

# Big data in tourism marketing: past research and future opportunities

Big data in  
tourism  
marketing

Sofía Blanco-Moreno and Ana M. González-Fernández  
*University of León, León, Spain, and*  
Pablo Antonio Muñoz-Gallego  
*University of Salamanca, Salamanca, Spain*

Received 6 June 2022  
Accepted 22 November 2022

## Abstract

**Purpose** – The purpose of this study was to uncover representative emergent areas and to examine the research area of marketing, tourism and big data (BD) to assess how these thematic areas have developed over a 27-year time period from 1996 to 2022. This study analyzed 1,152 studies to identify the principal thematic areas and emergent topics, principal theories used, predominant forms of analysis and the most productive authors in terms of research.

**Design/methodology/approach** – The articles for this research were all selected from the Web of Science database. A systematic and quantitative literature review was performed. This study used SciMAT software to extract indicators. Specifically, this study analyzed productivity and produced a science map.

**Findings** – The findings suggest that interest in this area has increased gradually. The outputs also reveal the innovative effort of industry in new technologies for developing models for tourism marketing. Ten research areas were identified: “destination marketing,” “mobility patterns,” “co-creation,” “gastronomy,” “sustainability,” “tourist behavior,” “market segmentation,” “artificial neural networks,” “pricing” and “tourist satisfaction.”

**Originality/value** – This work is unique in proposing an agenda for future research into tourism marketing research with new technologies such as BD and artificial intelligence techniques. In addition, the results presented here fill the current gap in the research since while there have been literature reviews covering tourism with BD or marketing, these areas have not been studied as a whole.

**Keywords** Big data, Tourism marketing, Literature review, SciMAT, Science mapping analysis, Future research agenda

**Paper type** Research paper

## Big Data en marketing turístico. Investigaciones pasadas y oportunidades futuras

### Resumen

**Propósito** – El objetivo de esta investigación fue descubrir nichos representativos de áreas emergentes y examinar el área de Marketing, Turismo y Big Data, evaluando cómo han evolucionado estas áreas temáticas durante un período de 27 años desde 1996–2022. Analizamos 1.152 investigaciones para identificar las

© Sofía Blanco-Moreno, Ana M. González-Fernández and Pablo Antonio Muñoz-Gallego. Published in *Spanish Journal of Marketing – ESIC*. 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>

This research was funded by Ministerio de Industria, Comercio y Turismo (Spain), AEI-010500–2020-253 (DTI^A Project: 4.0 technological tools for measurement, evaluation and monitoring of the Friendliness concept linked to the Smart Tourist Destinations)

*Declaration of interest:* None



principales áreas temáticas y temas emergentes, las principales teorías utilizadas, las formas de análisis predominantes y los autores más productivos en términos de investigación.

**Metodología** – Todos los artículos para esta investigación fueron seleccionados de la base de datos Web of Science. Realizamos una revisión sistemática y cuantitativa de la literatura. Utilizamos el software SciMAT para extraer indicadores. Específicamente, analizamos la productividad y elaboramos un mapeo científico.

**Hallazgos** – Los hallazgos sugieren que el interés en esta área ha aumentado gradualmente. Los resultados también revelan el esfuerzo innovador de la industria en nuevas tecnologías para desarrollar modelos de marketing turístico. Se identificaron diez áreas de investigación (“marketing de destinos”, “patrones de movilidad”, “co-creación”, “gastronomía”, “sostenibilidad”, “comportamiento turístico”, “segmentación de mercado”, “redes neuronales artificiales”, “precios”, y “satisfacción del turista”).

**Valor** – Este trabajo es único al proponer una agenda para futuras investigaciones en investigación de Marketing Turístico con nuevas tecnologías como Big Data y técnicas de Inteligencia Artificial. Además, los resultados presentados aquí llenan el vacío actual en la investigación ya que si bien se han realizado revisiones de literatura que cubren Turismo con Big Data o Marketing, estas áreas no se han estudiado como un conjunto.

**Palabras clave** Palabras Big data, Marketing turístico, Revisión de la literatura, SciMAT,

Análisis de mapeo científico, Agenda de investigación futura

**Tipo de artículo** Trabajo de investigación

大数据旅游营销: 过去的研究和未来的机会

## 摘要

**目的** – 这一特定研究领域的目标是发现具有代表性的新兴领域, 并考察市场营销、旅游和大数据研究领域, 以评估这些主题领域在1996年至2022年的27年间是如何发展的。我们分析了1152项研究, 以确定主要专题领域和新兴主题、使用的主要理论、主要的分析形式以及在研究方面最有成效的作者。

**方法** – 本研究的文章都是从Web of Science数据库中选出的。我们进行了系统化的定量文献审查, 并使用SciMAT软件来提取指标。具体来说, 我们分析了生产力并制作了一个科学研究地图。

**研究结果** – 研究结果表明, 人们对这一领域的兴趣已经逐渐增加。本文也揭示了工业界在开发旅游营销模式的新技术方面的创新努力。研究确定了十个研究领域: “目的地营销”、“流动模式”、“共同创造”、“美食”、“可持续性”、“游客行为”、“市场细分”、“人工智能”、“定价”和“游客满意度”。

**原创性** – 这项研究的独特之处在于提出了未来利用大数据和人工智能技术等新技术进行旅游营销研究的议程。此外, 本文的结果填补了目前的研究空白, 因为虽然有文献综述涉及旅游与大数据或市场营销, 但这些领域还没有被作为一个整体来研究。

**关键词** 关键词 大数据, 旅游营销, 文献综述, SciMAT, 科学映射分析, 未来研究议程

**文章类型** 研究型论文

## 1. Introduction

The field of tourism research is one of the most long-established areas, with more than 175,000 publications listed on the Web of Science (WoS) from 1940 until 2022 (Kontogianni and Alepis, 2020).

Unfortunately, researchers have not always been in possession of sufficiently advanced tools and techniques to process all this information. However, thanks to big data (BD), grounded in facilities for the massive storage of quality structured data, this issue is starting to be resolved.

BD and its tools have changed the ways in which we can analyze and process information. However, there is currently no literature giving a thorough overview of how BD techniques have been used in tourism marketing over the past 27 years of its existence.

In the past decade, several authors have undertaken bibliometric analyses of tourism research literature. Work has concentrated on three key areas in isolation: tourism (Hall, 2011; Köseoglu *et al.*, 2015, 2016), BD in tourism (Li *et al.*, 2018; Mariani and Baggio, 2021; Samara *et al.*, 2020) and tourism experience (Kim and So, 2022). To our knowledge, no bibliometric analyses exist dealing with BD, tourism and marketing. Such a study has great value enabling researchers to gain an understanding of how key areas of study have evolved over time.

Compared to existing literature reviews on the topic “BD and tourism,” our work is distinctive in three ways. First, while the two previous literature reviews have focused only on BD and tourism, this study performs queries related explicitly to BD, tourism and marketing. We feel that the inclusion of marketing is essential as there is currently a lack of research into the practical applications of BD in tourism product design and marketing.

Second, while previous work has reviewed articles published between 2007 and 2020, we have extended the time span of interest to include all articles published, from 1996 to 2022. In this way, we cover not only the inception of this field but also its most recent evolution, including the two-year period of the COVID-19 crisis.

Third, unlike the present study, none of the previous review articles mentioned application of the bibliometric techniques of productivity analysis and science mapping.

The aim of this study then is to fill the gap identified in the current literature by completing possibly the first exhaustive bibliometric analysis of research output in the combined areas of BD, tourism and marketing.

The scientific database, the WoS, was selected for our analysis of trends and prediction of future research paths in this field. The analysis itself was completed through a complete indexation of articles found and the use of the bibliometric research tool SciMAT (Science Mapping Analysis Tool).

