INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

14, Issue 12, (2024) E-ISSN: 2222-6990

Satisfaction and User Reviews in Tourism Using Big Data and Text Mining: A Bibliometric Study

Prahardika Prihananto, Lissa Rosdiana Noer, Sri Yayu Ninglasari, and Ni Gusti Made Rai

Departement of Business Management, Faculty of Creative Design and Digital Business, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Corresponding Author Email: prahardika@its.ac.id

To Link this Article: http://dx.doi.org/10.6007/IJARBSS/v14-i12/24181 DOI:10.6007/IJARBSS/v14-i12/24181

Published Date: 17 December 2024

Abstract

The integration of technology, particularly social media and user-generated content platforms, has shaped tourist behavior and decision-making processes. Tourists increasingly rely on online reviews or electronic word of mouth (eWOM) as key determinants in destination perception and selection. Sentiment analysis using big data and text mining techniques is conducted to understand and harness the power of eWOM in shaping tourist experiences. Bibliometric analysis is used in research to identify potential citation and co-citation patterns, facilitating exploration and explanation of key research content in specific fields. This study conducts bibliometric analysis on a collection of Scopus articles related to satisfaction and user reviews in the field of tourism using big data and text mining techniques. The results from analyzing 425 articles show that Asia is the most prominent region for this field of research. Citation analysis resulted in 10 clusters, whereas bibliometric coupling resulted in 6 clusters, with 1 cluster having newly emerging topics in this field of research, namely experience and cultural tourism. Keywords in this field of research are separated into 5 clusters, with "sustainable tourism," "topic modeling," and "hotel industry" being the newly emerging keywords in this field of research.

Keywords: Tourism, Satisfaction, User Reviews, Big Data, Text mining, Bibliometric

Introduction

In recent decades, the importance of tourism for global economies have steadily risen (Comerio & Strozzi, 2019). United Nation General Assembly acknowledges that the tourism sector has the capability to drive 'economic growth, social inclusion, and cultural and environmental (*UNWTO 2016 Annual Report*, 2017). The travel and tourism sector also gained considerable global significance in the context of a growing focus on service-oriented industries with the rise of sector's direct contribution to global GDP from 9.9% in 1995 to 10.3% by 2019 (Škare et al., 2021). However, the tourism sector has been hit hardest during

the pandemic (Bhatia et al., 2022). Tourism began to recover after the pandemic subsided in 2022. Figure 1 shows that global tourism revenue and the number of tourists increased sharply in 2023. In fact, the number of tourists and tourism revenue in 2023 have already surpassed those of 2019, before the pandemic occurred. Although the sector is recovering, the pandemic has introduced new trends in tourism (Morwinsky, 2023).

Tourism has become highly dependent on the digitalization and advancement of technology (Akhtar et al., 2021). The survival of the sector leveraged by modern advancements in technology and information technology (IT). Tourism is inseparable from the use of apps and social media. TripAdvisor sees more than 390 million visitors on its platform every month, with 435 million reviews and opinions encompassing nearly 7 million hotels, restaurants, and tourist sites globally (Tripadvisor, 2021). These statistics highlight how social media has the power to connect with every segment of the tourist population. Tourist behavior is now shaped by social media (Leelawat et al., 2022). Travelers turn to the internet to share their reviews and opinions. They use technology to gather information as a guide before selecting which travel destinations to visit (El-Said, 2020). Tourists need experiential information as a reference point because tourism products and services can't be experienced firsthand by tourists until they use them (Tran and Strutton, 2020). The availability of experiential information in the form of user reviews are accessible to a wide audience can either boost or damage the reputation of tourist destination (Alaei et al., 2019). User reviews are known as electronic Word of Mouth (*eWOM*) (Nilashi et al., 2022). Currently, 90% of tourists say they consider eWOM before deciding about traveling (Akhtar et al., 2019). eWOM is positively correlated with tourists' trust in a destination and their intent to travel (Abubakar, 2016). eWOM consists of a large and diverse set of data dimensions (Cheung & Thadani, 2012). The amount of textual data from eWOM continues to grow each year as more people use social media. The challenge of managing extensive textual data can be solved by using big data dan text mining techniques (Yan et al., 2022). Text mining involves various methods to explore large quantities of text, either automatically or semi-automatically, to identify trends, recurring patterns, and other insights that explain text behavior (Alzate et al., 2022). Using text mining to extract eWOM is useful for discovering hidden meanings in textual data, which is especially valuable in sentiment analysis (Bordoloi & Biswas, 2023). Processing eWOM data, such as sentiment analysis, is a crucial aspect of recent research in tourism (Alaei et al., 2019). Studies on sentiment analysis related to tourism topics have become increasingly popular, providing both empirical and theoretical contributions.

On the other hand, technological advancements have led to a new approach to literature review known as bibliometric analysis. Bibliometric analysis is a thorough examination of the characteristics of published literature, along with potential citation and co-citation patterns, which facilitates the exploration and clarification of key research content to advance a specific field (Ferreira et al., 2014). Bibliometric analysis can be used to study knowledge trends within any research area to uncover objective and hidden patterns (Hao et al., 2021; Pattnaik et al., 2020). The rise of user reviews in the tourism sector, coupled with the growing popularity of text mining and big data methods driven by the rapid growth of online user reviews, and the greater use of bibliometric analysis, has made bibliometric study that examines user reviews in tourism by using text mining and big data is fascinating to do.

Literature Review

Tourism

Tourism is defined as the temporary movement of people to destinations outside their usual place of residence, encompassing the activities they engage in during their stay, and the facilities created to meet their needs (Mtapuri & Giampiccoli, 2019). There are three core elements of tourism: the journey, the activities, and the infrastructure that supports them. It generates revenue through direct spending on accommodations, transportation, food, and entertainment, and has a ripple effect on local economies (Tribe, 2011). Tourism contributed about 10.4% of global GDP and supported over 300 million jobs worldwide (WWTC, 2019). In addition, tourism can also have a positive impact on maintaining the cultural integrity of local communities and preserving the natural environment. Discussing tourism always involves tourists. Tourists are individuals who travel to different regions primarily for vacation, relaxation, or recreation (George, 2021). They do not intend to settle at the destination (Crawley & Hagen-Zanker, 2019). Tourists are those who travel for recreational purposes and stay temporarily at the destination. Moreover, tourists are categorized into two groups: international tourists and domestic tourists.

