Archives of Business Research – Vol.6, No.4
Publication Date: Apr. 25, 2018
DOI: 10.14738/abr.64.4487.
Upadhyaya, N., Kenekar, A., Caicedo, M. A., & Berger, P. D. (2018). Determining The Key Factors That Influence Online Scores Given To Hotels On The Las Vegas Strip. *Archives of Business Research, 6(4),* 272-282.

Determining The Key Factors That Influence Online Scores Given To Hotels On The Las Vegas Strip

Nikita Upadhyaya
Bentley University
Waltham, MA 02452, U.S.A.

Apurva Kenekar
Bentley University
Waltham, MA 02452, U.S.A.

Maria Alejandra Caicedo
Bentley University
Waltham, MA 02452, U.S.A.

Paul D. Berger
Bentley University
Waltham, MA 02452, U.S.A.

ABSTRACT

The objective of this paper is to determine the relationship between the final scores (ratings) that customers give to hotels on the Las Vegas Strip and two sets of factors/independent variables: traveler-related attributes and hotel attributes. The data used in the paper has been retrieved by data mining from the TripAdvisor website. Excel and SPSS are used to conduct statistical analyses including descriptive statistics, correlation, and simple and stepwise [multiple] regression. The findings provide information which may be of help to hotels in considering improvement to their services and additions/changes to their facilities.

Keywords: Regression, Correlation, TripAdvisor, Las Vegas Strip

INTRODUCTION

Online reviews are being increasingly used as important decision-making factors when choosing hotels (Gretzel & Yoo, 2008). Websites, such as TripAdvisor, Expedia, etc., assist customers in making these decisions by allowing users to rate their experiences, and display the reviews and scores for everyone to see, registered users or non-registered users alike. When a hotel is reviewed, there are two main dimensions that are relevant - the characteristics of the traveler who is reviewing the hotel, and the facilities that the hotel offers and how well it offers them (Chu & Choi, 2000). The extent to which each dimension affects the overall score is a useful area on which to conduct research. The hotel industry is increasingly realizing the importance of these user ratings (Öğüt &Taş, 2012) and is investing resources in trying to improve these ratings. Hence, it would be of managerial significance to try to understand the main factors that affect the overall score/rating given to the hotels.

The new generation of millennials relies heavily on these ratings and also believes in contributing to these review portals to help others make more informed decisions (Mangold & Smith, 2012). Hence, it is imperative for the hoteliers to give due importance to these ratings and understand the driving forces behind a good and bad review. This study aims to shed light on these underlying factors based on data mined from the website, TripAdvisor.

TripAdvisor is a website which allows users to book travel and restaurant reservations, and also allows the users to provide ratings to review them. Information about registered users such as Name, Hometown, number of years since the user has been a member, number of places visited, and overall contribution of reviews, is also displayed. This information helps visitors to the website, who are reading these reviews, to also gauge the "qualifications" of those providing the reviews (Keates, 2007). For example, a visitor reading reviews might ascribe more importance to a review from a person who has been an active member of TripAdvisor for a long time than from a person who has recently joined TripAdvisor. Hence, this available information can also be taken into consideration while assessing the impact of a review being given to a hotel.

Las Vegas is a popular tourist destination attracting tourists from all over the world with different backgrounds and purposes of travel. "The Strip" is one of the most popular areas in Vegas where a lot of well known and well reviewed hotels are located. An interesting book about the history of hotels on the Las Vegas Strip is *Suburban Xanadu: The Casino Resort on the Las Vegas Strip and Beyond* by David Schwartz (2003). This particular area in Las Vegas has become a very frequent source of hotel reviews.

From secondary data which was mined by Moro, Rita, & Coelho (2017) from the site TripAdvisor, our study aims to provide research with the following objectives:

i) the effect of information about the traveler - i.e., who is rating the hotel - on the overall rating given to a hotel;
ii) the effect of facilities provided by the hotel - i.e., essentially "attributes" of the hotel - on the overall rating given to a hotel.

We first evaluate the extent to which the two blocks (traveler information and hotel facilities) impact the final score given to a hotel. When the independent variables from each of the two blocks are combined, the study aims to assess the extent of information provided by each variable, and the overall effect of all the selected variables, on the rating.

