Enlightening Tourism.

A Pathmaking Journal

Universidad de Huelva

Enlightening Tourism. A Pathmaking Journal
Universidad de Huelva

MODELLING TOURISM DEMAND USING GOOGLE ANALYTICS: A CASE STUDY OF PORTUGAL'S ALENTEJO REGION

Gorete Dinis
Polytechnic Institute of Portalegre (Portugal)
gdinis@ipportalegre.pt

Celeste Eusébio
University of Aveiro (Portugal)
celeste.eusebio@ua.pt

Zélia Breda
University of Aveiro (Portugal)
zelia@ua.pt

Ana Madaleno
University of Aveiro (Portugal)
ana.madaleno@ua.pt

ABSTRACT

The development of information and communication technologies, specifically the Internet, has changed the way tourists plan their trips, and is one of the most important information sources for tourism decision-making. However, a limited number of studies have been carried out to analyse the causal relationships between web interaction and tourism demand. Therefore, this paper intends to shed light on the usefulness of big data analytics to understand the tourism demand of a destination. More specifically, it aims to examine the causal relationship between website *visitor interactions and the tourism demand of a destination, and ascertain whether there are differences in this relationship according to the visitors' country of origin. In order to achieve the research objectives, the Alentejo region in Portugal was selected as a case study. Monthly data for the period between 2007 and 2017 was used to examine the long-run causal relationship between the sessions of the users to the official website of the Destination Management Organization of Alentejo (measured through Google Analytics) and tourism demand of this region (measured trough the number of guests in tourism accommodation establishments). To analyse whether there are differences in this relationship according to the country of origin of visitors, the most important tourism markets for this destination were selected. Cointegration (Johansen's maximum-likelihood method), the Granger causality test, a vector autoregression model, and a vector error correction model were used to examine the relationship. The results reveal a causal relationship between Internet searches and tourism demand. However, this relationship varies within the tourism market analysed.*

KEYWORDS
Tourism demand; Internet data; Google Analytics; Portugal; Destination Management Organization; Modelling

ECONLIT KEYS
Z3; Z33; B23

1. INTRODUCTION

With the widespread development and use of the Internet in recent years, very large sets of data are produced and stored, giving rise to a new era, called big data. This new era of data represents a huge opportunity, a game changer, and fuels competitiveness and innovation in tourism (Mariani, 2019). It results in more efficient and effective operations and enhanced decision-making (Sigala, Rahimi & Thewall, 2019). Moreover, tourist behaviour and tourism markets can be better explored and understood (Li, Xu, Tang, Wang & Li, 2018).

Although the origin of the big data concept appeared in the late 1990s (Cox & Ellsworth, 1997), there is still no consensus on this concept among researchers (Li et al., 2018). Nonetheless, one of the first definitions presented by Laney (2001) refers to the existence of three properties or characteristics of big data: volume (size of data), velocity (rapidity of data generation, modification and speed of data transfer), and variety (different types of data and different formats/structures). Later, a fourth feature was added by Gantz and Reinsel (2011) – value (the process of extracting valuable knowledge from data, known as big data analytics). More recently, Bello-Orgaz Bello-Orgaz, Jung and Camacho (2016) added a fifth feature – veracity (related to the governance of data and privacy concerns).

Therefore, among other aspects, talking about big data implies considering the source and variety of data. This involves new technologies that create, communicate, or are involved with data-generating activities, which produce different types of data (Kudyba & Kwatinetz, 2014), collected via sensors, smartphones or social networks, such as videos, images, text, audio and data logs (Bello-Orgaz et al., 2016).

Li et al. (2018) conducted a study about the literature review on different types of big data in tourism research, and they concluded that, from the point of view of data source, big data in tourism fall into three primary categories: (i) user-generated content (UGC) (user), (ii) device data, and (iii) transaction data (operations). UGC is the most used type of big data in tourism research (47% of the studies), and transaction data (17%) is the least used. Transaction data include, for example, web search data, web page visit data, and online booking; however, web search data make up the largest share (11%) (Li et al., 2018).

One explanation for this distribution of data sources is data availability. Transaction data, although being attractive to researchers, have been used far less in tourism because most of these data consist of private information, which is only in the possession or control of tourism organisations or governments. It means that there is still scope to develop tourism research using big data, and particularly transaction data, which had made a relatively small contribution to tourism research so far. Therefore, reciprocal cooperation between academia and the industry is needed, promoting research with this type of data source and, in return, effectively addressing practical problems in the tourism industry (Li et al., 2018).

The purpose of this paper is to examine the causal relationship between web interactions and tourism demand of a destination and to ascertain whether there are differences in this relationship according to visitors' country of origin. For this purpose, the Alentejo region in Portugal was selected as a case study and transaction data were used, namely, the website traffic data of destination management organisations (DMOs). Monthly data for a 10-year period were used to examine this long-run causal relationship. The data collected concerns the visitors to the official website of the Alentejo (measured through Google Analytics) and the tourism demand of this region (measured through the number of guests in tourism accommodation establishments). Data were subjected to econometric estimations, using the statistical software package Stata-Data Analysis and Statistical Software-12. Cointegration (Johansen's maximum-likelihood method), the Granger causality test, a vector autoregression model (VAR) and a vector error correction model (VECM) were used to examine the relationship between web interactions and tourism demand.

The paper is organised as follows: first, the literature on the importance of DMOs' web visiting data as an information source in tourism is reviewed, following which modelling tourism demand using web transaction data is addressed. Subsequently, the methodology used for the empirical study is described in more detail. Finally, the research findings are reported and discussed, and their implications for further research and applications are highlighted.

2. LITERATURE REVIEW

2.1) DMOS' WEB VISITING DATA AS AN INFORMATION SOURCE IN TOURISM

In recent years, the Internet has grown in popularity, making it an important tool for searching for information and product purchasing (Steinbauer & Werthner, 2007). In 2018, in the European Union (EU 28), 85% of individuals used the Internet in the previous three months (Eurostat, 2019a), with 82% of them using it to find information about services and goods (Eurostat, 2019b). According to a survey about the preferences of Europeans towards tourism, websites were mentioned as important information sources for travel decision-making by 44% of the respondents (EU 28) (European Union, 2016). Websites run by a service provider or by a destination scored differently according to the origin of travellers, especially for the Finns (39%), Dutch (26%) and Austrians (26%).

Outside Europe, DMO websites remain an important resource for travellers, being amongst the most trusted and valued, used by one-third of American leisure travellers to research and/or plan travel. According to the same study, 78.0% of DMO website users visited the destination after using the site (DMA West, n.d.). Similarly, So and Morrison (2003) found in their study that a significantly higher proportion of individuals visiting a destination for the first time were more likely to visit the DMO's website. As mentioned by Yang, Pan and Song (2014, p. 435), these results indicate that "the traffic volume of a DMO website could indicate the interests and purchase intentions of potential tourists".

Tourist consumer behaviour significantly changed due to the increasingly widespread use of information and communication technologies (Parra-López, Gutiérrez-Taño, Díaz-Armas & Bulchand-Gidumal, 2016). Online information sources have a strong influence on tourist behaviour, with tourists using the Internet in all phases of the travel cycle, from trip planning and organisation to experience sharing (Kim & Fesenmaier, 2008).

The most used sources to find information about tourism destinations are websites and social networks (Buhalis & Law, 2018). These online sources allow a large amount of data on consumers' needs, behaviour and perceptions, to be gathered; however, tourism destinations seldom use these valuable knowledge sources (Fuchs, Höpken & Lexhagen, 2014). Examples are tourists' website navigation, transaction or survey data (Höpken, Fuchs, Keil & Lexhagen, 2015), which organisations could make sense of through data analytics (Centobelli & Ndou, 2019). Analytics are important tools to convert data into valuable information for organisations. "These tools are a relevant source of information that should be exploited by DMOs, especially by small- and medium-sized ones, to define and measure the [key performance indicators] KPIs of the experiences" (Garcia, Linaza, Gutierrez & Garcia, 2018, p. 46).

Travellers desire a higher level of engagement with the tourism destination at all stages of the travel decision-making process, thus requiring innovative new approaches to tourism marketing, customer and network relationships. This means developing a "model based on communities, networks, openness, peering, sharing, collaboration, customer empowerment, thinking and acting globally" (Hamill, Stevenson & Attard, 2012, p. 112). It is therefore crucial that DMOs understand consumer choices, concerns and determinants. To that end, it is necessary to collect consumer information at different stages of the travel cycle (Buhalis & O´Connor, 2006).

The travel and hospitality industry has actively adopted the Internet as a new distribution channel, as well as an advertising medium, since the mid-1990s (Xiang, Wang, O'Leary & Fesenmaier, 2015). Tourism organisations are among the pioneers in adopting the Internet and e-commerce (Buhalis, 2003). Since that time, DMOs have embraced the Internet as a channel for promotion and communication, leading to the decline of traditional marketing channels (So & Morrisson, 2003).

