

## **Big data in action: An overview of big data studies in tourism and hospitality literature**

### **ABSTRACT**

Tourism research has marched into the big data arena and brought remarkable developments. Despite the promising role of big data and increasing volume of research, it is worth noting that current big data research in the tourism and hospitality field is rather vague or insufficient, particularly from the perspectives of its philosophical grounding, methodological approaches, and implications. This article aims to provide a comprehensive review of big data research from the tourism and hospitality literature. A content analysis of 146 big-data-related articles identifies the research and methodological trends in these fields. The findings reveal that big data have expanded the scope of tourism research. It offers useful and practical knowledge for destination, hotel, revenue, and reputation management. However, big data research is rather limited in methodologies, and there is a need for more solid theoretical and philosophical footings for significant knowledge generation.

*Keywords:* big data, content analysis, philosophy grounding, methodological approaches, research focus

## 1. Introduction

In today's high-tech world, a considerable volume and variety of data are offered and consumed on a routine basis. Development in data science has led to the formation of the concept of big data and has created a wealth of innovative research, unveiling the role of big data analytics in practical knowledge generation (Kambatla, Kollias, Kumar, & Grama, 2014). The potential value of big data analytics has been captured by the tourism and hospitality industries. They have built up the capability to turn big data into valuable practical knowledge and insights, for instance, to improve tourism and hospitality management and practices, which are considered significant competitive advantages (Liu, Teichert, Rossi, Li, & Hu, 2017; Xiang, Schwartz, Gerdes Jr, & Uysal, 2015).

As a data technology revolution, big data will have an omni bearing impact on knowledge discovery, mode of thinking, and the values and ethics of tourism and hospitality research. On the one hand, big data, involving enormous, exhaustive, and varied datasets, is beyond the ability of a conventional database tool to capture, store, manage, and analyse (Elragal & Klischewski, 2017). On the other hand, big data rely more on a new (inductive) statistical approach, which may engender a paradigm shift from theory-driven to data-driven in how we think about research (Kitchin, 2014b, Resnyansky, 2019). Furthermore, the research on big data analytics is evolving very fast: new approaches, methodologies, and algorithms applied to examine tourism and hospitality themes are continuously emerging (He, Deng, Li, & Gu, 2021; Ma, Kirilenko, & Stepchenkova, 2020). The complexity of analytical methods also increases the demand for expertise in interpreting the results appropriately and extracting valuable information (Kitchin, 2014b).

To date, only a few systematic literature reviews have been conducted on big data research in the tourism and hospitality industry, which underscores the importance of the issue under discussion. For instance, Mariani, Baggio, Fuchs, & Höepken (2018) undertook a review of big data and business intelligence by taking 77 big data articles in the 2000–2016 period, and Li, Xu, Tang, Wang, and Li (2018) reviewed big data research in tourism in terms of research focus, data characteristics, and analytic techniques during 2007–2017.

Overall, existing review studies of big data research in tourism and hospitality have paid little attention to the philosophical grounding at the levels of ontology, epistemology, and ethics. It is necessary to critically examine the understanding of big data philosophy and explore to what extent big data changes the research philosophy and theoretical expositions in tourism studies. Additionally, due to the rapid development of novel big data analytical methods in tourism and hospitality research in the past five years, an updated analysis of big data methods in terms of data type, data storage, data processing, and data analysis is deemed timely. In addition, there is a lack of synthesis regarding the practical implications of existing big data research for destination management, image building, climate change and welfare, hotel performance, etc. To address the above deficiencies, the overall goal of this study is to review the state of big data research in the scientific literature to better understand the ways it has been and can be used to inform tourism and hospitality management and practices. In this review, we address the following main questions pertaining to big data research in tourism and hospitality management:

1. What are the publication trends (growth of publications, journal categories, country of authorship, country of focus, sector of research) of big data research in tourism and hospitality?
2. What are the discussions and trends of big data research philosophy in tourism and hospitality studies?
3. What are the big data trends in terms of data, methodological approaches, and research themes in tourism and hospitality?
4. What are the practical implications of big data research in tourism and hospitality management?

Essentially, this research contributes to the literature in several ways: (1) it provides critical insights on tourism and hospitality scholars' understanding of big data philosophy, (2) it develops a framework that links data type, methods, and research focus, (3) it reveals the role of big data research in guiding tourism management practices, and (4) it offers several paths for future research.

## 2. Methodology

This study deployed content analysis to extract useful information from the body of literature in tourism and hospitality. The general goal of content analysis is to identify and record relatively objective features of messages in a set of documents (Neuendorf, 2002). It is a useful tool to review prior research, as it identifies, assesses, and creates high-quality research evidence as well as discovers future research potential, directions, and challenges (Hung & Lu, 2016). Using content analysis, researchers can analyse the existence, meaning, and relationships between words, concepts, and themes and draw inferences. It has been widely used in analysing tourism and hospitality big data. For example, Lu and Stepchenkova (2015) conducted an analysis of user-generated content of 122 articles, in terms of topics, methods, and software. They found that big data enhanced the reliability of the findings. Mariani, Di Felice, and Mura (2016) conducted a content analysis of big data and business intelligence in tourism, analysing 173 articles to identify research gaps and future research development agendas.

To find relevant research articles, we generated a list of keywords referring to either big data or tourism/hospitality. These keywords emerged from an in-depth analysis of the literature and interviews with subject experts. Then, we combined keywords pertaining to the two categories to generate the final search string: [(“Big Data” AND “Touris\*”) OR (“Big Data” AND “Hospitality”) OR (“Big Data” AND “Hotel”) OR (“Data analytics” AND “Touris\*”) OR (“Data analytics” AND “Hospitality”) OR (“Data analytics” AND “Hotel”) OR (“User-generated content” AND “Touris\*”) OR (“User-generated content” AND “Hospitality”) OR (“User-generated content” AND “Hotel”)]. The definition of this search string means that articles should contain at least one of the proposed combinations of keywords (\* denotes a wildcard) in their abstracts, titles, or keywords.

As big data research is very new to tourism and hospitality academics, we collected all the relevant literature published up until 31 March 2020, without specifying the start date. A Scopus search was performed to retrieve the data associated with the search terms. Scopus is considered one of the most comprehensive academic research repositories across disciplines (Vieira & Gomes, 2009). The Scopus search generated a

list of the relevant tourism and hospitality contributions, focusing on various subject areas. The search terms generated 1,643 hits consisting of journal articles, conference papers, book chapters, conference reviews, reviews, and short surveys. The initial screening followed the criteria in Table 1. Figure 1 shows the sequential steps of the article screening process adopted in this study.

Eventually, we obtained a final sample of 146 big data articles. The title and abstract of each article were then rechecked individually for relevance. We created a manual database in MS Excel containing the title, year of publication, focused industry, sample size, data format, data source, and targeted country/region. The selected articles were fully read to carefully evaluate each article's main findings, analytical methods, and implications. After reading each article, the findings, analytical methods applied for data analysis, and practical implications were manually entered into the MS-Word datasheet (see Appendix A).

<Insert Table 1 about here>

<Insert Figure 1 about here>

### **3. Results**

#### *3.1. Publication trends*

##### *3.1.1 Growth of publications and journal categories*

A general upward trend has been observed in the annual growth of big data publications in the tourism and hospitality sectors, as shown in Figure 2. The high annual growth rate indicates that big data has been receiving a high level of academic attention in the field of tourism and hospitality, particularly since 2014.

<Insert Figure 2 about here>

The high number of titles of journals indicates that many discipline and multidisciplinary journals have paid attention to big data research in tourism and hospitality. As shown in Figure 3, Tourism Management

published the highest number of articles, followed by the International Journal of Hospitality Management, Current Issues in Tourism, International Journal of Contemporary Hospitality Management, and the Journal of Travel Research.

<Insert Figure 3 about here>

### *3.1.2 Country of authorship/Country of focus*

Authors from 49 countries are involved in this type of research (Figure 4). The country of authorship refers to the affiliation of the first author. The findings suggest that the top five countries are China (125 articles), the U.S. (94), Spain (60), Italy (42), and the UK (38). The findings indicate that big data research has taken a greater hold in China, Europe, and the U.S.; apart from China, authors from the developing world have paid little attention to this type of research.

The overall trend suggests that scholars have been highly focused on East Asia, North America, and Europe (Figure 4). The developments in big data research show that a high number of researchers have focused on China's data, followed by the U.S. In addition, some tourism scholars have researched broader global issues.

<Insert Figure 4 about here>

### *3.1.3 Sector of research*

Figure 4 also reveals that most of the articles are focused on tourism-related issues, followed by the hotel and travel sector. The tourism sector is complex and offers researchers several areas of investigation, e.g., tourists' behaviour, satisfaction, mobility, tourists' experience, and social and commercial influence on residential communities. In addition, some big data research has specifically examined various sub-fields of tourism, e.g., culture tourism, special interest tourism, ecotourism, and dark tourism.

### *3.2. Philosophy of big data*

This sub-section aims to offer a critical review of the current discussion of the philosophy of big data science and reflect on its application in the tourism and hospitality sector. Big data philosophy encompasses three main domains: ontology, epistemology, ethics and aesthetics (Warburton, 1999). Based on this literature review, no article in the tourism and hospitality discipline focuses explicitly on the philosophy of science (ontology, epistemology, and ethics) in big data research. Some articles discuss or mention “ontology”, nevertheless, as a computer science terminology about data representation, collection and classification within a certain framework, such as in Fuchs, Höpken, and Lexhagen (2014) and Moreno, Valls, Isern, Marin, and Borràs (2013). The absence of discussion on what big data really means for tourism and hospitality research has also been identified in the earlier systematic literature review by Mariani et al. (2018). They labeled this gap strongly as terra incognita, where big data research faces epistemological dilemmas without efforts to theorize big data-driven findings. Have tourism scholars in big data research successfully addressed this criticism after Mariani et al. (2018)? “It remains insufficiently addressed” is a generous conclusion. This review finds limited progress in addressing this gap in the tourism literature. Despite potential practical implications, some studies hardly discussed, applied, or reflected on the philosophy of science in big data research.

Our analysis reveals that tourism big data research have ontological challenges due to the shift in analytical methods and the use of different scales to measure social phenomena (Mariani, 2019). Two critical ontological issues emerge from the literature: first, the relationships between various human life aspects and what big data can tell us about it. Second, the nature of big data interpretation (Wagner-Pacifici, Mohr, & Breiger, 2015). There are however some exceptions in recent publications venturing into discussing the challenges in the ontology. For example, Song and Liu (2017) have argued that veracity is the biggest challenge of tourism data by posing whether big data is meaningful to the existing analysed problem. Zhang (2018) has provided a short commentary. Xu, Nash, and Whitmarsh (2020) have offered a more extensive discussion but are confined to sustainable tourism research. It should also note that a number of tourism

researchers (e.g., Mariani et al., 2018; Li et al., 2018) have cited the works of scholars in other social sciences, such as Kitchin (2014b), Kitchin and McArdle (2016), and Swan (2015), in their discussions.

The epistemology challenge of big data research in tourism and hospitality mainly deals with the shift from theory-driven to data-driven prediction. Our findings revealed that 20 out of 146 articles used a theory to guide their research. We further analyzed these 20 articles that have used a theory to guide their research. We examined: If the articles posted any research questions? Have hypotheses been developed? What theoretical and conceptual paradigm has been applied? How do these articles contribute to the applied theory? Do these articles attempt to address philosophical issues of big data? Please see Appendix B for the detailed answers to the above questions.

Our findings reveal that most studies do frame the research with a theoretical lens and make a contribution to the said theory. For example, Ma, Kirilenko, and Stepchenkova (2020) are properly hypothesized and tested. They used 37652 user-generated 41747 geo-tagged Instagram photos to examine special interest tourism typologies. They empirically tested three hypotheses and found that visitation rates between hardcore and opportunists tourists differ based on travel distance, income, and education which contribute to consumer behaviour theory. Xu and Li (2016) identified the research gap, developed and tested hypotheses under the framework of the multi-attribute theory and Herzberg's motivation-hygiene theory. The authors noted that the findings of this study support Multi-attribute theory/Herzberg's motivation-hygiene theory and confirm expectation-disconfirmation models. Mellinas and Nicolau (2020) study is another example that successfully tested various hypotheses and the findings contributed to Herzberg's motivation-hygiene theory. Thus, our findings unveil that a minority of tourism and hospitality researchers have attempted to guide their studies by conceptual and theoretical frameworks, developed prior research questions, tested hypotheses, and provided logical reasoning to articulate their findings with relevant theory.

In addition, we also noticed that researchers have largely ignored big data ethics, although researchers have called for more research on data governance for ethical data management in tourism and hospitality (Yallop

& Seraphin, 2020). Some researchers have identified several ethical issues related to big data; for instance, big data has created a “data divide” and results in power inequality between organisations, businesses, individuals, academicians, and nations (Van Deursen & Helsper, 2015). The privatisation of data sources has severe implications for knowledge production and research. Usually, organisations and data banks (e.g., Twitter, Facebook, Weibo, and Flickr) release data that have less economic value or that need public assistance to be interpreted. Participants and citizens are encouraged to participate, interact, and interpret online databases that add additional economic value. These kinds of activities may be considered a type of social exploitation (Pasquale, 2016). Data ownership, protection, individual rights, and acquisition are of great vitality in the tourism and hospitality industry, where the sharing economy and growing consolidation have significantly altered the competitive landscape. However, legislation in different countries is underway to fill the regulatory gaps; please see Ferretti et al. (2021) for more details.

### 3.3. Trends in data, methodological approaches, and research themes

#### 3.3.1 *Big data types, sources, format, and size*

The big data used in tourism and hospitality research are derived from three leading sources: user-generated content (UGC) (72%), operations (web searching) (17%), and the Internet of Things (IoT) (10%) (see Figure 5). However, only a small proportion (3%) of the studies used a combination of UGC, operations, and IoT data.

<Insert Figure 5 about here>

Regarding the data sources and formats, a large proportion of articles (22%) collected data solely from one platform. The massive dependence of tourism and hospitality research on a single source (e.g., TripAdvisor) also raises serious questions regarding the representativeness and validity of the findings (see Figure 6). Data from Flickr and Chinese websites (Qunar and Weibo) were used in 11.5% of the articles. In addition,

data were also collected from various social media instruments (e.g., Twitter and Facebook) and IoT devices (e.g., mobile phones, Bluetooth, and GPS), as well as transactional websites.

In terms of data formats, 41% of the data were in the format of user (tourist)-generated reviews. Data with geotagged photos were also collected from various platforms, such as Flickr, Instagram, Facebook, and Twitter. Other data formats are tweets, posts, ratings, blogs, GPS readings, text, articles, mobile tracking records, and Google trends. The questions of bias and representativeness of UGC should be examined more extensively, e.g., what types of users do and do not generate UGC?

<Insert Figure 6 about here>

For sample size, the results indicate that advancements in digital technologies and their applications have helped overcome some of the sample size challenges of conventional sampling techniques. The sample size was determined based on the nature of the data (e.g., UGC, IoT, and operation), which varied from thousands to hundreds of thousands. Moreover, attitudes vary among scholars regarding the usefulness of a large sample size. Some scholars propose that big data enables researchers to answer any question regarding travellers' ideas, suggestions, recommendations, and behaviours (Mariani et al., 2018). Moreover, the same provides a powerful tool for developing innovative research designs that lead to advancing knowledge, supporting managerial decisions, and developing policy implications. Other researchers suggest that a large sample size does not necessarily provide valid information (Swan, 2015). In addition, to requiring a rational understanding, big data might need to be combined and recombined with other sources of information and different methodological approaches. Large sample use would lead to a reconsideration of statistical tools and inferential methods (Fan, Han, & Liu, 2014).

### *3.3.2. Methodological and analytical techniques*

Based on the nature of the data and research objectives, various methodological and analytical techniques have been used to analyse big data in tourism and hospitality research. We noticed that during the first decade, big data research was mostly focused on descriptive statistics, content analysis, frequency

distribution, and correlations. However, advanced analytical methods have been deployed more recently, e.g., spatial, social network, and artificial neural network analyses and machine learning modelling approaches (Giglio, Bertacchini, Bilotta, & Pantano, 2019; Guo et al., 2017). Antonio, de Almeida, and Nunes (2019), for example, used the data of eight hotel property management systems in combination with social reputation, special events, and weather datasets to develop predictive models through machine learning algorithms to minimise hotel losses.

Moreover, various economic models, such as the Var Model, Travel Cost Method, Forecasting Techniques, and Google Analytics, were also used for predictions (Gunter & Önder, 2016). Most articles have applied data and methodological triangulation techniques. Some articles are based on geographical methodologies (Ma et al., 2020; Versichele et al., 2014). Table 2 reveals that several articles applied more than one methodological approach for data analysis. Methodological triangulation accounted for 28%, followed by data mining and visual analytics, each with 11%, and content analysis with 10%. The other noteworthy methods are social networking, cluster analysis, spatial analysis, sentiment analysis, forecasting, and economic modelling.