Customer Satisfaction, User Review, and eWOM

Consumer satisfaction is the sense of fulfilment that arises when people evaluate the products or services that they receive against their initial expectations (Joudeh & Dandis, 2018). In tourism, measuring tourist satisfaction involves comparing their experiences to what they expected in terms of service quality, the performance of the products or services, and their perceptions of various destination attributes (Biswas et al., 2021). If what they receive falls short of their expectations, tourists feel dissatisfied; on the other hand, they are satisfied if the actual experience meets or exceeds their expectations. Understanding tourists' levels of satisfaction and gathering their opinions is vital for fostering the development of tourism regions. Meeting visitors' expectations is essential, not just for providing a positive experience but also for building a strong reputation and boosting the attractiveness of a tourism destination (Alaei et al., 2019). Tourist satisfaction has significant implications for tourism management because it can drive customer loyalty and lead to positive word-of-mouth marketing (Preko et al., 2021). Electronic Word of Mouth (eWOM) refers to user-generated experiences from previous purchases, along with opinions on the quality of services or products (Nilashi et al., 2022). EWOM can take the form of online reviews, recommendations, or other online opinions, making it a crucial component of marketing. The advantage of eWOM is that it spreads information more quickly and spontaneously than traditional word-of-mouth methods (Manes & Tchetchik, 2018). These days, 90% of consumers admit that they consider Electronic Word of Mouth (eWOM) before making a purchasing decision (Akhtar et al., 2019).

Big Data and Text Mining

The increasing use of the internet has led to a significant rise in the number of text-based documents online. A wide array of data and documents, including reviews, books, electronic messages, and other forms, has generated vast amounts of data, known as big data. This surge in data necessitates tools for processing it effectively and accurately. Technological advances have also paved the way for new methods of big data analysis, such as text mining. Text mining is a type of data mining that aims to discover interesting patterns within large collections of textual data (Berger et al., 2020). It involves a range of methods designed to

explore unstructured data sources and uncover potentially valuable structures, models, patterns, or principles within textual data (Nielbo et al., 2024; Yan et al., 2022). Text mining, or text extraction, is the process of converting unstructured text data into a structured format for further data mining (Jain et al., 2021). Extracting useful information and hidden knowledge from textual data is crucial for subsequent processing (Bordoloi and Biswas, 2023). Sentiment analysis is an application of text mining. It is a method for assessing public opinion, specifically whether the sentiment toward a product or service is favourable or unfavourable (Amira & Irawan, 2020). The goal of sentiment analysis is to determine the overall sentiment within a text document, review, opinion, or emotion, with classifications that may be positive, neutral, or negative (Alaei et al., 2019). Sentiment analysis is widely used today to address various research questions and provide new insights in tourism studies (Leelawat et al., 2022). The use of big data and text mining is very important in today's tourism, especially for tourist satisfaction. Online travel agency Travelocity leverages big data for pricing, inventory management, and personalized recommendations by analyzing vast amounts of data to offer the best deals and improve customer satisfaction (Chen, 2020). Jiangsu Province applies big data for smart tourism systems in cities like Nanjing and Zhenjiang by providing intelligent information and enhancing the tourist experience (Zhang et al., 2021). Platforms like Qunar, Ctrip, and Fliggy utilize big data to offer personalized travel services, reflecting the shift towards individualized tourism by analyzing user data and providing customized travel recommendations (Zhang et al., 2021). On the other hand, by using text mining, organizations can analyze online reviews and identify trending topics to craft targeted marketing campaigns and improve specific areas that need enhancement (Xu & Lv, 2022).

Bibliometric Analysis

Bibliometric analysis involves using mathematical and statistical approaches to study a book or a collection of articles (Liu et al., 2019). It has become a scientific method for examining the temporal evolution of research fields from a multidisciplinary viewpoint (Bhatt et al., 2020). This analysis helps to understand the boundaries of a research field, pinpoint influential authors, and explore new directions for future research (Donthu et al., 2020). Recently, bibliometric analysis has been used by experts to evaluate and predict trends in various research fields (Leung et al., 2019). In other words, bibliometric analysis is valuable for deciphering and mapping cumulative scientific knowledge and understanding the progression of literature among the public by analyzing large volumes of data from articles. Common methods used in bibliometric analysis include citation analysis and bibliometric coupling. Citation analysis is a fundamental science mapping technique based on the assumption that if one publication cites another, it indicates intellectual relevance between them (Mora et al., 2019). This method assesses a publication's impact by the number of citations it receives. Consequently, by tracking citations, one can determine the most influential publications in a given research field. Bibliometric coupling, on the other hand, measures the similarity or connection between two documents based on shared references (Rojas-Lamorena et al., 2022). If two documents cite some of the same references, they are considered bibliometrically coupled (Bernatović et al., 2022). The more references two documents share, the stronger their bibliometric coupling (Bernatović et al., 2022). While bibliometric coupling is less popular than citation analysis, it complements the latter by providing additional insights (Donthu et al., 2021). Bibliometric coupling has proven to be effective in identifying shifts in research topics (Andersen, 2021).

Research Methodology

This research begins by defining keywords for database searches. The keyword search technique allows researchers to selectively retrieve specific files, as database content is organized based on the frequency of keyword occurrence in documents (Meharwade & Patil, 2016). To construct the keyword query, Boolean operators such as "AND" and "OR" are utilized. "AND" is used to narrow the search by ensuring that results contain all specified criteria, thereby excluding documents that only partially meet the criteria. Conversely, "OR" broadens the search by including synonyms or similar phrases, ensuring that at least one of the specified terms appears in the output. This research analyzes satisfaction based on user reviews in the tourism industry. Therefore, the keywords used are "satisfaction," "review," and "tourism." "Satisfaction" and "review" are interchangeable because analyzing satisfaction involves using user reviews. Hence, "satisfaction" OR "review" can be used as keywords. Since tourism issues inherently involve tourist destinations, the keyword "destination" is also employed in this study. "Tourism" and "destination" are considered equivalent, so the keyword used is "tourism" OR "destination." Additionally, this research examines the application of big data and text mining approaches in assessing tourist satisfaction within the tourism industry. Therefore, the keywords "big data" and "text mining" are included, and "big data" OR "text mining" can be used as keywords. Consequently, all search keywords for existing articles are combined into: ("satisfaction" OR "review") AND ("tourism" OR "destination") AND ("big data" OR "text mining"). After establishing these keywords, the researchers selected a database for conducting the search, specifically utilizing the Scopus database.