Crouch (2000) mentions that few relevant tourism destinations currently do not have a website. DMOs' investment in websites has increased significantly in recent years, and the marketing activities of today's organisations are conducted online (UNWTO, 2005). Currently, websites provide a hub for DMO e-marketing activity, and web analytics represents a core business intelligence tool for anyone involved in marketing (UNWTO & ETC, 2014). However, probably only a small part of DMOs are able to know if the website is successful (Phippen, Sheppard, & Furnell, 2004), because the type of analysis they perform is superficial, focusing only on basic metrics. Very few use the information obtained to increase usability, improve the performance of their sites and/or to calculate the return on investment for online campaigns (UNWTO, 2005).

However, since consumers are the key for the website's success, it is necessary to understand how they interact with it and, subsequently, develop measures aimed at consumers (Phippen et al., 2004). Thus, DMO websites should smooth the path through the customer journey, and support the user through the stages of the travel cycle (UNWTO & ETC, 2014). To define a tourism strategy for a region, the DMO needs to know the visitor profile and behaviour, and destination web traffic volumes may indicate potential tourists' interests and act as leading indicators of tourist visits (Yang et al., 2014). Therefore, this information is of the utmost relevance to allow tourism demand forecasting.

2.2) MODELLING TOURISM DEMAND USING WEB TRANSACTION DATA

Consumers and tourism organisations increasingly use the Internet. The digital footprint that consumers leave on different online marketing channels can provide insights into their intentions, needs and behaviour (Ettredge, Gerdes & Karuga, 2005). This huge amount and variety of data can provide diverse and up-to-date information about the consumer in tourism that is relevant to different travel stages. "The challenge remains however on shifting the attention from a 'big' to a 'smart' usage of these data, adding layers of information, facilitating real-time usage and appropriate dissemination of trends" (Volo, 2019, p. 307).

Using big data in tourism has advantages over traditional methodologies, namely research based on surveys, since the data is real-time and reliable, providing high-frequency information (e.g., daily or weekly) and covering other types of information produced by tourists (Yang et al., 2014). For this reason, it can be used as a complement to traditional data sources (Choi & Varian, 2012), compensating for the limitations of tourism demand forecasting when using traditional data.

Enlightening Tourism. A Pathmaking Journal, Vol 12, No 1 (2022), pp. 177-212 ISSN 2174-548X

One of the most innovative uses of big data is nowcasting, which means that real-time data are used to describe contemporaneous activities before official data sources are made available (Song & Liu, 2017). It is a "special case of forecasting as it deals with the knowledge of the present, immediate past and very near future" (Antolini & Grassini, 2019, p. 2385).

Nowadays, the environment in which organisations operate is complex, fast and competitive, and they must adapt to change and seek to promote the future with some degree of precision (Hanke & Wichern, 2005). In this context, it is important to combine the latest computational capabilities with modern forecasting techniques (Hanke & Wichern, 2005), that help organisations to anticipate customer needs and prevent problems. The advanced analytics derived from big data can provide solutions with special added value in the same key areas, namely market forecasting (Vriens & Kidd, 2014 cited in Nicolau, 2017). Timely use of big data for forecasting and decision-making using proper approaches and methods is the best way to capitalise on the benefits of big data (Song & Liu, 2017).

Two primary types of Internet big data have appeared in the tourism demand forecasting literature, namely, search query data and website traffic data. The prevailing form of Internet big data used in tourism forecasting is search query data, specifically generated from search engines, such as Google and Baidu (Li, Hu & Li, 2020).

According to Li et al. (2018), the most popular web search data used in tourism demand forecasting is provided by Google Trends and Baidu (the latter specifically in the case of China) (Clark, Wilkins, Dagan, Powell, Sharp & Hillis, 2019; Sun, Wei, Tsui & Wang, 2019; Yang, Pan, Evans & Lv, 2015). The same conclusion was obtained by Dinis, Breda, Costa and Pacheco (2019), in a literature review study on the use of search engine data on tourism and hospitality research, namely Google Insights for Search and Google Trends. Their findings point out that most of the studies under analysis use Google Trends data to forecast or nowcast tourism demand. Although the methods used in the studies are diverse, they are consensual in stating that forecasting tourism demand with Google Trends data improves the prediction capacity of the models used (e.g. Bangwayo-Skeete & Skeete, 2015; Choi & Varian, 2012; Kim & Malek, 2018; Rivera, 2016; Siliverstovs & Wochner, 2018). These results are also in line with Li et al. (2018), who concluded that web search data could be used as powerful predictors for tourism demand. However, this research area is still at an early stage and there is still a lot of room for improvement, namely in terms of research area expansion and analytic technique innovation (Li et al., 2018).

Several scholars have used website data in their research, namely from Google Analytics. Google Analytics data has been used in studies related to healthcare websites (Clark, Nicholas & Jamili, 2014; Crutzen, Roosjen & Poelman, 2013; McCloskey, Johansson, Harvey & Compston, 2017; Song et al., 2018; Tell, Anderberg, Olander & Berglund, 2018; Vogel, Kleib, Davidson & Scott, 2016; Vona, Wilmoth, Jaycox & McMillen, 2014); libraries (Fang, 2007; Turner, 2010; Vélez & Pagán, 2011); e-commerce (Hasan, Morris & Probets, 2009); food composition (Pakkalaa, Presser & Christensen, 2012); tourism (Artola, Pinto & de Pedraza García, 2015; Dinis, Costa & Pacheco, 2012, 2016, 2017; Gunter & Önder, 2016; Pan & Yang, 2017; Plaza, 2011; Plaza, Gonzalez Casimiro, Moral Zuazo & Ostolaza, 2011; Yang et al., 2014); and art-related humanities and social sciences (Plaza, 2009a, 2009b, 2010).

In the field of tourism, Plaza et al. (2011) used the same methodology applied in the study of Plaza (2009a), which consists of combining Google Analytics data with time series methodology, namely an autoregressive model, to analyse whether visit behaviour and length of sessions depend on the traffic source. Dinis et al. (2012, 2016) explored Google Analytics data to understand the behaviour of potential tourists from analysis of the Alentejo DMO website. Furthermore, Dinis et al. (2017) correlated Google Analytics data with official tourism demand statistics and with the search volume index made available by Google Trends. They concluded that there is a strong correlation between visits to the website and the search volume index on the Alentejo region carried out by individuals located in Portugal (Google_Alentejo).

Yang et al. (2014) used the web traffic from a DMO, collected from Google Analytics, to predict demand for hotel rooms in a tourist area. This study validates the power of Google Analytics data, using the ARMAX model, which is particularly useful in short-run predictions of total room nights sold.

Pan and Yang (2017) combined different big data sources, including search engine and website traffic data, to forecast weekly hotel occupancy for Charleston, South Carolina. By integrating web traffic data into the traditional time series model, they demonstrated that short-term forecasts of hotel room demand could be significantly improved. The results revealed the high value of web traffic data in improving the accuracy of forecasting hotel demand.

Gunter and Önder (2016) concluded that Google Analytics data is useful not just for DMO website performance, but also to predict actual tourism demand to a destination. They investigated how website traffic data can be used by DMO managers to forecast tourism demand, and demonstrated that the forecast accuracy benefits from the inclusion of Google Analytics website traffic indicators. Another study highlighting the benefits of these data is that of Clark et al. (2014), who explored the general advantages of Google Analytics to replace log analysis systems as tools to obtain information about users. Moral, Gonzalez, and Plaza (2014) analysed the visibility and performance of a website of a regional association of tourism SMEs that advertises on Google. They measured KPIs to test the effectiveness of online marketing using the data provided by Google Analytics. Moreover, Kuo and Chuang (2016), aiming to measure the effect of a gamification initiative, analysed both surveys and Google Analytics to measure its impact on online academic dissemination, based on the behaviour of users when accessing the platform.

Both qualitative and quantitative methods may be used to examine the relationship between the website traffic and the tourism demand of a tourism destination (Frechtling, 2001; Sun et al., 2019). Quantitative methods have been more frequently used, mainly time series models. Recently, with the proliferation of the Internet and with technological development, artificial intelligence techniques and modern economic approaches have been used, such as cointegration and causality tests (Song & With, 2000; Sun et al., 2019; Yang et al., 2015). However, the number of studies that simultaneously use Granger causality and cointegration tests and big data to forecast tourism demand of a destination is still scarce. The studies of Yang et al. (2015) and of Sun et al. (2019) are the few studies that use these techniques to examine the relationship between the Internet search index and tourist arrivals. Both of these studies use Google Trends and Baidu to forecast the tourism demand in two tourism destinations in China, Hainan Province (Yang et al., 2015) and Beijing (Sun et al., 2019). To the best of our knowledge, no study uses Google Analytics to forecast tourism demand through cointegration and causality tests.

3. METHODOLOGY

This empirical research, which is quantitative in nature, aims to examine the causal relationship between website visitor interactions and the tourism demand of a tourism destination, and ascertain whether there are differences in this relationship according to the country of origin of the tourists. The Alentejo region in Portugal was selected as a case study.

3.1) BRIEF CHARACTERISATION OF THE REGION UNDER ANALYSIS

Portugal is divided into seven regional tourism areas for tourism purposes: Porto and North of Portugal; Centre of Portugal; Lisbon and Tagus Valley; Alentejo/Ribatejo; Algarve; Madeira; and Azores. The Alentejo is the region under analysis in this study. It is located in the centre-south of Portugal and coincides geographically with the administrative boundaries defined in Portugal for statistical purposes, i.e. according to the Common Nomenclature of Territorial Unit for statistical purposes (NUTS II – Alentejo). This region, in terms of area, is the largest of Portugal, integrating five subregions (NUTS III): Alentejo Litoral; Alto Alentejo; Alentejo Central; Baixo Alentejo; and Lezíria do Tejo.

