Do Online Ratings and Reviews Have Influence on Hotel Prices in Armenia?

Hermine Grigoryan

American University of Armenia

Introduction to Econometrics

Empirical Project

April 2018

Table of Contents

1. Abstract 3
2. Introduction 4
3. Literature Review 4
4. Theoretical Framework 6
5. Empirical Analysis 7
   5.1. Data 7
   5.2. Descriptive Analysis 9
   5.3. Regression Results 12
6. Conclusion 15
7. Appendices 17
   7.1. Data 17
   7.2. Graphs 18
8. References 22

1. Abstract

During the travel and tourism decision making process, online reviews and ratings are used as an important source of information by hotel guests. The main objective of this research is to analyze how online ratings and reviews affect the prices of hotels in Armenia. Data from TripAdvisor, an open online community, on 462 Armenian hotels were scraped to use for the analysis.

The research has shown that the number of reviews have a positive impact on the price. However, the impact is very small as if the number of reviews of a hotel increases by 100, then the price is predicted to increase by only 8.5%. Meanwhile, the online rating plays a big role in shaping hotel prices. If the hotel has a score 4 and higher, in a 5-point scale, then its price is more by 10% compared to those hotels that have a lower score. However, the findings have shown that hotel prices are not only determined by hotel ratings and reviews, but other factors affect hotel rates as well.

Keywords: online hotel ratings, hotel prices, Armenian hotels, TripAdvisor

2. Introduction

Many studies have revealed that social media has an increasingly important role on consumers' purchasing behavior and influences their purchasing patterns. Sparks, Perkins and Buckley (2013) in their paper have shown that consumers, such as tourists searching for accommodations, tend to trust online ratings and comments by other customers more than recommendations provided by official hotel websites. As it turns out, online reviews and ratings have high impact on the bottom line - especially in an industry as competitive as hotel and hospitality. According to UNWTO (2015, p. 6) "Before making an online hotel reservation, consumers visit on average almost 14 different travel-related sites with about three visits per site, and carry out nine travel-related searches on search engines".

The objective of this paper is to examine whether there is a causal link between online ratings, reviews and the prices of hotels in Armenian cities; or the prices for accommodations are not customer driven, meaning that online reviews by customers do not play any role in shaping prices.

This paper is organized as follows. Section 3 summarizes and describes papers and findings related to hotel industry done by other researches. In section 4 I discuss theoretical framework and approaches that I have employed in the paper. The "Empirical Analysis" section is comprised of three parts: data, descriptive analysis and regression results. Finally, the last section summarizes the main results along with concluding remarks.

3. Literature Review

Previous research has demonstrated that online ratings and reviews are not an effective source of information for consumers. First, online ratings do not represent customers' real preferences. Second, reviewers are not a randomly drawn sample from the population as extremely satisfied and extremely dissatisfied customers are more

likely to review a hotel. Moreover, empirical studies have proven that prior ratings are a source of bias for future ratings based on positive social influence (Aral & Walker, 2012, p. 337). In addition, Hu, Bose, Koh, and Liu (2012) have shown in their studies that interested parties can easily manipulate online forums by posting anonymous reviews.

However, none of these arguments discards the fact that customers rely on others' ratings and reviews. On the contrary, the trust towards online ratings and reviews has been increasing during the recent years, which in its turn, shapes the purchasing decisions of customers.

This section of the paper mainly focuses on 3 research papers done by Chris Anderson from Cornell University, Gregory Lewis (Microsoft Research) and Georgios Zervas (Boston University), and Conceição Castro and Fernanda A. Ferreira from Polytechnic Institute of Porto.

The Cornell University Study of 2011 called "The Impact of Social Media on Lodging Performance" has revealed the following findings:

- The percentage of consumers who consult online bookings, such as TripAdvisor and Booking, has steadily increased over time, as has the number of reviews that people are reading prior to booking a hotel room.
- If a hotel increases its online rating by 1 point on a 5-point scale, it is able to increase the price for hotel rooms by approximately 11.2%.
- 1% increase in a hotel's online reputation score is predicted to increase the hotel price by 0.89%. Likewise, this 1% increase leads to an occupancy increase of up to 0.54%. Moreover, one percent increase in the reputation score causes to up to a 1.42 percent increase per available room. (Cornell University, 2012)

Another study by Gregory Lewis of Microsoft Research and Georgios Zervas from Boston University's Questrom School of Business explores further the relationship between online reviews and hotel prices. The researchers used panel data of hotel prices from three review platforms: TripAdvisor, Expedia and Hotels.com, sales and reviews

from five US states over a 10-year period from 2005-2014 for the analysis. The study suggests that better reviews are not only correlated with higher demand (higher prices) for accommodations, but there is a causal link between the ratings and prices. According to the study, each additional "online star" increased demand by 25 percent which then led to a 9 percent increase in prices.

