African Journal of Hospitality, Tourism and Leisure, Volume 8 (3) - (2019) ISSN: 2223-814X

Social big data analysis of Five Star hotels: A case study of hotel guest experience and satisfaction in Marrakech

El haouta Imane*
Algebra Team and it's Applications, Department of Mathematics
Abdelmalek Essaadi University
Faculty of Sciences, Morocco
elhaouta.emy@gmail.com

Professor Idelhadj Abdelouahab
Algebra Team and its Applications, Department of Mathematics
Abdelmalek Essaadi University
Faculty of Sciences, Morocco

Corresponding author*

Abstract

The fast progress of the Internet and mobile devices facilitated the introduction of travel and hospitality review sites, producing a high number (big data) of customer opinion posts. While such comments may manipulate future demand of the specific hotels, they can also be used by hotel managers for bettering customer experience. In order to understand the hotel guest experience and satisfaction in Morocco, especially in Marrakech city during the period from 2015 to 2018, we collected data from "TripAdvisor.com", the online customer textual reviews, and rating of 7 five star hotels. A Hadoop cluster was configured to handle the social big data extracted, as the method we employed big data analytics and sentiment analysis for handling the sentiment-detection algorithms using tidytext package, as well as bing lexicon. The used techniques were reliable to analyze and identify different characteristics of hotel guest such as "types" of hotel guests; we found that the highest rated type of hotel guests was 'with family'. Concerning the "behavioral", the type of guest 'with family' seems to comment and expresses his satisfaction or dissatisfaction every month, except in January, September, and December. Moreover, concerning the "nationality" of hotel guests, France represented the highest rate by 70.07%. On the other hand for satisfaction, the "rating" of the field 'service' was decreased in the last two years. Finally, the "sentiment-experience" and the words 'limitations' and 'excellent' were clearly the main hotel attributes mentioned by the customers. The present study showed that big social data analytics and sentiment analysis could be good solutions to help the tourism and hospitality industry to gain insights into the characteristics, satisfaction, and sentiments of the guests to hotels during their experiences at the hotels. However, knowing more elements relevant to the hotel guest would be even more beneficial for hotels in order to enable them to improve their products/services ratio.

Keywords: TripAdvisor, tourism, hotel guest experience and satisfaction, social big data, rating, sentiment analysis.

Introduction

In recent years, tourists have relied on the use of some social platform and network such as TripAdvisor, Yelp and Facebook in order to obtain the most important travel information. This information has already been provided by other tourists that share their visit experiences via some platforms by describing the services used during the trip such as the hotels and

African Journal of Hospitality, Tourism and Leisure, Volume 8 (3) - (2019) ISSN: 2223-814X

restaurant. These actions create a vast amount of information every day and constitute a huge amount of the dataset to create a 'big data' for the tourist destination. However, the social media content, such as comment, post, review, and tweet have aided the creation of big data extensively from platform providers (Ghani et al., 2018). The emergence of big data from social media in the tourism sector has brought about a new wave of excitement into the field of artificial intelligence and big data analytics.

In addition, the increase in the use of social media has offered considerable opportunities and challenges for investigators and practitioners in this sector. It has been shown in previous research that the latter can be directly influenced in different destinations by electronic word of mouth and word online social media (Arsal et al., 2010; Yacouel & Fleischer, 2012). The data which is shared by tourists, on the travel websites, about their travel experience leads to a large number of hotel reviews. These online reviews have become the leading resource for prospective tourists (Chaves et al., 2012; Serra Cantallops & Salvi, 2014).

One of the most famous travel websites used in previous studies is TripAdvisor which emerged in 2004 as a Web 2.0 application for the tourism domain and as social media where one can ask questions and share information in the travel forum. Tourists can leave an opinion and a numeric rating in TripAdvisor to briefly indicate their evaluation of services rendered. In some other cases, tourists can even add text comments to give more details about their consumption experience which is helpful and useful for other tourists. Certainly, the reviewers can give a positive rating or they can give it negatively, and this should help firms to deal with some issues and questions and enable them work on making the negative things better so as to gain benefits and win new customers through better service provision etc. Nowadays, hotel guest satisfaction is a very important topic because it contributes to customer loyalty, higher profitability and also leads to desired repeat purchases.

Tourism is an important sector in Morocco. It has an impact on the development of other sectors of activity and offers many economic spin-offs. In 2010, His Majesty King Mohammed VI signed the tourism strategy “Vision 2020” which aims at the development of eight tourist destinations: Souss-Sahara Atlantic, Morocco Mediterranean, Marrakech Atlantic, Central Morocco, North Cape, Atlantic Center, Great South Atlantic, Atlas and Valleys. Furthermore, this vision focuses on doubling the size of the sector and the accommodation capacity, with the creation of 200,000 new beds in hotels. This new capacity is expected to double tourist arrivals from Europe and emerging countries (attract 20 million tourists). Also, the tourism receipts are estimated to reach 140 billion Dirhams in 2020 (Stratégie du secteur du tourisme : la Vision 2020, 2011).

However, Moroccan tourism is still a growing industry and Marrakech is one of the most important tourist destinations and one of the most famous cities to visit in Morocco. In 2017, more than 7.619.256 tourist visited the city and the tourist arrivals rate increased from less than 12 percent compared with 2016 (Arrivées des touristes, 2017). Therefore, we have selected this destination in order to analyze the new habits of tourists, both on how the select their destination and in finding a suitable hotel.

To achieve our goal, we used a big data analytics and sentiment analysis of the hotel guest to promote intelligent tourism and thus contribute to its economic development. Thereafter, we analyzed hotel guest experience, characteristics, satisfaction and sentiment, so that the economic decision-makers will use those elements for improving hotel services, increasing the number of tourists and maintaining their loyalty as well as their satisfaction, therefore, trying to achieve the objective of the tourism strategy “Vision 2020” of Morocco.

Research background

Hotel guest experience and satisfaction

Hotel guest satisfaction is a complicated human experience in a hospitality service setting. The study of guest satisfaction was introduced as early as the 1970s. Various definitions of guest satisfaction have appeared. Oh (1997) showed that satisfaction engages cognitive and affective methods, as well as other psychological and physiological impacts.

In the tourism literature, a variety of perspectives have been employed to contemplate the technique of tourist satisfaction including the expectation/disconfirmation attitude, the perceived overall performance, the norm view, including the equity view (Yoon & Uysal, 2005). Among the managerial points of view, it is essential to realize the components or antecedents of hotel guest satisfaction. As an example, it has been conceptualized that the hotel services and products consist of various levels. That is, the core product, i.e., the hotel room, deals exactly with what the customer receives from the purchase. Besides, the hotel product can also be symbolized as a set of attributes as suggested by Dolnicar and Otter, (2003) and also Qu et al., (2000). These attributes include inter alia, room, location, services price/value, and security. The mostly mentioned two-factor theory postulates that hygiene elements like cleanliness and maintenance do not positively contribute to satisfaction, while dissatisfaction results from their absence, and this even though motivator elements such as the experiential aspects of staying at a hotel give positive satisfaction (Herzberg, 1966).

Certain scholars have used the service-dominant logic reasoning and argue that guest experience is not to be limited to what the hotel offers, but instead, it is co-generated from the service provider and the hotel guest (Chathoth et al., 2013). Xiang et al. (2015) indicated that the association between guest experience and satisfaction seems strong, indicating that these two domains of consumer behaviour are inherently connected. Therefore, customer satisfaction can be noticed as the customer's evaluation of his or her experience through association with diverseness of services areas.

