

Using Big Data and Text Analytics to Understand How Customer Experiences Posted on Yelp.com Impact the Hospitality Industry

Pei-Ju Lucy Ting
National Taipei University
E-Mail: lucyting@gmail.com

Szu-Ling Chen
National Taipei University
E-Mail: nikilo1633@gmail.com

Hsiang Chen
National Taipei University
E-Mail: c650113@hotmail.com

Wen-Chang Fang
National Taipei University
E-Mail: wenchang.fang@gmail.com

ABSTRACT

This study combines programming and data mining to analyze consumer reviews extracted from Yelp.com to deconstruct the hotel guest experience and examine its association with satisfaction ratings. The findings show many important factors in customer reviews that carry varying weights and find the meaningful semantic compositions inside the customer reviews. More importantly, our approach makes it possible to use big data analytics to find different perspectives on variables that might not have been studied in the hospitality literature.

Keywords: Big Data, Text Analytics, Data Mining, Social Website, Guest Experience Satisfaction, Hotel Management

INTRODUCTION

Big data has become a popular area of research with the potential to add huge value of products and services to industry and business (Ang & Seng, 2016). The changing rate of generated data comes from the rapid growth of Internet of Things, cloud computing, and efficacy improvement of the search engine, leading to the growth of big data and more opportunities for data centers (Mehdipour, Noori, & Javadi, 2016). Big data analytics focuses on the collection of data with an unprecedented breadth, scale, and depth to solve actual problems (Mayer-Schonberger & Cukier, 2014). The increasing need to solve such problems is driven by companies such as Facebook, Google, LinkedIn that have to find a way to feed large-scale analytic engines to produce additional value services such as recommender systems, customer analytics, and social network analytics (Ang & Seng, 2016). At the same time, many social media websites, especially in the hospitality industry, that operate booking services produce large quantities of user-generated data. Researchers hope to use this data to gain insights into research problems which have not well been understood by traditional methods (Yang, Pan, & Song, 2014). Indeed, over the past few decades, businesses have generated more web data than they can or know how to use (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Therefore, analyzing primary data from individual consumers will lead to entirely new ways of understanding consumer behavior and formulating marketing strategy (Erevelles, Fukawa, & Swayne, 2016). The goal of this study is to explore the possible utility of online guest reviews by using big data analytics to finding the important text of different hotel that have not yet been studied. User-generated content in the hospitality industry, which in this study is online customer reviews, could cover many topics of interest to hotel managers because it contributes to electronic word-of-mouth (e-WOM) information (Bailey, 2005), customer loyalty, and reach purchase decisions in relation to tourism and hospitality products and services (Browning, So, & Sparks, 2013). The hotel industry offers highly competitive services and products, so each hotel must distinguish itself. Customer satisfaction has therefore become a key factor in measuring a company's competitiveness and success (Bitner & Hubbert, 1994). Although many studies have used customer reviews to explore comprehensive factors to understand the components of and the relationship of guest satisfaction (Wu & Liang, 2009), and suggest different viewpoints on e-mouth of mouth research, most studies have relied on conventional methods such as focus group interviews or customer surveys to gauge what leads to guest satisfaction (Xiang, Schwartz, Gerdes, & Uysal, 2015). Therefore, whether we can explore meaningful

insights into customer experience related to guest reviews, using big data as a new research method to decode the hospitality meaning in it is an important research question.

This study uses online customer reviews to gain behavioral insights about consumers and to find the useful keywords. Customer reviews are collected from Yelp.com and text analytics is used to explore the answer. These data mining methods could find the meaningful structure of customer reviews when customers assign a rating and post a comment about a hotel at which they have stayed.

This study is structured as follows. After the introduction, the second section reviews the literature on big data, text analytics, hotel guest experience and customer reviews. Two research questions are formulated and explored to understand the meaning of customer reviews. The section on research method details the data collection and the text mining techniques that are used to answer the research questions. The results are then presented. Finally, the study's contributions to practice and literatures as in addition to directions for future research and limitation are discussed.

LITERATURE REVIEW

Big Data and Text Analytics

Many people are familiar with the term “big data,” which can be defined as unprecedented volume, velocity, and the variety of primary data due to the advance of technology (Erevelles et al., 2016). Thousands of information datasets are generated in social media, mobile transactions and user-generated content; there is also purposefully generated content through sensor networks or business transactions such as sales queries and purchase transactions (George, Haas, & Pentland, 2014). As data become larger, more complex, and more inexplicable, the application of big data has been investigated in fields like genomics, healthcare, engineering, operations management, the industrial Internet, and finance. Unlike traditional analytics that are concerned with small data sets with limited analytics platforms, big data analytics work with much larger, unstructured, and complex data sets that require new and advanced technologies to store, manage, and analyze (Chen, Chiang, & Storey, 2012). The best-known example of application of big data analytics is Google's search engine, which uses keyword search trends to produce a daily estimate, the “now cast,” of the occurrence of flu among a population two weeks before the publication of official government data (Ginsberg et al., 2009). The methods used in this study made it possible to use search queries to detect influenza epidemics in areas with large populations of web

search users. Unlike a uniform method, big data takes advantage of several kinds of analytical tools to use large data to predict trends and gain business value.

Another application of big data analysis that is growing in popularity is business intelligence. Numerous e-commerce companies use highly scalable e-commerce platforms and social media platforms to expand the volume of unstructured data as digital text. Managers can then base their decisions on the information extracted from these sources (Ittoo, Nguyen, & van den Bosch, 2016). Web data posted on social media website express posters' opinions and their feelings and understand customers, competitors, products, business environment, impact of technologies, and strategic stakeholders such as alliance and suppliers (Xiang et al., 2015). In the past, scholars have demonstrated that Internet information can be used to predict product quality and improve product design (Finch, 1999). The study shows that web information postings have enough linguistic content to predict a firm's earnings and stock returns (Tetlock, Saar-Tsechansky, & Macskassy, 2008). In addition, many researchers have used user-generated online reviews to estimate their helpfulness and economic impact (Ghose & Ipeirotis, 2011) and change the way consumers shop online by using marketing tools such as search engines and recommendation systems (Ghose, Ipeirotis, & Li, 2012).