Another study related to the relationship between online reviews and hotel prices in Lisbon is a research conducted by Conceição Castro and Fernanda A. Ferreira. The results of the analysis suggest that star rating, convenient location and facilities provided have a positive impact on hotel room rates, whereas the number of comments affects hotel prices negatively. The impact of quality signaling factors, such as cleanliness, location, etc., differs among hotels with different categories.

4. Theoretical Framework

The aim of this research is to understand how online ratings and reviews affect prices of accommodations. In this framework, our dependent variable is price (USD). Research shows that it is common to use semilog (or log-linear) specification for the pricing function instead of the linear specification. Therefore, the model will be specified with the following formula:

$$\ln(price) = \beta_0 + \beta_i \sum_{i=1}^{n} x_i + u\,,$$

where x_i is the vector of variables, such as number of reviews, average score, free Wi-Fi, free parking, location of the hotel (Armenian provinces) and hotel category.

The initial hypothesis is such that all of the variables mentioned above will have a positive impact on the price. However, as regression results will show, location of a hotel has a negative impact on price, meaning that on average, hotels in Yerevan are cheaper than in other provinces holding all other variables constant. Moreover, it turned out free Wi-Fi, free parking and average score (provided by customers) are not significant

predictors of price. That is why average score will be specified as a categorical variable in the model with 3 categories: low score, medium score and high score.

In addition, I believe that the number of reviews will depend on the current score of a hotel, and vice versa, meaning that the average score of a hotel will depend on the current level of reviews. That is why I introduce an interaction term between variables "High score" and "Number of reviews". Another interaction term is introduced between two dummy variables: "Hotels" category and "Yerevan". I believe that the impact of being located in Yerevan is different for different hotel categories.

5. Empirical Analysis

5.1. Data

The data for the analysis was scraped from TripAdvisor, which is an open online community where businesses, such as restaurants and hotels, are listed and reviewed by millions of users all around the world. Statistical tools, such as Python and R were used for scraping and providing tables and graphs. Moreover, some of the variables used in the analysis were scraped from TripAdvisor by data scientist of UNDP Armenia.

For the analysis, the data was cleaned, missing observations were removed or replaced with appropriate values and hotels with average rating (given by reviewers) of 0 were removed as these hotels were not reviewed at all. After cleaning the scraped data set, overall, 462 hotels were obtained with the following features:

- *Price of the hotel (USD)* - Prices are the average nightly price for 1 room (2 adults) of each hotel.
- *Yerevan* – This dummy variable indicates if the hotel is located in Yerevan or not. It was obtained from a variable called "State" with a condition such that if the state is Yerevan, the label is 1, and 0 otherwise.
- *Number of reviews* – This variable shows the number of reviews that a hotel has.

- *Number of photos* – number of photos for a particular hotel submitted either by travelers, or hotel management. This variable is highly correlated with number of reviews (0.78 Spearman correlation coefficient).
- *Category* – This is a categorical variable with 3 unique categories identifying to which category an accommodation belongs to: B&B (Bread and Breakfast), Hotels and Specialty lodging. For the purpose of regression, each category was considered as a dummy variable.
- *Average score* – This variable is obtained by calculating the weighted average of 5 review scores: Number of Terrible ratings (weight=1), Number of Poor ratings (weight=2), Number of Average ratings (weight=3), Number of Good ratings (weight=4) and Number of Excellent ratings (weight=5). The average score will be a number with values ranging from 1 to 5. In order to include this variable into the model, I divide them into categories: High score (4-5), Medium score (2-4), Low score (1-2).
- *Free WiFi* – A dummy variable indicating if the hotel provides free WiFi.
- *Free parking* - A dummy variable indicating if the hotel provides free parking.

A snapshot of several rows for the data set is provided in Appendix 7.1: Data.

