

TEXT MINING IN THE HOSPITALITY SECTOR TO EXTEND THE MOTIVATION THEORY

Taşkın DİRSEHAN, PhD

Marmara University,
Department of Business Administration (Lectured in English),
Bahçelievler/Istanbul, Turkey, 34180
taskin.dirsehan@marmara.edu.tr

Abstract

In order to provide customer revisit intention or positive word of mouth, hospitality managers should focus on satisfying customers by listening to them. In today's information age, customers can be heard via their comments on travel websites that provide "big data." The main purpose of this paper is to present a text-mining procedure to be used in marketing strategies to discover critical elements in the hospitality sector from customers' online reviews by extending Herzberg's Two-Factor Theory of Motivation. The result of the text-mining procedure provides a list of words including how many times they occurred in satisfaction and dissatisfaction reviews. Critical ones are selected, and after conducting a correlation analysis, they are grouped into three categories: (1) motivators: words correlated with total review scores only in the positive reviews; (2) hygiene factors: words negatively correlated with total review scores only in the negative reviews; and (3) effective motivators: words correlated with total review scores in both positive and negative reviews.

Keywords: Hospitality Management, Motivation Theory, Text Mining, Tourism Marketing, Big Data

1. Introduction

As companies' upstream activities such as sourcing, production and logistics become commoditized, companies should shift their strategy downstream—i.e., from products to customers—to gain competitive advantages. For this purpose, they should accumulate customer data (Dawar, 2013). The accumulation of data from various sources results in so-called Big Data, which creates a new challenge that does not build on traditional marketing skills. The question is whether marketers will be capable of managing big data or whether the customer experience will be shaped by software engineers (MSI, 2015). Nowadays, marketers need to use more technology than ever before such as CRM systems and big data analytics. In the past, IT function undertook the management of this technology, but that is changing rapidly (Joshi & Giménez, 2014). Businesses retain data-mining experts, data scientists, data-visualization experts, data analysts, etc. in their marketing departments to produce and use information that will gain competitive advantages in today's information era (Köktürk & Dirsehan, 2012).

“Developing Marketing Analytics for a Data-Rich Environment” is considered a tier 1 research priority by Marketing Science Institute (MSI) for 2014-2016. It reveals a better understanding of customers and improved marketing decision-making. In order to understand consumer behavior, marketing managers should know about their motivations, since they lead people to behave as they do (Solomon, 2009). Thus, this study tries to find out travelers' motivations about visiting hotels via text mining.

The role of marketing intelligence in marketing strategy is firstly discussed in this study, and then text mining approach is mentioned to create such a strategy. The last part of the literature survey evokes Herzberg's motivation theory to understand travelers' motivation to visit hotels. An application of text mining follows the literature review. Then, the results are discussed to provide insights regarding marketing academics and practitioners.

2. Theoretical Background

2.1. Text Mining Approach to Determine Customer Preferences

In today's dynamic environment, companies are faced with huge amounts of data coming from various sources such as Web traffic, e-mail messages, and social media content, as well as machine-generated data from sensors. In addition, these data may be unstructured or semi-structured, so they are not suitable for relational databases organizing data in the form of columns and rows. This kind of huge volume of datasets is called “big data,” which are beyond the ability of typical database management systems to capture, store, and analyze (Laudon & Laudon, 2014). Data-mining tools are used to collect, organize, analyze, and visualize the data, and they can be considered as a new research paradigm to make inferences about reality using huge volumes of data. Data-mining techniques are needed to reveal hidden information from such large datasets, to discover and understand patterns in customer behavior (Lee & Siau, 2001; Hoontrakul & Sahadev, 2008). While traditional methods rely on a set of predefined hypotheses, big data analytics aim to explore new patterns or predict future trends from big data for knowledge discovery (Xiang et al., 2015; Wang & Wang, 2008). Brachman et al. (1993) call this process “data archaeology,” in which the analyst can discover interesting knowledge through an iterative, exploratory process. In the case of social media and consumer-generated content as sources of data including large datasets and unstructured data, text analytics play an important role in big data analytics (Xiang et al., 2015).