Data collection is performed based on predefined keywords. Following the search using these keywords, the data is filtered by setting specific criteria to refine the search results. The filtering process includes selecting articles written in English for easier interpretation. Additionally, articles without author names are excluded from the analysis. This involves narrowing the search to specific subject areas to select the most relevant fields for the research. Document type and language are also employed as additional filtering criteria. The articles selected by the system during the initial filtering phase undergo manual review. This entails individually reading each article, examining titles, abstracts, and conclusions to ensure alignment with the research focus. This manual filtering process resulted in a final dataset of 425 articles for use in the analysis stage.

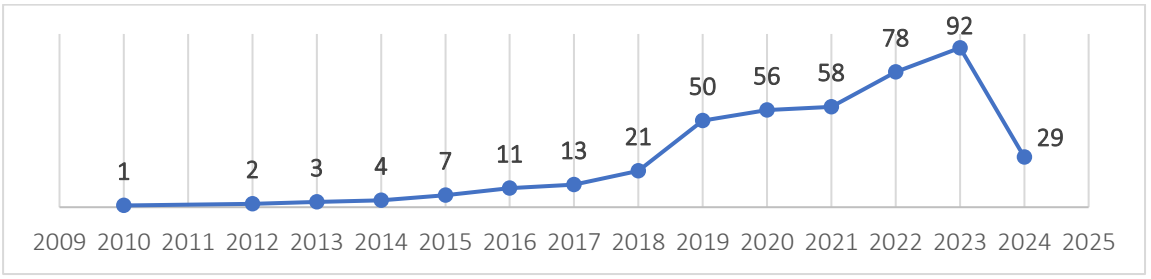
The analysis stage will be conducted by summarizing the data that was categorized in the previous step. The results will be displayed using statistical diagrams that illustrate the distribution of publications by year, country, and publisher. In addition, bibliometric analysis methods such as citation analysis, bibliometric coupling, and keyword analysis will be applied to extract further insights from the final dataset.

Result And Discussion

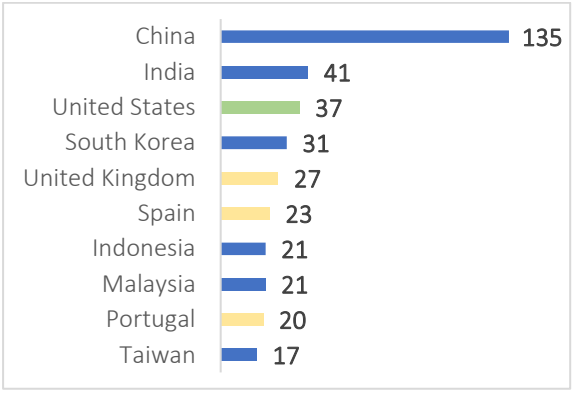
Descriptive Analysis

Figure 1 illustrates the development of scholarly publications on tourism satisfaction and reviews that use big data and text mining methods, based on a total of 425 articles. Publications in this field started to emerge in 2010 and began to grow in 2015. The highest increase occurred in 2019, with a 138% rise, while the second-highest increase happened in 2022, with a 34.4% rise. This suggests that research in this field entered a new phase in 2019,

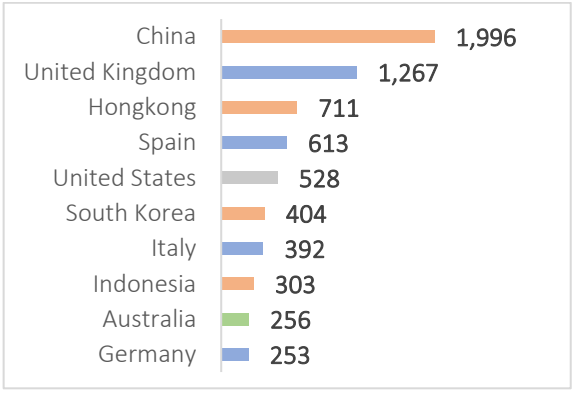
with significant growth in 2022 indicating the beginning of yet another phase. Top 10 countries contribute 87.8% of all articles out of 65 countries. China comes first with 31.8%, followed by India with 9.6%, and the United States with 8.7%. Although the United States ranks third and some European countries are in the top 10, this research is primarily popular in the Asian region. The country with the most citations is China, with 1,996, followed by the United Kingdom with 1,267, and Hong Kong with 711. Those result suggesting that this research area is quite popular in Asias region and warrants further investigation.



(a)



(b)



(c)

Figure 1. (a) Number of Publications 2010-2024 (b) Top 10 Country by Documents Number (c) Top 10 Country by Citations Number

Table 1
Top 10 Publisher by Number of Documents

Publisher	Documents
Sustainability (Switzerland)	18
International Journal of Contemporary Hospitality Management	16
Tourism Management	10
Current Issues in Tourism	9
Advances in Intelligent Systems and Computing	8
Information Technology and Tourism	6
Asia Pacific Journal of Tourism Research	5
Journal of Hospitality and Tourism Technology	5
Tourism Recreation Research	5
Annals of Tourism Research	4

An analysis was also conducted based on the publishers of the articles. Table 1 shows the top 10 publishers by the number of documents. The results indicate that the publisher Sustainability (Switzerland) released 18 documents, followed by the International Journal of

Contemporary Hospitality Management with 16 articles, Tourism Management with 10 documents, and Current Issues in Tourism with 9 articles. The results from Table 1 also indicate that there are several publication outlets available to accommodate research findings in this field. Table 2 represents the top 10 publishers by the number of citations. From this table, we can identify the publisher with the highest citation count. Tourism Management leads with 1,231 citations, followed by the International Journal of Contemporary Hospitality Management in second place with 448 citations. The third rank goes to Current Issues in Tourism, which has 403 citations. In fourth place is Sustainability (Switzerland) with 381 citations, and the Asia Pacific Journal of Tourism Research ranks fifth with 300 citations. The data suggests that publishers focusing on tourism and hospitality still dominate topics related to tourism, satisfaction, user reviews (eWOM), big data, and text mining.

