Alexander Bustamante Martínez / Ernesto Amaru Galvis Lista / Mayda Patricia Gonzalez Zabala

Research Profile on Data Science in the Field of Tourism

Abstract

This research examined trends in data science application within tourism using SCOPUS for bibliographic records and VOSViewer for metadata analysis. It highlighted the top 100 keywords based on their strength. Since 2012, there's been an exponential rise in related publications. There seems to be a link between a nation's economic tourism strength and its research output. Countries like Australia, Italy, and Spain, which are known for tourism, also dominate research contributions. China has led in this domain, consistently upping its publications since 2012. The study identified some key areas: Cluster 1 emphasizes Big Data's role in enhancing tourism services; Cluster 2 explores the intricacies of human language in mining tourist reviews; Cluster 3 delves into sentiment polarity detection in texts; while Cluster 4 presents metrics for gauging destination competitiveness.

Keywords: data analysis, data science, data engineering, smart tourism, tourism trends

1. Introduction

The World Tourism Organization (UNWTO) and the World Travel & Tourism Council (WTTC) affirm tourism's profound impact on the global economy (UNWTO, 2022). Preceding the COVID-19 pandemic, tourism has maintained a positive trajectory, evidenced by its substantial contributions to Gross Domestic Product (GDP) and employment generation. However, the pandemic inflicted significant setbacks, with international tourist numbers plummeting in 2020 and 2021, leading to a marked decline in GDP contribution. This downturn disproportionately affected economies heavily reliant on tourism, such as Spain and Mexico, which comprise a significant portion of GDP. Strategic interventions, including the intensified use of Information and Communication Technologies (ICT), are being pursued to mitigate these effects and facilitate recovery. Encouragingly, the sector demonstrates signs of rebound, with international tourist arrivals improving in the first quarter of 2023 (UNWTO, 2023).

Since the pandemic, destinations have been striving to enrich ICT-based services, empowering tourists to plan, manage, and share experiences. ICT's omnipresence throughout the tourist journey yields valuable data for service providers, tourists, and sector promoters, necessitating technological tools for data processing to inform decision-making processes.

The availability of large volumes of data is not exclusive to the tourism sector. On the contrary, many sectors have seen data as a new and valuable resource that has not been previously considered. However, effectively processing and exploiting increasingly large, varied, and frequent data requires knowledge and technologies

Alexander Bustamante Martínez, Corresponding Author, Associate Professor, Universidad del Magdalena, Santa Marta, Colombia; e-mail: abustamante@unimagdalena.edu.co

Ernesto Amaru Galvis Lista, PhD, Full Professor, Universidad del Magdalena, Santa Marta, Colombia; ORCID ID: https://orcid.org/0000-0002-6012-668X; e-mail: egalvis@unimagdalena.edu.co

Mayda Patricia Gonzalez Zabala, PhD, Full Professor, Universidad del Magdalena, Santa Marta, Colombia; ORCID ID: https://orcid.org/0000-0002-6847-7919; e-mail: mpgonzalez@unimagdalena.edu.co

Review Article
Vol. 72/ No. 4/ 2024/ 673 - 686
UDC: 338.482:303.4; https://doi.org/10.37741/t.72.4.13

from computing, statistics, and mathematical thinking. This has led to the emergence of a new interdisciplinary field called data science, which has evolved rapidly.

Due to its evolution and that of related areas, such as artificial intelligence, and the maturity of the tools that facilitate these analysis tasks, a significant range of opportunities have opened for leveraging the generated data to improve productivity, competitiveness, and user experience. In this sense, within the tourism sector, there are applications of data science to identify tourists, understand their preferences, and detect their routes within destinations, among others. Table 1 shows some examples of those as mentioned above.

Table 1
Some action fields of data science in tourism

Use	Description	References
Identification of tourists through secondary data present in social networks	It consists of using spatial data and opinions from networks such as Twitter, Instagram, and Facebook to determine which users at a specific time are tourists concerning a destination and which are not.	Beltrán, 2014
Identifying tourist preferences	It involves identifying the preferences and tastes of tourists through comments and opinions on social networks.	Leonardi & Elías, 2018; Feitosa et al., 2020; Atabay & Güzeller, 2021; Özdemir & Aslı Arzık, 2022
Detecting the routes or trajectories that a tourist follows in a destination	Use data from platforms such as FourSquare, Twitter, and OpenStreetMap, among others, to identify the places a tourist visits in a destination and the routes they take during their visits.	Comito et al., 2015
Suggesting places to visit for tourists according to their preferences	Based on the preferences expressed by tourists, the characteristics of the attractions and plans of a destination, as well as the plans developed by other people, make recommendations to tourists, indicating which places they are highly likely to enjoy.	Domínguez & Araújo, 2014; Caldevilla-Domínguez et al., 2021
Automatic retrieval of data regarding tourist destination attractions	With the development of Web 2.0, users can contribute to content creation. With the evolution of the semantic web, tools can be built today that automatically retrieve data about attractions from sources like Wikipedia and OpenGeoData.	Amaya Molinar et al., 2017
Monitoring tools	Organizations of all levels (supranational, national, and private sector) build integrated systems for monitoring performance indicators in the tourism sector.	Shafique & Ali, 2016; Guerrero-Millán et al., 2020

The examples in Table 1 offer a glimpse into various data science applications within the tourism domain. While this diversity initially seems promising, it can overwhelm individuals, hindering a clear understanding, particularly for those less familiar with handling vast amounts of data. Hence, methodologies like scientometric analysis help to construct research profiles within a specific field or domain by leveraging data and metadata from scholarly documents. Although not as exhaustive as systematic literature reviews, these profiles offer insights into prevalent techniques and trends. Consequently, this article delineates a research profile spanning the last decade, examining the utilization of data science in tourism. This endeavour addresses critical inquiries such as identifying top journals, countries, and authors regarding publication volume, tracking publication trends over time, delineating research themes, and exploring the interplay between data science components and research methodologies.