This study may be helpful to the hospitality industry to understand the key driving factors of the scores/ratings given to hotels on websites such as TripAdvisor. In addition to the role of facilities that a hotel provides, the management of a hotel can also understand how the characteristics and backgrounds of its guests influence the way they judge their experience at the hotel, and their final level of satisfaction with the hotel.

LITERATURE REVIEW

Studies of the impact of reviews on social media and websites such as TripAdvisor can be found (e.g., Molinillo, Ximenez-de-Sanoval, Fernandez-Morales, & Coca-Stefaniak, 2016). Generally, these are somewhat qualitative, with few advanced statistical analyses. Based on these studies, there have been changes in both, how travelers plan their trips - by reliance on reviews -and how the hotel industry manage their customers.

Qualitative studies have also been carried out to assess the *impact* of such online reviews, and how hotel managers can harness the wealth of information that lies in these reviews, to their advantage. A prominent example is: *The mechanism of online review management: A Qualitative study* (Ngyugen & Coudonris, 2015).

The impact of traveler information on scores given

Studies (e.g., Liu et al., 2015) have shown that when prospective travelers visit websites such as TripAdvisor, they ascribe higher importance to the reviews of experienced travelers than to

relatively new users. The source of this information about the amount of travel experience the reviewers have is from the number of years of active information that is publicly displayed by TripAdvisor to both registered users and non-registered users. There is also information about the total number of reviews a user has given overall (for hotels, restaurants, etc.), and the number of reviews given to hotels, specifically.

While studies have dealt with the influence that factors, such as total number of reviews given, have on the minds of people who are seeking information (e.g., Ye et al., 2011), little information exists on the effect these factors have on the *actual ratings given to hotels*. Would it be the case that the more experienced a traveler is, the more discerning their taste is considered to be, and thus, they would be less likely to give hotels higher scores? This is one of the issues that this study aims to explore.

The Impact of the type of facilities hotels provide on the score given to hotels

A hotel review and its overall score depends as much on the facilities provided by the hotels, as on the experience level of the traveler (Dolnicar & Otter, 2007). Whether a person is an experienced traveler or a new user, there are some basic facilities and services that all customers expect from the hospitality industry. A recent article from the Las Vegas Review Journal (Prince, 2017) states that thousands of rooms in all the big hotels at the Strip are getting upgraded. This renovation means thousands of dollars of investment; however, this type of investment is necessary if the hotels want its customer to stay and spend more while on the property.

The major challenges faced by these hotels is to keep a competitive advantage that can constantly lure its customers. With changing times, the aging baby boomers tend to spend less than the previous generation, and the new generation of millennials have a completely different set of expectations; thus, a lot of business and profit depends on the services and facilities provided by these hotels. What a customer expects and what he/she receives will have a big impact on the reviews and on the overall score/rating he/she gives to a particular hotel. Indeed, we explore the effect of all the variables associated with hotels on the final score provided on TripAdvisor.

METHODOLOGY

The data used for this paper come from a previous data mining study called "Stripping customers` feedback on hotel thorough data mining: the case of the Las Vegas Strip" by Moro, Rita, & Coelho (2017). Their work presents a data mining approach for modeling TripAdvisor scores using 504 reviews published in 2015 for the 21 hotels located on the Las Vegas Strip. The data for that study were collected from the TripAdvisor.com website. Las Vegas was selected because it is considered one of the most entertaining cities in the world, and is driven by tourism and gambling pressure (Rowley, 2015). The Las Vegas Strip is a very popular street, with many famous and large hotels; the study focuses on those 21 hotels. The data contained 504 records in total, 24 user reviews for each of the 21 hotels. However, there were missing values for two of the variables (Hotel's number of stars and Hotel's number of rooms) for four of the hotels, so those hotels were deleted from the final data set, leaving 408 records (unique users) in total. Each record contains 19 variables which are explained Table 1.