<Insert Table 2 about here>

Several text mining techniques have been widely adopted for deducting useful knowledge from the hidden patterns of textual data, such as sentiment analysis, cluster analysis, Latent Dirichlet allocation (LDA), statistical analysis, dependency modelling, and text summarisation. Different methods are used for specific purposes. For example, sentiment analysis is used to determine the traveller/tourist attitudes towards particular tourism, such as special interest tourism, hotels, or attraction products, by identifying negative, neutral, or positive feelings (Kirilenko & Stepchenkova, 2017). Various types of regression models are used to compare hotels, destination competitiveness, restaurants, and destination performances, comparing travellers' attitudes, behaviours, and feelings about specific tourism products. Clustering analysis is used in tourism big data for combining the same set of objects/texts in one group (e.g., Cheng, Fu, Sun, Bilgihan,

& Okumus, 2019). This extracted information offers valued knowledge and practical implications for hotel and tourism industry professionals (for more details, see Appendix A).

Our analysis of 146 studies shows that approximately 18% (geotagged + photos data format as in Figure 6) of data have been collected in the form of photos (e.g., Vu, Luo, Ye, Li, & Law, 2018). Cluster analysis was conducted from the spatial aspect to discover tourism destinations or specific attraction spots. Travel trajectories are used to capture the movements between tourism spots and tourists' time between tourist attractions. Various methodological and analytical techniques are used for route generations, for instance, the Markov chain technique.

In addition, GPS data were analysed using statistical techniques, clustering, movement prediction, and pattern mining. For example, Raun, Ahas, and Tiru (2016) used mobile GPS readings to capture tourists' movements and patterns while visiting Estonia. GPS mobile data were also used to investigate tourists' movements and patterns. Our analysis indicates that 3% of data have been collected through mobile devices to analyse tourism-related issues (see Figure 6).

<Insert Figure 7 about here>

We have summarised big data types, collections, storage, processing, and methodological approaches in tourism and hospitality in Figure 7. The process of translating big data into real knowledge can be divided into two domains: data management and data analytics (Labrinidis & Jagadish, 2012). Data management addresses the required technologies needed for acquisition, storage, and retrieval. Data analytics refers to the methods and techniques applied to analyse and extract intelligence from big data. Big data analytics can be classified as text analytics, audio analytics, video analytics, social analytics, and predictive analytics (Gandomi & Haider, 2015). Figure 7 would help researchers collect, process, and select the appropriate techniques for analysing big data in tourism and hospitality to guide decision-making.

Text analytics or text mining refers to the methods used for extracting knowledge from textual data. Text analytics encompassed machine learning, computational linguistics, and statistical analysis. Text analytics

are used for destination image analysis, location, facilities, tourist profiling, hotel and destination market segmentation, as well as the standardisation of tourism products ( Li, Li, Zhang, Hu, & Hu, 2019; Miranda et al., 2011). Audio analytics are applied to unstructured data to extract information. There are two main techniques and technological approaches used for audio analysis: transcript-based and phonetic-based approaches (Gandomi & Haider, 2015). In the tourism and hospitality sector, customer call centres are the principal area of audio analytics. Hotels, resorts, attractions, and destinations utilise these techniques to enhance customer experience, evaluate performance, improve sales, gain insight into customer behaviour, monitor policy compliance, and enhance products and services.

Video analytics or video content analysis comprises various techniques to analyse, monitor, and extract useful information from video streams. Video analytics are used for automated surveillance and security systems (Gandomi & Haider, 2015). In addition to surveillance, video analytics improve operations and customer experience by transforming videos into actionable insights and real-time alerts. Marketing and operation management is the critical area addressed by video analytics. For instance, intelligent algorithms can collect tourists' demographic information, measure the length of stay at a destination, count the number of visitors visited, track patterns of movement, calculate the dwelling time at different attractions, and capture promotion design, pricing, and other real-time information.

Social analytics refers to the analysis of unstructured and structured social media data. It comprises social networks, blogs, social bookmarks, media sharing, review sites, and microblogs ( Li et al., 2018). The major types of social analytics are content-based analytics (text, audio, and video analytics), structure-based analytics (social networking analysis), social influence analysis, and link prediction. These techniques help uncover tourist attractions, opinions, behavioural patterns, sentiments, networking, psychological aspects, peer influence, predicting emerging trends, collaboration between entities, etc. Predictive analytics encompasses various techniques used to predict the future by using historical data, for instance, to predict tourists' next move, what to buy, when, who is influencing their decisions, and their opinion on social

media. There are two methodological approaches: regression techniques and neural networks (machine learning). These techniques are used to predict airline demand, hotels, and destination demand.

### *3.3.3. Research focus*

The review unveiled several significant research focuses. We found that big data analytics primarily focused on tourists' psychology, big data and artificial intelligence in tourism and hospitality, and other related businesses. Under subthemes, big data in tourism focuses on various tourism and hospitality business dimensions, such as business performance, revenue management, supply chain management, attraction management, value creation, and many others (see Figure 8 for more details).

We found that most of the big data research in tourism and hospitality is consumer-centric, and less attention is given to organisational and entrepreneurial research. In addition, some problems, such as hotel occupancy, tourist patterns, tourist attraction, destination image, tourist satisfaction, and perceptions, are repeatedly researched.

We also noticed that researchers tend to work on the same topic with different methods that creates spotlight effects; for instance, Deng and Li (2018) investigated the image of destinations with data mining and visual analytics. Stepchenkova and Zhan (2013) analysed the image of destinations with comparative content analysis. Both studies had utilised geotagged photos.

In addition, we noticed that tourism and hospitality big data research possesses streetlight effects issues; for example, researchers look for answers on issues where there is more data, rather than focusing on novel research problems. For instance, tourist satisfaction, hotel revenue determinants, destination image, and attractiveness have already been discussed in the literature. Each major theme in Figure 8 is developed with the summation of several subthemes; details are provided in Table 3.

<Insert Figure 8 about here>

<Insert Table 3 about here>

#### *3.3.3.1. Textual data research focus*

It is observed that textual data primarily come from online reviews, travellers' blogs, tweets, and diaries. We observed that online reviews generally focused on psychological aspects such as tourist behaviours (spending and consumption patterns and experiences), attitudes towards hotel services, tourist attractions, perceptions of destinations, preferences about standards, and satisfaction attributes. For instance, Berezina et al. (2016) used TripAdvisor online reviews to explore the reasons for hotel customers' satisfaction. Huang, Coghlan, and Jin (2020) analysed tourists' psychological reasons for discontinuance with Airbnb and reported nine online and six offline elements that cause discontinuation, with control, regulation, and policy implications. Table 4 summarises some of the selected studies, providing relevant examples of data types and research focus.

#### *3.3.3.2. Photo data research focus*

Similar to online textual data, photo data are also widely used for analysing tourist behaviour, mobility patterns, experiences, digital footprints, and destination images. For example, Giglio et al. (2019) applied photo data to investigate the attractiveness of Italian cities and tourists' photographic activity trends and recommended a predictive model for formulating tourism circumstances. Agustí (2018) used photo data from Instagram, travel guides, and official tourist brochures to study Uruguay tourism promotion. Stepchenkova and Zhan (2013) compared the photos of Peru collected on Flickr and DMO websites to identify significant differences. The images from both sources reveal that tourists are highly interested in Peru's culture.

<Insert Table 4 about here>

#### *3.3.3.3. Device data focus*

Various devices are used for data collection in tourism and hospitality, such as GPS, mobile phones, and Bluetooth. Device data provide opportunities to measure travelling time between two points: locate tourists, identify most visited attractions, map routes, and study tourists' behaviours, mostly focusing on mobility. For example, Raun et al. (2016) applied mobile tracking data to measure tourists' flow in the destination and differentiated tourism destinations with mobile positioning data. Some of the articles based on device data introduced a conceptual model for studying tourist behaviour, destination management, and travel routes. For example, Park and Pan (2018) combined the gravity model and buying funnel theory to investigate hotel guests' travel behaviours, flight passengers, and mobile device users to identify the next flight route, hence determining the potential market for a destination.

#### *3.3.3.4. Web search data research focus*

The findings reveal that research using web search data focused on predictions, such as tourism and hotel demand forecasts and tourist flows and volume. The articles related to prediction are mostly based on economic modelling. For example, Gunter and Önder (2016) used 10 Google analytics website traffic indicators from DMO websites to predict tourist arrivals to Vienna. Pan, Xiang, Law, and Fesenmaier (2011) focused on web search data in the search engine marketing dynamic and posited that the relationship between tourists and search engines makes it an online destination strategic marketing tool.

### **3.5 Practical implications**

Our analysis also reveals that big data research in tourism and hospitality offers practical implications in destination management, image building, climate change and welfare, hotel performance, revenue management, experience management, supply chain management, and reputation management. The practical implications of all the studies are given in the Appendix A. For instance, the study by Huang et al. (2020) offers useful information for Airbnb managers on the main reasons for discontinuing Airbnb, e.g., lack of transparent policies, communication, check-in processes, cleanliness, and the post-stay

assessment system. Kirilenko et al. (2019)'s research provided insights into tourist segmentation and opportunities for co-programming, co-packaging, and copromotion based on visitors' knowledge and interest. The implications of this research would enable destination managers to use destination attractions as the basis for effective marketing. The study of Vu et al. (2018) offers museum practitioners an understanding of visitors' behaviour and presence, and recommends that managers improve their collections, incorporate new activities, promote visitors' on-site participation, and improve their artefact displays to enhance the visitor experience.

## **5. Discussion**

This study seeks to provide a recap of big data research in tourism and hospitality. Overall, it is evident that this area of research has expanded rapidly in the last 18 years (from 2002–2020). Most existing studies rely on UGC data, followed by location data (GPS and mobile data) and travel information web search data with respect to data type. Due to the growing significance of big data in understanding complex phenomena among different contexts, studies analysing large-scale data are increasing. In terms of research topics, it is observed that current research coverage is not wide enough and has focused on a few research topics, e.g., tourist satisfaction, mobility, destination images, and electronic word of mouth. Big data research evolves in tourism and hospitality over time. For example, the first published articles are more inclined towards data visualisation, and those more recently published are focused on the development of business intelligence. In terms of analytical techniques, traditional regression, text mining and analytics, and machine learning are deployed. However, advanced techniques such as Bayesian classification and artificial intelligence are used only minimally.

This study makes several contributions to tourism and hospitality research by revealing specific knowledge gaps regarding big data. The first knowledge gap lies in conceptualisation, understanding, and philosophically explaining big data in tourism. Big data critically explores new ways of experimentation and empiricism that declare “the end of theory,” such as the generation of data-driven knowledge rather

than knowledge-driven science (Kitchin, 2014b). Theories establish the ground for interpretation and consolidation. It is only a “theory that is capable of deducing inferences and conclusions from patterns in data in a self-consistent way”(Mariani et al., 2018). Hence, epistemologically, big data are nothing more than big algorithms that are completely disengaged from realism’s social dashes. In most studies, the conceptual framework is missing, although it is essential to pinpoint critical business issues and links the domain of big data to the tourism and hospitality sectors. As mentioned earlier, epistemological dilemmas and concerns for the theoretical development of big data-driven science are not yet out of the woods (Ekbja et al., 2015). However, big data research is in the building stage in tourism and hospitality and requires more attention. The findings indicate that scholars are aware of the potential impact of big data on these businesses, individual tourists, and society at large.

In addition, our analysis indicates that little attention has been given to the problematisation of issues in big data research. Big data research in tourism and hospitality is not well developed in terms of the relationships between data analytics and issues under discussion, such as smart tourism development (Gretzel, Sigala, Xiang, & Koo, 2015), mobility patterns (Versichele et al., 2014), smart marketing (María, 2011), and destination management (Pantano, Priporas, & Stylos, 2017). Researchers need to provide sound philosophical footing to big data because data without background knowledge, logic, theory, and predefined assumptions could be considered rather useless. The philosophical foundation would provide legitimacy, a conceptual position, and sense-making to big data. The avoidance of these philosophical inquiries weakens intellectual rigour and broadens the possibility of critiques of big data (Kitchin & Lauriault, 2015).

Another knowledge gap that exists in big data relates to methodologies. Our analysis observes that a limited set of methodological approaches have been used, such as preprocessing techniques, pattern-based methods, speed-based correction methods, clustering methods, frequency-based methods, and econometric modelling. Researchers believe that big data are facing many problems, such as “trajectory data-rich but activity information poor” (Gong, Liu, Wu, & Liu, 2016). For instance, the clustering methods used result in inferred activity locations but do not explain types of locations, and in regard to deriving types of

locations, simplistic methods such as frequency-based approaches (Alexander, Jiang, Murga, & González, 2015) or model-based approaches are usually applied (Chen, Bian, & Ma, 2014).

Another critical problem with big data research is data representativeness (Cynthia, Ma, Susilo, Liu, & Wang, 2016). Many studies reviewed tend to rely on a single platform (e.g., only using data from TripAdvisor). Moreover, our analysis indicates that many of the articles neglect the process of data screening, data cleaning, tokenisation, and word stemming techniques, which are critical for analysing such data (Mariani et al., 2018). Kitchin and Lauriault (2015) suggested that big data findings must complement official statistics at the national, international, and destination levels to increase its value and validity.

Furthermore, existing studies tend to replace small data with big data rather than complement it. It is also a fact that small data (survey) have implications at the firm level (Kitchin, 2014a). Thus, how big data would complement small data is a challenge. It is recommended that the horizon of research in tourism and hospitality be enlarged; for example, the tourist perception metrics of small data should be combined with the actual behaviour stemming from big data online records. The collaboration of big and small data will enrich destinations and hospitality managerial skills, insights, and vision (Cynthia et al., 2016).

The third knowledge gap lies in the data interpretation and implications. Interpretation and mindful knowledge extraction from voluminous, high velocity, and variety of datasets remain very challenging in tourism. First, big data are mostly unstructured, and data collection techniques are not well known by most researchers. Second, big data research is limited in methodologies and conceptual frameworks. It is a conceptual design that enables scholars to compare their findings with other researchers and associate the similarities and differences based on environment, culture, and economy in other world regions. Third, research on big data in the hospitality and tourism literature is scant and highly fragmented (Mariani et al., 2018).

Single research activities often occur in an isolated manner. Without background information and understanding the whole picture, scholars are attempting to discover technical solutions rather than smart

solutions. Such research practices are common during the emergence of new research areas (Knudsen, 2003). The implementation of big data and analytics also poses a great challenge to tourism and hospitality academicians and practitioners. It requires a complete shift in critical thinking, employee skills, technological shift, and organisational transformation in the areas of corporate culture, talent hunt, partnership, leadership, and decision-making processes (Mariani et al., 2018). Without making radical changes to the mentioned areas, big data transformation would cause significant management failures (Hausladen & Schosser, 2020).

## **6. Implications**

### **6.1 Combine theory-driven and data-driven approaches**

Overall, it is essential to use theory to guide the experimental and conceptual design of data collection and analytical techniques to develop reliable and valuable knowledge based on predictive models. The end of theory means the end of the mind (Mariani et al., 2018). Triangulation of data and methodological approaches through the lens of robust theoretical frameworks would help understand big data paradigms. Kitchen and Lauriault (2015) argued that big data should be combined with small data to extract more useful knowledge on established theoretical grounds. Theories are revised continuously, and new concepts are developed under cover of novel pragmatic evidence. Social theories, behavioural theories, economic theories, and causalities provide grounds for understanding the germane domains of tourism and hospitality. Thus, to understand tourists, psychological analyses always require real-time data collection that is expected to influence consumer behaviour (Fuchs et al., 2014).

Future research on big data in tourism and hospitality is recommended to focus on conceptual paradigms to establish a link between issues and solutions. There is a need for more theory-driven big data research rather than merely using complex algorithms and high-resolution visual analytics.

### **6.2 Development of integrated database**

This analysis reveals that many researchers mostly depend on data collected from a single platform, which might be limited in information and representation. In addition, we noticed that researchers continuously study the same phenomena or use similar datasets with different analytical techniques leading to streetlight effects. A potential solution is establishing highly scalable and enterprise-grade web data extraction platforms that scrap information from multiple sources and collate it into one structured dataset. This would help to obtain a thorough picture of the market, competitive strategies, and brand performance based on multiple sources. This shift from part to a whole would help explain industrial trends holistically. In addition, the development of multiple sources of big databases connecting transportation, hotels, accommodations, destinations, attractions, and food service facilities would reveal new relationships and casualties. This can improve network connectivity, cocreation and design, product improvement, insights into add-on services, travel patterns, network influence, tourism flows, the discovery of new opportunities, and identification of the existing problems.

### 6.3 Knowledge sharing

Knowledge sharing and cooperation between academia and industry are recommended to enhance research quality. Academic research attempts to determine causation, develop theory, and build philosophical grounds for big data. However, the industry is more concerned with predicting future trends, identifying operational risks, and influencing consumer behaviour. The collaboration of academia and industry would ensure the combination of diverse high-level competencies to solve complex research problems with the right theoretical direction, high-quality data, and appropriate analytical methods. The findings show that IoT and web search data have been used to a limited extent in tourism and hospitality research, as they are expensive and private. It is suggested that consultation and collaboration be developed between researchers and industry professionals to better use the IoT and web search data.