Text mining can be defined as discovering unknown facts and hidden patterns existing in the lexical, semantic or even statistical relations of text collections (Stavrianou et al., 2007).

In the literature, it is apparent that text-mining applications in marketing have increased in the recent decade. They include:

- The use of text mining to analyze competitors' online promotional text messages (Leong et al., 2004);
- Content analysis of travel related websites (Choi et al., 2007);
- Application of text mining to forecast fashion trends (Rickman & Cosenza, 2007);
- Identifying evaluation criteria of customers' intention to revisit restaurants (Yan et al., 2013);
- Evaluating consumers' sentiments toward well-known brands from tweets (Mostafa, 2013);
- Impact of text product reviews on sales (Moon et al., 2014);
- Segmenting customers' opinions toward the low-cost airlines or low-cost carriers (Liau & Tan, 2014); and
- Identifying and analyzing brand associations in an online community (Camiciottoli et al., 2014).

Although there are an increasing number of studies on application of text mining in marketing, there is still a research gap regarding the assessment of consumers' preferences through text-mining procedures.

In order to gain a competitive advantage, hotel managers should focus on their customer relations to create loyal customers. This involves customer lifetime value (CLV) analysis. When a CLV analysis is conducted for a company, it's obvious that acquiring new customers is much more costly than retaining profitable ones, since acquiring new customers involves advertising, promotion, and start-up operating expenses (Reichheld, 1996). So, the competitive advantage is provided by customer loyalty, and to develop it for a hotel, firms must learn their customers' wishes and needs (Tepeci, 1999). Data-mining techniques can be proposed to hotel managers to bolster their customer retention strategy to understand their customers' preferences and ways to interact with them (Min et al., 2002). In order to fill the gap to reveal consumers' preferences through text-mining procedures, Xiang et al. (2015) applies a text-mining approach to consumer reviews extracted from Expedia.com. However, they are limited to U.S. hotels, and they do not separate the words in the satisfaction and dissatisfaction reviews. So, it was not possible to find out the dual effect of the words. This study, on the contrary, investigates whether some words may become hygiene factors and motivators at the same time.

2.2. Online Booking Websites

In the United States and Europe, it is estimated that more than 50% of the travel reservations are made online, and online booking websites have gained importance (Hoontrakul & Sahadev, 2008). Moreover, online reviews are sources of information for consumers in making purchase decisions. Thus, marketing managers should understand the influential and predictive effects of online reviews (Tsang & Prendergast, 2009). Hotel industry provides a warehouse of customer comments, and precious knowledge can be extracted as a result of mining them (Dirsehan, 2015). This is an opportunity for hotel managers to understand their customers if they are able to conduct text mining. In this study, Booking.com is used for its rich information as shown in Figure 1.