To process large amounts of user-generated data in social media and extract meaningful knowledge and insights from it, text mining, sentiment analysis and social network analysis have been widely used (He, Wu, Yan, Akula, & Shen, 2015). Text mining can extract meaningful unstructured textual data (He, Zha, & Li, 2013; Hung & Zhang, 2008) and find useful models, trends, patterns, or rules (He, 2011; Romero, Ventura, & García, 2008). Text mining can improve the user experience (He, 2013) and improve the understanding of customer behavior and patterns (Chau & Xu, 2012). One famous example of text analytics is Netflix, which exploits huge volumes of data collected from members to improve the quality of the movies (Ittoo et al., 2016). In addition, opinion mining and sentiment analysis combines natural language processing (NLP), information retrieval (IR), structured and unstructured text analysis and computational linguistics to identify the text sentiment (Ravi & Ravi, 2015; Vinodhini & Chandrasekaran, 2016). They can be valuable in understanding customer satisfaction, and in predicting market trends and stock prediction (Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014). However, although much research has been done in text mining, it is important to explore the insight and meaning of unstructured data.

Customer Reviews

Consumer-generated content contains a variety of media forms and types (Gretzel, 2006). One form in which content is created online is as consumer reviews and ratings. Online reviews, which reflect the way consumers describe and share their experiences, is a valuable way to find out what consumers think. Due to the advance of technology, it is easy for consumers to post their experience with products and services online. Over 30 percent of Internet users have given the comment and rate products online (Boase et al., 2006). They also give product descriptions, reviews from experts, and personalized advice generated by automated recommendation systems. This type of customer review can have a significant effect on purchase decisions (Park, Lee, & Han, 2007). Many studies have demonstrated the strong impact of online customer reviews. For example, Dellarocas, Zhang, and Awad (2007) have demonstrated that online review metrics can accurately forecast movie revenue. Chevalier and Mayzlin (2006) used the number of book reviews to reveal that more online book reviews increased the same of the books that were reviewed. Mudambi and Schuff (2010) concluded that review depth has a greater positive effect on the helpfulness of the review for search goods than does experience goods or service. Minnema, Bijmolt, Gensler, and Wiesel (2016) have demonstrated that products returns have strong relationship with online customer reviews and the effect of it need to be considered. While researchers have demonstrated the effect of the presence of customer reviews to an online retailer, a part of unanswered research issue is what makes customer reviews helpful to a consumer in the process of making a purchase decision (Mudambi & Schuff, 2010). In the hospitality industry, websites like TripAdvisor.com, Expedia, and yelp.com allow customers to rate and review hotels and restaurants. The number of reviews is increasing more the past decades, and the literature has discussed the relationship between online guest reviews and hotel performance. For example, Ye, Law, and Gu (2009) argue that online reviews can predict hotel booking since a high number of reviews is related to increase in hotel booking. There is a strong relationship between online consumer reviews and business performance of hotels. Xiang et al. (2015) use online reviews to explore the nature and underlying structure of hotel guest experience and its relationship to satisfaction. However, they pointed out that future researchers needed more evidence and more sophisticated method to support the positive impact of the number of reviews on hotel performance.

Research Questions

1. What are the texts customers usually leave online in their e-word of mouth?
2. What keywords might help hotel business industry to manage the hotel and help customers to find a suitable hotel?

METHODOLOGY

Research Design

In this study, we use text analytics to understand hotel guest experiences that are posted on the social media website and its association with satisfaction. Created in 2004, Yelp.com invites customers to write reviews and rate their experiences with businesses. Yelp had received more than 95 million reviews by the end of 2015 (About Yelp.com, 2015). Yelp.com visitors discuss their preferences or their recommendations. They can also search for a business by name, location or type and read reviews. In this way, a business can learn what customers think and make changes accordingly. Yelp.com and its users thus can influence whether a business succeeds or fails (Sax, 2011).

Figure 1 shows the webpage of hotel taken from Yelp.com. The hotel (name concealed) can have several customer reviews. Reviews can be updated, and include a five-star rating system. Reviewers can supply a name, photo, and personal information. On Expedia and booking.com, only hotels that partner with the website can be reviewed, but Yelp collects reviews on almost all hotels in the United States. Therefore, the site has more reviews. Guests are asked to post their positive and negative reviews separately. Also, the average rating of each hotel is given by guests, and so it is more authentic.

Data Collection

The data for this study was collected from Yelp.com using a self-designed web crawler written in python. We use Scrapy as our web crawler framework. The program contains spider.py, item.py, and pipeline.py. The function of spider.py is to define crawling website domain, define data crawling rules, perform data preprocessing, and load crawled items. Data preprocessing used for text that we think is less meaningful. The second function of class of item.py is to define items to crawl. Finally, the third class of pipeline will define how to output results to files and define detailed output form format. The whole architecture of Scrapy is shown in Figure 2.