Table 2
Top 10 Publisher by Number of Citations

Publisher	**Citations**
Tourism Management	1,231
International Journal of Contemporary Hospitality Management	448
Current Issues in Tourism	403
Sustainability (Switzerland)	381
Asia Pacific Journal of Tourism Research	300
Journal of Destination Marketing and Management	256
Annals of Tourism Research	172
Information Technology and Tourism	144
International Journal of Environmental Research and Public Health	93
Journal of Hospitality and Tourism Technology	86

Citation Analysis

Table 3 identifies the top 10 most cited documents. The top three documents related to the topics of tourism, satisfaction, user reviews (eWOM), big data, and text mining are Guo, et al. (2017), in the first rank, with a citation count of 658; Marine-Roig and Clavé (2015), in the second rank, with 255 citations; and Li, et al. (2013), in the third rank, with 235 citations.

Table 3
Top 10 Citated Documents

Authors	**Citations**
Guo, et al. (2017)	658
Marine-roig & Clavé (2015)	255
Li, et al. (2013)	235
Liu, et al. (2017)	208
Suhartanto, et al (2020a)	157
Jia (2020)	134
Marine-roig (2019)	92
Park, et al (2020)	79
Centobelli & Ndou (2019)	76
Radojevic, et al (2017)	70

In Figure 2, you can observe the Citation Network of Articles, which illustrates the citation relationships among various authors. From the citation network, we can see that the articles are grouped into 10 clusters: the blue cluster, light brown cluster, pink cluster, greenish-blue cluster, orange cluster, yellow cluster, light red cluster, green cluster, red cluster, and purple cluster. Only 8 clusters have articles within the top 10 citations; the red and purple clusters do not contain articles within the top 10 citations. The study by Guo et al. (2017) belongs to the Blue Cluster and is the most cited. This research identifies the main dimensions of customer service expressed by hotel visitors using a data mining approach, specifically Latent Dirichlet Allocation (LDA). The dataset comprises 266,544 online reviews from 25,670 hotels located in 16 countries. Through LDA, 19 controllable dimensions were identified as crucial for hotels to manage their interactions with visitors. In the Light Brown Cluster, the study by Marine-Roig and Clavé (2015) aims to assess the efficacy of big data analysis in supporting smart destinations by scrutinizing the online image of Barcelona—a prominent smart city and tourist destination— communicated through social media. The analysis involved examining more than 100,000 travel blogs and online travel reviews (OTRs) in English authored by tourists who visited the city over the past decade. The findings suggest that user-generated content (UGC) data are beneficial for developing and evaluating marketing strategies, as well as for enhancing branding and positioning strategies within tourism and marketing organizations.

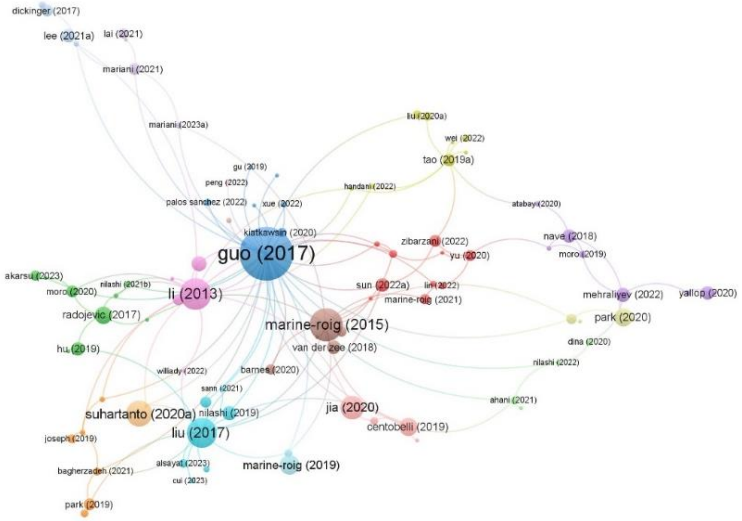


Figure 2. Citation Network of Articles

The research by Li et al. (2013), part of the Pink Cluster, demonstrates that analyzing online reviews can identify the factors influencing customer satisfaction in accommodation venues. The research employed text mining and content analysis on 42,668 traveler reviews from 774 star-rated hotels. The findings indicate that customers booking both luxury and budget hotels consider factors like transportation convenience, food and beverage management, proximity to tourist destinations, and budget are important. High performance in these areas can lead to customer satisfaction. However, customers pay more attention to, and are often less satisfied with, bed comfort, reception services, and room size and decoration. Most determinants of customer satisfaction are consistent between luxury and budget hotels, except for aspects related to lobbies and sound insulation. Additionally, the research by Liu et al. (2017) and Marine-Roig (2019) is included in the Greenish-Blue Cluster. The study by Liu et al. (2017) leverages user-generated reviews to gain new perspectives on

the factors influencing hotel customer satisfaction, specifically by segmenting customers based on language groups. By analyzing 412,784 reviews generated by users on TripAdvisor for 10,149 hotels in five cities in China, the researchers discovered that foreign tourists speaking various languages (such as English, German, French, Italian, Portuguese, Spanish, Japanese, and Russian) significantly vary in their prioritization of hotel attributes ("Rooms," "Location," "Cleanliness," "Service," and "Value") when assessing overall satisfaction with hotels. Domestic Chinese tourists exhibit distinct preferences for room-related hotel attributes compared to foreign tourists. The study reveals significant interaction effects between attributes such as "Rooms" and "Service," as well as "Value" and "Service." Meanwhile, Marine-Roig's (2019) research aims to demonstrate how online travel reviews (OTRs) can be valuable for analyzing the image or perception of a tourist destination. The theoretical and methodological framework for this research includes metrics that allow for the measurement of various aspects of the tourist image, such as descriptive, appreciative, and prescriptive elements. The research model is implemented in the Attica region of Greece using a random sample of 300,000 TripAdvisor reviews written in English between 2013 and 2018, covering attractions, activities, restaurants, and hotels. The findings provide insights into trends, preferences, assessments, and opinions from the perspective of travelers, offering valuable information for destination managers to optimize resource allocation and promote sustainability initiatives.

In the Orange Cluster, the research by Suhartanto et al. (2020) examines tourist loyalty in creative tourism, with a focus on experience quality, perceived value, satisfaction, and motivation as antecedents. It also assesses the mediating role of motivation in the development of tourist loyalty. The findings suggest that experience quality, perceived value, tourist satisfaction, and tourist motivation play pivotal roles in influencing tourist loyalty toward creative attractions. In the Yellow Cluster, the research by Park et al. (2020) explores hotel customer revisiting behavior using large-scale customer review data. This study provides insights into factors associated with customer revisiting behavior and the potential to attract more customers to reuse hotel services, as well as to predict future customer revisiting behavior. The findings from Park et al. (2020) indicate that feedback reviews from returning visitors tend to contain more words in a sentence and express more extreme sentiments—whether positive or negative—compared to reviews from one-time visitors.