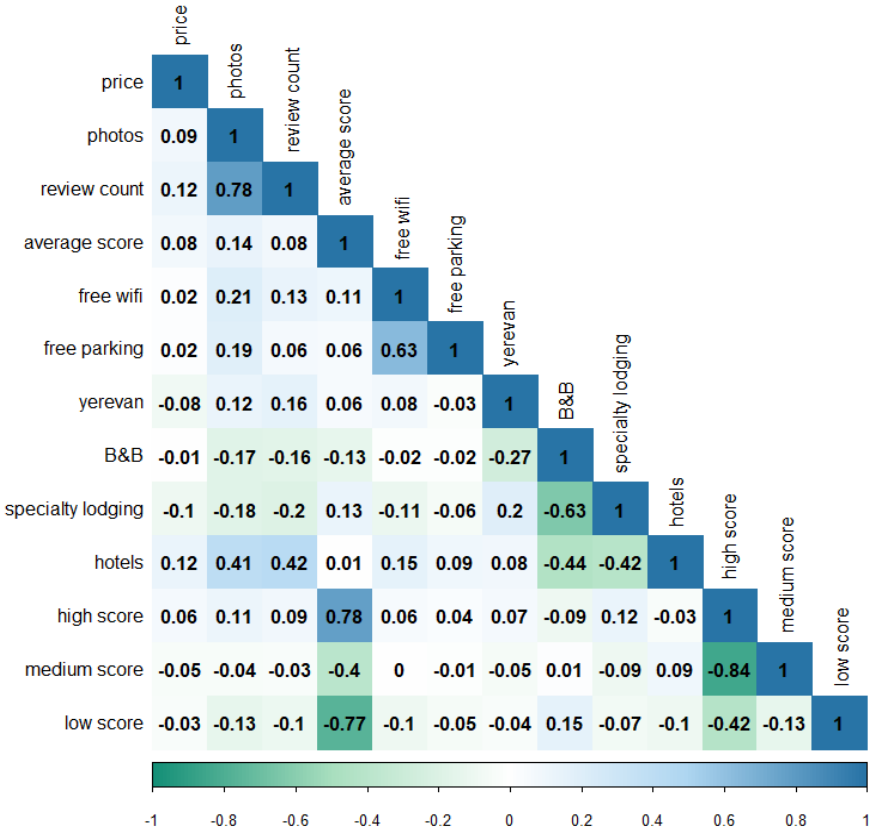
5.2. Descriptive Analysis

Table 1: Summary Statistics

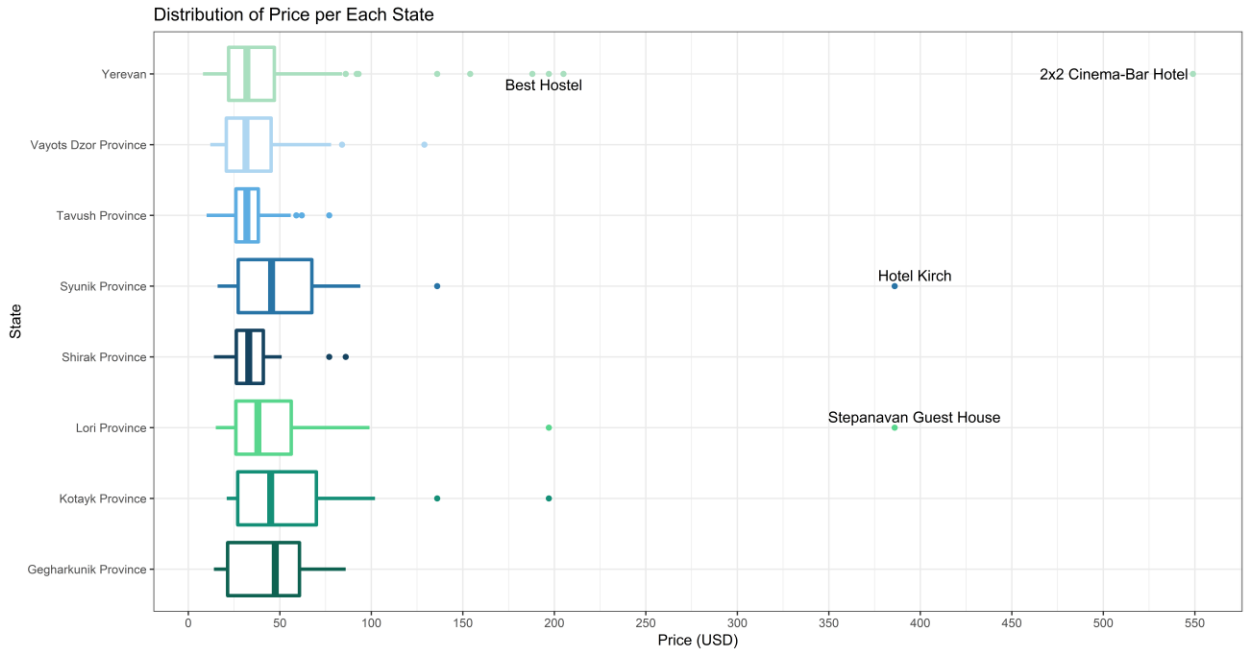
	High score (n=336)	Low score (n=29)	Medium score (n=97)	Overall (n=462)
Price (USD)				
Mean (SD)	45.8 (46.8)	38.9 (23.3)	40.1 (30.9)	***44.2 (42.8)***
Median [Min, Max]	36.0 [8.00, 549]	30.0 [15.0, 129]	34.0 [12.0, 205]	***35.0 [8.00, 549]***
Number of Reviews				
Mean (SD)	38.8 (90.4)	2.93 (2.93)	29.3 (57.0)	***34.6 (81.9)***
Median [Min, Max]	9.00 [1.00, 794]	2.00 [1.00, 11.0]	8.00 [1.00, 462]	***7.00 [1.00, 794]***
Number of Photos				
Mean (SD)	47.8 (76.6)	6.10 (9.64)	36.9 (54.0)	***42.9 (70.6)***
Median [Min, Max]	21.5 [0.00, 596]	0.00 [0.00, 30.0]	17.0 [0.00, 327]	***18.0 [0.00, 596]***
Average Score				
Mean (SD)	4.61 (0.367)	1.35 (0.467)	3.42 (0.401)	***4.15 (0.949)***
Median [Min, Max]	4.65 [4.00, 5.00]	1.00 [1.00, 2.00]	3.50 [2.50, 3.96]	***4.37 [1.00, 5.00]***
State				
Gegharkunik Province	6 (1.8%)	4 (13.8%)	6 (6.2%)	***16 (3.5%)***
Kotayk Province	29 (8.6%)	3 (10.3%)	7 (7.2%)	***39 (8.4%)***
Lori Province	23 (6.8%)	2 (6.9%)	7 (7.2%)	***32 (6.9%)***