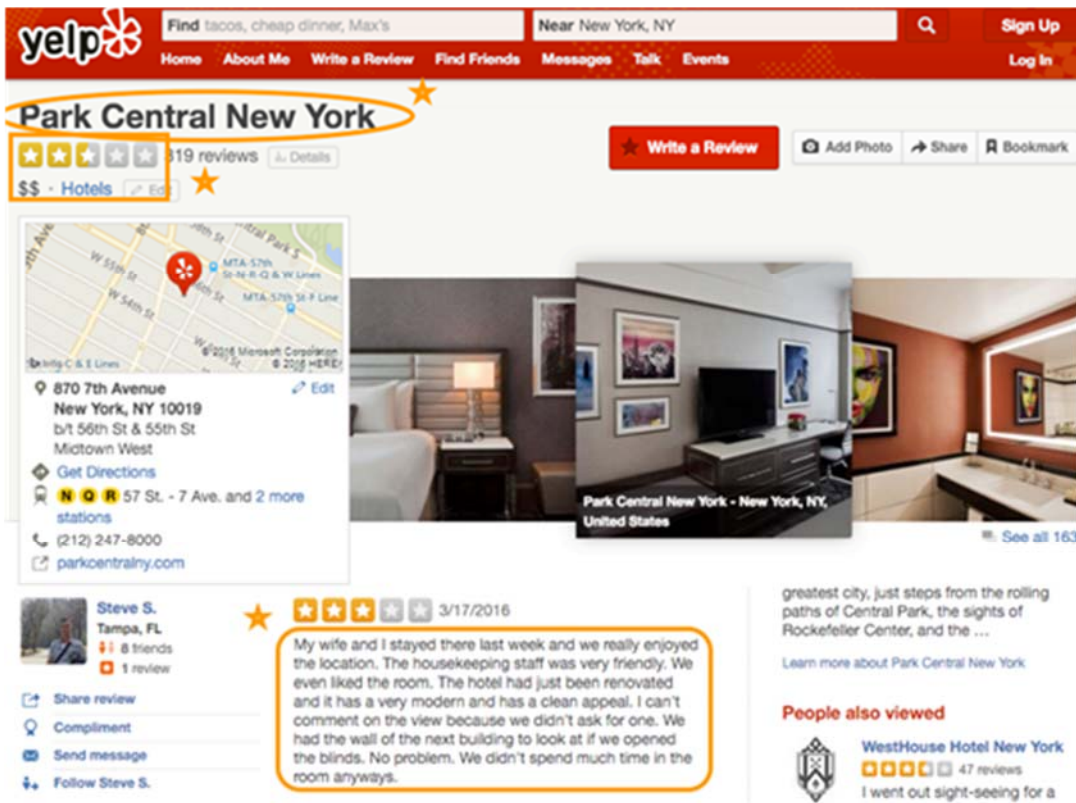


Figure 1 Web Pages of Hotel Taken from YELP.COM

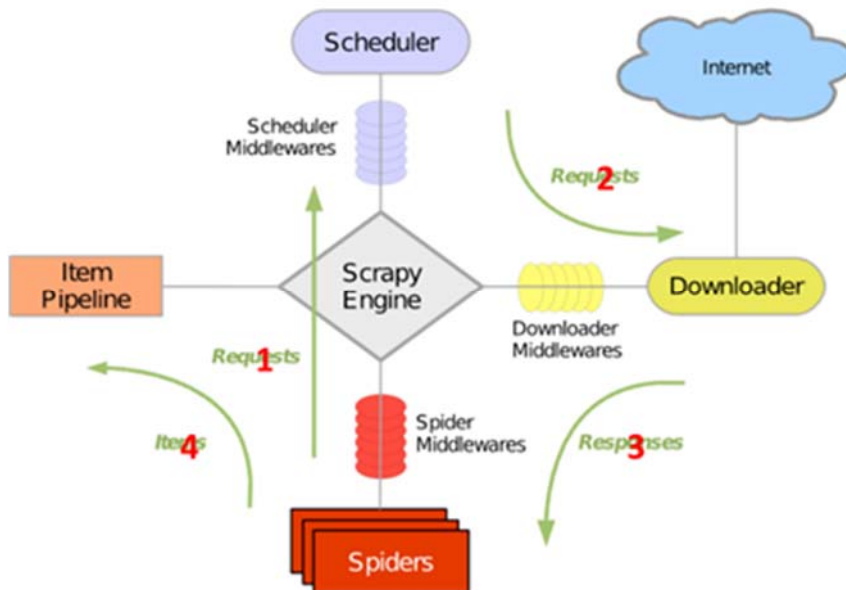


Figure 2 Scrapy Architecture

Scrapy is the main program that we sent the request to ask crawler start grabs the data. The spider will send the request to scheduler and then download the information from internet and send it back to spider. The process was divided into several stages, as depicted in Figure 3.