Bearing in mind the complexity of the guest experience, measuring and managing hotel guest satisfaction is a challenging task. In fact, there is an improving effort in utilizing consumer-generated content to gauge guest satisfaction. Even though these studies make beneficial contributions to enrich our comprehension of guest satisfaction, they focused upon a relatively small sample of online data to make inferences, and therefore, they are restricted from the big data analytics perspective. Although these studies may have high levels of internal credibility, they may suffer to some extent from external credibility issues since it would be confusing to generalize their results on the grounds of relatively limited samples compared to huge data sets used in big data analytics.

Big data analytics in the hotel industry

Recently, modern tourism companies are applying Big data analytics-related techniques to get hidden insights into their data or to better know what customers like and do not like (Xu et al., 2017). Nonetheless, the big data analytics strategies in the social media context are associated with sentiment analysis, social media analysis, natural language processing, and opinion mining. These techniques play an essential role in improving the business and decision-making through analysis (Ghani et al., 2018). For instance, social media help businesses to obtain customers feedback relating to their products, which can be used to modify decisions and to purchase value out of their business (Xindong Wu et al., 2014).

In addition, existing studies have found that by taking advantage of big data analytics infrastructure, core-stakeholders can get “real-time information on tourists' on-site behaviors at tourism destinations" (Fuchs et al., 2014). The authors developed an interesting theoretical background for a "knowledge destination pattern," which includes an architecture that describes a series of components for the analysis of their data. However, other studies have introduced user-generated content and big data analytics of tourism-related research. For Tourist arrivals or hotel sales, the estimation is more precise if we use big data sources (Blal and Sturman, 2014). Yang et al. (2014) predicted hotel demand by combining traditional econometric models with web traffic volumes and exhibited the incorporation of web volumes to predict hotel occupancy in a tourist holiday destination. Q. Ye et al. (2009) inspected the effects of online consumer-generated reviews collected from the largest Chinese travel website concerning hotel room sales. The findings showed a considerable relationship between online reviews and business performance in hotels. G. Li et al. (2015) have shown that by using online review data from TripAdvisor we can also detect emergent hotel features of interests of international travelers. Their findings assisted hotel managers to gain insights into travelers' interests and offer a deeper understanding of tourist preferences. Z. Xiang et al. (2015) checked out the way big data analytics handle the relationship between hotel guest experiences and satisfaction. Thus, big data and text analytics can ascertain customer behavior and feedback.

Different from traditional data sources in tourism research, big data analytics includes a large amount of data without any sampling bias. Along with these new data sources, academia and businesses can improve and better comprehend visitor behavior in the tourism and hospitality fields. More specifically, big data analytics is a set of techniques that can be applied effectively to contribute to the objectives of the hotel industry and assist the organization to obtain competitive improvements or substantial gains.

Sentiment analysis in the hotel industry

Sentiment analysis inside the hotel industry websites and application, specifically online reviews' websites which are maintained, underpinned a new position in which the consumer is at the center of the network with different motivations (El haouta & Idelhadj, 2018) and preferences, leading to sharing of opinions (Liburd, 2012). In social media as an example, sentiment analysis has several purposes. More specifically, this analysis can be used to discover the emotions of consumers in a marketing and customer service department, which ends into finding whether consumers are satisfied or not with a service or/and product they receive (Povoda et al., 2017).

In tourism and hospitality industries several studies are based on online reviews which applied a sentiment analysis involving the process of extracting, classifying online user feedback mentioned in web content or digital documents, and obtaining resulting data against well-developed lexicons, and a variety of words or sentences grouped into neutral, positive, and negative (Aydogan & Akcayol, 2016). Park and Nicolau (2015) noticed that based on online reviews for tourism and hospitality, exchanges of information impact directly upon the consumer choices of hotels, concluding that for an online hotel positive review can increase the average probability of that consumer booking a room in the same hotel. Elements such as the number of stars have been revealed to favorably affect the score given by users on online reviews (Li et al., 2015). In fact, according to Phillips et al. (2015), to have more positive reviews, users expect higher rated hotels. Woo (2017), analyzed TripAdvisor scores and traditional customer satisfaction through travel intermediaries, and found that online reviews play an important role in clarifying hotel performance metrics and more so than traditional user's opinions.

In general, sentiment analysis contains three fundamental challenges: aspect detection (Bhattacharjee & Petzold, 2016), opinion word recognition, and sentiment placement identification (Paltoglou & Thelwall, 2012). Figure 3 shows a typical hotel review and an overall view of rating based on a variety of features of the hotel, including location, service, and check-in/out. However, it would be confusing for decision-makers to determine the reviewer's test of individual aspects. Knowing the distinctive aspect of each review is essential because the overall score may be similar across consumers' reviews. Therefore, using sentiment analysis would lead to a better understanding of customers' opinions and satisfaction.

Research questions

Online customer reviews have an important factor in the online sales of the hospitality and tourism industry in most countries of the world. Indeed, they have been generally considered important for comprehending consumer behaviour using the most powerful types of consumer-generated content, and therefore determine performance in the tourism industry (Browning et al., 2013; Serra Cantallops & Salvi, 2014; Sparks & Browning, 2011).

In different websites of travel such as for example TripAdvisor.com, and online travel agencies such as Expedia, users are allowed to post their ratings and reviews concerning their experiences with hotel characteristics they have stayed at in the past. Users reviews indicate the way consumers express, reconstruct, share their experiences, and relive them. Because other consumers are utilizing this information for travel planning purposes, customer reviews can have a huge influence on travel planning and eventually impact on their attitude and behaviour (Gretzel & Yoo, 2008). The rate of visitors' reviews has increased significantly in the last few years. For example, TripAdvisor in 2018 generated approximately 730 million user reviews and opinions covering over eight million listings for hotels, vacation rentals, restaurants, and attractions (TripAdvisor - Statistics & Facts, 2018). This prosperity of consumer-generated big data presents opportunities to explain and make statistical inferences about customers' behaviours in tourism and the hospitality industry in general. As a finding of the previous discussion, the following study questions were developed to guide the study:

RQ1.Through the big data extracted what are the hotel customer characteristics represented in guest reviews?

RQ2.Can hotel customer rating be employed to describe guest satisfaction?

RQ3.Can hotel customer experience presented in customer reviews be used to describe sentiment?