The Alentejo region was chosen because, between 2013 and 2016, this region and the North of Portugal were those that presented the highest relative growth in terms of the number of overnight stays (INE, 2017). In 2017, the Alentejo region received approximately 1.4 million guests and 2.5 million of overnight stays (INE, 2018). This region has a unique and genuine territory, owing its recent tourism development, in part, to the efforts of the DMO Turismo do Alentejo. Its communication strategy and online marketing actions carried out in 2019 were developed according to the different market niches of Alentejo. An intervention on the website (turismodoalentejo.com.br) is planned regarding the Brazilian market. Although this region has greatly improved its tourist performance in recent years, reducing its dependence on the national market, in the period under analysis, this continues to be significant (representing about 67% of guests) (INE, 2018).

The promotion of Alentejo is of the responsibility of Turismo do Alentejo and the Regional Agency for Tourism Promotion of Alentejo (RATPA). These public organisations work closely together, and they are the main providers of tourist information about the destination via the web, and are therefore responsible for the official tourism destination website (www.visitalentejo.pt), which will be the focus of this study.

According to four surveys conducted in 2011 and 2012 with a stratified sample of visitors in 14 locations in the region, this website is the most used online source of information by visitors for their trips to Alentejo (Milheiro & Dinis, 2013). The website was launched in 2005 and, at the end of 2011, it underwent a major overhaul. It was transformed into an innovative platform, allowing visitors to create their own tours and share information on social networks. It is available in six languages (Portuguese, Spanish, English, French, German and Dutch) and has a mobile version. Since then, no significant changes have been observed (Dinis, Barreiro & Breda, 2020).

3.2) DATA COLLECTION

This study combines two types of data sources: (i) official tourism statistics, disseminated by the national statistics institute (Statistics Portugal); and (ii) website visitor transactions, namely data gathered in the Alentejo tourism promotional website (http://www.visitalentejo), which has been tracked by Google Analytics since 2007.

Statistics Portugal collects and publishes monthly data from accommodation establishments. One of the indicators is the number of guests, which is used in this study as a proxy to measure the tourism demand in the Alentejo region. To examine if there are differences in the relationship between website interaction and tourism demand, the most important tourism markets were selected: the domestic market, which in 2017 represented 67% of the total number of guests in tourism accommodation establishments, and the most important foreign markets (representing 73% of the total foreign guests) – Spain (20%), France (11%), Brazil (10%), Germany (9%), United States of America (USA) (7%), United Kingdom (UK) (6%), the Netherlands (5%) and Italy (5%) (INE, 2018).

The transaction data was collected from the Google Analytics account of the DMO website. Google Analytics is a free tool, offered by Google Inc., that can be used by anyone that owns a website, allowing them to measure and report visitor activity on the website, and evaluate the organisation's online activities (Tonkin, Whitmore & Cutroni, 2010). In terms of website traffic analytics tools, Google Analytics is the market leader, presenting an 85.3% market share (W3Techs, 2019).

After placing the tracking code on the various website pages, Google Analytics starts the data collection process that begins when the visitor browses this page. These data are processed, using the visitor's computer identification (the IP address), and transformed into reports (about 80) (Teixeira, 2010), which are generated on demand, that is, when requested by the administrator or owner of the site (Tonkin et al., 2010). In this study, we used data from the “target audience” report by geographical location, and the metrics on “sessions”. A session is the period of time that a user is actively interacting with the website; in case of inactivity for 30 minutes or more, any future activity is considered a new session (Google, 2019). Only the owner or administrator of the website can access these reports, and this study was only possible because the DMO allowed access to the data for research purposes. Data used in the study spans from 2007 to 2017, corresponding, respectively, to the first year when Google Analytics was introduced to the website and the last year with official statistics available at the time the research started.

Website traffic data has an advantage over other types of data, such as search query data available through Google Trends, because: (i) it is raw data; (ii) it does not vary depending on the day of the sample collection; and (iii) it is not necessary to define which queries are most suitable and in which languages. Therefore, website visitor interactions might possess stronger predictive power than search engine query volumes (Yang et al., 2014).

Monthly data between April 2007 and December 2017 were used to examine the long-run causal relationship between web interactions with the official DMO website and tourism demand. To analyse if there are differences in this relationship according to the country of origin of visitors (based on the IP address), the most important tourism markets for this destination were selected, according to the official statistics. As often used in similar studies, all the series are expressed in logarithms to facilitate the interpretation and to reduce the impact of outliers (Sun et al., 2019; Yang et al., 2015).

3.3) DATA ANALYSIS

In order to examine the relationship between the level of interaction with the DMO website and the tourism demand, the methodological procedure described in Figure 1 was used.

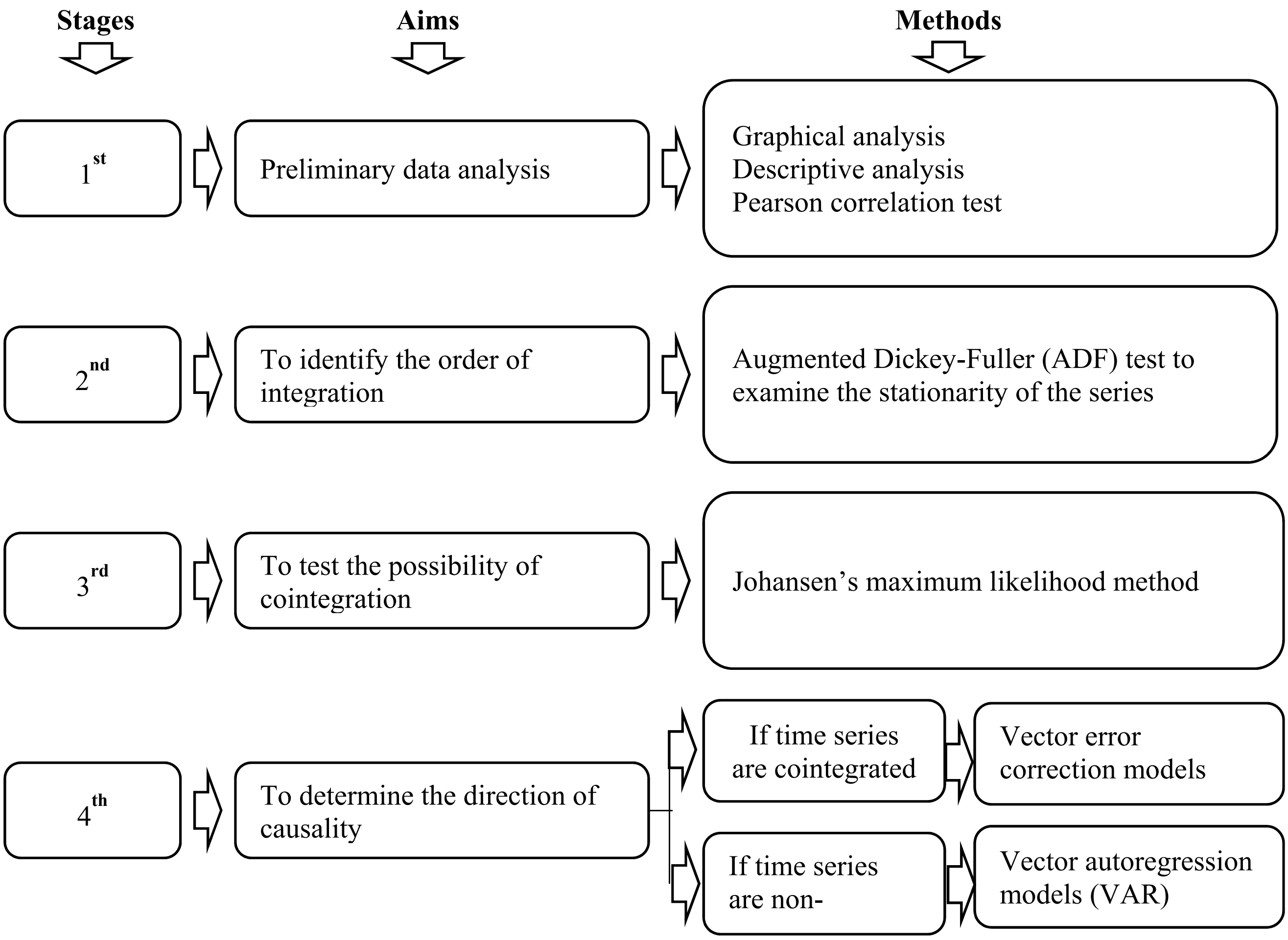

Figure 1: Methodological process used.
Source: Own elaboration.

First, an exploratory analysis, using graphical representations, descriptive analysis and Pearson correlation tests, was carried out (i) to provide a better understanding of the time series used, (ii) to identify behavioural differences among the tourism markets analysed, and (iii) to examine whether there are significant statistical associations between the level of interaction with the DMO website and the tourism demand of the region.