2. Method

The method used to carry out this work is an adaptation and combination of elements applied in similar studies conducted in other research areas. The process generally consists of searching and obtaining bibliographic records and analyzing the metadata received from the consulted source.

The search and retrieval phase of bibliographic records was carried out using SCOPUS as the source. First, the two search expressions presented in Table 2 were iteratively constructed, consisting of phrases, logical operators, and proximity operators offered by SCOPUS. Expression 1 comprised terms associated with the field of tourism, and expression 2 with terms from the field of data science.

Table 2
Search expressions

Id	Expression
1	(*touris* OR hospitality* OR restaurant* OR vacation* OR (hotel* AND NOT hotelling) OR hostel* OR ((destination* OR attraction) W/1 (advertis* OR area OR attach* OR attractiv* OR attribut* OR beach OR beautiful OR booking OR brand* OR choice OR choos* OR coastal OR competit* OR creative OR cultural OR customer OR decision* OR develop* OR experience OR familiarity OR future OR governance OR heritage OR holiday OR image* OR "life cycle" OR lifecycle OR local OR loyalty OR manage* OR market* OR mature OR natural OR perception* OR place OR planning OR pleasure OR preference OR product OR rating OR religious OR reservation OR revisit* OR risk OR rural OR safe* OR satisfact* OR seaside OR smart OR spiritual OR sport* OR stakeholder* OR sustain* OR *touris* OR travel* OR trip OR trust OR urban OR vacation* OR value OR visit* OR wildlife)) OR ((accommodation OR holiday OR leisure OR travel*) W/1 (agen* OR beach OR booking OR business OR coastal OR company OR cost OR customer OR decision OR destination OR enterprise OR establishment* OR facility OR industry OR information OR manage* OR market* OR "peer-to-peer" OR price OR provider OR rating OR rent* OR reservation OR rural OR seaside OR sector OR selection OR sharing OR smart OR *touris* OR trust OR urban OR vacation*)))
2	((data PRE/0 (scien* OR enginee* OR mining OR analytics OR technology* OR visualization OR acquisition OR integration OR warehouse* OR privacy OR infrastructure OR preservation OR security OR architecture OR lifecycle)) OR "big data" OR ((advanced OR predictive OR business OR text) PRE/0 analytics) OR "knowledge discovery" OR "text mining" OR "business intelligence" OR OLAP)

The search to obtain basic bibliographic records, especially the electronic document identifier or EID, was performed using the functionalities provided by the SCOPUS web interface. For this, an equation was formed by combining the expressions from Table 2 as follows:

((TITLE (1) AND TITLE-ABS-KEY (2)) OR

(SRCTITLE (1) AND TITLE-ABS-KEY (2)) OR

(AUTHKEY (1) AND TITLE-ABS-KEY (2)) OR

(TITLE-ABS-KEY (1) AND SRCTITLE (2))) AND

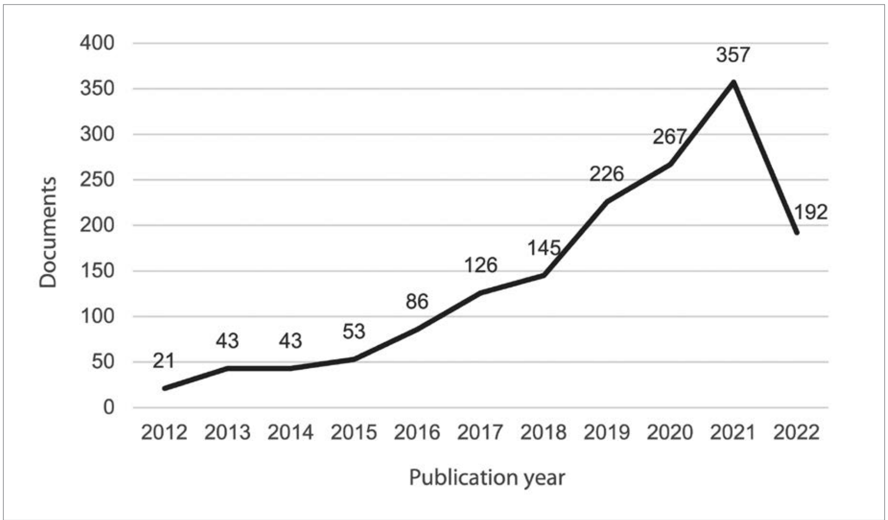
PUBYEAR > 2011 AND SRCTYPE(j)

The bibliographic records were retrieved on July 1, 2022, using the SCOPUS web interface and the Elsevier API. The complete metadata of 1559 files was obtained. Preprocessing of metadata was conducted using Python and Pandas. Data was organized into tables, and descriptive summaries were generated based on publication year, journal rankings, author affiliations, and countries. Content analysis was performed using VOSViewer software. Co-occurrence maps were created based on author-assigned keywords, with a minimum occurrence threshold of 5. This reduced the keywords from 4042 to 182, from which the top 100 by total link strength indicator were selected. VOSViewer's default normalization and layout parameters were employed, yielding a graph with 5 clusters for focused analysis.

3. Results and discussion

The search equation allowed us to identify 1559 documents between 2012 and 2022 according to the distribution shown in Figure 1. This result indicates that data science research in tourism has experienced significant growth. Specifically, the results evidenced that 21 papers were published in 2012 and 357 papers were registered in 2021, which shows that the number of publications multiplied 17 times in 10 years. For the first six months of 2022, at the time of data download, 192 works were recorded; this is 53% of documents published the previous year, which suggests that the number of works this year may be equal to or greater than those of the immediately preceding year.

Figure 1
Documents by publication year



Following our analysis, we identified that 1,559 documents were published across 533 journals worldwide. Remarkably, ten journals accounted for 27% of these publications. These journals are positioned in the first quartile according to the Scimago index, indicating a substantial number of publications and a significant impact on the tourism research community. Leading the top 10 list is the prestigious "Tourism Management" journal, while the remaining eight are specialized in the field of tourism, with two addressing broader topics. These two journals are "Sustainability," which addresses sustainability issues across various domains, and "IEEE Transactions on Knowledge and Data Engineering," which focuses on research involving data in multiple fields, including tourism. Furthermore, our analysis revealed that out of the 533 identified journals, 327 had only one document published during the study period. Table 3 provides a detailed breakdown of the top 10 journals with the most published documents.

