Natural de Les Balears, Vol. 31, pp. 541-552, ISBN: 978-84-09-22881-2.

-
- López, L., Lois González, R.C. and Fernández Castro, M.B. (2017), "Spiritual tourism on the way of Saint James the current situation", *Tourism Management Perspectives*, Vol. 24, pp. 225-234, doi: [10.1016/J.TMP.2017.07.015](https://doi.org/10.1016/J.TMP.2017.07.015).
- Luna, A., Nunez-del-Prado, M., Talavera, A. and Holguin, E.S. (2017), "Power demand forecasting through social network activity and artificial neural networks", *2016 IEEE ANDESCON*, pp. 1-4, doi: [10.1109/ANDESCON.2016.7836248](https://doi.org/10.1109/ANDESCON.2016.7836248).
- Ma, S., Li, H., Hu, M., Yang, H. and Gan, R. (2023), "Tourism demand forecasting based on user-generated images on OTA platforms", *Current Issues in Tourism*, pp. 1-20, doi: [10.1080/13683500.2023.2216882](https://doi.org/10.1080/13683500.2023.2216882).
- Mariani, M.M. and Baggio, R. (2022), "Big data and analytics in hospitality and tourism: a systematic literature review", *International Journal of Contemporary Hospitality Management*, Vol. 34 No. 1, pp. 231-278, doi: [10.1108/IJCHM-03-2021-0301](https://doi.org/10.1108/IJCHM-03-2021-0301).
- Meehan, K., Lunney, T., Curran, K. and McCaughey, A. (2016), "Aggregating social media data with temporal and environmental context for recommendation in a mobile tour guide system", *Journal of Hospitality and Tourism Technology*, Vol. 7 No. 3, pp. 281-299, doi: [10.1108/JHTT-10-2014-0064/FULL/XML](https://doi.org/10.1108/JHTT-10-2014-0064/FULL/XML).
- Mittal, R. and Sinha, P. (2021), "Framework for a resilient religious tourism supply chain for mitigating post-pandemic risk", *International Hospitality Review*, Vol. 36 No. 2, pp. 322-339, doi: [10.1108/IHR-09-2020-0053](https://doi.org/10.1108/IHR-09-2020-0053).
- Nadeau, J., Wardley, L.J. and Rajabi, E. (2022), "Tourism destination image resiliency during a pandemic as portrayed through emotions on Twitter", *Tourism and Hospitality Research*, Vol. 22 No. 1, pp. 60-70, doi: [10.1177/14673584211038317](https://doi.org/10.1177/14673584211038317).
- Navío-Marco, J., Ruiz-Gómez, L.M. and Sevilla-Sevilla, C. (2018), "Progress in information technology and tourism management: 30 years on and 20 years after the internet - revisiting Buhalis & Law's landmark study about eTourism", *Tourism Management*, Vol. 69, pp. 460-470, doi: [10.1016/j.tourman.2018.06.002](https://doi.org/10.1016/j.tourman.2018.06.002).
- Nickerson, R., Austreich, M. and Eng, J. (2014), "Mobile technology and smartphone apps: a diffusion of innovations analysis", *20th Americas Conference on Information Systems*, Savannah, Georgia, pp. 1-12.
- Nofer, M. and Hinz, O. (2015), "Using twitter to predict the stock market: where is the mood effect?", *Business and Information Systems Engineering*, Vol. 57 No. 4, pp. 229-242, doi: [10.1007/S12599-015-0390-4/TABLES/5](https://doi.org/10.1007/S12599-015-0390-4/TABLES/5).
- Park, H., Seo, S. and Kandampully, J. (2015), "Why post on social networking sites (SNS)? Examining motives for visiting and sharing pilgrimage experiences on SNS", *Journal of Vacation Marketing*, Vol. 22 No. 4, pp. 307-319, doi: [10.1177/1356766715615912](https://doi.org/10.1177/1356766715615912).
- Park, E., Park, J. and Hu, M. (2021), "Tourism demand forecasting with online news data mining", *Annals of Tourism Research*, Vol. 90, 103273, doi: [10.1016/J.ANNALS.2021.103273](https://doi.org/10.1016/J.ANNALS.2021.103273).
- Payntar, N.D., Hsiao, W.L., Covey, R.A. and Grauman, K. (2021), "Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru", *Tourism Management*, Vol. 82, 104165, doi: [10.1016/J.TOURMAN.2020.104165](https://doi.org/10.1016/J.TOURMAN.2020.104165).
- Philander, K. and Zhong, Y.Y. (2016), "Twitter sentiment analysis: capturing sentiment from integrated resort tweets", *International Journal of Hospitality Management*, Vol. 55, pp. 16-24, doi: [10.1016/j.ijhm.2016.02.001](https://doi.org/10.1016/j.ijhm.2016.02.001).
- Piramanayagam, S. and Seal, P.P. (2022), "Geographical Indication (GI) tagged foods and promotion of gastronomic tourism: a developing country perspective", in *Current Issues in Tourism, Gastronomy, and Tourist Destination Research*, Routledge, pp. 393-399, doi: [10.1201/9781003248002-52](https://doi.org/10.1201/9781003248002-52).
- Polyzos, S., Samitas, A. and Spyridou, A.E. (2020), "Tourism demand and the COVID-19 pandemic: an LSTM approach", *Tourism Recreation Research*, Vol. 46 No. 2, pp. 175-187, doi: [10.1080/02508281.2020.1777053](https://doi.org/10.1080/02508281.2020.1777053).