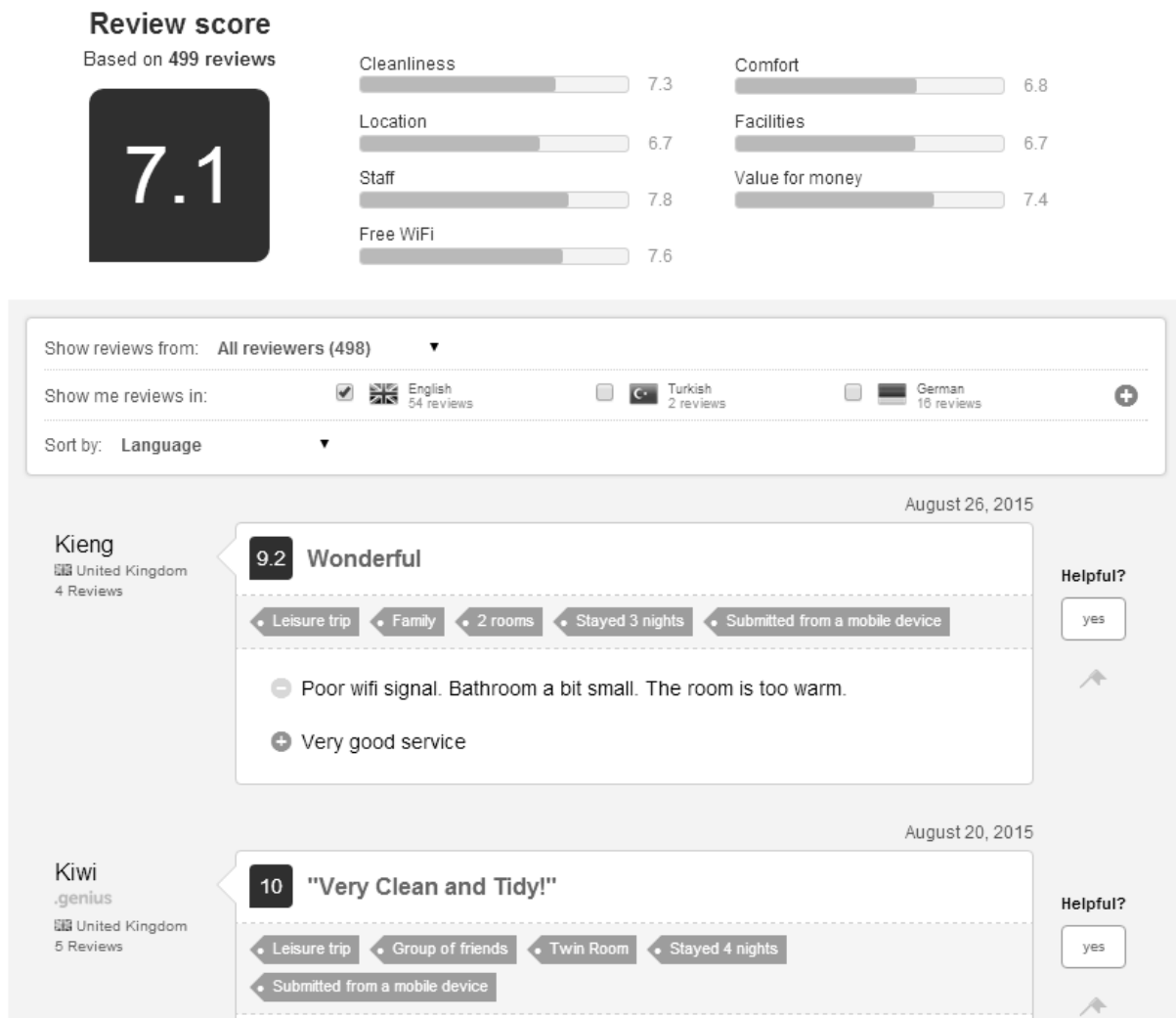


Figure 1. Screenshot of customer reviews on Booking.com

2.3. Understanding Consumer Motivation: Herzberg's Two-Factor Theory

Herzberg's two-factor theory identifies two categories of motivational forces (Herzberg et al., 1962; Herzberg, 1975):

- (1) Motivator/Satisfiers (when motivators are present, employees feel satisfied); and
- (2) Hygiene/Dissatisfiers (when hygiene factors are lacking, employees experience dissatisfaction, but their existence does not necessarily experience satisfaction).

Tuten and August (1998) extend the model to consumer services and state that service hygiene factors are the core service experience—including price, policies surrounding purchases and rebates on the service, availability of service representatives to answer questions, tangible environmental conditions (cleanliness, etc.), and opportunities for social interaction. On the other hand, service motivator factors include opportunities to feel appreciated for purchasing or using the service (Tuten & August, 1998).

Chan and Baum (2007) apply the theory in the tourism and hospitality sector. They conduct an interview with 29 guests in Malaysia, and they group the findings according to Herzberg's two factors. Satisfiers are considered as related to personal experiences from the natural environment and attractions, physical sites and leisure activities. On the other hand,

dissatisfiers are considered to be constructs related to the performance and availability of facilities, amenities, and maintenance (Chan & Baum, 2007). Similarly, in the quantitative research of Gu and Ryan (2008), Chinese guests expect a hotel to be comfortable and clean, but they do not generate high levels of satisfaction because they are considered as a core hotel service and can be interpreted as hygiene factors.