Table 3
Top 10 sources by documents

Source	Documents
Tourism Management	76
Sustainability (Switzerland)	74
International Journal of Contemporary Hospitality Management	52
International Journal of Hospitality Management	45
Current Issues in Tourism	44
IEEE Transactions on Knowledge and Data Engineering	36
Journal of Hospitality and Tourism Technology	28
Information Technology and Tourism	28
Tourism Management Perspectives	26
Annals of Tourism Research	21

Institutional affiliation analysis varied research productivity across institutions globally, reflecting the worldwide distribution of research activity. Notably, Chinese institutions like the University of Hong Kong, with the most papers published (75); Sun Yat-Sen University, Harbin Institute of Technology, and Peking University contribute significantly, as did Portuguese universities Iscte – Instituto Universitário de Lisboa and Universidade Nova de Lisboa (Table 4).

Table 4
Top 10 institutions by documents

Affiliation	Country	Documents
Hong Kong Polytechnic University	Hong Kong	75
Iscte – Instituto Universitário de Lisboa	Portugal	37
Griffith University	Australia	24
Universidade Nova de Lisboa	Portugal	23
Sun Yat-Sen University	China	22
MODUL University Vienna	Austria	17
Harbin Institute of Technology	China	16
University of Florida	United States	16
Deakin University	Australia	16
Peking University	China	16

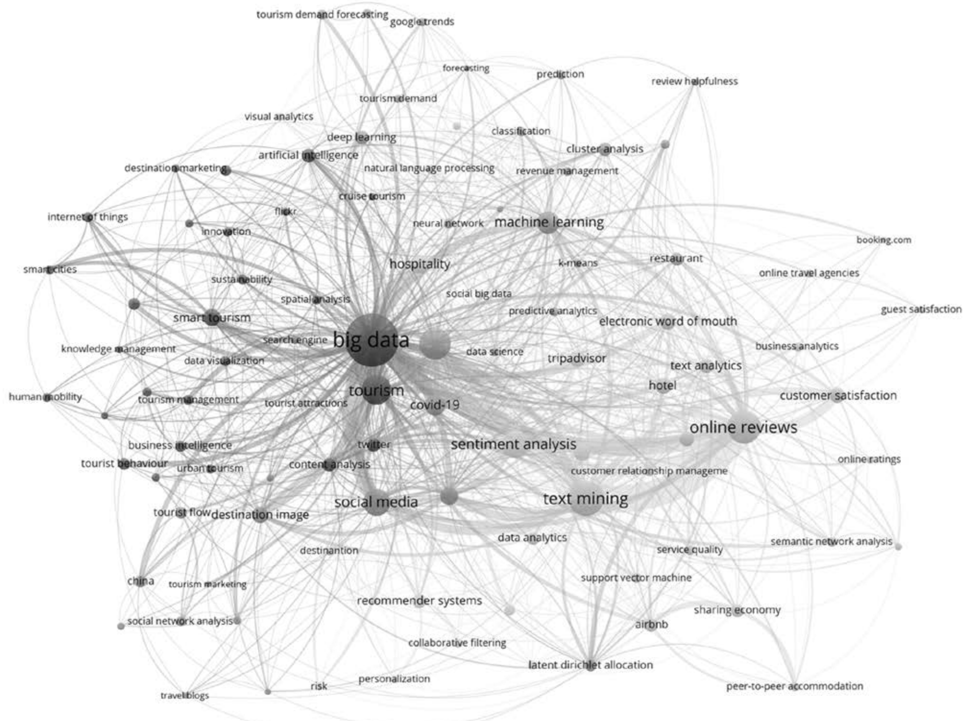
During the analysis, 68 countries were identified as contributors, with China and the United States leading the pack (Table 5). Chinese institutions accounted for 31% of all publications, reflecting China's significant role in global economics and trade. Similarly, the USA published 16% of works. Its prolific research output is likely linked to its leadership in various fields and robust tourist industry. Comparing countries' research output with their rankings in the World Economic Forum's tourism competitiveness index revealed a correlation, suggesting that nations investing in data science for tourism enhancement will likely be more competitive in the tourism sector.

Table 5
Top 10 countries with the highest number of documents per year between 2012 and 2022

País	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022*	Total documents
China	2	5	9	13	28	40	38	61	91	126	76	489
United States	6	8	9	11	22	26	25	36	44	41	28	256
United Kingdom	1	1	3	2	7	7	11	14	22	29	12	109
Spain	1	2	-	4	6	7	14	24	12	27	11	108
South Korea	1	-	3	6	3	7	4	15	18	23	18	98
Hong Kong	2	8	4	7	2	4	5	12	22	19	11	96
Australia	2	4	4	5	4	4	8	17	14	20	3	85
Italy	-	1	2	3	4	6	10	14	10	17	8	75
Taiwan	2	3	3	3	6	5	6	8	5	16	11	68
India	-	-	-	1	3	1	10	8	7	20	10	60
Portugal	-	3	-	-	1	8	8	10	10	13	7	60