This bibliometric analysis makes a longitudinal study of how the research areas of BD, tourism and marketing have evolved with the general aim of identifying the structure of relationships between past and present research themes, as well as determining emerging tendencies. Thus, our research objectives are:

- RO1.* To uncover specific research niches representing emergent areas in the tourism marketing field.
- RO2.* To analyze the body of research in terms of principal authors, volume of publications and most productive categories.
- RO3.* To help academics and professionals gain a better understanding using a schema showing the evolution between 1996 and 2022.
- RO4.* To identify the key thematic areas that have drawn most research interest during the past 27 years.

We believe that one major contribution of this bibliometric analysis is the identification of 10 key themes in the past 27 years of BD research in tourism marketing. Furthermore, this study offers researchers useful information concerning the significance of BD in the development of tourism marketing strategies, both in the present and the future, and it highlights the emerging tendencies on which future investigation should be focused.

Our study begins with an overview of the evolution of BD in tourism marketing and goes on to explain our methods of bibliometric analysis, before giving a detailed explanation of the results of our empirical analysis and future research trends. We conclude with a description of the study’s limitations and its implications for the future.

## **2. The evolution of big data in tourism marketing**

BD first emerged in 1989 with the birth of the World Wide Web. The term refers to the massive volumes of data produced online that are processed at high velocity, have a high level of veracity and comprise huge variety being both complex and diverse.

In the area of tourism, BD enables consumer profiling to create personalized services and make forecasts. Furthermore, recent research shows a clear tendency toward its use in the

---

field of sustainable tourism; thus it has become an essential element in the United Nations plans to achieve its Sustainable Development Goals.

The use of BD in tourism marketing strategies can be explained through the classical resource-based view theory. This arises naturally from the fact that the use of BD requires physical resources such as sufficiently powerful computers; human resources, such as data scientists; and, because it is essential that organizations and corporate processes should be able to adapt to new technologies, intellectual resources like organizational capital.

The three major sources of BD for the tourism industry are as follows (Li *et al.*, 2018): user data or user-generated content (UGC) like text and photos; device data, including that from the global positioning system (GPS) or Bluetooth; and transaction data such as Web searches and online bookings among others.

In the area of tourism marketing, research is dominated by studies that use online ratings and reviews to measure tourist satisfaction. Indeed, there are numerous studies concerning how hotels use electronic word of mouth (eWOM) due to the importance of this phenomenon in attracting tourists.

Furthermore, while the analysis of textual data is still important, photos are beginning to acquire prominence thanks to the development of web 2.0 and social networking platforms such as Instagram, Pinterest, Flickr and Facebook. These data have a diversity of uses, for example, to analyze the attitudes of tourists toward a particular destination, as well as tourist behavior, given that a photo greatly simplifies the process by which travelers can communicate their tourist experiences online. In this way, industry specialists can make recommendations to potential clients, and design marketing strategies to promote particular services or tourist destinations.

### 3. Research design and data collection

To gain an understanding of the themes of BD, tourism and marketing, we performed a bibliometric analysis of academic articles indexed in one of the most important academic databases: WoS. Bibliometric analysis was used as it has several advantages and enables the evaluation of academic research according to objective criteria. It is used as a tool, and it facilitates the identification of new lines of research.

Because the aim of our bibliometric analysis was to evaluate key themes explored by researchers, and identify thematic clusters, it was vital to have a holistic overview of the BD, tourism and marketing themes. We selected WoS over other sources for three reasons.

First and foremost, even though WoS and Scopus are the two most commonly used sources for bibliometric analysis, the WoS database is the only large-scale literature database from as early as 1940 (Calof *et al.*, 2022) and also contains articles from journals identified as having the highest impact factor according to the *Journal Citation Report* index. Second, WoS, compared to Scopus, has the advantage of having its own tourism category. Third, and finally, the WoS database is the most frequently used source of scientific information (Kim and So, 2022).

Several search criteria were deployed to retrieve the articles. In line with Mariani and Baggio (2021) we developed multiple search queries entailing a combination of the focal keywords “big data,” “artificial intelligence,” “machine learning,” “marketing” and “consumer behavior,” with hospitality and tourism words “travel\*,” “touris\*” and “hotel” in the text, abstract and keywords.

As the data used for this study was collected between 1996 and 2022, the search was conducted from the beginning of the coverage up to March 31, 2022. We eliminated articles which were not directly related to the topic of the analysis. The final data set used for the analyses contains 1,152 papers for WoS.

To execute the bibliometric analysis, our sample of articles was grouped into four time periods, each addressing a particular era in the evolution of research into BD techniques, tourism and information technology (Xiang, 2018).

The first time period (1996–2006) corresponds to a phase of explosion and digitalization of information. It is composed of 12 papers and 147 keywords. The second time period (2007–2016) corresponds to a phase of acceleration in the use, storage and processing of massive digital data. It is composed of 112 papers and 387 keywords. The third period (2017–2020) constitutes the most recent phase in which this type of data and its associated technologies are established, and the research field has matured. It is composed of 426 papers and 811 keywords. The last period (2020–2022) corresponds to the two years of the COVID-19 pandemic. It is composed of 602 papers and 578 keywords.

The research methodologies used in this work are in line with the other well-known principles used in bibliometric analyses and quantitative literature reviews (Cobo *et al.*, 2012; Tranfield *et al.*, 2003) (Figure 1).

#### 4. Method: bibliometric analysis using SciMAT

There are two principal methods of bibliometric analysis: *productivity analysis*, which evaluates the impact of academic research, and *science mapping* which enables the visualization of the structure and evolution of concepts within an academic field. This investigation combines both types of analysis to present the most important conceptual domains.

The first stage of our investigation involved a *retrieval of publications* related to BD, tourism and marketing on the WoS database.

Following this, the search was revised for possible errors, and the relevant documents were extracted to begin constructing our thematic network, in this instance using keywords (Cobo *et al.*, 2012). We then constructed a word-network based on keyword co-occurrence, that is, when words appear together in a document this implies a relationship (Cobo *et al.*, 2011a).

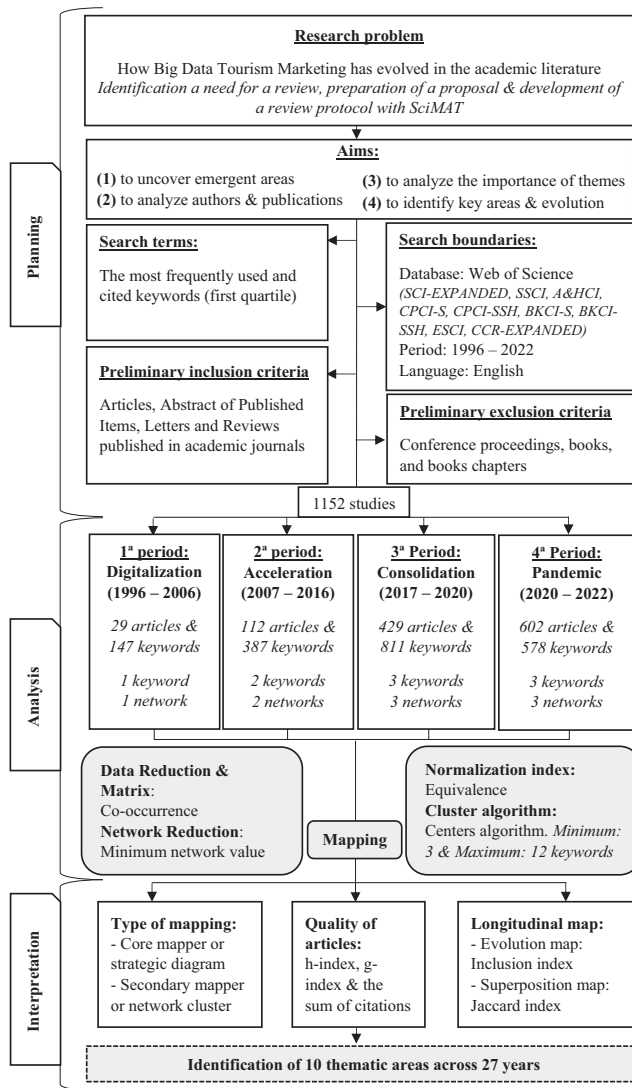
The next step was *relationship network normalization* via the equivalence index, with the aim of calculating the degree of similarity between keywords. This is deemed to be the most appropriate way to normalize co-occurrence frequencies (Cobo *et al.*, 2011b).