Most big data research in tourism is focused on online reviews or photo data. However, other types of valuable big data, such as videos, audio, podcasts, and cross-domain data, could also be introduced to the

current stream of research. In addition, these data are more expressive and hence offer new outlooks in understanding the tourism market, traveller behaviour, and other psychological dimensions.

Tourism and hospitality big data research is primarily focused on tourist sentiments, tourist behaviour, and demand forecasting. Other vital issues have been largely ignored, e.g., tourism and happiness, wellbeing, online tourism marketing, product design, emotional attachment, tourism emergency, tourism crisis management, and destination rebuilding.

#### 6.4 Small data and big data

Big data samples almost represent a population, and the notion of statistical significance is not compatible with big data. Small data statistical techniques and methods are based on statistical significance. The output is compared with evaluating a specific relationship's significance and generalising the conclusion to a population. The computational efficiency of conventional methods for small samples does not meet the requirement of big data. The accumulated error for various parameters may dominate the variables' magnitude with large effects (Gandomi & Haider, 2015). Therefore, some significant explanatory variables might be ignored due to accumulated noise in big data.

The massive size of big data creates spurious correlation problems. The correlation between explanatory variables increased with the sample size. The high dimensionality of big data generates erroneous correlations among scientifically unrelated variables (Fan et al., 2014). Various statistical techniques and methods used in regression analysis depend on the exogeneity assumption, suggesting that the explanatory variables are independent of the residual term (error). Endogeneity or the dependence of the residual term on some of the independent variables undermines the statistical methods' validity for regression analysis. Small samples usually meet the exogeneity requirements; however, big data always poses incidental endogeneity, making the findings doubtful (Gandomi & Haider, 2015).

#### 6.5 More precise technique

Big data offer a large sample size that has addressed sample selection biases, leading to appropriate analysis and interpretations that are distinct from the real situation. In the meantime, big data in travel and tourism face reliability issues; for instance, there is a good probability of deceptive reviews (Martinez-Torres & Toral, 2019). Thus, it is highly recommended to develop such a mechanism to detect fake reviews. The textual review data are unstructured and complex and are obtained from various sources in different languages. Its processing and methodological framework development is a considerable challenge for researchers (Liu et al., 2017). Analysing data from multiple sources in multiple languages is a huge challenge, where scholars are unaware of the specific cultural behaviours, preferences, and travelling benefits expected (Schuckert, Liu, & Law, 2015). Moreover, considerable attention is required in interpreting tourist sentiments, such as the degree of intensity of positive and negative emotions (Li et al., 2018).

Online photo data analysis in tourism also required a higher methodological precision and analytical perfection. Online photo data are mainly used to examine tourists' interests, such as most visited attractions. The current stream of tourism research using online photos focuses on destination comparison and spatial patterns (e.g., Stepchenkova & Zhan, 2013; Vu et al., 2018). However, little attention has been given to visiting time, seasonality, and duration (temporal dimensions). It is recommended that scholars focus on the temporal aspects of tourist behaviours in the future. Cluster and sequential analyses are heavily used as analytical tools for online photo data; hence, other methods, such as mining, predictive models, classification, and other reliable techniques, are needed for analysis.

The application of IoT data in tourism and hospitality requires precise methodological approaches and analytical techniques. IoT data devices such as GPS data for tourism focus on tourist behaviours; these studies only explored such data's feasibility for tourism and are silent about its practical implications, such as designing tourist routes and products (Hallo et al., 2012). The contemporary literature analysis suggests that volunteers and GPS loggers were the primary sources for data collection. In this advanced world, we can use other GPS-enabled devices, such as vehicles, yachts, and watches, to obtain less biased and a greater

variety of additional useful information. Research can deliver a high level of insight to professionals about tourists' attention towards specific attractions or products by using web search data. The manager can use web searching for designing tourism products, marketing decision-making, and destination management in the future.

## 6.6 Performance and networking

The tourism supply chain is highly diverse (e.g., hotels, restaurants, travel agents, tour operators, insurance agencies, airlines, transport, lodges, attractions, and destinations). Big data analytics, powerful techniques, and tools can enhance supply chain decisions through data-driven evidence. They can help organisations satisfy customer needs by interpreting huge datasets better and faster and improving their supply chain design by reducing costs and risks (Govindan, Cheng, Mishra, & Shukla, 2018). Big data analytics can achieve coordination, ensuring better performance (Wamba et al., 2017). The transportation and logistics industry heavily depends on supplier relationships and collects a huge amount of data, such as sensor data, customer profiles, customer relations, and real-time locations for predicting future demand (Hausladen & Schosser, 2020). Hence, big data offer supply networks with unlimited data clarity, accuracy, and insights, leading to more contextual communal intelligence across supply chains. Thus, various types of big data, such as those collected from UGC, IoT, and web searching, enable hotels, restaurants, destinations, airlines, and transport firms to predict future demand, resulting in better management of prices and services (Cheng et al., 2019).

## 7. Conclusions

This study provides a comprehensive review of big data research in tourism and hospitality by focusing on its philosophical foundations, data sources and formats, methodological and analytical approaches, research themes and focus, and general trends. Furthermore, this study also revealed the knowledge gaps and research challenges of big data. There are statistical and computational challenges in terms of the analytical methods, e.g., spurious correlation, measurement errors, noise accumulation, algorithmic instability,

multiple sources, and endogeneity. In addition, theorisation remains weak in the majority of big-data research. These challenges and research gaps highlight future directions in the tourism and hospitality discipline.

Although this study offers several useful insights, many limitations must be acknowledged. First, Scopus is considered one of the widely adopted databases for searching research documents; however, we cannot rule out the possibility of missing data. This study is based on a sample determined through the proposed methodological approach instead of on all published literature. Therefore, our findings are constrained by the available literature of the Scopus database. Furthermore, we encourage scholars to collect publications from various databases to refine our findings in the context of upcoming studies related to big data in the tourism and hospitality field. Second, this study focuses on the big data literature in tourism and hospitality in terms of data types, research focus, methodological approaches, and implications. However, big data is a diverse phenomenon and covers hundreds of topics in tourism. Third, we excluded book chapters, commentaries, opinion letters, reports, and non-SSCI or ESCI articles. These sources reported pioneering work on big data and concepts; therefore, researchers are encouraged to expand their work to encompass these sources.

## References

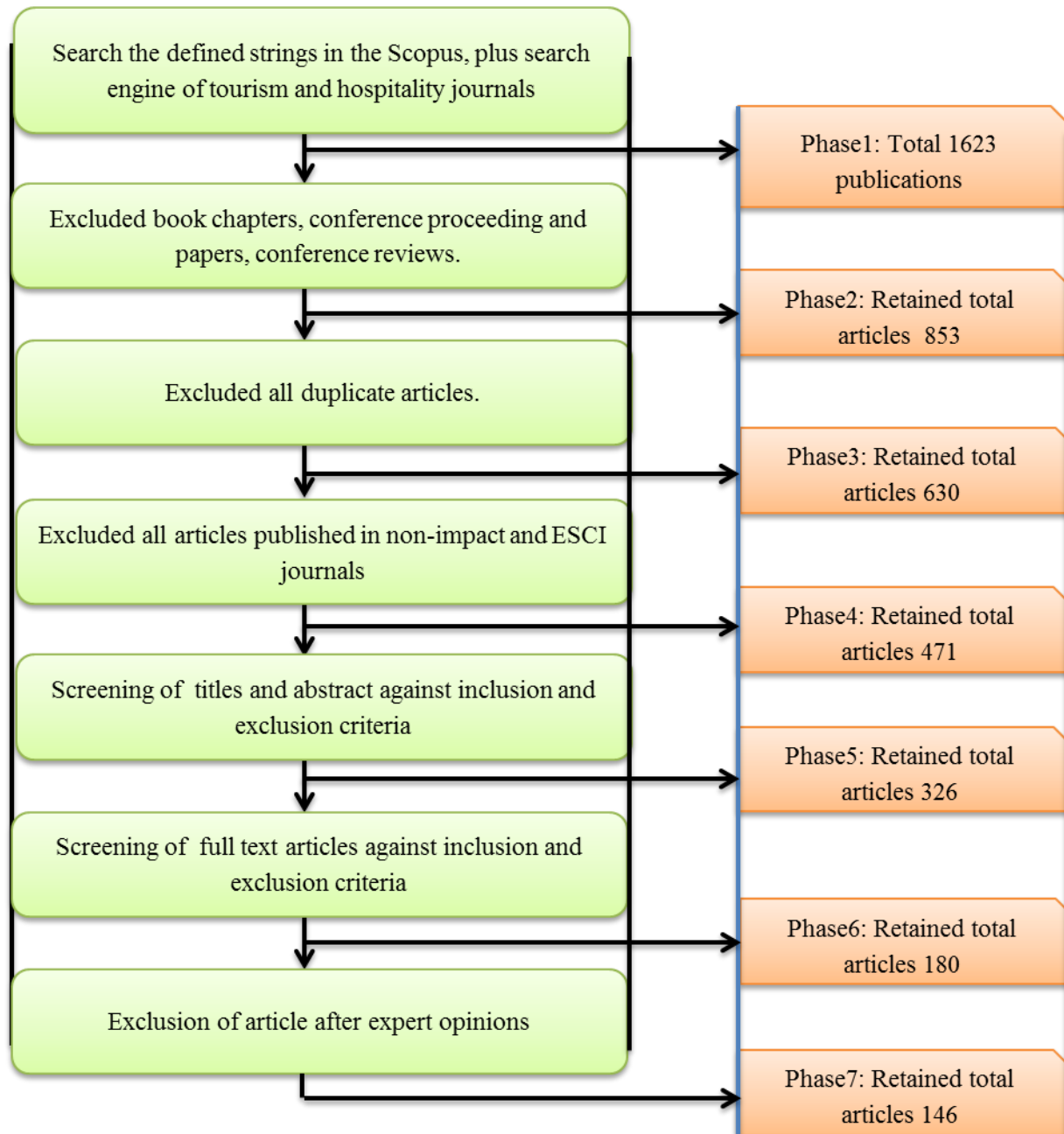
- Agustí, D. P. (2018). Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides. *Annals of tourism research*, 73, 103-115.
- Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, 240-250.
- Antonio, N., de Almeida, A., & Nunes, L. (2019). Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. *Cornell Hospitality Quarterly*, 60(4), 298-319.
- Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, 53, 125-131.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24.

- Black, S. P. (2018). The ethics and aesthetics of care. *Annual Review of Anthropology*, 47, 79-95.
- Chareyron, G., Da-Rugna, J., & Raimbault, T. (2014). *Big data: A new challenge for tourism*. Paper presented at the 2014 IEEE International Conference on Big Data (Big Data).
- Chen, C., Bian, L., & Ma, J. (2014). From sightings to activity locations: how well can we guess the locations visited from mobile phone sightings. *Transp Res Part C*, 46(10), 326-337.
- Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management*, 71, 366-377.
- Cynthia, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285-299.
- Deng, N., & Li, X. R. (2018). Feeling a destination through the “right” photos: A machine learning model for DMOs’ photo selection. *Tourism Management*, 65, 267-278.
- Ekbja, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., . . . Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, 66(8), 1523-1545.
- Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data*, 4(1), 1-20.
- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, 1(2), 293-314.
- Ferretti, A., Ienca, M., Sheehan, M., Blasimme, A., Dove, E. S., Farsides, B., . . . Kleist, P. (2021). Ethics review of big data research: What should stay and what should be reformed? *BMC Medical Ethics*, 22(1), 1-13.
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations—A case from Sweden. *Journal of Destination Marketing & Management*, 3(4), 198-209.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, 72, 306-312.
- Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103-114.
- Govindan, K., Cheng, T. E., Mishra, N., & Shukla, N. (2018). Big data analytics and application for logistics and supply chain management: Elsevier.
- Gransche, B. (2016). The oracle of big data—prophecies without prophets. *The International Review of Information Ethics*, 24.
- Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, 25(3), 179-188.
- Gunter, U., & Önder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of tourism research*, 61, 199-212.
- 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.
- Hallo, J. C., Beeco, J. A., Goetcheus, C., McGee, J., McGehee, N. G., & Norman, W. C. (2012). GPS as a method for assessing spatial and temporal use distributions of nature-based tourists. *Journal of Travel Research*, 51(5), 591-606.
- Hausladen, I., & Schosser, M. (2020). Towards a maturity model for big data analytics in airline network planning. *Journal of Air Transport Management*, 82, 101721.
- He, Z., Deng, N., Li, X., & Gu, H. (2021). How to “read” a destination from images? machine learning and network methods for DMOs’ image projection and photo evaluation. *Journal of Travel Research*, 61(3), 597-619.
- Hernández, J. M., Kirilenko, A. P., & Stepchenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Annals of tourism research*, 73, 35-47.

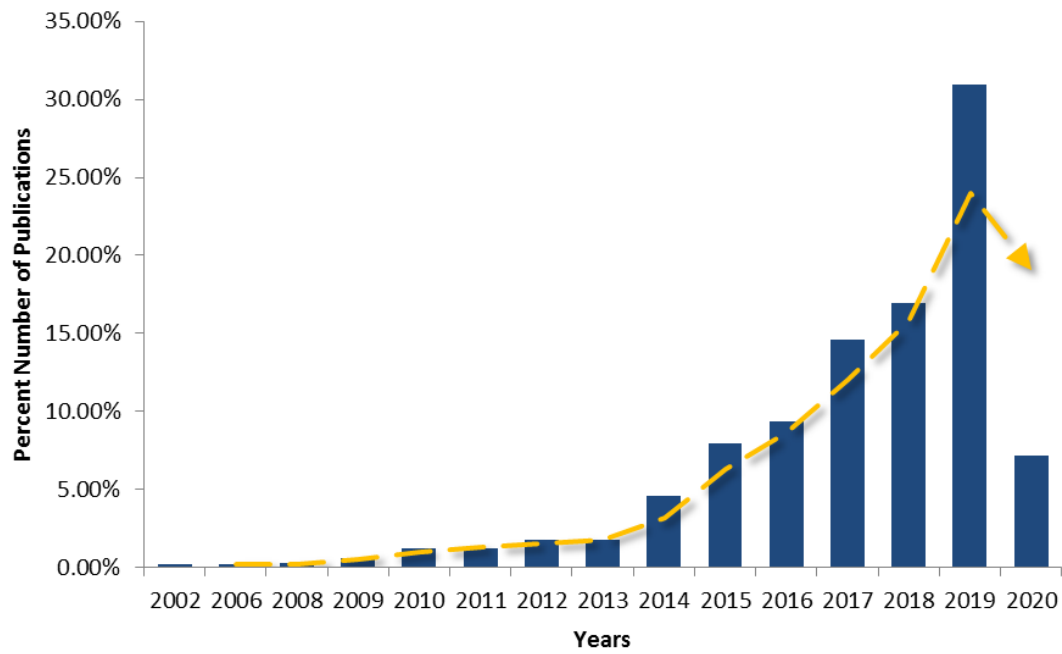
- Himes, H., & Schulenberg, J. (2013). Theoretical reflections: Theory and philosophy should always inform practice. *VOICES OF THE CLOBAL COMMUNITY*.
- Huang, D., Coghlan, A., & Jin, X. (2020). Understanding the drivers of Airbnb discontinuance. *Annals of tourism research*, 80, 102798.
- Hung, K., & Lu, J. (2016). Active living in later life: An overview of aging studies in hospitality and tourism journals. *International Journal of Hospitality Management*, 53, 133-144.
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.
- Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tourism Management*, 63, 54-65.
- Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management*, 72, 400-410.
- Kitchin, R. (2014a). Big data should complement small data, not replace them. *LSE Impact blog*, 27.
- Kitchin, R. (2014b). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1), 2053951714528481.
- Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, 80(4), 463-475.
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big data & society*, 3(1), 2053951716631130.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323.
- Li, Q., Li, S., Zhang, S., Hu, J., & Hu, J. (2019). A review of text corpus-based tourism big data mining. *Applied sciences*, 9(16), 3300.
- Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: evidence from text mining of online reviews. *Tourism Management*, 66, 38-46.
- 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.
- Lu, W., & Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management*, 24(2), 119-154.
- Ma, S. D., Kirilenko, A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse. *Tourism Management*, 77, 104021.
- María, M. A. (2011). Tourist-created content: rethinking destination branding. *International journal of culture, tourism and hospitality research*, 5(3), 291-305.
- Mariani, M. (2019). Big data and analytics in tourism and hospitality: a perspective article. *Tourism Review*, 75(1), 299-303.
- Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of contemporary hospitality management*, 30(12), 3514-3554.
- Mariani, M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, 54, 321-343.
- Martinez-Torres, M., & Toral, S. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403.
- Mellinas, J. P., & Nicolau, J. L. (2020). Let's hook up fast! Hotel reviews and Wi-Fi flaws. *Annals of tourism research*, 80, 102842.
- McFarland, D. A., & McFarland, H. R. (2015). Big data and the danger of being precisely inaccurate. *Big data & society*, 2(2), 2053951715602495.

