



**Universitat de les  
Illes Balears**

Facultat de Turisme

**Memòria del Treball de Fi de Grau**

# Analyzing Big Data to discover new touristic products

Julia Johanna Edith Kraus

**Grau de Turisme**

Any acadèmic 2019-20

DNI de l'alumne: L5JK047X7

Treball tutelat per Francisco Rejón Guardia  
Departament d'economia de l'empresa

Paraules clau del treball:

Big Data, social media, Twitter, alternative tourism, new touristic products

---

**TABLE OF CONTENTS**

<b>List of abbreviations .....</b>	<b>IV</b>
<b>List of tables .....</b>	<b>V</b>
<b>List of figures.....</b>	<b>VI</b>
<b>Abstract.....</b>	<b>VIII</b>
<b>1 Introduction .....</b>	<b>1</b>
<b>2 Big Data .....</b>	<b>2</b>
2.1 The three primary sources of Big Bata in tourism.....	2
2.1.1 User Generated Content .....	2
2.1.2 Device data .....	4
2.1.3 Transaction data .....	4
<b>3 Alternative tourism .....</b>	<b>4</b>
3.1 Sustainable tourism.....	5
3.2 Product diversification .....	5
3.3 Alternative touristic products .....	5
<b>4 Social Media .....</b>	<b>5</b>
4.1 Social Media in tourism .....	6
4.1.1 Word-of-Mouth .....	6
4.1.2 Social media impact on customer decision-making.....	6
4.1.3 Demographic differences .....	6
4.2 Benefits for tourism and research.....	7
4.3 Twitter.....	7
<b>5 Analyzing Big Data.....</b>	<b>8</b>
5.1 Big Data analytics approach.....	8
5.2 Text mining.....	8
5.2.1 Data pre-processing techniques.....	9
5.2.2 Pattern discovery .....	9

---

5.2.3 Keyword selection .....	9
5.3 Challenges in Big Data Analysis.....	10
5.3.1 Research language and word selection .....	10
5.3.2 Data Quality .....	10
5.3.3 Data Costs .....	10
5.3.4 Data Privacy.....	11
5.3.5 Solution Approach.....	11
5.4 Usefulness of Big Data Analysis.....	11
5.4.1 Decision-making.....	11
5.4.2 Developing new business opportunities .....	12
<b>6 Twitter analysis .....</b>	<b>12</b>
6.1 Keyword selection .....	12
6.2 Program used.....	12
6.3 Techniques used .....	13
6.3.1 Data cleaning .....	13
6.3.2 Statistical analysis.....	13
6.3.3 Manual analysis .....	13
6.4 Discoveries.....	13
6.4.1 Languages analyzed.....	13
6.4.2 Touristic activities.....	14
6.5 Create new touristic products .....	14
6.6 Future directions for analyzing data .....	15
<b>7 Conclusion.....</b>	<b>15</b>
<b>List of cited literature .....</b>	<b>17</b>

---

**LIST OF ABBREVIATIONS**

AI	Artificial Intelligence
GPS	Global Positioning System
POST	Part-of-Speech Tagging
RFID	Radio-frequency identification
UGC	User Generated Content
USP	Unique Selling Point
WOM	Word-of-Mouth

---

**LIST OF TABLES**
**Table 1** List of Keywords

<b>English Keywords</b>		
active tourism	ecologic tourism	mountain biking
alternative tourism	ecotourism	mountaineering
alternative travel	ecotravel	mountain climbing
bike tourism	future tourism	nature tourism
biking	green tourism	nature travel
biking tourism	green travel	new tourism
cycling tourism	healthy tourism	new travel
cycling	hiking	sport tourism
challenge ciclista mallorca	hiking tourism	sustainable tourism
ecofriendly tourism	innovative tourism	sustainable travel
ecofriendly travel	mountain bike	
<b>English Hashtags</b>		
#activetourism	#cyclingspain	#mountainbiking
#bicycle	#discovermallorca	#mountainclimbing
#biketour	#ecofriendlytourism	#naturetourism
#biking	#ecofriendlytravel	#newtourism
#bikingmallorca	#ecotourism	#roadbiking
#challengeciclistamallorca	#ecotravel	#roadcycling
#cycling	#greentourism	#sustainabletourism
#cyclingibiza	#greentravel	#sustainabletravel
#cyclinglife	#healthytourism	#tourismfortomorrow
#cyclingmallorca	#hiking	#trekking
#cyclingmenorca	#hikingtour	
<b>Spanish Keywords</b>		
bicicleta	senderismo	turismo deportivo
bicicleta de montaña	turismo activo	turismo en bicicleta
caminata	turismo de senderismo	turismo saludable
<b>Spanish Hashtags</b>		
#ciclismo	#senderismo	
#bicicleta	#vueltademallorca	
<b>German Keywords</b>		
aktiver tourismus	mountainbike	radfahrtourismus
fahrradtourismus	radeln	sporttourismus
gesunder tourismus	radfahren	wandertourismus
<b>German Hashtags</b>		
#bergsteigen	#mountainbiken	#wandertour
#fahrrad	#radeln	#wandertourismus
#fahrradfahren	#radfahren	#wanderwege
#fahrradmallorca	#radtour	
#gesundheitstourismus	#wandern	





---

**ABSTRACT**

The topic of Big Data is gaining more and more attention in several business fields. For the tourism and hospitality industry it offers many opportunities and gathered the interest of researchers. With this study I want to provide an overview about the opportunities of Big Data analysis for companies, organizations and research in the tourism industry. The main focus will be on the explanation of Big Data analysis techniques and how the analysis can contribute to the decision-making process in companies and organizations and help in the development of new business opportunities. This thesis will be finalized with a practical analysis of Big Data collected from social media, aimed to discover new alternative touristic products on the Balearic Islands using geo-tagged Twitter data from 2015 to 2017.

---

## 1 INTRODUCTION

Applying Big Data in tourism research is still at a young stage considering that it has been under examination just since 2007 and has a rather small amount of 30 annual research publications (J. Li, Xu, Tang, Wang, & Li, 2018). But there is a rapid growth in interest regarding this topic and the number of research articles increases every year (J. Li et al., 2018).

The massive amount of Big Data in social media platforms has a growing impact on the tourism and hospitality industry (Xiang & Gretzel, 2010). Analyzing this data offers new approaches to understand consumer behavior, improve marketing operations and consumer relationship and solve real life problems in the industry (J. Li et al., 2018; Philander & Zhong, 2016; Xiang, Schwartz, Gerdes, & Uysal, 2014). It provides new approaches to optimize processes and also to create new business opportunities (Robles & Barragán, 2017).

I chose this topic because technology and alternative tourism are both topics with increasing importance for tourism. The internet surrounds us in nearly every part of our life and has profoundly changed the way travel is planned and consumed (Buhalis & Law, 2008). Big Data analysis is a quite new approach and already offers many opportunities for improvement in the industry.

The goal for this study was to provide an overview on how Big Data analysis techniques work and to examine how an analysis can help in the decision-making process of different businesses areas and in the discovery of new business opportunities. As a practical approach, I analyzed geo-tagged Twitter data to discover new alternative touristic products in the Balearic Islands.

---

## 2 BIG DATA

There is no consistent definition of the term Big Data yet, but it can be described as a set of data, too big for traditional systems to analyze in an acceptable time period (Gandomi & Haider, 2015; Laney, 2001). The first definition describes Big Data as the 3V of Volume, Variety and Velocity, focusing on its technological perspective (Laney, 2001; McAfee & Brynjolfsson, 2012). As the overall Volume of data increases, each single data point loses in value (Laney, 2001). In addition, Variety increases, offering more and more different data formats, data structures and data semantics in incompatible, disorganized and unallocated forms (Laney, 2001). Thirdly, the Velocity, or speed of points of interaction of Big Data will result in a form of competitive distinction (Laney, 2001). This definition later got extended to 4V, adding Value (Gantz & Reinsel, 2011) and then Veracity and Variability, forming 6V (McAfee & Brynjolfsson, 2012). The extension results from the ongoing research in Big Data, showing that it can create significant value by offering a huge variety of data, being more accurately and functional (Manyika et al., 2011; McAfee & Brynjolfsson, 2012).

