Data Science in Finance and Economics

DSFE, 4 (1): 65–82.
DOI: 10.3934/DSFE.2024003
Received: 04 November 2023
Revised: 18 January 2024
Accepted: 19 January 2024
Published: 18 February 2024

https://www.aimspress.com/journal/dsfe

Research article

Exploring the influence of online word-of-mouth on hotel booking prices: insights from regression and ensemble-based machine learning methods

Sangjae Lee[1,*] and Joon Yeon Choeh[2]

[1] College of Business Administration, Sejong University, Seoul, 05006, Korea
[2] Department of Software, Sejong University, Seoul 05006, Korea

* **Correspondence:** Email: sangjae@sejong.ac.kr; Tel: +82-10-7925-7035; Fax: +82-2-3408-4310.

Abstract: Previous studies have extensively investigated the effects of online word-of-mouth (eWOM) factors such as volume and valence on product sales. However, studies of the effect of eWOM factors on product prices are lacking. It is necessary to examine how various eWOM factors can either explain or affect product prices. The objective of this study is to suggest explanatory and predictive analytics using a regression analysis and ensemble-based machine learning methods for eWOM factors and hotels booking prices. This study utilizes publicly available data from a hotel booking site to build a sample of eWOM factors. The final study sample was comprised of 927 hotels. The important eWOM factors found to affect hotel prices are the review depth and the review rating, which are moderated by a number of reviews to affect prices. The effect of the number of positive words is moderated by the review helpfulness to affect the price. The review depth and rating, along with the number of reviews, should be considered in the design of hotel services, as these provide the rationale for adjusting the prices of various aspects of hotel services. Furthermore, the comparison results when applying various ensemble-based machine learning methods to predict prices using eWOM factors based on a 46-fold cross-validation partition method indicated that ensemble methods (bagging and boosting) based on decision trees outperformed ensemble methods based on k-nearest neighbor methods and neural networks. This shows that bagging and boosting methods are effective ways to improve the prediction performance outcomes when using decision trees. The explanatory and predictive analytics using eWOM factors for hotel booking prices offers a better understanding in terms of how the accommodation prices of hotel services can be explained and predicted by eWOM factors.

Keywords: explanatory analytics; predictive analytics; hotel booking prices; online word of mouth (eWOM); determinants of hotel booking prices; ensemble based machine learning methods; ensemble methods based on a decision tree

JEL Codes: C12, C31, M15, M20

1. Introduction

From a consumer perspective, social media has improved price transparency to an unprecedented level (McGuire, 2016). Consumers depend on online reviews to obtain information about the quality of hotels and decide whether to make a booking (Liu and Park, 2015). Managers should consider appropriate pricing strategies in the dynamic condition of price transparency because users are able to try a series of alternative online bookings and instantly compare the prices provided by the alternative services. Additionally, online feedback on social media platforms represents a crucial factor that hotel managers should not neglect (Calheiros et al., 2017).

Hotel booking prices can be likely affected by online word-of-mouth (eWOM) factors, as tourists tend to access an online review platform to evaluate others' opinions before booking a hotel and acquire further information about the sentiments of online reviews. Online consumer reviews show a widely used form of eWOM, and information sources are becoming especially relevant in service contexts such as the hotel industry. The hospitality sector is among the sectors most influenced by eWOM factors (Cantallops and Salvi, 2014). For instance, online reviews have effects on hotel sales and hotel performance (Xie et al., 2014).

Furthermore, there are several studies which use data mining in tourism devoted to forecasting tourism demand. These include the data-mining applications in bank telemarketing (Moro et al., 2015) and social media (Moro et al., 2016). From a managerial perspective, forecasting tourism demand is one of the most prolific and interesting domains.

Our study intends to accomplish the following research objectives. First, the objective of this study is to suggest explanatory analytics using a regression analysis for eWOM factors and hotel booking prices. Among the various methods or measures devised to explain financial dealings such as price (Li et al., 2020; Liao et al., 2021), our study intends to use eWOM factors for hotel booking prices. Given the lack of studies regarding the effects of the volume and valence of eWOM on hotel booking prices, our study examines the effects of the number of reviews and the review ratings of various aspects of hotel services on hotel booking prices, which is based on studies of the effect of eWOM factors on product sales. While earlier work studied the impact of eWOM factors on profits and sales (Zhu and Zhang, 2010), the effect of eWOM factors on product prices has not received substantial research attention, especially in the hotel industry (Wu and Wu, 2016). The eWOM factors associated with the consumer sensitivity to price are regarded as an important research opportunity Cantallops and Salvi, 2014). Online information about a product enhances the price premium by the lowering price consciousness and improving trust (Huang et al., 2013).

Second, our study intends to apply ensemble-based machine learning methods to examine their performance for predicting hotel booking prices using the volume and valence of eWOM factors. While a number of studies have investigated the predictive performance of eWOM factors for predicting sales (Lee et al., 2020), there is a lack of studies that predict product prices, especially in the hotel industry. Given that several techniques are available for modeling prices and financial measures

(Chang et al., 2020; Corradin et al., 2022; Ren et al., 2022), we intend to apply the most sophisticated data-mining techniques.

2. The eWOM factors affecting hotel booking prices and hotel bookings

Price has been a crucial factor that determines a consumer's behavioral intentions (Moon et al., 2008). For example, Masiero and Nicolau (2012) showed that price sensitivity plays a significant role in affecting tourists' choices of tourism products.

The price differences between hotel booking prices result in various booking reconsiderations by consumers. It is necessary to determine the price variations based on the customers' reactions toward hotel services, which represents how consumers respond to price levels and changes of these levels. Moreover, with regard to hotel bookings, it is necessary to assess a customer's consideration of the price, which indicates the customer's evaluation of the value or utility of certain factors when choosing their accommodations, even to lower perceived risks, such as financial risk and performance risk (Liang et al., 2018). Based on the notion that hotel booking prices are associated with the value perceived by customers with regard to hotel services, our study focuses on hotel booking prices instead of hotel bookings as a dependent measure due to the importance attached to price during a customer's valuation of the overall utility of the hotel accommodation.

eWOM factors can be described as any positive or negative statement posted by consumers regarding a product or service, which can be a source of information as a basis for future purchase decision making, and are likely to be in line with the suggestions in online reviews (Filieri and McLeay, 2014).

Pricing comprises a crucial marketing strategy in the hospitality industry, thereby leveraging new paradigms of revenue management to incorporate a technological evolution (McGuire, 2016). Thus, it is supported by a number of online platforms, with hotels.com and TripAdvisor being examples of promising online media where users can post evaluations of hotels and compare other users' opinions (Ayeh et al., 2016). Such widespread online information can be appropriately utilized by managers to create pricing initiatives.

