

Article

# Investigating Key Attributes in Experience and Satisfaction of Hotel Customer Using Online Review Data

Hyun-Jeong Ban <sup>1</sup>, Hyeon Choi <sup>2</sup>, Eun-Kyong Choi <sup>2</sup>, Sanghyeop Lee <sup>3</sup> and Hak-Seon Kim <sup>1,\*</sup> 

<sup>1</sup> School of Hospitality and Tourism Management, Kyungsoong University, 309 Suyoungro, Nam-Gu, Busan 48434, Korea; helenaban@naver.com

<sup>2</sup> Department of Nutrition and Hospitality Management, The University of Mississippi, Oxford, MS 38677, USA; hchoi9@go.olemiss.edu (H.C.); Echoi2@olemiss.edu (E.-K.C.)

<sup>3</sup> Department of Tourism Management, Keimyung University, 1095 Dalgubeol-daero, Dalseo-gu, Daegu 42601, Korea; leesanghyeop@gw.kmu.ac.kr

\* Correspondence: kims@ks.ac.kr; Tel.: +82-51-663-4473

Received: 31 October 2019; Accepted: 20 November 2019; Published: 21 November 2019



**Abstract:** With the development of social media, customers are sharing their experiences, and it is rapidly spreading as a form of online review. That is why the online review has become a significant information source affecting customers' purchase intention and behavior. Therefore, it is important to understand the customer's experience shown in the online review in order to maintain sustainable customer satisfaction and loyalty. The purpose of this study is to investigate what are the key attributes and the structural relationship of those key attributes. To accomplish this purpose, a total of 6596 hotel reviews were collected from Google (google.com). A frequency analysis using text mining was performed to figure out the most frequently mentioned attributes. In addition, semantic network analysis, factor analysis, and regression analysis were applied to understand the experience and satisfaction of the hotel customer. As a result, the top 99 keywords were divided into four groups such as "Intangible Service", "Physical Environment", "Purpose", and "Location". The factor analysis reduced the dimension of the original 64 keywords to 22 keywords, and grouped them into five factors, which are "Access", "F&B (Food and Beverage)", "Purpose", "Tangibles", and "Empathy". Based on these results, theoretical and practical implications for sustainable hotel marketing strategies are suggested.

**Keywords:** customer experience; customer satisfaction; selective attribute; online hotel review; eWOM; big data; semantic network analysis; hospitality marketing

## 1. Introduction

Nowadays, reviewing a product or service online is common for customers. In the online review, customer leave text comments with a numeric rating to briefly indicate their evaluation of the product or service [1]. The review can be called word of mouth (WOM). In the Internet era, the effect of WOM has been further enhanced in the form of electronic word of mouth (eWOM), and it has developed substantially [2]. The eWOM is a new method to identify the main attributes of service quality from a customer's perspective. The eWOM is the result of a summary of the customer's experience and is usually written voluntarily without any economic cost or external stimulus [3]. This eWOM is used by customers who have experienced a particular service to help other customers make the right choice. Therefore, the experience of the services mentioned in the eWOM implies the main attributes and quality levels of the product or service that the customer considers [4]. Accordingly, online review mining research is actively under way to extract customer information needed to develop new products

or improve existing products [5] from online review in various service industries such as medical service, airline [6], dining [7], and hotel [8].

Previously, many studies have focused on identifying key attributes that investigate customer satisfaction with the survey method. Cheng et al. [9] investigated the impact of service recovery dimensions on customer satisfaction and subsequently on customer loyalty in the context of the hotel industry. Nysveen et al. [10] examined the influence of a brand's innovativeness and green image with hotel brand satisfaction. Pizam et al. [11] discussed customer satisfaction and its application to the hospitality and tourism industries. However, the majority of prior studies using the survey method were done a few years ago [12–16]. Nowadays researchers are using the text mining method for analyzing eWOM [4,8,16–19].

A hotel is an industry in which consumers themselves are the property of the enterprise and engage in production activities through direct contact with the customer. That is why understanding the opinions of customers and their experiences is significant. In addition, the hotel is a special product that combines the tangible product of the facility with the intangible product of human service [20]. Therefore, it is difficult to identify the main attributes because there are so many service attributes and the customer's needs vary depending on the type, purpose of visit, and local characteristics. In view of these characteristics, this study focuses on understanding the customer through online review mining and examines customer experience and satisfaction to find out what key attributes were affecting the customer intention.

## 2. Literature Review

### 2.1. Customer Experience and Satisfaction

Customer satisfaction is a complex customer experience in the service industry and can be defined as an evaluation of what the customers have experienced [5]. Understanding what customers expect from a service industry is important in order to provide a standard of comparison against which customers consider an organization's performance regarding the expectation [7]. Service quality can be defined as a customer's overall impression of the relative efficiency of the organization [21]. In addition, customer satisfaction can be defined as an experience made on the basis of a specific service encounter, and it contributes to loyalty, repeat purchase, positive WOM, and ultimately higher profitability [22].

Hotel selection attributes are what a customer considers important among the many attributes a hotel has when choosing a hotel [23]. Hotel selectivity is a determinant that is particularly preferred and considered by the customer and is defined as a target for assessing satisfaction and dissatisfaction [20,24]. Therefore, identifying a customer's hotel selection attributes is part of the effort to improve the quality of service and increase customer satisfaction to gain competitive advantage [25].

### 2.2. Electronic Word of Mouth (eWOM)

eWOM is an Internet-based development in which the delivery of reviews takes place on the Internet [19]. This refers to the online exchange of product information and experience [21]. Unlike WOM, eWOM is very influential to customers as it can be written and edited without the limitations of time and space. Through eWOM, customers share their product reviews with others through social network service (SNS) and online communities regardless of commercial interests. Unlike one-sided advertising, an actual review or specific information from customers with experience is greatly influenced by customers' selection attributes [26]. Through the Internet, people can freely produce and share their experiences and opinions with potential customers at all times.

The service features inseparability where purchases and consumption take place at the same time and intangibility being invisible, unlike ordinary products [27]. Therefore, customers want more specific and practical information to minimize failure. At this time, people will listen to more diverse and specific opinions from a wider group of experienced people through Internet searches, beyond the limited information they get from the people around them [28]. According to Klein [29], this is because

customers rely more on the reviews of the actual customers when purchasing experience products than when buying general merchandise. Hu et al. [17] said that online reviews are far more interesting, reflect the best information, and are more reliable than information provided by the company. In other words, customers are more sympathetic to the experiences they have left behind than the promotional information provided by the company. This online review is used as a new source of information for other customers looking for information about the hotel. Accordingly, reviewers will play a role as opinion leaders, whether or not they are meant to be.