This value creation provided by Big Data is a key source and presents new ways of research (Vecchio, Mele, Ndou, & Secundo, 2018). It offers opportunities to analyze patterns, behaviors, social dynamics and the impact of social networks on individuals, groups and organizations (George, Haas, & Pentland, 2014). It can also be used as a predictive tool to forecast the probability of a certain event occurring (George et al., 2014). In tourism, Big Data can increase tourists' satisfaction, identify new and unexplored market opportunities and create profitable businesses (Vecchio et al., 2018).

The power and influence of Big Data should not be underestimated: A single message on social media platforms like Twitter or Instagram can cause a catalytic chain reaction and damage or profit in billion range and become a hugely discussed topic within hours (George et al., 2014). It has huge influence on making decisions in real time, is able to transform whole societies and communities and could change the way policy and research work in the near future (George et al., 2014).

### 2.1 The three primary sources of Big Bata in tourism

Big Data can be divided into 2 main types: physical world data, generated by sensors, scientific experiments, etc., and human society data collected from the internet, social network platforms and other forms of human communication and interaction (Jin, Wah, Cheng, & Wang, 2015).

Both data types are generated by many different sources including mobile and business transactions, web traffic, social media platforms and also purposely through sensor networks (George et al., 2014). In tourism research, three primary sources are relevant: user-generated content (UGC), device data and transaction data (J. Li et al., 2018).

#### 2.1.1 User Generated Content

UGC is mainly spread on Social Media platforms through online photo data and online textual data, sharing the users feelings, experiences, sentiments and moods (J. Li et al., 2018; Xiang, Du, Ma, & Fan, 2017). For tourism research, this type of data provides information to understand customer behaviors, their attitude towards tourism products and their satisfaction (J. Li et al., 2018; Xin Li & Law,

2020). It has huge influence on customers' view of a brand's image and reputation and their decision-making process when acquiring a product or service (Browning, So, & Sparks, 2013; Leung, Law, Hoof, & Buhalis, 2013; Vermeulen & Seegers, 2009; Ye, Law, & Gu, 2009; Zhang, Ye, Law, & Li, 2010).

UGC changes the way people purchase products and services online and it is often used for holiday planning and information gathering in tourism, where details about travel offers play an important role. (Chung & Buhalis, 2008; Zeng & Gerritsen, 2014). The main reason for its importance is the fact that touristic products and services are intangible and cannot be rated before consumption (Dellaert, 2000; Rabanser & Ricci, 2005; Senecal & Nantel, 2004). Especially when visiting a new or an international destination, it is used by many travelers as a source to gain knowledge about the destination beforehand (Simms, 2012).

Opinions expressed through UGC have a higher validity than information from traditional sources given by travel agencies, official websites, advertising in mass media or the destination itself (Fotis, Buhalis, & Rossides, 2012; Sarks, 2007). More than 20% of users perceive it as a trustworthy source (Sarks, 2007). In a certain degree, consumers even value information offered by UGC comparable to a friends' or family members' recommendation (Bray, Schetzina, & Steinbrink, 2006; Y. Wang, Yu, & Fesenmaier, 2002; K.-H. Yoo, Lee, Gretzel, & Fesenmaier, 2009).

#### **2.1.1.1 Online photo data**

Online photo data is increasingly known as a useful tool to create an image about a destination in the mind of potential tourists (Deng & Li, 2018). In addition, this type of UGC offers a lot of background information for researchers in the form of metadata, like photo and user ID, geographical information, the date the photo was taken and uploaded as well as textual information in the form of title and description (J. Li et al., 2018).

#### **2.1.1.2 Online textual data**

Online textual data in the form of social media, reviews and blogs offers insights to potential customers about the touristic product or service, partly forming their opinion about it and influencing their decision to either use it or not (Fronzetti Colladon, Guardabascio, & Innarella, 2019; Gavilan, Avello, & Martínez-Navarro, 2018; Hudson, Roth, Madden, & Hudson, 2015; Llodrà-Riera, Martínez-Ruiz, Jiménez-Zarco, & Izquierdo-Yusta, 2015). Reviews focus on tourist's opinion and point of view towards touristic products with the main goal of measuring their satisfaction (Xiang et al., 2014). Review data can be very important for touristic products like attractions and may help the management to improve constantly by analyzing the critique in this data (Fang, Ye, Kucukusta, & Law, 2016; Pearce & Wu, 2015). Blog data on the other hand provides information about tourist feelings and emotions, presenting the experience they made on their travel, which can be used in tourism recommendation and also for sentiment analysis (J. Li et al., 2018). The public exchange among tourists provided by these types of UGC can help to forecast tourism demand, customer behavior and needs, as well as identify unexplored markets (Hudson et al., 2015; Llodrà-Riera et al., 2015; Vecchio et al., 2018).

UGC is a useful tool in hospitality since it enables deeper insights into research problems, where traditional methods reached their limit in explaining them (Xiang et al., 2014). But online textual data has also a major disadvantage: it is

---

generated by users who can share their opinion about a product or service and might give false information or fake reviews (Schuckert, Liu, & Law, 2015b). Therefore, the reliability of this data type cannot be guaranteed and will influence the quality of the analyzed data and might change the results (Schuckert et al., 2015b).

### **2.1.2 Device data**

The temporal data passively recorded by devices is called device data or tracking data (Hardy et al., 2017; Shoval & Ahas, 2016). It includes GPS, Mobile Roaming, Bluetooth, WIFI, RFID and also meteorological data, particularly relevant for tourism, since weather and climate can play an important role in travel decisions (J. Li et al., 2018). Especially GPS data can have very high value, since it has the ability to not only track tourist locations with high accuracy, but can also be combined with survey data to segment tourists, analyze their behavior and give recommendations (East, Osborne, Kemp, & Woodfine, 2017; J. Li et al., 2018; Yoon, Zheng, Xie, & Woo, 2010).

### **2.1.3 Transaction data**

Operations like web searching, online booking and purchasing are transactions common in tourism (J. Li et al., 2018). This type of data tracked by search engines like Google gives deep insights into tourist online behavior, reflecting accurately public attention, interest and trends, and is therefore a powerful tool in predicting tourism demand (J. Li et al., 2018; Xiaoxuan Li, Wu, Peng, & Lv, 2016). As it is a main tool for information procurement, online search data reflects the intentions and behaviors of people and can be used to find patterns, observe their decision-making processes and even predict their future behavior (Xiaoxuan Li et al., 2016). Transaction data is not only useful for exploring and understanding the tourism market, but can also improve (online) marketing through Search Engine Optimization (SEO) and Search Engine Marketing (SEM) (J. Li et al., 2018).

## **3 ALTERNATIVE TOURISM**

Since the 1980s, alternative tourism has been a buzzword in the industry as an alternative to the widely discredited mass tourism (Butler, 1990). The main idea behind alternative tourism is to decrease the negative effects on the environment and the population in the destinations without diminishing its positive economic effects (Tezcan, 2004). In addition, it provides support to achieve a moral economy with social and environmental awareness, sustainable trading and fair working conditions (Gibson, 2010).

Alternative tourism is rapidly growing because the negative effects caused by mass tourism might affect the attractiveness of destinations in the future (Moscardo, 2001). While mass tourism concentrates on economic profit and develops rapidly, alternative tourism focuses more on slowly developing sustainability and a good relationship with locals and their economic and social needs (Tezcan, 2004). It's main objective is to maintain and save the social, environmental and historical assets destinations can offer visitors (Tezcan, 2004).

---

### **3.1 Sustainable tourism**

Sustainable tourism can be seen as the main goal for the development of tourism as it is described as the type of development that is able to fulfill the needs of the current generation, without ruling out the possibility for the future generations to fulfill their needs equally (Tezcan, 2004; Weaver, 2001). Touristic development is not possible without nature and instead of exploiting, it should focus on improving the natural attributes of the destination and contribute to its growth socially and economically (Tezcan, 2004). With the increasing awareness regarding the environment in touristic activities, mass tourism became a criticized topic and alternative forms of tourism received more attention (Tezcan, 2004).

### **3.2 Product diversification**

Alternative tourism is seen as a tool for product diversification since the changing consumer demand resulted in searching for new opportunities apart from the mass tourism basing on beaches, sea and sun (Kiliç & Kurnaz, 2010; Pektaş, 2018). The goal of achieving profit pursued by the mass tourism is replaced by focusing on connecting people and places in a pure and authentic way (E. Cohen, 1987; Germann Molz, 2013). Alternative forms of tourism attracts those growing amounts of tourists who prefer personalized, impressive experiences and a real and deep connection with the destination and its inhabitants (Germann Molz, 2013; McIntosh & Zahra, 2007; Steylaerts & Dubhghaill, 2012).