Subsequently, to estimate the causal relationships between website visitor interactions and the tourism demand, cointegration and Granger causality tests were carried out. Before these analyses, it was necessary to identify the order of integration. The stationarity proprieties of the data were checked through the Augmented Dickey-Fuller test (ADF Test). After, Johansen's maximum-likelihood method was used to test the possibility of cointegration between the variables used. Finally, to determine the direction of the causality, two different models were used. When the hypothesis of non-cointegration was not rejected, the VAR was used; and when the time series were cointegrated, the VECM was carried out. To undertake the econometric estimations, statistical software package Stata-Data Analysis and Statistical Software-12 was used (Baum, 2009; StataCorp., 2011).

4. RESULTS AND DISCUSSION

4.1) PRELIMINARY DATA ANALYSIS AND PEARSON CORRELATION

A descriptive analysis of the variables used in this study (Table 1) reveals differences among the tourism markets under analysis in terms of the number of guests in accommodation establishments, as well as concerning the number of website interactions. The domestic market has the highest mean values both in the number of guests and in the number of web interactions. On the other hand, the Italian market has the lowest averages both in terms of the number of guests and in terms of the number of web interactions. However, it is also interesting to note that, although the British market ranks seventh in terms of the average number of guests, it ranks sixth concerning the number of web interactions. Moreover, a large dispersion of the data is observed.

Tourism markets	Guests in the Alentejo region (thousands)				Number of interactions with the DMO website of Alentejo			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Domestic (Portugal)	50.1	19.8	23.5	128.1	9736	6130	179	31658
Spain	4.5	2.4	1.7	16.3	1210	886	1	4362
France	2.5	2.2	0.3	10.6	474	444	0	2336
Brazil	1.7	1.1	0.3	5.9	369	254	0	885
Germany	1.9	1.4	0.2	7.4	280	173	0	825
USA	1.7	1.1	0.2	5.9	190	167	0	835
United Kingdom (UK)	1.2	0.8	0.2	4.3	196	92	3	416
Netherlands	1.1	0.8	0.2	3.5	121	86	0	557
Italy	1.0	1.2	0.2	7.1	97	73	8	73

Table 1: Descriptive statistics of the original data series.
Source: Own elaboration.

A graphical representation of temporal evolution of the variables under analysis (Figure 2) highlights three important conclusions: (i) both variables reveal a seasonal pattern in all tourism markets; (ii) an association between the number of interactions and the number of guests is evidenced; and (iii) there are differences according to the tourism market, in terms of the evolution of the variables under analysis and also in terms of potential associations between the number of web interactions and the number of guests. For example, the number of website interactions in the domestic market registered a considerable increase until 2014, and, since then, it has been decreasing. A similar pattern can be seen in the French, British, German and Italian tourism markets. On the other hand, in the case of Brazilian and USA tourism markets, the number of website interactions was almost constant between 2007 and 2012, increasing in both markets after this period.

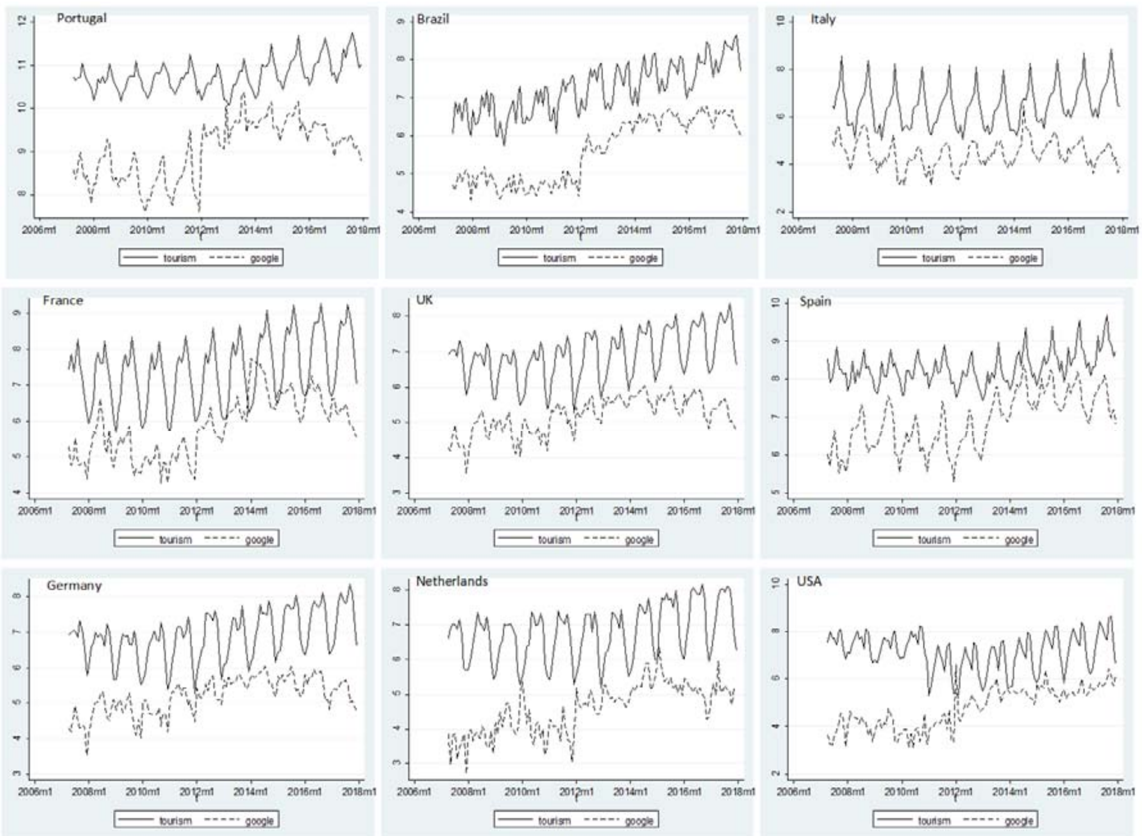

Figure 2: Google Analytics website data and the number of guests in tourism accommodation establishments, by country of origin of visitors.
Source: Own elaboration.

These results clearly show that tourism markets are heterogeneous concerning visitor interactions with the DMO website. The identification of these differences is of utmost relevance for the DMO to better understand the diverse tourism markets to design specific strategies for each of them. Trying to understand why the interaction with the website is decreasing is essential for the DMO to define strategies to boost visitor interactions in order to increase the tourism demand.

In order to observe whether there is a statistically significant association between the number of website interactions and the number of guests, the Pearson correlation test was used. Table 2 reveals that there are statistically significant positive associations between website interactions and the number of guests in all tourism markets under analysis. However, considerable differences are observed among tourism markets in terms of the magnitude of the association coefficients. The highest coefficients occur in the case of Brazil and Spain, while the lowest are detected in the case of the USA, Italy and France.

Tourism markets	Pearson correlation values
Domestic (Portugal)	0.401**
Spain	0.647**
France	0.298**
Brazil	0.714**
Germany	0.400**
USA	0.238**
UK	0.450**
Netherlands	0.344**
Italy	0.224**
Notes: ** Correlation is significant at the 0.01 level	

Table 2: Pearson correlation results between the number of interactions with the DMO website of Alentejo and the number of guests in the tourism accommodation establishments.
Source: Own elaboration.

4.2) COINTEGRATION AND CAUSALITY

4.2.1) THE ORDER OF INTEGRATION

Before testing cointegration and causality between the time series, the stationarity properties of the data were examined through unit root tests. In this analysis, the ADF test was performed. The was that the time series has a unit root against the alternative hypothesis that the time series is stationary (Dickey & Fuller, 1979). To support the selection of the correct specification of the unit root tests, Akaike's and Schwartz's Bayesian information criteria were used to determine the lag length. The results of the stationarity test show that the variables are non-stationary, but, when the stationarity test was applied to the first difference of the data series, the results reveal that the hypothesis of non-stationarity is rejected (Table 3). Therefore, the variables are integrated of order one (I[1]). Results suggest that the use of cointegration techniques is suitable to proceed with the long-run analysis.

	$lnNG^t$		$lnGV^t$	
Tourism Markets	**Level**	**1st Dif.**	**Level**	**1st Dif.**
Portugal *lags*	-2.719 (0.0708) [6]	-5.728 (0.0000)* [3]	-2.455 (0.1267) [2]	-6.966 (0.0000)* [4]
Spain *lags*	-2.047 (0.2662) [7]	-6.000 (0.0000)* [3]	-2.767 (0.0632) [5]	-6.714 (0.0000)* [2]
France *lags*	-2.252 (0.1878) [7]	-7.853 (0.0000)* [4]	-2.614 (0.0901) [1]	-9.399 (0.0000)* [1]
Brazil *lags*	-2.307 (0.1697) [3]	-8.223 (0.0000)* [2]	-1.206 (0.6712) [2]	-10.269 (0.0000)* [1]
Germany *lags*	-1.002 (0.7525) [9]	-7.059 (0.0000)* [3]	-2.182 (0.2130) [2]	-7.163 (0.0000)* [4]
USA *lags*	-2.752 (0.0654) [8]	-6.821 (0.0000)* [4]	-1.499 (0.5340) [3]	-7.973 (0.0000)* [4]
UK *lags*	-2.176 (0.2152) [8]	-7.739 (0.0000)* [2]	-2.756 (0.0649) [3]	-9.686 (0.0000)* [1]
Netherlands *lags*	-2.110 (0.2404) [8]	-5.695 (0.0000)* [3]	-2.131 (0.2322) [3]	-9.133 (0.0000)* [4]
Italy *lags*	-1.625 (0.4702) [9]	-7.358 (0.0000)* [1]	-2.451 (0.1279) [8]	-7.719 (0.0000)* [1]

Note: $lnNG_t$ is the proxy of tourism demand of a tourism destination; $lnGV_t$ is the proxy of website visitor interactions. Hypotheses - ADF test: Ho: Non-stationary series | Ha: Stationary series. ADF test: the values in brackets are MacKinnon approximate p-value for Z(t); * indicate rejection of a unit root hypothesis based on MacKinnon critical values at 5%.