Shirak Province	16 (4.8%)	0 (0%)	6 (6.2%)	***22 (4.8%)***
Syunik Province	31 (9.2%)	2 (6.9%)	13 (13.4%)	***46 (10.0%)***
Tavush Province	20 (6.0%)	3 (10.3%)	5 (5.2%)	***28 (6.1%)***
Vayots Dzor Province	17 (5.1%)	1 (3.4%)	4 (4.1%)	***22 (4.8%)***
Yerevan	194 (57.7%)	14 (48.3%)	49 (50.5%)	***257 (55.6%)***
Category				
B&B	125 (37.2%)	20 (69.0%)	40 (41.2%)	***185 (40.0%)***
Hotels	73 (21.7%)	2 (6.9%)	29 (29.9%)	***104 (22.5%)***
Specialty Lodging	138 (41.1%)	7 (24.1%)	28 (28.9%)	***173 (37.4%)***
Free WiFi				
0	70 (20.8%)	11 (37.9%)	22 (22.7%)	***103 (22.3%)***
1	266 (79.2%)	18 (62.1%)	75 (77.3%)	***359 (77.7%)***
Free Parking				
0	114 (33.9%)	13 (44.8%)	35 (36.1%)	***162 (35.1%)***
1	222 (66.1%)	16 (55.2%)	62 (63.9%)	***300 (64.9%)***

The table above provides summary statistics, such as mean, standard deviation, median, minimum and maximum value for each variable. We see that average price of Armenian hotels is $44 ranging from 8 to 549 dollars. By further exploring the other variables, we see that none of them has a constant value. Therefore, assumption about sample variation in explanatory variables is met.