After the normalization process, a *science map* was constructed to show the knowledge structure of this research area through its key concepts. The present study used an analysis of co-words in a longitudinal framework (Cobo *et al.*, 2011a). A clustering algorithm was applied to the networks of co-words generated for each of our selected time periods, to identify the most significant word in each cluster.

The *visualization techniques* available in SciMAT enable the representation of the science map with the evolution of thematic areas, through a diagram that allows the representation of two Callon's centrality and Callon's density (Cobo *et al.*, 2011b).

Callon's centrality measures the degree of interaction between one network and other networks. It is defined as:  $c = 10 \times \sum e_{k|h}$ , where  $k$  refers to a keyword belonging to a theme in one network, and  $h$  refers to a keyword belonging to themes in other networks. Callon's density measures the internal strength of the network and is defined as:  $d = 100 (\sum e_{ij} / \omega)$ , where  $i$  and  $j$  are keywords belonging to a given theme, and  $\omega$  is the number of keywords in that theme. Two measures can represent the detected networks. On the strategic diagram, centrality and density are represented on the horizontal and vertical axes, respectively (Figure 2). In this way, the diagram is divided into four categories:

- (1) Driving themes (upper right quadrant): those that are very interrelated, developed in great depth and highly relevant.

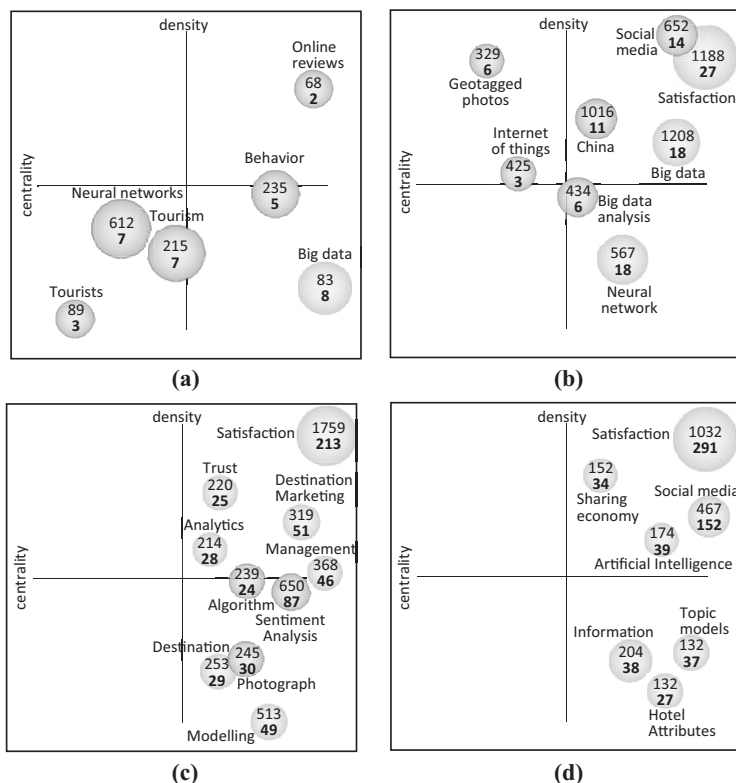


**Figure 1.**  
Analytical process implemented

- (2) Underlying and transversal themes (lower right quadrant): important general themes in the research field but which are less well developed.
- (3) Emerging themes or those in decline (lower left quadrant): under-developed topics.
- (4) Specialized or peripheral themes (upper left quadrant): marginal themes having little relevance to the research field as a whole.

The last step is the *productivity analysis* which incorporates indicators such as the citation number, and the h and g indices. It enables an understanding of which topics are most productive and have the greatest impact.





**Figure 2.** Strategic diagrams between 1996 and 2022 (cites and papers): (a) 1996–2006; (b) 2007–2016; (c) 2017–2020 (March); (d) 2020 (April)–2022 (April)

## 5. Mapping the co-word analysis

### 5.1 Productivity analysis and science mapping

BD has made a significant impact in the field of tourism marketing research. Since 2017, the number of academic articles published in this area has seen a fivefold increase. More than 89% of the articles were published in the past six years.

Of the 446 journals included in the database, only 6% are directly related to tourism marketing, that is to say, 26 journals containing 71 articles.

The majority of the articles are not published in tourism marketing journals but are distributed across a variety of journals focusing on other disciplines such as management, sustainability and technology. The category of Hospitality, Leisure, Sport and Tourism itself contains 476 articles and Tourism Management is the second most productive category, with 224 articles published in this area. Finally, the most productive authors are Rob Law (School of Hotel and Tourism Management, Hong Kong) and Zheng Xiang (Virginia Tech, Beijing Union University).

Certain themes have established their intrinsic importance throughout the 27 years studied here and we will discuss their development in what follows (see Table 1).

### 5.2 First period: digitalization of information (1996–2006)

Only 12 relevant articles appear in this 11-year period (Figure 2a).

5.2.1 *Driving themes: “website,” “photographs,” “performance,” “online reviews” and “tourism patterns.”* The most highly related and most relevant driving themes are “website” and “online reviews” (Cobo *et al.*, 2011a).

The “website” cluster demonstrates the growing importance of three areas of research: traveler experiences recorded on blogs and Facebook; consumer perspectives on the personalization of products and services; and smart cities in Asia via the Internet of Things. The “online reviews” topic is connected with sentiment analysis for segmenting the international tourist market.

Time periods	The period	Theories	Big data tourism marketing			
			Research	Analysis	Platform database	Country database
<i>Digitalization of information (1996–2006)</i>	Big data to explore traveler experiences and perspectives	Cognitive theories	Prediction of tourist demand; segmentation of the tourism market and decision-making through websites	Sentiment; cluster, semantic and textual, multiple linear regression; neural networks and data mining	Sina Weibo; Facebook; travel blogs; and Google	China; USA; Europe; and Japan
<i>Acceleration (2007–2016)</i>	Big data to explore destination marketing and image and traveler motivations	Sign theory; attribution theory; transaction cost theory; expectancy theory; and RTB perspective	Tourist satisfaction and WOM through geo-tagged data, UGC, social media	Spatial; cluster analysis; texts; time series; and photos	Facebook; Twitter; Flickr; and CRM	China
<i>Consolidation (2017–2020)</i>	Big data becomes a fundamental knowledge creation tool	Resource and capabilities theory; long short-term memory; and service-dominant logic	Connection between WOM, experience reviews and hotel attributes	Semantic; spatial; sentiment analysis; cluster; machine learning; and content analysis	Facebook; Booking.com; Airbnb; Flickr; TripAdvisor; Instagram; Twitter; and phone data	China
<i>COVID-19 (2020–2022)</i>	Big data to predict the purchase intention of tourists and to analyze user experience	Social exchange theory; service management theory	Behavior and WOM; social media; engagement; destination marketing; and smart tourism	Sentiment; texts; spatial, cluster; Bayesian; semantic; machine learning; neural networks	Twitter; Airbnb; TripAdvisor; Facebook; Phone data; Sina Weibo	China; USA; and Spain

**Table 1.** Summary of the most important aspects of the four periods



---

“Performance” and “tourism patterns” are concerned with forecasting in the tourism sector which studies segmentation strategies and the results in terms of performance (Curry *et al.*, 2001) using social networks such as Sina Weibo.