Table 3. Results of the unit root test.
Source: Own elaboration.

4.2.2) TESTING FOR COINTEGRATION

Since the variables are integrated of the same order (order one), the next procedure was to test the possibility of cointegration among the variables selected. For this purpose, Johansen's maximum-likelihood method, which tests the number of cointegrating relationships and estimates their parameters, was used in this research. A comprehensive description of estimating cointegrating vectors and testing the hypothesis can be found in Johansen (1988, 1991, 1995) and Johansen and Juselius (1990). The results of this test are reported in Table 4.

Countries	r: number of cointegrating vectors	Lags	Trace Statistic	5% critical value
Portugal	r =1	[5]	2.3430*	3.76
Spain	r = 0	[7]	12.0686*	15.41
France	r = 1	[4]	2.3577*	3.76
Brazil	r = 1	[4]	1.0010*	3.76
Germany	r = 1	[3]	3.0288*	3.76
USA	r = 1	[3]	2.9105*	3.76
UK	r = 0	[8]	14.4624*	15.41
Netherlands	r = 1	[6]	2.9610*	3.76
Italy	r = 1	[7]	3.5555*	3.76

Note: * by the trace statistic indicates that this is the value of r selected by Johansen's multiple-trace test procedure. Akaike information criterion (AIC) is used to select the lag length (values in brackets).

Table 4. Johansen maximum likelihood cointegration tests.
Source: Own elaboration.

Results show differences among the tourism markets under analysis. Considering that the trace statistic at r = 0 is less than its critical value for Spain and UK, the null hypothesis of non-cointegration is not rejected. Conversely, because the trace statistic at r = 1 is less than its critical value for the Brazilian, French, German, Italian, Dutch, domestic (Portugal) and USA markets, the hypothesis that there is one or fewer cointegrating relations is not rejected. Therefore, except for the Spanish and the British markets, there is a presence of a stable long-run equilibrium linear relationship between the tourism demand and DMO website interactions, corroborating the results of other studies carried out in other tourism destinations (e.g., Sun et al., 2019; Yang et al., 2015). The existence of this stable long-run relationship between these two variables means that they are causally related, at least, in one direction. However, the number of studies that use cointegration tests to examine the relationship between the number of DMO website interactions and the tourism demand are very scarce in the literature. Moreover, to the best of our knowledge, there is no study examining this relationship in different tourism markets.

4.4) CAUSAL RELATIONSHIPS

Finally, the third and final step consists of analysing the causal relationship between the number of website visitor interactions and the number of guests in tourism accommodation establishments, as well as analysing whether there are differences in this relationship according to the tourism market under analysis. Are interactions with the DMO website triggering tourism demand for Alentejo, or is the tourism demand producing visitor interactions with the DMO website? To answer this question, two methods were used. On the one hand, when the hypothesis of non-cointegration is not rejected (as is the case of Spain and UK), the VAR was carried out (Gujarati, 2003). The model can be written as follows:

- Relationship between the tourism demand of Alentejo and visitor interactions with the official DMO website:

$$\Delta \ln \mathrm{NG}_t = \sum_{i=1}^{p} . \varphi_{1i} \Delta \ln \mathrm{NG}_{t-i} + \sum_{i=1}^{p} . \zeta_{1i} \Delta \ln \mathrm{GV}_{t-i} + u_{1t} \qquad (1)$$

$$\Delta \ln \mathrm{GV}_t = \sum_{i=1}^{p} . \varphi_{2i} \Delta \ln \mathrm{NG}_{t-i} + \sum_{i=1}^{p} . \zeta_{2i} \Delta \ln \mathrm{GV}_{t-i} + u_{2t} \qquad (2)$$

Where NGt is the proxy of tourism demand of a tourism destination; lnGVt is the proxy of website visitor interactions, both endogenous variables and (I(1)); p is the lag length; and u_{1t} and u_{2t} are the residuals.

After fitting the VAR, the short-run Granger causality test was performed. This test consists of the regression equation (1) on its own lagged values and on lagged values of the variable GVt, and tests the null hypothesis that the coefficients estimated on the lagged values of website visitor interactions (ζ1i) are jointly zero. If the null hypothesis is not rejected, this means that, in the short run, website visitor interactions do not Granger-cause tourism demand. A similar testing procedure was applied to equation (2) (Adkins & Hill, 2008; Gujarati, 2003; StataCorp., 2011).

Table 5 displays the results of the short-run Granger causality test for two traditional markets: Spain and the UK. The null hypothesis is rejected, so there is bidirectional short-run causality between the tourism demand of the destination and DMO website visitor interactions. Although the effects are not immediate, these results suggest that these behaviours are related and impact each other. It shows (i) the usefulness of big data analytics, namely web page visiting data, to promote the tourism demand of a destination; and (ii) the interest of these international visitors in consulting the official website to improve their knowledge about a destination.

	$\Delta lnNG_t$	$\Delta lnGV_t$
Countries	*Wald Test*	
Spain	$chi^2(7)$ = 81.261 (0.000)*	$chi^2(7)$ = 61.238 (0.000)*
UK	$chi^2(8)$ = 57.881 (0.000)*	$chi^2(8)$ = 66.868 (0.000)*

Note: $\Delta lnNG_t$ - VAR model for tourism demand of a tourism destination; $\Delta lnGV_t$ - VAR model for sessions of the users to the official Destination Management Organization (DMO) website: $\Delta lnNG_t$: Ho: $\zeta_{11} = \ldots = \zeta_{1p} = 0$ Users of the website do not Granger-cause tourism / Ha: $\zeta_{11} = \ldots = \zeta_{1p} \neq 0$ Users of the website Granger-cause tourism | $\Delta lnGV_t$: Ho: $\varphi_{21} = \ldots = \varphi_{2p} = 0$ Tourism does not Granger-cause use of the website / Ha: $\varphi_{21} = \ldots = \varphi_{2p} \neq 0$ Tourism Granger-causes users of the website. Values in brackets correspond Prob > chi^2; * indicates significance at the 5% level. Akaike information criterion (AIC) is used to select the lag length.

Table 5. Results of Granger causality test.
Source: Own elaboration.

On the other hand, since time series are cointegrated in the case of Brazil, France, Germany, Italy, the Netherlands, Portugal, and the USA, the VECM was performed (Gujarati, 2003). Overall, the VECM model for the relationship between tourism demand and website interactions can be written as follows:

$$\Delta \mathrm{lnNG}_t = \varphi_1 + \sum_{i=1}^{p} \alpha_{1i} \Delta \mathrm{lnNG}_{t-i} + \sum_{i=1}^{p} \beta_{1i} \Delta \mathrm{lnGV}_{t-i} + \gamma_1 ECM_{t-1} + u_{1t} \quad (3)$$

$$\Delta \mathrm{lnGV}_t = \varphi_2 + \sum_{i=1}^{p} \alpha_{2i} \Delta \mathrm{lnNG}_{t-i} + \sum_{i=1}^{p} \beta_{2i} \Delta \mathrm{lnGV}_{t-i} + \gamma_2 ECM_{t-1} + u_{2t} \quad (4)$$

Where NG_t is the proxy of tourism demand of a tourism destination; GV_t is the proxy of website visitor interactions, both endogenous variables and (I(1)); p is the lag length; u_{1t} and u_{2t} are the residuals; φ_1 and φ_2 are the constants; α_{1i}, β_{1i}, α_{2i}, β_{2i} are the parameters to be estimated; and γ_1 and γ_2 are the coefficients of error correction term (ECM).

The estimated coefficients (γ_1 and γ_2) of ECM measure the speed of adjustment to restore equilibrium in the dynamic model. Given the statistical significance of the ECM in a VECM, it suggests the existence of a long-run equilibrium relationship between the variables. The *t*-test for those coefficients provides the long-run Granger causality result. The coefficients of the lagged values (α_{1i}, β_{1i}, α_{2i}, β_{2i}) represent the short-run effects; thus, the test of joint significance of the lagged terms for each variable provides the short-run Granger causality. Since the coefficients of β_{1i} are jointly significant in equation (3), the null hypothesis that the users of the website do not Granger-cause tourism is rejected. A similar testing procedure is applied to equation (4) (Adkins & Hill, 2008; Gujarati, 2003; StataCorp., 2011).