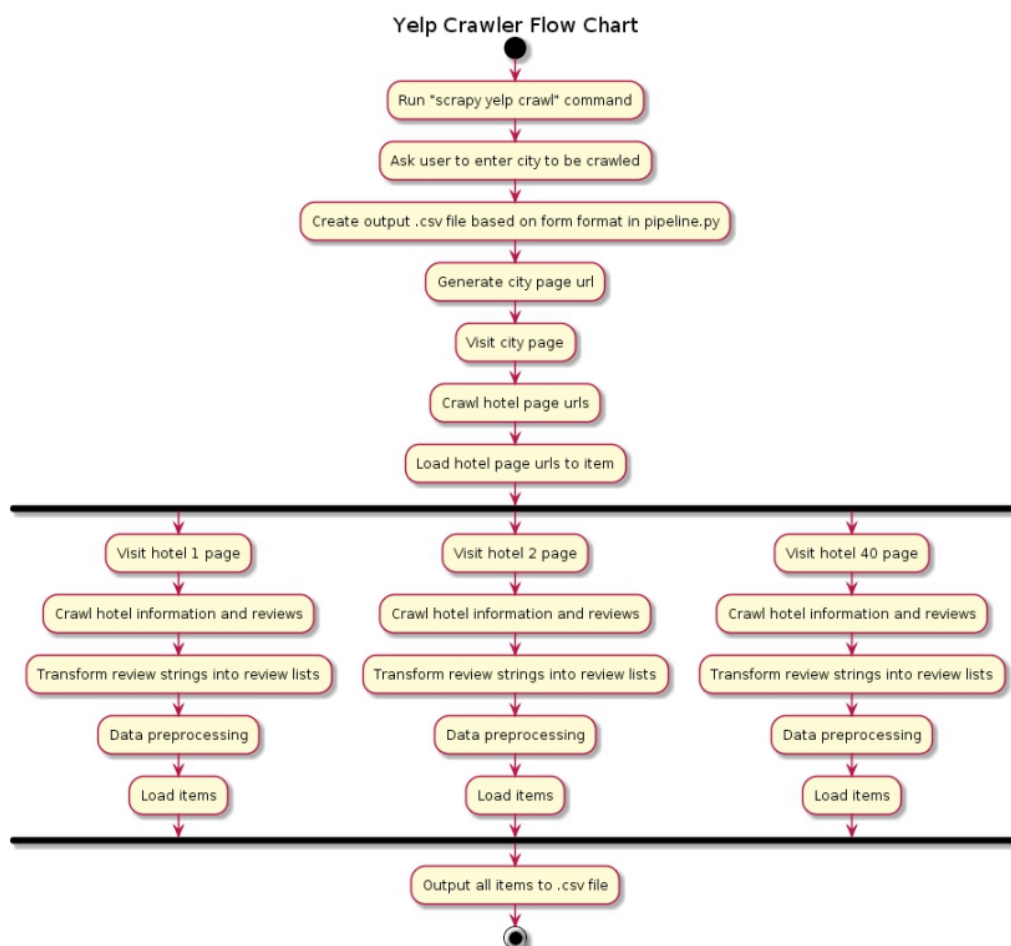


Figure 3 YELP Crawler Flow Chart

The crawler visited Yelp.com and extracted the URLs of hotel page of each city. It then grabbed information about the hotel from the page, including star rating, hotel name, price range, according the hotel page link in the default Excel file. Next, the written program saved each city's information and customer reviews in each city's Excel file, producing 43 files, including New York, Boston, Chicago, and Austin. Figure 4 shown the New York file which contains the name of the hotels, ratings, and customer reviews.

0	Hotel_name	Hotel_rating	...	Reviews1	Reviews2	..	Review100
1	Row NTC	3.5	...	I was so happy to stay here ...	I rate them 3 stars because our room was filthy	..	
2	The POD 51	4	:	I'm obsessed with Kimpton hotels, and ...	302	..	
3	:	:	:	:	:	..	:

Figure 4 New York File

The 43 cities shown in Table 1 were chosen in United States according to the popular city listed on the wiki website and 25 most popular cities to visit. The crawler collected data on 1,716 hotels, resulting in 171,600 customer reviews, meaning that each hotel had an average of 100 reviews.

Table 1 List of 43 Cities

Atlanta	Dallas	Las Vegas	Oklahoma	San Francisco
Austin	Denver	Los Angeles	Orlando	San Jose
Baltimore	Detroit	Louisville	Palm Springs	Savannah
Boston	El Paso	Memphis	Philadelphia	Seattle
Branson	Fort Worth	Miami	Phoenix	Sedona
Charlotte	Honolulu	Napa	Portland	St. Louis
Charlotte	Houston	Nashville	Saint Augustine	Washington
Chicago	Indianapolis	New Orleans	San Antonio	
Columbus	Jacksonville	New York	San Diego	

Once the data of the 43 cities were collected, we used our program to separate each sentence of the guest reviews into a series of words and then saved each word in Excel to prepare for frequency calculation.

Data Analysis and Pre-processing

Customer reviews are unstructured information. Therefore, we have to use text mining to extract useful, meaningful, and nontrivial information from unstructured texts to (Netzer, Feldman, Goldenberg, & Fresko, 2012). Normally, data analysis consists of data collection, pre-processing, domain identification, classification, and statistical association analysis (Fan, Wallace, Rich, & Zhang, 2006). Data pre-processing and domain identification establish content validity that extracts meaningful words from the corpus (Krippendorff, 2012). Next, the analysis examines the associations between the dependent variable and the identified domain-related words, which in this study is hotel satisfaction. Typical data pre-processing contains a series of stages such as stemming, reducing words to their roots so that grammatical forms or declinations of verbs are identified and counted as the same meaning (Raychaudhuri, Schütze, & Altman, 2002), misspelling identification, and identification and removal of stop words (i.e., definite and indefinite articles). The purpose of domain identification is to classify guest and non-guest experience-related words, which in this study is hospitality-related words. However, since all the words come from hotel guest reviews, there is no need to process the domain identification.

Data pre-processing consists of several stages. First, considering the sheer size of the word bank, our coding design excludes the words that are not meaningful, and that are misspelled. For example, there were many variations for a word like “hotel” so our coding design handled this first. Most words that reflected those aspects of guest experience were included (Xiang et al., 2015): (1) stop words such as verbs and conjunctions; (2) adverbs and adjectives such as “really,” “next,” “back,” and “still;” and, (3) words with low frequencies such as hotel brands or location. As part of the coding process, all possible variations of a word (e.g., plurals and misspellings) were manually searched and identified. When all of the sentences are divided, the program will calculate the frequency of words. This coding process was done iteratively from words with high frequencies to those with low frequencies. Our spreadsheet saved data from words with high frequencies to those with low frequencies and the top 500 frequency words were chosen to process statistic analytics. A frequency table with the hotel information that includes name, satisfaction ratings, price range, as well as guest comments for each hotel was generated. This final word frequency table was then imported into SPSS for statistical association analysis.

Statistical Analysis

In this part, we choose the hotel that is out of two standard deviations that rank in different variable, including CP, rating, price. Our CP value is defined as rating / price. However, our CP and rating data is non-normalize distribute data. Only the group of price data is normal distributed data (Figure 5 and Table 2). Therefore, we apply a transform of the data. The new, transformed data is normally distributed. The frequency in the Figure 5 presents the number of hotels.