The “photographs” topic is connected with analysis of smart tourism and ecotourism, and how to segment the market through self-organizing maps. Here, investigation predominantly focuses on the tourist motivations which have the greatest weight in buying decisions in the senior-tourist market segment (Kim *et al.*, 2003).

*5.2.2 Underlying and transversal themes: “behavior” and “big data.”* Tourist behavior is the most relevant of all the themes identified. Articles belonging to this cluster focus on environmental behavior, post-buying behavior, and forecasting tourist behavior. In addition, work in this area relies on two cognitive theories: the theory of reasoned action and its extension the theory of planned behavior. These theories are considered to offer the best framework for understanding tourist behavior (Hsu and Huang, 2012).

The “behavior” theme is, in turn, related to others such as loyalty, market segmentation, mobility, demand and tourism forecasting. The majority of this research strand comes from the USA.

The application of human–computer interaction theory is another important topic here. This theory establishes the fundamentals for an understanding of tourists’ behavior in terms of how they search for and plan their trips (Xiang, 2018).

To understand “consumer behavior,” researchers have used BD techniques such as time series (Pattie and Snyder, 1996), and lexicon and text mining or modeling (Bloom, 2004), and have predicted things like loyalty, sales and tourist satisfaction.

*5.2.3 Emergent themes: “neural networks” and “tourism and hospitality.”* The theme “neural networks” is associated with predicting trends in “tourism demand” through the use of BD. Specifically, it links to how BD can improve models used in econometric forecasting (Witt and Witt, 1995) through the use of artificial neural networks and so enable the development of improved tourism demand models (Palmer *et al.*, 2006). Japan, China and Spain are connected to this theme. The most common types of analysis are cluster and multiple linear regression.

### *5.3 Second period: acceleration (2007–2016)*

The total number of articles belonging to this period is 112, so is evidence of the huge growth index for publications in this field (Figure 2b). Topics such as “tourist satisfaction,” “big data,” “neural networks,” “China” and “social media” achieved 5,350 citations.

*5.3.1 Principal driving theme: “tourist satisfaction”.* This is the most important driving theme in the field, leading in terms of number of documents, citations and values of h and g indices. It is strongly linked to WOM as recorded in reviews left by travelers describing their experiences in hotels, and the impact of these reviews on sales is also a topic of study.

This decade is characterized as an era of acceleration due to the enormous increase in UGC on the internet. This factor, among others, has enabled the in-depth study of eWOM (Ghose *et al.*, 2012). UGC, comprising any online data either in the form of text or images, makes up almost 50% of BD in connection with tourism (Li *et al.*, 2018). The reason for its extensive use lies in the fact that it can be easily accessed and processed, and indeed, it is very low cost (Karimi *et al.*, 2020).

The predominant theoretical frameworks applied in this era include sign theory, attribution theory, transaction cost theory and expectancy theory. This demonstrates the impact of reviews in the description of consumer experience.

Online reviews are one of the significant elements in eWOM which can influence future demand from other clients, and as a result, has important commercial value (Xie *et al.*, 2014).

---

This is due to the way it can enable forecasting of future profits for hotels, decisions concerning the location of accommodation and room rates, as well as the improvement of results based on performance (Pan and Yang, 2017).

A predominant trend here is articles addressing new ways of categorizing hotels based on the mean perceived utility of specific hotel features (Berezina *et al.*, 2016). Other important work involves identification of which sorts of messages posted on social media enabled the greatest user interaction or the possibility of virality (Mariani *et al.*, 2016). In this respect, Facebook and Twitter stand out.

*5.3.2 Driving and transversal themes: “big data” and “neural networks.”* Alongside “tourist satisfaction,” these are the other driving themes in the second period. Both these concepts are cornerstones of marketing, due to their capacity to positively influence the performance of an organization. In this way, they are very interrelated terms and, in addition, are linked to the themes “perceived quality of service” and “loyalty,” which in turn are strongly connected to “tourist satisfaction.”

A large proportion of articles addresses the theme of “performance” and analyzes which variables affect tourism-business outcomes within a competitive environment. Among the areas that have received most attention in this regard are the quality of hotel services, and hotel attributes and efficiency, in addition to the identification of factors determining tourist satisfaction and appropriate strategic decision-making (Moutinho *et al.*, 2015). The most common types of analysis are spatial (Supak *et al.*, 2015), cluster (Brida *et al.*, 2012), textual (Krawczyk and Xiang, 2016), time series (Claveria and Torra, 2014), fuzzy system (Shahrabi *et al.*, 2013) and photo-sharing analysis (García-Palomares *et al.*, 2015).

*5.3.3 Secondary underlying and transversal themes: “administration and management,” “destination marketing” and “social media analyses.”* These three topics constitute the underlying transversal themes of research in this second period.

“Administration and management,” which began as a driving theme moves to being a transversal theme, that is, we see its consolidation. In the course of this theme’s evolution, BD research can be seen to undergo significant development, enabling it to encompass the problems of tourism management (Xiang, 2018). In addition, this topic is aligned with the evolution in tourism demand. In this area, three big powers stand out: China, the USA and Europe, specifically Spain. In fact, “Europe” moves from being an emergent theme to become integrated into an essential cluster.

The topic of “destination marketing” is linked to the study of tourism destinations and traveler motivations. Of great importance here is the use of images and websites that guide traveler management (Xiang, 2018). It is a fundamental theme from the resource-based theory, because online visibility is a differentiating factor leading to superior business performance because it potentially helps attract more tourists enabling increased rates of occupancy (Smithson *et al.*, 2011).

Finally, the “analysis of social media” appears as an underlying theme. Understanding clients through the reviews left on social media platforms such as Twitter constitutes a key factor for success in the era of BD (Park *et al.*, 2016). The principal techniques used in this field include neural networks and data mining.

*5.3.4 Emergent areas: “pricing” and “geo-tagged data.”* These two themes are considered emergent areas. In contrast to the first period, these terms are now important, and they will have importance in the following (third) time period.

The “pricing” theme shows strong links to airlines through revenue management, pricing strategies and tourist satisfaction with low-cost or full-service carriers (Leong *et al.*, 2015).

---

Through the use of geographic information systems, “geo-tagged data” has enabled the use of photos obtained principally from the Flickr social media platform (Levin *et al.*, 2015).

#### 5.4 Third period: consolidation (2017–2020)

Over these four years, the research field has grown with 426 articles (Figure 2c). Over this time period, tourism research undergoes a dramatic change as BD becomes a fundamental knowledge creation tool. This transformation is without precedent in academic research, and is thanks to ever more efficient management of the millions of bytes of data generated (Batista e Silva *et al.*, 2018).

*5.4.1 Principal driving theme: “tourist satisfaction.”* This is the highest central theme in the third period and is a topic that has gained importance with respect to the previous period. Tourism literature establishes general tourist satisfaction, and indeed tourists’ intention to return to a given destination is effected by many different destination attributes (Alegre and Garau, 2010). For instance, consumers gain a specific degree of satisfaction as a function of their perceptions concerning the various attributes of hotels, thus perceptions represent one dimension of satisfaction (Guo *et al.*, 2017).