### **3.3 Alternative touristic products**

Alternative touristic products include all types of touristic activity excluding mass tourism (Pektaş, 2018). They are not dependent on seasonal climate and rely on the active participation and decision-making of the tourists, who are mainly smaller groups, families or individual travelers (Tezcan, 2004). Alternative touristic activities can be divided into several categories like health tourism, nature tourism, entertainment tourism, sport tourism or cultural and historical tourism (Tezcan, 2004).

## **4 SOCIAL MEDIA**

As social media and its users are constantly evolving and changing, there is no clear definition of the term (H. Cohen, 2020; Zeng & Gerritsen, 2014). Still, all definitions made for it so far agree on some characteristics (Zeng & Gerritsen, 2014):

Social media includes several internet-based tools, platforms and applications for creating and sharing UGC among individuals based on Web 2.0 and accordingly, it is dependent on information and communication technology (H. Cohen, 2020; Kaplan & Haenlein, 2010; Zeng & Gerritsen, 2014). It is interactive and works through peer-to-peer communication channels, therefore changing fundamentally how individuals, groups and organizations communicate (H. Cohen, 2020). This communication socially connects people and creates an entire environment of virtual online communities, built on users interactions and input, influencing their behavior and also their real life (H. Cohen, 2020; Zeng & Gerritsen, 2014).

---

## **4.1 Social Media in tourism**

The way travel is planned and also consumed by tourists and how they search, find and also trust information has changed radically with to the growth of social media (Buhalis & Law, 2008; Gretzel & Yoo, 2008; Hudson & Thal, 2013; Zeng & Gerritsen, 2014). Due to the ability of sharing travel experiences and exchanging tourism related information among individuals, social media and online reviews are an important source in travel planning and play a key role in the decision-making process of potential tourists (Gavilan et al., 2018; Zeng & Gerritsen, 2014).

### **4.1.1 Word-of-Mouth**

Individuals can share reviews, opinions and information about products, services, destinations and touristic suppliers on social media (Zeng & Gerritsen, 2014). This powerful informal communication among users, called Word-of-Mouth (WOM) can be easily spread through social media platforms on a large scale and is perceived by consumers as a commercially independent and trustworthy information source (Leung et al., 2013; Litvin, Goldsmith, & Pan, 2008). As it is able to influence a touristic business positively and also negatively, managing electronic WOM in social media properly can be essential for companies and organizations to constantly improve their products and services, avoid negative consequences and build a good relationship management with customers (Hede & Kellett, 2012; Thevenot, 2007; Zeng & Gerritsen, 2014).

### **4.1.2 Social media impact on customer decision-making**

Since touristic products are intangible, experiential and often costly, potential tourists prefer to inform themselves beforehand in the best possible way: by asking those, who already made the experience and share their information about it on social media (Fotis et al., 2012; Leung et al., 2013; Litvin et al., 2008). But these experience-sharing users vary in the influence they have on potential travelers (Zeng & Gerritsen, 2014). Only a small group made of very active and experienced “travel opinion leaders”, also called “central travelers” have the reach to significantly impact the decision-making process on potential tourists with their information provided (Vasiliki & Kostas, 2010; K. H. Yoo, Gretzel, & Zach, 2011).

In addition, it is very important to provide credible UGC and reliable high quality information on social media platforms to attract potential travelers and influence their decision-making process (Zeng & Gerritsen, 2014). Especially older generations are more skeptical regarding the trustworthiness of UGC, while younger generations value the credibility of such data higher and are more influenceable (Rodríguez, 2009, as cited in Zeng & Gerritsen, 2014).

### **4.1.3 Demographic differences**

People differ in how they use social media depending on their nationalities (Wilson, Murphy, & Fierro, 2012). Cultural, social and economic factors influence their behaviors, motivations and also the platforms they use and have to be considered in the tourism and hospitality industry and research (Wilson et al., 2012; Zeng & Gerritsen, 2014).

In addition, the language plays an important role in communication and information exchange on social media platforms (Zeng & Gerritsen, 2014). As the world is dominated by the English language mainly, tourists and touristic businesses who don't speak English are at a disadvantage in international

communication, whether it is promoting their business internationally or planning an international trip (Hsu, 2012).

#### **4.2 Benefits for tourism and research**

Social media is proven to be a powerful strategy in marketing and for promoting tourism products (Fotis et al., 2012; Zeng & Gerritsen, 2014). The participation of consumers in social media communities positively influences their attitude towards using the products or services provided by firms and also increases the possibility of recommending them to others (Casaló, Flavián, & Guinalíu, 2010).

Social media does not only provide information about tourists' mindsets regarding topics like alternative tourism, but can also be used to leverage their opinion and spread awareness about problems surrounding tourism and for example inform about more sustainable and eco-friendly alternatives (Zeng & Gerritsen, 2014).

The huge amount of Big Data in social media is of high value for research in tourism, especially in exploring tourists' decision-making behaviors and information search, forecasting consumer behavior and their needs, as well as analyzing business communication and interaction with customers through social media channels (Fotis et al., 2012; Vecchio et al., 2018; Zeng & Gerritsen, 2014).

#### **4.3 Twitter**

As information and communication technologies develop, the amount of UGC in social media platforms and blogs increases every second and is an enormous data base for research (Thelwall, Buckley, & Paltoglou, 2011).

The microblogging platform Twitter is one of the most popular social networks among individuals and companies with about 340 million monthly active users (Akehurst, 2008; Philander & Zhong, 2016; Statista, 2020; Thelwall et al., 2011). Since its formation in 2006, the platform's influence grew constantly and its users show a very high level of active engagement, posting more than 500 million messages, so called "tweets" every single day (Krikorian, 2013; Philander & Zhong, 2016).

In hospitality and tourism Twitter is a frequently used platform for product promotion, marketing, communicating with (potential) customers and conducting market research (Leung et al., 2013). Twitter data is widely used in research because it has a more liberal availability compared to platforms like Facebook, and most other social networks have a smaller userbase and are therefore less representative (Philander & Zhong, 2016; J. Wang, Gu, & Wang, 2013). Another reason why Twitter is so interesting is because the platform structure makes its users work like sensors and their interactions among each other represent public opinions and reactions towards external and social events (Kirilenko & Stepchenkova, 2014; Philander & Zhong, 2016). That means, the platform is very suitable for sentiment and content analysis, since it offers a huge amount of authentic customer opinions, real-time in-the-moment experiences and sentiment (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011; Kirilenko & Stepchenkova, 2014; Philander & Zhong, 2016). Firms use it to extract feedback and information from their customers for marketing purposes and building competitive benchmarks (Philander & Zhong, 2016).

When sharing their content, users have a maximum of 280 characters, which makes these tweets very compact and forces them to include only the essential meaning of what they want to express, leaving out irrelevant information

(Philander & Zhong, 2016). To discuss or inform about a certain topic, hashtags, “#”, are a useful tool to highlight the essential keyword in the message and find other posts talking about the same topic (Philander & Zhong, 2016). Due to this high amount of condensed data, an analysis can be way more exact and detailed than other information extraction techniques like surveys (Asur & Huberman, 2010; Kirilenko & Stepchenkova, 2014).

But Twitter also has its limits: The platform is widely used, but cannot represent the whole human population (Philander & Zhong, 2016). It should also be considered in research that the data is also not representative for the whole Twitter user community either, since about 40% of users don't upload their own content (Boyd & Crawford, 2012; Philander & Zhong, 2016).

## **5 ANALYZING BIG DATA**

With the constantly increasing volume and variety of data, manual data analysis has become way too time consuming and inefficient (Gandomi & Haider, 2015; Provost & Fawcett, 2013). In order to extract useful information from these extremely large databases, high-performance computer-aided techniques are necessary (George et al., 2014).

In most research, the averages in an analysis reveal tendencies of behaviors in certain situations, but due to the massive amount of Big Data, nearly every information has value (George et al., 2014). Especially the outliers can be of significant importance as they reveal upcoming trends, disturbances and revolutionary behaviors that might become a big issue in the future and could have huge consequences (George et al., 2014).