Next, we need to check the assumption about perfect collinearity between the explanatory variables. As the correlation matrix shows, there is no perfect correlation between the variables. Therefore, this assumption is met.



A few interesting plots can be developed showing how the data looks like. For example, the boxplot below shows the distribution of price per each state. We see that the median price for each state ranges between $30 and $45.



For finding other figures of visual analysis please refer to Appendix 7.2: Graphs.

5.3. Regression Results

In this section regression results of five different models are presented, as well as intuition behind the variables is discussed.

Table 2: Regression output

	Dependent variable - Logarithm of Price (USD)				
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Yerevan (0,1)	-0.207 ***	-.211 ***	-0.298 ***	-0.315 ***	-0.315 ***
Number of reviews	0.002 ***	0.002 ***	0.001 **	0.001	0.001 **
Average score (0-5)	0.036	0.036			
Number of photos		0.000			
Free WiFi (0,1)		0.049			
Free parking (0,1)		-0.029			
Low score (1-2)			0.094		
High score (4-5)			0.125 *	0.106 *	0.096 **
High score*Number of reviews				-0.001	
Hotels*Yerevan			0.519 ***	0.498 ***	0.505 ***
Hotels			0.025		
B&B			0.09		
Adjusted R-squared	0.075	0.069	0.137	0.135	0.137

*Note: *** Statistically significant at 1%; ** statistically significant at 5%; * statistically significant at 10%*

The first model was constructed using independent variables Yerevan, number of reviews and average score. As we see, the first two variables are significant, whereas average score is not significant even at 10% significance level.

Therefore, it was decided to turn "average score" into a categorical variable with levels High (4-5), Medium (2-4) and Low (1-2). Model 3 was constructed using this categorical variable and medium score was used as a base level. Here, dummy variables related to the hotel category (Hotels, B&B) specialty lodging being the base level are included in the model. Moreover, an interaction term between Hotels and Yerevan was introduced into the model. However, some of the variables turned out not to be significant.

After constructing different models, the best one among them turned out to be model 5 with the following expression:

$$\ln(price) = \beta_0 + \beta_1 * yerevan + \beta_2 * (number\ of\ reviews) + \beta_3 * (high\ score) + \beta_4 * (hotels * yerevan)$$

Table 3: Regression output of Model 5

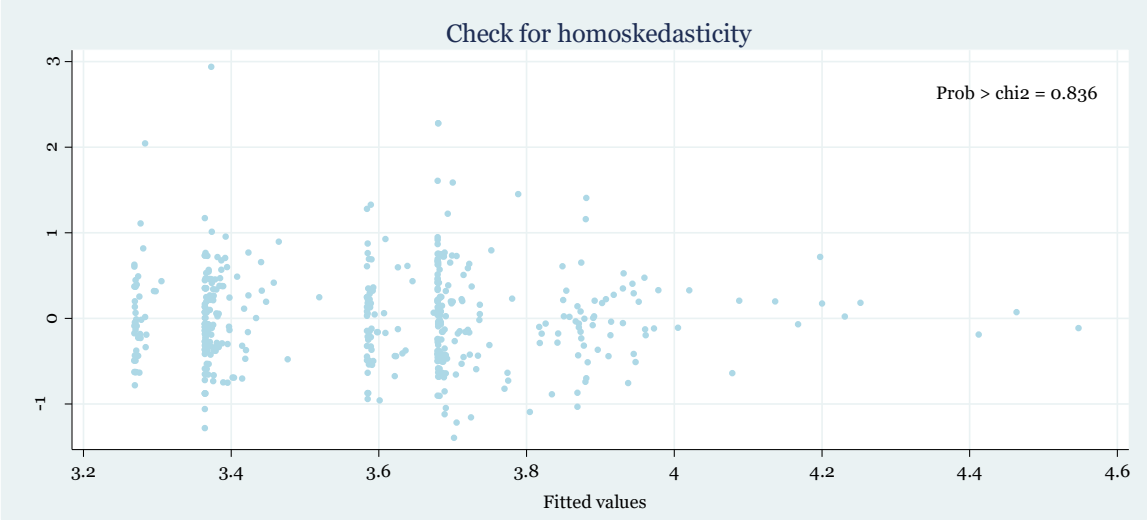
Source	SS	df	MS
Model	23.1826446	4	5.79566116
Residual	137.855089	457	.301652274
Total	161.037734	461	.349322633