Two articles in the Light Red Cluster rank among the top 10 in citations: Jia (2020) and Centobelli and Ndou (2019). The research by Jia (2020) aims to compare the motivations and satisfaction levels of restaurant tourists from China and the United States by analyzing their online ratings and reviews. Data were gathered from two prominent online review platforms, where customer ratings and reviews were collected, quantitatively analyzed, subjected to text-mining techniques, and interpreted using statistical methods, Latent Dirichlet Allocation, and frequency analysis. The findings indicate that Chinese tourists are less prone to giving low ratings to restaurants and emphasize food quality, while US tourists tend to prioritize enjoyment and are less concerned about crowdedness (Jia, 2020). Meanwhile, the research by Centobelli and Ndou (2019) provides a systematic literature review of the use of big data in tourism, identifying current issues and suggesting future research directions. The findings indicate that there is a growing number of contributions on this topic, certain areas still require further exploration and development. In the Green Cluster, the study by Radojevic et al. (2017) aims to conduct a multilevel analysis to investigate the factors influencing customer

satisfaction in the global hotel industry. The findings suggest that hotel attributes and visitors' personal characteristics are among the most significant factors influencing customer satisfaction. However, the evaluation of hotels is also significantly influenced by the purpose of the trip, the characteristics of the destination, and the nationality of the visitor.

Each cluster has a distinct thematic focus based on the articles it encompasses. The blue cluster focuses on various dimensions and characteristics of tourist services, while the light brown cluster discusses the impact of user-generated content on destination image. The pink cluster examines factors contributing to tourist satisfaction, such as support services and complementary attributes, and the greenish-blue cluster analyzes destination attributes and their perceived images. The orange cluster pertains to experiential tourism, delving into the nature of tourist experiences. The yellow cluster concentrates on factors that affect tourists' intentions to revisit, whereas the red cluster is related to cultural tourism. Lastly, the green cluster focuses on the characteristics of tourist destinations and their visitors.

We know that the original research themes come from the blue cluster, and most clusters have close relationships with it. The pink, green, and yellow clusters have a similar focus to the blue cluster. This is supported by the citation network, showing that these four clusters have close relationships. The greenish-blue cluster is closely associated with both the blue and the light-brown clusters in the citation network. The theme for the greenish-blue cluster is a blend between the blue cluster and the light-brown cluster. We know that the light-brown and red clusters have a close position to the blue cluster; however, their themes are different from that of the blue cluster. This shows the evolution of research themes for those clusters. The orange cluster has a distant relationship with the blue cluster in terms of both position and discussed theme. From this explanation, it is evident that research in this field is starting to evolve, although it still has a relationship with the original theme. Additionally, there is an emergence of completely new themes for this field of research. From this, we can understand that themes such as experience tourism, cultural tourism, and destination image are interesting to explore in future research.

Bibliometric Coupling

Figure 3 illustrates the Bibliometric Coupling of Articles, which was employed to detect changes in research themes. Six clusters are derived from this bibliometric coupling: the blue cluster, green cluster, purple cluster, yellow cluster, greenish blue cluster, and red cluster. Only four clusters contain the top 10 most cited articles; the greenish blue cluster and red clusters do not. Research by Guo et al. (2017), Li et al. (2013), Liu et al. (2017), Radojevic et al. (2017), and Park et al. (2020) are included in the blue cluster. Meanwhile, articles from Marine-Roig and Clavé (2015) and Centobelli and Ndou (2019) are categorized under the green cluster. The purple cluster comprises articles from Suhartanto et al. (2020) and Jia (2020), while Marine-Roig (2019) research is situated in the yellow cluster. Overall, two primary clusters are identified in this research field: the blue cluster and green cluster, with the former slightly more prevalent than the latter. The purple and yellow clusters may represent newly emerging topics in the field, as they differ from the blue and green clusters but still feature articles within the top 10 citations. However, while the greenish blue cluster and red clusters also deviate from the previous four clusters, they cannot yet be classified as new emerging topics.

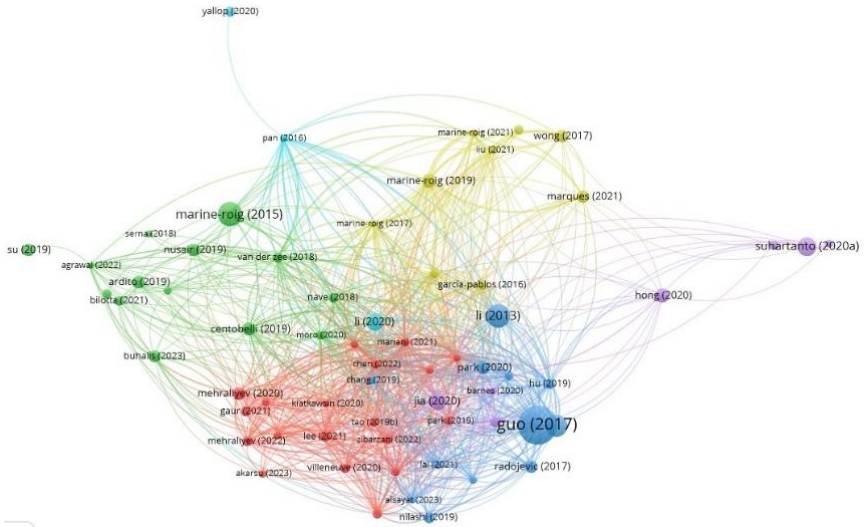


Figure 3. Bibliometric Coupling of Articles

The blue cluster mainly discusses the dimensions of services and the factors, attributes, and characteristics of services that affect tourist satisfaction and revisit intention. The green cluster focuses on user-generated content and destination images. The yellow cluster mainly discusses destination image. Although it is separated from the green cluster, the topic still has similarities with the green cluster. Because of these similarities, the yellow cluster cannot be said to represent a new emerging topic in this field of research. The purple cluster focuses on experience tourism and cultural tourism. These topics differ from those of the blue, green, and yellow clusters. Consequently, research by Suhartanto et al. (2020) and Jia (2020) may encapsulate the newly emerging topics in this research field, although the topics from the original clusters continue to exert significant influence. Suhartanto et al. (2020) explores the factors influencing tourist loyalty in creative tourism, finding that the quality of experiences that connect tourists with the local culture and community has the most impact on tourist loyalty. Jia (2020) studies cross-cultural behavior in tourists, which can assist tourism managers in enhancing customer satisfaction and increasing business revenue. From this explanation, it can be understood that research related to experience and cultural tourism is interesting to pursue in the future.