Big Data analysis is not a single method but rather a pattern or paradigm of diverse analytical tools profoundly leading to a new way of understanding knowledge, research processes, reality and how to work with large data sets (Boyd & Crawford, 2012; Xiang et al., 2014). It can be used to understand businesses in many aspects: not only customers, products and general market characteristics, but also competitors, suppliers and other stakeholders as well as the influence of modern technology (Xiang et al., 2014).

### **5.1 Big Data analytics approach**

By using the so-called Big Data analytics approach it is possible to collect and analyze data in a unique new way for solving real-life problems (Ginsberg et al., 2009; Manyika et al., 2011; Mayer-Schönberger & Cukier, 2013). It has the ability to remarkably complement traditional methods and give new, deeper insights into statistics, questionnaires and other mainly static data types (George et al., 2014; Xiang et al., 2014). It is also a very interesting tool for research with survey data since it can solve the problem of having too few samples by offering data on a large scale (Yang, Pan, Evans, & Lv, 2015).

### **5.2 Text mining**

For online textual data especially text mining techniques are common tools in tourism research (J. Li et al., 2018). The process usually begins with the data collection from the selected social media platforms using a web crawling technology (Xiang, Tussyadiah, & Buhalis, 2015; H. Xu, Yuan, Ma, & Qian, 2015). To extract the important information from this large data set collected, the data will be pre-processed in different sub-stages (J. Li et al., 2018):

---

### 5.2.1 Data pre-processing techniques

Data cleaning is an important sub-stage to eliminate irrelevant and non-usable data like misspellings, stop words, uncommon words or those in languages not considered by the research (Guo, Barnes, & Jia, 2017; J. Li et al., 2018; H. Xu et al., 2015; X. Xu & Li, 2016). Splitting data into information parts like words, sentences or other forms of grouping is called tokenization and enables tourism relevant data to be extracted and assigned to the different information types like tourist feelings, opinions or locations (Guo et al., 2017; Xiang et al., 2017; X. Xu & Li, 2016). After that, word stemming can be used to link all words with the same base for further simplification (X. Xu & Li, 2016). Part-of-speech Tagging (POST) is the last sub-stage used to mark each word with its word type (adjectives, verbs, nouns, etc.) removing all irrelevant words which are not classified with the selected relevant types. (Guo et al., 2017).

### 5.2.2 Pattern discovery

After pre-processing the data, several techniques of pattern discovery are applied to filter the most interesting information from the textual data (J. Li et al., 2018).

For tourism, especially the sentiment analysis is a very useful technique to gain knowledge about tourists' mindset regarding certain touristic products or destination locations by clustering the data into sentiment categories of positive, neutral and negative attitude (J. Li et al., 2018). A big advantage compared to traditional research methods like surveys is that it is less expensive and often more time efficient (Philander & Zhong, 2016). In addition, it does not interfere in the customer's opinion and enables extracting data in real-time, solving the problem of recall biases addressed by Rylander, Propst, & Mcurmury (1995) and providing the means for researchers to examine sentiment profiles second by second (Philander & Zhong, 2016).

With a statistical analysis it is possible to evaluate basic information like descriptive statistics from textual data to gain information about for example tourism demand or destination preferences (Gardini, 2008; J. Li et al., 2018). Clustering and categorization are tools to combine data elements with similarities into groups to achieve higher data value and gain new, often meaningful information (J. Li et al., 2018). For extracting key information from one or several documents of data, text summarization can be used to analyze for example hotel reviews (Hu, Chen, & Chou, 2017; J. Li et al., 2018).

### 5.2.3 Keyword selection

In order to analyze web search data properly in tourism prediction, the selection of keywords to search for is crucial (Peng, Liu, Wang, & Gu, 2017). Before starting the analysis it is recommended to divide the data into indexes instead of using the raw material to avoid multicollinearity in keyword selection (J. Li et al., 2018; Yang et al., 2015). There are three methods used mainly to select the right keywords: the empirical (also named experiential), the territorial and the technological approach (J. Li et al., 2018).

The empirical method uses only keywords the researcher knows from knowledge and experience, therefore being the most uncomplicated form of selection (J. Li et al., 2018). The main disadvantage of this subjective method is the high chance of missing out important keywords or even choosing wrong ones (J. Li et al., 2018).

---

The territorial method attaches to the empirical one by adding keywords that fit both, the already determined base of keywords from experience and the recommended words (J. Li et al., 2018). The problem here is that it might add disproportionate or unimportant words as well (J. Li et al., 2018).

By taking advantage of its prediction ability the technological method selects the most relevant keywords from a large amount of data (J. Li et al., 2018). Combining the territorial approach to offer a range of possible keywords with the technological approach of selecting the most promising, can provide a quite systematic method in picking useful keywords to analyze web search data (J. Li et al., 2018).

### **5.3 Challenges in Big Data Analysis**

Analyzing Big Data for research in tourism faces several challenges regarding not only the analytic techniques but also the data in general (J. Li et al., 2018). It is important to know that Big Data can't and shouldn't replace all sources of data in the industries (Xin Li & Law, 2020). Especially when making decisions, it should be analyzed and used carefully and traditional tools like observation and expertise should be taken into account as well (Xin Li & Law, 2020; Xiang et al., 2017).

#### **5.3.1 Research language and word selection**

While researchers often try to get the most accurate information, most studies so far focused their attention on data in only one language, ignoring or deleting for example reviews in other languages (Guo et al., 2017). This causes the data to be less representative, since tourists from other countries and cultures might value a different travel experience and would give different feedback (Schuckert, Liu, & Law, 2015a). To avoid information loss and provide accurate information, methods for multi-language analysis should be preferred (J. Li et al., 2018).

In addition to the language, the important words reviewed in tourist sentiment analysis are not only nouns, adjectives and negative adverbs, but also verbs (like love and hate) and the degree of strength they have, which should be equally valued in research (Hu et al., 2017; J. Li et al., 2018).

#### **5.3.2 Data Quality**

Regarding the data itself, several concerns, still affecting research are mentioned (J. Li et al., 2018). The first concern is data quality in several aspects like the accuracy of location data from photo or roaming data, which is known to be less precise than GPS data, the reliability of online textual data due to possible fake reviews by users or the possibility of biased google trends data caused by approximation methods and data sampling (J. Li et al., 2018; Pan, Xiang, Law, & Fesenmaier, 2012; Schuckert et al., 2015b).

#### **5.3.3 Data Costs**

Another challenge are the high costs mainly regarding device data, especially in contrast to data types like web search data and UGC. Sensor devices to obtain device data can be quite expensive and also finding volunteers to participate can become tricky and costly (J. Li et al., 2018).

---

#### **5.3.4 Data Privacy**

Privacy concerns is the third and for researchers most problematic challenge, since tourist movements, any kind of transaction data or sensitive information about customers, stakeholders or companies are usually kept secret within companies to provide trust, safety and privacy to their customers (J. Li et al., 2018). For research purpose, this data has very high value, but due to those privacy concerns it is often rarely available (J. Li et al., 2018).

#### **5.3.5 Solution Approach**

To find a solution for these challenges it is considered a promising step to start a collaboration between industries and academy to reduce costs in data collection, improve data availability and approach problems constructively (J. Li et al., 2018). To improve the issues regarding privacy, confidential agreements and options to remove sensitive data could be practical solutions (J. Li et al., 2018).

### **5.4 Usefulness of Big Data Analysis**

Big Data analysis is used by companies in nearly every industry to improve performance, create significant value and achieve a market based competitive advantage (Line et al., 2020; Manyika et al., 2011; Mariani, Baggio, Fuchs, & Höepken, 2018; Provost & Fawcett, 2013). For tourism and hospitality, it is particularly useful to predict outcomes like tourist arrivals or the occupancy rate in hotels, reveal patterns and understand customer experience and its relation to customer satisfaction (Xin Li & Law, 2019; Xin Li, Pan, Law, & Huang, 2017; Xiang et al., 2015).

Online textual data in the form of online reviews is a useful and important source to understand and predict online bookings, accommodation demand and tourism marketing (Xin Li & Law, 2020). By analyzing and observing customer interactions and behaviors, companies gain deeper insights into their consumer segments and can beneficially connect and engage with them (Line et al., 2020). It is also possible to select the most valuable and profitable customers and generate value from good customer relationship management (Line et al., 2020).

#### **5.4.1 Decision-making**

Analyzing data can improve all decisions made by people in organizations and provides possibilities for all sizes of companies and businesses (Robles & Barragán, 2017).