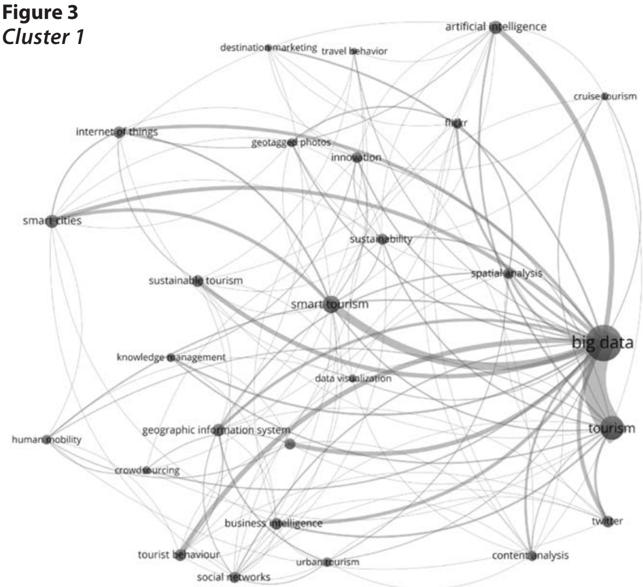
Figure 2 depicts a keyword map illustrating the evolution of data science and Web 2.0 terms from 2012 to 2022. Notably, the term "data science" turned diffuse, overshadowed by terms like "data mining," "text mining," and "Big Data," reflecting the expanded scope of data analysis techniques. "Big Data" stands out for its capacity to process vast and diverse data sources, facilitating deeper insights previously untapped. However, its assimilation of terms like business intelligence and data visualization complicates terminological clarity. Within Web 2.0, "user-generated content" emerged prominently, particularly within social networks and online reviews, providing invaluable insights into tourist preferences and behaviours. Platforms like Booking.com and Airbnb transcend their original purpose, becoming rich data sources for understanding tourist dynamics. The proliferation of user-generated content and advanced data processing techniques facilitates innovation and targeted marketing strategies, fostering intelligent tourism conducive to sustainable development across economic, environmental, and cultural dimensions.

Figure 2
Keyword map of author's published documents between 2012 and 2022



Having presented this general overview, it is time to focus on each cluster. Firstly, the cluster shown in Figure 3 reveals a close relationship between the terms Big Data and tourism, and the circle's diameter representing them confirms that they appear in the most significant number of documents. A first clarification before continuing is: what should be understood by the concept of Big Data? It should be understood as the use of data storage, processing, and visualization technologies to work with amounts of data that exceed the capacities of traditional equipment (Volume), allowing work with data of different types and formats (Variety), and that these data are generated at high frequency (Velocity).

Figure 3
Cluster 1



With a clear understanding of the Big Data concept, it becomes evident in Figure 4 that it has opened up the possibility of getting to know tourists better, providing improved services by providers, and understanding the overall dynamics of the sector by utilizing this vast amount of data, of different types, and high frequency generated today. For example, the graphic clearly shows that two widely used sources are Twitter (which provides data on location, preferences, and opinions of tourists) and Flickr (which contains photos of the experiences shared by users). Additionally, these two sources provide analysts with geotagged data that allows them to know the places tourists visit, their time at a location, and perform different types of spatial analysis.

Besides human-generated content, there are data created by machines, the interconnection between them, and the connection of all machines to the Internet, known today as the Internet of Things. This enables access to data, almost in real time, about the attractions a person visits and the vehicles they use (such as bicycles), thereby, for example, opening the possibility of better distributing the allocation of public bicycles available at a destination for visitors and locals.

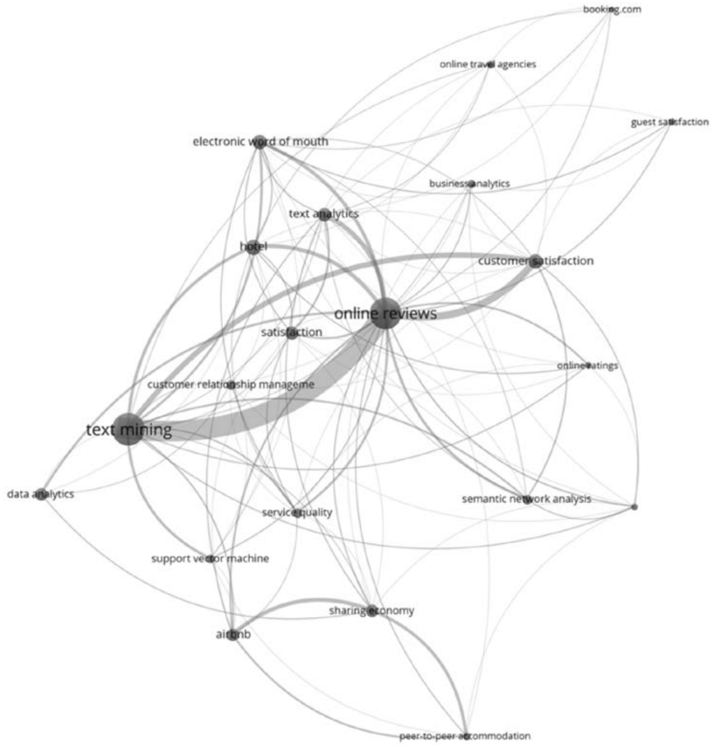
In summary, it can be confidently stated that in this cluster, data science enables intelligent tourism, allowing for more significant and more genuine interaction with tourists, as these platforms become a medium where tourists plan, experience, and share their opinions about their destination, both in general and at the level of different services, attractions, and places they visit. Next, Table 6 lists some action fields that may emerge from this cluster.

Table 6
Some action fields derived from Cluster 1

Use	Description	References
Identification of tourists through secondary data available on social media	This involves collecting spatial data and opinions from platforms such as Twitter, Instagram, and Facebook to determine which users are tourists at a specific destination and which are not at a given time.	Beltrán, 2014
Identifying tourist preferences using shared photos	This approach employs artificial vision techniques to process multimedia files, such as photos, to identify where the tourist took the picture, ultimately creating profiles of tourist preferences.	Feitosa et al., 2020
Detecting routes or trajectories followed by a tourist at a destination	By using data from platforms like FourSquare, Twitter (Malik & Kim, 2019), OpenStreetMap, and others, this method identifies the places tourists visit at a destination and the routes they take during their sightseeing trips.	Malik & Kim, 2019; Shafique & Ali, 2016
Real-time visual monitoring of tourist activity	Destinations seek to integrate data from various sources to create dashboards that allow quick insights into the sector's performance, enabling agile decision-making.	Macário de Oliveira et al., 2013; Guerrero-Millán et al., 2020

The cluster presented in Figure 4 can be explained starting from the strong relationship between the term's online reviews (which correspond to comments, opinions, and expressions made by users on platforms available on the Internet) and text mining or text analytics. In other words, their relationship is easily explainable because online reviews are a source of text mining. The figure also shows terms such as online travel agencies (OTAs) and peer-to-peer accommodation and terms referring to platform names such as Booking.com and Airbnb. These terms identify the sources of online reviews, meaning, for example, that on Airbnb, one can find reviews about person-to-person accommodation sites, and on Booking.com, hotel reviews. The texts from these sources are then used in text-mining tasks to extract useful information, such as the most frequent qualifiers tourists use for the least visited attractions in Santa Marta.