### 2.3. Text Mining and Semantic Network Analysis

The text mining refers to discovering unknown useful patterns and knowledge in text using information retrieval, information extraction, and natural language processing techniques [30]. The general process of text mining consists of the steps of data collection, data refining, data analysis, and management information system as shown in Figure 1. Data collection identifies and clarifies the type of information that researchers seek to find. Then, it is important to limit the range of data you want to collect and to be familiar with the characteristics of keywords. The data refining process is the process of converting unstructured text data into structured forms. When the analysis stage is used to analyze text based on technologies such as information extraction, clustering, and categorization, it is utilized as a management information system and accumulated as knowledge [31,32].

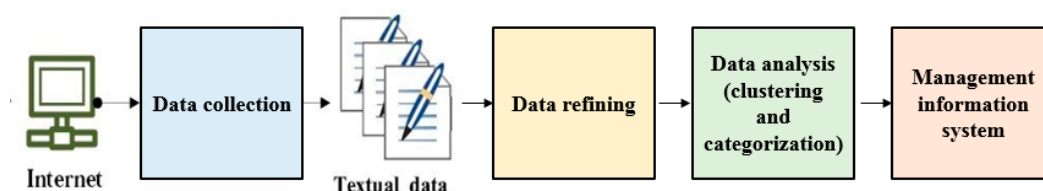


Figure 1. Text mining process [15].

Research using text mining in the hospitality industry has been active recently. Joseph and Varghese [33] analyzed user reviews to understand various aspects that drive customer satisfaction using text mining analysis. Hu et al. [17] analyzed 27,864 hotel reviews in New York City, and the results showed that customer complaints for high-end hotels are mainly related to service issues. Kuhzady and Ghasemi [34] used text mining to analyze online hotel reviews for Mazandaran province in Iran. Berezina et al. [35] analyzed the factors of satisfaction and dissatisfaction of U.S. hotel customers, and Philander and Zhong [36] used text mining to analyze customer's feelings about a Las Vegas resort on Twitter.

Semantic network analysis can be described as the use of network analytics techniques based on shared meaning as opposed to paired associations of behavioral or perceived communication links [37–39]. In other words, text is coded into a network to identify the structural relationship between words. Semantic network analysis can identify the inherent meaning of the context that has not been revealed. Moreover, it is possible to understand the degree of influence of a particular word through frequency and cluster analysis of words and to analyze how a particular word affects the relationships between groups [40,41]. Semantic network analysis, as a method of quantitative textual analysis, provides an impressive theoretical and methodological foundation with which to describe the semantic nature of the online review [42].

## 3. Methodology

### 3.1. Data Collection

The data collection procedure used in this study was as follow. The first step was to collect the research data. The online hotel reviews were obtained from google.com, which is the largest search engine in the world. That is why it is easy to access and share the online hotel review. The data was

collected by web crawling, and the web crawler was written in the Python 2.7. The server operating system was Ubuntu 16.04 LTS from google.com. Several types of data were collected including hotel brands, writer's identification code, written date, overall score, and review as shown in Figure 2. The overall score was defined as the customer satisfaction in this study.

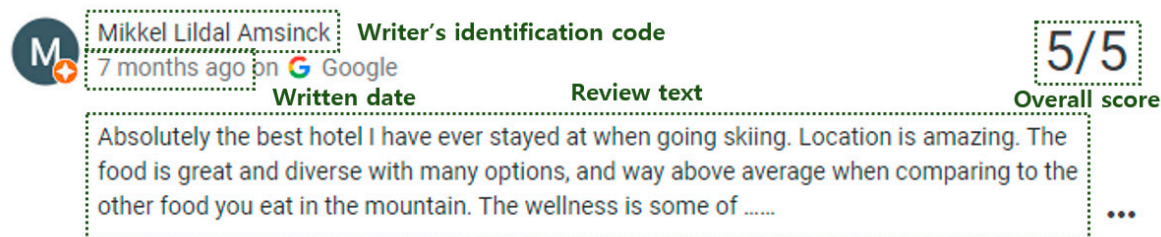


Figure 2. Google review data screen.

Initially, 19,332 reviews were collected in top 25 hotels in the world. The ranking information was from TripAdvisor (tripadvisor.com). The top 25 hotels and numbers of reviews are presented in Table 1. After deleting online reviews of languages other than English, there were 6597 online reviews left and a total of 189,571 words were extracted. The period of collected data was for five years from August 2014 to July 2019.

Table 1. Number of reviews according to hotel brands.

Rank	Brands	No. of Reviews	Rank	Brands	No. of Reviews
1	Tulemar Bungalows and Villas	181	14	Hotel Amira Istanbul	332
2	Hotel Belvedere	230	15	Hotel 41	217
3	Viroth's Hotel	253	16	Ikos Oceania	449
4	Kenting Amanda Hotel	97	17	Mandapa, a Ritz-Carlton Reserve	466
5	Hotel Alpin Spa Tuxerhof	113	18	Nayara Springs	147
6	French Quarter Inn	339	19	Rosewood Mayakoba	209
7	The Resort at Pedregal	487	20	Valle D'incanto Midscale	77
8	Belmond Palacio Nazarenas	139	21	Hotel Spadai	271
9	Kayakapi Premium Caves	245	22	Constance Prince Maurice	268
10	Hanoi La Siesta Hotel and Spa	278	23	O'Gallery Premier Hotel	377
11	Golden Temple Retreat	220	24	The Nantucket Hotel and Resort	188
12	Quinta Jardins do Lago	170	25	AYADA Maldives	219
13	The Oberoi Rajvilas	625			
	Total/Average			6597/263	

### 3.2. Data Analysis

As for the data analysis of this study, first, the text mining technique was applied to obtain the word frequency from online hotel reviews. The articles, prepositions, and pronouns that are meaningless were excluded, and words that pertained to hotel experience were contained in the refined data. The top 99 frequent words were manually selected after refining the collected data. Furthermore, the distribution of hotel experience evaluation based upon overall satisfaction score was performed to generally ascertain the satisfaction level of hotel experience, and this overall score was used as a dependent variable because its value can be treated as a main output variable. In addition, the word matrix (keywords  $\times$  keywords) was deduced for further data analysis.