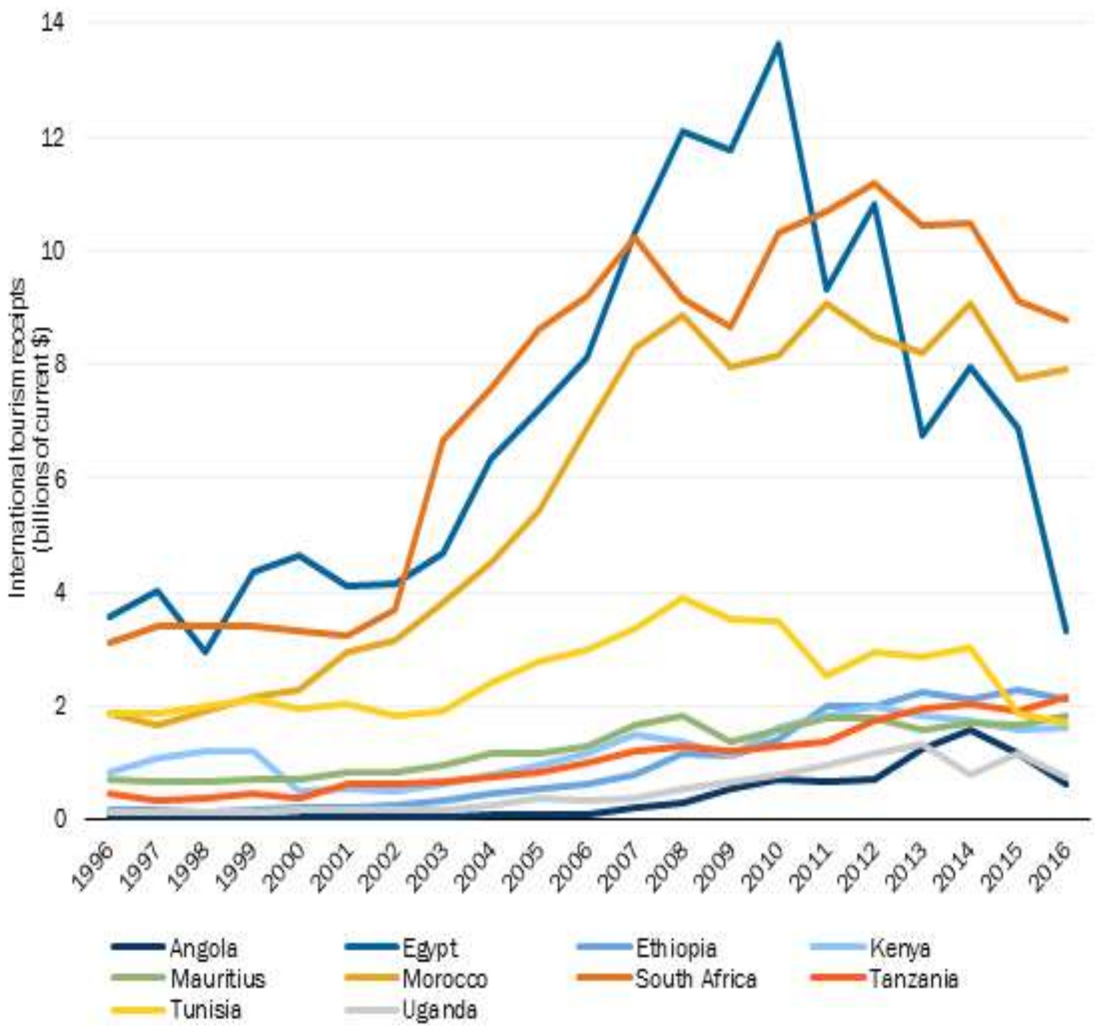
Research methodology

Data Collection

In the present work, we have exploited the largest travel website in the world, TripAdvisor.com which has more than 600 million travellers (TripAdvisor, 2018).

Firstly, we collected links from 7 five-star hotels in Marrakech, as shown in Figure 1. Then, we selected the language which contains a maximum of user reviews. Figure 2 represents a sample of one 5 star hotel, and it illustrates that French is the most dominant language of reviews in all selected hotels with more than 2.843 reviews. The same conclusion was noticed for the other hotels. Thereafter, we extracted “Review Text” combined with “Review Title”, “User Location”, “Travellers Type”, “Total Rating”, “Date and Hotel Services Rating”. Figure 3 demonstrates Screenshot of one review sample on Tripadvior.com. Finally, we filtered data by date extracting reviews from the period between 2015 and 2018.

The graph below demonstrates the upward trajectory of Morrocan tourism, making research such as this vital for future growth and sustainability.



Graph 1. Growth in tourism-related revenues in Africa’s largest markets, 1996-2016
Source: Signe, L. (2018). Africa's tourism potential - Trends, drivers, opportunities, and strategies, Brookings institution

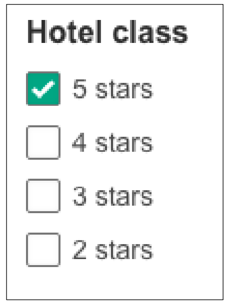


Figure 1. Classification of hotels by stars in TripAdvisor

Figure 2. The number of reviews and language on TripAdvisor (sample of one 5 star hotel)

Figure 3. Screenshot of one review sample on Tripadvior.com

Extracting big data

In this step, we used Selenium Webdriver with Python binding in ChromeDriver. Selenium WebDriver "allows commands transmitted in Selenese, or possibly via a Client API and transmits them to a browser" (Kudikala, 2014). The Python programming language used to extract data is justified by its rapidity, reliability and the minimum number of codes. Also, it is authorized and acceptable.

In order to extract and collect the elements, we used the XML Path Language (XPATH) and CSS selectors. The XPath is a query language for selecting nodes from an XML file (XML and XPath, n.d.). Figure 4 shows XPATH and CSS selectors' source code.

```
for elem in browser.find_elements_by_css_selector('div:not(.mgrRspnInline
    reviwe_user_from = ""
    try:
        reviwe_user_from = elem.find_element_by_xpath('.//parent::div[@cl
    except:
        pass
    review_txt = elem.find_element_by_xpath('.//p[@class="partial_entry"]
    review_title = elem.find_element_by_xpath('.//parent::div[@class="ui_
    review_date = datetime.strptime(elem.find_element_by_xpath('.//parent
```

Figure 4. XPATH and CSS selectors source code

Big data storage

The dataset of each hotel is saved in a CSV file which is saved in a folder named by the five-star hotel name. In this study, we have selected only 7 hotels for this category. The CSV file contains the dataset as shown in Table1.

Table 1. Extract of dataset from CSV file

Language	Total Rating	Traveller Type	Review	Cleanliness	Sleep Quality	Service	Date	User From	Location	Value	Rooms
French	4	Travelled as a couple	Des vacances aux calmes Nous avons passé un très bon moment dans cet hôtel l'accueil excellent, les chambres propres, les 2 piscines les jardins très propres et reposants les 2 restaurants sont très correctes	5	5	4	13/10/2018	Marne la vallée	3	3	4
French	5	Travelled with family	Magnifique séjour avec un personnel aux petits soins nous avons séjourné dans cet hôtel au mois de juin et des notre arrivée, l'accueil que chaleureux	3	4	5	14/08/2018	Mons;Belgique	4	5	4
French	3	Travelled as a couple	Peux mieux faire Hôtel situé à la palmeraie en retrait de l'effervescence néanmoins, bâtiments vieillissent, peignoirs au SPA	2	1	1	11/10/2018		5	2	5
French	3	Travelled with family	Plus jamais là bas, Ça fait 5 années de suite que je m'y rends à la même	5	1	1	11/10/2018	Paris, France	5	4	3
French	5	Travelled as a couple	Merveilleuses vacances au Pullman Marrakech ! nous avons passé un excellent moment dans cet hôtel Le lieu est magnifique, il y a une piscine réservé aux adultes	4	3	4	09/10/2018	Avellin, France	3	5	5
French	5	Travelled as a couple	Un week-end fabuleux Nous avons passès4 nuits dans l'hôtel Pullman et Nous avons eu le droit	5	3	4	07/10/2018	Paris, France	5	4	4

			à un accueil parfait, dès notre Nous avons profité du spa pour faire un message Le soir on a également profité du bar de l'hôtel,									
French	4	Travelled on business	Top service Le personnel de l'hôtel est souriant, accueillant,	3	2	4	05/10/2018	Brussels, Belgium	5	5	5	5

In the last step, we exploited the software library Apache Hadoop that "…allows the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage" (Apache Hadoop, n.d.). However, we loaded data from CSV file into Apache Hadoop Distributed File System (HDFS) that is the distributed file system designed to run on commodity hardware. Figure 5 summarizes social big data processing e.g., TripAdvisor, from the collection of the data until the end user.