Table 1- List of Variables

Name	Description	Response
User_country	Which country is the user from	Actual country
Nr_reviews	How many reviews the user has given	Actual value
Nr_hotel_reviews	How many hotel reviews the user has provided	Actual value
Helpful_votes	How many of those reviews has been considered helpful for other users	Actual value
Score	The score that the customers gives to the Hotel	Actual value
Period _of_stay	Period of stay (Dec-Feb, Mar-May, Jun-Aug, Sep-Nov)	Actual time period
Traveler_type	What type of traveler is the user (Business, Couples, Families, Friends, Solo)	Actual category
Pool	If the hotel has outside pool	Yes/no
Gym	If the hotel has a gym	Yes/no
Tennis_court	If the hotel has a tennis court	Yes/no
Spa	If the hotel has a spa	Yes/no
Casino	If the hotel has a Casino inside	Yes/no
Free_internet	If the hotel provides free internet	Yes/no
Hotel_stars	Hotel's number of stars	Actual value
Nr_rooms	Hotel's number of rooms	Actual value
User_continent	Continent where the user`s country is located	Actual continent
Member_years	Number of years the user is been member of TripAdvisor	Actual value
Review_month	Month when the review was written	Actual month
Review_weekday	Day of the week when the review was written	Actual day

The main objective of this paper is to understand which of these variables are key in affecting the score that a user will give to a hotel. Thus, the dependent variable for the analyses is *Score* and the independent variables are: User Country, Number of reviews, Number of hotels reviews, Helpful votes, Period of stay, Traveler type, Pool, Gym, Tennis court, Spa, Casino, Free Internet, Hotels stars, Number of rooms, User continent, Member Years, Review Month and Review Week - the 19 variables listed in Table1.

First, we performed a descriptive analysis of the variables. Next, we considered a correlation analysis to explore if there is a significant correlation between certain variables. This correlation analysis was divided in two blocks. One included the variables that are related to user/traveler information and the other included the variables related to hotel facilities. Then, we performed a stepwise linear regression analysis.

The study first analyzes the data in two different blocks and then puts them together in a final model. One block considers the effect of traveler information variables on the rating given to hotels. The second block examines the effect of facilities provided by the hotel on the rating given to them. Finally, there is a final model that includes all the variables and eliminates redundancy between variables of different blocks.

ANALYSIS AND DISCUSSION OF RESULTS

Descriptive statistics

Table 2 shows the minimum, maximum, mean and standard deviation for the quantitative variables included in the data set.

Table 2: Descriptive Statistics- Quantitative Variables

	N	Minimum	Maximum	Mean	Std. Deviation
Nr_reviews	408	1	775	43.85	72.075
Nr_hotel_reviews	408	0	263	14.76	23.086
Helpful_votes	408	0	365	29.77	46.143
Score	408	1	5	4.08	1.02
Hotel_stars	408	3	5	4.24	0.808
Nr_rooms	408	315	4027	2540.53	1177.605
Member_years	408	0	13	4.36	2.942

For the Qualitative Variables:

The highest representation of the travelers is from North America (56.9%).
The distribution of period of travel and the month in which the travelers have submitted their reviews is almost equally distributed throughout the four quarters.

The major category of travelers are couples (43.9%).

94% of the hotels have pools and casinos while only 23% have tennis courts.
The variable Gym is a constant (and thus, deleted), since 100% of the hotels indicate that they have a gym.

Travelers show no particular preference for the day of the week in which the review is written.

Correlation Analysis

We use the strength scale suggested in Evans (1996), which is as follows:

Strength	r value
very weak	0 - less than .20
weak	0.20 - less than .40
moderate	0.40 - less than .60
strong	0.60 - less than .80
very strong	≥ .80

Table 3 provides the correlation matrix for block 1, the user-information variables. We include the dependent variable, *score* (row and column highlighted in the table.) Cells with correlations that are moderate and strong are also highlighted. (And, * = .01 < p-value ≤ .05, ** = p ≤ .01.)

Table 3: Correlations (Traveler Information Block)