Keywords Analysis

Figure 4 illustrates the co-occurrence network of keywords, categorized by group and year. When examined by group, it is apparent that frequently used keywords are divided into 5 clusters. The green cluster comprises keywords such as "sentiment analysis," "tourist satisfaction," "tourist experience," and "tourist attraction." The red cluster includes keywords like "smart tourism," "sustainable tourism," and "tourism management." The yellow cluster consists of "destination image," "service quality," "user-generated content," and "topic modeling." The purple cluster encompasses "online reviews" and "social media." The blue cluster contains keywords related to "e-WoM," the "hotel industry," and "hospitality."

We can observe that the yellow cluster focuses on experiences and tourist destinations visited by tourists. The red cluster focuses on tourism management and the development of the tourism industry. The yellow cluster mainly discusses destination image and quality. The purple cluster is more related to reviews on social media. The blue cluster is focused on the hotel industry and hospitality.

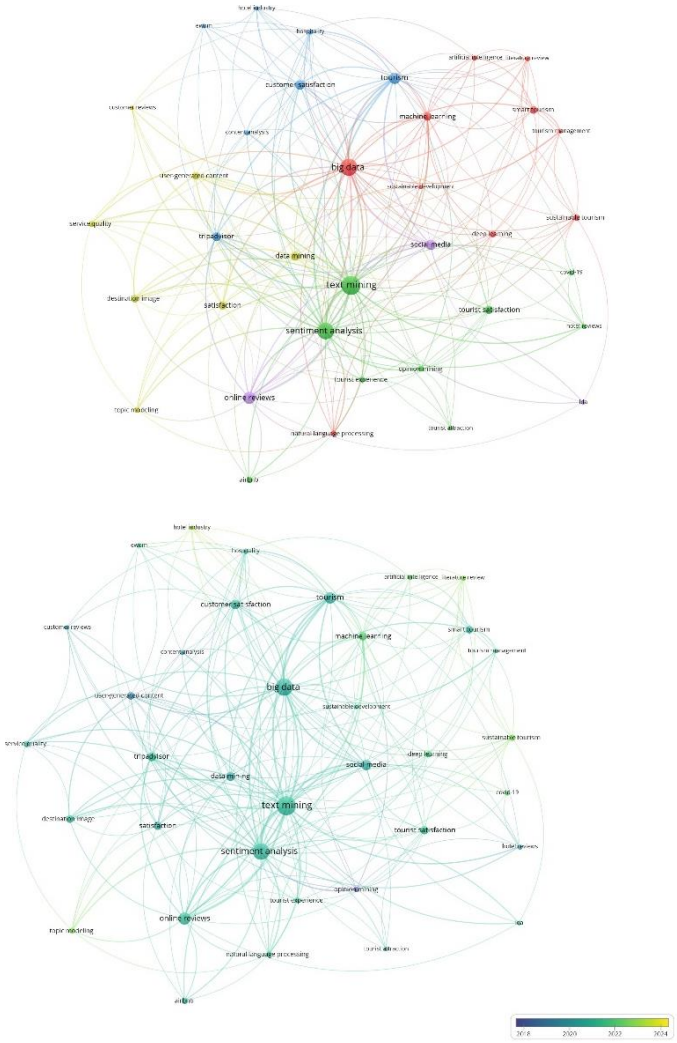


Figure 4. Co-Occurrence Network of Keywords: a) by group b) by year

Analyzing the co-occurrence network of keywords based on year reveals that "sustainable tourism," "topic modeling," and "hotel industry" are among the new keywords that have emerged in recent years. Topic modeling is a method to uncover abstract topics within a collection of textual documents (Habbat & Nouri, 2024). This keyword has become a trend due to the vast amount of review data in the tourism sector, making topic modeling popular for research in tourism that utilizes text mining and big data. The hotel industry has emerged in recent years due to the increasing number of user reviews for hotels and platforms that enable user-generated content about hotels. Sustainable tourism has emerged as a focus on the sustainability of the environment and tourist destinations in recent years. This keyword is related to the results of citation analysis and bibliographic coupling that identified the newly emerging topics in this field of research as experience and cultural marketing. These two topics are closely related to sustainable tourism as they benefit from preserving cultural heritage, conserving natural resources, and respecting local communities. Based on these findings, it is recommended to conduct research using a topic modeling approach related to the hotel industry or sustainable tourism in the future.

Conclusions

This study delves into the evolution of tourism, particularly within the framework of technological advancements. The integration of technology, notably through platforms like social media and user-generated content sites such as TripAdvisor, has significantly influenced tourist behavior and decision-making processes. Travelers increasingly rely on online reviews and opinions to select their destinations, making electronic Word-of-Mouth (eWOM) a critical factor in destination perception and selection. Consequently, sentiment analysis, facilitated by big data and text mining techniques, has become an essential tool for comprehending and leveraging the impact of eWOM on shaping tourist experiences. It is noted that research analyzing tourism satisfaction and reviews using big data and text mining techniques has entered a new phase since its inception. While this field is predominantly led by the Asian Region, contributions from the United States and European countries are also noteworthy, with China leading in both the number of documents and citations. Publishers focusing on tourism and hospitality continue to drive discussions on topics related to satisfaction and user reviews in tourism using big data and text mining methods. Citation analysis reveals the formation of ten clusters, with eight of them containing the top ten cited articles in the field. Among these, four clusters share close relationships based on topic similarities, three have distinct topics, and one exhibits a combination of vastly different topics. Bibliometric coupling identifies six clusters, with two original research clusters and potential emerging trends in this research field. However, only one cluster truly represents new emerging topics, namely experience and cultural tourism. Hence, future research on experience and cultural tourism holds significant promise. Co-occurrence network analysis delineates keywords into five clusters, with "sustainable tourism," "topic modeling," and "hotel industry" emerging as recent keywords. This underscores the importance of pursuing research that utilizes topic modeling related to hotels and sustainable tourism.