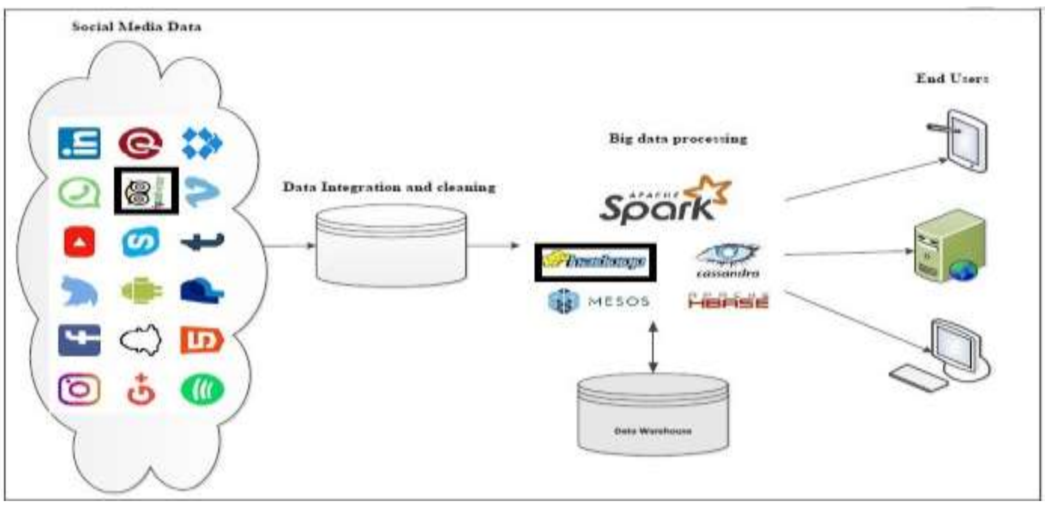


Figure 5. Big TripAdvisor data processing. Source: inspired by (Ghani et al., 2018)

Statistical analysis

In order to analyze the extracted data, we used Apache Hadoop and the R language. Subsequently and with the aim to set up a Hadoop cluster, we exploited IBM Bluemix. Then, the data were loaded into the Hadoop Distributed File System (HDFS).

In this part, the steps performed were as follows:

1. We installed RStudio that is an integrated development environment dedicated to R development.

2. To analyze the services rating during 2015 to 2018, we retrieved the FHSS data using the RHDFS R software package and we used a barplot (ggplot2 R Package) with the service name as an argument. This step is important to understand the evolution of services such as location and sleep quality.

3. In order to detect the origin countries of online users, we used ggplot package and tidytext package to format the text in the country row fields. In this step, we started to define the revision rate per country from 2015 to 2018.

4. We used a ggplot2 package to evaluate the calculated rate of hotels' guest types during all months of 2015 up until 2018.

5. Finally, the opinions of the hotel users were analyzed by exploiting the positive and negative expressions extracted from the review text. The tools used in this step are Tidytext Package with “bing” Lexicon considered as a collection of information about the words of a language and the lexical categories to which they belong. There are a variety of methods and dictionaries for evaluating the opinion or emotion in text. The tidytext package contains several sentiment lexicons.

To achieve this step we contrasted tidy text with our data structure. Then, we unset tokens using Tidytext unnest_tokens function. Thereafter, we tidied up the data frame by removing stop words and mutated the data; Loading bing lexicon, joining bing sentiment with our data and plot the data using ggplot package. In this step, we used the tools of text mining to approach the emotional content of text programmatically, as indicated in figure 6.

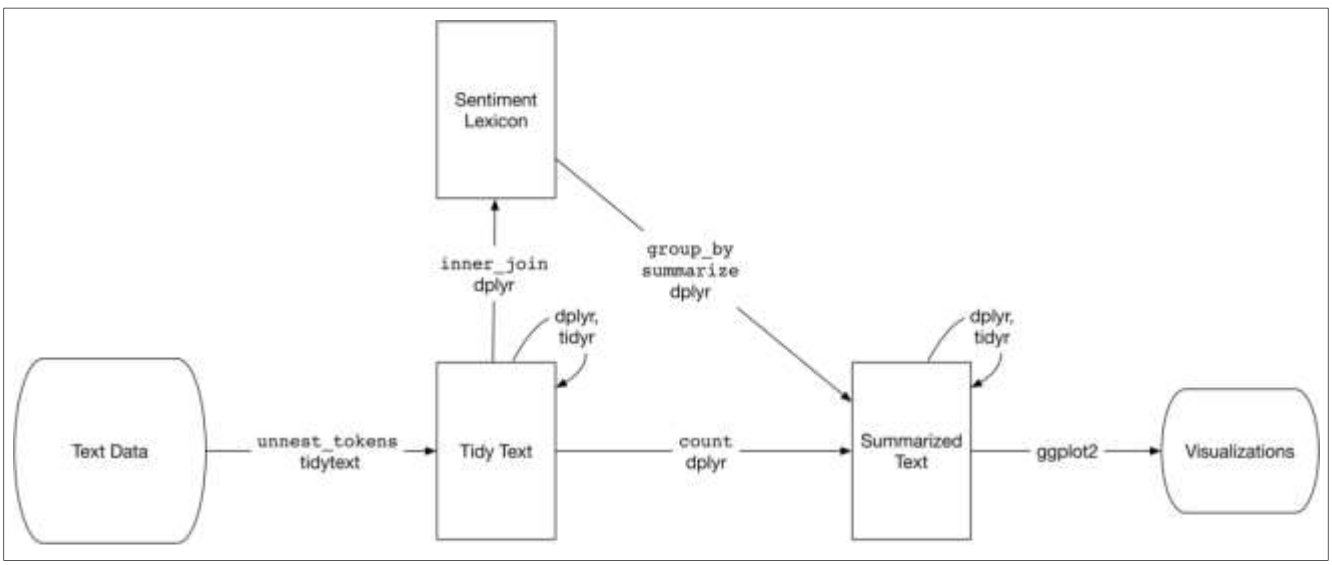


Figure 6. Tools of text mining for the emotional text content

We also created a WordCloud using R wordcloud library which uses base R graphics to tag positive and negative words. This step is done with joins, piping, dplyr and reshape2 packages and functions.
An example of code Tidytext Package:

```
“ bing_word_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ggplot(aes(reorder(word, n), n, fill = sentiment)) +
  geom_bar(alpha = 0.8, stat = "identity", show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Contribution to sentiment", x = NULL) +
coord_flip()
## Selecting by n ”
```

We can view this visually to assess the top n words for each sentiment (Text Mining: Sentiment Analysis,n.d.).

Results

Figure 7 illustrates that the type of hotel guests with 'family' represents the largest growth rate for every month of the period from 2015 to 2018 excluding January, September, and December. Furthermore, it was noted that the type 'as a couple' fast increases in January.The type 'with friends', that seems to be present in specific months from March to July, October and November, is a considerable rate during October. Moreover, people who came for business are more expressive in particular months like May, September, November, and December.
Finally, singles hotel guests expressed with type 'solo' became more active in April and from September to November.