-
- Punel, A. and Ermagun, A. (2018), "Using Twitter network to detect market segments in the airline industry", *Journal of Air Transport Management*, Vol. 73, pp. 67-76, doi: [10.1016/j.jairtraman.2018.08.004](https://doi.org/10.1016/j.jairtraman.2018.08.004).
- Qiu, R.T.R., Liu, A., Stienmetz, J.L. and Yu, Y. (2021), "Timing matters: crisis severity and occupancy rate forecasts in social unrest periods", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2044-2064, doi: [10.1108/IJCHM-06-2020-0629/FULL/XML](https://doi.org/10.1108/IJCHM-06-2020-0629/FULL/XML).
- Rita, P., Vong, C., Pinheiro, F. and Mimoso, J. (2022), "A sentiment analysis of Michelin-starred restaurants", *European Journal of Management and Business Economics*, Vol. 32 No. 3, pp. 276-295, doi: [10.1108/EJMBE-11-2021-0295/FULL/PDF](https://doi.org/10.1108/EJMBE-11-2021-0295/FULL/PDF).
- Roy, K.C., Hasan, S., Culotta, A. and Eluru, N. (2021), "Predicting traffic demand during hurricane evacuation using Real-time data from transportation systems and social media", *Transportation Research C: Emerging Technologies*, Vol. 131, 103339, doi: [10.1016/j.trc.2021.103339](https://doi.org/10.1016/j.trc.2021.103339).
- Saz, G. (2011), "The efficacy of SARIMA models for forecasting inflation rates in developing countries: the case for Turkey", *International Research Journal of Finance and Economics*, Vol. 62, pp. 111-142.
- Sigala, M. (2015), "Social media marketing in tourism and hospitality", *Information Technology & Tourism*, Vol. 15 No. 2, pp. 181-183, doi: [10.1007/S40558-015-0024-1](https://doi.org/10.1007/S40558-015-0024-1).
- Smeral, E. and Song, H. (2015), "Varying elasticities and forecasting performance", *International Journal of Tourism Research*, Vol. 17 No. 2, pp. 140-150, doi: [10.1002/JTR.1972](https://doi.org/10.1002/JTR.1972).
- Song, H. and Li, G. (2021), "Editorial: tourism forecasting competition in the time of COVID-19", *Annals of Tourism Research*, Vol. 88, 103198, doi: [10.1016/j.annals.2021.103198](https://doi.org/10.1016/j.annals.2021.103198).
- Song, H., Qiu, R.T.R. and Park, J. (2019), "A review of research on tourism demand forecasting", *Annals of Tourism Research*, Vol. 75, pp. 338-362, doi: [10.1016/j.annals.2018.12.001](https://doi.org/10.1016/j.annals.2018.12.001).
- Stojanovic, I., Andreu, L. and Curras-Perez, R. (2018), "Effects of the intensity of use of social media on brand equity: an empirical study in a tourist destination", *European Journal of Management and Business Economics*, Vol. 27 No. 1, pp. 83-100, doi: [10.1108/EJMBE-11-2017-0049/FULL/PDF](https://doi.org/10.1108/EJMBE-11-2017-0049/FULL/PDF).
- Stylos, N., Zwiegielaar, J. and 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*, Vol. 33 No. 3, pp. 1015-1036, doi: [10.1108/IJCHM-07-2020-0644](https://doi.org/10.1108/IJCHM-07-2020-0644).
- Sulong, Z., Abdullah, M. and Chowdhury, M.A.F. (2022), "Halal tourism demand and firm performance forecasting: new evidence from machine learning", *Current Issues in Tourism*, Vol. 26 No. 23, pp. 1-17, doi: [10.1080/13683500.2022.2145458](https://doi.org/10.1080/13683500.2022.2145458).
- Teixeira, J.P. and Gunter, U. (2023), "Editorial for special issue: 'tourism forecasting: time-series analysis of world and regional data.'", *Forecasting*, Vol. 5 No. 1, pp. 210-212, doi: [10.3390/FORECAST5010011](https://doi.org/10.3390/FORECAST5010011).
- Terzidou, M., Scarles, C. and Saunders, M.N.K. (2018), "The complexities of religious tourism motivations: sacred places, vows and visions", *Annals of Tourism Research*, Vol. 70, pp. 54-65, doi: [10.1016/j.annals.2018.02.011](https://doi.org/10.1016/j.annals.2018.02.011).
- Tromble, R. (2019), "In search of meaning: why we still don't know what digital data represent", *Journal of Digital Social Research*, Vol. 1 No. 1, pp. 17-24, doi: [10.33621/JDSR.V1I1.8](https://doi.org/10.33621/JDSR.V1I1.8).
- Utkarsh and Sigala, M. (2021), "A bibliometric review of research on COVID-19 and tourism: reflections for moving forward", *Tourism Management Perspectives*, Vol. 40, 100912, doi: [10.1016/j.tmp.2021.100912](https://doi.org/10.1016/j.tmp.2021.100912).
- Vázquez, C.R., Lozano, F.B. and Pollán, M.M. (2020), "Cultural Tourism in Social Media, the paradigm of the Camino de Santiago Francés", *15th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE Computer Society, pp. 1-6, doi: [10.23919/CISTI49556.2020.9140955](https://doi.org/10.23919/CISTI49556.2020.9140955).
- Vila, N.A., Cardoso, L., de Araújo, A.F. and Fraiz Brea, J.A. (2020), "Pilgrimage or tourism? Travel motivation on way of Saint James", *International Journal of Tourism Anthropology*, Vol. 8 No. 1, pp. 1-21, doi: [10.1504/IJTA.2020.113922](https://doi.org/10.1504/IJTA.2020.113922).
- Wang, Y., Liu, J., Huang, Y. and Feng, X. (2016), "Using hashtag graph-based topic model to connect semantically-related words without Co-occurrence in microblogs", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28 No. 7, pp. 1919-1933, doi: [10.1109/TKDE.2016.2531661](https://doi.org/10.1109/TKDE.2016.2531661).