Secondly, the semantic network analysis of these top 99 frequent words was performed using the UCINET 6.0 package with the visualization tool named NetDraw. Furthermore, Freeman's degree centrality and Eigenvector centrality were chosen for illustrating the semantic network. Finally, CONCOR (CONvergence of iterated CORrelation) analysis was conducted to obtain the subgroups of these words so as to understand these interwoven correlations with each other and figure out the facets that customers are interested in.

At last, the results of semantic network analysis were synthesized to select words for further factor analysis and linear regression analysis with a dummy variable. Factor analysis was performed to derive the main factors affecting hotel satisfaction using 64 words out of the 99 top-frequency words, which will be considered as variables for the linear regression analysis. Moreover, the linear regression analysis comprised six independent variables retrieved from the factor analysis and the overall score for each review as a dependent variable was employed to test the hypothesis: The hotel experience represented in online reviews can be used to explain customer satisfaction.

## 4. Result

### 4.1. Frequency Analysis

To find the words most frequently used in customer reviews, Table 2 listed the top 99 frequent words associated with the hotel experience. The top five words were ‘staff’, ‘service’, ‘room’, ‘place’, and ‘food’. The distribution of frequently used words is shown in Figure 3, and the result of visualizing the network that reflects the frequency is Figure 4. There were words describing the food, such as ‘food’, ‘breakfast’, ‘restaurant’, ‘dining’, ‘bar’, ‘drink’ and describing the service, such as ‘staff’, ‘service’, ‘hospitality’, ‘care’. Likewise, there were the words related to facility, such as ‘room’, ‘resort’, ‘view’, ‘pool’, ‘spa’, ‘villa’, ‘facility’, ‘beach’, ‘garden’, ‘bathroom’, ‘lake’ and the words related to the location or name of the hotel, such as ‘place’, ‘locate’, ‘Hanoi’, ‘Quarter’, ‘Ayada’, ‘Mandapa’, ‘distance’, ‘Jaipur’, ‘Istanbul’, ‘Nantucket’, ‘Amira’, ‘Charleston’.

**Table 2.** Top 99 frequent words from the online hotel review.

Rank	Word	Freq.	Rank	Word	Freq.	Rank	Word	Freq.
1	staff	2332	34	people	168	67	bathroom	99
2	service	1709	35	suite	158	68	price	98
3	room	1708	36	garden	156	69	lake	97
4	place	1024	37	quality	154	70	vacate	96
5	food	835	38	detail	152	71	Quarter	94
6	locate	784	39	city	148	72	street	94
7	resort	640	40	Hanoi	146	73	care	92
8	stay	636	41	class	145	74	part	90
9	view	628	42	holiday	145	75	Ayada	90
10	breakfast	534	43	visit	144	76	reception	90
11	everything	496	44	minute	140	77	walk	87
12	time	488	45	airport	138	78	butler	87
13	night	486	46	home	134	79	Mandapa	86
14	experience	479	47	hospitality	134	80	check	86
15	restaurant	451	48	bar	131	81	kid	85
16	pool	407	49	custom	129	82	distance	85
17	day	362	50	amenity	127	83	manage	84
18	family	295	51	guest	127	84	kind	83
19	star	293	52	wife	127	85	water	81
20	property	264	53	everyone	126	86	thank	79
21	trip	245	54	honeymoon	124	87	conciierge	77
22	year	242	55	drink	122	88	heart	77
23	area	216	56	town	120	89	desk	76
24	world	214	57	expectation	120	90	birthday	76
25	moment	210	58	attention	118	91	anniversary	74
26	weekend	204	59	husband	112	92	cave	73
27	spa	203	60	level	110	93	Jaipur	73
28	villa	202	61	boutique	108	94	Istanbul	72
29	facility	192	62	accommodate	105	95	Nantucket	71
30	arrival	190	63	friend	104	96	Amira	71
31	beach	179	64	tour	103	97	notch	71
32	luxury	178	65	ground	101	98	choice	71
33	dining	173	66	team	99	99	Charleston	70



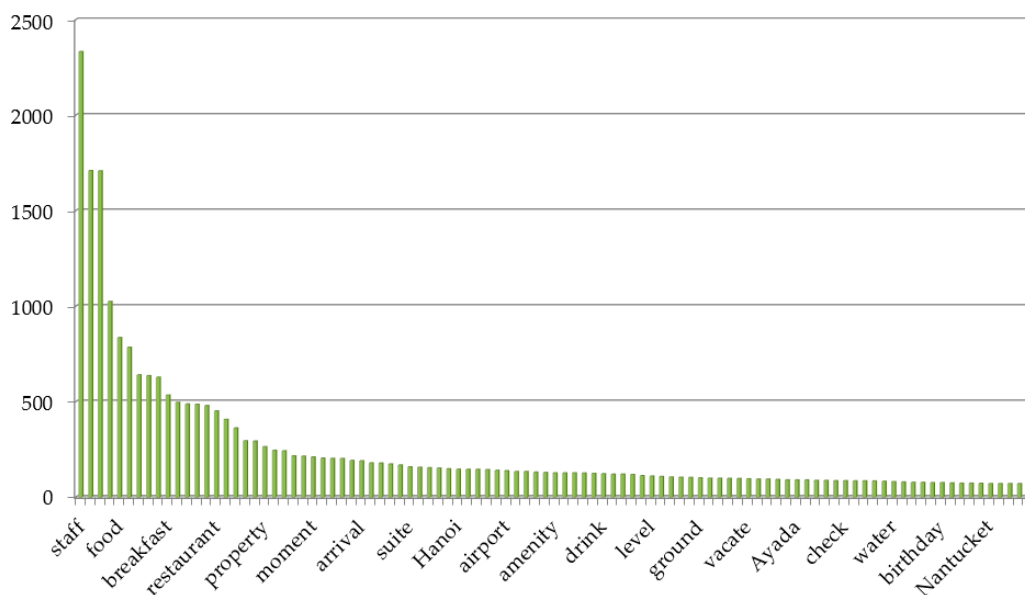


Figure 3. Distribution of the frequently used words in the hotel online review.