Number of observations	462
F (4, 457)	19.21
Prob > F	0.0000
R-squared	0.1440
Adj R-squared	0.1365
Root MSE	0.54923

Log(price)	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
Yerevan	-.3150969	.0553545	-5.69	0.000	-.4238778	-.2063161
Number of reviews	.0008555	.0003538	2.42	0.016	.0001602	.0015507
High Score	.0955978	.0577104	1.66	0.098	-.0178128	.2090084
Hotels*Yerevan	.5045306	.0874006	5.77	0.000	.3327738	.6762875
Intercept	3.583566	.0554338	64.65	0.000	3.474629	3.692503

$$\ln(price) = 3.6 - 0.32 * yerevan + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * (hotels * yerevan)$$

In order the standard errors, t-statistics, their corresponding p-values, as well as F statistic to be reliable, we need to check for homoskedasticity of residuals for model 5.



As the Breusch-Pagan test and the residuals plot show, the errors are homoscedastic. Therefore, we can safely use the statistics.

The model, as measured by the adjusted R-squared, shows that 13.65% of the variance in Ln(price) is explained by the variables included in the analysis. The F-statistic is significant at the 0.01 cutoff value. This provides evidence of the existence of a linear relationship between the dependent and the explanatory variables.

The interpretation of the model coefficients is the following:

As we see, the dummy variable, Yerevan, has a negative sign meaning that if the hotel is located in Yerevan, the price is estimated to be less by $(e^{-.315} - 1) * 100 = 27\%$. My initial hypothesis was that hotels in Yerevan would have higher prices, on average. However, this may be connected with the fact that there can actually be more cheaper hotels in Armenian districts, but they are not being listed in TripAdvisor and reviewed by tourists. Whereas, most of the hotels in Yerevan are being recorded in TripAdvisor's database.

Number of reviews have a positive impact on the price. If we consider that the number of reviews of a hotel increases by 100, then the price is predicted to increase by only 8.5% holding all other variables constant, which is not a significant impact.

The coefficient of the variable "High score" shows that on average, the price of a hotel is estimated to be more by $(e^{.0956} - 1) * 100 = 10\%$ compared if the score is low or medium. This proves that hotels take into account users' feedback about the hotel when setting the price.

Interaction between Hotels and Yerevan variables is being introduced into the model. This variable is significant and the coefficient can be interpreted in the following way:

$$\ln(price) = 3.6 - 0.32 * yerevan + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * (hotels * yerevan)$$

The accommodation is located in Yerevan and the type of it is "hotel":

$$\ln(price) = 3.6 - 0.32 * 1 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * 1 * 1 = 3.78 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score)$$

The accommodation is located in Yerevan and the type of it is NOT "hotel":

$$\ln(price) = 3.6 - 0.32 * 1 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * 0 * 1 = 3.28 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score)$$

The accommodation is NOT located in Yerevan and the type of it is "hotel":

$$\ln(price) = 3.6 - 0.32 * 0 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * 1 * 0 = 3.6 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score)$$

The accommodation is NOT located in Yerevan and the type of it is NOT "hotel":

$$\ln(price) = 3.6 - 0.32 * 0 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score) + 0.5 * 0 * 0 = 3.6 + 0.00086 * (number\ of\ reviews) + 0.096 * (high\ score)$$

$$\ln(price)\widehat{_{Yerevan_{hotel}}} - \ln(price)\widehat{_{NotYerevan_{hotel}}} = 0.18$$

$$\ln(price)\widehat{_{Yerevan_{hotel}}} - \ln(price)\widehat{_{Yerevan_{NotHotel}}} = 0.5$$

$$\ln(price)\widehat{_{NotYerevan_{hotel}}} - \ln(price)\widehat{_{Yerevan_{NotHotel}}} = 0.32$$

These findings show that the price of accommodations which are located in Yerevan and those located outside differs by 18% if the hotel category is "hotel". The other coefficients can be interpreted with the same approach.

6. Conclusion

The findings have shown that online ratings and reviews have a linear relationship between hotel prices in Armenia. However, online ratings and reviews are not the only variables that influence hotel prices. The location of the hotel as well as other variables can shape hotel prices as the adjusted R-squared of the model is low, meaning that the variation of independent variables is not explaining the variation of hotel prices very well.