-
- Wickramasinghe, K. and Ratnasiri, S. (2021), "The role of disaggregated search data in improving tourism forecasts: evidence from Sri Lanka", *Current Issues in Tourism*, Vol. 24 No. 19, pp. 2740-2754, doi: [10.1080/13683500.2020.1849049](https://doi.org/10.1080/13683500.2020.1849049).
- Wu, B., Wang, L., Wang, S. and Zeng, Y.R. (2021), "Forecasting the U.S. oil markets based on social media information during the COVID-19 pandemic", *Energy*, Vol. 226, 120403, doi: [10.1016/j.ENERGY.2021.120403](https://doi.org/10.1016/j.ENERGY.2021.120403).
- Wu, E.H.C., Hu, J. and Chen, R. (2022), "Monitoring and forecasting COVID-19 impacts on hotel occupancy rates with daily visitor arrivals and search queries", *Current Issues in Tourism*, Vol. 25 No. 3, pp. 490-507, doi: [10.1080/13683500.2021.1989385](https://doi.org/10.1080/13683500.2021.1989385).
- Wu, X.X., Shi, J. and Xiong, H. (2023), "Tourism forecasting research: a bibliometric visualization review (1999-2022)", *Tourism Review*, Vol. 79 No. 2, pp. 465-486, doi: [10.1108/TR-03-2023-0169/FULL/XML](https://doi.org/10.1108/TR-03-2023-0169/FULL/XML).
- Xin, Y. and MacEachren, A.M. (2020), "Characterizing traveling fans: a workflow for event-oriented travel pattern analysis using Twitter data", *International Journal of Geographical Information Science*, Vol. 34 No. 12, pp. 2497-2516, doi: [10.1080/13658816.2020.1770259](https://doi.org/10.1080/13658816.2020.1770259).
- Xu, H., Zhang, N. and Zhou, L. (2020), "Validity concerns in research using organic data", *Journal of Management*, Vol. 46 No. 7, pp. 1257-1274, doi: [10.1177/0149206319862027/ASSET/IMAGES/LARGE/10.1177_0149206319862027-FIG1.JPEG](https://doi.org/10.1177/0149206319862027/ASSET/IMAGES/LARGE/10.1177_0149206319862027-FIG1.JPEG).
- Yang, M. and Han, C. (2021), "Revealing industry challenge and business response to Covid-19: a text mining approach", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 4, pp. 1230-1248, doi: [10.1108/IJCHM-08-2020-0920/FULL/PDF](https://doi.org/10.1108/IJCHM-08-2020-0920/FULL/PDF).
- Yang, J.S., Zhao, C.Y., Yu, H.T. and Chen, H.Y. (2020), "Use GBDT to predict the stock market", *Procedia Computer Science*, Vol. 174, pp. 161-171, doi: [10.1016/j.PROCS.2020.06.071](https://doi.org/10.1016/j.PROCS.2020.06.071).
- Yang, Y., Fan, Y., Jiang, L. and Liu, X. (2022), "Search query and tourism forecasting during the pandemic: when and where can digital footprints be helpful as predictors?", *Annals of Tourism Research*, Vol. 93, 103365, doi: [10.1016/j.annals.2022.103365](https://doi.org/10.1016/j.annals.2022.103365).
- Zhang, H., Song, H., Wen, L. and Liu, C. (2021), "Forecasting tourism recovery amid COVID-19", *Annals of Tourism Research*, Vol. 87, 103149, doi: [10.1016/j.ANNALS.2021.103149](https://doi.org/10.1016/j.ANNALS.2021.103149).