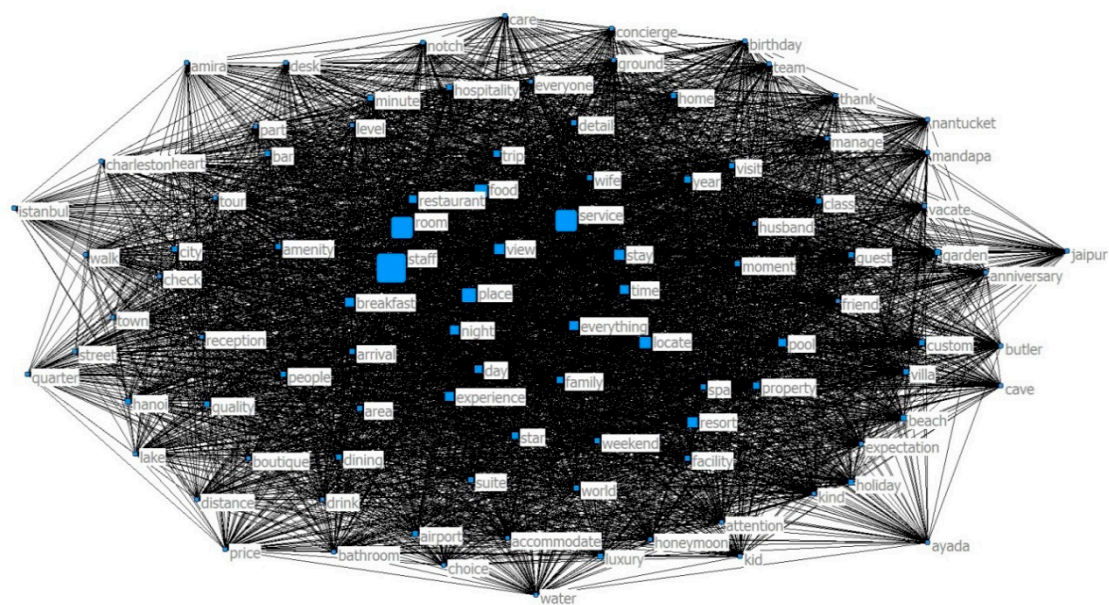


Figure 4. Keyword visualization of network analysis.

4.2. Semantic Network Analysis

The semantic network analysis identifies the relationship between words and expresses the connection between them. The centrality and CONCOR analyses of keywords were performed. In the online hotel review, among the top 99 frequent words, the results of an analysis of the degree and eigenvector centrality of the words are described in Table 3.

The degree centrality is a simple centrality measure that counts how many neighbors a node has and refers to the degree to which a word has many connections and becomes central, and the more connections it has, the greater its impact on other words and the more dominant it can be [43]. The eigenvector centrality extends the concept of connective centrality by considering not only the number of words connected, but also how important a connected relationship is. Thus, it is a useful indicator for finding the most influential central node in the network [44]. It is sometimes used to measure a node’s influence in the network. It performs matrix calculations to determine adjustments.

**Table 3.** Comparison of keyword frequency and centrality analysis.

	Frequency		Degree		Eigenvector			Frequency		Degree		Eigenvector	
	Freq	Rank	Coef.	Rank	Coef.	Rank		Freq	Rank	Coef.	Rank	Coef.	Rank
staff	2332	1	0.086	1	0.469	1	weekend	204	26	0.009	25	0.052	25
service	1709	2	0.063	3	0.372	3	spa	203	27	0.010	22	0.065	20
room	1708	3	0.070	2	0.425	2	villa	202	28	0.009	26	0.047	33
place	1024	4	0.029	6	0.184	6	facility	192	29	0.008	31	0.055	24
food	835	5	0.036	4	0.240	4	arrival	190	30	0.008	28	0.047	34
locate	784	6	0.031	5	0.222	5	beach	179	31	0.008	30	0.049	29
resort	640	7	0.024	9	0.136	12	luxury	178	32	0.006	42	0.034	51
stay	636	8	0.026	7	0.164	8	dining	173	33	0.007	32	0.048	30
view	628	9	0.025	8	0.154	9	people	168	34	0.007	36	0.040	39
breakfast	534	10	0.023	10	0.176	7	suite	158	35	0.007	34	0.043	36
everything	496	11	0.021	13	0.137	11	garden	156	36	0.007	37	0.044	35
time	488	12	0.021	11	0.126	15	quality	154	37	0.007	35	0.052	26
night	486	13	0.021	12	0.132	14	detail	152	38	0.007	33	0.048	31
experience	479	14	0.017	16	0.105	16	city	148	39	0.006	45	0.038	43
restaurant	451	15	0.020	14	0.139	10	Hanoi	146	40	0.006	39	0.042	37
pool	407	16	0.020	15	0.135	13	class	145	41	0.006	38	0.039	40
day	362	17	0.016	17	0.098	17	holiday	145	42	0.005	55	0.032	55
family	295	18	0.013	18	0.074	19	visit	144	43	0.006	41	0.036	44
star	293	19	0.012	19	0.077	18	minute	140	44	0.006	47	0.032	56
property	264	20	0.010	20	0.061	22	airport	138	45	0.006	46	0.036	45
trip	245	21	0.010	21	0.063	21	home	134	46	0.006	49	0.034	52
year	242	22	0.009	23	0.049	27	hospitality	134	47	0.005	57	0.031	60
area	216	23	0.009	24	0.057	23	bar	131	48	0.006	40	0.042	38
world	214	24	0.008	29	0.047	32	custom	129	49	0.005	52	0.039	41
moment	210	25	0.008	27	0.049	28	amenity	127	50	0.006	48	0.039	42

The result is that ‘staff’, ‘service’, ‘room’ recorded top in both degree and eigenvector centrality. The word ‘breakfast’ recorded a lower rank in frequency and degree centrality than in eigenvector centrality. The words ‘spa’, ‘facility’, ‘quality’, ‘bar’, ‘amenity’ recorded higher in degree and eigenvector centrality than in frequency. However, the word ‘luxury’ recorded higher in frequency than in degree and eigenvector centrality.

The CONCOR analysis is the connection of the relationship and discovering patterns between words, and the greater the similarity of the connection relationship patterns, the greater the degree of structural equivalence of the other words. It forms clusters that include keywords with similarities to each other [45]. In other words, the CONCOR analysis is a method of repeatedly analyzing correlations to search certain levels of similarity groups. This study identifies the blocks of nodes according to the correlation coefficient of the matrix of the concurrent keywords and forms clusters that include keywords with similarities [5]. The keywords extracted from the frequency histogram according to the frequency ranking were used and a matrix was constructed. To visualize the results, NetDraw in UCINET 6.0 program was applied. The nodes are presented as blue squares and their sizes indicate their frequency, and the network shows the connectivity between them.

The result of the CONCOR analysis is shown in Figure 5 with visibility. There are four groups that were intricately interwoven with each other. After looking at the words in the group, the group was named as “Intangible Service”, “Physical Environment”, “Location”, and “Purpose”.