	User_ country	Nr_ reviews	Nr_hotel _reviews	Helpful _votes	Score	Period_ of_stay	Traveler_type	User_continent	Member_years	Review_ month	Review_ weekday
User_ country	1	.137**	0.083	0.089	0.034	-.100*	-0.007	-.143**	0.075	-0.047	-0.058
Nr_ reviews	.137**	1	.594**	.771**	-0.025	-.112*	0.001	0.003	.106*	-.130**	0.006
Nr_hotel_ reviews	0.083	.594**	1	.731**	0.013	-0.003	-0.012	-0.032	.280**	-0.029	0.067
Helpful_votes	0.089	.771**	.731**	1	0.019	-0.091	-0.047	-0.021	.252**	-.110*	-0.006
Score	0.034	-0.025	0.013	0.019	1	0.002	-.101*	.100*	0.078	0.011	.104*
Period_of _stay	-.100*	-.112*	-0.003	-0.091	0.002	1	-0.037	-0.007	-0.082	.580**	0.072
Traveler_ type	-0.007	0.001	-0.012	-0.047	-.101*	-0.037	1	-0.021	-0.07	0.045	0.08
User_continent	-.143**	0.003	-0.032	-0.021	.100*	-0.007	-0.021	1	-0.018	-0.053	-0.009
Member_ years	0.075	.106*	.280**	.252**	0.078	-0.082	-0.07	-0.018	1	-0.008	0.025
Review_ month	-0.047	-.130**	-0.029	-.110*	0.011	.580**	0.045	-0.053	-0.008	1	0.076
Review_ weekday	-0.058	0.006	0.067	-0.006	.104*	0.072	0.08	-0.009	0.025	0.076	1

From Table 3 (again: for the Traveler block - i.e., all the variables related to the details of the traveler), we can note that we have two strong positive correlation: 1) Number of Reviews and the Number of helpful reviews 2) Number of hotel reviews and Number of helpful reviews. We also have two moderate (just short of "strong") correlations: 1) Number of reviews and Number of hotel reviews, 2) Review month and Period of stay. None of these results are particularly surprising. Also, *score* had a significant correlation (p < .05) with Traveler type, User continent, and Review weekday. (We have noted 4 correlations, even though there are 8 cells appropriately highlighted; the same is true for our next table. This, of course, is due to the correlation matrix being a symmetric matrix.)

Table 4 presents the correlation matrix for the hotel-information block.

Table 4: Correlations (Hotel Information Block)

	Score	Pool	Tennis_court	Spa	Casino	Free_internet	Hotel_stars	Nr_rooms
Score	1	.215**	0.068	.179**	0.031	.194**	.263**	-0.05
Pool	.215**	1	.139**	.685**	-0.063	-0.063	.383**	-.262**
Tennis_court	0.068	.139**	1	.203**	.139**	.139**	-.162**	-.282**
Spa	.179**	.685**	.203**	1	.685**	-0.091	.559**	.154**
Casino	0.031	-0.063	.139**	.685**	1	-0.063	.383**	.473**
Free_internet	.194**	-0.063	.139**	-0.091	-0.063	1	0.073	-.098*
Hotel_stars	.263**	.383**	-.162**	.559**	.383**	0.073	1	.270**
Nr_rooms	-0.05	-.262**	-.282**	.154**	.473**	-.098*	.270**	1

From Table 4 (again: for the Hotel block - i.e., all the variables related to the attributes of the hotel), we can note that we (again) have two strong positive correlation: 1) Having a spa and Having a pool, 2) Having a spa and Having a casino. We also (again) have two moderate correlations: 1) Having a spa and How many stars the hotel is assigned, 2) Number of rooms and Having a casino. None of these results are particularly surprising, although it may not have been obvious that having a spa plays such a role in how many stars the hotel is assigned. Also, *score* had a significant positive correlation ($p < .05$) with Having a pool, Having a spa, Having free internet, and How many stars the hotel is assigned. None of these results are especially surprising, but it is noteworthy that these 4 variables were significant, while variables such as Having a casino, Having a tennis court, and Number of rooms are not significantly correlated with *score*.

Regression Analysis

Our data have been used to understand how the dependent variable (Score) is related to all the independent variables. We perform a stepwise [multiple] regression. The variables that are a part of the final model are Hotel stars, Free internet, Pool, Review weekday, Traveler type and User continent. The details of the model summary is shown in Fig. 1 below.