For tourism especially the ability of forecasting demand is of high value, impacting the decision-making process of policy makers, managers and other stakeholders in the industry (Fronzetti Colladon et al., 2019).

The UGC shared on social media platforms can be used to discover trends in destination choices and track changes in tourists' attitudes, behaviors and preferences (Fronzetti Colladon et al., 2019). Online reviews and customer feedback through social media not only enable tourists to exchange information, but also give the company a tool to connect to their (potential) customers and by analyzing the data, they can continuously improve their service, marketing and management (Fang et al., 2016; J. Li et al., 2018).

Big Data analysis provides support for destination management in strategic decision-making and can help forecasting tourists' choices in destination and

---

arrivals (Matthias Fuchs, Höpken, & Lexhagen, 2014; Gavilan et al., 2018; Miah, Vu, Gammack, & McGrath, 2017; Yang et al., 2015).

#### **5.4.2 Developing new business opportunities**

Big Data is already used by companies to create new products, services and business ideas and its importance is continuously growing (Robles & Barragán, 2017). By analyzing the information in Big Data generated by algorithms and analytical methods it is possible to improve and optimize current processes and also develop new business opportunities (Robles & Barragán, 2017).

### **6 TWITTER ANALYSIS**

For this practical analysis, the main goal was to discover new alternative touristic products and business opportunities in the Balearic Islands via the social media platform Twitter. For that, a database of tweets, geo-tagged from the Balearic Islands, was collected from 2015 to 2017.

I decided to specify my analysis on the quite popular hiking and cycling tourism to achieve a great amount of data. Hiking and cycling tourism is considered a sustainable alternative to mass tourism as it promotes ecofriendly forms of transport and reduces air pollution and the amount of traffic on the islands (Moon & Cj, 2018).

#### **6.1 Keyword selection**

In order to extract related tweets, the first step is to prepare a list of relevant keywords connected to the topic.

To select the keywords, I followed the empirical approach first and thought of words and hashtags connected to cycling and hiking in English, Spanish and German. I continued with the territorial method to avoid missing out important words by searching on Twitter for the words I already had. As users often include more than one hashtag, it was easy to discover new relevant words. This technique also helped to select words and hashtags that were used often on the platform and furthermore delete those, that were too specific and not relevant enough for the research.

Since some English words were not possible to translate properly into the other languages or the translated equivalent was rarely used on the platform, I ended up with less keywords and hashtags in German and Spanish than I had in English. The final selection contained a total of 50 keywords and 49 hashtags listed in table 1.

#### **6.2 Program used**

To extract the relevant tweets from the database, a Python script was used to filter for the selected keywords. This extraction resulted in 4511 messages containing at least one word or hashtag in English, 1301 containing at least one in Spanish and 430 containing at least one in German.

Furthermore, the integrated development environment R Studio provided a diagram and a list of the most used words, starting at a minimum word appearing of 5 times. This helped me to get an overview about the most important words and topics in the tweets.

---

## 6.3 Techniques used

I applied several text mining techniques to be able to analyze the data.

### 6.3.1 Data cleaning

The first step was to pre-process the data. For this I used the technique of data cleaning and deleted all tweets, which:

- included no text but only hashtags, as they mostly referred to online photo data, or were irrelevant.
- were in a language different from English, Spanish and German.
- were off topic. This, for example, often occurred with the word “cycling” as it was part of recycling and upcycling and furthermore used as a synonym for “switching” or “changing” in some tweets and therefore irrelevant.

### 6.3.2 Statistical analysis

To discover the most interesting information from the data, a statistical analysis was applied. This analysis provided me with information about how often words appeared in the tweets. The 200 most frequently used ones in each language extraction then got visualized with R Studio, as shown in the figures 1, 2 and 3.

The by far most used word was “cycling” with 2465 mentions, followed by “Mallorca” and its English notation “Majorca”, which all were not very surprising.

The third most used word was “Endomondo”, which is a mobile sport software that tracks distances via GPS (Endomondo, 2020). Users of the app can share their daily progress on social media by sending a message, for example when they start their first workout using the app, or when they finished cycling, sharing the exact distance they went. It appears to be a widely used application among hiking, cycling and mountain biking tourists in the Balearic Islands as it is being mentioned 925 times.

“Ibiza” and its Catalan translation “Eivissa” also appeared very often, showing that, after Mallorca, it is the second most talked about island for hiking and biking in the Balearic Islands. Also, “mountainbike” and “mountainbiking” were mentioned 142 and 175 times, revealing that there is a quite active segment of interested people for this activity.

### 6.3.3 Manual analysis

After preparing the data I also manually read through the tweets to discover topics that were rarely talked about. As George et al. (2014) mentions, in Big Data, every element can be of value and not only the averages are important. Especially in discovering new upcoming trends or business opportunities, the focus should not only lie on already widely known topics but also consider less familiar ones. Therefore, I focused my attention on the outliers which could lead to new business ideas and new touristic products.

## 6.4 Discoveries

### 6.4.1 Languages analyzed

English is the most used language on Twitter, and I noticed many people communicating online in this language, even if wasn't their mother tongue (Statista, 2013). The main purpose for tourists to use Twitter is to share experiences and exchange opinions and information with others (Philander & Zhong, 2016; Zeng & Gerritsen, 2014). Communicating in English is the best way

to reach people from different countries in Europe, as many learn it as a second language (Parker, 2015).

The amount of results varied a lot among the three languages considered. Although the Balearic Islands are part of Spain, the by far biggest number of tweets was in English. The fewer number in Spanish and German are caused by several factors, for example the fact that the keyword selection contained more English words and that some people prefer to write their messages in English instead of their first language.

For this particular analysis the consideration of 3 instead of only one language definitely influenced the results and discoveries. With a bigger analysis and a broader topic, I assume the results will depend even more on the amount of languages considered. The option of implementing that into an analysis depends on the researcher's knowledge and ability to work with several languages. The ultimate goal would be to include all languages into an analysis to achieve the most accurate results.

#### **6.4.2 Touristic activities**

I discovered many interesting topics through the analysis. For example, a user stated that their favorite activity is to go hiking and horse trekking in Mallorca. Also, beach biking, climbing, downhill biking, all-terrain biking and cliff diving were activities of great interest among rather action-oriented users. Furthermore, I noticed several users talking about spending their holidays in training camps for cycling as they preferred to use their leisure time to improve their body constitution and do sports instead of relaxing at the beach. Also, paracycling athletes shared that they enjoy visiting Mallorca for its huge offerings for cyclists. Several users talked about spending the time preparing for a triathlon or a race competition. Several hikers mentioned that they enjoyed the landscapes full of olive trees. This led me to the new emerging topic of "olive oil tourism", which appears to be a booming form of tourism in rural areas with great amounts of olive trees (Pulido-Fernández, Casado-Montilla, & Carrillo-Hidalgo, 2019).

Another remarkably interesting discovery among these were some users mentioning the combination of meditation and yoga with hiking and cycling.

#### **6.5 Create new touristic products**

The different activities mentioned in the tweets can be used directly or as an inspiration to create new business opportunities. To see if a topic is interesting as a touristic product, the market needs to be analyzed regarding potential customers and furthermore to check if it is already existing and how it is marketed by the competition. Big Data analysis could be applied to get an overview about the topic and how it is implemented by other companies, understand which potential customer segment is interested in it and how this product could be marketed to the public.

An interesting point is that these discoveries can also be used as business opportunities for different regions. Getting inspired by touristic products mentioned for the Balearic Islands could also be an interesting opportunity for other countries or regions as long as they provide the similar relevant environmental, cultural and social factors.

The results from the twitter analysis for hiking and cycling tourism in the Balearic Islands show for example that there is a group of people interested in meditation and yoga combined with hiking and cycling. After some research I

discovered that this form of tourism seems to be growing and is offered for individuals or smaller groups in the form of camps or guided tours by a small, yet increasing amount of tour operators (BookYogaRetreats, 2020; Chillouttreats, 2020; Roxybikemallorca, 2020). This type of touristic product is not limited to the Balearic Islands, but also offered for example in Italy, the Canary Islands or Portugal (BookCyclingHolidays, 2020). As cycling and hiking are both outdoor sport activities, it is important to have good roads, quality routes for hiking, cycling and mountain biking and a rather warm climate with few rainfalls. The customer segment appears to be adult men and women, ages about 25-50, who are active and enjoy a lifestyle with healthy food, dieting and sport activities (BPlans.com, 2020). Their objective is to use these holidays to discover a destination while being active, finding a balance between body work out and relaxation of mind and soul. The touristic product should therefore offer the participants to continue with their lifestyle and provide well organized yoga and meditation classes as well as guided hiking or cycling tours where they are able to discover and enjoy the natural environment of the destination.