Previous studies show that the online reputation of a hotel can affect a traveler's hotel booking intentions (Hernandez-Maskivker et al., 2017) and hotel prices (Nieto-García et al., 2017). If hotels have positive eWOM reviews, their online bookings are less influenced by room prices and star ratings (Wang et al., 2015). Negative reviews exert a crucial effect on hotel booking intentions (Wen et al., 2021). El-Said (2020) showed that online reviews provide some moderation of the brand image, star category, and price, all of which affect booking intentions. Social media, web visibility, and online reservations are among the major factors that affect hotel prices (Moro et al., 2018). Tourists can utilize an online customer review platform to evaluate others' reviews before booking a hotel to acquire additional information about the positive or negative responses provided, thus further affecting the fluctuations of hotel booking prices.

Global digital players such as hotels.com and Expedia have appeared to dominate markets by providing digital hotel bookings, which expedite the purchasing of third-party hotel services by customers (Alt & Zimmermann, 2015). Within the hospitality industry, it is particularly relevant to design customer online review forums where customers can post their comments, score hotel units, and observe others' opinions (Ayeh, et al., 2016). Such information can be utilized by marketers to establish pricing strategies.

3. Research model

According to social exchange theory, social exchange is facilitated through social engagement via online reviews (Amblee et al., 2011). In relation to this, eWOM factors provide social influence as online customer reviews create distinct forms of media for social influence (Park and Lee, 2009, Lee and Choeh, 2020). Social influence refers to the process by which consumers alter their behaviors based on the responses of other consumers. Online review ratings can create either positive or negative reputations of products, as eWOM is an important market indicator of product reputation. eWOM offers signals of social proof, through which customers rely on the shared experiences of other hotel users to make their purchasing decisions (Premzaai et al., 2010). Thus, as structured eWOM factors in hotels indicate social capital and facilitate the delivery of product knowledge, eWOM can expedite changes in hotel booking prices. Our study posits three eWOM factors that affect hotel booking prices: the review volume (depth), valence (rating), and sentiment (the number of positive words). Furthermore, our study suggests the number of reviews and the degree of review helpfulness as moderators among the three eWOM factors and hotel booking prices. The research model is shown in Figure 1.

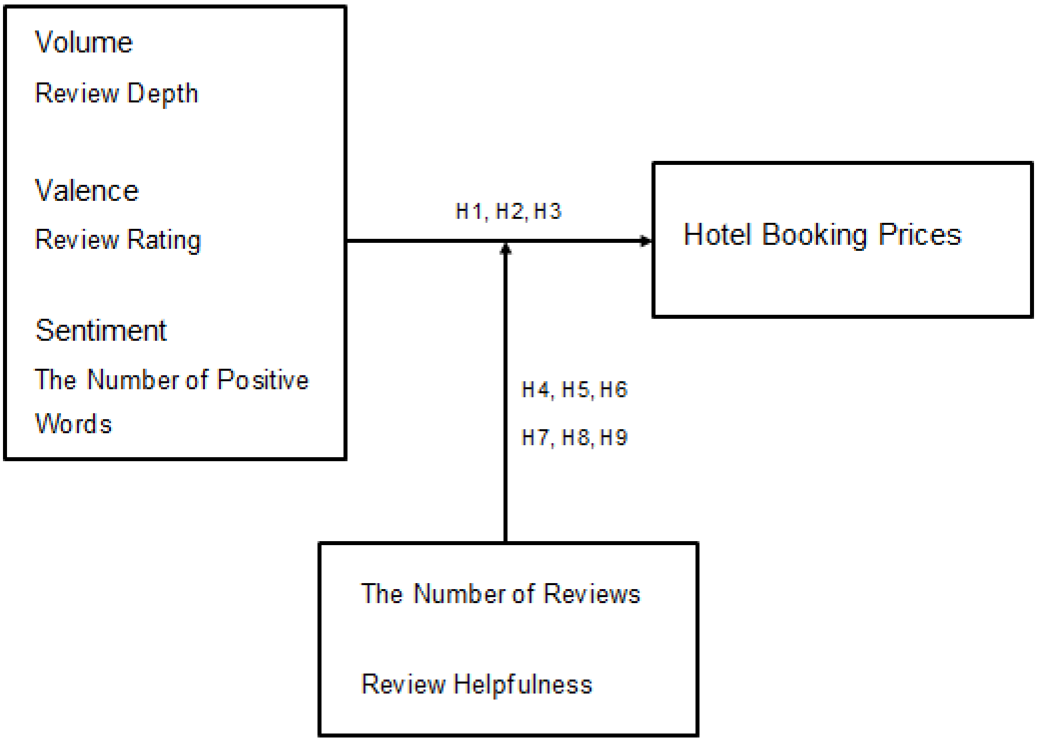

Figure 1. Research model.

Given that the perceived product value is influenced by eWOM factors providing informational cues regarding consumers' beliefs (Cheung et al., 2009) and price sensitivity (Liang et al., 2018), it can be assumed that eWOM has a positive effect on booking prices, which can represent a hotel's value. The eWOM volume depicts the consumer awareness of products, thus facilitating their subsequent purchase behaviors (Liu, 2006). As the eWOM volume is positively related to a product's sales (Cui et al., 2012), the eWOM volume levels are positively associated with the price level due to the eWOM awareness effect on the dissemination of hotel reputations.

Our study assumes that the eWOM volume is composed of two dimensions: the depth and the number of reviews. The eWOM depth, which greatly expedites the volume of reviews, represents the number of words in a review; the review volume normally exerts an important effect on the consumer's evaluations of the products they recommend (Duan et al., 2008). The volume of reviews can show the informative facet of social influence, as the acceptance of information from others (positive or negative) is evident by reality, thus reducing their sense of uncertainty (Park and Lee, 2009). This results in an increase in the consumer's behavioral intentions due to the informative nature of such reviews. The eWOM volume contains a credibility signal, especially for experience goods with a uncertain quality, which facilitates a consumer's intention to buy (Yang et al., 2012). More comments increase the consumer awareness effect for the product (Duan et al., 2008). There is a direct effect of the volume on hotel booking prices due to the monetary value of an increase in the eWOM volume (Jang et al., 2012). Thus, as the review volume can deliver a positive cue pertaining to the good quality of a hotel, the review volume can increase the hotel prices. Lee et al. (2020) also used the number of reviews with machine learning methods to predict box office outcomes. Considering its capacity to influence the perceived value, we propose that the review volume can have a positive effect on room booking prices.