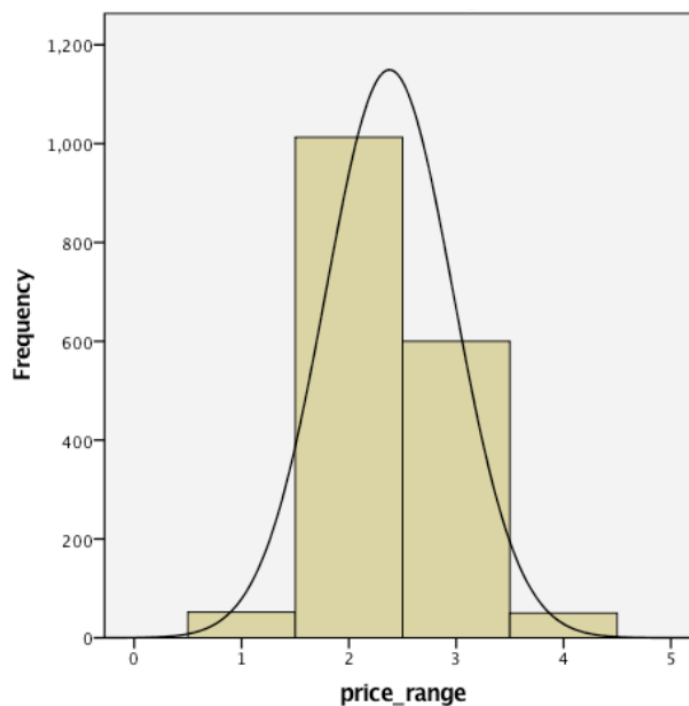


Figure 5 Distribution of Price Range

As can be seen in Figure 6 and Table 3, we transform data by using log nature and the kurtosis value became lower. Hence the distribution of CP approach normal distribution. Figure 7 and Table 4 show the distribution of the group of rating data that we transform by squaring the data. So the kurtosis value becomes lower and the distribution of rating approaches normal distribution.

Table 2 SPSS Data of Price Range

		Price_Range
N	Valid	1715
	Missing	4
Mean		2.38
Std. Deviation		0.595
Skewness		0.463
Std. Error of Skewness		0.059
Kurtosis		0.000
Std. Error of Kurtosis		0.118
Percentiles	25	2
	50	2
	75	3

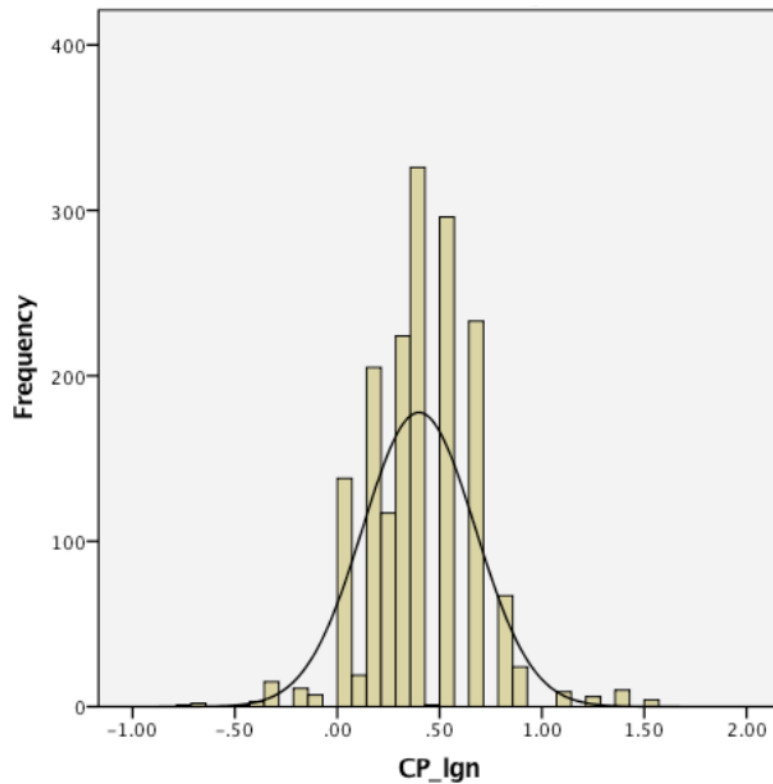


Figure 6 Distribution of CP Range

Table 3 SPSS Data of CP

	CP-E	CP_lgn
N Valid	1719	1719
Missing	0	0
Mean	1.55017	0.3992
Std. Deviation	0.465276	0.27537
Skewness	1.889	0.257
Std. Error of Skewness	0.059	0.059
Kurtosis	7.845	1.318
Std. Error of Kurtosis	0.118	0.118
Percentiles 25	1.25	0.2231
50	1.5	0.4055
75	1.75	0.5596

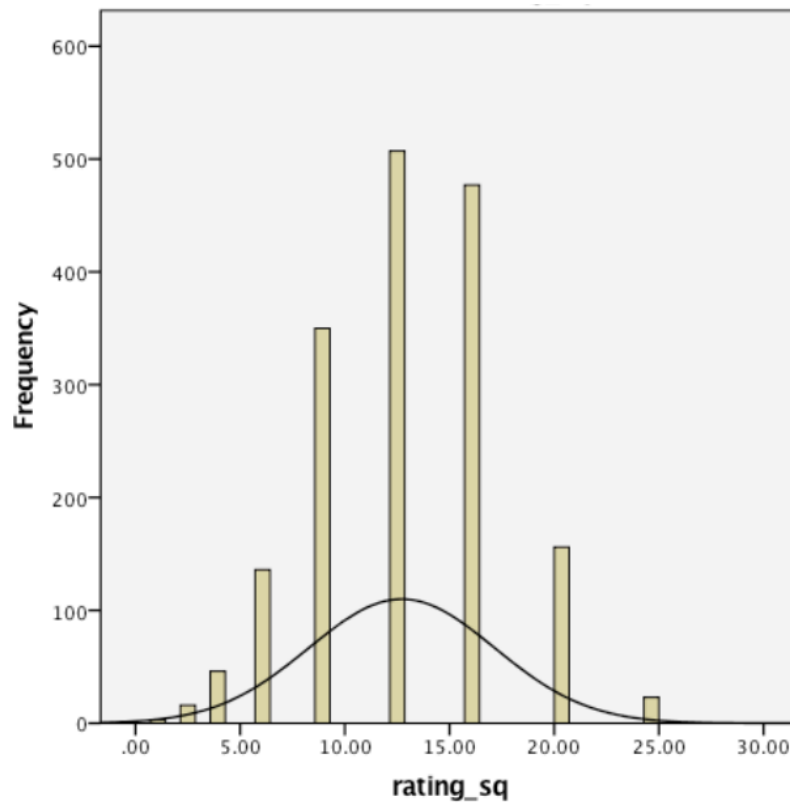


Figure 7 Distribution of Rating Range

Table 4 SPSS Data of Rating

	Hotel_Rating	Rating_sq
N Valid	1714	1714
Missing	5	5
Mean	3.506	12.7186
Std. Deviation	0.6558	4.4392
Skewness	-0.474	0.117
Std. Error of Skewness	0.059	0.059
Kurtosis	0.387	-0.139
Std. Error of Kurtosis	0.118	0.118
Percentiles 25	3	9
50	3.5	12.25
75	4	16