To make it easier to see which words belong to each group, the words grouped in the cluster and the ones to be noted are listed in Table 4. As shown in Table 4, the group name was chosen as follows considering the characteristics of the words. The “Intangible Service” group consists of ‘staff’, ‘service’, ‘food’, ‘hospitality’, ‘concierge’, ‘dining’, ‘breakfast’, ‘drink’, ‘reception’, ‘moment’, ‘desk’, ‘quality’ which are the significant terms for the hotel industry. “Physical Environment” comprises of ‘room’, ‘view’, ‘resort’, ‘restaurant’, ‘spa’, ‘bar’, ‘beach’, ‘facility’, ‘cave’, ‘garden’, ‘pool’, ‘suite’, ‘bathroom’, ‘amenity’ which contains various facilities. “Location” refers to several hotel brands such as ‘Amira’,

‘Mandapa’, ‘Quarter’, ‘Ayada’, ‘Jaipur’, and this group also contains words such as ‘city’, ‘street’, ‘walk’, ‘locate’, ‘distance’. The last group, “Purpose” is composed of words relating with purpose such as ‘honeymoon’, ‘vacate’, ‘anniversary’, ‘holiday’, ‘weekend’, and also contains words relating to company such as ‘people’, ‘wife’, ‘husband’, ‘kid’, ‘family’, ‘friend’. Accordingly, after the CONCOR analysis, 64 words that intimately associate with hotel experience and satisfaction were extracted for further research.

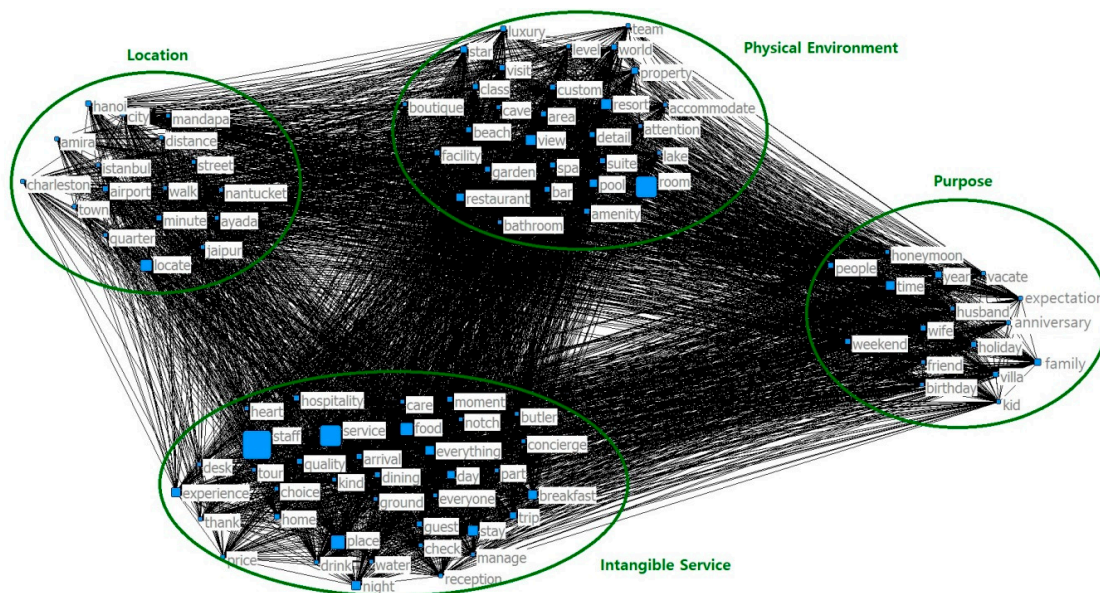


Figure 5. Visualization of CONvergence of iterated CORrelation (CONCOR) analysis.

Table 4. Result of CONCOR analysis.

	Extracted Words	Significant Words
Intangible Service	staff/service/food/place/night/everything/experience/day/breakfast/trip/stay/desk/heart/hospitalityquality/tour/choice/home/thank/price/kind/arrival/dining/ground/water/drink/guest/check/reception/manage/care/moment/notch/butler/concierge/part	staff/service/food/experience/breakfast/hospitality/quality/thank/dining/water/drink/guest/care/notch/butler/concierge
Physical Environment	room/view/resort/pool/restaurant/property/world/team/level/luxury/visit/star/class/cave/facility/boutique/beach/garden/spa/bar/bathroom/amenity/detail/suite/accommodate/attention/lake/custom/area	room/view/resort/pool/restaurant/luxury/facility/garden/cave/boutique/beach/bathroom/bar/spa/suite/amenity/suite/accommodate/lake
Location	locate/minute/walk/airport/Quarter/Istanbul/distance/street/Nantucket/Ayada/Jaipur/Mandapa/city/Hanoi/Amira/Charleston/town	Locate/minute/walk/airport/Quarter/Istanbul/distance/street/Nantucket/Ayada/Jaipur/Mandapa/city/Hanoi/Amira/Charleston
Purpose	time/honeymoon/people/weekend/wife/husband/year/vacate/expectation/anniversary/holiday/family/villa/kid/friend/birthday	honeymoon/people/weekend/wife/husband/vacate/expectation/anniversary/holiday/family/kid/friend/birthday

### 4.3. Factor Analysis

The factor analysis can discover the commonalities among these keywords and show connection of variables through the variance of keywords within the same online hotel review. The purpose of the factor analysis is to reduce a large number of variables into smaller factors using an oblique rotation process. Common factorial criteria were used in extracting the factors, and the minimum factor loading was set to 0.400 in the final model. The factors also had to have Eigenvalues greater than 1.0 and had to explain a substantial percentage of total variance. Eight elements had low scores, which is less than 0.3. Ten elements loaded on two factors. Therefore, a total of 18 elements were excised from the 64 keywords. Finally, five factors with 22 keywords covering 22.099% of all variance to elicit the hotel experience were generated through the process of elimination of five times.