About the authors

Adrián Mendieta-Aragón. He obtained his Ph.D. in Economics from UNED. His research focuses on different fields of Digital Tourism, with a particular interest in consumer behaviour and social networks. He has published in international refereed journals, including, *Tourism Review*. Adrián Mendieta-Aragón is the corresponding author and can be contacted at: amendieta@cee.uned.es

Julio Navío-Marco. MSc in Telecommunications Engineering; B.A and Ph.D. in Economics and Business Administration at the UNED; Postgraduate in IESE Business School. Julio Navío is Professor of Business Organization and Digital Economy at the UNED in Spain. EU Jean Monnet Chairholder in Digital Economy. Dr Navío is also Expert for the EC Directorate-General for Regional and Urban Policy (DG REGIO) and H2020. Dr Navío was Deputy Dean of the Spanish College of Telecommunication Engineers and Vice-President of the Spanish Association of Telecommunication

Teresa Garín-Muñoz. Full Professor of Economics at the National Distance Education University (UNED) in Spain. She has been Visiting Scholar at the University of California, San Diego. Her research interests are in microeconomics (demand modelling, consumer satisfaction and consumer protection). Most of her research has been devoted to the areas of tourism and telecommunications. She has published in international refereed journals, including, *Tourism Management*, *Tourism Economics*, *International Journal of Tourism Research*, *Telecommunications Policy*, *Information Economics and Policy* and *Applied Economics*.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com