When creating a touristic product, it is important to consider the characteristics of the customer segment and the given environmental, cultural, and social circumstances of the region. Furthermore, the competition should be kept in mind, comparing the offers and their prices, and thinking about unique selling points (USP) to make the own product stand out from the competition.

## **6.6 Future directions for analyzing data**

With the number of tweets obtained from the database, it was possible to read through them manually to discover ideas for new touristic products in the Balearic Islands. But when using a bigger amount of data, the manual analysis of reading through all messages will be too time consuming and ineffective. Therefore, future research relies on the technology development. Machine learning is a sub-topic of artificial intelligence (AI) and is used in data analytics, data mining and to discover patterns, insights and trends (Duggal, 2020). It might be an interesting direction for tourism research to improve the data analysis for discovering new business opportunities.

## **7 CONCLUSION**

Using Big Data analytics for tourism research is still at a young stage, but already provides lots of opportunities for the industry. The goal of my thesis was to provide an overview about how Big Data already helps in the tourism industry, focusing on how it can improve the decision-making process in companies and how it can help to discover new business opportunities.

Analyzing Big Data can be used to understand and predict tourism demand and consumer behavior. It can help to improve marketing, product promotion, customer satisfaction and customer relationship management. The online textual data in the form of UGC in blogs and reviews can help the company to analyze customer feedback and improve their products and services. Destinations can benefit from Big Data by predicting tourist's decision-making in destination choice and therefore forecast arrivals.

By offering these new and deep insights into the business, Big Data can highly influence the decision-making process in companies and organizations and offers new approaches to improve business processes and solve real life problems.

---

The results of the Twitter analysis show, that it is possible to discover new alternative touristic products using UGC from social media platforms. The analysis process still requires a lot of time and depending on the data size it can be inefficient as it has to be manually analyzed to discover new opportunities in the tourism industry. But with the development of technology this problem might be solved in the future.

With the analysis of the data obtained from the Balearic Islands, it was possible to discover several interesting opportunities like the combination of yoga, meditation and hiking or cycling. With further analysis of the market, competition and customer segment, businesses can get inspired for own ideas and decide if they want to create a new touristic product themselves. Big Data analysis can be used as a starting point to discover and understand upcoming trends and their customer segments, which might offer great opportunities for a future business.

---

**LIST OF CITED LITERATURE**

- Akehurst, G. (2008). User generated content: The use of blogs for tourism organisations and tourism consumers. *Service Business*, 3(1), 51.  
<https://doi.org/10.1007/s11628-008-0054-2>
- Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. *Applied Energy*, 112, 1536–1543.  
<https://doi.org/10.1016/j.apenergy.2013.03.027>
- BookCyclingHolidays. (2020). Top 10 Cycling and Yoga Holidays Worldwide. Retrieved May 10, 2020, from BookCyclingHolidays.com website:  
<https://www.bookcyclingholidays.com/all/c/cycling-yoga>
- BookYogaRetreats. (2020). Top 10 Vinyasa Yoga Cycling Holidays in Mallorca. Retrieved April 29, 2020, from BookYogaRetreats.com website:  
[https://www.bookyogaretreats.com/all/c/yoga-cycling/s/vinyasa-yoga/d/europe/spain/mallorca?\\_=1588172195](https://www.bookyogaretreats.com/all/c/yoga-cycling/s/vinyasa-yoga/d/europe/spain/mallorca?_=1588172195)
- Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon: *Information, Communication & Society*: Vol 15, No 5. Retrieved April 1, 2020, from <https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>
- BPlans.com. (2020). Yoga Center Business Plan Sample—Market Analysis | Bplans. Retrieved May 10, 2020, from [https://www.bplans.com/yoga\\_center\\_business\\_plan/market\\_analysis\\_summary\\_fc.php](https://www.bplans.com/yoga_center_business_plan/market_analysis_summary_fc.php)
- Bray, J., Schetzina, C., & Steinbrink, S. (2006). 6 Travel Tech Trends for 2006. *Hospitality Upgrade*, 150–152.

- 
- Browning, V., So, K. K. F., & Sparks, B. (2013). The Influence of Online Reviews on Consumers' Attributions of Service Quality and Control for Service Standards in Hotels. *Journal of Travel & Tourism Marketing*, 30(1–2), 23–40. <https://doi.org/10.1080/10548408.2013.750971>
- Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management*, 29(4), 609–623. <https://doi.org/10.1016/j.tourman.2008.01.005>
- Butler, R. W. (1990). Alternative Tourism: Pious Hope Or Trojan Horse? *Journal of Travel Research*, 28(3), 40–45. <https://doi.org/10.1177/004728759002800310>
- Casaló, L. V., Flavián, C., & Guinalíu, M. (2010). Determinants of the intention to participate in firm-hosted online travel communities and effects on consumer behavioral intentions. *Tourism Management*, 31(6), 898–911. <https://doi.org/10.1016/j.tourman.2010.04.007>
- Chillouttreats. (2020). MALLORCA CYCLE & YOGA - chillout Retreats. Retrieved April 29, 2020, from <http://www.chilloutretreats.com/mallorca-cycle-yoga/4592243787>
- Chung, J. Y., & Buhalis, D. (2008). Web 2.0: A study of online travel community. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 70–81). [https://doi.org/10.1007/978-3-211-77280-5\\_7](https://doi.org/10.1007/978-3-211-77280-5_7)
- Cohen, E. (1987). "Alternative Tourism"—A Critique. *Tourism Recreation Research*, 12(2), 13–18. <https://doi.org/10.1080/02508281.1987.11014508>

- 
- Cohen, H. (2020, March 1). Social Media Definition: The Guide You Need To Get Results. Retrieved April 14, 2020, from Heidi Cohen website: <https://heidicohen.com/social-media-definition/>
- Dellaert, B. G. C. (2000). Tourists' Valuation of Other Tourists' Contributions to Travel Web Sites. In D. R. Fesenmaier, S. Klein, & D. Buhalis (Eds.), *Information and Communication Technologies in Tourism 2000* (pp. 293–302). [https://doi.org/10.1007/978-3-7091-6291-0\\_26](https://doi.org/10.1007/978-3-7091-6291-0_26)
- Deng, N., & Li, X. (Robert). (2018). Feeling a destination through the “right” photos: A machine learning model for DMOs' photo selection. *Tourism Management*, 65, 267–278. <https://doi.org/10.1016/j.tourman.2017.09.010>
- Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter. *PLoS ONE*, 6(12). <https://doi.org/10.1371/journal.pone.0026752>
- Duggal, N. (2020, April 21). Top 8 Technology Trends for 2020. Retrieved May 2, 2020, from Simplilearn.com website: <https://www.simplilearn.com/top-technology-trends-and-jobs-article>
- East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, 61(C), 307–320.
- Endomondo. (2020). Endomondo. Retrieved April 28, 2020, from <http://www.endomondo.com/>
- 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.  
<https://doi.org/10.1016/j.tourman.2015.07.018>
- Fotis, J. N., Buhalis, D., & Rossides, N. (2012). Social media use and impact during the holiday travel planning process. In M. Fuchs, F. Ricci, & L. Cantoni (Eds.), *Information and Communication Technologies in Tourism 2012* (pp. 13–24). Retrieved from <http://www.springerlink.com>
- Fronzetti Colladon, A., Guardabascio, B., & Innarella, R. (2019). Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems*, 123, 113075.  
<https://doi.org/10.1016/j.dss.2019.113075>
- Fuchs, Matthias, 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.  
<https://doi.org/10.1016/j.jdmm.2014.08.002>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC IView*, (1142), 1–12. Retrieved from Scopus.
- Gardini, A. (2008). Statistical analysis of tourism destination competitiveness. *Statistica*, 68(2), 153–166. <https://doi.org/10.6092/issn.1973-2201/3527>
- Gavilan, D., Avello, M., & Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, 53–61. <https://doi.org/10.1016/j.tourman.2017.10.018>