Model	R	R square	Adjusted R Square	Std. Error of the Estimate
1	0.263	0.069	0.067	0.985
2	0.316	0.1	0.095	0.97
3	0.346	0.12	0.113	0.96
4	0.363	0.132	0.123	0.955
5	0.379	0.144	0.133	0.949
6	0.391	0.153	0.14	0.945

Predictors: (Constant), Hotel_stars

Predictors: (Constant), Hotel_stars, Free_internet

Predictors: (Constant), Hotel_stars, Free_internet, Pool

Predictors: (Constant), Hotel_stars, Free_internet, Pool, Review_weekday

Predictors: (Constant), Hotel_stars, Free_internet, Pool, Review_weekday, Traveler_type

Predictors:(Constant), Hotel_stars, Free_internet, Pool, Review_weekday, Traveler_type, User_continent

Figure 1: Stepwise models

The coefficients in the final model are depicted in Figure 2:

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**
		B	**Std. Error**	**Beta**		
6	(Constant)	1.268	0.408		3.105	0.002
	Hotel_stars	0.255	0.063	0.202	4.025	0
	Free_internet	0.777	0.202	0.18	3.85	0
	Pool	0.648	0.217	0.15	2.986	0.003
	Review_weekday	0.06	0.023	0.119	2.567	0.011
	Traveler_type	-0.089	0.037	-0.109	-2.362	0.019
	User_continent	0.115	0.055	0.096	2.073	0.039
Dependent Variable: **Score**						

Figure 2: Coefficient table for stepwise regression

The final equation of the model is:

Predicted Score = 1.268 + 0.255*Hotel Stars + 0.777*Free_internet + 0.648*Pool
+ 0.06*Review_Weekday -0.089*Traveler_type + 0.115*User_continent

We can see from the final model equation that Predicted Score goes up as the number of stars goes up, if the hotel has free internet, and if it has a pool. It also varied by the day of the week, by the traveler type, and by the continent of the traveler; these latter three variables are, of course, not under the control of the hotel management

The other eleven[1] independent variables did not enter the model. These eleven independent variables are User Country, Number of reviews, Number of hotel reviews, Helpful votes, Period of stay, Tennis Court, Spa, Casino, Number of rooms, Member for number of years and Review month. A simple linear-regression was run between the dependent variable and each of these eleven variables to check if the variables were excluded from the final stepwise- regression model due to redundancy or due to irrelevancy. The simple regressions showed that none of these eleven variables have a statistically-significant relationship (all had $p > .05$), with the dependent variable, essentially eliminating redundancy as a reason for any of the 11 variables to be excluded from the final stepwise-regression model. Among these 11 excluded variables, the one with the lowest p-value was Having a tennis court ($p = 0.17$).

LIMITATION AND DIRECTION OF FUTURE RESEARCH

User continent is one of the variables in the final model of the stepwise regression. On further analysis, it was found that 232 Scores out of 408 were given by users in North America, about

[1] Recall that there were originally 18 independent variables; however, the variable, "Having a gym," was constant, and, hence, dropped from the analysis; thus, we now have 17 variables that were eligible to enter the stepwise regression.

58%. One could argue that the respondents are not evenly distributed across the world's travelers. Figure 3 displays respondents by continent.

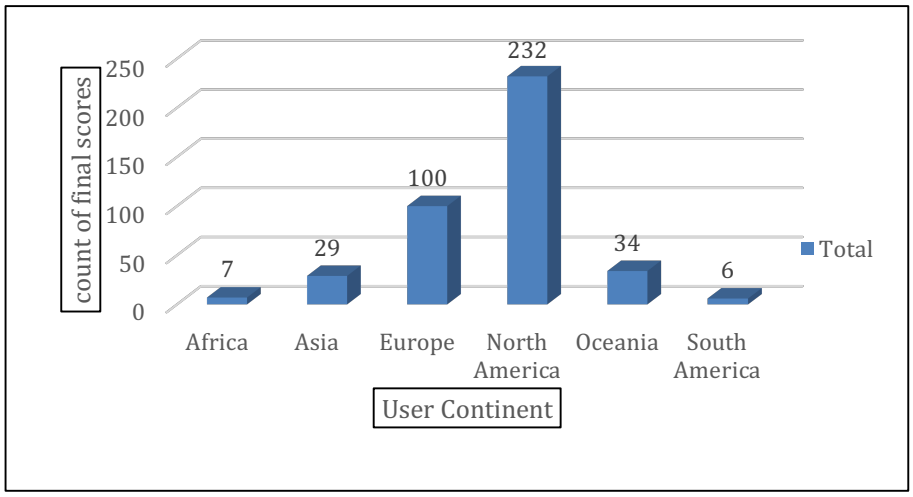

Figure 3: Respondents by continent

One might want to do a follow up study that is not so heavily weighted toward North American respondents.