- Miranda, N., Raminhos, R., Seabra, P., Gonçalves, T., Saias, J., & Quaresma, P. (2011). *Information extraction for standardization of tourism products*. Paper presented at the Conference of the Spanish Association for Artificial Intelligence.
- Moreno, A., Valls, A., Isern, D., Marin, L., & Borràs, J. (2013). Sigtur/e-destination: ontology-based personalized recommendation of tourism and leisure activities. *Engineering applications of artificial intelligence*, 26(1), 633-651.
- Neuendorf, K. A. (2002). : The content analysis guidebook, Sage Publications. Inc., Thousand Oaks, 141.
- Pan, B., Xiang, Z., Law, R., & Fesenmaier, D. R. (2011). The dynamics of search engine marketing for tourist destinations. *Journal of Travel Research*, 50(4), 365-377.
- Pantano, E., Priporas, C.-V., & Stylos, N. (2017). 'You will like it!' using open data to predict tourists' response to a tourist attraction. *Tourism Management*, 60, 430-438.
- Park, S. Y., & Pan, B. (2018). Identifying the next non-stop flying market with a big data approach. *Tourism Management*, 66, 411-421.
- Pasquale, F. (2016). Two narratives of platform capitalism. *Yale L. & Pol'y Rev.*, 35, 309.
- Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, 57, 202-212.
- Ruiz-Martinez, J. M., Minarro-Giménez, J. A., Castellanos-Nieves, D., Garcia-Sánchez, F., & Valencia-Garcia, R. (2011). Ontology population: an application for the e-tourism domain. *International Journal of Innovative Computing, Information and Control (IJICIC)*, 7(11), 6115-6134.
- Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5), 608-621.
- Song, H., & Liu, H. (2017). Predicting tourist demand using big data *Analytics in smart tourism design* (pp. 13-29): Springer.
- Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management*, 36, 590-601.
- Swan, M. (2015). *Philosophy of big data: Expanding the human-data relation with big data science services*. Paper presented at the 2015 IEEE First International Conference on Big Data Computing Service and Applications.
- Van Deursen, A. J., & Helsper, E. J. (2015). The third-level digital divide: Who benefits most from being online? *Communication and information technologies annual*: Emerald Group Publishing Limited.
- Versichele, M., De Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, 44, 67-81.
- Vieira, E. S., & Gomes, J. A. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, 81(2), 587.
- Vu, H. Q., Li, G., & Law, R. (2020). Discovering highly profitable travel patterns by high-utility pattern mining. *Tourism Management*, 77, 104008.
- Vu, H. Q., Luo, J. M., Ye, B. H., Li, G., & Law, R. (2018). Evaluating museum visitor experiences based on user-generated travel photos. *Journal of Travel & Tourism Marketing*, 35(4), 493-506.
- Wagner-Pacifi, R., Mohr, J. W., & Breiger, R. L. (2015). *Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences*: SAGE Publications Sage UK: London, England.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356-365.
- Warburton, N. (1999). *Philosophy: the basics*: Psychology Press.
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.
- Xu, F., Nash, N., & Whitmarsh, L. (2020). Big data or small data? A methodological review of sustainable tourism. *Journal of Sustainable Tourism*, 28(2), 144-163.
- Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, 55, 57-69.

- Yallop, A., & Seraphin, H. (2020). Big data and analytics in tourism and hospitality: opportunities and risks. *Journal of Tourism Futures*.
- Zhang, J. (2018). Big data and tourism geographies—an emerging paradigm for future study? *Tourism geographies*, 20(5), 899-904.



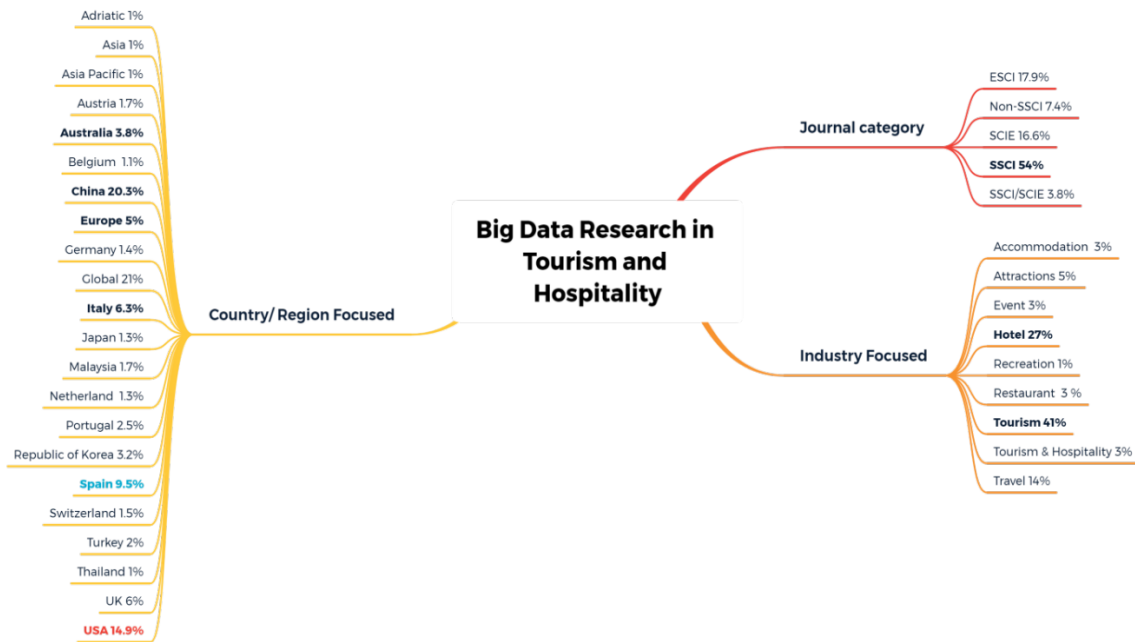
**Figure. 1.** Data Collection Process



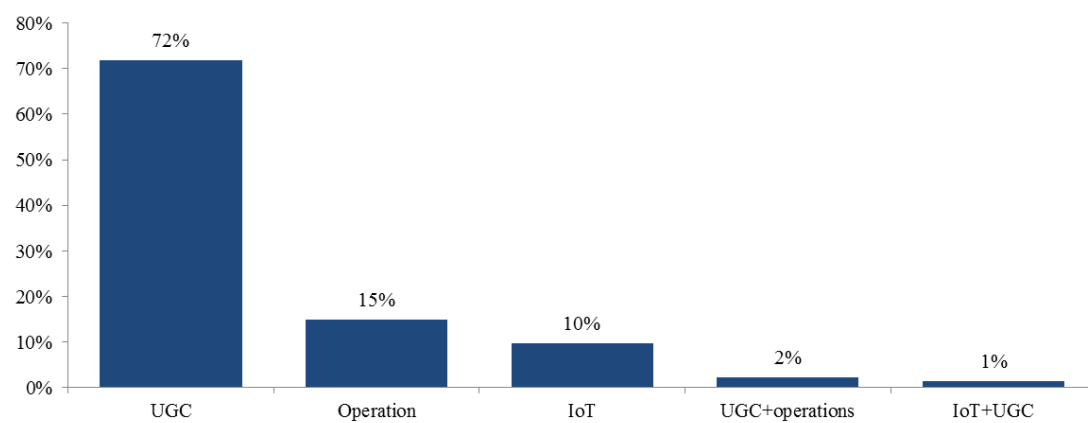
**Figure. 2.** Big Data Year-wise Publications in Tourism and Hospitality (Note: Data in 2020 is until the end of March)

Journal Name	# Publications	Country	# Publications
Anatolia	4	Australia	25
Annals of Tourism Research	11	Austria	11
Applied Geography	6	Belgium	7
Asia Pacific Journal of Tourism Research	7	China	125
Boletín Técnico/Technical Bulletin	7	Denmark	5
Cluster Computing	4	Finland	6
Cornell Hospitality Quarterly	4	Germany	9
Current Issues in Tourism	18	India	22
e-Review of Tourism Research	5	Italy	42
European Journal of Tourism Research	4	Japan	8
Information Processing and Management	4	Malaysia	11
Information Technology and Tourism	10	Netherlands	8
International Journal of Contemporary Hospitality Management	15	Portugal	16
International Journal of Hospitality Management	26	Republic of Korea	20
International Journal of Information Management	4	Singapore	4
International Journal of Recent Technology and Engineering	4	Spain	60
International Journal of Tourism Cities	4	Sweden	5
International Journal of Tourism Research	4	Switzerland	9
ISPRS International Journal of Geo-Information	6	Taiwan	8
Journal of Destination Marketing and Management	11	Thailand	4
Journal of Hospitality and Tourism Research	4	Turkey	8
Journal of Hospitality and Tourism Technology	9	United Kingdom	38
Journal of Hospitality Marketing and Management	8	United States of America	94
Journal of Travel and Tourism Marketing	14		
Journal of Travel Research	15		
Journal of Vacation Marketing	6		
Kybernetes	4		
Scandinavian Journal of Hospitality and Tourism	4		
Sustainability	15		
Tourism Analysis	6		
Tourism Economics	4		
Tourism Management	73		
Tourism Management Perspectives	11		
Tourism Review	9		

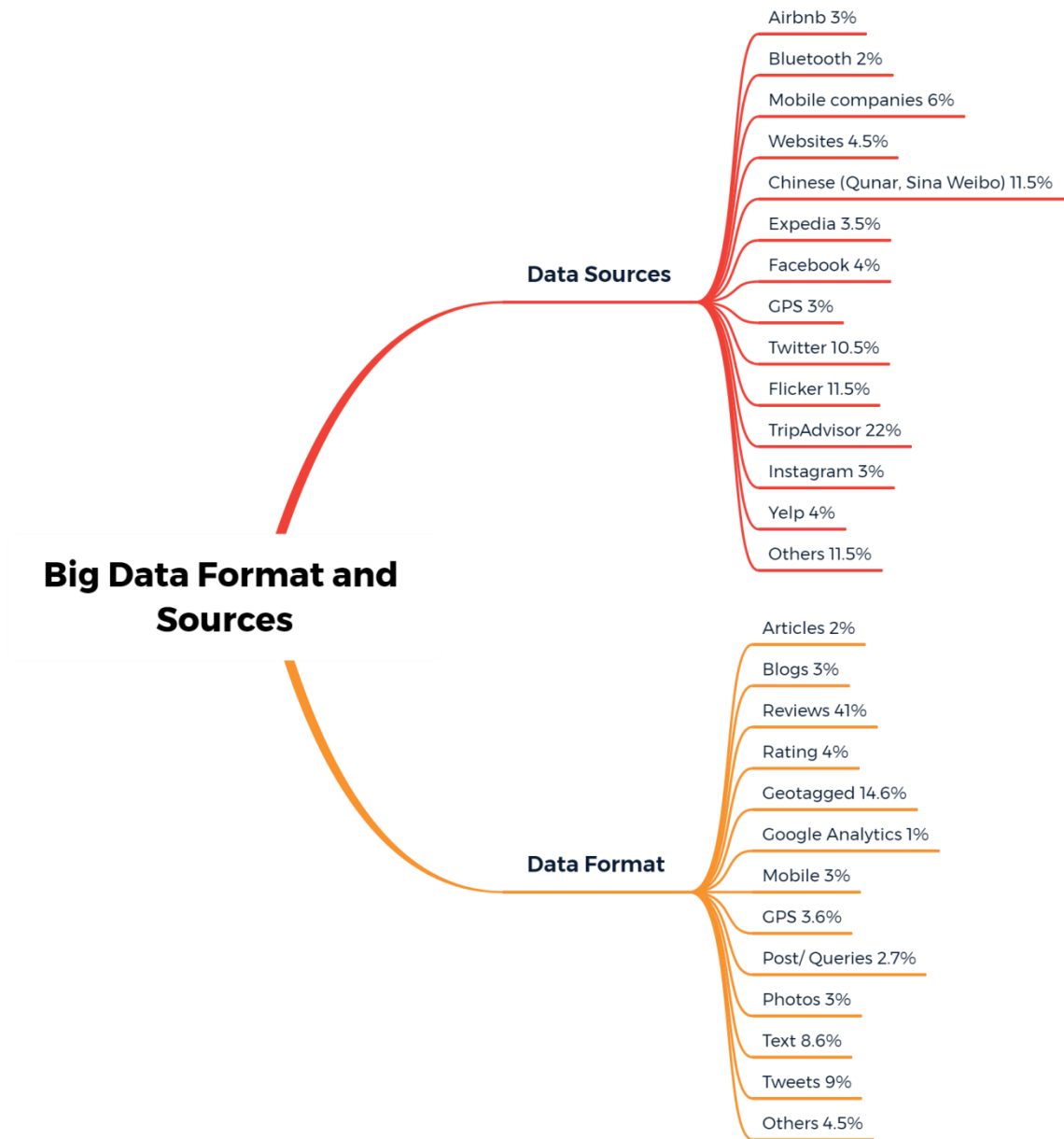
**Figure. 3.** Big Data in Tourism and Hospitality, Number of Publications in Journals, Country of Authorship — Country of Authorship is based on the first authorship (Only journals published more than four publications for the total period are displayed)



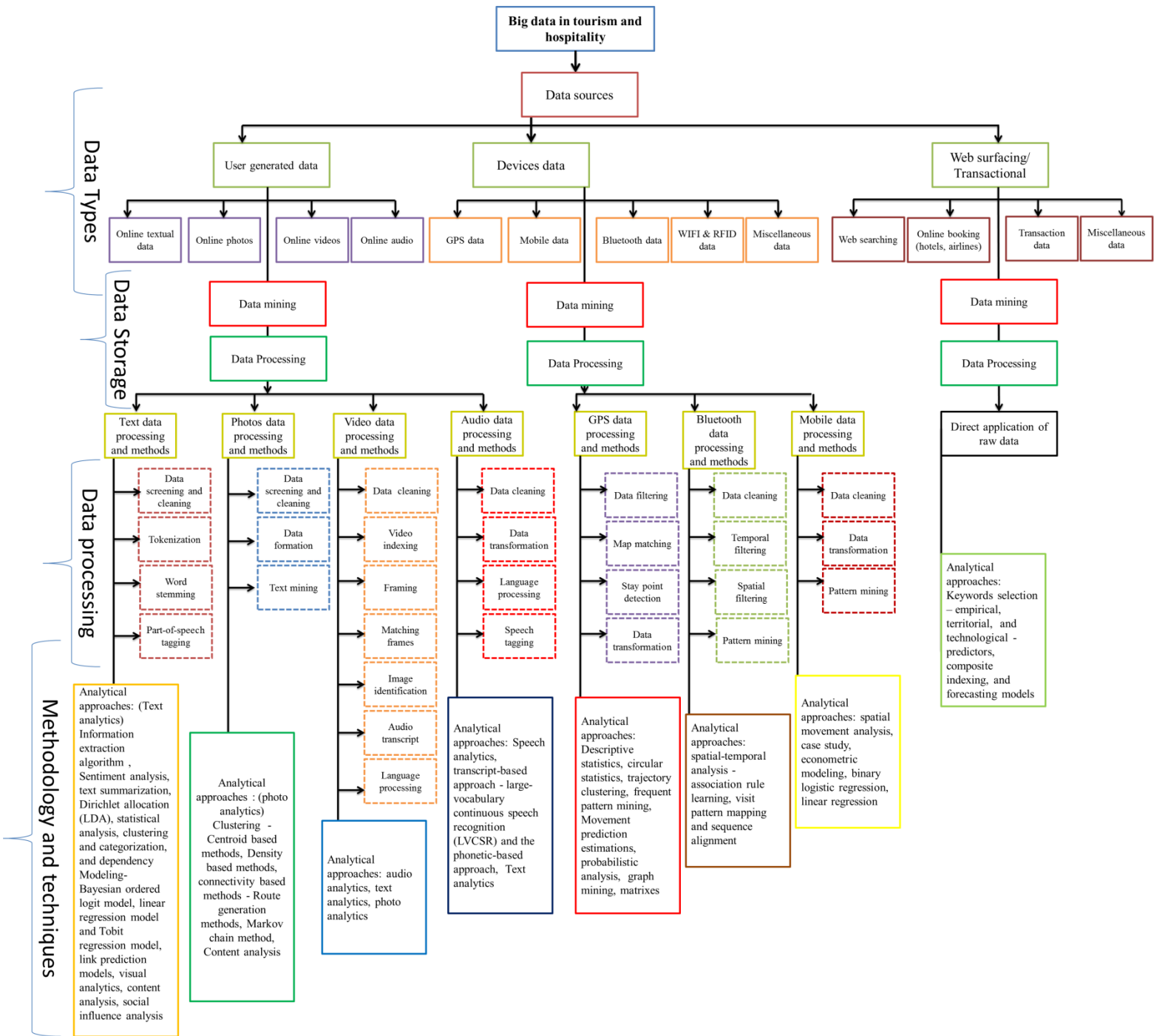
**Figure. 4.** Big data in Tourism and Hospitality, Journal category, Industry Focus and Country/Regional Focus — Country/ Regional Focus is based on data collection site  
 [Note: SSCI = Social Science Citation Index, SCIE = Science Citation Index Expanded, ESCI = Emerging Sources Citation Index]