- 
- George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management: From the Editors. *Academy of Management Journal*, 57(2), 321–326.  
<https://doi.org/10.5465/amj.2014.4002>
- Germann Molz, J. (2013). SOCIAL NETWORKING TECHNOLOGIES AND THE MORAL ECONOMY OF ALTERNATIVE TOURISM: THE CASE OF COUCHSURFING.ORG. *Annals of Tourism Research*, 43, 210–230.  
<https://doi.org/10.1016/j.annals.2013.08.001>
- Gibson, C. (2010). Geographies of tourism: (Un)ethical encounters. *Progress in Human Geography*, 34(4), 521–527.  
<https://doi.org/10.1177/0309132509348688>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014.  
<https://doi.org/10.1038/nature07634>
- Gretzel, U., & Yoo, K. H. (2008). Use and Impact of Online Travel Reviews. In P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 35–46).  
[https://doi.org/10.1007/978-3-211-77280-5\\_4](https://doi.org/10.1007/978-3-211-77280-5_4)
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59, 467–483.  
<https://doi.org/10.1016/j.tourman.2016.09.009>
- Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., & Eccleston, R. (2017). Tracking tourists' travel with smartphone-based GPS

- 
- technology: A methodological discussion. *Information Technology & Tourism*, 17(3), 255–274. <https://doi.org/10.1007/s40558-017-0086-3>
- Hede, A.-M., & Kellett, P. (2012). Building online brand communities: Exploring the benefits, challenges and risks in the Australian event sector. *Journal of Vacation Marketing*, 18(3), 239–250. <https://doi.org/10.1177/1356766712449370>
- Hsu, Y.-L. (2012). Facebook as international eMarketing strategy of Taiwan hotels. *International Journal of Hospitality Management*, 31(3), 972–980. <https://doi.org/10.1016/j.ijhm.2011.11.005>
- Hu, Y.-H., Chen, Y.-L., & Chou, H.-L. (2017). Opinion mining from online hotel reviews – A text summarization approach. *Information Processing & Management*, 53(2), 436–449. <https://doi.org/10.1016/j.ipm.2016.12.002>
- Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management*, 47, 68–76. <https://doi.org/10.1016/j.tourman.2014.09.001>
- Hudson, S., & Thal, K. I. (2013). *The Impact of Social Media on the Consumer Decision Process: Implications for Tourism Marketing*. <https://doi.org/10.1080/10548408.2013.751276>
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59–64. <https://doi.org/10.1016/j.bdr.2015.01.006>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>

---

Kiliç, B., & Kurnaz, A. (2010). *Alternative Tourism and Ecological Farms on Creating Diversification of Tourism Product: Example of Pastoral Valley*.

3.

Kirilenko, A. P., & Stepchenkova, S. O. (2014). Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change*, 26, 171–182. <https://doi.org/10.1016/j.gloenvcha.2014.02.008>

Krikorian, R. (2013, August 16). New Tweets per second record, and how! Retrieved April 16, 2020, from Blog.twitter.com website: [https://blog.twitter.com/engineering/en\\_us/a/2013/new-tweets-per-second-record-and-how.html](https://blog.twitter.com/engineering/en_us/a/2013/new-tweets-per-second-record-and-how.html)

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety* (p. 4). Retrieved from Meta Group website: <https://www.bibsonomy.org/bibtex/742811cb00b303261f79a98e9b80bf49?lang=de#export>

Leung, D., Law, R., Hoof, H. van, & Buhalis, D. (2013). Social Media in Tourism and Hospitality: A Literature Review. *Journal of Travel & Tourism Marketing*, 30(1–2), 3–22. <https://doi.org/10.1080/10548408.2013.750919>

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>

Li, Xiaoxuan, Wu, Q., Peng, G., & Lv, B. (2016). Tourism forecasting by search engine data with noise-processing. *African Journal of Business Management*, 10(6), 114–130. <https://doi.org/10.5897/AJBM2015.7945>

- Li, Xin, & Law, R. (2019). Forecasting Tourism Demand with Decomposed Search Cycles. Retrieved April 1, 2020, from <https://journals.sagepub.com/doi/abs/10.1177/0047287518824158>
- Li, Xin, & Law, R. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives*, 33, 100608. <https://doi.org/10.1016/j.tmp.2019.100608>
- Li, Xin, Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, 59, 57–66. <https://doi.org/10.1016/j.tourman.2016.07.005>
- Line, N. D., Dogru, T., El-Manstrly, D., Buoye, A., Malthouse, E., & Kandampully, J. (2020). Control, use and ownership of big data: A reciprocal view of customer big data value in the hospitality and tourism industry. *Tourism Management*, 80, 104106. <https://doi.org/10.1016/j.tourman.2020.104106>
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. <https://doi.org/10.1016/j.tourman.2007.05.011>
- Llodrà-Riera, I., Martínez-Ruiz, M. P., Jiménez-Zarco, A. I., & Izquierdo-Yusta, A. (2015). A multidimensional analysis of the information sources construct and its relevance for destination image formation. *Tourism Management*, 48, 319–328. <https://doi.org/10.1016/j.tourman.2014.11.012>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011, May). Big data: The next frontier for innovation, competition, and productivity | McKinsey. Retrieved April 1, 2020, from <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>

- 
- 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. <https://doi.org/10.1108/IJCHM-07-2017-0461>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- McAfee, A., & Brynjolfsson, E. (2012). *Big Data: The Management Revolution*. 9.
- McIntosh, A. J., & Zahra, A. L. (2007). *A cultural encounter through volunteer tourism: Towards the ideals of sustainable tourism?* 15(5), 541–556. <https://doi.org/10.2167/jost701.0>
- Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management*, 54(6), 771–785. <https://doi.org/10.1016/j.im.2016.11.011>
- Moon, C. J., & Cj, M. (2018). *Mass Tourism vs. Sustainable Tourism in the Balearic Islands: Measuring Social and Environmental Impact in Mallorca*. 2018(03), 13.
- Moscardo, G. (2001). Cultural and heritage tourism: The great debates. In B. Faulkner, G. Moscardo, & E. Laws (Eds.), *Tourism in the 21st Century: Lessons from experience* (pp. 3–17). Retrieved from <https://researchonline.jcu.edu.au/14350/>
- Pan, B., Xiang, Z., Law, R., & Fesenmaier, D. R. (2012). The Dynamics of Search Engine Marketing for Tourist Destinations. Retrieved April 1, 2020, from ResearchGate website: [https://www.researchgate.net/publication/228709534\\_The\\_Dynamics\\_of\\_Search\\_Engine\\_Marketing\\_for\\_Tourist\\_Destinations](https://www.researchgate.net/publication/228709534_The_Dynamics_of_Search_Engine_Marketing_for_Tourist_Destinations)

- Parker, B. (2015, October 8). European youths learn English as foreign language more than any other. Retrieved April 26, 2020, from Pew Research Center website: <https://www.pewresearch.org/fact-tank/2015/10/08/more-than-any-other-foreign-language-european-youths-learn-english/>
- Pearce, P. L., & Wu, M.-Y. (2015). Entertaining International Tourists: An Empirical Study of an Iconic Site in China. Retrieved April 1, 2020, from [https://journals.sagepub.com/doi/full/10.1177/1096348015598202?casa\\_token=07OSWqAOPHkAAAAA%3AC5BaXvPi3t6STxLkioW1piD4r0IHBuQvxhHMEd2n3Em-2sU0032NpUtXjtWE6nH0\\_vn1tApiJyYqTpw](https://journals.sagepub.com/doi/full/10.1177/1096348015598202?casa_token=07OSWqAOPHkAAAAA%3AC5BaXvPi3t6STxLkioW1piD4r0IHBuQvxhHMEd2n3Em-2sU0032NpUtXjtWE6nH0_vn1tApiJyYqTpw)
- Pektaş, F. (2018). THE EFFECT OF LIFESTYLE ON THE DEMAND FOR ALTERNATIVE TOURISM. *International Journal of Management Economics and Business*, 14(1), 187–198. <https://doi.org/10.11122/ijmeh.2014.14.1.1140>
- Peng, G., Liu, Y., Wang, J., & Gu, J. (2017). Analysis of the prediction capability of web search data based on the HE-TDC method – prediction of the volume of daily tourism visitors. *Journal of Systems Science and Systems Engineering*, 26(2), 163–182. <https://doi.org/10.1007/s11518-016-5311-7>
- Philander, K. S., & Zhong, Y. Y. (2016). *Twitter sentiment analysis: Capturing sentiment from integrated resort tweets*. <https://doi.org/10.1016/j.ijhm.2016.02.001>
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51–59. <https://doi.org/10.1089/big.2013.1508>