Figure 4
Cluster 2



The complexity of human language is not easily matched by algorithms' ability to automate processes. One challenging aspect is the multitude of terms human language uses to give meaning to the same thing. This complexity is transferred to the moment of text mining on tourists' online reviews, compounded by the fact that tourists may speak different languages. In this context, semantic networks and their analysis allow for the relating of terminology, the disambiguation of concepts, and the finding of synonyms and antonyms, all to improve the information obtained through analyzing texts written by tourists. Table 7 shows some examples of action fields derived from this cluster.

Table 7
Some action fields derived from Cluster 2

Use	Description	References
Classification of opinions expressed by tourists about attractions and the destination	Using data from social networks and knowledge bases, supervised or semi-supervised classification models can be created to determine the category where a tourist's comment about a specific attraction or the overall destination falls.	Amaya Molinar et al., 2017; Von Matuschka, 2021; Herrera & Reyes, 2022
Extraction of attractions visited by tourists from text written on social media	They employed entity recognition techniques to identify user comments referencing attractions at a destination and the sentiment expressed about them.	Rossetti et al., 2015; Von Matuschka, 2021
Prediction of tourist influx to destinations based on the perception expressed by tourists on social media	Developing predictive models to estimate the number of tourists visiting a destination, area, or attraction	De Lucca, 2017; Amaya Molinar et al., 2017
Categorization of tourist areas within a destination according to their image	Segmenting areas of a destination based on tourists' opinions, for example, into areas perceived as unsafe, having unpleasant odors, etc.	Rossetti et al., 2015; Su et al., 2020; Chen et al., 2022

With a similar meaning to the previous cluster is the one presented in Figure 5. In this cluster, one facet of text analysis is the detection of the polarity of sentiment expressed in a text string, ranging from identifying whether a comment is positive or negative to identifying what sentiment can be inferred from the text (anger, happiness, hilarity, distress, among others). Also present is the term "opinion mining," a more generic concept than sentiment analysis, which can be said to go a step beyond detecting feelings (although it contains them) and is used to denote the detection of positions regarding discourse, for example, determining what tourists think about a particular product or attraction, which characteristics they value the most and which they consider less important.

Figure 5
Cluster 3

The recent surge in social networks has empowered individuals, mainly tourists, to openly share their feelings, perceptions, and thoughts on various experiences, giving rise to social network analytics and Social Big Data. This large-scale data analysis is facilitated by cloud computing and leverages advancements in natural language processing (NLP) technologies, such as named entity recognition (NER) and semantic analysis, to interpret and extract meaningful information from digital communications. Additionally, the overwhelming abundance or lack of services and attractions at tourist destinations necessitates collaborative filtering and preference-based strategies to manage visitor experiences effectively. Combined with new data sources on opinions and emotions, these techniques enable tourists to make informed choices, helping tourism service providers better understand public sentiment regarding their offerings enhancing service personalization and overall offerings, as illustrated by various examples in Table 8 of this cluster.

Table 8
Some action fields derived from Cluster 3

Use	Description	References
Creation of travel planner based on tourist preferences	Developing recommendation systems to suggest activities and attractions for tourists to visit during their stay at a destination based on their preferences.	Batet et al., 2012; Valenzuela Sabogal et al., 2020
Suggesting places to visit for tourists according to their preferences	Using expressed tourist preferences, the characteristics of a destination's attractions and activities, as well as itineraries created by others, to make recommendations to tourists, indicating which places are highly likely to be appealing to them.	Hsu et al., 2009; González-Suárez et al., 2017

Social networks have not only become a medium where people express their opinions but also a space for marketing and positioning the image of tourist destinations. That is why vital tourism competitiveness indices, such as the one produced by the World Economic Forum, include indicators to measure the competitiveness of a destination, such as the number of searches made for the destination in search engines and the destination's ability to sell its services digitally. Cluster 4 focuses on these points, as shown in Figure 6.

In the figure, two approaches used in the literature to take advantage of data from social networks also stand out. One of these is correspondence analysis, which is used, for example, to visualize the relationship between variables associated with tourists and their experiences. A case where correspondence analysis can be used is to graphically represent the relationship between tourists' nationalities and their decision to visit specific attractions. Similarly, Latent Dirichlet Allocation operates, allowing the grouping of documents according to the topics they address, in this case, comments made on social networks according to their content, enabling an automatic and fast way to "know" what is being expressed on social networks about a product or destination. These two techniques are just examples of those that can be used to take advantage of the explosion of data for managing destination images and more efficiently directing marketing efforts. Table 9 shows some examples of action fields derived from this cluster.