3. Methodology

The text-mining approach is applied to reveal key information from travelers' comments. Then, the revealed characteristics are analyzed in terms of their relationships with travelers' review scores. In this process, the following data-mining steps are performed:

3.1. Collecting Raw Data from Data Source

Booking.com is used as a data source. It is a website including millions of hotels all around the world, and it allows online booking. On the other hand, its most precious content is the travelers' reviews that are made after individuals leave the hotel. These reviews represent unstructured data, and they provide a "mine" for hotel marketing managers. If they are able to mine the information, hotel market managers obtain valuable knowledge in the competitive environment.

Data that can be extracted from booking.com include:

- (1) Traveler's name (if provided)
- (2) Traveler's nationality
- (3) Travel type
- (4) Traveler's gender (if provided)
- (5) Traveler's age (if provided)
- (6) Comment dates
- (7) Traveler's review scores
- (8) Travelers' positive comments
- (9) Traveler's negative comments
- (10) Names of the hotels to be reviewed
- (11) Hotel stars
- (12) Hotel's total review scores

3.2. Data Cleaning

After collecting the mentioned data, they are cleaned based on several criteria. For instance, the comments written in another language are translated into English by using Portals' translator services. Here, the main purpose is to reveal the words, rather than grammatically correct sentences.

3.3. Building Data Warehouse and Data Mart

For this study, a data warehouse is built after organizing the mentioned data from booking.com. Destinations are selected according to "Global Destination Cities Index." So, the numbers of reviews from the destinations of the hotels are as follows:

- 399 reviews from Amsterdam
- 296 reviews from Bangkok

- 100 reviews from Barcelona
- 153 reviews from Berlin
- 301 reviews from Dubai
- 200 reviews from Kuala Lumpur
- 397 reviews from London
- 196 reviews from Los Angeles
- 199 reviews from Madrid
- 100 reviews from Miami
- 399 reviews from Munich
- 401 reviews from New York
- 385 reviews from Paris
- 100 reviews from Seoul
- 399 reviews from Singapore

So, a total of 4025 positive and 4025 negative comments from 15 destinations are included in the data warehouse. Three points are considered while choosing the reviews:

- (1) Travelers who wrote both positive and negative comments are included in order to reveal critical words that occurred in both types of comments;
- (2) A maximum of 30 comments for the same hotel is considered in the dataset in order not to be directed;
- (3) Comments from the most- and least-reviewed hotels are included in order to increase the variance (the range of hotels' total review scores is between 4.30 and 9.7; the median is 8.1).

Then, a data mart is prepared including traveler ID, destination, traveler's review score, positive comments and negative comments.

3.4. Using Text Mining as Analytical Tool

A text-mining tool is used to apply text mining to the prepared data mart. The issues or decisions that should be considered during the text-mining process can be summarized thusly (Stavrianou et al., 2007):

- Stop list (decision to take into account stop words);
- Stemming (decision to reduce the words to their stems);
- Noisy data (clarity of text from noisy data);
- Word sense disambiguation (decision to clarify the meaning of words in text);
- Tagging (considering data annotation and/or part of speech characteristics);
- Collocations (considering compound or technical terms);
- Grammar/Syntax (Decision to make a syntactic or grammatical analysis);
- Tokenization (Considering tokenization of words or phrases);
- Text representation (Determining important terms, words or phrases, nouns or adjectives, word order, context, and background knowledge); and
- Automated learning (Decision to use categorization, application of similarity measures).

In this study, satisfaction stories and dissatisfaction stories are analyzed separately using text-mining software. The first step is to list all of the words that occurred in satisfaction and dissatisfaction stories. Stop words—which are high frequency words such as “a,” “the” or “of”—are not excluded from the analysis. They will be ignored at the end of the process. Noisy data are eliminated in the data-cleaning step as explained before (by translating them).

However, miswritten words are not excluded from the texts. Their occurrences are detected and added at the end of the process.

In the second step, an operator is placed to list combined words (such as “internet connection” in addition to “internet” and “connection” separately). The reason is that the meaning of the whole is greater than the sum of its parts (Stavrianou et al., 2007). Grammar correction is not needed in this study since the words, rather than phrases, are analyzed.