Hypothesis 1: Review depth has a positive effect on hotel book prices.

Given that review valence is positively associated with either sales or consumers' purchase intentions, the seller's reputation is likely to promote consumer interest in bidding and increase the product's prices (Van Der Heide et al., 2013). Our study is based on the premise that consumers who've experienced high price bookings are those who recently used the services of high-quality accommodations, since price represents a quality preference indicator (Nieto-García et al., 2017). Previous studies have strongly supported the effect of the review valence and volume on hotel prices (Nieto-García et al., 2017). eWOM valence presents a signal that is crucial with regard to the pre-purchase evaluations of and preferences for a product or service (Chevalier and Mayzlin, 2006, Duan et al., 2008). The average rating of online reviews shows the average customer satisfaction with a product or service (Chintagunta et al., 2010). The rationale for the relationship between the review rating and price was suggested by Daniel et al. (2016), who posited an estimated increase in the hotel booking prices due to a one-star increase in the average online customer rating. Using a hedonic pricing model, Schamel (2012) suggested that popularity ratings as a crucial determinant of hotel room prices. Lee et al. (2020) also utilized review ratings to predict box office outcomes.

Thus, eWOM valence is a signal of what consumers will obtain to the extent that it may affect their perceived value, and therefore increase hotel booking prices. Given its ability to affect the perceived value, this study suggests that review ratings increase the booking prices for potential customers who have visited the review sites. Thus, the following hypothesis can be suggested.

Hypothesis 2: Review ratings have a positive effect on hotel booking prices.

Positive online reviews facilitate hotel consideration (Vermeulen and Seegers, 2009) and purchase intentions (Ladhari and Michaud, 2015). A positively worded message enhances consumer preferences for the product, whereas the proportion of negative online consumer reviews negatively affects consumer attitudes (Duan et al., 2008). Positive words contain information that creates a price premium for hotels with good reputations (Yacouel and Fleischer, 2012). Positive reviews increase price premiums, while negative ones decrease them (Houser and Wooders, 2006). Lee et al. (2020)

employed various eWOM factors, including review sentiment, to predict box office outcomes. Thus, the following hypothesis can be suggested.

Hypothesis 3: The number of positive words has a positive effect on hotel book prices.

The number of reviews can provide an informative facet of social influence, in that people consider information from others as evidence of reality. With a large number of available reviews, the behavioral intentions of consumers are enhanced. Because they consider those reviews as more informative and difficult to ignore, this indicates that a large number of reviews is a credibility cue (Nieto-García et al., 2017; Wu, 2017; Yang et al., 2012), thus possibly increasing a consumers willingness to buy (Grewal et al., 1994).

For hotels with a large number of reviews, the difference in hotel booking prices, as related to high or low eWOM valence and sentiment, could be greater than those for hotels with low review volumes. Thus, it can be suggested that there is a difference in consumer responses to the information provided by the review depth, valence and sentiment, depending on the number of reviews in relation to hotel booking prices for upcoming accommodations.

The eWOM volume can have a moderating effect on the relationship between the valence and hotel price by improving consumer persuasiveness, which is based on the impact of online reviews on the product choice (Khare et al., 2011). As the number of reviews can represent the eWOM diagnostic value, this factor can enhance the impact of eWOM on hotel booking prices because customers are likely to assign more credibility to the corresponding hotel's online reviews. Customers tend to obtain a more informative value from the corresponding hotel's eWOM factors when instances of eWOM are shown to them in a large quantity, thus resulting in a stronger association between other instances of eWOM and hotel booking prices; this subsequently supports their decision making regarding their hotel accommodation choice based on other eWOM factors and hotel booking prices data. Thus, the following hypotheses can be suggested.

Hypothesis 4: The number of reviews exerts a moderating effect on the relationship between the depth of reviews and hotel booking prices.

Hypothesis 5: The number of reviews exerts a moderating effect on the relationship between the review valence and hotel booking prices.

Hypothesis 6: The number of reviews exerts a moderating effect on the relationship between the number of positive words and hotel booking prices.

Based on the association between eWOM and review helpfulness (Siering et al., 2018; Wu, 2017), review helpfulness can also serve as a moderator between eWOM factors (volume, valence, and sentiment) and product sales (Lee and Choeh, 2016, 2023). Hotels tend to have higher prices with high review volumes, ratings, and sentiment levels, and moreover when the reviews for these hotels are helpful. The review volume, ratings, and positive sentiments are strongly associated with hotel prices when the average reviews for these hotels are comprehensive and helpful, thus indicating that helpfulness can moderate the effects of the review depth, valence, and sentiment on hotel booking prices.

Furthermore, as eWOM factors affect repurchase intentions and negatively effect the perceived risk (Liang et al., 2018), review helpfulness is one of the eWOM variables that can interact with other eWOM variables to affect box office revenue (Lee and Choeh, 2020). Review helpfulness can promote the informative value of a corresponding hotel's online reviews and can strengthen its impact on the relationship between eWOM factors and hotel booking prices; consequently, this supports the customer's decision making regarding their hotel accommodation based on eWOM factors and the hotel booking prices. Thus, we suggest the following hypotheses.

Hypothesis 7: Review helpfulness exerts a moderating effect on the relationship between the review depth and hotel booking prices.

Hypothesis 8: Review helpfulness exerts a moderating effect on the relationship between review valence and hotel booking prices.

Hypothesis 9: Review helpfulness exerts a moderating effect on the relationship between the number of positive words and hotel booking prices.

4. Methods

To investigate the influence of two specific eWOM factors, namely the volume and the valence, we crawled data from one of the most prominent tourism sites (https://www.booking.com). This study used publicly available data from booking.com to build measures of eWOM factors and booking prices. To investigate this, we collected data pertaining to economic transactions on booking.com and investigated the associated review system. The final sample consisted of 927 hotels. The dependent variable was the mean room price. The provision of mean prices for each hotel is necessary due to the real differences in prices across rooms. Table 1 shows the eWOM factors used in the study; they consist of the review volume, valence, and sentiment. All eWOM factors represent the average values for a specific hotel. For instance, the review volume is measured according to the average number of words in a review for a hotel. The review depth is the sum of positive and negative words in a review, as reviewers are instructed to post either positive or negative reviews in the positive or negative review box, respectively. Review helpfulness is calculated as the total number of helpfulness votes divided by the total number of reviews for a hotel. Table 2 shows the distribution of the sample. The major class of hotel is class 3 (62.4%). The average hotel booking price is \$ 96.97.