This study aims to explore and analyze previous research utilizing Big Data and Text Mining to assess tourist satisfaction from user reviews related to the tourism industry and destinations. However, this study does not specifically explore technical aspects of Big Data and Text Mining, such as data privacy, sentiment analysis accuracy, and system integration, as it focused on satisfaction and user reviews in the tourism industry. However, given the abundance of existing studies utilizing Big Data and Text Mining, further research that analyzes the research focused on technical aspects of Big Data and Text Mining is recommended.

Acknowledgments

This research is funded by the Department of Business Management Research Grant, Institut Teknologi Sepuluh Nopember.

References

2020 Annual Report. (2021). *Massachusetts.* Retrieved from https://ir.tripadvisor.com/static-files/fef1a79b-0b14-40b3-ae35-da7ee030aca4

Akhtar, Nadeem, Khan, N., Mahroof Khan, M., Ashraf, S., Hashmi, M. S., Khan, M. M., & Hishan, S. S. (2021). Post-covid 19 tourism: Will digital tourism replace mass tourism? *Sustainability*, *13*(10). https://doi.org/10.3390/su13105352

Akhtar, Naeem, Sun, J., Akhtar, M. N., & Chen, J. (2019). How attitude ambivalence from conflicting online hotel reviews affects consumers' behavioural responses: The

moderating role of dialecticism☆. *Journal of Hospitality and Tourism Management*, *41*, 28–40. https://doi.org/10.1016/j.jhtm.2019.09.003

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: Capitalizing on Big Data. *Journal of Travel Research*, *58*(2), 175–191. https://doi.org/10.1177/0047287517747753

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. *Journal of Retailing and Consumer Services*, *67*. https://doi.org/10.1016/j.jretconser.2022.102989

Amira, S. A., & Irawan, M. I. (2020). Opinion analysis of traveler based on tourism site review using sentiment analysis. *IPTEK The Journal for Technology and Science*, *31*(2), 223. https://doi.org/10.12962/j20882033.v31i2.6338

Andersen, N. (2021). Mapping the expatriate literature: A bibliometric review of the field from 1998 to 2017 and identification of current research fronts. *International Journal of Human Resource Management*, *32*(22), 4687–4724. https://doi.org/10.1080/09585192.2019.1661267

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25. https://doi.org/10.1177/0022242919873106

Bernatović, I., Slavec Gomezel, A., & Černe, M. (2022). Mapping the knowledge-hiding field and its future prospects: A bibliometric co-citation, co-word, and coupling analysis. *Knowledge Management Research and Practice*, *20*(3), 394–409. https://doi.org/10.1080/14778238.2021.1945963

Bhatia, A., Roy, B., & Kumar, A. (2022). A review of tourism sustainability in the era of Covid-19. *Journal of Statistics and Management Systems*, *25*(8), 1871–1888. https://doi.org/10.1080/09720510.2021.1995196

Bhatt, Y., Ghuman, K., & Dhir, A. (2020). Sustainable manufacturing. Bibliometrics and content analysis. *Journal of Cleaner Production*, *260*, 120988. https://doi.org/10.1016/j.jclepro.2020.120988

Biswas, C., Deb, S. K., Hasan, A. A.-T., & Khandakar, Md. S. A. (2021). Mediating effect of tourists' emotional involvement on the relationship between destination attributes and tourist satisfaction. *Journal of Hospitality and Tourism Insights*, *4*(4), 490–510. https://doi.org/10.1108/JHTI-05-2020-0075

Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *56*(11), 12505–12560. https://doi.org/10.1007/s10462-023-10442-2

Centobelli, P., & Ndou, V. (2019). Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism*, *22*(15), 1862–1882. https://doi.org/10.1080/13683500.2018.1564739

Chen, N. (2020). Application of big data technology in smart tourism. *Journal of Physics: Conference Series*, *1648*(4), 1648–1655. IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/1648/4/042101

Cheung, C. M. K., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems*, *54*(1), 461–470. https://doi.org/10.1016/j.dss.2012.06.008

Comerio, N., & Strozzi, F. (2019). Tourism and its economic impact: A literature review using bibliometric tools. *Tourism Economics*, *25*(1), 109–131. https://doi.org/10.1177/1354816618793762

Crawley, H., & Hagen-Zanker, J. (2019). Deciding where to go: Policies, people and perceptions shaping destination preferences. *International Migration*, *57*(1), 20–35. https://doi.org/10.1111/imig.12537

Donthu, N., Kumar, S., & Pattnaik, D. (2020). Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of Business Research*, *109*, 1–14. https://doi.org/10.1016/j.jbusres.2019.10.039

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

El-Said, O. A. (2020). Impact of online reviews on hotel booking intention: The moderating role of brand image, star category, and price. *Tourism Management Perspectives*, *33*, 100604. https://doi.org/10.1016/j.tmp.2019.100604

Ferreira, M. P., Santos, J. C., de Almeida, M. I. R., & Reis, N. R. (2014). Mergers & acquisitions research: A bibliometric study of top strategy and international business journals, 1980-2010. *Journal of Business Research*, *67*(12), 2550–2558. https://doi.org/10.1016/j.jbusres.2014.03.015

George, R. (2021). *Marketing tourism and hospitality: Concepts and cases*. Springer Nature.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Habbat, N., & Nouri, H. (2024). Unlocking travel narratives: A fusion of stacking ensemble deep learning and neural topic modeling for enhanced tourism comment analysis. *Social Network Analysis and Mining*, *14*(82), 1–24. https://doi.org/10.1007/s13278-024-01256-3

Hao, A. W., Paul, J., Trott, S., Guo, C., & Wu, H. H. (2021). Two decades of research on nation branding: A review and future research agenda. *International Marketing Review*, *38*(1), 46–69. https://doi.org/10.1108/IMR-01-2019-0028

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, *41*, 100413. https://doi.org/10.1016/j.cosrev.2021.100413

Jia, S. (2020). Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management*, *78*, 104071. https://doi.org/10.1016/j.tourman.2019.104071

Joudeh, J. M. M., & Dandis, A. O. (2018). Service quality, customer satisfaction and loyalty in an internet service providers. *International Journal of Business and Management*, *13*(8), 108–120. https://doi.org/10.5539/ijbm.v13n8p108

Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, *8*(10), e10894. https://doi.org/10.1016/j.heliyon.2022.e10894