Thirdly, a tokenization operator is placed to split the texts into units, which are words in this case. Then, an operator is used to count the word stems (such as “clean-“). This way, all the affixed words are counted at a time (such as cleaned, cleaning, cleaner, cleanliness, etc.) At the end of this procedure, thousands of words are listed with their occurrences in the satisfaction and dissatisfaction reviews. So, it was decided to continue with the 45 words that occurred most often, as listed in the Appendix 1.

3.5. Data Improvement for Further Investigation

After exploring the critical words, the following question arises: “What is the relationship between the existence of these words and travelers’ review scores?” To answer this question, new columns are added to the data list (as word occurrences). So, each column is coded with binary variables 0 or 1 (0 representing non-existence of the word and 1 representing the existence of the word in the review). The same procedure is repeated for the negative comments too. Then, a correlation analysis is conducted between travelers’ review score, word occurrences in the positive reviews and word occurrences in the negative reviews. The research finding is summarized in Appendix 2 and Appendix 3.

4. Conclusion

4.1. Text Representation to Extend Herzberg’s Two-Factor Motivation Theory

According to the results, the words can be divided into three classes:

- (1) Words that occurred only in the positive reviews;
- (2) Words that occurred only in the negative reviews; and
- (3) Words that occurred in both positive and negative comments.

Classes (1) and (2) can be applicable to Herzberg’s Theory of Needs. The first class includes the words that are important to create traveler satisfaction such as view, décor, breakfast, etc. It corresponds to the motivators in Herzberg’s theory. In the second category, there are words indicating travelers’ disappointment. This class represents the hygiene factor of the theory. So, some critical features such as smell, reception, and toilet are subject to customer dissatisfaction. However, there are some critical characteristics that show both classes. So, there are words that occur in both satisfaction and dissatisfaction reviews such as “staff”, “help-”, “comfort-”, “friendl-”, etc. So, this class may be called “effective motivators” since they can eliminate the dissatisfaction of the customer and create satisfaction at the same time. So, the opposite situation may be called neutral effect (neither satisfaction nor dissatisfaction). This explanation is summarized in Table 1. The corresponding words in this study are shown in the Table 2.

Table 1. Proposed Extension of Theory of Motivation

		Critical Words		Possible Outcomes
		If They Occurred	If They Did Not Occur	
Herzberg's Two-Factor Theory of Motivation	Hygiene Factors	0 (not necessarily satisfaction)	- (creating dissatisfaction)	Eliminating dissatisfaction
	Motivators	+	0 (not necessarily dissatisfaction)	Creating satisfaction
Proposed Extension	Effective Motivators	+	- (creating dissatisfaction)	Eliminating dissatisfaction and creating satisfaction at the same time
	Neutral Effects	0 (neither satisfaction nor dissatisfaction)	0 (neither satisfaction nor dissatisfaction)	No outcome

Table 2. Summary of the Critical Words according to the Proposed Theory of Needs

	Critical Words	Average Correlation Coefficient with Travelers' Review Score
Hygiene Factors	smell, book-, nois-, reception, toilet, shower, TV, polite, luggage, sleep, pay-, wifi	-0.0993
Motivators	view, décor, breakfast, pool, food, design, bus, tea	0.0537
Effective Motivators	staff, help, comfort-, friendl-, clean-, service-, facilit-, bar, bed, bathroom, money, towel, window	0.1283

4.2. Academic and Practical Implications

The average correlation coefficient with travelers' review score is also indicated in the last table. Accordingly, hygiene factors have an average correlation of -0.0993, meaning that if dissatisfaction factors are not eliminated, this may be accompanied with a decrease in the review score. On the other hand, if motivators are placed, customers' comments may be accompanied by an increase of 0.0537. In terms of effective motivators, their existence accompanies an increase of 0.1283 in the customer reviews.

Moreover, these words can be grouped according to their functions. For instance, “toilet”, “shower”, “bathroom”, and “towel” can be grouped as “bath components.” As these characteristics are considered to be “standard requirements” by travelers, the problems associated with them may cause dissatisfaction, but their existence is not a motivator. Another class may be “staff characteristics,” including “polite”, “help”, and “friendly”.

In summary, the four categories presented in this research may be applied by academicians for other sectors by using text-mining procedures explained in this study. The features in the categories can reduce travelers’ dissatisfaction and improve their satisfaction. This framework should be considered by hospitality marketing managers to enhance quality.