Concerning long-run causality (Table 6), the value of the ECM has statistical significance in almost all models, thus suggesting the existence of a long-run Granger causality between visitor interactions with the official DMO website and tourism demand, revealing the relevance of the use of these data in modelling tourism demand. In other words, when the coefficient of the ECM is significant in the equation $\Delta lnNG_t$, it means that website visitor interactions Granger-cause tourism demand in the long run. The same analysis applies to equation $\Delta lnNV_{t}$.. A significant ECM coefficient implies that past equilibrium errors play a role in determining current outcomes. The size of the coefficient of the error correction term indicates the speed of the adjustment towards equilibrium. The negative value shows that whenever the variable moves from equilibrium, the overall effect is to force it to converge back to the long-run equilibrium (Gujarati, 2003).

Countries	Equation	Short-run coefficients	ECM_{t-1}
Portugal	$\Delta lnNG_t$	$chi^2(4) = 20.61$ Prob > chi2 = (0.0004)*	$\gamma_1 = -0.3050444$ z = (-4.36)*
	$\Delta lnGV_t$	$chi^2(4) = 5.54$ Prob > chi2 = (0.2366)	$\gamma_2 = -0.4552475$ z = (-3.65)*
France	$\Delta lnNG_t$	$chi^2(3) = 17.54$ Prob > chi2 = (0.0005)*	$\gamma_1 = -0.3764542$ z = (-6.05)*
	$\Delta lnGV_t$	$chi^2(3) = 15.78$ Prob > chi2 = (0.0013)*	$\gamma_2 = -0.3403097$ z = (0.000)*
Brazil	$\Delta lnNG_t$	$chi^2(3) = 14.88$ Prob > chi2 = (0.0019)*	$\gamma_1 = -0.4216696$ z = (-3.88)*
	$\Delta lnGV_t$	$chi^2(3) = 13.79$ Prob > chi2 = (0.0032)*	$\gamma_2 = -0.0820941$ z = (-1.20)
Germany	$\Delta lnNG_t$	$chi^2(2) = 23.61$ Prob > chi2 = (0.0000)*	$\gamma_1 = -0.3280071$ z = (-5.03)*
	$\Delta lnGV_t$	$chi^2(2) = 4.23$ Prob > chi2 = (0.1207)	$\gamma_2 = -0.2027381$ z = (-3.81)*
USA	$\Delta lnNG_t$	$chi^2(2) = 4.80$ Prob > chi2 = (0.0909)**	$\gamma_1 = -0.3775008$ z = (-5.13)*
	$\Delta lnNV_t$	$chi^2(2) = 0.64$ Prob > chi2 = (0.7246)	$\gamma_2 = -0.0954849$ z = (-1.47)
Netherlands	$\Delta lnNG_t$	$chi^2(5) = 7.12$ Prob > chi2 = (0.2118)	$\gamma_1 = -0.5177547$ z = (-6.01)*
	$\Delta lnGV_t$	$chi^2(5) = 12.50$ Prob > chi2 = (0.0285)*	$\gamma_2 = -0.2859053$ z = (-3.06)*
Italy	$\Delta lnNG_t$	$chi^2(6) = 94.70$ Prob > chi2 = (0.0000)*	$\gamma_1 = -0.0765858$ z = (-0.93)
	$\Delta lnGV_t$	$chi^2(6) = 18.12$ Prob > chi2 = (0.0059)**	$\gamma_2 = -0.2885604$ z = (-3.97)*

Note: $\Delta lnNG_t$ - VEC model for tourism demand of a tourism destination; $\Delta lnGV_t$ - VEC model for sessions of the users to the official Destination Management Organization (DMO) website. Hypotheses for short-run Granger causality: $\Delta lnNG_t$: Ho: $\beta_{11} - ... - \beta_{1p} = 0$ Users of the website do not Granger-cause tourism / Ha: $\beta_{11} - ... - \beta_{1p} \neq 0$ Users of the website Granger-cause tourism | $\Delta lnGV_t$: Ho: $\alpha_{21} - ... - \alpha_{2p} = 0$ Tourism does not Granger-cause users of the website / Ha: $\alpha_{21} - ... - \alpha_{2p} \neq 0$ Tourism Granger-causes users of the website. γ_1 and γ_2 estimated coefficients of error correction terms (ECM_{t-1}); z = z test; Akaike information criterion (AIC) is used to select the lag length; * and ** indicate significance at the 5% and 10% levels, respectively.

Table 6. Results of the VECM estimation.
Source: Own elaboration.

The results for the short-run causality test reveal that there is a potential relationship between tourism demand and visits to official DMO website, but the effects are time-lagged, and there are differences among the tourism markets analysed. As presented in Table 7, bidirectional short-run causality between tourism demand and website visitor interactions is observed for five markets: Spain, France, Brazil, France, the UK and Italy. These results show that interaction with the DMO website influences tourism demand, and when the visitors of these tourism markets travel to Alentejo they will then interact with the official website. For the case of Germany, Portugal, and the USA, there is a unidirectional relationship between website interactions and tourism demand, providing strong evidence in support of the hypothesis that website visitor interactions drive tourism demand. Conversely, for the Dutch market, it is tourism demand that promotes visits to the official Alentejo website.

These differences reveal heterogeneity in the behaviour of the most important tourism markets of Alentejo, as well as in the relationship between website interactions and tourism demand. The knowledge of these differences, it is of utmost relevance to the DMO of Alentejo to design specific strategies for each market. The absence of literature in this field makes the discussion of these results difficult. However, based on the knowledge of these tourism markets and the historical, geographical and cultural ties between them and the region under analysis, some reasons may be explored. Intense historical or commercial relationships, due to colonial ties or geographical proximity, are likely to justify the strong bidirectional short-run causality in the case of Brazil, Spain and the UK. In the case of France, these might be explained by the importance of this market to inbound tourism and due to the considerable number of Portuguese emigrants in France.

Concerning the domestic tourism market, interactions with the official DMO website not only improve knowledge about the tourism supply of the Alentejo region but also stimulate the interest of the Portuguese in visiting this region. The small size of the country and the transportation infrastructure networks facilitate mobility.

In the case of the German and the USA tourism markets, the official website could play a significant role in increasing the presence of these two tourism markets in the region. Specifically, in the case of the USA market, the physical and cultural distance may justify the relevance of visitors collecting information through the official website of the region prior to carrying out a tourism trip. Regarding the Dutch market, visiting the Alentejo region triggers the search for more information about the destination. DMO official websites are usually an accessible and reliable tool for consultation or virtual interaction with the destination. Finally, in the case of the Italian tourism market, although it is still of little relevance to the Alentejo region and the number of website interactions have decreased in recent years, the bidirectional long-run relationship suggests that the DMO should make more effort to attract this market and the official website can be a powerful tool to achieve this goal.

Tourism Markets	$\Delta lnGV_t$	Direction	$\Delta lnNG_t$
Domestic (Portugal)		→	
Spain		↔	
France		↔	
Brazil		↔	
Germany		→	
USA		→	
UK		↔	
Netherlands		←	
Italy		↔	

Table 7. Granger causality tests (summary).
Source: Own elaboration.

The results obtained in this research confirm the existence of interconnection between tourism demand and visitor interactions with the DMO website for a significant number of tourism markets (Spain, France, Brazil, the UK and Italy). This relationship suggests not only that visiting the destination is preceded by interactions with the official website (to decide where to go or what to visit at the destination), but that when the visit occurs, the number of interactions with the website also increase following the visit.

5. CONCLUSIONS AND IMPLICATIONS

The Internet is used by consumers in tourism at all stages of travel. The official websites for promotion and dissemination of tourism destinations are considered an important source of information for potential consumers. Tracking the movements of users on these websites can reveal consumers' interests and trends in tourism. DMOs need to make decisions based on up-to-date and relevant information and should, therefore, consider official statistical data, but also a wide range of data that can be obtained online. In recent years, Internet data has been increasingly used in tourism studies; however, very few have used data related to visitor interactions with DMO websites.

This study aimed to ascertain whether there is a relationship between DMO website data and tourism demand in the respective tourism destination, and if this changes according to the country of origin of the tourists. For this purpose, an approach which has very seldom been employed in previous studies was used, consisting of simultaneously using Granger causality and cointegration tests and website data to model the tourism demand of a destination. The methodological procedures used in this study may also be of interest for future studies.

Results indicate that there is a causal relationship between website data and tourism demand, which varies according to the tourism market under analysis. The differences may be justified not only by the characteristics of the market in terms of the use of web platforms in the tourism planning process, but also by the importance of these markets for the destination. Historical, commercial and cultural links can also justify these differences, as well as geographical proximity. The adjustment of the website to the needs of the different markets may also justify the observed differences. These results may support private and public entities in implementing coordinated strategies that are designed to maximise tourism-website interdependence.

Big data proved to be of highly relevant in analysing the tourism demand of a destination, corroborating previous studies that have shown that online data is a useful source to forecast tourism demand (Höpken, Eberle, Fuchs & Lexhagen, 2018; Park, 2017; Sun et al., 2019; Yang et al., 2015). This study demonstrates the value of using Google Analytics data to model the tourism demand of a destination through econometric models. Therefore, it provides relevant insights for DMOs related to the decision-making process in tourism. It allows improved understanding of the behaviour and intentions of tourists from the main tourism markets for the region, thus allowing the DMO to develop strategies and adjust marketing and promotional actions to take advantage of each market, and to better plan and manage the development of tourism. Therefore, empirical research using data disaggregated by tourism market is of the utmost relevance. In addition, results suggest that DMOs can use website search data by country to improve the performance of forecasting models, thus anticipating the tourism demand.