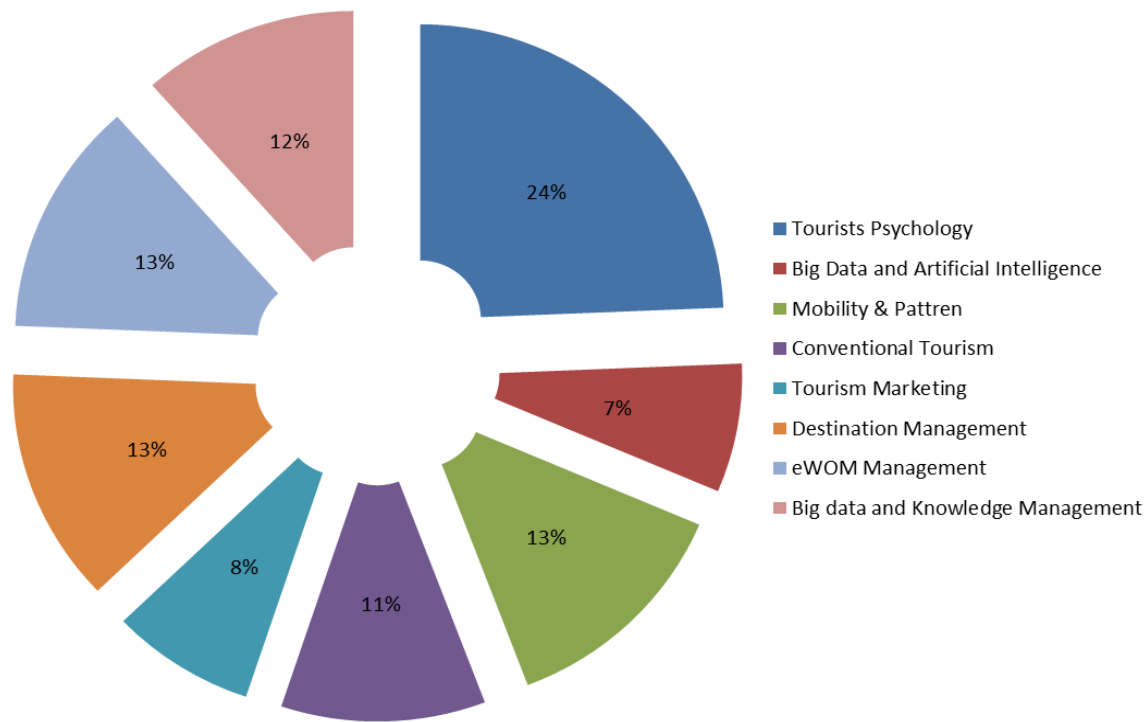
**Figure. 5.** Types of Big Data used in Tourism and Hospitality Research



**Figure. 6.** Big Data Format and Sources



**Figure. 7.** Summary of big data collection, storage, processing, methodologies, and techniques



**Figure. 8.** Big Data Research Focus in Tourism and Hospitality

**Appendix A:Articles on the Application of Big Data in Tourism and Hospitality.**

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Agustí, 2018)	"Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides."	Annals of Tourism Research	37,000	Spatial analyses	Destination image theory	"The results indicated a partial overlap between the locations captured in user-generated images and those promoted by official tourist brochures and travel guides. The findings suggest a territorial distribution of tourism images that are clearly differentiated according to the source(s)."	Tourism promotion	This provides useful insights for applications of social media into future geographical and image management research.
(Hernández, Kirilenko, & Stepchenkova, 2018)	"Network approach to tourist segmentation via user-generated content."	Annals of Tourism Research	14,025	Network analysis	N/A	"Two segmentation solutions are provided: a posteriori, in which only review information is taken into account, and mixed, in which tourist groups are defined a priori by their travel interests and age, and this information is combined with visitation information."	Tourists segmentation	This study contributes to the tourist segmentation literature; an approach to segmenting tourists using network analysis with user-generated content is presented.
(Vu, Li, & Law, 2020)	"Discovering highly profitable travel patterns by high-utility pattern mining."	Tourism Management	3,200	High-utility pattern mining	N/A	Revealed valuable outbound travel patterns of Turkish tourists to several countries in Europe, North America, and Asia.	Travel pattern	The findings are useful for tourism managers to develop valuable travel packages and earn high profits.
(Kubo et al., 2020)	"Mobile phone network data reveal nationwide economic value of coastal tourism under climate change."	Tourism Management	536	Travel Cost Method	Consumer theory	The estimated national economic value loss is higher than the expected national physical beach loss. The results establish regional differences in recreational values.	Tourism climate change and welfare	These conclusions would empower policymakers to discuss management priorities under climate change.
(Ma, Kirilenko, & Stepchenkova, 2020)	"Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse."	Tourism Management	41,747	Cluster analysis	Serious leisure theory/ SIT model	"The large share of tourists involved in what is traditionally understood as special interest tourism activities exhibit behavior and profile characteristic of mass tourists seeking novelty but conscious about risks and comforts."	Special interest in tourism +Tourist behavior + segmentation	This study offers a framework for the development and management of the potential of rural and urban special interest tourism.
(Alaei, Becken, & Stantic, 2019)	"Sentiment Analysis in Tourism: Capitalizing on Big Data"	Journal of Travel Research		Sentiment analysis	N/A	"Big Data and deep learning approaches can help tourism research to discover dynamics based on large interconnected sets of data and getting more insight from different aspects of Big Data."	Use of big data in tourism	Sentiment analysis of tourist's behavior will enable managers, stakeholders to be aware more precisely about the tourist's needs and demand.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Banerjee & Chua, 2016)	"In search of patterns among travellers' hotel ratings in TripAdvisor."	Tourism Management	37,652	Mix methods	N/A	"Travellers' rating patterns for independent and chain hotels are examined. Five traveller profiles are considered: business, couple, family, friend, and solo. Four regions are considered: America, Asia-Pacific, Europe and Middle East-Africa. Data were collected from the popular hotel review website TripAdvisor. Rating patterns for both hotel types differed across profiles and regions."	travellers' pattern + hotel ratings	This paper offers implications to hotel managers in terms of resource allocation, for designers of hotel review websites, and how destination marketers could identify strengths and weaknesses of accommodation options in a given region.
(Berezina, Bilgihan, Cobanoglu, & Okumus, 2016)	"Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews"	Journal of Hospitality Marketing & Management	2,510	Text mining	N/A	Place of business, room, furnishing, members, and sports are used in both positive and negative reviews. "Satisfied customers who are willing to recommend a hotel to others refer to intangible aspects of their hotel stay, such as staff members, more often than unsatisfied customers. Dissatisfied customers mention more frequently the tangible aspects of the hotel stay, such as furnishing and finances".	Tourists psychology	Managerial should consider the e-review categories revealed by the study. "Lack of feedback about customer experience is very costly for the firms. Thus, gaining feedback via text mining is of great value to the firm, and customer reviews of products and services on electronic media and social networks provide valuable managerial information".
(Cheng, Fu, Sun, Bilgihan, & Okumus, 2019)	"An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust"	Tourism Management	1485	Content analysis	N/A	Six thematic features of accommodation experiences were reported as salient themes from online review contents, "the salient cognitive themes are room aesthetics; host attributes, location, and repurchase intention, overall evaluation and room description. These themes govern trust perception: Host attributes positively lead to benevolence, location and room aesthetics positively affect ability, room description positively affects integrity, and location and host attributes determine overall trust perception".	Sharing economy + Trust perceptions	The results have implications on "host attributes and can serve as references for hosts to improve their service quality, increase demand for rooms, and help with social interactions through sharing economy platforms".

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Chua, Servillo, Marcheggiani, & Moere, 2016)	"Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy"	Tourism Management	72,031	Data mining, Visual analytics, Flow analysis, geo-visualizations	N/A	An approach to analyze geotagged social media data from Twitter to "characterize spatial, temporal and demographic features of tourist flows in Cilento is presented. The analysis of geotagged social media data yields more detailed spatial, temporal and demographic information of tourist movements, in comparison to the current understanding of tourist flows".	Tourist flow	The understandings gained from this case study demonstrate the prospective of the projected methodology.
(Duverger, 2013)	"Curvilinear Effects of User-Generated Content on Hotels' Market Share: A Dynamic Panel-Data Analysis"	Journal of Travel Research	138	Mix methods	Social Judgmental theory	UGC has a "positive impact on market share. A curvilinear mechanism explains the relationship between ratings and market share, discouraging lower-tiered firms from seeking a high rating. Moderating factors were also found to diminish the impact of a review length on market share".	eWOM + Hotel market share+ performance	Managers should concentrate on quality to drive ratings. Managers of Economy hotels should encourage ratings and reviews and should try to mitigate the negative reviews.
(Ganzaroli, De Noni, & van Baalen, 2017)	"Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice"	Tourism Management	575	GPS Visualizer & descriptive	Theory of the vicious circle	"The rank of restaurants in TripAdvisor is more related to quality than popularity. TripAdvisor ranking weakly affects the ex-post popularity of restaurants. The ex-post popularity of restaurants mainly depends on their ex-ante popularity. TripAdvisor does not have a strong effect on the vicious circle of tourism in Venice."	quality of restaurants + Tourist decision making	Consumer-generated media (CGM) may help tourists to enhance their capacity to share the value of their experiences. CGM could help municipalities in heritage cities to monitor their local quality of services.
(Giglio, Bertacchini, Bilotta, & Pantano, 2019)	"Using social media to identify tourism attractiveness in six Italian cities"	Tourism Management	26392	Machine Learning Modeling approach	N/A	"The results of this study show maps of the users' behavior identify the annual trend of photographic activity in cities and highlight the effectiveness of the proposed methodology that is able to provide with the place and user information. The study underlines that the analysis of social data can create a predictive model to formulate tourism scenarios."	Tourists behavior & Attractions	Marketers and destination managers can formulate new tourism-cultural scenarios and provide customized services for individuals.
(Guo, Barnes, & Jia, 2017)	"Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation."	Tourism Management	266,544	Text mining, latent Dirichlet analysis (LDA), mapping	N/A	"Latent Dirichlet analysis uncovers 19 controllable dimensions, key for hotels to manage their interactions with visitors. The results suggest variances according to demographic segments. Perceptual mapping identifies the most important dimensions according to the star-rating of hotels."	Tourists satisfaction	By extracting value from UGC, hotel, and destination managers and public administrations officers will be able to hear the voice of the customer more correctly and effectively, thus, providing practical help for business owners and investors.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Hou, Cui, Meng, Lian, & Yu, 2019)	"Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis."	Tourism management	165429	Semantic association analysis	Social network theory	There are obvious inconsistencies in OTAs, UGC platforms in terms of "thematic words, distribution of topics, structural properties, and community relationships. In particular, the results of network visualization can clearly identify hot topics and the social network relationships of thematic words".	Travel opinion mining + tourism decision-making	The article expands the knowledge on methodological challenges and offers novel insights for mining the opinions for the benefit of tourists, hotels and tourism enterprises, and online travel agencies.
(Huang, Coghlan, & Jin, 2020)	"Understanding the drivers of Airbnb discontinuance"	Annals of Tourism Research	582	Content analysis	Innovation diffusion theory	It differentiates between online and offline service issues as key elements of discontinuance, a graphical demonstration of discontinuance drivers provide knowledge about the phenomenon. Introduced discontinuance as a new research stream and offers a grounded understanding of its antecedents.	Consumer discontinuation of Airbnb	Suitable mechanisms to control host behavior are of key significance to the enduring accomplishment of Airbnb. "These mechanisms include developing fair and transparent policies, communicating the policies, host performance in communication, check-in processes, booking acceptance, safety ratings could be auxiliary to the current post-stay assessment system along with accuracy, location, communication, check-in, cleanliness, and value".
(Jin, Cheng, & Xu, 2018)	"Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility"	Journal of Travel Research	1,424	Social network analysis and spatial interaction model	N/A	"The research reveals temporal variations with the length of the trip. The paper highlights the role of time in the tourism study by incorporating a temporal dimension into the analyses and availability of new data".	Tourists mobility	Tourist's mobility offer implications for infrastructure, product development, transportation development, destination, and attractions planning.
(Kirilenko & Stepchenkova, 2017)	"Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests"	Tourism Management	400,000	Sentiment analysis	N/A	This study examines whether sentiments toward the Games changed as the event unfolded, and the event succeeded in producing a positive image. It was found that the positive attitudes expressed in the tweets about the Sochi Olympics improved throughout the course of the Games, only for the hosts' segment of the sample, and smaller improvement in the guests' segment.	Mega sport event	The managerial implications are visible by using Big Data analytics. The potential use of such information will increase if the knowledge of Big Data on mega sports events will be used. Thus, this study provides a useful starting point for the host countries of future Olympic events.
(Kirilenko, Stepchenkova, & Hernandez, 2019)	"Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews"	Tourism Management	254,571	Network analysis, spatial analysis, and geo-visualizations	N/A	"Found three different groups of tourists: those primarily visiting amusement parks (Excitement cluster), those concentrating on cultural attractions (Heritage cluster), and those visiting beach and other nature-based locations (Nature cluster). There are large-scale geographical differences between the locational preferences of the tourist segments."	destination attractions, Visitation patterns	The concrete implication of this "scholarship lies in increased opportunities for co-promoting, co-packaging, and co-programming groups of attractions for different tourist markets based on tourists' interest and knowledge level and, therefore, achieving more effective marketing of destination attractions".