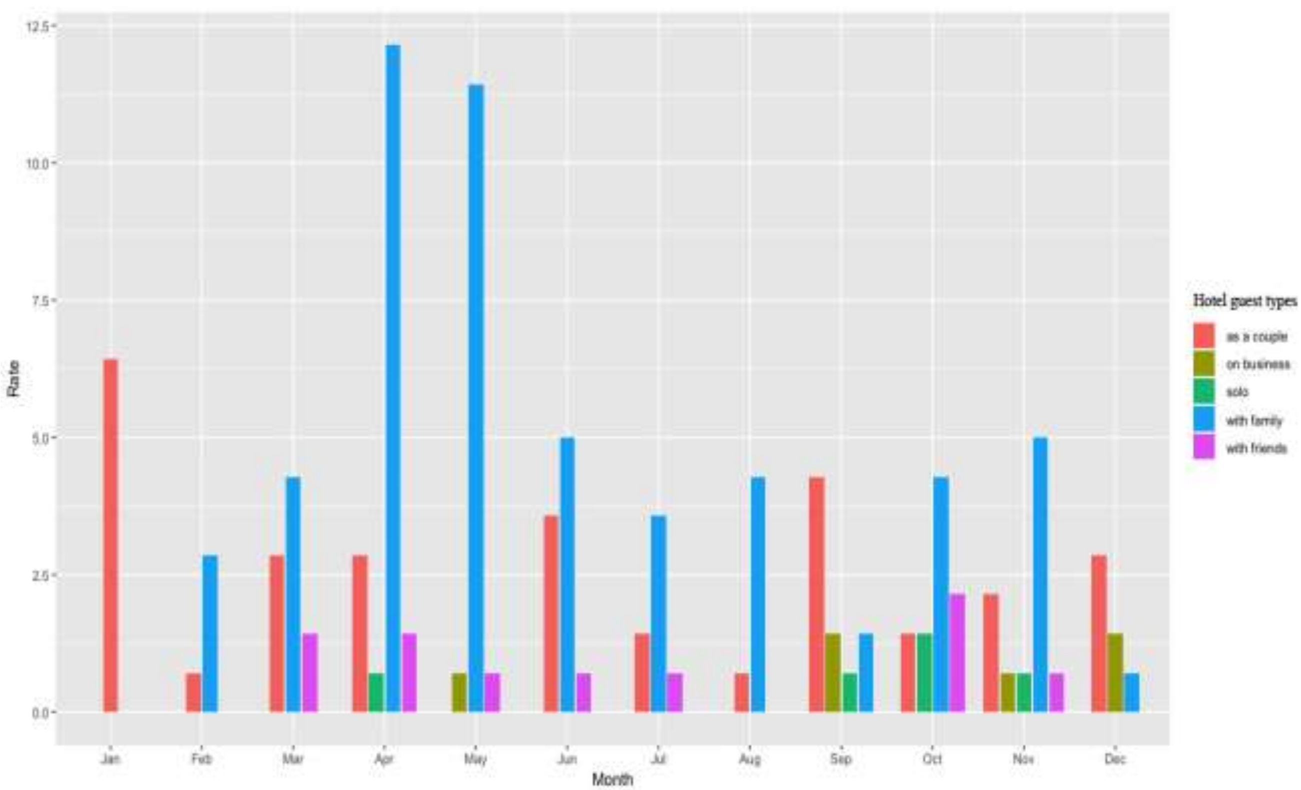


Figure7. The rate of hotel guest types in the period 2015 to 2018

Figure 8 shows that during the period from 2015 to 2016, more than half of the guest rating was from France. We note that this rate decreased in 2018. Turkey and Switzerland were added to reviews in 2016.Similarly to Belgium country which added reviews with a low rate in 2017.

African Journal of Hospitality, Tourism and Leisure, Volume 8 (3) - (2019) ISSN: 2223-814X

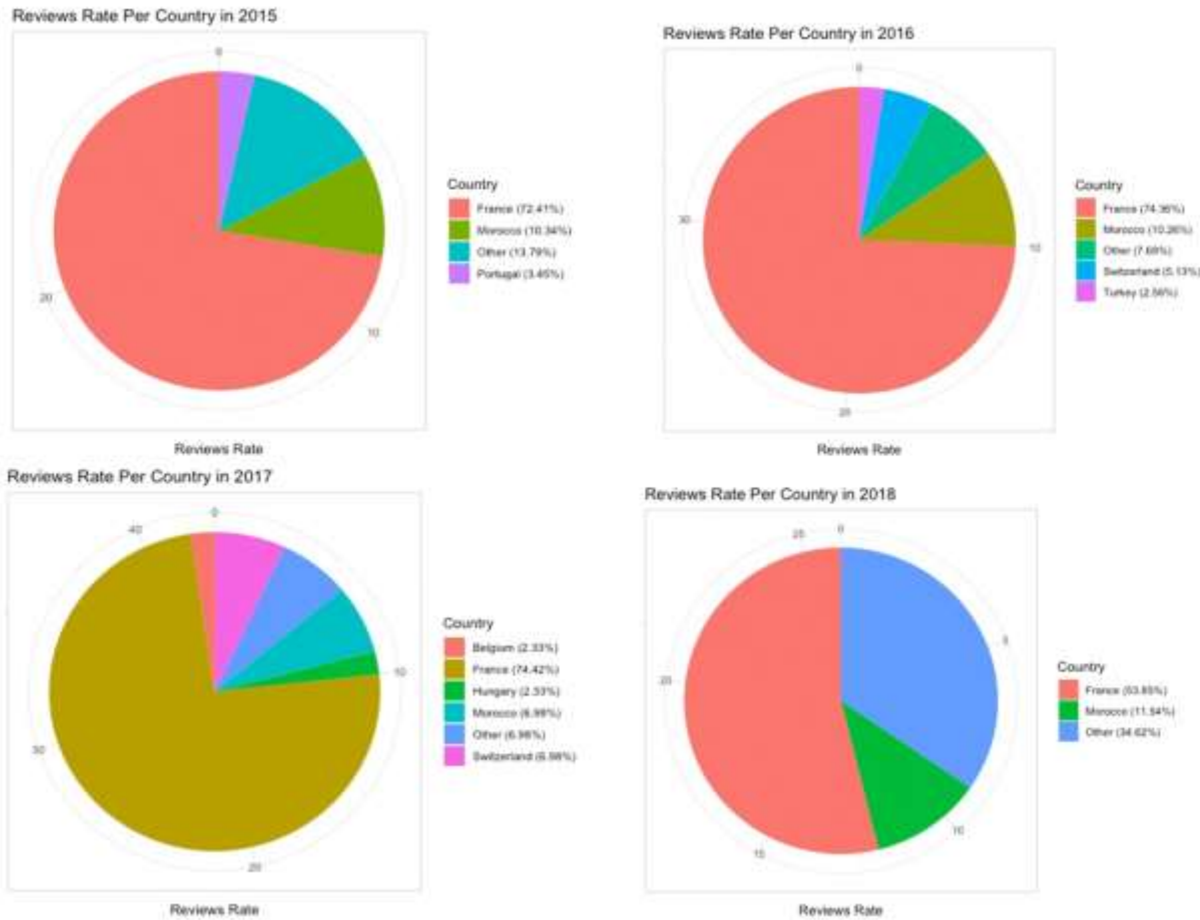


Figure8. Reviews rate per country from 2015 to 2018

As demonstrated in figure 9, France is the country with the highest reviews rate by 70.07%, flowed by other countries (13.87 %) and Morocco (9.49 %).However, Switzerland (3.65 %), Turkey (0.73 %), Portugal (0.73 %), Hungary (0.73 %) and Belgium (0.73 %) have the lower reviews rate.