Figure 6
Cluster 4

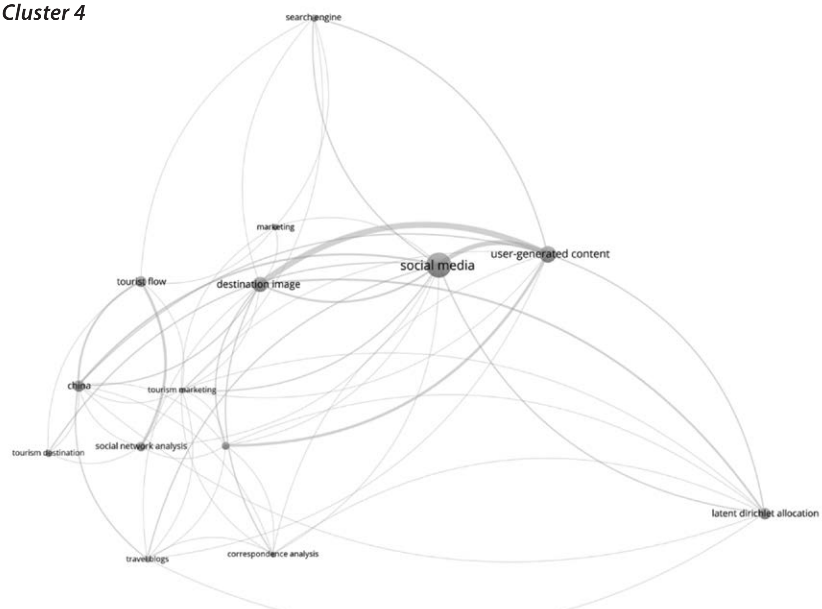
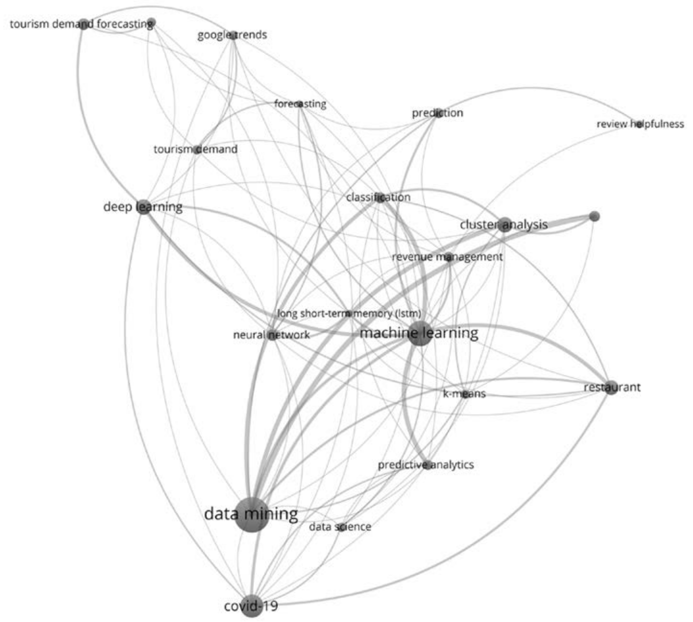


Table 9
Some action fields are derived from Cluster 4

Use	Description	References
Classification of attractions and types of activities undertaken by tourists according to their nationality	Understanding the relationship between tourists' nationality and their choice to visit specific types of attractions to personalize the tourism offerings based on the tourists' country of origin.	Ortega & Rodríguez, 2005; Kim et al., 2011
Understanding what is being expressed on social media about a product or destination	Review comments on social media, identify when a specific product or destination is being discussed, and understand the sentiment expressed, whether positive or negative.	Ordieres-Meré & Franco Riquelme, 2017; Caldevilla-Domínguez et al., 2021
Targeted marketing using user-generated content	Employing the knowledge from data collected across various social media platforms on tourists' opinions, sentiments, and expectations regarding a destination and applying it to targeted tourism marketing processes.	Hsu, 2012; Icoz et al, 2018

Following the previously mentioned thread, Cluster 5, Figure 7, focuses primarily on displaying some tasks and techniques commonly used for automatic data analysis within the tourism sector. The figure prominently features two key terms: "data mining" and "machine learning," which are often confused. It is essential to remember that while the latter is a subfield of artificial intelligence focusing on enabling machines to learn, the former, data mining, searches for patterns in data, as the name suggests. The relationship occurs because data mining techniques often automate machine learning. A more specific term, predictive analysis, focuses solely on the predictive component of data mining.

Figure 7
Cluster 5



For this reason, the figure generally presents tasks that can be achieved through data mining techniques, such as classification, clustering, and prediction, and more specifically, techniques such as neural networks (including deep learning and short-term memory) and the K-Means clustering algorithm. This set of tasks, methods, and algorithms allows operational data from tourism service providers and secondary sources like social networks, for example, to predict demand and tourist inflows so that internal logistics and investments can be organized. Table 10 mentions some action fields derived from this cluster.

Table 10
Some action fields are derived from Cluster 5

Use	Description	References
Identification of "bots"	Various identification models and natural language processing techniques detect fake accounts on social media platforms like Twitter, eliminating the noise these accounts create in data analysis.	Ojo, 2019 Sahoo & Gupta, 2019; Jabardi & Hadi, 2020
Prediction of satisfaction and demand for services offered by tourism providers	Predicting user demand and satisfaction for destinations and attractions based on online reviews from sources like TripAdvisor, Twitter, etc., using machine learning and deep learning techniques.	Chen et al., 2017 Ahani et al., 2019
Tourist segmentation	Implementing machine learning techniques, neural networks, and predictive analysis to perform tourist and market segmentation, ultimately proposing strategies based on knowledge of the sector and tourists.	Bloom, 2004; Ahani et al., 2019

4. Conclusions

In conclusion, it is essential to highlight that throughout the research conducted, it has been observed how social networks have become, over the years, a source of great importance for data analysis in different areas, especially in tourism. People's opinions, posts, and comments about the various products they consume or services they acquire, as well as the destinations and attractions they visit, are essential inputs for the analysis and improvement of tourism indicators, personalization of experiences, prediction of satisfaction, and demand for destinations and hotels, among others.

The field of tourism is witnessing a data-driven transformation, where the customization of the tourist experience has become an essential element for the competitiveness of destinations. Future research should optimize tourist personalization, using sentiment analysis derived from social media to offer experiences that resonate with cultural expectations and individual traveler preferences. Furthermore, the application of machine learning techniques to improve the quality of lodging services is proposed, increasing customer satisfaction and driving competitiveness in the global market.

Looking towards the imminent future of the tourism sector, the presence, use, and utilization of the previously mentioned techniques are fundamental parts of intelligent tourism, where the vast volume of data generated from different sources allows. They will continue to enable a much clearer vision of the current context and, in turn, make predictions and projections.