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Li, Lin, Tsai, & Wang, 2015)	“Traveller-generated contents for destination image formation: Mainland China travellers to Taiwan as a case study”	Journal of Travel & Tourism Marketing	1,033	Text mining and content analysis	N/A	It is found that the “allure to images of activities and positive/ negative impressions are tied with five categories (attraction, shopping, food and beverage, accommodation and transportation) form travellers’ destination image. Traveller-generated web content can be beneficial for destination image analyses in marketing and management. Highlighted the importance of understanding destination image formation from traveller-generated contents and the challenges for those in tourism marketing to narrow the gap”.	Destination image	This study results provide new insight into the leisure and cultural activities of Hong Kong and Taiwan. It provides marketing implications for destination managers.
(Liu, Teichert, Rossi, Li, & Hu, 2017)	"Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews."	Tourism Management	412,784	Descriptive and ANOVA	N/A	"Found that foreign tourists, who speak diverse languages (English, German, French, Italian, Portuguese, Spanish, Japanese, and Russian), differ substantially in terms of their emphasis on the roles of various hotel attributes ("Rooms," "Location," "Cleanliness," "Service," and "Value") in forming their overall satisfaction rating for hotels."	Tourists satisfaction	Segmentation based on the customers' language is essential if one obtained a more precise understanding of hotel customers' unique preferences and their satisfaction with hotels.
(Liu, Schuckert, & Law, 2018)	"Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews"	Tourism Management	19,674	Text mining, Correlation analysis.	Goal-setting theory	This study "investigates the effect of online incentives on travelers' writing behavior during the status-seeking process. Empirical findings drawn from the data of 19,674 TripAdvisor members suggest that 1) at a lower status or earlier stage, members are more eager for quick promotion and utilitarianism results in fewer words per review; and 2) members' knowledge grows as their status rises."	Utilitarianism and knowledge and status + Tourists psychology	This yield managerial implication for the operation of UGC websites, especially online travel agencies, and for tourism management.
(Lu & Stepchenkova, 2012)	"Ecotourism experiences reported online: Classification of satisfaction attributes"	Tourism Management	373	Content Analysis	N/A	"26 attributes that influence eco-tourists' satisfaction with their eco-lodge stays were identified and further aggregated into seven categories: eco-lodge settings, room, nature, service, food, location, and value for money."	Ecotourism and satisfaction attributes	This article offers practical implications for eco-lodges and consumer satisfaction attributes.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Martinez-Torres & Toral, 2019)	"A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation."	Tourism Management	1,600	Content Analysis	N/A	"Distinguish positive and negative deceptive and non-deceptive reviews based on unique attributes and sentiment."	Experience-based on deceptive reviews	A system should be developed by "the operator to detect deceptive reviews, Need for advanced artificial intelligence for tourism and hospitality reviews."
(Mellinas & Nicolau, 2020)	"Let's hook up fast! Hotel reviews and Wi-Fi flaws"	Annals of Tourism Research	4,800	Heckit model	Herzberg's motivation–hygiene theory	"Wi-Fi is regarded as a hygiene factor, whereas speed is considered a motivator favor, WIFI influence guest satisfaction."	Tourists psychology	"Hotels should secure a certain level of WIFI connection and speed as it is essential."
(Moghavvemi et al., 2017)	"Connecting with prospective medical tourists online: A cross-sectional analysis of private hospital websites promoting medical tourism in India, Malaysia and Thailand"	Tourism Management	51	Content analysis	N/A	"5 categories (Hospital Information and facilities, Admission and medical services, Interactive online services, External activities, Technical items) and 38 items were identified on hospital websites for promoting medical tourism."	Influence of WIFI flaws on hotel reviews and rating	"Pointing to the need for hospital managers to improve their hospitals' online presence and interactivity."
(Pantano, Priporas, & Stylos, 2017)	"‘‘You will like it!’ using open data to predict tourists' response to a tourist attraction’’	Tourism Management	250	Predictive modeling	N/A	Online data make it more possible to predict tourists future preferences	Big data role in tourists prediction	‘‘Destination marketers can evaluate tourists' responses to a certain destination in advance, and can potentially influence the final destination choice by improving marketing strategies’’
(Seunghyun ‘‘Brian’’ Park, Ok, & Chae, 2016)	‘‘Using Twitter Data for Cruise Tourism Marketing and Research’’	Journal of Travel & Tourism Marketing	42,785	Content analysis and network analysis	N/A	‘‘Caribbean, Galapagos, and the Bahamas are among the most popular cruise destinations, and use twitter accounts for marketing.’’	Cruise tourism marketing	Established the likelihood of applying social media analytics tools and methods in cruise research.
(Seunghyun "Brian" Park, Kim, & Ok, 2018)	‘‘Linking emotion and place on Twitter at Disneyland’’	Journal of Travel & Tourism Marketing	56,418	Text mining, Sentiment analysis, hot spot analysis	N/A	Tracking consumers’ emotional response, spatial patterns, highly liked attractions	theme park + Tourists psychology	Methodological triangulation, offer practical implications for tourists management and attraction marketing
(Park & Pan, 2018)	‘‘Identifying the next non-stop flying market with a big data approach’’	Tourism Management		Buying funnel theory, Gravity Models	Buying funnel theory and gravity model	‘‘Impact of the direct flight route, non-stop route has a significant positive relationship with the number of visitors, five potential cities were identified.’’	Airlines non-stop services + Potential market identification	Provide valuable implications for the airline industry based on destination interest. Forecasting the best potential markets for the airlines.
(Philander & Zhong, 2016)	‘‘Twitter sentiment analysis: Capturing sentiment from integrated resort tweets’’	International Journal of Hospitality Management	31,550	Sentiment analysis	N/A	Twitter data is a low-cost and real-time measure of hospitality customer attitudes/perceptions	customer attitudes/perceptions	Methodological contribution by demonstrating a new and innovative analytical method.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Phillips, Zigan, Silva, & Schegg, 2015)	“The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis”	Tourism Management	59,688	Artificial Neural Network	N/A	“Revenue per available room is negatively affected by room quality, positive regional review, hotel regional reputation and positively impacted by regional room star rating.”	Determinants of hotel performance	Quantify the impact of social media on hotel performance, utilization of an aggregated evaluation score in order to evaluate UGC.
(Raun, Ahas, & Tiru, 2016)	“Measuring tourism destinations using mobile tracking data”	Tourism Management	361,853	Spatial analyses, Binary logistic regression	N/A	Smaller destination areas can be distinguished inside the whole country by the geographical, temporal and compositional parameters of the visits	Spatial, temporal, compositional, social and dynamic of destination	Applications of “BIG” data in destination management
(Salas-Olmedo, Moya-Gómez, García-Palomares, & Gutiérrez, 2018)	“Tourists' digital footprint in cities: Comparing Big Data sources”	Tourism Management	234159, 307062	Spatial analysis, Cluster analysis, Regression	N/A	Tourists movement in the cities cannot be assessed and analyzed by one type of data, therefore, a triangulation of data should be made for accurate assessment	tourists behavior + pattern	Triangulation of data. The results are significant for locating new opportunities of business for the private sector, identifying areas with economies for locating retail tourism.
(Stepchenkova & Zhan, 2013)	“Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography”	Tourism Management	500	Comparative content analysis	Theory of hermeneutic circle of destination representation	‘Level of congruity between the projected and perceived images of this destination, DMO may consider expanding on existing ethnographical attractions or engaging in new ones”	Destination image & marketing	Destination marketing and development through visualization
(Stringam & Gerdes Jr, 2010)	“An Analysis of Word-of-Mouse Ratings and Guest Comments of Online Hotel Distribution Sites”	Journal of Hospitality Marketing & Management	60,648	Difference between proportions method (DBPM)	N/A	Travellers or guest showed concern about the bed, bathroom, amenities, food, and beverages, facilities to the attractions	Word of Mouth+ online hotel rating	Allocation of resource to improve hotel rating based on travellers opinion
(Su, Stepchenkova, & Kirilenko, 2019)	“Online public response to a service failure incident: Implications for crisis communications”	Tourism Management	10,245	Descriptive analysis, visualization, principle component analysis, Hotspot analysis	Situational Crisis Communication Theory and Attribution Theory	The study finds evidence that the effect of the incident on the general public is moderated by spatial and personal proximity of Weibo users	crisis communication in tourism +image restoration perspectives	Hotels should pay time detailed response to the customer damages and complaints, monitor time-specific social media and should take action before negative eWOM.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Villamediana, Küster, & Vila, 2019)	“Destination engagement on Facebook: Time and seasonality”	Annals of Tourism Research	639 posts, 178,913 audience reactions	Regression analyses	N/A	“Posting time is the most potent predictor of positive/negative engagement, posting before the high tourism season increases positive engagement, social media influences consumers' purchase decision journeys”	Destination engagement on social media	DMO managers are recommended to consider posting time in their marketing strategies,
(Vu, Li, Law, & Ye, 2015)	“Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos”	Tourism Management	29,443	Data mining, clustering and the Markov chain	N/A	“Discover the locations in which tourists are most interested and the routes, spatial and temporal movements of tourists.”	Inbound tourist travel behavior	The study benefits destination development, transportation planning, and impact management.
(Vu, Luo, Ye, Li, & Law, 2018)	“Evaluating museum visitor experiences based on user-generated travel photos”	Journal of Travel & Tourism Marketing	2843	Spatial analysis, Descriptive statistics, Temporal analysis, Photo content analysis	N/A	“Content in the History & Science Museum and Heritage Museum are more attractive compared with that of the Museum of Arts and the Space Museum, because it is located near to the other central attractions. Space Museum managers may consider improving their collection, enhancing their artifacts, or incorporating new activities and promoting visitor participation and attraction sites.”	Tourists' behavior and travel pattern	“Benefit tourism researchers and practitioners in understanding visitor behavior and preferences and allow museum practitioners to provide better products to museum visitors.”
(Vu et al., 2020)	“Discovering implicit activity preferences in travel itineraries by topic modeling”	Tourism Management	141,377	Itinerary analysis, Probabilistic topic modeling, Latent Dirichlet allocation technique to text data, Exploratory Analysis	N/A	The findings provided insights into the activity patterns of tourists, tourist activity preferences for developing appropriate travel and tour packages are important	Outbound travel behavior+ preferences	The results are useful for travel and tourism managers in developing travel and tour packages.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Xiang & Pan, 2011)	“Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations”	Tourism Management	19,016	Text mining, Text analysis, correspondence analysis	N/A	The article offers insights into the way tourism destinations are searched online	Online travelling pattern across tourists' destination	Offers implications for search engine marketing for destinations
(Xiang, Du, Ma, & Fan, 2017)	“A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism”	Tourism Management	438,826 351,182 30,770	Text analytics, semantic analysis	N/A	The findings suggest online reviews vary considerably in their “linguistic characteristics, semantic features, sentiment, rating, usefulness as well as the relationships between these features on different online platform. This study is the first to comparatively explore data quality in social media studies in hospitality and tourism”.	Comparison between online hotel review	This study offers a foundation for understanding the methodological challenges and identifies several research directions for social media analytics in hospitality and tourism
(Zhang, Zhang, & Yang, 2016)	“The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior”	Tourism Management	61,127	Regression analyses	cognitive-processing capacity theory	Positive effect of expert hotel reviews on travelers' rating behavior, The findings help guide travel website design and improve e-word of mouth of a hotel.	expert review influence on Traveller online rating behavior	Provide implications for hotel managers and travelling websites, to provide expert reviews to influence traveller behavior and rating
(Zhang, Chen, & Li, 2019)	“Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing”	Tourism Management	137,265	Cluster analysis, spatial analysis, Descriptive analysis	N/A	“Tourism destinations should extend the tourism attractions from traditional scenic spots to more modern scenic spots, and lead tourists to take more experiential activities”	Tourists' behavior and perception in tourism destination	Provide implications for destination managers to innovate their product and marketing strategies
(Zhou, Ye, Pearce, & Wu, 2014)	“Refreshing hotel satisfaction studies by reconfiguring customer review data”	International Journal of Hospitality Management	1345	Text Mining, descriptive and ANOVA	N/A	“Twenty three satisfaction attributes were identified that influence customer satisfaction, identified the influence of hotels with different star ratings and different ownership on tourists’ satisfaction appraisal.”	Hotel attributes influencing customer satisfaction	The attributes defining hotel customer satisfaction adopted in this study are potentially applicable to hotels
(Zhu, Cheng, Wang, Ma, & Jiang, 2019)	“The construction of home feeling by Airbnb guests in the sharing economy: A semantics perspective”	Annals of Tourism Research	42,085	Semantic analysis	N/A	“Home feeling dimensions of Airnab are reported i.e. physical, social and affective, safety, privacy, and engagement and fun, Hospitality dimension, Helpful, cultural experience, etc..”	Airbnb guests + home feeling	Important “managerial guidance for Airbnb developers, operators and/or hosts to further strengthen their own competitive advantages. Managerial implications for Airbnb hosts to increase return business by creating stronger social and affective dimensions of home feeling. It also offers insight for hotels to reflect on their core value in a commercial setting.”

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Hernández et al., 2018)	“Network approach to tourist segmentation via user generated content”	Annals of Tourism Research	14,025	Cluster analysis, Network analysis, Geo-visualization	N/A	Two networks are reported, i.e., “A posteriori segmentation: tourists-attractions network, Mixed a priori-a posteriori segmentation.”	Interrelationships between tourists and attractions at the destination	This study contributes to the tourist segmentation literature, both theoretically and method-wise, by proposing and demonstrating an approach to segmenting tourists using network analysis with user-generated content.
(e Silva et al., 2018)	“Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources”	Tourism Management	843000	Visual analytics, statistical indexing, Mapping	N/A	“Complete and consistent dataset describing tourist density at high spatial resolution with monthly breakdown for the whole of the European Union. The relative impact of tourism and its seasonality show marked uneven spatial distributions. Key spatiotemporal patterns and characteristics of tourism in Europe at both regional and local scales.”	Tourist density, Tourism intensity	Potential to further exploit seasonal information from online booking services exists.
(Xiang, Schwartz, Gerdes Jr, & Uysal, 2015)	“What can big data and text analytics tell us about hotel guest experience and satisfaction?”	International Journal of Hospitality Management	60,648	Text mining, factor analysis, regression	Herzberg's motivation–hygiene theory	The relationship between guest experience and satisfaction looks strong, proposing that these two areas of consumer behavior are fundamentally connected	utility of big data analytics in hospitality	This study reveals that big data analytics can produce new insights into variables that have been broadly studied in existing hospitality.
(Fang, Ye, Kucukusta, & Law, 2016)	“Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics”	Tourism Management	41,061	Text mining, regression analyses	N/A	“The effect of text readability is confirmed in attraction reviews. Text readability and reviewers characteristics influence affect the perceived value of reviews.”	Text readability and reviewer characteristics	The study results benefit to attraction management.
(Deng & Li, 2018)	“Feeling a destination through the “right” photos: A machine learning model for DMOs’ photo selection”	Tourism Management	20,974	Data mining, Visual analytics	N/A	Tourists taken images can be used for destination promotion. The destination manager can promote the destination attraction through the selection of proper images for feeling the real sense of the destination. The advantage of the approach is to shrink the gap between the projected and received image	Destination Image + marketing	Offers practical implications for destination/ attraction promotion and marketing
(Hallo et al., 2012)	“GPS as a Method for Assessing Spatial and Temporal Use Distributions of Nature-Based Tourists”	Journal of Travel Research	4,389	Spatial analytical	N/A	This study finds that “GPS offers advantages over traditional methods for tracking visitors, including more reliable, accurate, and precise data. GPS makes measuring and understanding visitor use patterns even more valuable since <i>actual</i> movements can be investigated instead of <i>reported</i> movements.”	Tourists distribution	Visitor-based GPS surveys signify a useful and potentially powerful tool to help better plan, manage, and monitor nature-based tourists

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Hunter, 2013)	“China’s Chairman Mao: A visual analysis of Hunan Province online destination image”	Tourism Management	995	Content analysis-semiotic analysis	N/A	Findings indicate that “representations of Chairman Mao play a significant part in the definition of the Hunan online destination image”.	Destination Image	These findings are valuable for destination development policymakers and destination image marketing
(Li, Xu, Tang, Wang, & Li, 2018)	“Big data in tourism research: A literature review”	Tourism Management	165	Content and descriptive analysis	N/A	“Various big data have been applied to tourism research, making great improvement. Carrying different information, different types address different tourism issues. Each type research focuses, data characteristics and analytic tools are analyzed.”	Big data in tourism and hospitality	This survey facilitates a systematic understanding of big data research and offers valuable insights into its future prospects
(Kim, Kim, Lee, Lee, & Andrada, 2019)	“Quantifying nature-based tourism in protected areas in developing countries by using social big data”	Tourism Management	168	Visual analytics, Mapping, regression	N/A	“Hotspots of high visitation were identified, while revealing the local spatial impact of distributed attributes”.	Visitation to nature based destination	This study offers insights into the applicability of social big data to protected-area management and its potential in strengthening current field-based participatory methods.
(Li, Law, Vu, Rong, & Zhao, 2015)	“Identifying emerging hotel preferences using Emerging Pattern Mining technique”	Tourism Management	118,300	Data mining	N/A	This study attempt to find international hotel preference through big data. Club lounge, pool, food, view, parking area, services, location, taxi services are identified as emerging trends that are getting more attention.	Hotel preferences and big data	The methods and findings can help hotel managers to gain deep understandings of travelers' interests, permitting the former to gain a better understanding of the fast variations in tourist preferences.
(Mariani, Di Felice, & Mura, 2016)	“Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations”	Tourism Management	33,597	Data triangulation	N/A	“Visual content and the average length of posts have a positive impact on engagement. Post frequency has a negative impact on engagement.”	Destination marketing	This study can only be generalized to countries “where regional levels of government, or individual municipalities, play a predominant role in destination marketing”.
(Nieto, Hernández-Maestro, & Muñoz-Gallego, 2014)	“Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments”	Tourism Management	29,507	Regression analyses	N/A	“Price and advertising expenditures affect customer reviews. Customer reviews affect business performance. Establishment owner’s experience moderates the link between reviews and performance.”	eWOM +Marketing decision	The study offers destinations' manager an insight for making marketing decision based on big data