Table 1. eWOM used in this study.

Variables	Description	Number of possible values
Review volume (average number of words)	Shows the number of words in a review on average	Real values
Review valence (average review rating)	Shows the review rating on average	Real values
Review sentiment (the number of positive words)	Represents the average percentage of emotional reviews	Real values
Average number of reviews	Shows the number of reviews on average	Real values
Helpfulness of review	Shows the helpfulness of reviews for each hotel and is the number of helpfulness votes divided by the total number of reviews	Real values

Table 2. Distribution of the study sample.

Category		Number of hotels	Percent of hotels
Hotel class	3	96	10.3
	4	253	27.3
	5	579	62.4
		Mean	Standard deviation
Number of words in a review		15.08	6.15
Review rating		8.07	0.69
Number of positive words		8.03	3.49
Number of negative words		7.77	3.65
Number of reviews		493.68	808.12
Helpfulness		0.14	0.14
Rating of staff		8.45	0.75
Rating of comfort		8.13	0.87
Rating of location		8.32	0.82
Rating of facilities		8.01	0.84
Rating of cleanliness		8.12	0.90
Rating of value with respect to cost		8.01	0.82
Rating of WI-FI		8.10	1.78
Price (USD)		96.97	104.57
Days of accommodation		1.93	0.69

5. Results

Table 3. Regression analysis results for the impact of eWOM on price.

Terms	Standardized Coefficient	Standard error	t value	p-value	Hypothesis Test Results
Review depth	.400	1.82	3.74	.000	Hypothesis 1: accepted
Review rating	.248	7.93	4.72	.000	Hypothesis 2: accepted
Number of positive words	−.373	3.32	−3.37	.001	Hypothesis 3: rejected
Number of reviews	−2.817	.147	−2.49	.013	N/A
Review helpfulness	.015	191.24	.06	.954	N/A
Review depth X the number of reviews	1.228	.00	2.92	.004	Hypothesis 4: accepted
Review rating X the number of reviews	3.095	.02	2.79	.005	Hypothesis 5: accepted
Number of positive words X the number of reviews	−1.519	.01	−3.92	.000	Hypothesis 6: rejected
Review depth X review helpfulness	−.367	3.37	−2.23	.026	Hypothesis 7: rejected
Review rating X review helpfulness	−.068	20.05	−.33	.739	Hypothesis 8: rejected
Number of positive words X review helpfulness	.410	7.68	2.32	.021	Hypothesis 9: accepted

F= 6.871, p = 0.000

In order to test the research hypotheses regarding direct or moderation effects, we conducted a multiple regression analysis with interaction terms (Table 3). The significance of the coefficient estimates for these terms in this regression can be used to test our research hypotheses. Furthermore, our study utilized data-mining techniques to examine the predictive performance for hotel booking prices using eWOM factors. The effects of the review depth and review ratings on hotel booking prices are significant and positive, thus confirming hypotheses 1 and 2, which state that the volume and valence of eWOM are crucial factors that determine hotel booking prices, thus representing the value of service. The effects of the review depth, the review rating, and depth on prices are moderated by the number of reviews, which is another dimension of the eWOM volume, meaning that hypotheses 4 and 5 can be accepted. Given identical review depth and review rating levels, this shows that the number of reviews is a crucial positive weighting factor that affects price differences. The number of positive words is moderated by review helpfulness to have an effect on price, thus confirming hypothesis 9. Given that there is a certain level of positive review sentiment, this shows that review helpfulness provides an important weighting effect on the price.

Negative and significant direct effects are found regarding the number of reviews and the number of positive words (hypothesis 6). In addition, negative moderating effects exist between the review depth and review helpfulness (hypothesis 7). These negative direct and moderating effects indicate that these coefficients are negative and significant in the multiple regression analysis due to their reverse effects on price relative to the positive and significant factors. It can be assumed that these factors and their pairs operate to lower hotel booking prices. The ultimate price increase depends on how much the positive effects of the eWOM factors overwhelm the effects of the negative eWOM factors.

Previous studies show that data mining is a promising method in tourism and hospitality, especially when based on a big-data approach using manifold sources (Schuckert et al., 2015). This research adopts ensemble methods based on machine learning to predict hotel booking prices. Decision trees, k-nearest neighbors, and neural networks are used as machine learning methods. This study utilizes ensemble methods (i.e., bagging and boosting) applied to decision trees, k-nearest neighbors, and neural networks to improve the forecasting efficiency. The root mean squared errors of the validation sample for each subsample in a 45-fold cross-validation process were obtained and used to compare the prediction performance capabilities of the machine learning methods.

The decision tree method using ensemble methods was compared with the k-nearest neighbors and neural networks using the ensemble methods. In the ensemble methods, individual forecasts using three decision trees were combined and averaged, after which they were compared with the k-nearest neighbors and neural networks using the ensemble methods. The predicted values were averaged to produce a single prediction from the combined model. The steps used in this study are data collection, variable selection, and the application of predictive analytics based on decision trees, k-nearest neighbors, and neural networks using the ensemble methods to compare the prediction accuracy of each method.

The model performance assessment was conducted with these steps for data partition to produce prediction errors in the validation sample, where the study sample was divided into training and validation samples. The input variables are as follows: the average number reviews, the average review rating, the average review helpfulness, the average number of positive words, the average number of negative words, the average review depth, and the review volume (i.e., depth multiplied by the number of reviews).

While the training sample was used to estimate the parameters or weights in the suggested multiple models, the validation sample was adopted to show the predictive performance of each learned model from the training sample. The predictive models were composed of seven input variables and one output variable (i.e., price). The prediction errors in the validation sample were suggested in terms of the root mean square error (RMSE).

The tool utilized for the model training and testing step is the analytical solver platform, version 2022, which enables predictive analyses using different ensemble methods. Twenty hotels were partitioned as the validation sample, while the remaining 907 hotels served as the training sample. In addition, 46-fold cross-validation partition samples were used, for which the machine learning methods provided a prediction of the price, after which these predicted price values were compared with the actual price values in the validation sample. The validation samples in each of the 46-fold sample set produced 46 errors when using predictive analytics methods.

Table 4. T-test results for the difference in the RMSEs of predictive analytics.