After we use standard deviation to determine the number of hotels in each group according to price, rating, and CP, we run our program to calculate the most frequent words about the hotels that are one standard deviation of the top and bottom according to different ranking of items. We generated six files, including high and low price of hotels, expensive and inexpensive hotels, and low CP and high CP hotels to apply our final analysis. The information is shown in Table 5.

Table 5 Number of Each Group

	High	Low
CP	52	38
Price	50	52
Quality	24	19

Data Exploration

Thinking of different needs of customers, we compare the words in high part and low part of the file under the same variables. So we compared words between each group, which consists with two files under same ranking items. The high price file and low price file are in the same group. The expensive hotel file and inexpensive file are in the same group. The high CP hotel file and low CP file are also in the same group. This gives us a total of three groups. The goal is to find the different words in each group and produces different sets of words. We separated each group into two-word sets. As can be seen in Figure 8, we define Set (E-S) as

expensive hotel, Set (IE-S) as inexpensive hotel and set (SP) as the basic elements in hotels that are associated with price. For quality group, we define Set (HQ-SQ) as high quality hotel, Set (LQ-SQ) as low quality hotel and set (SQ) as the basic elements in hotels that are associated with quality shown in Figure 9. As can be seen in Figure 10, we define Set (HCP-SCP) as high CP hotel, Set (HCP-SCP) as low CP hotel and set (SCP) as the basic elements in hotels that are associated with CP. The results are presented in the next section to see what words can help customers find the right hotel.

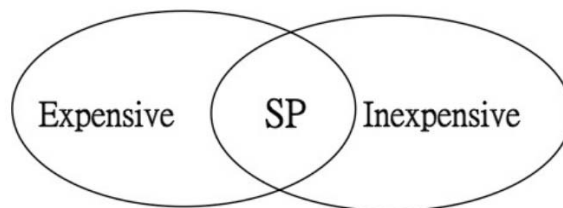


Figure 8 Words Set of Price Hotel

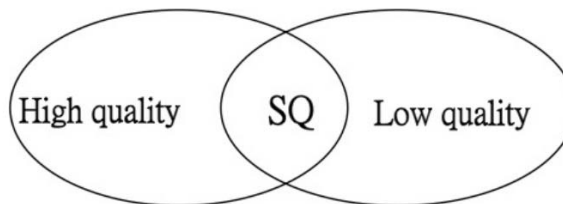


Figure 9 Words Set of Rating

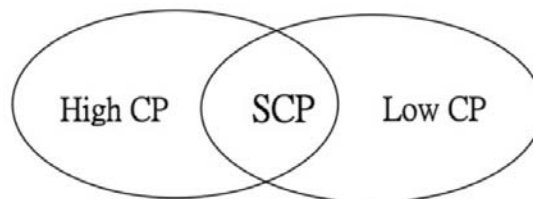


Figure 10 Words Set of CP Group

RESULTS

The first part of this section describes the words that pertain to hotel price, quality and CP (Table 6). The second part focuses on the different groups that responded to the second research question. The number of each word in Table 6 represents its frequency in the CP group. Words shown in Table 7 are found in all three groups. There are 26 words of the total 300 words in Table 8. This was

achieved in an exploratory fashion from the “angles” that many of these words also had relatively low total frequencies. In fact, more than 80% of these words had occurred in less than 2% of all customer reviews. It was clearly a problem for our analysis when one of the assumptions is to find the most useful words. As such, words with very low frequencies were removed. These final words sets were optimized by adjusting the type of word to maximize the explanatory power on searching keywords. As a result, the dataset was reduced to around 20 guest experience-related words.

Table 6 Words in Common (36 WORDS)

(‘room’, 2019)	(‘bed’, 466)	(‘people’, 265)
(‘place’, 1102)	(‘staff’, 449)	(‘staying’, 263)
(‘hotel’, 1066)	(‘pool’, 420)	(‘everything’, 263)
(‘great’, 914)	(‘first’, 401)	(‘bathroom’, 262)
(‘breakfast’, 796)	(‘day’, 398)	(‘check’, 262)
(‘clean’, 716)	(‘friendly’, 396)	(‘service’, 255)
(‘nice’, 614)	(‘location’, 378)	(‘around’, 242)
(‘night’, 605)	(‘area’, 365)	(‘way’, 238)
(‘like’, 558)	(‘free’, 342)	(‘never’, 203)
(‘back’, 528)	(‘little’, 307)	(‘see’, 195)
(‘time’, 522)	(‘experience’, 281)	(‘ever’, 169)
(‘good’, 483)	(‘next’, 278)	(‘water’, 168)

The second part provides the content of the three word sets: price, quality, and CP group. Table 7 presents 25 words that are related to cost; customers use these words to describe their stay in high-priced hotels. Table 8 lists 21 words that are related to inexpensive hotels.

Table 9 shows 35 words that are related to high-quality hotel. Table 10 shows 27 words that are related with low quality hotels. Tables 11 and 12 list the words that are related to CP factor of hotels.