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Park & Nicolau, 2015)	“Asymmetric effects of online consumer reviews”	Annals of Tourism Research	5,090	Count models	theory of information diagnosticity	The results show that “people perceive extreme ratings (positive or negative) as more useful and enjoyable than moderate ratings, giving rise to a U-shaped line, with asymmetric effects: the size of the effect of online reviews depends on whether they are positive or negative.”	eWOM +tourist decisions	Managers should identify that measures applied to take advantage of positive reviews and the actions developed to defend the firm from negative reviews should weigh differently.
(Rahmani, Gnoth, & Mather, 2018)	Hedonic and eudaimonic well-being: A psycholinguistic view	Tourism Management	10,912	Data mining, SEM	self-determination theory, cognitive theory	“An increase in Hedonic relates: positively to an increase in Eudemonia. For all three destinations, a decrease in Hedonic appears with an increase in words signaling a decrease of Eudemonia. The results strongly support the idea that Hedonic and Eudemonia together lead to the highest level of well-being.”	Tourists' experience +Hedonic and eudemonic well-being	The findings assist managers in designing destinations according to the desired type of well-being. Helps tourists to make a more informed decision about their holiday destination
(Talón-Ballester, González-Serrano, Soguero-Ruiz, Muñoz-Romero, & Rojo-Álvarez, 2018)	“Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector”	Tourism Management	4,935,806	Big Data analytics on proportion tests from ratios	N/A	“Strong uniformity was found on the most demonstrative feature of repeaters being traveling without children. Profiles were more similar for British and German clients, and their main differences with Spanish clients were in the stay duration and in age.”	Client profiling	Big Data technologies can be enormously valuable for analyzing indoor data accessible in CRM information systems from the hospitality industry.
(Versichele et al., 2014)	“Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium”	Tourism Management	235,597	Exploratory analysis, Data visualization	N/A	“Tourists in Ghent, Belgium tracked over 14 tourist attractions, 14 hotels and tourist inquiry desk during 15 days. Visit pattern analysis through mining of association rules between tourist attractions. Visualization of discovered patterns by visit pattern maps.”	mass event +Festival	Tourist segmentation in mass events. The potential of Bluetooth tracking in tourism management.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Xie, Zhang, & Zhang, 2014)	“The business value of online consumer reviews and management response to hotel performance”	International Journal of Hospitality Management	4994	Descriptive, Regression	N/A	The results show that “overall rating, attribute ratings of the purchase value, location and cleanliness, variation and volume of consumer reviews, and the number of management responses are significantly associated with hotel performance. In addition, variation and volume of consumer reviews moderate the relationship between overall rating and hotel performance. Management responses, together with variation and volume of consumer reviews, moderate the relationship between certain attribute ratings and hotel performance.”	eWOM + hotel performance	Utilizing consumer reviews and management responses to leverage hotel business is the main implication of this study.
(Xu & Li, 2016)	“The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach”	International Journal of Hospitality Management	3480	Semantic analysis, regression analyses	multi-attribute theory	“Determinants that create either customer satisfaction or dissatisfaction toward hotels are diverse and are specific to particular types of hotels, including full-service hotels, limited-service hotels, suite hotels with food and beverage, and suite hotels without food and beverage.”	eWOM +tourist's psychology	the study provides evidence for hoteliers to improve customer satisfaction and lessen customer dissatisfaction by improving service and satisfying the customers’ needs for the various kinds of hotels.
(Ye, Law, & Gu, 2009)	“The impact of online user reviews on hotel room sales”	International Journal of Hospitality Management	3625	Regression analyses	N/A	“Results indicate a significant relationship between online consumer reviews and business performance of hotels.”	eWOM +hotel performance	Offers practical implications for hotel and restaurant managers
(Zhang, Ye, Law, & Li, 2010)	“The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews”	International Journal of Hospitality Management	1242	Regression analyses	N/A	This study shows that “consumer-generated ratings about the quality of food, environment and service of restaurants, and the volume of online consumer reviews are positively associated with the online popularity of restaurants; whereas editor reviews have a negative relationship with consumers’ intention to visit a restaurant’s webpage.”	eWOM +Restaurant image	The findings will help hospitality researchers and practitioners better understand the impact of electronic word-of-mouth on purchase decisions.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Zhang & Cole, 2016)	“Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts”	Tourism Management	543	Penalty-Reward Contrast Analysis, Content analysis	three-factor theory of customer satisfaction	This study captures the real-life tourist service evaluation criteria with improved accuracy and reliability.	tourist's psychology	This study methods report the practical meaning of the recommended attribute development order or priority for lodging businesses' service management.
(Zhao, Lu, Liu, Lin, & An, 2018)	“Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China.”	Tourism Management	1250000	Mixed methods, Spatial pattern analysis, visualization, Fine-grained Travel Party Partition	N/A	“Aggregating travel parties according to the size, tourist movement patterns are compared across different party sizes from demographic, spatial and temporal aspects. the obtained insights can help the stakeholders in travel package improvement, connectivity enhancement among attractions, attraction planning and management, and personalized next-attraction recommendation.”	Tourist movement patterns and party size	“Travel package improvement for different-sized parties (Attraction managers), Adjustment of the schedule of vehicles on public tour lines, Activity planning and infrastructure construction improvement”.
(Padma & Ahn, 2020)	“Guest satisfaction & dissatisfaction in luxury hotels: An application of big data”	International Journal of Hospitality Management	800	Content analysis, critical incident technique	N/A	“The main themes of luxury hotel service quality include hotel-related attributes, room-related attributes, staff-related attributes, travel-related attributes, and possible outcomes. quality of rooms and interaction with employees have been determined as major drivers of customers word of mouth and revisit intentions”	tourist psychology	This study contributes with an empirical analysis of particular features of textual context and discussion of the concept of luxury service in the developing countries
(Zhao, Xu, & Wang, 2019)	“Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews”	International Journal of Hospitality Management	127,629	Descriptive statistics, regression analysis	Signal theory	This study found “a higher level of subjectivity and readability and a longer length of textual review lead to lower overall customer satisfaction and a higher level of diversity and sentiment polarity of textual review leads to higher overall customer satisfaction. We also find that customers’ review involvement positively influences their overall satisfaction.”	Tourists satisfaction	It provides implications for hoteliers to better comprehend customer online review behavior and implement effective online review administration actions to use eWOM and improve hotels’ performance.
(Cheng & Jin, 2019)	“What do Airbnb users care about? An analysis of online review comments”	International Journal of Hospitality Management	170,124	Text mining, sentiment analysis, visualization	N/A	Findings reveal “Airbnb users tend to evaluate their experience based on a frame of reference derived from past hotel stays. Three key attributes identified in the data include ‘location’, ‘amenities’ and ‘host’. Surprisingly, ‘price’ is not identified as a key influencer.”	Tourists satisfaction	This study can refine its rating items and also advance good practices and strategies to assist Airbnb and its accommodation providers with enhancements to its services.

Year	Title	Journal	Sample size	Analytical Method	Theoretical Framework	Main findings	Issue Under discussion	Practical Implications/ Contributions
(Ahani, Nilashi, Ibrahim, Sanzogni, & Weaven, 2019)	“Market segmentation and travel choice prediction in Spa hotels through TripAdvisor’s online reviews”	International Journal of Hospitality Management	4930	text mining, Clustering, SOM, HOSVD, and CART techniques, visualization	N/A	Findings confirm that “the proposed hybrid machine learning methods can be implemented as an incremental recommendation agent for spa hotel/resort segmentation through effectively utilizing ‘big data’ procured from online social media contexts”.	Customer segmentation via online reviews	The results of this research will advantage “hotel and travel managers, as well as tourist social media sellers, to better understand customer preferences and behaviors and evaluative assessments of each customer segment to ensure targeted and efficient marketing spend”.
(Ma, Xiang, Du, & Fan, 2018)	“Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning”	International Journal of Hospitality Management	24,960 and 3064	Recurrent neural networks, Descriptive analysis, deep learning models, machine learning techniques	N/A	Findings show that “deep learning models were more useful in predicting review helpfulness than other models. While user-provided photos alone did not have the same impact as review texts, combining review texts and user-provided photos produced the highest performance.”	information value of online hotel review	This study offers implications for the” theoretical development of analytics research related to user-generated content in hospitality and tourism. Website developers and marketers could develop specific means to encourage photo-sharing behavior to enhance user interactions and experiences in an online community setting”.
(Hu & Trivedi, 2020)	“Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content”	International Journal of Hospitality Management	111,986	Text mining, Repertory grid analysis	N/A	“The evaluation of detected brand attributes involves both customer preferences and perceptual performance uniquely in our research. The former indicates the importance of brand attributes that are used to identify key competitive groups and dimensions, while the latter implies the performance of brand attributes to uncover the competitive strategy for detected competitive groups.”	Brand positioning and competitive landscapes	The contents of this study aware specialists and marketers in developing brand positioning strategies to fight competitors within and across hotel categories.
(Jang & Moutinho, 2019)	“Do price promotions drive consumer spending on luxury hotel services? The moderating roles of room price and user-generated content”	International Journal of Hospitality Management	1019 reviews, 21,281 sentences and 10,194 noun words	Text mining, Descriptive and regression analysis	cue-utilization theory	The results indicate that, “overall, price promotion negatively influences consumer spending on luxury hotel services and its negative effect is strengthened when the room is higher-priced or the valence of UGC is high. Furthermore, a larger volume of intrinsic attribute-related UGC–amenity and location–with price promotion leads to more consumers spending than a larger volume of extrinsic attribute-related UGC–food and staff.”	Price, promotion and consumer spending	This study provides useful managerial implications in the field of luxury hotel marketing. To increase luxury traveler spending, hotel managers should allocate more promotional expenditures in an effective way by focusing on less expensive offerings.

(Van der Zee, Bertocchi, & Vanneste, 2020)	Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management	Current Issues in Tourism	NA	Spatial correlation, hotspot analysis, focus group, cluster mapping	N/A	This focused on tourist behavior by analyzing restaurant reviews in four art cities of Belgium. The findings suggest that tourists like to visit mature destinations and try to find a facility near to the main attractions. The success of a restaurant partly depends on its location. Tourist behavior may be influence by some reviews on the web.tourists might show herd behavior by going where the masses go	Tourist behavior and travel pattern	This study offers little practical implication; however, it can be used for knowledge-based policies and planning. The information of this study suggests that facilities should be developed near the main attraction.
(McCreary, Seekamp, Davenport, & Smith, 2020)	Exploring qualitative applications of social media data for place-based assessments in destination planning	Current Issues in Tourism	194 photos	Content analysis	N/A	Determined key components of destination image by UGC data analysis. This study suggests that photographs are reliable indicators of visitors' destination image.	Destination image	Inform place-based management and marketing
(Park, Kang, Choi, & Han, 2020)	Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews	Current Issues in Tourism	105,126	Sentiment analysis	N/A	By comparing one-time visitors and re-visitors, our analysis shows that the feedback reviews of re visitors tend to (i) contain more words in a sentence and (ii) reveal more positive/negative sentiments than those of one-time visitors. On the other hand, the feedback reviews of one-time visitors tended to include more analytical and anxious words than those of re-visitors.	Revisit intentions	The authors suggest no practical implications of the study. However, from this study, hoteliers can benefit from studying the reviews to develop some knowledge about revisiting intentions. Besides, it would be helpful in co-creation services and products by studying negative reviews.
(Centobelli & Ndou, 2019)	Managing customer knowledge through the use of big data analytics in tourism research	Current Issues in Tourism	109 articles	Citation Network Analysis, content analysis	N/A	The findings highlight the research gaps of big data in tourism and ask for questions such as the impacts of big data on the performance of firms, new emerging business models, impacts of big data on current business models, and implications of big data in tourism sectors.	Use of big data in tourism	This study does not offer practical implications; however, this study is valuable for understanding the concept of big data and its relationship to performance.
(Villamediana-Pedrosa, Vila-Lopez, & Küster-Boluda, 2019)	Secrets to design an effective message on Facebook: an application to a touristic destination based on big data analysis	Current Issues in Tourism	180 posts by brand Spain, 57526 reactions, 1368 commentaries	Content analysis	N/A	The findings suggest that the use of a messaging tool (video) and the use of information cues predict positive and negative engagements.	Destination branding and image	This study offers practical implications to destination managers on how to send a message to the audience and what medium can be used for positive engagements.
(Park, Kim, & Choi, 2019)	Application of social media analytics in tourism crisis communication.	Current Issues in Tourism	2353 and 1080 before and during comments IRMA	Social network analysis, Semantic network analysis	N/A	Public communication and media strategies in times of tourism crisis are very critical because the impact of the crisis can be disproportionate to the actual disruptive effects unless the communication strategies are well-planned and	Tourism crisis management, destination crisis management	Provide an analytical framework for the study of social media contents in the tourism and hospitality sector. This study provides several important implications to theme park managers and tourism organizations' crisis teams. Addresses problems faced by theme park

						effectively integrated to the overall crisis management practices		managers and provide strategies on how to handle it. During and after the crisis, visitors' information are of great value for theme park manager for co-creation and to respond time accurately to the visitor demand for a more valuable experience.
(Giglio et al., 2019)	Machine learning and point of interests: typical tourist Italian cities	Current Issues in Tourism	26,392 photos of six Italian cities	Cluster analysis through machine learning	N/A	Findings show that social media datasets are valuable data to understand tourist behavior and mobility within a location. The scope is to delineate famous or unpopular places and propose new touristic scenarios, highlighting how the social part covers the main role in the points of interest (POIs') recommendation process in the touristic field.	Tourists behavior	This study does not offer any specific practical implications; however, the study is useful for the cultural destinations managers. A deep study of the article may enable them to how to use big data information in promoting culture destinations.
(Nilashi et al., 2019)	A Hybrid Method with TOPSIS and Machine Learning Techniques for Sustainable Development of Green Hotels Considering Online Reviews	Sustainability	152 hotels, 17024 reviews	Multi-Criteria Decision Making (MCDM) and learning techniques, SOM, LDA, TOPSIS and ANFIS	N/A	From a marketing perspective, the results of our study demonstrated that green initiatives lead to greater consumer satisfaction. The level of satisfaction in green spa hotels is higher than in green non-spa hotels. There is a positive relationship between the quality of services in green hotels and their satisfaction level.	Tourist segmetation/ tourist satisfaction.	This study does not provide any practical implications. However, the content provides a tourist segmentation methodology. Besides, suggest hotel managers that customers are more inclined towards green hotels; hence, improvement in green services can enhance hotel business.
(Marine-Roig, 2019)	Destination image analytics through traveller-generated content	Sustainability	300,000	Content analysis	N/A	This study results and findings are wired. It does not provide any proper sense. However, reveling that restaurants are getting more reviews than hotels and the attractions.	Destination Image	The study does not mention any specific practical implications.
(Mariné-Roig, 2017)	Measuring Destination Image through Travel Reviews in Search Engines	Sustainability	387,414	Text mining and content analysis	N/A	This study focused on the spatial, temporal distribution of tourists. Evaluative dimension analysis has revealed a high degree of satisfaction among tourists with the main attractions, and affective dimension analysis has shown a massive use of positive adjectives. The analysis of cognitive and spatial components found a high concentration of attractions in the metropolis, constituting more than 90% of the entire region.	Destination image	Allow destination management organizations (DMOs) to compare various attractions or services from the same or other destinations, in addition to places, territorial brands, or whole regions. The temporal dimension also allows for an analysis of the evolution of the tourist destination over time and the change (or permanence) of images perceived during different years

## List of References

- Agustí, D. P. (2018). Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides. *Annals of tourism research*, 73, 103-115.
- 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.
- Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.
- Banerjee, S., & Chua, A. Y. (2016). In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management*, 53, 125-131.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing & Management*, 25(1), 1-24.
- Centobelli, P., & Ndou, V. (2019). Managing customer knowledge through the use of big data analytics in tourism research. *Current Issues in Tourism*, 1-22.
- Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, 76, 58-70.
- Chua, A., Servillo, L., Marcheggiani, E., & Moore, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, 57, 295-310.
- Deng, N., & Li, X. R. (2018). Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. *Tourism Management*, 65, 267-278.
- Duverger, P. (2013). Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis. *Journal of Travel Research*, 52(4), 465-478.
- e Silva, F. B., Herrera, M. A. M., Rosina, K., Barranco, R. R., Freire, S., & Schiavina, M. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, 101-115.
- Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52, 498-506.
- Ganzaroli, A., De Noni, I., & van Baalen, P. (2017). Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. *Tourism Management*, 61, 501-510.
- Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, 72, 306-312.
- 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.
- Hallo, J. C., Beeco, J. A., Goetcheus, C., McGee, J., McGehee, N. G., & Norman, W. C. (2012). GPS as a method for assessing spatial and temporal use distributions of nature-based tourists. *Journal of Travel Research*, 51(5), 591-606.
- Hernández, J. M., Kirilenko, A. P., & Stepchenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Annals of tourism research*, 73, 35-47.
- Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management*, 74, 276-289.
- Hu, F., & Trivedi, R. H. (2020). Mapping hotel brand positioning and competitive landscapes by text-mining user-generated content. *International Journal of Hospitality Management*, 84, 102317.
- Huang, D., Coghlan, A., & Jin, X. (2020). Understanding the drivers of Airbnb discontinuance. *Annals of tourism research*, 80, 102798.
- Hunter, W. C. (2013). China's Chairman Mao: A visual analysis of Hunan Province online destination image. *Tourism Management*, 34, 101-111.
- Jang, S., & Moutinho, L. (2019). Do price promotions drive consumer spending on luxury hotel services? The moderating roles of room price and user-generated content. *International Journal of Hospitality Management*, 78, 27-35.
- Jin, C., Cheng, J., & Xu, J. (2018). Using user-generated content to explore the temporal heterogeneity in tourist mobility. *Journal of Travel Research*, 57(6), 779-791.
- Kim, Y., Kim, C.-k., Lee, D. K., Lee, H.-w., & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*, 72, 249-256.
- Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi 2014 Olympics on Twitter: Perspectives of hosts and guests. *Tourism Management*, 63, 54-65.
- Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management*, 72, 400-410.
- Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., & Shirayama, Y. (2020). Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. *Tourism Management*, 77, 104010.
- Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. R. (2015). Identifying emerging hotel preferences using emerging pattern mining technique. *Tourism Management*, 46, 311-321.
- Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: evidence from text mining of online reviews. *Tourism Management*, 66, 38-46.

- Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: Classification of satisfaction attributes. *Tourism Management*, 33(3), 702-712.
- Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *International Journal of Hospitality Management*, 71, 120-131.
- Mariani, M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, 54, 321-343.
- Marine-Roig, E. (2019). Destination image analytics through traveller-generated content. *Sustainability (Switzerland)*, 11(12). doi: 10.3390/su10023392
- Mariné-Roig, E. (2017). Measuring destination image through travel reviews in search engines. *Sustainability*, 9(8), 1425.
- Martinez-Torres, M., & Toral, S. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403.
- McCreary, A., Seekamp, E., Davenport, M., & Smith, J. W. (2020). Exploring qualitative applications of social media data for place-based assessments in destination planning. *Current Issues in Tourism*, 23(1), 82-98.
- Mellinas, J. P., & Nicolau, J. L. (2020). Let's hook up fast! Hotel reviews and Wi-Fi flaws. *Annals of tourism research*, 80, 102842.
- Moghavvemi, S., Ormond, M., Musa, G., Isa, C. R. M., Thirumoorthi, T., Mustapha, M. Z. B., & Chandy, J. J. C. (2017). Connecting with prospective medical tourists online: A cross-sectional analysis of private hospital websites promoting medical tourism in India, Malaysia and Thailand. *Tourism Management*, 58, 154-163.
- Nieto, J., Hernández-Maestro, R. M., & Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments. *Tourism Management*, 45, 115-123.
- Nilashi, M., Mardani, A., Liao, H., Ahmadi, H., Manaf, A. A., & Almkadi, W. (2019). A hybrid method with TOPSIS and machine learning techniques for sustainable development of green hotels considering online reviews. *Sustainability*, 11(21), 6013.
- Padma, P., & Ahn, J. (2020). Guest satisfaction & dissatisfaction in luxury hotels: An application of big data. *International Journal of Hospitality Management*, 84, 102318.
- Pantano, E., Priporas, C.-V., & Stylos, N. (2017). 'You will like it!' using open data to predict tourists' response to a tourist attraction. *Tourism Management*, 60, 430-438.
- Park, S. B., Kim, H. J., & Ok, C. M. (2018). Linking emotion and place on Twitter at Disneyland. *Journal of Travel & Tourism Marketing*, 35(5), 664-677.
- Park, S. B., Ok, C. M., & Chae, B. K. (2016). Using Twitter data for cruise tourism marketing and research. *Journal of Travel & Tourism Marketing*, 33(6), 885-898.
- Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55(2016), 16-24.
- Phillips, P., Zigan, K., Silva, M. M. S., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, 50, 130-141.
- Rahmani, K., Gnoth, J., & Mather, D. (2018). Hedonic and eudaimonic well-being: A psycholinguistic view. *Tourism Management*, 69, 155-166.
- Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, 57, 202-212.
- Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13-25.
- Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management*, 36, 590-601.
- Stringam, B. B., & Gerdes Jr, J. (2010). An analysis of word-of-mouth ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management*, 19(7), 773-796.
- Su, L., Stepchenkova, S., & Kirilenko, A. P. (2019). Online public response to a service failure incident: Implications for crisis communications. *Tourism Management*, 73, 1-12.
- Talón-Ballester, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2018). Using big data from customer relationship management information systems to determine the client profile in the hotel sector. *Tourism Management*, 68, 187-197.
- Van der Zee, E., Bertocchi, D., & Vanneste, D. (2020). Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management. *Current Issues in Tourism*, 23(2), 175-196.
- Versichele, M., De Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, 44, 67-81.
- Villamediana-Pedrosa, J. D., Vila-Lopez, N., & Küster-Boluda, I. (2019). Secrets to design an effective message on Facebook: an application to a touristic destination based on big data analysis. *Current Issues in Tourism*, 22(15), 1841-1861.
- Villamediana, J., Küster, I., & Vila, N. (2019). Destination engagement on Facebook: Time and seasonality. *Annals of tourism research*, 79, 102747.

- Vu, H. Q., Li, G., & Law, R. (2020). Discovering highly profitable travel patterns by high-utility pattern mining. *Tourism Management*, 77, 104008.
- Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46, 222-232.
- Vu, H. Q., Luo, J. M., Ye, B. H., Li, G., & Law, R. (2018). Evaluating museum visitor experiences based on user-generated travel photos. *Journal of Travel & Tourism Marketing*, 35(4), 493-506.
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65.
- Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management*, 32(1), 88-97.
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.
- Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1-12.
- Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, 55, 57-69.
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180-182.
- Zhang, Y., & Cole, S. T. (2016). Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts. *Tourism Management*, 53, 13-27.
- Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111-121.
- Zhou, L., Ye, S., Pearce, P. L., & Wu, M.-Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1-10.
- Zhu, Y., Cheng, M., Wang, J., Ma, L., & Jiang, R. (2019). The construction of home feeling by Airbnb guests in the sharing economy: A semantics perspective. *Annals of tourism research*, 75, 308-321.

Appendix B

Tourism and hospitality big data articles contribution to theory and philosophy

Authors	Research questions	Hypotheses	Use of theory	Contribution to theory	Contribution to philosophy
Hou, Cui, Meng, Lian, and Yu (2019)	✓	✗	Social network theory (SNT)	The online reviews thematic words are treated as nodes, and the semantic relationships among nodes are regarded as connections. Together nodes and connections construct a social network. This research has expanded the application and scope of SNT.	This study identifies the problems by raising prior questions (related to ontology – philosophy) and uses the SNT to guide the process and arrive at solutions (new knowledge – epistemology) for those problems.
Huang, Coghlan, and Jin (2020)	✓	✗	Innovation diffusion theory (IDT)	Identified online and offline services issues with 15 sub-categories; and attempted to connect the elements of innovation diffusion theory.	Identified factors of discontinuance with Airbnb and used IDT and attempted to connect the different elements of IDT.
Jang and Moutinho (2019)	✓	✓	cue-utilization theory (CUT)	Identified multiple cues that influence customer evaluation of product quality by using the hotels’ big data. Besides, providing strong support to CUT	Questions are raised, CUT is used to guide the process and reach solutions (what – is ontology, and broadening the scope of Cut is epistemology
Kubo et al. (2020)	✗	✗	Consumer theory	We failed to identify how this study contributed to the mentioned theory, nor did the authors discuss the theoretical contributions.	No prior problem identification or research questions
Ganzaroli, De Noni, and van Baalen (2017)	✓	✗	Theory of the vicious circle	The findings satisfy the author’s proposed conceptual framework; however, we did not find any argument or content proving that this study contributes to the mentioned theory.	This study defined prior research questions. However, this study did not mention any mechanism to combine the different identified elements in a conceptual or theoretical frame.

Duverger (2013)	✓	✓	Social judgment theory	The findings satisfy the authors' proposed UGC and hotel market shares	This study developed hypotheses and guided the process through social judgment theory for hypotheses testing; the author provided inductive reasoning to support the findings.
Agustí (2018)	✗	✗	Destination image theory	The authors did not claim any theoretical contributions nor did we identify how this study is contributing to the theory.	The study did not posit any prior question and led to the data-driven conclusion instead of based on any theory or logic.
Liu, Schuckert, and Law (2018)	✗	✓	Goal-setting theory, status hierarchy theories	The findings articulate the components of goal-setting and status hierarchy theories.	This study properly developed the hypotheses by reviewing the available literature within the context of the mentioned theories. Besides, also satisfy the condition of epistemology by creating new knowledge; however, sample representations become a serious epistemological issue.
Ma, Kirilenko, & Stepchenkova, (2020)	✓	✓	Serious leisure theory/ SIT model	The study successfully tested the proposed hypothesis in the context of SIT and found that tourists visitation rate between two groups differs on travel distance, gender, income, and education	This study properly articulates the component of SIT such as cost sensitivity, risk sensitivity, and time efficiency. The authors noted \
Mellinas and Nicolau (2020)	✓	✓	Herzberg's motivation-hygiene theory	The study identified the factor that leads to customer dissatisfaction with WiFi services in hotels and articulated them under the context of the mentioned theory	This study successfully tested the posit hypotheses, and the findings contributed to Herzberg's motivation-hygiene theory

Park & Nicolau, 2015	✓	✗	Theory of information diagnosticity	This study applied the theory of information diagnosticity, which refers to the extent to which a consumer believes the product information is helpful to understand and evaluate purchase alternatives. Where the study findings satisfy the mentioned theoretical paradigms.	This study has a conceptual which is tested under the umbrella of the mentioned theory. Besides, all the variables are adequately defined and operationalized.
Park & Pan, 2018	✓	✗	Buying funnel theory and gravity model	This article certainly contributed to the theory as well viable for identifying potential market segments.	Buying funnel theory connects the aggregated gravity model with an individual visitor's decision-making.
Rahmani, Gnoth, and Mather (2018)	✓	✗	self-determination theory, cognitive theory	This study contributes to the mentioned theories through the developed conceptual model.	This study applied a new methodological approach to generalise experiences and compare the effect of destination countries on well-being while allowing more theoretically founded generalisations from qualitative data. The study philosophically articulates the different components of the conceptual model and offers new knowledge.
Stepchenkova and Zhan (2013)	✓	✗	Theory of hermeneutic circle of destination representation	This study partially supports the theory of the hermeneutic circle of destination representation.	The study is looking to answer specific prior questions; however, it is felt that the component of the mentioned theoretical framework is marginally articulated.

Su, Stepchenkova, and Kirilenko (2019)	✓	✗	Situational Crisis Communication Theory and Attribution Theory	The study's findings provide full support to the conceptual premises of situational crisis theory which states that crisis managers should match strategic crisis responses to the level of crisis responsibility and the reputational threat posed by a crisis. Besides, it also identifies the causes of the crisis, thus, meeting the assumptions of Attribution theory.	Philosophically this study starts with a proper research question (ontology). Then epistemologically provide evidence and reasoning to connect the findings to the mentioned theories and add new knowledge by applying the advanced methodology. Further, data cleaning and screening has been performed.
Xiang, Schwartz, Gerdes Jr, and Uysal (2015)	✓	✗	Herzberg's motivation-hygiene theory	The findings of this study contribute to the mentioned theory. Besides, the authors' suggested that theory should be used to guide big data analytics and attested that domain knowledge is critical in guiding the data processing and analytical process before reaching the point where meaningful relationships emerge.	To the best of our knowledge, it is the first study that emphasized using the conceptual and theoretical framework to guide big data collection and analysis to reach a meaningful and critical conclusion. Whereas philosophy raises questions, specific theoretical frames are used to answer them logically.
Xu and Li (2016)	✓	✓	Multi-attribute theory/ Herzberg's motivation-hygiene theory	The authors noted that the findings of this study support Multi-attribute theory/ Herzberg's motivation-hygiene theory and confirm expectation-disconfirmation models.	There is no doubt that this study attempted to identify the determinants of satisfaction and dissatisfaction in the hotel industry; however, the contents of this article missed the mechanism of how these different factors satisfy the overall premises of various theories.
Zhang and Cole (2016)	✗	✗	three-factor theory of customer satisfaction	This study has categorized the hotel attributes into three large factors, i.e., basic, excitement, and performance. The findings also reveal the relative critical different role of each factor in customer satisfaction. Thus, providing full support to the three-factor theory of customer satisfaction.	This study added value to the tourism and hospitality literature by using an advanced methodology, Penalty-Reward Contrast Analysis (PRCA), by using big data.

Zhang, Zhang, & Yang, (2016)	✓	✓	cognitive-processing capacity theory	The hypotheses have been supported by literature; however, the discussion section articulate the components of cognitive processing capacity theory; hence, this paper contributes to the theory.	This study has identified the research problem and, with the help of literature, developed hypotheses to test the influence of expert review on travelers' online ratings. However, the study provides reasoning for connecting the various finding within a certain framework.
Zhao, Xu, and Wang (2019)	✓	✓	Signal theory	The findings of the study support the theoretical premises of the signal theory. The author tested the hypotheses, and the results confirm the signal behavior between two parties where information asymmetry exists. Thus, this study shows what customers write (i.e., the contents) and how customers write (i.e., the linguistic style) signals their satisfaction or dissatisfaction with hotel product and service attributes.	This study developed hypotheses with the help of available literature under the theoretical frame of signal theory and designed a comprehensive research design model to guide the data collection and analysis.

#### List of References:

- Agustí, D. P. (2018). Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides. *Annals of tourism research*, 73, 103-115.
- Duverger, P. (2013). Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis. *Journal of Travel Research*, 52(4), 465-478.
- Ganzaroli, A., De Noni, I., & van Baalen, P. (2017). Vicious advice: Analyzing the impact of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice. *Tourism Management*, 61, 501-510.
- Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management*, 74, 276-289.
- Huang, D., Coghlan, A., & Jin, X. (2020). Understanding the drivers of Airbnb discontinuance. *Annals of tourism research*, 80, 102798.
- Jang, S., & Moutinho, L. (2019). Do price promotions drive consumer spending on luxury hotel services? The moderating roles of room price and user-generated content. *International Journal of Hospitality Management*, 78, 27-35.
- Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., & Shirayama, Y. (2020). Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. *Tourism Management*, 77, 104010.

- Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: evidence from text mining of online reviews. *Tourism Management*, 66, 38-46.
- Mellinas, J. P., & Nicolau, J. L. (2020). Let's hook up fast! Hotel reviews and Wi-Fi flaws. *Annals of tourism research*, 80, 102842.
- Rahmani, K., Gnoth, J., & Mather, D. (2018). Hedonic and eudaimonic well-being: A psycholinguistic view. *Tourism Management*, 69, 155-166.
- Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management*, 36, 590-601.
- Su, L., Stepchenkova, S., & Kirilenko, A. P. (2019). Online public response to a service failure incident: Implications for crisis communications. *Tourism Management*, 73, 1-12.
- Xiang, Z., Schwartz, Z., Gerdes Jr, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120-130.
- Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management*, 55, 57-69.
- Zhang, Y., & Cole, S. T. (2016). Dimensions of lodging guest satisfaction among guests with mobility challenges: A mixed-method analysis of web-based texts. *Tourism Management*, 53, 13-27.
- Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111-121.

**Table 1** Inclusion criteria for selecting articles

<i>No.</i>	<i>Criteria</i>	<i>Reason for inclusion</i>
1.	Type of publication	Only journal articles were included.
2.	Type of journals	Particular tourism and hospitality journals in the SSCI and SCIE lists, generally the articles, are of a higher quality. Besides, articles from multi-disciplinary journals were also included if the articles are focused on contribution to the knowledge in tourism and hospitality.
3.	Theoretical and review articles	These papers were included because they offer foundations for analysis, interpretation, summarization, and integration of evidence.
4.	Qualitative and quantitative articles	These papers provide empirical evidence for the relationship of various phenomena with each other.
5.	Perspective	Only those articles were included that focused on the methodology, research design, or the interplay between various tourism and hospitality aspects.
6.	Implications	Those articles were included, which offer some practical implications in the field of tourism and hospitality.
7.	Expert opinion	Experts' opinions were taken for final inclusion and exclusion.

**Table 2** Methodological and Analytical Techniques

<i>Analytical Method</i>	<i>Percentage</i>
Methodological Triangulation	28%
Data Mining and Analytics	11%
Visual Analytics	11%
Content Analysis	10%
Descriptive, Regression, ANOVA	8%
Social Networking	5.1%
Spatial Analysis	5.2%
Cluster Analysis	5.4%
Sentiment Analysis	5.3%
Forecasting and Economic Modeling	4.1%
Artificial Neural Network	2.2%
Case Study Analysis	1%
Machine Learning Approaches	1.5%
Flow Analysis	1.2%
Factor Analysis	1%

**Table 3** Research Focus Major and Sub-themes

<i>Major themes</i>	<i>Sub-themes</i>
Big data and artificial intelligence in tourism and hospitality	Includes big data, value creation, information value, business intelligence, and decision support, supply chain management
Conventional tourism	Includes adventure, recreation, museums, attractions, volunteerism, Mega sports events, festivals, eco-tourism, medical tourism, special interest tourism, theme park & vacations tourism management
Tourism marketing	Includes segmentation, promotion, branding, demand forecasting, automated product ranking, consumer spending, positioning, pricing, trends, competitive landscape
Destination management	Includes destination image, branding, smart destinations, crisis management, tourists hot spot, destination marketing, destination learning, seasonality, climate
Electronic Word Of Mouth (eWOM)	Includes eWOM and hotel performance, restaurant quality, reputation management, hotel share, online experience, WIFI influence, online hotel rating, smart traveling, online rating, tourists' decision
Big data and knowledge management	Includes client profiling, customer relationship management, profitable traveling pattern identification, travel knowledge co-production, artificial intelligence, and business analytics, online travel queries, the value of online information, emerging trends, learning destination and customers, customer knowledge and value creation
Other big data themes	Includes Airbnb management, big data use in airlines, travel planning, google analytics and trust perception

**Table 4** Data Types and Research Focus

<i>Data Types and Research Focus</i>	<i>Relevant Study</i>
<b>Reviews</b>	
Tourist satisfaction attributes	(Lu and Stepchenkova, 2015)
Guest experience	(Xiang <i>et al.</i> , 2015)
eWOM impacts on hotels and restaurants	(Hu <i>et al.</i> , 2017; Zhang <i>et al.</i> , 2010)
Attraction management	(Guo <i>et al.</i> , 2017)
<b>Blogs</b>	
Mapping tourist locations	(Yuan <i>et al.</i> , 2016)
Tourist sentiments	(Philander and Zhong, 2016)
Where tourists like to visit	(Xu, <i>et al.</i> , 2015)
<b>Photos</b>	
Tourist behavior	(Vu <i>et al.</i> , 2015)
Tourist origin	(Da Rugna <i>et al.</i> , 2012)
Point of interest	(Zhou <i>et al.</i> , 2015)
Traveling routes	(Lu <i>et al.</i> , 2010)
Traveling duration and stay at tourism destinations	(Popescu and Grefenstette, 2009; Lee <i>et al.</i> , 2014)
<b>IoT</b>	
GPS, mobile roaming, WIFI, and Bluetooth	(Shoval <i>et al.</i> , 2015)
Spatial behavior, temporal behavior, and spatial-temporal behavior	(Birenboim <i>et al.</i> , 2013; Grinberger <i>et al.</i> , 2014)
<b>Web searching (Google, Baidu)</b>	
Forecasting and prediction	(Peng <i>et al.</i> , 2017)
Tourists' online behavior and decision making	(Li <i>et al.</i> , 2016)