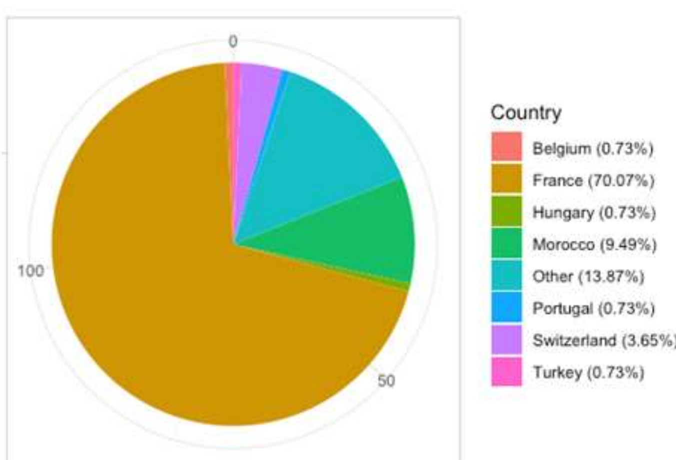


Figure 9. Reviews rate per country from 2015 to 2018

As Figure 10 has shown, all services ratings have reduced in the period from 2015 to 2018:

- Value (Value default ranking) rate decreased from 4.15 to 3.59.
- Sleep Quality (Bad/Good sleep quality) rate decreased from 4.53 to 4.03.
- Rooms (Bad/Good room) rate decreased from4.40 to 4.03
- Location rate decreased from 4.20 to 3.79
- Cleanliness rate decreased from 4.38 to 4.14.
- Except for the rate of the field Service, which is increased in 2016 (until 4.09) and then, during the period of 2017-2018, it is decreased.

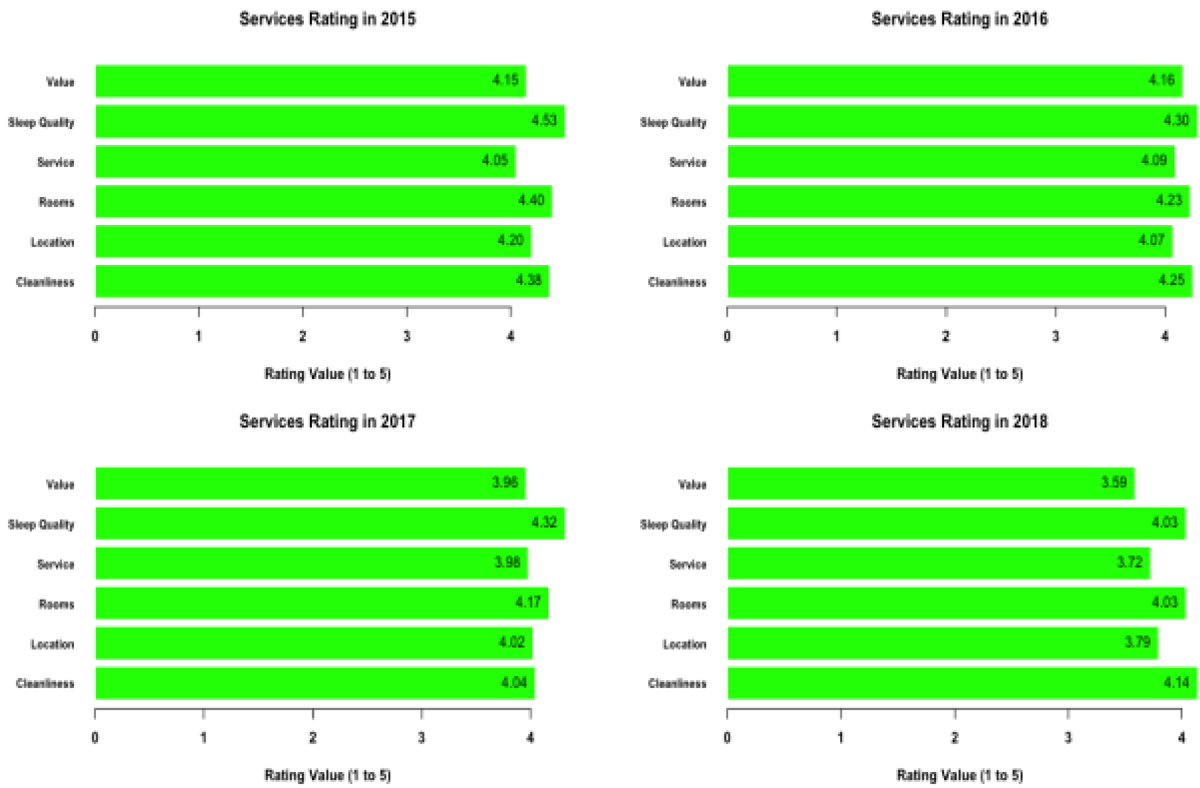


Figure 10. Services rating in 20 hotels (5-star) – Rating value from 1 to 5

According to figure 11 which provides the list of 20 guest experience-related words, the dominant language of customer reviews is French and the translation of words into English is in brackets. These words were used to explain (dis)satisfaction based on the online reviews in 7 five-star hotels inside Marrakesh and reflect a wide spectrum of aspects related to the hotel guest experience, including (1) hotel staff-related descriptors such as “bravo”, “adorable” , and “correct (adequate) ”, (2) evaluation of experience such as “sale (dirty) ”, and “infect (foul)”; (3) expression of (dis)satisfaction such as “super”, “top", "formidable (great)”, “excellent ”, “terrible”, and “médiocre (mediocre)”,(4) rooms attributes such as “grand (huge) ”; and (5) actions such as “limitations”. Therefore, this list does not reflect certain aspects of guest experience such as stay at the hotel due to word-of-mouth (recommendations).