Leung, A., Burke, M., Cui, J., & Perl, A. (2019). Fuel price changes and their impacts on urban transport–a literature review using bibliometric and content analysis techniques, 1972–2017. *Transport Reviews*, *39*(4), 463–484. https://doi.org/10.1080/01441647.2018.1523252

Li, H., Ye, Q., & Law, R. (2013). Determinants of customer satisfaction in the hotel industry: An Application of online review analysis. *Asia Pacific Journal of Tourism Research*, *18*(7), 784–802. https://doi.org/10.1080/10941665.2012.708351

Liu, W., Wang, J., Li, C., Chen, B., & Sun, Y. (2019). Using bibliometric analysis to understand the recent progress in agroecosystem services research. *Ecological Economics*, *156*, 293–305. https://doi.org/10.1016/j.ecolecon.2018.09.001

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, *59*, 554–563. https://doi.org/10.1016/j.tourman.2016.08.012

Manes, E., & Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: An empirical investigation of online hotel booking. *Journal of Business Research*, *85*, 185–196. https://doi.org/10.1016/j.jbusres.2017.12.019

Marine-Roig, E. (2019). Destination image analytics through traveller-generated content. *Sustainability*, *11*(12), 3392. https://doi.org/10.3390/su11123392

Marine-Roig, E., & Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing and Management*, *4*(3), 162–172. https://doi.org/10.1016/j.jdmm.2015.06.004

Meharwade, A., & Patil, G. A. (2016). Efficient keyword search over encrypted cloud data. *Physics Procedia*, *78*, 139–145. Elsevier B.V. https://doi.org/10.1016/j.procs.2016.02.023

Mohammed Abubakar, A. (2016). Does eWOM influence destination trust and travel intention: A medical tourism perspective. *Economic Research-Ekonomska Istrazivanja* , *29*(1), 598–611. https://doi.org/10.1080/1331677X.2016.1189841

Mora, L., Deakin, M., & Reid, A. (2019). Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities. *Technological Forecasting and Social Change*, *142*, 56–69. https://doi.org/10.1016/j.techfore.2018.07.019

Morwinsky, S. (2023). *Industries & markets - travel & tourism: Market data & analysis - market insights report*. New York. Retrieved from https://www.statista.com/study/40460/travel-tourism/

Mtapuri, O., & Giampiccoli, A. (2019). Tourism, community-based tourism and ecotourism: A definitional problematic. *South African Geographical Journal*, *101*(1), 22–35. https://doi.org/10.1080/03736245.2018.1522598

Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., & Tahmasebi, N. (2024). Quantitative text analysis. *Nature Reviews Methods Primers*, *4*(1), 1–16. https://doi.org/10.1038/s43586-024-00302-w

Nilashi, M., Ali Abumalloh, R., Alrizq, M., Alghamdi, A., Samad, S., Almulihi, A., … Mohd, S. (2022). What is the impact of eWOM in social network sites on travel decision-making during the COVID-19 outbreak? A two-stage methodology. *Telematics and Informatics*, *69*, 101795. https://doi.org/10.1016/j.tele.2022.101795

Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Current Issues in Tourism*, *23*(5), 605–611. https://doi.org/10.1080/13683500.2018.1549025

Pattnaik, D., Hassan, M. K., Kumar, S., & Paul, J. (2020). Trade credit research before and after the global financial crisis of 2008 – A bibliometric overview. *Research in International Business and Finance*, *54*, 101287. https://doi.org/10.1016/j.ribaf.2020.101287

Preko, A., Mohammed, I., Gyepi-Garbrah, T. F., & Allaberganov, A. (2021). Islamic tourism: Travel motivations, satisfaction and word of mouth, Ghana. *Journal of Islamic Marketing*, *12*(1), 124–144. https://doi.org/10.1108/JIMA-04-2019-0082

Radojevic, T., Stanisic, N., & Stanic, N. (2017). Inside the rating scores: A multilevel analysis of the factors influencing customer satisfaction in the hotel industry. *Cornell Hospitality Quarterly*, *58*(2), 134–164. https://doi.org/10.1177/1938965516686114

Rojas-Lamorena, Á. J., Del Barrio-García, S., & Alcántara-Pilar, J. M. (2022). A review of three decades of academic research on brand equity: A bibliometric approach using co-word analysis and bibliographic coupling. *Journal of Business Research*, *139*, 1067–1083. https://doi.org/10.1016/j.jbusres.2021.10.025

Škare, M., Soriano, D. R., & Porada-Rochoń, M. (2021). Impact of COVID-19 on the travel and tourism industry. *Technological Forecasting and Social Change*, *163*, 120469. https://doi.org/10.1016/j.techfore.2020.120469

Suhartanto, D., Brien, A., Primiana, I., Wibisono, N., & Triyuni, N. N. (2020). Tourist loyalty in creative tourism: The role of experience quality, value, satisfaction, and motivation. *Current Issues in Tourism*, *23*(7), 867–879. https://doi.org/10.1080/13683500.2019.1568400

Tran, G. A., & Strutton, D. (2020). Comparing email and SNS users: Investigating e-servicescape, customer reviews, trust, loyalty and E-WOM. *Journal of Retailing and Consumer Services*, *53*, 101782. https://doi.org/10.1016/j.jretconser.2019.03.009

Travel & Tourism Economic Impact 2019. (2019). Retrieved from https://www.un.org/development/desa/en/news/population/2018-revision-of-worldurbanization-prospects.html

Tribe, J. (2011). *The economics of recreation, leisure and tourism* (4th ed.). Oxford: Elsevier.

UNWTO 2016 Annual Report. (2017). Madrid. https://doi.org/https://doi.org/10.18111/9789284418725

Xu, H., & Lv, Y. (2022). Mining and application of tourism online review text based on natural language processing and text classification technology. *Wireless Communications and Mobile Computing*, *2022*(1), 1–13. https://doi.org/10.1155/2022/9905114

Yan, H., Ma, M., Wu, Y., Fan, H., & Dong, C. (2022). Overview and analysis of the text mining applications in the construction industry. *Heliyon*, *8*(12), e12088. https://doi.org/10.1016/j.heliyon.2022.e12088

Zhang, H., Guo, T., & Su, X. (2021). Application of big data technology in the impact of tourism e-commerce on tourism planning. *Complexity*, *2021*(1), 1–10. https://doi.org/10.1155/2021/9925260