Sustainability remains a priority on the tourism development agenda. Therefore, integrating data from various sources to measure and promote environmentally responsible tourism practices is imperative. This integration should be complemented by creating a framework for smart tourism, which uses advanced analytics to enhance the experience at various tourist destinations. Likewise, the need for multifactorial predictive models that consider socio-economic and behavioral variables to forecast tourist demand is highlighted more accurately.

Finally, there is an opportunity to assess the real impact of digital marketing in tourism through the exhaustive analysis of social media data and to conduct systematic reviews that map the evolution of sentiment analysis methods in tourism literature. These reviews will provide a deeper understanding of the influence of social media on tourist decisions and guide the implementation of artificial intelligence for the customization of the tourist experience, thus paving the way for a more dynamic and customer-centric industry.

References

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management, 80*, 52-77. https://doi.org/10.1016/j.ijhm.2019.01.003

Amaya Molinar, C.M., Magaña Espinoza, P., & Ochoa Llamas, I. (2017). Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos [Evaluation of tourism destinations using data science technology]. *Estudios y Perspectivas en Turismo, 26*(2), 286-305.

Atabay, E., & Güzeller, C.O. (2021). A bibliometric study on eye-tracking research in tourism. *Tourism: An International Interdisciplinary Journal, 69*(4), 595-610. https://doi.org/10.37741/t.69.4.8

Batet, M., Moreno, A., Sánchez, D., Isern, D., & Valls, A. (2012). Turist@: Agent-based personalised recommendation of tourist activities. *Expert Systems with Applications, 39*(8), 7319-7329. https://doi.org/10.1016/j.eswa.2012.01.086

Beltrán, G. (2014). *Geomarketing: Geolocalización, redes sociales y turismo* [Geomarketing: Geolocation, social networks and tourism]. Gersón Beltrán López.

Bloom, J.Z. (2004). Tourist market segmentation with linear and non-linear techniques. *Tourism Management, 25*(6), 723-733. https://doi.org/10.1016/j.tourman.2003.07.004

Caldevilla-Domínguez, D., Barrientos-Báez, A., Pérez-García, Á., & Gallego-Jiménez, M.G. (2021). El uso de las redes sociales y su relación con la decisión de compra del turista [The use of social networks and its relationship with the tourist's purchasing decision]. *Vivat Academia. Revista de Comunicación*, 443-458. https://doi.org/10.15178/va.2021.154.e1360

Chen, J., Becken, S., & Stantic, B. (2022). Assessing destination satisfaction by social media: An innovative approach using importance-performance analysis. *Annals of Tourism Research, 93*, Article 103371. https://doi.org/10.1016/j.annals.2022.103371

Chen, N.C., Xie, W., Welsch, R.E., Larson, K., & Xie, J. (2017). Comprehensive predictions of tourists' next visit location based on call detail records using machine learning and deep learning methods. In *2017 IEEE International congress on big data (BigData congress)* (pp. 1-6). IEEE. https://doi.org/10.1109/BigDataCongress.2017.10

Comito, C., Falcone, D., & Talia, D. (2015). Mining popular travel routes from social network geo-tagged data. In E. Damiani, R. Howlett, L. Jain, L Gallo, & G. De Pietro (Eds.), *Intelligent interactive multimedia systems and services. Smart innovation, systems and technologies* (Vol. 40, pp. 81-95). Springer. https://doi.org/10.1007/978-3-319-19830-9_8

De Lucca, M. (2017). *Data mining y turismo: Un enfoque integral para potenciar la imagen de marca Argentina* [Data mining and tourism: A comprehensive approach to enhance the Argentine brand image, Degree thesis]. Universidad Nacional del Sur. Departamento de Geografía y Turismo. https://repositoriodigital.uns.edu.ar/handle/123456789/4035

Domínguez, T., & Araújo, N. (2014). Gestión de las redes sociales turísticas en la web 2.0 [Management of tourist social networks in web 2.0.]. *Vivat Academia. Revista de Comunicación*, 57-78. https://doi.org/10.15178/va.2014.129.57-78

Estevão, C., Garcia, A., Filipe, S., & Fernandes, C. (2017). Convergence in tourism management research: A bibliometric analysis. *Tourism & Management Studies, 13*, 30-42. https://doi.org/10.18089/tms.2017.13404

Feitosa, W.R., Tolentino, D., Pedroso, M., & Gianocário, G. (2020). El uso de las redes sociales para posicionar un destino turístico: El caso de la ciudad de Guararema y sus visitantes del día [The use of social networks to position a tourist destination: The case of the city of Guararema and its visitors of the day]. *Turismo: Visão e Ação, 21*, 398-422. https://doi.org/10.14210/rtva.v21n3.p398-422

González-Suárez, G., Delgado-Fernández, T., Capote-Fernández, J.L., & Cruz-Iglesias, R. (2017). Método para sistemas de recomendaciones sensibles al contexto en entornos turísticos [Method for context-sensitive recommendation systems in tourism environments]. *Ingeniería Industrial, 38*(1), 68-80.

Guerrero-Millán, M., Villagómez-Méndez, J., & Herrera-Miranda, M. A. (2020). Sistema de indicadores para la gestión del turismo sostenible en el destino turístico de Acapulco, Guerrero, (México) [System of indicators for the management of sustainable tourism in the tourist destination of Acapulco, Guerrero, (Mexico)]. *Ciencias Administrativas. Teoría y Praxis, 16*(2), Article 2. https://doi.org/10.46443/catyp.v16i2.262

Herrera, N., & Reyes, N. (2022). Evaluación de sitios turísticos mediante análisis de sentimientos de comentarios emitidos por usuarios en redes sociales [Evaluation of tourist sites through sentiment analysis of comments issued by users on social networks]. *Revista Tecnológica - ESPOL, 34*, 124-138. https://doi.org/10.37815/rte.v34n2.921

Hsu, T.-K., Tsai, Y.-F., & Wu, H.-H. (2009). The preference analysis for tourist choice of destination: A case study of Taiwan. *Tourism Management, 30*, 288-297. https://doi.org/10.1016/j.tourman.2008.07.011