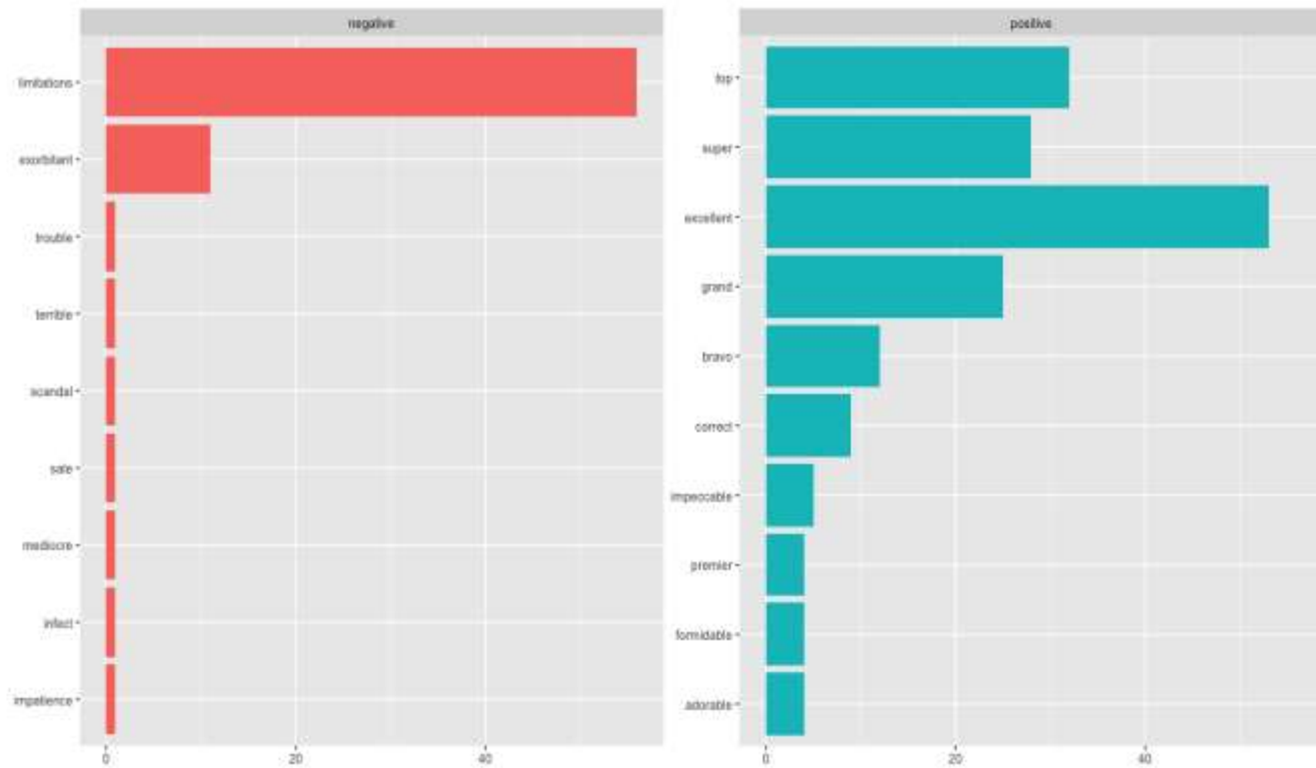


Figure 11. Top 20 primary words in hotels customer reviews

Figure 12 shows the Word Cloud produced from TripAdvisor data about five-star hotels in Marrakech, providing a visual presentation of the results.

Figure12. Word Cloud for in hotels customer reviews

Discussion

Hotel guest satisfaction was extensively examined in the documentation which demonstrated that the guest experience is certainly a complicated construction. Depending upon the study pattern and methods, researchers might find very different illustrations of what comprises guest experience and what essentially leads to guest satisfaction (Crotts et al., 2009; Lockyer, 2005). Since traditional methods generally depend on a set of predetermined hypotheses, justified in utilizing the previous and existing body of knowledge, the attempt is created in the direction of either confirming or disconfirming such hypotheses. Nevertheless, this is not really the case with big data analytics.

In the present study, we used the big data to display hotel guest characteristics such as type, nationality and behavior, in order to answer question 1 (RQ1) about the hotel customer characteristics represented in guest reviews. However, several types of hotel guests were detected: 'with family' represents the largest rate, followed by 'as a couple',' with friends', 'for business 'and 'solo'. Concerning the behavior of clients, we found a diversity that depends on their type, e.g., the type 'with family' seems to comments and expresses his (dis)satisfaction every month, except in January, September, and December. Our result about the hotel guest's nationality, illustrated in figures 8 and 9, showed that France is the country with the highest rate (70.07 %). This result is similar to the statistics published by the Ministry of Tourism of Morocco which indicates that the highest rate of overnight stays in tourist accommodation establishments in Morocco are made per French tourists (Fréquentation hôtelière, n.d.).

In response to question 2 (RQ2), about the hotel guest satisfaction through the online ratings, we found, as presented in figure 10, that the field of 'Service' was decreased in the last two years (2017 and 2018). However, we tried to make sense and add meaning to the inferences by getting appropriate results to shed light on and clarify guest sentiment-experience (RQ3) as we have shown in figure 11 and figure 12. Thus, we extracted 20 guests experience-related words which were used to explain (dis)satisfaction and remarks in online reviews. The words "excellent" and "limitations" were clearly the main hotel attributes mentioned by the customers, especially in measuring the main reasons for guest satisfaction in this tourism destination.

The traditional methods which discuss the findings are a part of the epistemology of producing and creating knowledge employing big data (George et al., 2014). This knowledge also confirms the capabilities of big data analytics to determine the nature and level of customer satisfaction that is not measured in the conventional sense. The three areas of guest behaviour, experience and satisfaction, are inevitably and naturally associated. The guest satisfaction is determined as the average rating of all guest hotels who reviewed the same hotel. The influence of these words in guest reviews could have been evened out. The connection could be even more powerful if the analysis had been utilizing cases involving several categories of hotels.

Furthermore, there is an increasing amount of literature which uses online reviews to understand how different product and service elements lead to hotel guest satisfaction (Crotts et al., 2009; Zhang and Cole, 2016). A research based on a website for travel review data showed that in online hotel reviews, the mention of "hygiene" elements such as room maintenance and cleanliness usually take place in a negative context, and have the effect of blocking the guest from posting any positive experience (Xiang et al., 2015).

This shows that it is possible to determine, through user-generated content such as online reviews, the significant structures among a variety of elements and characteristics. The present study is one of the encouraging study aspects of social media analytics in hospitality and tourism in Morocco.

We have successfully clarified the pattern of guest experience as mentioned in online reviews and rating to investigate guest satisfaction. The results revealed that the big data analytics is expected to improve our existing knowledge about hotel customer experience and satisfaction levels.

Conclusion

Although big data analytics has been suggested as a new investigation paradigm in many specialties, we have observed very few applications in the discipline of hospitality that entirely examine its abilities. In this paper, the social big data analytics were used to identify a large number of online customer reviews, examine the quality of these data, as well as to detect inherent associations between hotel management and satisfaction; also, understanding the nature and behavior of the guests in five-star hotels in Marrakesh. The originality of this investigation lies in the use of a large data and delineation of guest experience drivers on a level that was not presented in traditional guest survey studies.

However, this study is a basic effort in big data analytics; it offered a significant insight into some extensively studied produces in hospitality. These findings are very useful because when the hotel manager has information, such as the types of guests and their periods of stay, he becomes able to market this feature to maximize their interest. Our results confirm the effectiveness of big data analytics and sentiment analysis. Thus, several "types" of hotel guests were detected. The representative type is the "with the family" type. In terms of hotel guest behavior, it has been varied according to their type. For the "nationality" of hotel guests, the most dominant rate was France.

On the other hand, the “rating” was par excellence in expressing the satisfaction for different fields such as service, location, sleep quality, etc. However the “sentiment-experience” expressed 20 guests experience-related words which were used to explain (dis)satisfaction and remarks in the online reviews. The words “excellent” and “limitations” were clearly the main hotel attributes mentioned by the customers.

Thus, at least, the findings can be seen as obvious elements based on the big data analysis. This study is a starting-point that will be a basis for further studies including for example, all categories of hotels and other information about hotel guests such as their ages. In addition, future research may choose to extract social big data from multiple online travel platforms in order to obtain more information about the customer experiences in order to improve knowledge about customer satisfaction using big data analytics.

References

Apache Hadoop. (n.d). [Accessed 14/11/2018]. Available online at: https://hadoop.apache.org.

Arrivées des touristes. (2017). Ministère du Tourisme, du Transport Aérien, de l’Artisanat et de l’Economie Sociale - Département du Tourisme. [Accessed 20/11/2018]. Available online at: https://www.tourisme.gov.ma/fr/tourisme-en-chiffres/arrivees-des-touristes.