Table 5 shows the results of the factor analysis with 0.651 of KMO (Kaiser Meyer Olkin), which is higher than 0.6. Therefore, it was verified that the use of the factor analysis was suitable for this study. The Bartlett's test of sphericity value ( $X^2$ ) was 113,025.397, with overall significance of the correlation matrix ( $p < 0.001$ ). This result showed that the data did not produce an identity matrix, and it was multivariate normal and fit for using exploratory factor analysis. The five factors were named as "Access (Factor 1)", "F&B (Factor 2)", "Purpose (Factor 3)", "Tangibles (Factor 4)", "Empathy (Factor 5)". Factor 1 contains 'locate', 'town', 'Amira', 'walk', which are related to the accessible location. Factor 2 has 'food', 'breakfast', 'drink', 'dining' which is related with F&B in the hotel. Factor 3 was about purpose containing 'weekend', 'birthday', 'anniversary', 'honeymoon'. In addition, Factor 4 consisted of aspects concerning intangibles, such as 'pool', 'spa', 'beach', 'restaurant', 'bar', 'room'. Factor 5 has 'staff', 'care', 'service' and 'friend', which is related with empathy.

**Table 5.** Result of the factor analysis.

	Words	Factor Loading	Eigen Value	Variance (%)
Access	Locate	0.962	3.118	16.409
	Town	0.949		
	Amira	0.940		
	Walk	0.933		
Food and Beverage (F&B)	Food	0.928	2.818	14.832
	Breakfast	0.912		
	Drink	0.875		
	Dining	0.773		
Purpose	Weekend	0.817	2.306	12.135
	Birthday	0.815		
	Anniversary	0.685		
	Honeymoon	0.627		
Tangibles	Pool	0.747	2.004	10.547
	Spa	0.687		
	Beach	0.682		
	Restaurant	0.480		
	Bar	0.475		
	Room	0.437		
Empathy	Staff	0.844	1.858	9.778
	Care	0.738		
	Service	0.470		
	Friend	0.444		
Total variance (%) = 22.099				
KMO (Kaiser Meyer Olkin) = 0.657				
Bartlett chi-square(p) = 113,025.397 (p < 0.001)				

#### 4.4. Linear Regression Analysis

After the factor analysis, a customer experience and satisfaction was derived by using regression analysis as shown in Table 6. It has five independent variables: Access (A), F&B (FB), Purpose (P), Tangibles (T), Empathy (E), and one dependent variable: Customer Satisfaction (CS). The overall variance explained by the five predictors was 12% ( $R^2 = 0.120$ ) and the standard error of the estimated value was calculated as 0.51021. The correlation between the independent and dependent variables was rather low because many factors affecting customer experience and satisfaction might not have been included among the five factors due to their low frequency in the online hotel reviews. In regression models related to opinion mining it is impossible to include all relevant variables to estimate output variables such as opinion from text mining data. Therefore, the  $R^2$  value can be low [46–48]. There is a

prior study saying that the  $R^2$  value is 12.5%. This study also analyzed online review data for washing machines and used regression analysis and factor analysis [5].

“F&B (FB,  $\beta = 0.036$ ,  $p = 0.003$ )”, “Purpose (P,  $\beta = 0.029$ ,  $p = 0.017$ ), and “Empathy (E,  $\beta = 0.087$ ,  $p = 0.000$ )” are significant at the  $p < 0.05$  level and positively related with customer satisfaction. In order to estimate the possible correlations among the predictors, a multicollinearity statistic was conducted. The tolerance level was less than 1.00, and the variance inflation factor (VIF) of the predictors was between 1.00 and 1.10, respectively, that is, the predictors were not significantly correlated to each other. Therefore, based on unstandardized  $\beta$ , the regression equation can be expressed as:

$$CS = 4.861 + 0.010A + 0.019FB^{**} + 0.015P^* - 0.002T + 0.045E^{***}$$

The “Empathy (E)” factor holds the highest standardized coefficients, which means this experience aspect of the hotel customer is the most important factor associated with customer satisfaction significantly. Reviews such as “Fantastic natural surroundings and friendly staff give you the best time” and “The thing that made this place so great for us was the staff service, especially our personal Concierge” are related to the hotel experience based upon “Empathy” attributes.

**Table 6.** Results of linear regression analysis.

Model	Unstandardized Coef.		Standardized Coef.	t
	B	Std. error	Beta	
(Constant)	4.861	0.006		773.569
Access (A)	0.010	0.006	0.019	1.556
Food and beverage (FB)	0.019	0.006	0.036	2.947 **
Purpose (P)	0.015	0.006	0.029	2.377 *
Tangibles (T)	−0.002	0.006	−0.004	−0.359
Empathy (E)	0.045	0.006	0.087	7.114 ***

Notes: Dependent variable: Customer Satisfaction (CS);  $R^2 = 0.120$ ; adjusted  $R^2 = 0.100$ ;  $F = 13.894$ ; \*  $p < 0.05$ ; \*\*  $p < 0.005$ ; \*\*\*  $p < 0.001$ .

## 5. Discussion

This study was conducted to enhance the customer’s experience and satisfaction using online hotel reviews. For the online hotel review data analysis, the first process was extracting keywords by text mining and the second was calculating the frequency of words used by customers. Based on the frequency analysis, the degree and eigenvector centrality of top 99 frequent words were analyzed to search their connection and the most affected keywords among them. The CONCOR analysis was performed for grouping them. As a result, the top 99 keywords were divided into four groups, namely, “Intangible Service”, “Physical Environment”, “Purpose”, and “Location”. Moreover, they were visualized by drawing networks and nodes using NetDraw in UCINET 6.0. In addition, the study conducted factor analysis and linear regression analysis to understand the relationship between extracted factors and customer satisfaction. The factor analysis reduced the dimension of the original 64 keywords to 22 keywords and grouped them into five factors, which are “Access”, “F&B”, “Purpose”, “Tangibles”, and “Empathy”. The clusters can be related between CONCOR analysis and factor analysis, such as “Intangible Service” with “Empathy”, “Physical Environment” with “Tangibles”, “Location” with “Access”.

First of all, the group representing the highest beta coefficient was “Empathy” in the linear regression analysis, and the related words were ‘staff’, ‘service’, ‘care’, and ‘friend’ through the factor analysis. Especially, ‘staff’ and ‘service’ were most frequent words in the online hotel review, and through the CONCOR analysis, “Intangible Service” group was the biggest group compared with “Physical Environment”, “Purpose”, and “Location”. According to Lee [49] the service from hotel staff is the most important attribute to making customers satisfied rather than the other luxurious or new

facilities. In addition, Han and Chung [50] examined customers satisfied with excellent service by hotel staff rather than physical environment, such as room cleanness, comfort, and room condition. The results of this study show the same results as many prior studies show that intangible service has the greatest impact on customer experience and satisfaction. Therefore, Service by staff is an essential key attribute to create a good reputation in the service industry and can still be seen as a part of the hotel that must be managed at all times to keep up the image of the hotel. Therefore, it is important to improve the attitude of employees through systematic service training. In addition, providing an appropriate working environment to enhance employee satisfaction to produce better service to customers can be another way.