RMSE difference for predictive analytics		Average difference	t value	# of test samples compared	p value
Decision trees with bagging	Neural networks with boosting	−49.27	−5.38	45	.000
Decision trees with bagging	Automated neural networks	−102.15	−13.07	45	.000
Decision trees with bagging	Neural networks with bagging	−136.44	−15.42	45	.000
Decision trees with bagging	Neural networks	−202.30	−18.97	45	.000
Decision trees with bagging	k-nearest neighbors with best k	−9.12	−3.94	45	.000
Decision trees with bagging	k-nearest neighbors with bagging	−23.19	−6.28	45	.000
Decision trees with bagging	k-nearest neighbors with boosting	−37.73	−6.99	45	.000
Decision trees with boosting	Neural networks with boosting	−45.39	−4.43	45	.000
Decision trees with boosting	Automated neural networks	−98.27	−12.29	45	.000
Decision trees with boosting	Neural networks with bagging	−132.56	−14.82	45	.000
Decision trees with boosting	Neural networks	−198.42	−18.54	45	.000
Decision trees with boosting	k-nearest neighbors with best k	−5.24	−1.52	45	.134
Decision trees with boosting	k-nearest neighbors with bagging	−19.30	−4.22	45	.000
Decision trees with boosting	k-nearest neighbors with boosting	−33.85	−5.65	45	.000
k-nearest neighbors with best k	k-nearest neighbors with boosting	−28.61	−7.45	45	.000
k-nearest neighbors with best k	Automated neural networks	−93.03	−10.89	45	.000
k-nearest neighbors with best k	Neural networks with bagging	−127.32	−13.43	45	.000
k-nearest neighbors with best k	Neural networks	−193.18	−17.08	45	.000
k-nearest neighbors with bagging	Neural networks with boosting	−26.08	−2.57	45	.013
k-nearest neighbors with bagging	Automated neural networks	−78.96	−8.81	45	.000
k-nearest neighbors with bagging	Neural networks with bagging	−113.25	−11.64	45	.000
k-nearest neighbors with bagging	Neural networks	−179.12	−15.46	45	.000
k-nearest neighbors with boosting	Neural networks with boosting	−11.54	−1.05	45	.301
k-nearest neighbors with boosting	Automated neural networks	−64.42	−7.05	45	.000
k-nearest neighbors with boosting	Neural networks with bagging	−98.71	−10.13	45	.000
k-nearest neighbors with boosting	Neural networks	−164.58	−14.22	45	.000
k-nearest neighbors	Neural networks with boosting	−9.70	−0.87	45	.388
k-nearest neighbors	Automated neural networks	−62.58	−6.81	45	.000
k-nearest neighbors	Neural networks with bagging	−96.87	−9.90	45	.000
k-nearest neighbors	Neural networks	−162.73	−14.03	45	.000

The t-test results for the comparison of the RMSEs between two different data-mining methods showed that decision trees with bagging methods are the best methods among the ensemble methods based on bagging and boosting based on decision trees, k-nearest neighbor methods, and neural networks (Table 4). Figure 2 shows the results of the comparison of RMSEs among the ensemble-based methods employed in this study. This shows that decision trees with bagging or boosting methods outperform the other methods.

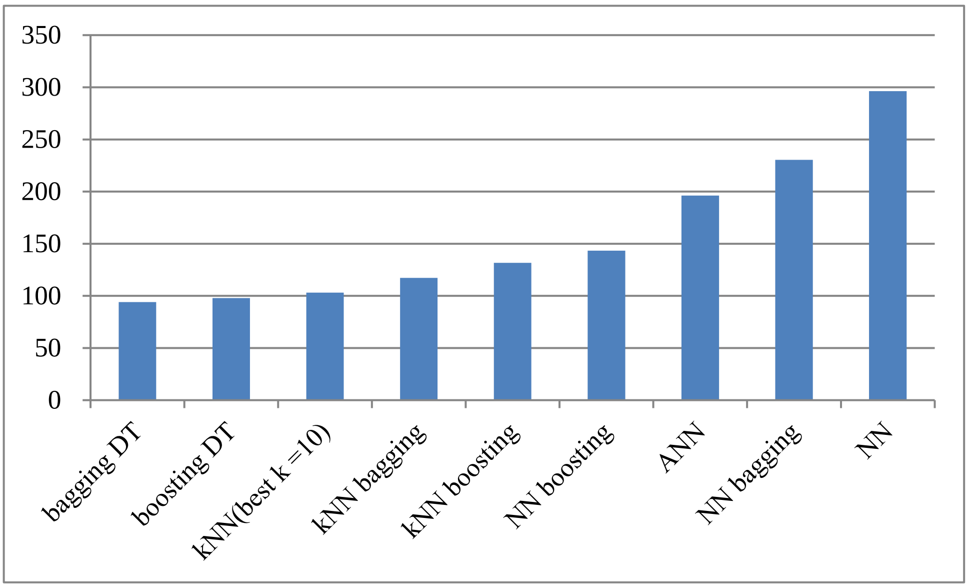

Figure 2. Comparison of RMSE among ensemble-based machine learning methods (Y axis is RMSE).

6. Discussion

Our study results indicate that the review depth and rating have a direct effect on hotel booking prices, thus indicating that price changes can be partially predicted by the average length of online reviews and how often customers provide a good evaluation of a hotel service. There were significant effects of eWOM factors on booking prices, which can be regarded as representing the perceived value of the hotel service. Given that the hotel booking prices greatly affect a consumer's purchase decisions (Mazumdar et al., 2005), especially in the hotel industry, which is characterized by price (Viglia et al., 2016), examining the effects of eWOM factors on hotel booking prices will crucially improve our understanding of the effects of eWOM on hotel performance outcomes.

Furthermore, a significant moderating relationship exists between the review depth and the number of reviews and between the review rating and the number of reviews; this shows that the number of reviews is a crucial weighting factor, which strengthens the effects of the review depth and rating on price. This supports earlier findings, such as those by Khare et al. (2011), Nieto-García et al. (2017), and Wu (2017), which collectively reviewed the volume representing both the informative value and a credibility cue. When eWOM factors are shown to customers in a large quantity, it can result in a stronger association between other eWOM factors and hotel booking prices, thus subsequently supporting the consumer's decision making regarding their hotel accommodation based on the other eWOM factors and hotel booking prices.

In a similar vein, the impact of the number of positive words on prices is affected by the review helpfulness, which shows that review helpfulness serves as a weighting factor for the effect of eWOM sentiment on price. This is in line with previous studies regarding the association between eWOM factors and the review helpfulness (Siering et al., 2018; Wu, 2017), and with the findings of

Lee and Choeh (2020, 2023), who suggested that the review helpfulness can promote the informative value of the corresponding hotel's online reviews. Our results show that the review helpfulness can strengthen the impact of the relationship between eWOM factors and hotel booking prices, thus supporting a customer's decision making regarding their hotel accommodation.