4.3. Limitations and Further Research

Even though this study provides some insights for extending Herzberg’s motivation theory into the hospitality sector, it has some limitations. As tourism destinations may affect tourists’ visiting intention, the revealed words may differ according to different locations. In addition, further studies may consider gender, travel types, and age to reveal the differences in terms of traveler motivation.

References

- Brachmann, R., Selfridge, P., Terveen, L., Altman, B., Borgida, A., Halper, F., et al. (1993). Integrated Support for Data Archaeology. *International Journal of Cooperative Information Systems* , 2 (2), 159-185.
- Camiciottoli, B., Ranfagni, S., & Guercini, S. (2014). Exploring brand associations: an innovative methodological approach. *European Journal of Marketing* , 48 (5/6), 1092-1112.
- Chan, J., & Baum, T. (2007). Researching Consumer Satisfaction: An Extension of Herzberg’s Motivator and Hygiene Factor Theory. *Journal of Travel & Tourism Marketing* , 23 (1), 71-83.
- Choi, S., Lehto, X., & Morrison, A. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management* , 28, 118-129.
- Dawar, N. (2013). When marketing is strategy. *Harvard Business Review* , 91 (12), 101-108.
- Dirsehan, T. (2015). An Application of Text Mining to Capture and Analyze eWOM: A Pilot Study on Tourism Sector In: S. Rathore, & A. Panwar (2015), *Capturing, Analyzing, and Managing Word-of-Mouth in the Digital Marketplace* (pp. 168-186). USA: IGI Global.
- Gu, H., & Ryan, C. (2008). Chinese clientele at Chinese hotels—Preferences and satisfaction. *International Journal of Hospitality Management* , 27, 337-345.
- Herzberg, F. (1975). *Work and the nature of man*. New York: T.Y. Crowell.
- Herzberg, F., Mausner, B., & Snyderman, B. (1962). *The motivation to work*. New York: John Wiley and Sons, Inc.
- Hoontrakul, P., & Sahadev, S. (2008). Application of data mining techniques in the on-line travel industry: A case study from Thailand. *Marketing Intelligence & Planning*, 26 (1), 60-76.
- Joshi, A., & Giménez, E. (2014). Decision-driven marketing. *Harvard Business Review* , July-August, 65-71.
- Köktürk, M., & Dirsehan, T. (2012). *Veri Madenciliği ile Pazarlama Etkileşimi (Interaction between Data Mining and Marketing)*. Ankara: Nobel.

- Laudon, K., & Laudon, J. (2014). *Management Information Systems: Managing the Digital Form* (13th Global Edition b.). USA: Pearson Education Limited.
- Lee, S., & Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems* , 101 (1), 41-46.
- Leong, E., Ewing, M., & Pitt, L. (2004). Analysing competitors' online persuasive themes with text mining. *Marketing Intelligence & Planning* , 187-200.
- Liau, B., & Tan, P. (2014). Gaining customer knowledge in low cost airlines through text mining. *Industrial Management & Data Systems* , 114 (9), 1344-1359.
- Min, H., Min, H., & Emam, A. (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of Contemporary Hospitality Management* , 14 (6), 274-285.
- Moon, S., Park, Y., & Kim, Y. (2014). The impact of text product reviews on sales. *European Journal of Marketing* , 48 (11/12), 2176-2197.
- Mostafa, M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications* , 40, 4241-4251.
- MSI (Marketing Science Institute). Big Data. <http://www.msi.org/topics/big-data/> , date of access: 29th August 2015.
- MSI (no date). 2014-2016 Research Priorities. http://www.msi.org/uploads/files/MSI_RP14-16.pdf , 1-16, date of access: 29th August 2015.
- Reichheld, F. (1996). *The Loyalty Effect*. Boston, MA: Harvard Business School Press.
- Rickman, T., & Cosenza, R. (2007). The changing digital dynamics of multichannel marketing. *Journal of Fashion Marketing and Management: An International Journal* , 11 (4), 604-621.
- Solomon, M. (2009). *Consumer Behavior: Buying, Having, and Being* (8th International Edition b.). New Jersey: Pearson Education Inc.
- Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007). Overview and Semantic Issues of Text Mining. *SIGMOD Record* , 36 (3), 23-34.
- Tepeci, M. (1999). Increasing brand loyalty in the hospitality industry. *International Journal of Contemporary Hospitality Management* , 11 (5), 223-229.
- Tsang, A., & Prendergast, G. (2009). Is a "star" worth a thousand words? *European Journal of Marketing* , 43 (11/12), 1269-1280.
- Tuten, T., & August, R. (1998). Understanding consumer satisfaction in services settings: a bidimensional model of service strategies. *Journal of Social Behavior and Psychology* , 13 (3), 553-564.
- Wang, H., & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. *Industrial Management & Data Systems* , 108 (5), 622-634.
- Xiang, Z., Schwartz, Z., Gerdes, J., & 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.
- Yan, X., Wang, J., & Chau, M. (2015). Customer revisit intention to restaurants: Evidence from online reviews. *Information Systems Frontiers* , 17, 645-657.