The second highest beta value was “F&B” in the linear regression analysis, and the related words were ‘food’, ‘breakfast’, ‘drink’, and ‘dining’ through the factor analysis and those keywords are recorded a very high position in the frequency analysis. In hotels with a high satisfaction index, F&B facilities should also be well equipped and restaurants providing high-quality food to customers are significant. Especially, as keywords for breakfast are derived, it will be helpful for hotel business to focus on breakfast rather than on other meals.

This study shows the academic implication that the study has extended its application area of the semantic network analysis. While given the significance of the hotel segment in the tourism industry, this study empirically explores hotel experience and satisfaction by big data analytics. Along the way, the hotel industry has the opportunity to gain an understanding of attributes on the online review, so as to infiltrate into this market and investigate corresponding marketing strategies for their significant advantages. Understanding online reviews as a manifestation of customers’ experiences can help the hotel industry to identify the main attributes required to achieve positive post-purchase behaviors and to minimize negative intentions. Thus, the online reviews not only provide an efficient way for the hotel industry to collect feedback from hotel customers, but also provide an opportunity to discover how to generate positive intent after the experience. To create a high satisfaction score and a positive eWOM, the hotel industry should consider “F&B”, “Purpose”, and “Empathy”. Among them, “Empathy” was the most influential attribute in the regression analysis. These key factors may be used to examine the customer satisfaction or to test theoretical models to have a better understanding of a hotel customer’s behavior.

In practice, the analysis of online reviews can be used as a marketing tool by managers since customer review is an important source for a hotel to improve service and to create some promotion regarding profit. The analysis also provides the level of importance of these service attributes so the hotel industry can allocate their resources accordingly. The online review analyses can provide reliable satisfaction assessment. The hotel industry can also use this method to analyze their competitors’ customer reviews so that they can benchmark themselves against competitors in terms of customer satisfaction. These reviews can be used for sustainable strategic marketing decisions against competitors.

However, this study shows limitations in data collection. Firstly, the data collected in this study is limited, because this study focuses on only the top 25 hotels in the world for the sample. That is why future research should collect online review data from more hotel industries to generalize the findings. Secondly, the collected text was analyzing based on the frequency of individual words, therefore, it is difficult to understand the additional meaning of words. In future studies, further analysis of positives and negatives and sentimental analysis is expected to be carried out to better understand the customer’s experience and satisfaction. Therefore, it can provide stronger strategies to the hotel industry.

**Author Contributions:** H.-J.B. and H.-S.K. designed the research model, analyzed online review data and wrote the paper. H.C., E.-K.C., and S.L. were in charge of review and editing the paper.

**Funding:** This research was supported by Kyung Sung University Research Grants in 2019 [grant number KSU-Grants2019]. Additionally, this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A2A03928029).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Siering, M.; Deokar, A.V.; Janze, C. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decis. Support Syst.* **2018**, *107*, 52–63. [[CrossRef](#)]
2. Kim, W.H. The impact of online reviews on customer satisfaction: An application of the American customer satisfaction index (ACSI). *Int. J. Tour. Hosp. Res.* **2017**, *32*, 65–78. [[CrossRef](#)]
3. Rodríguez-Díaz, M.; Rodríguez-Voltes, C.I.; Rodríguez-Voltes, A.C. Gap analysis of the online reputation. *Sustainability* **2018**, *10*, 1603. [[CrossRef](#)]
4. Gao, B.; Li, X.; Liu, S.; Fang, D. How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience. *Tour. Manag.* **2018**, *65*, 176–186. [[CrossRef](#)]
5. Kim, H.S.; Noh, Y. Elicitation of design factors through big data analysis of online customer reviews for washing machines. *J. Mech. Sci. Technol.* **2019**, *33*, 2785–2795. [[CrossRef](#)]
6. Ban, H.J.; Kim, H.S. Understanding customer experience and satisfaction through airline passengers' online review. *Sustainability* **2019**, *11*, 4066. [[CrossRef](#)]
7. Ban, H.J.; Kim, H.S. A study on the TripAdvisor review analysis of restaurant recognition in Busan 1: Especially concerning English reviews. *Culi. Sci. Hosp. Res.* **2019**, *25*, 1–11.
8. Ban, H.J.; Kim, H.S. Semantic network analysis of hotel package through the big data. *Culi. Sci. Hosp. Res.* **2019**, *25*, 110–119.
9. Cheng, B.L.; Gan, C.C.; Imrie, B.C.; Mansori, S. Service recovery, customer satisfaction and customer loyalty: Evidence from Malaysia's hotel industry. *Int. J. Qual. Serv. Sci.* **2019**, *11*, 187–203. [[CrossRef](#)]
10. Nysveen, H.; Oklevik, O.; Pedersen, P.E. Brand satisfaction: Exploring the role of innovativeness, green image and experience in the hotel sector. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 2908–2924. [[CrossRef](#)]
11. Pizam, A.; Shapoval, V.; Ellis, T. Customer satisfaction and its measurement in hospitality enterprises: A revisit and update. *Int. J. Contemp. Hosp. Manag.* **2016**, *28*, 2–35. [[CrossRef](#)]
12. Torres, E.N.; Kline, S. From customer satisfaction to customer delight: Creating a new standard of service for the hotel industry. *Int. J. Contemp. Hosp. Manag.* **2013**, *25*, 642–659. [[CrossRef](#)]
13. Dominici, G.; Guzzo, R. Customer satisfaction in the hotel industry: A case study from Sicily. *Int. J. Mark. Stud.* **2010**, *2*, 3–12. [[CrossRef](#)]
14. Olorunniwo, F.; Hsu, M.K.; Udo, G.J. Service quality, customer satisfaction, and behavioral intentions in the service factory. *J. Serv. Mark.* **2006**, *20*, 59–72. [[CrossRef](#)]
15. Vijayadurai, J. Service quality, customer satisfaction and behavioural intention in hotel industry. *J. Mark. Comm.* **2008**, *3*, 14–26.
16. Gundersen, M.G.; Heide, M.; Olsson, U.H. Hotel guest satisfaction among business travelers: What are the important factors? *Cornell Hotel Restaur. Admin. Q.* **1996**, *37*, 72–81. [[CrossRef](#)]
17. Hu, N.; Zhang, T.; Gao, B.; Bose, I. What do hotel customers complain about? Text analysis using structural topic model. *Tour. Manag.* **2019**, *72*, 417–426. [[CrossRef](#)]
18. Saura, J.; Reyes-Menendez, A.; Alvarez-Alonso, C. Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability* **2018**, *10*, 3016. [[CrossRef](#)]
19. Belarmino, A.M.; Koh, Y. How E-WOM motivations vary by hotel review website. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 2730–2751. [[CrossRef](#)]
20. Kim, D.K.; Kim, I.S. An analysis of hotel selection attributes present in online reviews using text mining. *J. Tour. Sci.* **2017**, *41*, 109–127.
21. Saha, G.C.; Theingi. Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in Thailand. *Managing Service Quality. Int. J.* **2009**, *19*, 350–372.
22. Browning, V.; So, K.K.F.; Sparks, B. The influence of online reviews on consumers' attributions of service quality and control for service standards in hotels. *J. Travel Tour. Mark.* **2013**, *30*, 23–40. [[CrossRef](#)]
23. Kotler, P.; Bowen, J.T.; Makens, J.; Baloglu, S. *Marketing for Hospitality and Tourism*; Pearson: London, UK, 2017.
24. Knutson, B.J. Frequent travelers: Making them happy and bringing them back. *Cornell Hot. Rest. Admin. Quart.* **1988**, *29*, 82–87. [[CrossRef](#)]