Table 7 Words Represent Expensive Hotels (25 words)

('food', 1453)	('seasons', 681)	('wonderful', 490)
('bar', 1267)	('love', 676)	('dinner', 489)
('beautiful', 1161)	('valet', 622)	('suite', 488)
('best', 1031)	('star', 613)	('new', 488)
('view', 1019)	('drinks', 592)	('worth', 483)
('beach', 892)	('tea', 543)	('walk', 482)
('restaurant', 865)	('wedding', 516)	('excellent', 481)
('four', 842)	('perfect', 511)	('property', 469)
('spa', 757)		

Table 8 Words Represent Inexpensive Hotels (21 words)

('price', 354)	('cheap', 208)	('sleep', 155)
('parking', 258)	('close', 193)	('dirty', 154)
('downtown', 256)	('beds', 188)	('reviews', 153)
('old', 238)	('morning', 179)	('money', 150)
('motel', 227)	('need', 178)	('helpful', 149)
('bad', 223)	('street', 168)	('super', 148)
('coffee', 213)	('work', 163)	('minutes', 148)

Table 9 Words Represent High Quality Hotels (35 words)

('inn', 462)	('weekend', 202)	('fresh', 149)
('wine', 370)	('love', 201)	('private', 146)
('amazing', 329)	('modern', 196)	('trip', 145)
('perfect', 304)	('coffee', 190)	('lovely', 142)
('beautiful', 301)	('recommend', 185)	('small', 141)
('comfortable', 289)	('tub', 177)	('enjoyed', 141)
('wonderful', 277)	('husband', 175)	('relaxing', 136)
('best', 265)	('restaurants', 175)	('things', 133)
('house', 258)	('guests', 174)	('complimentary', 132)
('definitely', 248)	('happy', 171)	('super', 131)
('food', 236)	('quiet', 162)	('owners', 128)
('home', 233)	('helpful', 158)	
('delicious', 205)	('downtown', 150)	

Table 10 Words Represent Low Quality Hotels (27 words)

('desk', 211)	('airport', 88)	('lobby', 67)
('front', 188)	('bad', 87)	('pay', 66)
('dirty', 129)	('worst', 83)	('minutes', 66)
('door', 116)	('manager', 76)	('cheap', 64)
('work', 111)	('price', 76)	('close', 64)
('old', 104)	('smell', 73)	('phone', 61)
('shower', 98)	('parking', 71)	('horrible', 61)
('floor', 97)	('sleep', 69)	('towels', 60)
('shuttle', 89)	('park', 68)	('smoking', 60)

Table 11 Words Represent High CP Hotels (26 WORDS)

('inn', 391)	('helpful', 207)	('guests', 177)
('morning', 276)	('feel', 199)	('happy', 176)
('downtown', 258)	('close', 198)	('beautiful', 173)
('coffee', 255)	('weekend', 195)	('love', 171)
('house', 249)	('super', 194)	('motel', 167)
('perfect', 239)	('tub', 192)	('husband', 166)
('amazing', 224)	('wonderful', 191)	('restaurants', 165)
('recommend', 219)	('quiet', 188)	('trip', 163)
('home', 215)	('wine', 185)	

Table 12 Words Represent Low CP Hotels (17 WORDS)

('bar', 517)	('small', 299)
('lobby', 490)	('bad', 298)
('floor', 371)	('manager', 275)
('work', 348)	('walk', 259)
('view', 339)	('minutes', 257)
('hotels', 336)	('dirty', 255)
('valet', 321)	('star', 251)
('better', 314)	('think', 250)
('restaurant', 304)	

CONCLUSION

Finding

Table 6 lists 36 words that were associated with the basic requirements of customers in terms of quality, price and CP and that answer the first research question. These words reflect a cross-section of the hotel guest experience: (1) the core product such as “room,” “bed,” and “bathroom”; (2) hotel amenities such as “pool,” “parking,” “lobby,” “shower,” “TV,” “bar,” and “amenities,” (3) hotel attributes such as “location,” “area,” “service,” “quiet”; (4) staff-related descriptors such as “staff,” “friendly,” and “nice”; and (5) evaluation of experience such as “clean,” “good.” Most of these words are functional and objective, while a handful were subjective evaluations. Overall these 36 words reflect a diverse array of amenities, attributes, and service encounters shaped by hotel guests’ unique expectations and evaluations in quality, price, and CP.

The words shown in Tables 7 to 12 are in response to the first research question. The words in different groups reflect different situations and types of hotels. To answer the second research question, those words can be used as keywords to help customers search for a suitable hotel. For example, for people who care more about price than quality, they can use words that pertain to cost. People who are price-conscious but want to find a high-quality hotel, can use the words that represent high CP. What’s more, it is interesting to note that some words listed in different groups are not usually thought of. Such as the word “better” represents the low CP hotel, which people usually think of as representing the positive meaning. Other findings, such as “wedding” that represents the expensive hotel, also provide interesting findings, showing people have a special goal of finding an expensive hotel for wedding. The word such as “husband” also tells us that people who book a hotel with their husband will tend to find a high CP hotel. Therefore, our findings show the detail of the words that help us to distinguish what types of hotel these words represent.

The words also present pictures that depict a hotel that customers think. For example, the words of the high CP group tell us that customers think a high CP hotel will make people feel at home. Also, people tend to visit these hotels on weekends. The hotels need to be close to downtown. Those words can tell managers how to run a hotel that customers will like.

Implications and Contributions

The results of the analysis reveal patterns reflective of consumers’ evaluation of their experiences with hotels. We make sense of and attach meaning to the

inferences by applying theories to explain patterns revealed by the data. Unlike traditional methods, our way of explaining the findings is part of the epistemology of generating and creating knowledge using big data (George et al., 2014). Although there is no previous study to benchmark against, the validity of our study, like many others based upon big data, was established by the precisely devised analytical process that followed both theory and common practices in text mining. For example, we use content analysis and the definition of hotel product and guest experience to make the analysis more actually and veraciously.