Cased on cross-validation partition, the comparison results when applying various ensemble-based machine learning methods to predict prices using eWOM factors show that ensemble methods (bagging and boosting) based on decision trees outperform ensemble methods based on k-nearest neighbor methods and neural networks. Given its novel approach to support a customer's decision making regarding hotel accommodations, our study can augment the literature on predictions using eWOM factors for product sales, thus presenting effective ensemble-based methods to predict hotel booking prices. This indicates that bagging and boosting methods are promising ways to improve prediction performance outcomes when using decision trees.

7. Conclusions

Our study suggests explanatory and predictive analytics using a regression analysis and ensemble-based machine learning methods for eWOM factors and hotel booking prices. This study utilized publicly available data from https://www.booking.com to build a sample of eWOM factors. The final study sample consisted of 927 hotels. The study comprised crucial eWOM determinants from previous studies of eWOM factors and online revenue. The three important variables are the review depth, reviewer ratings, and the number of positive words. The number of reviews and the degree of review helpfulness are moderators that affect the relationship between eWOM factors and the prices of hotel services. Based on a multivariate regression analysis with interaction terms, the important eWOM factors that affect hotel prices are the review depth and review rating, which are moderated by the number of reviews to affect the price. The effect of the number of positive words is moderated by the review helpfulness to affect the price. The review depth and rating factors, along with the number of reviews, should be considered when designing a hotel service, as these provide a rationale for adjusting the prices of various aspects of hotel services. The results of this study offer us a better understanding of accommodation prices of hotel services as affected by eWOM factors. Furthermore, the comparison results when applying various ensemble-based machine learning methods to predict prices using eWOM factors based on 46-fold cross-validation partition indicate that ensemble methods (bagging and boosting) based on decision trees outperform ensemble methods based on k-nearest neighbor methods and neural networks. This shows that bagging and boosting methods are effective ways to improve prediction performance outcomes when using decision trees. Our study can expand the literature on predictions using eWOM factors for hotel booking prices by introducing a novel approach to aid a customer's decision-making activities regarding their hotel accommodation, thus showcasing effective ensemble-based methods to predict hotel booking prices.

8. Implications for researchers

Our study builds on previous studies regarding eWOM factors and hotel booking prices by showing the moderating effects of the eWOM volume and helpfulness votes on the relationship between eWOM factors (i.e., valence and sentiment) and booking prices.

Regarding the eWOM volume and helpfulness votes, a large number of comments or helpful reviews are desirable for hotels with positive ratings and reviews that satisfy their clients' expectations. For such hotels, the positive effect of valence and sentiment can be enhanced by the volume and helpfulness of votes.

eWOM is defined as any words or discussions regarding certain goods, a service, or an enterprise, either positive or negative, and that are accessible by anyone online (Hennig-Thurau, Gwinner, Walsh, & Gremler, Citation2004). Litvin, Goldsmith, and Pan (Citation2008) adapted Westbrook's (Citation1987, p. 461) definition of WOM to the electronic world, since "all informal communications directed at consumers through Internet-based technology related to the usage or characteristics of particular goods and services, or their sellers." This study adopted the latter definition, referring to all informal communications by Airbnb consumers through the Internet as eWOM, which is related to the usage or characteristics of booking and living in Airbnb accommodations.

Our study can contribute to the research of the effects of eWOM factors on the booking prices of hotels. Given that eWOM is positively related to perceived value (Liang et al., 2018), it can have an effect on the accommodation prices. Our results are in line with previous studies that examined the effects of eWOM factors on the perceived value. For example, Gruen et al. (2006) found a direct positive influence of eWOM factors on the perceived value, and Cheung et al. (2009) posited that eWOM plays an informational and normative role by influencing a consumer's beliefs and conformity, thus affecting their perceived value. Since there was a lack of studies regarding the relationships among eWOM factors, the perceived value, and prices, our study provides insights to fill this gap.

Our study also has limitations that would benefit from future studies. First, although we developed the hypotheses regarding the impact of three eWOM factors on hotel booking prices, our study does not include hotel accommodation rates or sales in our model. There is a need to examine how hotel booking sales are affected by eWOM factors, and it is necessary to apply machine learning methods to predict hotel sales using eWOM factors in a future study. Second, our study is in the exploratory stage to set hotel bookings as a dependable variable of eWOM; in future work, it will be necessary to test other moderators based on hotel characteristics in the relationship between eWOM factors and hotel booking prices. Third, other adapted types of machine learning methods can also be employed to improve the prediction performance for hotel booking prices.

9. Implications for practitioners

Our results suggest that hospitality managers need to evaluate and adjust their prices according to the eWOM volume, valence, helpfulness, and the sentiments that online reviews provide about their accommodations. Online reputations should play an important role in price decisions (Abrate and Viglia, 2016). When eWOM factors for a hotel exist in a large quantity or are very helpful, it can result in a stronger association between other eWOM factors and hotel booking prices, thus subsequently supporting a marketing manager's decisions regarding their price strategy based on the other eWOM factors and hotel booking prices.

Therefore, hospitality operators should make an effort to meet their clients and promote their desire to post feedback, which is preferred to be helpful. Then, the resulting online reputation can facilitate a price increase, which will directly affect their net profit. Our study can provide a viable prediction approach that uses eWOM factors for hotel booking prices in the form of a novel ensemble-based machine learning method to support a marketing manager's decisions regarding their price strategy and hotel services.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5A2A03052219).