No dummy variables were created in this study because we believed that there were limitations due to over- representation and under-representation various categories of certain variables. Also, using dummy variables for all the levels of the categorical variables might have adversely affected the power of the results, by limiting the [error] degrees of freedom. However, this renders the exact interpretation of the impact of Day of the week of the review, Traveler type, and Continent of reviewer somewhat unclear. In retrospect, the inclusion of dummy variables would likely have been a superior choice, even with the concomitant downsides. Still, these variables did show up as affecting Score, as a function of the independent variable value.

CONCLUSION

From this study, it can be concluded that the variables, such as the number of stars of a hotel, whether the hotel provides free internet, whether it is equipped with a pool, the day of the week on which the review is submitted online, the reason for travel, and the continent from which the traveler belongs, are the most important variables in predicting the Score that a traveler gives the hotel on a site such as TripAdvisor.

The managerial use of this model could be for the hotel officials to predict the Scores or the satisfaction levels of their guests based on the amenities that they provide plus a few details of the travelers. Management can, ideally, use this information for producing targeted marketing-communication and to develop superior marketing strategies. Sites such as TripAdvisor can also make use of this information for predictions as to how their users will score hotels and can also make future recommendations on the basis of these variables.

While the study by Moro et al. (the source of our data) comes to the conclusion that the variables related to the traveler have more relevance than the variables related to the hotel, itself; this paper differs somewhat in its findings. Out of the six variables that are a part of the stepwise regression final model, three are related to the traveler and three are related to the hotel. In fact, the first three variables that entered the stepwise regression were hotel-related. Still, we must admit that this may have been influenced by our lack of dummy variable

construction for many of the traveler-related variables.

References

Chu R., and T. Choi, T. (2000), "An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: A comparison of business and leisure travelers," *Tourism Management*, 21(4), p. 363-377.

Dolnicar, S., and Otter, T (2003), Which Hotel Attributes Matter? A Review of Previous and a Framework for Future Research, in Griffin, T & Harris, R (eds.) *Proceedings of the 9th Annual Conference of the Asia Pacific Tourism Association (APTA)*, University of Technology Sydney, 1, 176-188.

Gretzel U., and Yoo, K. (2008), "Use and Impact of Online Travel Reviews," In: O'Connor P., Höpken W., Gretzel U. (eds) *Information and Communication Technologies in Tourism 2008*, Springer, Vienna

Keates, N. (2007), "Deconstructing TripAdvisor," *Weekend Journal*, June 1, p.4.

Liu, Z. , Le Calvé, A. , Cretton, F. , Glassey Balet, N. , Sokhn, M. and Délétroz, N. (2015), "Linked Data Based Framework for Tourism Decision Support System: Case Study of Chinese Tourists in Switzerland," *Journal of Computer and Communications*, 3, 118-126

Mangold, G. and Smith, K., (2012), "Selling to Millennials with Online Reviews (2012)," *Business Horizons*, 55(9), 141-153.

Molinillo, S., Ximénez-de-Sandoval, J., Fernández-Morales, A. and Coca-Stefaniak, A. (2016), "Hotel Assessment through Social Media: The case of Trip Advisor," *Tourism & Management Studies*, 12(1), p. 15-24.

Moro, S., Rita, P., and Coelho, J. (2017), "Stripping customers' feedback on hotels through data mining: The case of Las Vegas Strip," *Tourism Management Perspectives*, 23, 41-52.

Nguyen, K and Coudounaris, D. (2015), "The mechanism of online review management: A qualitative study," *Tourism Management Perspectives*, 16, 163-175.

Öğüt, H., Taş, B. 2012. "The Influence of Internet Customer Reviews on the Online Sales and Prices in Hotel Industry," *Service Industries Journal* 32(2), 197–214.

Prince, T. (2017), "Thou**s**ands of Las Vegas Strip hotel rooms getting major upgrades in Las Vegas," *Las Vegas Review-Journal*. Retrieved from https://www.reviewjournal.com/business/casinos-gaming/thousands-of-las-vegas-strip-hotel-rooms-getting-major-upgrades-photos/

Schwartz, D. (2003), "Suburban Xanadu: The Casino Resort on the Las Vegas Strip and Beyond," Routledge, New York.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011), The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," Computers in Human Behavior, 27(2), 634–639.