Appendixes

Appendix 1. Critical Words Revealed by Text Mining from Travelers' Reviews

Critical Words	Times They Occurred in Positive Comments	Times They Occurred in Negative Comments
Locat-	1963	135
Staff	1071	302
Help-	602	103
Friendl-	586	41
Clean-	533	217
Breakfast	480	547
Comfort-	448	70
Servic-	274	261
Bed	196	224
Pool	160	128
Restaurant	86	62
Food	93	89
Facilities	113	78
Price	108	130
Wifi	168	432
City	103	38
Shop	147	23
Metro	92	13
Bar	88	81
Money	86	71
Bathroom	81	214
Central	49	1
Reception	77	129
Bus	45	14
Parking	43	86
Polite	43	2
Design	41	18
Lobby	41	64
Décor	37	21
Park	36	7
Shower	58	161
Book-	36	161
Taxi	14	60
Pay	8	104
Nois-	17	125
Toilet	16	75
Sleep	16	58
Tea	23	57
TV	11	52
Luggage	17	48
View	210	64
Expens-	17	164

Appendix 1 (continued).

Critical Words	Times Occurred in the Positive Comments	Times Occurred in the Negative Comments
Window	34	114
Smell	11	100
Towel	18	97

Appendix 2. Significant Correlation Scores between Word Occurrences in the Positive Reviews and Travelers' Review Score

Word Occurrences in the Positive Reviews	Pearson Correlation Coefficient (with Travelers' Review Score)	Sig.
staff	.239**	.000
help-	.139**	.000
comfort	.125**	.000
friendl-	.118**	.000
clean-	.104**	.000
service-	.100**	.000
view	.079**	.000
facilit-	.068**	.000
bar	.068**	.000
decor	.060**	.000
bed	.052**	.001
breakfast	.050**	.001
pool	.050**	.003
bathroom	.049**	.002
food	.047**	.003
design	.047**	.003
bus	.043**	.006
money	.039*	.014
towel	.038*	.017
window	.037*	.018
tea	.033*	.036

* significant at the $p < .05$

** significant at the $p < .01$

Appendix 3. Significant Correlation Scores between Word Occurrences in the Negative Reviews and Travelers' Review Score

Word Occurrences in the Negative Reviews	Pearson Correlation Coefficient (with Travelers' Review Score)	Sig.
clean-	-.187**	.000
staff	-.184**	.000
smell	-.124**	.000
bed	-.122**	.000
service-	-.108**	.000
book-	-.108**	.000
friendl-	-.102**	.000
nois-	-.100**	.000
reception	-.097**	.000
toilet	-.092**	.000
bathroom	-.089**	.000
money	-.085**	.000
towel	-.084**	.000
shower	-.075**	.000
TV	-.073**	.000
polite	-.062**	.000
comfort-	-.058**	.000
help-	-.053**	.001
luggage	-.053**	.001
window	-.053**	.001
sleep	-.049**	.002
bar	-.044**	.005
facilit-	-.042**	.008
pay-	-.039*	.012
wifi	-.031*	.048

* significant at the $p < .05$

** significant at the $p < .01$