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflicts of Interest

The authors declare no conflicts of interest

References

Abrate G, Viglia G (2016) Strategic and tactical price decisions in hotel revenue management. *Tourism Manage* 55: 123–132. https://doi.org/10.1016/j.tourman.2016.02.006

Alt R, Zimmermann HD (2015) Editorial 25/3: Electronic markets on ecosystems and tourism. *Electron Mark* 25: 169–174. https://doi.org/10.1007/s12525-015-0197-7

Amblee N, Bui T (2011) Harnessing the influence of social proof in online shopping: The effect of electronic word of mouth on sales of digital microproducts. *Int J Electron Commer* 16: 91–114. https://doi.org/10.2753/JEC1086-4415160205

Ayeh JK, Au N, Law R (2016) Investigating cross-national heterogeneity in the adoption of online hotel reviews. *Int J Hosp Manag* 55: 142–153. https://doi.org/10.1016/j.ijhm.2016.04.003

Bezzubtseva A, Ignatov DI (2013) The Early Booking Effect and Other Determinants of Hotel Room Prices in Europe. *International Conference on Data Mining Workshops* Dollas. https://doi.org/10.1109/ICDMW.2013.32

Calheiros AC, Moro S, Rita P (2017) Sentiment classification of consumer-generated online reviews using topic modeling. *J Hosp Market Manag* 26: 675–693. https://doi.org/10.1080/19368623.2017.1310075

Cantallops AS, Salvi F (2014) New consumer behavior: a review of research on eWOM and hotels. *Int J Hosp Manag* 36: 41–51. https://doi.org/10.1016/j.ijhm.2013.08.007

Chang A, Yang L, Tsaih R, et al. (2022) Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. *Quant Financ Econ* 6: 303–325. https://doi.org/10.3934/QFE.2022013

Conroy SJ, Toma N, Gibson GP (2020) The effect of the Las Vegas Strip on hotel prices: A hedonic approach. *Tourism Econ* 26: 622–639. https://doi.org/10.1177/1354816619858259

Corradin F, Billio M, Casarin R (2022) Forecasting economic indicators with robust factor models. *Natl Account Rev* 4: 167–190. https://doi.org/10.3934/NAR.2022010

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: online book reviews. *J Market Res* 43: 345–354. https://doi.org/10.1509/jmkr.43.3.345

Cheung MY, Luo C, Sia CL, et al. (2009) Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *Int J Electron Commer* 13: 9–38. https://doi.org/10.2753/JEC1086-4415130402

Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. *Market Sci* 29: 944–957. https://doi.org/10.1287/mksc.1100.0572

Cui G, Lui HK, Guo X (2012) The effect of online consumer reviews on new product sales. *Int J Electron Commer* 17: 39–58. https://doi.org/10.2753/JEC1086-4415170102

Daniel K, Reiner J, Natter M, et al. (2016) Decomposing the effects of online customer reviews on brand, price, and product attributes. *Int J Res Market* 33: 11–26. https://doi.org/10.1016/j.ijresmar.2014.12.004

Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—an empirical investigation of panel data. *Decis Support Syst* 45: 1007–1016. https://doi.org/10.1016/j.dss.2008.04.001

El-Said OA (2020) Impact of online reviews on hotel booking intention: The moderating role of brand image, star category, and price. *Tourism Manag Perspect* 33: 100604. https://doi.org/10.1016/j.tmp.2019.100604

Filieri R, McLeay F (2014) E-WOM and accommodation an analysis of the factors that influence travelers' adoption of information from online reviews. *J Travel Res* 53: 44–57. https://doi.org/10.1177/0047287513481274

Ghose A, Ipeirotis PG (2010) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE T Knowl Data En* 23: 1498–1512. https://doi.org/10.1109/TKDE.2010.188

Gruen TW, Osmonbekov T, Czaplewski AJ (2006) EWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *J Bus Res* 59: 449–456. https://doi.org/10.1016/j.jbusres.2005.10.004

Houser D, Wooders J (2006) Reputation in auctions: theory, and evidence from eBay. *J Econ Manag Strateg* 15: 353–369. https://doi.org/10.1111/j.1530-9134.2006.00103.x

Huang M, Zhu H, Zhou X (2013) The effects of information provision and interactivity on e-tailer websites. *Online Inf Rev* 37: 927–946. https://doi.org/10.1108/OIR-07-2012-0117

Hung WT, Shang JK, Wang FC (2010) Pricing determinants in the hotel industry: Quantile regression analysis. *Int J Hosp Manag* 29: 378–384. https://doi.org/10.1016/j.ijhm.2009.09.001

Jang S, Prasad A, Ratchford BT (2012) How consumers use product reviews in the purchase decision process. *Market Lett* 23: 825–838. https://doi.org/10.1007/s11002-012-9191-4

Khare A, Labrecque LI, Asare AK (2011) The assimilative and contrastive effects of word-of-mouth volume: an experimental examination of online consumer ratings. *J Retailing* 87: 111–126. https://doi.org/10.1016/j.jretai.2011.01.005

Kim J, Jang S, Kang S, et al. (2020) Why are hotel room prices different? Exploring spatially varying relationships between room price and hotel attributes. *J Bus Res* 107: 118–129. https://doi.org/10.1016/j.jbusres.2018.09.006

Ladhari R, Michaud M (2015) eWOM effects on hotel booking intentions, attitudes, trust: and website perceptions. *Int J Hosp Manag* 46: 36–45. https://doi.org/10.1016/j.ijhm.2015.01.010

Lee S, Choeh JY (2016) The determinants of helpfulness of online reviews. *Behav Inform Technol* 35: 853–863. https://doi.org/10.1080/0144929X.2016.1173099

Lee S, Choeh JY (2020) Using the social influence of electronic word-of-mouth for predicting product sales: the moderating effect of review or reviewer helpfulness and product type. *Sustainability* 12: 7952. https://doi.org/10.3390/su12197952

Lee S, Bikash KC, Choeh JY (2020) Comparing performance of ensemble methods in predicting movie box office revenue. *Heliyon* 6: e04260. https://doi.org/10.1016/j.heliyon.2020.e04260

Li Z, Zhong J (2020) Impact of economic policy uncertainty shocks on China's financial conditions. *Financ Res Lett* 35: 101303. https://doi.org/10.1016/j.frl.2019.101303

Liang LJ, Choi HSC, Joppe M (2018) Understanding repurchase intention of Airbnb consumers: Perceived authenticity, electronic word-of-mouth, and price sensitivity. *J Travel Tour Mark* 35: 73–89. https://doi.org/10.1080/10548408.2016.1224750

Liao G, Hou P, Shen X, et al. (2021) The impact of economic policy uncertainty on stock returns: The role of corporate environmental responsibility engagement. *Int J Financ Econ* 26: 4386–4389. https://doi.org/10.1002/ijfe.2020

Liu Z, Park S (2015) What makes a useful online review?: Implication for travel product websites. *Tourism Manag* 47: 140–151. https://doi.org/10.1016/j.tourman.2014.09.020

Masiero L, Nicolau JL (2012) Price sensitivity to tourism activities: Looking for determinant factors. *Tourism Econ* 18: 675–689. https://doi.org/10.5367/te.2012.0143

Mathur S, Dewani PP (2016) Influence of cultural heritage on hotel prices, occupancy and profit: Theory and evidence. *Tourism Econ* 22: 1014–1032. https://doi.org/10.5367/te.2015.0487