This topic is strongly related to themes in the “tourist satisfaction” cluster from the second period, such as online and offline reviews, hotels and tourist intentions. Topics such as loyalty, and hotel attributes and service quality that were previously related to perceptions are now linked with satisfaction. Furthermore, terms such as “Twitter” and “UGC” have disappeared. Research is no longer so focused on general social networks, but rather on those that are specifically concerned with tourism such as TripAdvisor.

Data from reviews and blogs are now principally used in studies of satisfaction, recommendations and tourist opinion (Deng and Li, 2018).

*5.4.2 Secondary driving themes: “management,” “mobility,” “trust” and “destination marketing.”* Together with “tourist satisfaction,” these are among the driving themes of the third period. “Management” is a topic of relatively high importance in all the periods studied and, in the third period is once again a driving theme.

This cluster is related to other topics such as “social networks,” “Facebook” and “engagement.” The investigations in which these terms appear focus on the strategic use of Facebook to promote and market destinations (Mariani *et al.*, 2018); on the analysis of opinions using texts (Zola *et al.*, 2019); and the generation of commitment (Villamediana-Pedrosa *et al.*, 2019).

The topic of “mobility” involves examples of the use of data obtained from GPS, social media and mobile telephones used between cities, and at open-air venues hosting sporting events or festivals (Salas-Olmedo *et al.*, 2018). The theme of “trust,” on the other hand, exemplifies the growth of concerns and problems associated with engagement in the so-called trust economy (Xiang, 2018), specifically Airbnb and Booking.com. Variables such as reputation, communication and pricing strategies are found to be moderating factors in the “trust” theme.

With respect to the “destination marketing” theme, here UGC predominates, as do marketing strategies on social networks and their analysis. In this way, organizations can understand the perceptions of users and develop strategies to promote revisiting.

In all, 73% of the articles look at tourist destination image. This theme has evolved from being dominated by the destination marketers, to become a dynamic process of interaction between tourists and promotion, before finally reaching a new era in which destination management organizations examine and modify their projected destination image based principally on behavior, perceptions, experiences and the diffusion of information by tourists on social networking platforms.

---

“Destination marketing” is related to heritage too, as well as rural tourism in protected areas and National Parks. Two basic objectives dominate: developing branding strategies and extracting trends in this area of tourism, with sustainability and ecological protection high on the agenda. The most common type of analysis is content analysis.

*5.4.3 Underlying and transversal themes: “tourism destinations” and “photographs.”* Besides tourist satisfaction, these constitute the most important underlying and transversal themes in this period. Both are related to the analysis of geo-tagged text and images obtained from social media platforms such as Facebook, Twitter, TripAdvisor and Sina Weibo.

To improve their business intelligence, “tourist destinations” are supported by tools such as customer relationship management (CRM). The surge in social networks challenges traditional notions of how to manage client relationships, and thus social-CRM has appeared on the scene (Chan *et al.*, 2018).

In terms of size, the “social networks” cluster clearly stands out. Current literature concerning CRM focuses on the analysis of BD and the use of social networking platforms to capture huge amounts of data and take advantage of customers’ improved interactivity to personalize services (Sota *et al.*, 2020). TripAdvisor appears as the most widely used platform in terms of marketing strategies. Another area of high research activity is applied studies concerning China and sport tourism.

“Photographs” in conjunction with “tourism destinations” constitute the underlying and transversal themes of the third period of study.

This topic is highly related to the management and promotion of hotel rooms and online bookings, as well as attempts to better understand client profiling via BD (Liu *et al.*, 2019). Furthermore, the availability of large sets of photos from trips shared online provides an accessible source of data for tourism researchers (Ma *et al.*, 2020). This type of content can be interpreted through semiotic theory. The principal origin of online photographic content is social media such as Twitter, Instagram and Flickr, as well as blogs. These enable study of the discovery and development of tourist routes, marketing strategies and tourism patterns, and can be differentiated into two types: concerning travelers or trips. At present, tourism research related to photos is dominated by Flickr, despite the fact that Instagram has more users and contains more images.

*5.4.4 Emergent themes: “market segmentation” and “internet.”* The “internet,” understood as the tool that provides the raw data on which the techniques of BD can operate, is starting to manifest as an emergent theme in the context of tourism marketing because it enables accommodation providers to adapt, for example, room characteristics and pricing strategies.

A further area of high interest is “market segmentation,” related to recommendation systems via the “internet” cluster. Both of these themes are themselves strongly linked to co-creation which enables, among other things, the personalization of products through market segmentation using traveler preference data and geo-localized data extracted from mobile phones. The use of BD techniques to segment the tourism market, in fact, continues to be recognized as a key source of value creation in the fourth time period.

### *5.5 Fourth period: COVID-19 (2020–2022)*

To supplement this investigation in the wake of the global COVID-19 pandemic, a further 602 articles published during the pandemic were added to our database. This additional, newly published work constitutes 50% of our database (Figure 2d).

*5.5.1 Principal driving themes: “tourist satisfaction,” “social media,” “sharing economy,” “consumer” and “artificial intelligence.”* The theme “tourist satisfaction” continues to be the most important theme despite the COVID-19 pandemic. During these two last years studied,

---

the number of studies dealing with BD see continued growth, particularly in reviews concerning the prediction of customer purchase preferences and its impact, and in looking at user experiences and perceptions through content analysis or making use of data gathered from platforms such as TripAdvisor. Specifically, areas being investigated include consumer behavior and social media marketing (Nilashi *et al.*, 2021), and engagement with social exchange theory (Song *et al.*, 2020).

The most extensively studied theme in this respect is sentiment analysis applied to text-based and photographic UGC shared on social media platforms, particularly Twitter. This analysis has allowed researchers to deepen and advance their understanding of destination marketing in the promotion of products and services.

The “sharing economy” is another theme that has gained importance in this last time period, with most research focusing on the social media site Airbnb (Canziani and Nemati, 2021).

In addition, during this period, AI has become a consolidated topic with machine learning emerging as the most widely used technique to study the tourism ecosystem. Several Spanish authors specialize in the use of these techniques (Marine-Roig and Huertas, 2020; Sánchez-Martin *et al.*, 2020; Valls and Roca, 2021) and they have been applied particularly successfully in the areas of tourism innovation and forecasting, decision-making and the analysis of performance and strategy.

*5.5.2 Underlying and transversal themes: “hotel attributes” and “deep learning”.* These two themes are consolidated during the two years of the COVID-19 pandemic becoming transversal topics. In particular, “hotel attributes” have been studied in relation to competitiveness, rating and the effect they have on WOM. The forms of data gathering most widely used include text and data mining which enable the analysis of language and emotions through text. “Deep learning” is another important tool as it facilitates visual analysis, the prediction of occupancy and opinion classification (Gómez *et al.*, 2021), all of which help tourism managers to develop and promote appropriate response strategies informed by service management theory (Zhu *et al.*, 2021). In this area, China appears to be the most visible.

*5.5.3 Emergent themes: “sustainability,” “tourist recommendation,” “social media analysis,” “values,” “prices” and “gastronomy.”* The bibliometric analysis undertaken has allowed us to identify the emergent themes that are likely to become increasingly important in the future.

*Sustainability.* The number of studies concerning profitability and perceptions in ecotourism is growing exponentially. The principal sources of data for this work are Google data and geo-tagged photographs. Analyzing trends in ecotourism is part of a strategic approach to assessing progress toward the UN’s Sustainable Development Goals (Go *et al.*, 2020).

*Tourist recommendation.* An emergent theme in the third time period, market segmentation continues to be important in this time period, and as before, it is driven by tourist recommendation. Researchers continue to use BD to analyses tourist recommendations, and additionally we see this source of data being applied to new variables such as types of tourism, length of stay, attachment and quality of service (Penagos-Londoño *et al.*, 2021).