One limitation of this study lies in the choice of a single destination, although Alentejo represents a good case study. In further research, this type of data and methodology should be applied to other tourism destinations to analyse the relationship between website interactions and tourism demand. The results are also sensitive to the sample period, cointegration and causality approaches used, and variables that are omitted in the model.

Compared to traditional large-scale surveys, one of the advantages of applying web search data in tourism demand analysis lies in its timeliness, as well as low cost. However, big data studies should employ and triangulate a variety of sources and formats of data (Sun et al., 2019), with the use of multiple types of data being recommended to capture the characteristics of the complex nature of tourism. Despite being a valuable and timely complement to traditional statistics, further research is needed to better understand how website search data translates into actual consumption. Nonetheless, the findings of the study can contribute to broadening the knowledge of important aspects for the planning, management and monitoring of tourism by both public and private organisations, since most official statistics are released with a lag, making it difficult for policymakers to make an accurate assessment of current conditions. Obtaining information in real-time is of great use for decision making, reducing the risk of associated uncertainty. Data gathered through DMO websites also offer the possibility of directing marketing campaigns to a specific segment or time of the year. In addition, the level of geographical breakdown it allows is difficult to obtain in official statistics.

Acknowledgements

The authors would like to thank *Agência Regional de Promoção Turística do Alentejo* and *Turismo do Alentejo* for providing access to Google Analytics data of the website “www.visitalentejo.pt”, which was essential to obtain the data for conducting this research.

This work was financially support by the research unit on Governance, Competitiveness and Public Policy (UIDB/04058/2020), funded by national funds through FCT – Fundação para a Ciência e a Tecnologia.

References

Adkins, L.C.; Hill, R.C. *Using Stata for principles of econometrics*. New York: John Wiley & Sons, 2008.

Antolini, F.; Grassini, L. Foreign arrivals nowcasting in Italy with Google Trends data. *Quality & Quantity,* No 53, 2019, pp. 2385-2401.

Artola, C.; Pinto, F.; de Pedraza García, P. Can internet searches forecast tourism inflows? *International Journal of Manpower,* Vol. 36, No 1, 2015, pp. 103-116.

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

Baum, C.F. *An introduction to Stata programming*. College Station, Texas: Stata Press. 2009.

Bello-Orgaz, G.; Jung, J.J.; Camacho, D. Social big data: Recent achievements and new challenges. *Information Fusion*, No 28, 2016, pp. 45-59.

Buhalis, D. *eTourism: Information technology for strategic tourism management*. London: Pearson (Financial Times/Prentice Hall), 2003.

Buhalis, D.; Law, R. Progress in information technology and tourism management: 20 years on and 10 years after the Internet – The state of eTourism research. *Tourism Management,* Vol. 29, No 4, 2008, pp. 609-623.

Buhalis, D.; O'Connor, P. Information communication technology: Revolutionising tourism, In: Buhalis, D.; Costa, C. (Eds). *Tourism management dynamics: Trends, management and tools*. London: Elsevier, 2006, pp. 196-209.

Centobelli, P.; Ndou, V. Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism,* Vol. 22, No 15, 2019, pp.1862-1882.

Choi, H.; Varian, H. Predicting the present with Google Trends. *Economic Record, 88*(SUPPL.1), 2012, pp. 2-9.

Clark, D.; Nicholas, D.; Jamili, H.R. Evaluating information seeking and use in the changing virtual world: the emerging role of Google Analytics. *Learned Publishing*, Vol. 27, No 3, 2014, pp.185-194.

Clark, M.; Wilkins, E.J.; Dagan, D.T.; Powell, R.; Sharp, R.L.; Hillis, V. Bringing forecasting into the future: Using Google to predict visitation in US national parks. *Journal of Environmental Management,* Vol. 243, 2019, pp. 88-94.

Cox M.; Ellsworth D. Managing big data for scientific visualization. *ACM Siggraph, MRJ/NASA Ames Research Center,* No 5, 1997, pp. 1-17.

Crouch, G. Services research in destination marketing: A retrospective and prospective appraisal. *International Journal of Hospitality & Tourism Administration*, Vol.1, No 2, 2000, pp. 65-86.

Crutzen, R.; Roosjen, J.L.; Poelman, J. Using Google Analytics as a process evaluation method for Internet-delivered interventions: An example on sexual health. *Health Promotion International*, Vol. 28, No 1, 2013, p. 36.

Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Vol. 74, No 366, 1979, pp. 427-431.

Dinis. G.; Breda, Z.; Barreiro, T. Digital marketing strategies of destination management organizations: An exploratory study, In: Carvalho, L.C.; Calisto, L.; Gustavo, N. (Eds.). *Strategic business models to support demand, supply, and destination management in the tourism and hospitality industry*. Hershey: IGI Global, 2020, pp. 266-285.

Dinis, G.; Breda, Z.; Costa, C.; Pacheco, O. Google Trends in tourism and hospitality research: A systematic literature review. *Journal of Hospitality and Tourism Technology,* Vol.10, No 4, 2019, pp. 747-763.

Dinis, G.; Costa, C.; Pacheco, O. Google Analytics as a tool for understanding visitors behaviour: The website of Alentejo as a tourist destination case. *Journal of Tourism and Development,* No 17/18, 2012, pp. 321-331.

Dinis, G.; Costa, C.; Pacheco, O. Profile of the national Alentejo online visitor: Analysis of the official tourism website using google analytics. *Tourism and Hospitality International Journal*, Vol. 6, No 1, 2016, pp. 23-34.

Dinis, G.; Costa, C.; Pacheco, O. Aplicação da análise da web no turismo: O caso do Turismo do Alentejo. *International Journal of Marketing, Communication and New Media,* Special No 2 – Marketing and Digital Business, 2017, pp. 67-87.

DMA West. *The impact of DMO websites: DMO website user & conversion study*. (n.d.). Retrieved from https://bit.ly/3z7v3jL [accessed 3 May 2020].

Ettredge, M., Gerdes, J.; Karuga, G. *Using web-based search data to predict macroeconomic statistics*. *Communications of the ACM,* Vol. 48, No 11, 2005, pp. 87-92.

European Union. *Flash Eurobarometer 432: Preferences of Europeans towards tourism*. 2016. Retrieved from https://bit.ly/2CT42nG [accessed 2 September 2019].

Eurostat. *Individuals: Internet use*. 2019a. Retrieved from https://bit.ly/32Th7Ii [accessed 10 October 2019].

Eurostat. *Individuals: Internet activities*. 2019b. Retrieved from https://bit.ly/33Tv0Yb [accessed 10 October 2019].

Fang, W. Using Google Analytics for improving library website content and design: A case study. *Library Philosophy & Practice*, Paper 121. 2007. Retrieved from http://bit.ly/1tKuQf9 [accessed 30 October 2011].

Frechtling, D.C. *Forecasting tourism demand: Methods and strategies*. London: Routledge Taylor & Francis Group, 2001.

Fuchs, M., Höpken, W.; Lexhagen, M. Big data analytics for knowledge generation in tourism destinations: A case from Sweden. *Journal of Destination Marketing and Management,* Vol. 3, No 4, 2014, pp. 198-209.

Gantz, L.; Reinsel, D. Extracting value from chaos. *IDC iView*, 2011, pp. 1-12.

Garcia, A., Linaza, M.T., Gutierrez, A.; Garcia, E. Gamified mobile experiences: Smart technologies for tourism destinations. *Tourism Review*, Vol. 74, No 1, 2018, pp. 30-49.

Google. *How a web session is defined in Analytics*. 2019. Retrieved from https://bit.ly/2D3x701 [accessed 1 October 2019].

Gujarati, D. *Basic econometrics*. Boston, MA: McGraw-Hill, 2003.

Gunter, U.; Önder, I. Forecasting city arrivals with Google Analytics. *Annals of Tourism Research,* Vol. 61, 2016, pp. 199-212.

Hamill, J., Stevenson, A.; Attard, D. National DMOs and Web 2.0, In: Sigala, M.; Christou, E.; Gretzel, U. (Eds.). *Social media in travel, tourism and hospitality: Theory, practice and cases*. Farnham: Ashgate. 2012, pp. 99-120.

Hanke, J.; Wichern, D. *Business forecasting*. New Jersey: Pearson Prentice Hall. 2005.

Hasan, L., Morris, A.; Probets, S. Using Google Analytics to evaluate the usability of e-commerce sites, In: Kurosu, M. (Ed.). *Human centered design*. Cham: Springer Berlin Heidelberg, 2009, pp. 697-706.

Höpken, W.; Eberle, T.; Fuchs, M.; Lexhagen, M. Search engine traffic as input for predicting tourist arrivals, In: Stangl, B.; Pesonen, J. (Eds). *Information and communication technologies in tourism*. Cham: Springer, 2018, pp. 381-393.

Höpken, W., Fuchs, M., Keil, D.; Lexhagen, M. The knowledge destination: A customer information-based destination management information system, In: Law, R.; Fuchs, M.; Ricci, F. (Eds.). *Information and communication technologies in tourism 2011*. Wien: Springer/Verlag, 2015, pp. 417-429.

INE. *Retrato territorial de Portugal*. Lisboa: Instituto Nacional de Estatística. 2017.