Although big data analytics has been considered as a latest research exemplification in many disciplines, we have seen very few applications in the field of hospitality that fully explore its capabilities. This study uses text analytics to classify online customer reviews, assess their quality, and to identify different kinds of text for different types of hotels. The uniqueness of this study lies in the use of big data and drivers of guest experience on a scale that was not available in conventional guest survey studies. This study is an abecedarian effort in big data analytics; however, we have gained substantial insights into some of the extensively studied constructs in hospitality. We hope that it sets an example for the development of business analytics in hospitality marketing and management. Showing important keywords in customer reviews can give managers a reference to buy Google ad words.

Limitations and Future Study Directions

Although this research has some managerial implications and contributions for academia and the hospitality industry, the findings still need verification to ensure that our findings meet our assumption that people use those words as key words can help customers find the hotels that they want. Our findings can be extended to explore more useful meaning words for hospitality or tourism industry that give us a reference in buying Google ad words.

REFERENCES

- Ang, L.-M., & Seng, K. P. (2016). Big Sensor Data Applications in Urban Environments. *Big Data Research*, 4, 1–12. <https://doi.org/10.1016/j.bdr.2015.12.003>
- Bailey, A. A. (2005). Consumer awareness and use of product review websites. *Journal of Interactive Advertising*, 6(1), 68-81. <http://dx.doi.org/10.1080/15252019.2005.10722109>
- Bitner, M. J., & Hubbert, A. R. (1994). Encounter satisfaction versus overall

- satisfaction versus quality. *Service quality: New directions in theory and practice*, 34, 72-94. <http://dx.doi.org/10.4135/9781452229102.n3>
- Boase, J., Horrigan, J.B., Wellman, B. & Raine, L. (2006). The internet and email aid users in maintaining their social networks and provide pathways to help when people face big decisions. *Pew Internet and American Life Project*. Retrieved July 26, 2011 (<http://pewinternet.org/Reports/2006/The-Strength-of-InternetTies.aspx>).
- 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. <http://dx.doi.org/10.1080/10548408.2013.750971>
- Chau, M., & Xu, J. (2012). Business intelligence in blogs: Understanding consumer interactions and communities. *MIS quarterly*, 36(4), 1189-1216.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, 36(4), 1165-1188.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354. <http://dx.doi.org/10.1509/jmkr.43.3.345>
- Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4), 23-45. <http://dx.doi.org/10.1002/dir.20087>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82. <http://doi.org/10.1145/1151030.1151032>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- Finch, B. J. (1999). Internet discussions as a source for consumer product customer involvement and quality information: an exploratory study. *Journal of Operations Management*, 17(5), 535-556. [https://doi.org/10.1016/S0272-6963\(99\)00005-4](https://doi.org/10.1016/S0272-6963(99)00005-4)
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57(2), 321-326. <http://doi.org/10.5465/amj.2014.4002>

- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498-1512. <http://doi.org/10.1109/TKDE.2010.188>
- Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31(3), 493-520. <http://dx.doi.org/10.1287/mksc.1110.0700>
- 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. <http://doi.org/10.1038/nature07634>
- Gretzel, U. (2006). Consumer generated content—trends and implications for branding. *E-review of Tourism Research*, 4(3), 9-11.
- He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. *British Journal of Educational Technology*, 42(1), 40-49. <http://doi.org/10.1111/j.1467-8535.2009.00980.x>
- He, W. (2013). Improving user experience with case-based reasoning systems using text mining and Web 2.0. *Expert Systems with Applications*, 40(2), 500-507. <https://doi.org/10.1016/j.eswa.2012.07.070>
- He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, 52(7), 801-812. <https://doi.org/10.1016/j.im.2015.04.006>
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472. <https://doi.org/10.1016/j.ijinfomgt.2013.01.001>
- Hung, J.-L., & Zhang, K. (2008). "Revealing Online Learning Behaviors and Activity Patterns and Making Predictions with Data Mining Techniques in Online Teaching". *MERLOT Journal of Online Learning and Teaching*, 4(4), 426-437.
- Ittoo, A., Nguyen, L. M., & van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. *Computers in Industry*. 78, 96-107. <https://doi.org/10.1016/j.compind.2015.12.001>
- Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*: Sage.
- Mayer-Schonberger, V., & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, And Think*. Houghton Mifflin Harcourt.

- Mehdipour, F., Noori, H., & Javadi, B. (2016). Energy-efficient big data analytics in datacenters. *Advances in Computers*, 100, 59-101. <http://doi.org/10.1016/bs.adcom.2015.10.002>
- Minnema, A., Bijmolt, T. H., Gensler, S., & Wiesel, T. (2016). To Keep or Not to Keep: Effects of Online Customer Reviews on Product Returns. *Journal of Retailing*, 92(3), 253-267. <https://doi.org/10.1016/j.jretai.2016.03.001>
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon. com. *MIS quarterly*, 34(1), 185-200.
- Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16), 7653-7670. <https://doi.org/10.1016/j.eswa.2014.06.009>
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543. <http://dx.doi.org/10.1287/mksc.1120.0713>
- Park, D.-H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125-148.
- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46. <https://doi.org/10.1016/j.knosys.2015.06.015>
- Raychaudhuri, S., Schütze, H., & Altman, R. B. (2002). Using text analysis to identify functionally coherent gene groups. *Genome Research*, 12(10), 1582-1590. <http://doi.org/10.1101/gr.116402>
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- Sax, D. (2011). Yelp's online reviewing mafia. *Bloomberg Businessweek* (June 2). Retrieved from http://www.businessweek.com/magazine/content/11_24/b4232083260194.htm
- Tetlock, P. C., SAAR-TSECHANSKY, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437-1467. <http://doi.org/10.1111/j.1540-6261.2008.01362.x>
- Vinodhini, G., & Chandrasekaran, R. (2016). A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University-Computer and Information Sciences*, 28(1), 2-12. <https://doi.org/10.1016/j.jksuci.2014.03.024>

- Wu, C. H.-J., & Liang, R.-D. (2009). Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants. *International Journal of Hospitality Management*, 28(4), 586-593. <https://doi.org/10.1016/j.ijhm.2009.03.008>
- 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. <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433-447. <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(1), 180-182. <https://doi.org/10.1016/j.ijhm.2008.06.011>