*Social media analysis.* A particular use of this type of analysis is to look at revisit intentions in hospitality. This concept is integral to the relationship between marketing and customer loyalty, and has traditionally been investigated largely through customer surveys using closed-ended questions (Liu and Beldona, 2021). Currently, there is an exponential growth in revisit intention analysis, particularly to look at decision making in hotel management, with researchers now turning to supervised machine learning rather than using social media analysis.



---

*Values.* Little is known about the influence of cultural factors in consumers' evaluations of review helpfulness, and as a result, research into values, particularly using the theory of dominant logic, must be categorized as an emergent theme (Filieri and Mariani, 2021).

*Prices.* Researchers are beginning to apply BD techniques to understanding how differences in market perception and information create a price differential (Casamatta et al., 2022). Until now, setting the price for new accommodation has been often based largely on location, number of beds and type of house, among other physical factors. However, the use of machine learning and intention analysis is beginning to take over as the means for price prediction in online booking systems (Trang et al., 2021).

*Gastronomy.* In the third time period studied, there were only three articles considering this topic and thus, it was considered isolated and highly specialized. In the fourth time period, however, we identified 14 articles concerning gastronomy, and thanks to this increased research interest, it must now be considered an emerging theme. Particular work worth highlighting includes a study using neural networks, an otherwise rarely used technique in the tourism sector, to construct gastronomic tourist profiles through behavioral analysis (Moral-Cuadra et al., 2021). In addition, new research is emerging concerning the design of gastronomic experiences based on consumer opinion, that is, involving co-creation (Lin et al., 2022). The exponential growth in co-creation strategies has already been pointed out by other authors.

#### 5.6 Ten thematic areas across 27 years

Here, we give a structural analysis of the evolution of an academic field that has matured over the past 27 years. This analysis shows the development of 10 key areas (shaded with 10 different colors in Figure 3): "destination marketing," "mobility patterns," "co-creation," "gastronomy," "sustainability," "tourist behavior," "market segmentation," "artificial neural networks," "pricing" and "tourist satisfaction." The literature demonstrates a solid cohesion because many of the same themes appear in all four of the different periods of development identified, showing the consolidation of these themes.

In *the first period* we examined, there are two thematic areas which might be described as classic: "mobility patterns" (81 papers and 988 citations) and "tourist behavior" (81 papers and 1,474 citations). In *the second period*, two further topics are added to the list: "tourist satisfaction" (541 papers and 4,379 citations) and "pricing" (181 papers and 1,195 citations). In *the third period*, two further topics are added to the list: "destination marketing" (220 papers and 1,450 citations) and "co-creation" (40 papers and 639 citations). These three periods represent the basis of BD tourism marketing research and show a highly developed line of investigation: the prediction of behavior patterns based on geo-tagged content enabling the improvement of strategies for destination marketing.

*The fourth period of study*, composed of articles published most recently (2020–2022) and thus affected by the COVID-19 pandemic, contains several emergent themes that may well gain importance in the future. These topics include, "gastronomy" (17 papers and 86 citations), "market segmentation" (75 papers and 1,577 citations), "sustainability" (55 papers and 768 citations) and "artificial neural networks" (158 papers and 2,447 citations). Artificial neural networks in particular have been in use from the beginnings of applied artificial intelligence (AI) in tourism marketing. However, it is only in recent years that their use has become widespread, and they should now be considered among the most important tools in tourism marketing (Mariani and Baggio, 2021).

The two themes that stand out most in terms of impact indices are tourist satisfaction and destination marketing. These topics can, therefore, be considered as those of central importance are fundamental to the development of the whole field.



**Figure 3.**  
Thematic map of big  
data tourism  
marketing literature  
(1996–2022)

The “tourist satisfaction” theme shows a definitive upward trend with respect to relevant indices and citation numbers. This theme starts with a very small footprint which has grown and reflects the rapid development of this topic such that it is now considered as one of the leading areas of research. On the other hand, topics such as “astro-tourism” initially achieved high impact, but this has not grown over time. Other areas exist that have maintained their relevance throughout the 27 years studied, for example, “pricing” and still others, such as “co-creation” and “gastronomy” that have expanded, branching into new themes and gaining relevance in each subsequent time period.



---

The fourth period indicates the expanding use of BD in the field of tourism marketing and the increasing multidisciplinary of the areas under investigation.

## 6. Discussion

There are several conclusions in the present study. Among the most important of these is revealing the direction of future research trends as well as identifying the structure of relationships between current and past themes in the research areas of BD, tourism and marketing.

This is the first study to apply a bibliometric approach to a clear gap in the research, in that it covers these three thematic areas simultaneously. In addition, it is unique in covering such a wide time period, from 1996 to 2022; thus, it includes the two years corresponding to the COVID-19 pandemic. This two-year period is significant as it was particularly productive and saw the emergence of several new themes.

In this way, we have been able to identify tools, types of BD techniques, authors and most importantly, conceptual themes that have played the most vital roles in this research field throughout the 27 years studied. Thus, as explained previously, this work constitutes a significant contribution to the field by uniquely covering BD, tourism and marketing.

We developed a schematic diagram to show the evolution of principal research themes from 1996 to 2022, divided into four individual time periods. To this end, we used the SciMAT to make an initial, exhaustive bibliometric search of the literature with 1,152 articles published on WoS. This constitutes the entire academic output in this field to date and publications can be divided into four categories corresponding to different periods: digitalization of information (1996–2006); acceleration (2007–2016); consolidation (2017–2020); and COVID-19 (2020–2022).

To aid analysis, the body of research considered in this study was separated into ten major thematic areas: “destination marketing,” “mobility patterns,” “co-creation,” “gastronomy,” “sustainability,” “tourist behavior,” “market segmentation,” “artificial neural networks,” “pricing” and “tourist satisfaction.”

A particularly important area was “tourist satisfaction,” which shows an upward trend through the full 27-year span of this study, reaching what might be called its golden era in the third time period considered. Tourism research defines the general concept of tourist satisfaction and also identifies several dimensions, among which one of the most important is visitor perceptions of hotel attributes. The analysis of tourist satisfaction has been assisted primarily by marketing platforms on social media networks. In recent times, certain networks, such as Twitter, have declined in importance, giving way to other UGC platforms like TripAdvisor which allows access to tourists’ opinions through the reviews they leave.

The most important aspect of this work has been the identification of future lines of investigation and where there is a need to deepen our understanding in certain fields.

## 7. Implications

This investigation highlights the relevance of BD in tourism marketing research, demonstrates its importance to business and offers relevant and empirical information to tourism-related organizations and private businesses.

In the first place, this review suggests that researchers are interested in BD, tourism and marketing in many different disciplines. In fact, our analysis shows that many of the academics contributing to the field of BD and tourism do not publish in marketing journals. Thus, we would suggest that more interdisciplinary collaboration would help advance the field and, perhaps, this observation constitutes one of the principal contributions of this work. Through this analysis, we hope to provide information concerning new opportunities for research and help to strengthen

lines of investigation that may be of potential interest both for academics and practitioners in this field. This is especially important for establishing possible collaborations between these two groups.

In the second place, marketing professionals should invest in more research into the problems they wish to solve using BD and AI since, as we have seen, their current uses are many and varied: predicting tourism demand, analyzing tourist satisfaction, or market segmentation. On the basis of such research, businesses could obtain a variety of appropriate data for every type of analysis or purpose proposed.

In the third place, while the tourism industry is making effective investment in the management of BD and its analysis of AI, this bibliometric analysis demonstrates that the contribution of academic research is also significant. Thus, collaboration between industry and academia would further invigorate this area of research and facilitate its advance.