McGuire KA (2016) *Hotel pricing in a social world–Driving value in the digital economy*, NewYork City: Wiley. https://doi.org/10.1002/9781119162285

Moon J, Chadee D, Tikoo S (2008) Culture, product type, and price influences on consumer purchase intention to buy personalized products online. *J Bus Res* 61: 31–39. https://doi.org/10.1016/j.jbusres.2006.05.012

Moro S, Cortez P, Rita P (2015) Using customer lifetime value and neural networks toimprove the prediction of bank deposit subscription in telemarketing campaigns. *Neural Comput Appl* 26: 131–139. https://doi.org/10.1007/s00521-014-1703-0

Moro S, Rita P (2016) Forecasting tomorrow's tourist. *Worldwide Hospitality and Tourism Themes* 8: 643–653. https://doi.org/10.1108/WHATT-09-2016-0046

Moro S, Rita P, Coelho J (2017) Stripping customers' feedback on hotels through datamining: The case of Las Vegas Strip. *Tourism Manag Perspect* 23: 41–52. https://doi.org/10.1016/j.tmp.2017.04.003

Moro S, Rita P, Oliveira C (2018) Factors influencing hotels' online prices. *J Hosp Mark Manag* 27: 443–464. https://doi.org/10.1080/19368623.2018.1395379

Nieto-García M, Muñoz-Gallego PA, Óscar González-Benito Ó (2017) Tourists' willingness to pay for an accommodation: The effect of eWOM and internal reference price. *Int J Hosp Manag* 62: 67–77. https://doi.org/10.1016/j.ijhm.2016.12.006

Park D, Lee J (2009) eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electron Commer Res Appl* 7: 386–398. https://doi.org/10.1016/j.elerap.2007.11.004

Premzaai K, Castaldo S, Grosso M, et al. (2010) Customer information sharing with e-vendors: The roles of incentives and trust. *Int J Electron Commer* 14: 63–74. https://doi.org/10.2753/JEC1086-4415140304

Ren X, Xu W, Duan K (2022) Fourier transform based LSTM stock prediction model under oil shocks. *Quant Financ Econ* 6: 342–358. https://doi.org/10.3934/QFE.2022015

Rezvani E, Rojas C (2020) Spatial price competition in the Manhattan hotel market: The role of location, quality, and online reputation. *Manag Decis Econ* 41: 49–63. https://doi.org/10.1002/mde.3092

Sanchez-Perez M, Illescas-Manzano MD, Martinez-Puertas S (2019) Modeling hotel room pricing: A multi-country analysis. *Int J Hosp Manag* 79: 89–99. https://doi.org/10.1016/j.ijhm.2018.12.014

Schamel G (2012) Weekend vs. midweek stays: modelling hotel room rates in a small market. *Int J Hosp Manag* 31: 1113–1118. https://doi.org/10.1016/j.ijhm.2012.01.008

Schuckert M, Liu X, Law R (2015) Hospitality and tourism online reviews: Recent trends andfuture directions. *J Travel Tourism Mark* 32: 608–621. https://doi.org/10.1080/10548408.2014.933154

Siering M, Muntermann J, Rajagopalan B (2018) Explaining and predicting online review helpfulness: the role of content and reviewer-related signals. *Decis Support Syst* 108: 1–12. https://doi.org/10.1016/j.dss.2018.01.004

Singh A (2017) The effects of securitization, foreclosure, and hotel characteristics on distressed hotel prices, resolution time, and recovery rate. *Cornell Hosp Q* 58: 39–52. https://doi.org/10.1177/1938965516639124

Van Der Heide B, Johnson BK, Vang MH (2013) The effects of product photographs and reputation systems on consumer behavior and product cost on eBay. *Comput Hum Behav* 29: 570–578. https://doi.org/10.1016/j.chb.2012.11.002

Vermeulen IE, Seegers D (2009) Tried and tested: the impact of online hotel reviews on consumer consideration. *Tourism Manag* 30: 123–127. https://doi.org/10.1016/j.tourman.2008.04.008

Viglia G, Mauri A, Carricano M (2016) The exploration of hotel reference prices under dynamic pricing scenarios and different forms of competition. *Int J Hosp Manag* 52: 46–55. https://doi.org/10.1016/j.ijhm.2015.09.010

Vives A, Jacob M (2020) Dynamic pricing for online hotel demand: The case of resort hotels in Majorca. *J Vacat Mark* 26(2): 268–283. https://doi.org/10.1177/1356766719867377

Vives A, Jacob M (2021) Dynamic pricing in different Spanish resort hotels. *Tourism Econ* 27: 398–411. https://doi.org/10.1177/1354816619870652

Wang MH, Lu Q, Chi RT, et al. (2015) How word-of-mouth moderates room price and hotel stars for online hotel booking: an empirical investigation with Expedia data. *J Electron Commer Res* 16: 72–80.

Wang XR, Sun JX, Wen HZ (2019) Tourism seasonality, online user rating and hotel price: A quantitative approach based on the hedonic price model. *Int J Hosp Manag* 79: 140–147. https://doi.org/10.1016/j.ijhm.2019.01.007

Wen J, Lin ZB, Liu X, et al. (2021) The interaction effects of online reviews, brand, and price on consumer hotel booking decision making. *J Travel Res* 60: 846–859. https://doi.org/10.1177/0047287520912330

Wu Y, Wu J (2016) The impact of user review volume on consumers' willingness-to-pay: a consumer uncertainty perspective. *J Interact Mark* 33: 43–56. https://doi.org/10.1016/j.intmar.2015.11.001

Xie KL, Zhang Z, Zhang Z (2014) The business value of online consumer reviews and management response to hotel performance. *Int J Hosp* 43: 1–12. https://doi.org/10.1016/j.ijhm.2014.07.007

Xue P, Jo W, Bonn MA (2021) Online hotel booking decisions based on price complexity, alternative attractiveness, and confusion. *J Hosp Tourism Manag* 45: 162–171. https://doi.org/10.1016/j.jhtm.2020.08.013

Yacouel N, Fleischer A (2012) The role of cybermediaries in reputation building and price premiums in the online hotel market. *J Travel Res* 51: 219–226. https://doi.org/10.1177/0047287511400611

Yang J, Kim W, Amblee N, et al. (2012) The heterogeneous effect of WOM on product sales: why the effect of WOM valence is mixed? *Eur J Mark* 46: 1523–1538. https://doi.org/10.1108/03090561211259961

Zhu F, Zhang X (2010) Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. *J Mark* 74: 133–148. https://doi.org/10.1509/jm.74.2.133