Hsu, Y.-L. (2012). Facebook is an international e-marketing strategy for Taiwanese hotels. *International Journal of Hospitality Management, 31*(3), 972-980. https://doi.org/10.1016/j.ijhm.2011.11.005

Icoz, O., Kütük, A., & Icoz, O. (2018). Social media and consumer buying decisions in tourism: The case of Turkey. *PASOS. Revista de Turismo y Patrimonio Cultural, 16*, 1051-1066. https://doi.org/10.25145/j.pasos.2018.16.073

Jabardi, M., & Hadi, A. (2020). Twitter fake account detection and classification using ontological engineering and semantic web rule language. *Karbala International Journal of Modern Science, 6*(4), Article 8. https://doi.org/10.33640/2405-609X.2285

Jun, S.-P., Yoo, H.S., & Choi, S. (2018). Ten years of research change using Google trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change, 130*, 69-87. https://doi.org/10.1016/j.techfore.2017.11.009

Kim, S.S., Timothy, D.J., & Hwang, J. (2011). Understanding Japanese tourists' shopping preferences using the decision tree analysis method. *Tourism Management, 32*(3), 544-554. https://doi.org/10.1016/j.tourman.2010.04.008

Leonardi, V.S., & Elías, S.R. (2018). Redes sociales en Internet en el marco del turismo cultural: Los museos en Bahía Blanca (Argentina) [Social networks on the Internet in the framework of cultural tourism: Museums in Bahía Blanca (Argentina)]. *Redes. Revista hispana para el análisis de redes sociales, 29*(1), Article 1.

Macário de Oliveira, V., Pasa Gómez, C., & Ataíde Cândido, G. (2013). Indicadadores de sustentabilidad para la actividad turística: Una propuesta de monitoreo usando criterios de análisis [Sustainability indicators for tourism activity: A monitoring proposal using analysis criteria]. *Estudios y Perspectivas en Turismo, 22*(2), 177-197.

Malik, S., & Kim, D. (2019). Optimal travel route recommendation mechanism based on neural networks and particle swarm optimization for efficient tourism using tourist vehicular data. *Sustainability, 11*(12), Article 12. https://doi.org/10.3390/su11123357

Ojo, A.K. (2019). Improved model for detecting fake profiles in online social network: A case study of Twitter. *Journal of Advances in Mathematics and Computer Science, 33*(4), Article JAMCS.50575. https://doi.org/10.9734/jamcs/2019/v33i430187

Ordieres-Meré, J., & Franco Riquelme, J.N. (2017). Análisis de opinión en redes sociales y valor para el análisis de producto [Opinion analysis in social networks and value for product analysis]. *Revista Economía Industrial, 405*, Article 405.

Ortega, E., & Rodríguez, B. (2005). Importancia de la comunicación en los destinos turísticos internacionales: La percepción de los turistas extranjeros en los destinos españoles [Importance of communication in international tourist destinations: The perception of foreign tourists in Spanish destinations]. *Cuadernos de Turismo, 15*, 169-188. https://revistas.um.es/turismo/article/view/18451

Özdemir, G.Ö., & Aslı Arzık, V. (2022). Segmentation of social media users for destinations: A clustering approach. *Tourism: An International Interdisciplinary Journal, 70*(1), 53-66. https://doi.org/10.37741/t.70.1.4

Rossetti, M., Stella, F., Cao, L., & Zanker, M. (2015). Analysing user reviews in tourism with topic models. In I. Tussyadiah & A. Inversini (Eds.), *Information and communication technologies in tourism 2015* (pp. 47-58). Springer International Publishing. https://doi.org/10.1007/978-3-319-14343-9_4

Sahoo, S.R., & Gupta, B.B. (2019). Hybrid approach for detection of malicious profiles on Twitter. *Computers & Electrical Engineering, 76*, 65-81. https://doi.org/10.1016/j.compeleceng.2019.03.003

Shafique, S., & Ali, M.E. (2016). Recommending most popular travel path within a region of interest from historical trajectory data. In *Proceedings of the 5th ACM SIGSPATIAL international workshop on mobile geographic information systems* (pp. 2-11). Association for Computing Machinery. https://doi.org/10.1145/3004725.3004728

Su, L., Lian, Q., & Huang, Y. (2020). How do tourists' attribution of destination social responsibility motives impact trust and intention to visit? The moderating role of destination reputation. *Tourism Management, 77*, Article 103970. https://doi.org/10.1016/j.tourman.2019.103970

Szpilko, D. (2017). Tourism supply chain – Overview of selected literature. *Procedia Engineering, 182*, 687-693. https://doi.org/10.1016/j.proeng.2017.03.180

Valenzuela Sabogal, G.M., Sosa Cortes, J.S., & Torres Rojas, J.D. (2020). Sistema de recomendación turístico mediante aplicación de minería de datos en la provincia del Gualivá, Cundinamarca [Tourism recommendation system through data mining application in the province of Gualivá, Cundinamarca]. *Revista ONTARE, 8*, 71-87. https://doi.org/10.21158/23823399.v8.n0.2020.2644

Von Matuschka, C. (2021). Análisis de sentimientos usando la red social Twitter ¿qué sintieron los turistas que volaron en 2020 con seleccionadas aerolíneas sudamericanas? [Sentiment analysis using the social network Twitter, what did tourists who flew in 2020 with selected South American airlines feel?]| *Revista de Turismo e Identidad, 2*(1), 55-71. https://revistas.uncu.edu.ar/ojs/index.php/turismoeindentidad/article/view/4991

VOSviewer. (2023). *VOSviewer — Visualizing scientific landscapes*. https://www.vosviewer.com/

World Tourism Organization. (2022, January). *UN report underscores importance of tourism for economic recovery in 2022*. https://www.unwto.org/news/un-report-underscores-importance-of-tourism-for-economic-recovery-in-2022

World Tourism Organization. (2023). World Tourism Barometer and Statistical Annex, May 2023. *UNWTO World Tourism Barometer, 21*(2). https://doi.org/10.18111/wtobarometereng.2023.21.1.2

Submitted: October 08, 2023
Revised: April 18, 2024
Accepted: May 14, 2024