Arsal, I., Woosnam, K.M., Baldwin, E.D. & Backman, S.J. (2010). Residents as Travel Destination Information Providers: An Online Community Perspective. *Journal of Travel Research*, 49, 400–413. Available online at https://doi.org/10.1177/0047287509346856

Aydogan, E. & Akcayol, M.A. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques, in: 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, Sinaia, Romania, pp. 1–7. Available online at https://doi.org/10.1109/INISTA.2016.7571856

Bhattacharjee, K. & Petzold, L. (2016). What Drives Consumer Choices? Mining Aspects and Opinions on Large Scale Review Data Using Distributed Representation of Words, in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), IEEE, Barcelona, Spain, pp. 908–915. Available online at https://doi.org/10.1109/ICDMW.2016.0133

Blal, I. & Sturman, M.C. (2014). The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hospitality Quarterly*, 55, 365–375. Available online at https://doi.org/10.1177/1938965514533419

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, 23–40. Available online at https://doi.org/10.1080/10548408.2013.750971

Chathoth, P., Altinay, L., Harrington, R.J., Okumus, F. & Chan, E.S.W. (2013). Co-production versus co-creation: A process based continuum in the hotel service context. *International Journal of Hospitality Management* 32, 11–20. Available online at https://doi.org/10.1016/j.ijhm.2012.03.009

Chaves, M.S., Gomes, R. & Pedron, C. (2012). Analysing reviews in the Web 2.0: Small and medium hotels in Portugal. *Tourism Management*, 33, 1286–1287. Available online at https://doi.org/10.1016/j.tourman.2011.11.007

Crotts, J.C., Mason, P.R. & Davis, B. (2009). Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry: An Application of Stance-Shift Analysis to Travel Blog Narratives. *Journal of Travel Research*, 48, 139–151. Available online at https://doi.org/10.1177/0047287508328795

Dolnicar, S. & Otter, T. (2003). Which Hotel attributes Matter? A review of previous and a framework for future research. *Faculty of Commerce - Papers (Archive)* 23.

El haouta, I. & Idelhadj, A. (2018). Tracking community motivation to use online social networks: a case study of Facebook. *Journal of Theoretical and Applied Information Technology*, 96, 13.

Fréquentation hôtelière. (n.d). Ministère du Tourisme, du Transport Aérien, de l'Artisanat et de l'Economie Sociale - Département du Tourisme. [Accessed 20/01/2019]. Available online at: https://www.tourisme.gov.ma/fr/tourisme-en-chiffres/frequentation-hoteliere.

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, 198–209. Available online at https://doi.org/10.1016/j.jdmm.2014.08.002

George, G., Haas, M.R. & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal* 57, 321–326. Available online at https://doi.org/10.5465/amj.2014.4002

Ghani, N.A., Hamid, S., Targio Hashem, I.A. & Ahmed, E. (2018). Social media big data analytics: A survey. *Computers in Human Behavior*. Available online at https://doi.org/10.1016/j.chb.2018.08.039

Gretzel, U. & Yoo, K.H. (2008). Use and Impact of Online Travel Reviews, in: O'Connor, P., Höpken, W., Gretzel, U. (Eds.), Information and Communication Technologies in Tourism 2008. Springer Vienna, Vienna, pp. 35–46. h Available online at ttps://doi.org/10.1007/978-3-211-77280-5_4

Herzberg, F.I. (1966). Work and the nature of man, Work and the nature of man. World, Oxford, England.

Kudikala, R. (2014). Selenium automation framework: Selenium Testing Complete Tutorials. Selenium automation framework. [Accessed 13/02/2019]. Available online at: https://simpleseleniumwithjava.blogspot.com/2014/01/selenium-testing-complete-tutorials.html

Li, G., Law, R., Vu, H.Q., Rong, J. & Zhao, X. (2015). Identifying emerging hotel preferences using Emerging Pattern Mining technique. *Tourism Management*, 46, 311–321. Available online at https://doi.org/10.1016/j.tourman.2014.06.015

Liburd, J.J.(2012). Tourism research 2.0. *Annals of Tourism Research*, 39, 883–907. Available online at https://doi.org/10.1016/j.annals.2011.10.006

Lockyer, T. (2005). The perceived importance of price as one hotel selection dimension. *Tourism Management,* 26, 529–537. Available online at https://doi.org/10.1016/j.tourman.2004.03.009

Oh, H. (1997). Customer Satisfaction and Service Quality: A Critical Review of the Literature and Research Implications for the Hospitality Industry. *Hospitality Research Journal,* 20, 30.

Paltoglou, G. & Thelwall, M. (2012). Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems and Technology,* 3, 1–19. Available online at https://doi.org/10.1145/2337542.2337551

Park, S. & Nicolau, J.L. (2015). Asymmetric effects of online consumer reviews*. Annals of Tourism Research* 50, 67–83. Available online at https://doi.org/10.1016/j.annals.2014.10.007

Phillips, P., Zigan, K., Santos Silva, M.M. & 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. Available online at https://doi.org/10.1016/j.tourman.2015.01.028

Povoda, L., Burget, R., Dutta, M.K. & Sengar, N. (2017). Genetic optimization of big data sentiment analysis, in: 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, Noida, Delhi-NCR, India, pp. 141–144. Available online at https://doi.org/10.1109/SPIN.2017.8049932

Qu, H., Ryan, B. & Chu, R. (2000). The Importance of Hotel Attributes in Contributing to Travelers' Satisfaction in the Hong Kong Hotel Industry. *Journal of Quality Assurance in Hospitality & Tourism,* 1, 65–83. Available online at https://doi.org/10.1300/J162v01n03_04

Serra Cantallops, A., Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management,* 36, 41–51. Available online at https://doi.org/10.1016/j.ijhm.2013.08.007

Signe, L. (2018). Africa's tourism potential - Trends, drivers, opportunities, and strategies, Brookings institution. Available online at https://www.brookings.edu/wp

Sparks, B.A. & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32, 1310–1323. https://doi.org/10.1016/j.tourman.2010.12.011

Stratégie du secteur du tourisme : la Vision 2020. (2011). portail sud maroc. [Accessed 20/01/2019]. Available online at: https://www.portailsudmaroc.com/actualite/vision-2020-maroc.

Text Mining: Sentiment Analysis. (n.d). UC Business Analytics R Programming Guide. [Accessed 15/02/2019]. Available online at: https://uc-r.github.io/sentiment_analysis.

TripAdvisor - Statistics & Facts. (2018). Statista. [Accessed 30/02/2019]. Available online at: https://www.statista.com/topics/3443/tripadvisor/.

TripAdvisor : lisez des avis, comparez les prix et réservez. (2018). TripAdvisor. [Accessed 30/02/2019]. Available online at: www.tripadvisor.fr/.

Woo, G.K. (2017). Social media review rating versus traditional customer satisfaction. *International Journal of Contemporary Hospitality Management*. 29(2), 784-802. Available online at https://doi.org/10.1108/IJCHM-11-2015-0627

Xiang, Z., Schwartz, Z., Gerdes, 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. Available online at https://doi.org/10.1016/j.ijhm.2014.10.013

Xindong Wu, Xingquan Zhu, Gong-Qing Wu, & Wei Ding. (2014). Data mining with big data. IEEE *Transactions on Knowledge and Data Engineering* 26, 97–107. https://doi.org/10.1109/TKDE.2013.109

XML and XPath. (n.d). w3schools.com. [Accessed 30/03/2019]. Available online at https://www.w3schools.com/xml/xml_xpath.asp.

Xu, X., Wang, X., Li, Y. & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management*, 37, 673–683. Available online at https://doi.org/10.1016/j.ijinfomgt.2017.06.004

Yacouel, N. & Fleischer, A. (2012). The Role of Cybermediaries in Reputation Building and Price Premiums in the Online Hotel Market. *Journal of Travel Research,* 51, 219–226. https://doi.org/10.1177/0047287511400611

Yang, Y., Pan, B. & Song, H. (2014). Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data. *Journal of Travel Research,* 53, 433–447. Available online at https://doi.org/10.1177/0047287513500391

Ye, Q., Law, R. & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management,* 28, 180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

Yoon, Y. & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management*, 26, 45–56. Available online at https://doi.org/10.1016/j.tourman.2003.08.016

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. Available online at https://doi.org/10.1016/j.tourman.2015.09.001