25. Verma, V.K.; Chandra, B. Sustainability and customers' hotel choice behaviour: A choice-based conjoint analysis approach. *Environ. Dev. Sustain.* **2018**, *20*, 1347–1363. [CrossRef]
26. Kim, S.; Kandampully, J.; Bilgihan, A. The influence of eWOM communications: An application of online social network framework. *Comput. Hum. Behav.* **2018**, *80*, 243–254. [CrossRef]
27. Hunt, J.D. Image as a factor in tourism development. *J. Travel Res.* **1975**, *13*, 1–7. [CrossRef]
28. Bhandari, R.S.; Bansal, A. A study of characteristics of e-WoM affecting consumer's social media opinion behavior. *J. Indian Manag. Strateg.* **2019**, *24*, 33–43. [CrossRef]
29. Klein, L.R. Evaluating the potential of interactive media through a new lens: Search versus experience goods. *J. Bus. Res.* **1998**, *41*, 195–203. [CrossRef]
30. Gaikwad, S.V.; Chaugule, A.; Patil, P. Text mining methods and techniques. *Int. J. Comput. Appl.* **2014**, *85*, 42–45.
31. Hassani, H.; Huang, X.; Silva, E.S.; Ghodsi, M. A review of data mining applications in crime. *Stat. Anal. Data Min. ASA Data Sci. J.* **2016**, *9*, 139–154. [CrossRef]
32. Sirsat, S.R.; Chavan, D.V.; Deshpande, D.S.P. Mining knowledge from text repositories using information extraction: A review. *Sadhana* **2014**, *39*, 53–62. [CrossRef]
33. Joseph, G.; Varghese, V. Analyzing Airbnb customer experience feedback using text mining. In *Big Data and Innovation in Tourism, Travel, and Hospitality*; Springer: The Gateway, Singapore, 2019.
34. Kuhzady, S.; Ghasemi, V. Factors influencing g customers' satisfaction and dissatisfaction with hotels: A text-mining approach. *Tour. Anal.* **2019**, *24*, 69–79. [CrossRef]
35. Berezina, K.; Bilgihan, A.; Cobanoglu, C.; Okumus, F. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *J. Hosp. Mark. Manag.* **2016**, *25*, 1–24. [CrossRef]
36. Philander, K.; Zhong, Y. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *Int. J. Hosp. Manag.* **2016**, *55*, 16–24. [CrossRef]
37. Kweon, S.H. A semantic network analysis of 4th industry revolution: News frame for on drone, AI, big data and IoT news. *Int. J. Adv. Sci. Tech.* **2018**, *110*, 115–128. [CrossRef]
38. Shuting, T.; Kim, H.S. Cruising in Asia: What can we dig from online cruiser reviews to understand their experience and satisfaction. *Asia Pac. J. Tour. Res.* **2019**, *24*, 514–528.
39. Kim, Y.J.; Ban, H.J.; Kim, H.S. An exploratory study on the semantic network analysis of Busan tourism: Using Google web and news. *Culin. Sci. Hosp. Res.* **2019**, *25*, 126–134.
40. Lim, S.; Tucker, C.S.; Jablokow, K.; Pursel, B. A semantic network model for measuring engagement and performance in online learning platforms. *Comp. App. Eng. Edu.* **2018**, *26*, 1481–1492. [CrossRef]
41. Zhang, L.; Yun, H.J. Tourism information contents and text networking (focused on formal website of Jeju and Chinese personal blogs). *J. Korea Cont. Asso.* **2018**, *18*, 19–30.
42. Shuting, T.; Kim, H.S. A study of comparison between cruise tours in China and USA through big data analytics. *Culin. Sci. Hosp. Res.* **2017**, *23*, 1–10. [CrossRef]
43. Kim, H.S. A semantic network analysis of big data regarding food exhibition at convention center. *Culin. Sci. Hosp. Res.* **2017**, *23*, 257–270.
44. Bonacich, P. Some unique properties of eigenvector centrality. *Soc. Netw.* **2007**, *29*, 555–564.
45. Lee, Y.J.; Yoon, J.H. Exploring ways to utilize SNS big data in the tourism sector. *Int. J. Tour. Hosp. Res.* **2014**, *28*, 5–14.
46. Neter, J.; Wasserman, W.; Kutner, M.; Nachtsheim, C. *Applied Linear Statistical Models*, 4th ed.; McGraw-Hill: Chicago, IL, USA, 1996.
47. Freund, R.J.; Littell, R.C. *SAS System for Regression*, 3rd ed.; SAS: Cary, NC, USA, 2000.
48. Gennaioli, N.; Martin, A.; Rossi, S. Banks, government bonds, and default: What do the data say? *J. Monet. Econ.* **2018**, *98*, 98–113. [CrossRef]
49. Lee, J.H. Understanding customer values by analyzing the contents of online hotel reviews. *J. Korea Cont. Assoc.* **2013**, *13*, 533–546. [CrossRef]
50. Han, J.H.; Chung, N. Opportunities of big data applications to develop global marketing strategies: Analyzing hotel guest online reviews. *J. Tour. Lie Res.* **2017**, *7*, 233–251.