INE. *Estatísticas do turismo 2017*. Lisboa: Instituto Nacional de Estatística. 2018.

Johansen, S. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, Vol. 12, No 2/3, 1988, pp. 231-254.

Johansen, S. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, Vol. 59, No 6, 1991, pp. 1551-1580.

Johansen, S. *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press. 1995.

Johansen, S.; Juselius, K. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, Vol. 52, No 2, 1990, pp. 169-210.

Kim, H.; Fesenmaier, D.R. Persuasive design of destination web sites: An analysis of first impression. *Journal of Travel Research*, Vol. 47, No 1, 2008, pp. 3-13.

Kim, W.H.; Malek, K. Forecasting casino revenue by incorporating Google Trends. *International Journal of Tourism Research,* Vol. 20, No 4, 2018, pp. 424-432.

Kudyba, S.; Kwatinetz, M. Introduction to the big data era, In: Kudyba, S. (Ed.). *Big data, mining, and analytics*. Boca Raton: CRC Press and Taylor and Francis. 2014, pp. 1-15.

Kuo M.-S.; Chuang T.-Y. How gamification motivates visits and engagement for online academic dissemination: An empirical study. *Computers in Human Behavior,* Vol. 55, No 3652, 2016, pp. 16-27.

Laney, D. 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, No 6, 2001, p. 70.

Li, H.; Hu, M.; Li, G. Forecasting tourism demand with multisource big data. *Annals of Tourism Research,* Vol. 83, 2020, p. 102912.

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

Mariani, M. Big data and analytics in tourism and hospitality: A perspective article. *Tourism Review,* Vol. 75, No 1, 2019, pp. 299-303.

McCloskey, E.V.; Johansson, H.; Harvey, N.C.; Compston, J. Access to fracture risk assessment by FRAX and linked National Osteoporosis Guideline Group (NOGG) guidance in the UK: An analysis of anonymous website activity. *Osteoporosis International*, Vol. 28, No 1, 2017, pp. 71-76.

Milheiro, E.; Dinis, G. Caracterização da procura turística – Alentejo. Observatório Regional do Turismo do Alentejo. 2013.

Moral, P.; Gonzalez, P.; Plaza, B. Methodologies for monitoring website performance: Assessing the effectiveness of AdWords campaigns on a tourist SME website. *Online Information Review*, Vol. 38, No 4, 2014, pp. 575-588.

Nicolau, J. Travel demand modeling with behavioral data, In: Xiang, Z.; Fesenmaier, D. (Eds.). *Analytics in smart tourism design concepts and methods*. Cham: Springer International Publishing, 2017, pp. 31-42.

Pakkalaa, H.; Presser, K.; Christensen, T. Using Google Analytics to measure visitor statistics: The case of food composition Websites. *International Journal of Information Management*, Vol. 32, No 6, 2012, pp. 504-512.

Pan, B.; Yang, Y. Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research,* Vol. 56, No 7, 2017, pp. 957-970.

Park, D.-H. The development of travel demand nowcasting model based on travelers' attention: Focusing on web search traffic information. *The Journal of Information Systems,* Vol. 26, No 3, 2017, pp. 171-185.

Parra-López, E.; Gutiérrez-Taño, D.; Díaz-Armas, R.; Bulchand-Gidumal, J. Travellers 2.0: Motivation, opportunity and ability to use social media, In: Sigala, M.; Christou, E.; Gretzel, U. (Eds.). *Social media in travel, tourism and hospitality: Theory, practice and cases*. UK: Ashgate, 2012, pp. 171-187.

Phippen, A.D., Sheppard, L.; Furnell, S. A practical evaluation of Web analytics. *Internet Research,* Vol. 14, No 4, 2004, pp. 284-293.

Plaza, B. Monitoring Web traffic source effectiveness with Google Analytics. An experiment with time series. *Aslib Proceedings*, Vol. 61, No 5, 2009a, pp. 474-482.

Plaza, B. Using Google Analytics for measuring inlinks and effectiveness, *Munich Personal RePEc Archive* (MPRA Paper No. 19676). 2009b.

Plaza, B. Google Analytics intelligence for information professionals. *Online*, Vol. 34, No 5, 2010, pp. 33-37.

Plaza, B. Google Analytics for measuring Website performance, *Tourism Management*, Vol. 32, No 3, 2011, pp. 477-481.

Plaza, B.; Gonzalez Pilar, C.; Moral Zuazo, M.P.; Ostolaza, I. Validating Google analytics tips for micro-firms. *African Journal of Business Management*, Vol. 5, No 14, 2011, pp. 5681-5689.

Rivera, R. A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management,* No 57, 2016, pp. 12-20.

Sigala, M.; Rahimi, R.; Thewall, M. (Eds.). *Big data and innovation in tourism, travel, and hospitality*. Singapore: Springer, 2019.

Siliverstovs, B.; Wochner, D.S. Google Trends and reality: Do the proportions match? Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. *Journal of Economic Behavior and Organization, No 145*, 2018, pp. 1-23.

So, S.; Morrisson, A. Destination marketing organisations' web site users and nonusers: A comparison of actual visits and revisit intentions. *Information Technology & Tourism,* Vol. 6, No 2, 2003, pp. 129-139.

Song, H.; Liu, H. Predicting tourist demand using big data; In: Z. Xiang; D. Fesenmaier (Eds.). *Analytics in smart tourism design concepts and methods*. Cham: Springer International Publishing, 2017, pp. 13-30.

Song, H.; Witt, S.F. *Tourism demand modelling and forecasting: Modern econometric approaches*. New York: Pergamon. 2000.

Song, M.; Ward, J.; Choi, F.; Nikoo, M.; Frank, A.; Shams, F.; Tabi, K.; Vigo, D.; Krausz, M. A process evaluation of a web-based mental health portal (WalkAlong) using Google Analytics. *JMIR Mental Health,* Vol. 5, No 3, 2018, p. e50.

StataCorp. *Stata: Release 12. Statistical software*. College Station, TX: StataCorp LP. 2011.

Steinbauer, A.; Werthner, H. Consumer behaviour in e-tourism. In: Sigala M.; Mich L.; Murphy J. (Eds.), *Information and Communication Technologies in Tourism 2007*. Vienna: Springer, 2007, pp. 65-76.

Sun, S.; Wei, Y.; Tsui, K.L.; Wang, S. Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management,* Vol. 70, 2019, pp. 1-10.

Teixeira, J. *Your google game plan for success: Increasing your web presence with Google AdWords, Analytics and Website Optimizer*. Indianapolis: Wiley Publishing, Inc. 2010.

Tell, J.; Anderberg, P.; Olander, E.; Berglund, J. The usage of web-based national guidelines for child healthcare: A web analytic study. 2018. Retrieved from https://bit.ly/2qnuymr [accessed 5 September 2019].

Tonkin, S.; Whitmore, C.; Cutroni, J. *Performance marketing with Google Analytics: Strategies and techniques for maximizing online ROI*. Indianapolis: John Wiley Publishing, Inc. 2020.

Turner, S.J. Website statistics 2.0: Using 2010 Google Analytics to measure library website effectiveness. *Technical Services Quarterly*, Vol. 27, No 3, 2010, pp. 261-278.

UNWTO. *Evaluating and improving websites: The destination web watch*. Madrid: World Tourism Organization. 2005.

UNWTO; ETC. *Handbook on e-marketing for tourism destinations: Fully revised and extended version 3.0*. Madrid: World Tourism Organization. 2014.

Vélez, J.; Pagán, L. *Usage of statistics analysis specialised libraries websites*. Paper presented at the World Library and Information Congress: 77th IFLA General Conference and Assembly, San Juan, Puerto Rico, 13-18 August. 2011.

Vogel, T.K.; Kleib, M.; Davidson, S.J.; Scott, S.D. Parental evaluation of a nurse practitioner-developed pediatric neurosurgery website. *JMIR Research Protocols,* Vol. 5, No 2, 2016, p. e55.

Volo, S. Tourism statistics, indicators and big data: A perspective article. *Tourism Review,* Vol. 75, No 1, 2019, pp. 304-309.

Vona, P.; Wilmoth, P.; Jaycox, L.H.; McMillen, J.S. A web-based platform to support an evidence-based mental health intervention: Lessons from the CBITS web site. *Psychiatric Services,* Vol. 65, No 11, 2014, pp. 1381-1384.

W3Techs. Usage of traffic analysis tools for websites. 2019. Retrieved from https://bit.ly/37d6xPN [accessed 3 October 2019].

Xiang, Z.; Wang, D.; O'Leary, J.; Fesenmaier, D. Adapting to the Internet: Trends in travelers' use of the web for trip planning. *Journal of Travel Research*, Vol. 54, No 4, 2015, pp. 511-527.

Yang, X.; Pan, B.; Evans, J.A.; Lv, B. Forecasting Chinese tourist volume with search engine data. *Tourism Management,* Vol. 46, 2015, pp. 386-397.

Yang, Y.; Pan, B.; Song, H. Predicting hotel demand using destination marketing organisations' web traffic data. *Journal of Travel Research*, Vol. 53, No 4, 2014, pp. 433-447.

Article info: Received 31/10/2021. Accepted 23/03/2022. Refereed anonymously.