Finally, given that the rate of evolution in marketing strategies based on new technologies is extremely fast moving, leading hotel and tourism businesses, and indeed, marketing consultants, must make use of AI to improve, innovate and extract the maximum value from data. Furthermore, this may be even more important in the wake of the COVID-19 pandemic, as this work demonstrates that the correct management of data is increasingly invaluable to the industry being able to respond and adapt to external shocks. This information can then be used to plan more efficient business strategies focused on specific types of clients.

## 8. Limitations and future research

It is necessary to address the limitations of this study. The use of other databases such as Scopus or Google scholar might have provided additional results. Thus, WoS was considered adequate for our purposes.

Despite this limitation, we feel this investigation is of undoubted interest. It provides a novel, possibly the only, presentation of the major trends in this area of research and as a result provides a point of departure for academics and practitioners to discover new avenues of investigation, as well as strengthening already established lines of research, for example, the “sustainability” theme in which it recommends considering the profitability of hotel businesses and tourist perceptions; or “gastronomy,” where there is a large gap in the literature concerning the gastronomical profiling of tourists, and this could be solved by the use of techniques such as neural networks. Other emergent themes are “social media analysis” to study tourist decision-making, “values” and “prices.”

## References

- Alegre, J. and Garau, J. (2010), “Tourist satisfaction and dissatisfaction”, *Annals of Tourism Research*, Vol. 37 No. 1, pp. 52-73.
- Batista e Silva, F., Marín Herrera, M.A., Rosina, K., Ribeiro Barranco, R., Freire, S. and Schiavina, M. (2018), “Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources”, *Tourism Management*, Vol. 68, pp. 101-115.
- Berezina, K., Bilgihan, A., Cobanoglu, C. and Okumus, F. (2016), “Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews”, *Journal of Hospitality Marketing and Management*, Vol. 25 No. 1, pp. 1-24.
- Bloom, J.Z. (2004), “Tourist market segmentation with linear and non-linear techniques”, *Tourism Management*, Vol. 25 No. 6, pp. 723-733.
- Brida, J.G., Disegna, M. and Osti, L. (2012), “Segmenting visitors of cultural events by motivation: a sequential non-linear clustering analysis of Italian christmas market visitors”, *Expert Systems with Applications*, Vol. 39 No. 13, pp. 11349-11356.

- 
- Calof, J., Söilen, K.S., Klavans, R., Abdulkader, B. and Moudni, I.E. (2022), "Understanding the structure, characteristics, and future of collective intelligence using local and global bibliometric analyses", *Technological Forecasting and Social Change*, Vol. 178, p. 121561.
- Canziani, B. and Nemati, H.R. (2021), "Core and supplemental elements of hospitality in the sharing economy: insights from semantic and tonal cues in airbnb property listings", *Tourism Management*, Vol. 87, p. 104377.
- Casamatta, G., Giannoni, S., Brunstein, D. and Jouve, J. (2022), "Host type and pricing on airbnb: seasonality and perceived market power", *Tourism Management*, Vol. 88, p. 104433.
- Chan, I.C.C., Fong, D.K.C., Law, R. and Fong, L.H.N. (2018), "State-of-the-art social customer relationship management", *Asia Pacific Journal of Tourism Research*, Vol. 23 No. 5, pp. 423-436.
- Claveria, O. and Torra, S. (2014), "Forecasting tourism demand to Catalonia: neural networks vs. time series models", *Economic Modelling*, Vol. 36, pp. 220-228.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2011a), "An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the fuzzy sets theory field", *Journal of Informetrics*, Vol. 5 No. 1, pp. 146-166.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2011b), "Science mapping software tools: review, analysis, and cooperative study among tools", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 7, pp. 1382-1402.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E. and Herrera, F. (2012), "SciMAT: a new science mapping analysis software tool", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 8, pp. 1609-1630.
- Curry, B., Davies, F., Phillips, P., Evans, M. and Moutinho, L. (2001), "The Kohonen self-organizing map: an application to the study of strategic groups in the UK hotel industry", *Expert Systems*, Vol. 18 No. 1, pp. 19-31.
- Deng, N. and Li, X.R. (2018), "Feeling a destination through the 'right' photos: a machine learning model for DMOs' photo selection", *Tourism Management*, Vol. 65, pp. 267-278.
- Filieri, R. and Mariani, M. (2021), "The role of cultural values in consumers' evaluation of online review helpfulness: a big data approach", *International Marketing Review*, Vol. 38 No. 6, pp. 1267-1288.
- García-Palomares, J.C., Gutiérrez, J. and Mínguez, C. (2015), "Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS", *Applied Geography*, Vol. 63, pp. 408-417.
- Ghose, A., Ipeiritos, P.G. and Li, B. (2012), "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content", *Marketing Science*, Vol. 31 No. 3, pp. 493-520.
- Go, H., Kang, M. and Nam, Y. (2020), "The traces of ecotourism in a digital world: spatial and trend analysis of geotagged photographs on social media and Google search data for sustainable development", *Journal of Hospitality and Tourism Technology*, Vol. 11 No. 2, pp. 183-202.
- Gómez, M., Tinoco Guerrero, N.S. and Tinoco Guerrero, L.M. (2021), "The influence of airbnb on hotel occupancy in Mexico: a big data analysis (2007-2018)", *Revista Cimexus*, Vol. 16 No. 1, pp. 9-32.
- Guo, Y., Barnes, S.J. and Jia, Q. (2017), "Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation", *Tourism Management*, Vol. 59, pp. 467-483.
- Hall, C.M. (2011), "Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism", *Tourism Management*, Vol. 32 No. 1, pp. 16-27.
- Hsu, C.H.C. and Huang, S. (2012), "An extension of the theory of planned behavior model for tourists", *Journal of Hospitality and Tourism Research*, Vol. 36 No. 3, pp. 390-417.
- Karimi, S., Shakery, A. and Verma, R. (2020), "Online news media website ranking using user-generated content", *Journal of Information Science*, Vol. 47 No. 3, pp. 340-358.