- 
- Pulido-Fernández, J. I., Casado-Montilla, J., & Carrillo-Hidalgo, I. (2019). Introducing olive-oil tourism as a special interest tourism. *Heliyon*, 5(12), e02975. <https://doi.org/10.1016/j.heliyon.2019.e02975>
- Rabanser, U., & Ricci, F. (2005). Recommender Systems: Do They Have a Viable Business Model in e-Tourism? In A. J. Frew (Ed.), *Information and Communication Technologies in Tourism 2005* (pp. 160–171). [https://doi.org/10.1007/3-211-27283-6\\_15](https://doi.org/10.1007/3-211-27283-6_15)
- Robles, V. M., & Barragán, D. (2017, March 15). Big Data: A Critical Path to Develop New Business Opportunities. Retrieved April 30, 2020, from IE Insights website: <https://www.ie.edu/insights/articles/big-data-critical-path-to-develop-new-business-opportunities/>
- Rodríguez, I. L. (2009). *Social media in tourism behavior* (Master's degree dissertation). Bournemouth University and Université de Savoie.
- Roxybikemallorca. (2020). MTB Events Spain, Mallorca, Costa Blanca and South Tyrol. Retrieved April 29, 2020, from <https://roxybikemallorca.com/mountain-bike-events-coaching-english/>
- Rylander, R. G., Propst, D. B., & Mcurtry, T. R. (1995). Nonresponse and Recall Biases in a Survey of Traveler Spending. *Journal of Travel Research*, 33(4), 39–45. <https://doi.org/10.1177/004728759503300406>
- Sarks, G. (2007, April). *Consumer generated content in travel: A roadtrip for bringing consumers into the conversation*. Web-seminar presentation.
- Schuckert, M., Liu, X., & Law, R. (2015a). A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *International Journal of Hospitality Management*, 48, 143–149. <https://doi.org/10.1016/j.ijhm.2014.12.007>

- 
- Schuckert, M., Liu, X., & Law, R. (2015b). Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*, 32(5), 608–621. <https://doi.org/10.1080/10548408.2014.933154>
- Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169. <https://doi.org/10.1016/j.jretai.2004.04.001>
- Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies*, 18(5), 587–606. <https://doi.org/10.1080/14616688.2016.1214977>
- Simms, A. (2012). Online user-generated content for travel planning—Different for different kinds of trips? *Faculty of Business - Papers*, 76–85.
- Statista. (2013, September). Most-used languages on Twitter as of September 2013. Retrieved April 26, 2020, from Statista website: <https://www.statista.com/statistics/267129/most-used-languages-on-twitter/>
- Statista. (2020, January). Most popular social networks worldwide as of January 2020, ranked by number of active users. Retrieved April 16, 2020, from Statista website: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Steylaerts, V., & Dubhghaill, S. O. (2012, February 16). CouchSurfing and authenticity: Notes towards an understanding of an emerging phenomenon [Text]. [https://doi.org/info:doi/10.1386/hosp.1.3.261\\_1](https://doi.org/info:doi/10.1386/hosp.1.3.261_1)
- Tezcan, B. (2004). *DEVELOPING ALTERNATIVE MODES OF TOURISM IN TURKEY* (Thesis, Middle East Technical University). Retrieved from

---

[https://www.semanticscholar.org/paper/DEVELOPING-ALTERNATIVE-MODES-OF-TOURISM-IN-TURKEY-](https://www.semanticscholar.org/paper/DEVELOPING-ALTERNATIVE-MODES-OF-TOURISM-IN-TURKEY-Tezcan/c563a4a940f2d4bcc895efbe3c0e2afc79e6ffb4)

[Tezcan/c563a4a940f2d4bcc895efbe3c0e2afc79e6ffb4](https://www.semanticscholar.org/paper/DEVELOPING-ALTERNATIVE-MODES-OF-TOURISM-IN-TURKEY-Tezcan/c563a4a940f2d4bcc895efbe3c0e2afc79e6ffb4)

- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406–418. <https://doi.org/10.1002/asi.21462>
- Thevenot, G. (2007). Blogging as a Social Media. *Tourism and Hospitality Research*. <https://doi.org/10.1057/palgrave.thr.6050062>
- Vasiliki, V., & Kostas, Z. (2010). Locating central travelers' groups in travel blogs' social networks. *Journal of Enterprise Information Management*, 23(5), 595–609. <https://doi.org/10.1108/17410391011083056>
- Vecchio, P. D., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing & Management*, 54(5), 847–860. <https://doi.org/10.1016/j.ipm.2017.10.006>
- Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127. <https://doi.org/10.1016/j.tourman.2008.04.008>
- Wang, J., Gu, Q., & Wang, G. (2013, April 1). Potential Power and Problems in Sentiment Mining of Social Media [Article]. Retrieved April 16, 2020, from International Journal of Strategic Decision Sciences (IJSDS) website: [www.igi-global.com/article/content/78345](http://www.igi-global.com/article/content/78345)
- Wang, Y., Yu, Q., & Fesenmaier, D. R. (2002). Defining the virtual tourist community: Implications for tourism marketing. *Tourism Management*, 23(4), 407–417. [https://doi.org/10.1016/S0261-5177\(01\)00093-0](https://doi.org/10.1016/S0261-5177(01)00093-0)

- 
- Weaver, D. (2001). Sustainable tourism: Is it sustainable? *Tourism in the Twenty-First Century: Reflections on Experience*, 300–311.
- Wilson, A., Murphy, H., & Fierro, J. C. (2012). Hospitality and Travel: The Nature and Implications of User-Generated Content. *Cornell Hospitality Quarterly*.  
<https://doi.org/10.1177/1938965512449317>
- 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(C), 51–65.
- Xiang, Z., & Gretzel, U. (2010). Role of Social Media in Online Travel Information Search. *Tourism Management*, 31, 179–188.  
<https://doi.org/10.1016/j.tourman.2009.02.016>
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2014). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120–130.  
<https://doi.org/10.1016/j.ijhm.2014.10.013>
- Xiang, Z., Tussyadiah, I., & Buhalis, D. (2015). Smart destinations: Foundations, analytics, and applications. *Journal of Destination Marketing & Management*, 4. <https://doi.org/10.1016/j.jdmm.2015.07.001>
- Xu, H., Yuan, H., Ma, B., & Qian, Y. (2015). Where to go and what to play: Towards summarizing popular information from massive tourism blogs. *Journal of Information Science*, 41(6), 830–854.  
<https://doi.org/10.1177/0165551515603323>
- 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.  
<https://doi.org/10.1016/j.ijhm.2016.03.003>
- Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46, 386–397.  
<https://doi.org/10.1016/j.tourman.2014.07.019>
- 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.  
<https://doi.org/10.1016/j.ijhm.2008.06.011>
- Yoo, K. H., Gretzel, U., & Zach, F. (2011). Travel Opinion Leaders and Seekers. *Faculty of Commerce - Papers (Archive)*, 525–535.  
[https://doi.org/10.1007/978-3-7091-0503-0\\_42](https://doi.org/10.1007/978-3-7091-0503-0_42)
- Yoo, K.-H., Lee, Y., Gretzel, U., & Fesenmaier, D. R. (2009). Trust in Travel-related Consumer Generated Media. In W. Höpken, U. Gretzel, & R. Law (Eds.), *Information and Communication Technologies in Tourism 2009* (pp. 49–59). [https://doi.org/10.1007/978-3-211-93971-0\\_5](https://doi.org/10.1007/978-3-211-93971-0_5)
- Yoon, H., Zheng, Y., Xie, X., & Woo, W. (2010). Smart Itinerary Recommendation Based on User-Generated GPS Trajectories. In Z. Yu, R. Liscano, G. Chen, D. Zhang, & X. Zhou (Eds.), *Ubiquitous Intelligence and Computing* (pp. 19–34). [https://doi.org/10.1007/978-3-642-16355-5\\_5](https://doi.org/10.1007/978-3-642-16355-5_5)
- Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27–36.  
<https://doi.org/10.1016/j.tmp.2014.01.001>
- Zhang, Z., Ye, Q., Law, R., & Li, Y. (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*, 29(4),  
694–700. <https://doi.org/10.1016/j.ijhm.2010.02.002>