- Kim, H. and So, K.K.F. (2022), "Two decades of customer experience research in hospitality and tourism: a bibliometric analysis and thematic content analysis", *International Journal of Hospitality Management*, Vol. 100, p. 103082.
- Kim, J., Wei, S. and Ruys, H. (2003), "Segmenting the market of west Australian senior tourists using an artificial neural network", *Tourism Management*, Vol. 24 No. 1, pp. 25-34.
- Kontogianni, A. and Alepis, E. (2020), "Smart tourism: state of the art and literature review for the last six years", *Array*, Vol. 6, p. 100020.
- Köseoglu, M.A., Sehitoglu, Y. and Craft, J. (2015), "Academic foundations of hospitality management research with an emerging country focus: a citation and co-citation analysis", *International Journal of Hospitality Management*, Vol. 45, pp. 130-144.
- Köseoglu, M.A., Rahimi, R., Okumus, F. and Liu, J. (2016), "Bibliometric studies in tourism", *Annals of Tourism Research*, Vol. 61, pp. 180-198.
- Krawczyk, M. and Xiang, Z. (2016), "Perceptual mapping of hotel brands using online reviews: a text analytics approach", *Information Technology and Tourism*, Vol. 16 No. 1, pp. 23-43.
- Leong, L.Y., Hew, T.S., Lee, V.H. and Ooi, K.B. (2015), "An SEM-artificial-neural-network analysis of the relationships between SERVPERF, customer satisfaction and loyalty among low-cost and full-service airline", *Expert Systems with Applications*, Vol. 42 No. 19, pp. 6620-6634.
- Levin, N., Kark, S. and Crandall, D. (2015), "Where have all the people gone? Enhancing global conservation using night lights and social media", *Ecological Applications*, Vol. 25 No. 8, pp. 2153-2167.
- Li, J., Xu, L., Tang, L., Wang, S. and Li, L. (2018), "Big data in tourism research: a literature review", *Tourism Management*, Vol. 68, pp. 301-323.
- Lin, M.-P., Marine-Roig, E. and Llonch-Molina, N. (2022), "Gastronomic experience (co)creation: evidence from Taiwan and Catalonia", *Tourism Recreation Research*, Vol. 47 No. 3, pp. 277-292.
- Liu, Y. and Beldona, S. (2021), "Extracting revisit intentions from social media big data: a rule-based classification model", *International Journal of Contemporary Hospitality Management*, Vol. 33 No. 6, pp. 2176-2193.
- Liu, P., Zhang, H., Zhang, J., Sun, Y. and Qiu, M. (2019), "Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: from online to arrival", *Tourism Management*, Vol. 73, pp. 105-114.
- Ma, S., Kirilenko, A.P. and Stepchenkova, S. (2020), "Special interest tourism is not so special after all: big data evidence from the 2017 great American solar eclipse", *Tourism Management*, Vol. 77, p. 104021.
- Mariani, M. and Baggio, R. (2021), "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.
- Mariani, M., Di Felice, M. and Mura, M. (2016), "Facebook as a destination marketing tool: evidence from Italian regional destination management organizations", *Tourism Management*, Vol. 54, pp. 321-343.
- Mariani, M., Mura, M. and Di Felice, M. (2018), "The determinants of Facebook social engagement for national tourism organizations' Facebook pages: a quantitative approach", *Journal of Destination Marketing and Management*, Vol. 8, pp. 312-325.
- Marine-Roig, E. and Huertas, A. (2020), "How safety affects destination image projected through online travel reviews", *Journal of Destination Marketing and Management*, Vol. 18, p. 100469.
- Moral-Cuadra, S., Solano-Sánchez, M.Á., Menor-Campos, A. and López-Guzmán, T. (2021), "Discovering gastronomic tourists' profiles through artificial neural networks: analysis, opinions and attitudes", *Tourism Recreation Research*, Vol. 47 No. 3, pp. 347-358.

- 
- Moutinho, L., Caber, M., Silva, M.M. and Albayrak, T. (2015), "Impact of group package tour dimensions on customer satisfaction (an ANNs application)", *Tourism Analysis*, Vol. 20 No. 6, pp. 619-629.
- Nilashi, M., Asadi, S., Minaei-Bidgoli, B., Ali Abumalloh, R., Samad, S., Ghabban, F. and Ahani, A. (2021), "Recommendation agents and information sharing through social media for coronavirus outbreak", *Telematics and Informatics*, Vol. 61, p. 101597.
- Palmer, A., Montaña, J. and Sesé, A. (2006), "Designing an artificial neural network for forecasting tourism time series", *Tourism Management*, Vol. 27 No. 5, pp. 781-790.
- Pan, B. and Yang, Y. (2017), "Forecasting destination weekly hotel occupancy with big data", *Journal of Travel Research*, Vol. 56 No. 7, pp. 957-970.
- Park, S.B., Ok, C.M. and Chae, B.K. (2016), "Using twitter data for cruise tourism marketing and research", *Journal of Travel and Tourism Marketing*, Vol. 33 No. 6, pp. 885-898.
- Pattie, D.C. and Snyder, J. (1996), "Using a neural network to forecast visitor behavior", *Annals of Tourism Research*, Vol. 23 No. 1, pp. 151-164.
- Penagos-Londoño, G.I., Rodriguez-Sanchez, C., Ruiz-Moreno, F. and Torres, E. (2021), "A machine learning approach to segmentation of tourists based on perceived destination sustainability and trustworthiness", *Journal of Destination Marketing and Management*, Vol. 19, p. 100532.
- Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C. and Gutiérrez, J. (2018), "Tourists' digital footprint in cities: comparing big data sources", *Tourism Management*, Vol. 66, pp. 13-25.
- Samara, D., Magnisalis, I. and Peristeras, V. (2020), "Artificial intelligence and big data in tourism: a systematic literature review", *Journal of Hospitality and Tourism Technology*, Vol. 11 No. 2, pp. 343-367.
- Sánchez-Martín, J.M., Gurria-Gascón, J.L. and Rengifo-Gallego, J.I. (2020), "The distribution of rural accommodation in Extremadura, Spain-between the randomness and the suitability achieved by means of regression models (OLS vs. GWR)", *Sustainability*, Vol. 12 No. 11, p. 4737.
- Shahrabi, J., Hadavandi, E. and Asadi, S. (2013), "Developing a hybrid intelligent model for forecasting problems: case study of tourism demand time series", *Knowledge-Based Systems*, Vol. 43, pp. 112-122.
- Smithson, S., Devece, C.A. and Lapiedra, R. (2011), "Online visibility as a source of competitive advantage for small- and medium-sized tourism accommodation enterprises", *The Service Industries Journal*, Vol. 31 No. 10, pp. 1573-1587.
- Song, S., Park, S.B. and Park, K. (2020), "Thematic analysis of destination images for social media engagement marketing", *Industrial Management and Data Systems*, Vol. 121 No. 6, pp. 1375-1397.
- Sota, S., Chaudhry, H. and Srivastava, M.K. (2020), "Customer relationship management research in hospitality industry: a review and classification", *Journal of Hospitality Marketing and Management*, Vol. 29 No. 1, pp. 39-64.
- Supak, S., Brothers, G., Bohnenstiehl, D.W. and Devine, H. (2015), "Geospatial analytics for federally managed tourism destinations and their demand markets", *Journal of Destination Marketing and Management*, Vol. 4 No. 3, pp. 173-186.
- Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a methodology for developing evidence-informed management knowledge by means of systematic review", *British Journal of Management*, Vol. 14 No. 3, pp. 207-222.
- Trang, L.H., Huy, T.D. and Le, A.N. (2021), "Clustering helps to improve price prediction in online booking systems", *International Journal of Web Information Systems*, Vol. 17 No. 1, pp. 45-53.
- Valls, F. and Roca, J. (2021), "Visualizing digital traces for sustainable urban management: mapping tourism activity on the virtual public space", *Sustainability*, Vol. 13 No. 6, p. 3159.
- Villamediana-Pedrosa, J.D., Vila-Lopez, N. and Küster-Boluda, I. (2019), "Secrets to design an effective message on Facebook: an application to a touristic destination based on big data analysis", *Current Issues in Tourism*, Vol. 22 No. 15, pp. 1841-1861.

- 
- Witt, S.F. and Witt, C.A. (1995), "Forecasting tourism demand: a review of empirical research", *International Journal of Forecasting*, Vol. 11 No. 3, pp. 447-475.
- Xiang, Z. (2018), "From digitization to the age of acceleration: on information technology and tourism", *Tourism Management Perspectives*, Vol. 25, pp. 147-150.
- Xie, K.L., Zhang, Z. and Zhang, Z. (2014), "The business value of online consumer reviews and management response to hotel performance", *International Journal of Hospitality Management*, Vol. 43, pp. 1-12.
- Zhu, J.J., Chang, Y.C., Ku, C.H., Li, S.Y. and Chen, C.J. (2021), "Online critical review classification in response strategy and service provider rating: algorithms from heuristic processing, sentiment analysis to deep learning", *Journal of Business Research*, Vol. 129, pp. 860-877.
- Zola, P., Cortez, P., Ragno, C. and Brentari, E. (2019), "Social media cross-source and cross-domain sentiment classification", *International Journal of Information Technology and Decision Making*, Vol. 18 No. 5, pp. 1469-1499.

**Corresponding author**

Sofía Blanco-Moreno is the corresponding author and can be contacted at: [sblanm@unileon.es](mailto:sblanm@unileon.es)

---

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

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)