Based on the constructed model, we can make some predictions and predict a hotel price and see if the number makes sense.

We can take the following hotel as an example:

The hotel is located in Yerevan, its category is “specialty lodging”, it has 11 reviews and its average score is 4.8 (high score). Let’s see what will be the predicted price of the model:

$$\text{price} = e^{3.6-0.32*1+0.00086*11+0.096*1+0.5*(0*1)} = \$29.5$$

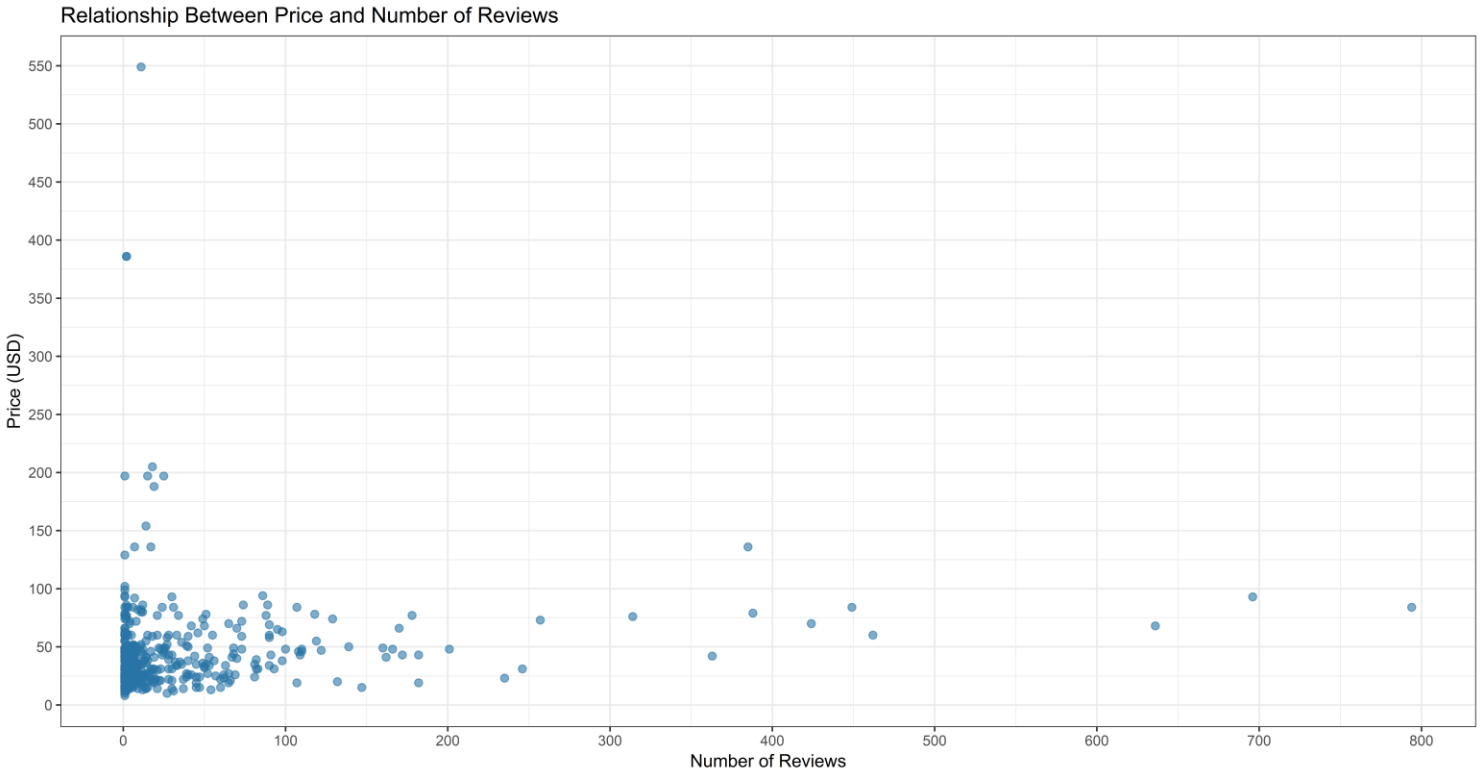
The model predicted that the price of the hotel is $29.5, whereas the actual price of this hotel is $46. We see that the model made a poor prediction. This is connected with the fact that other factors influence hotel prices in Armenia as well.

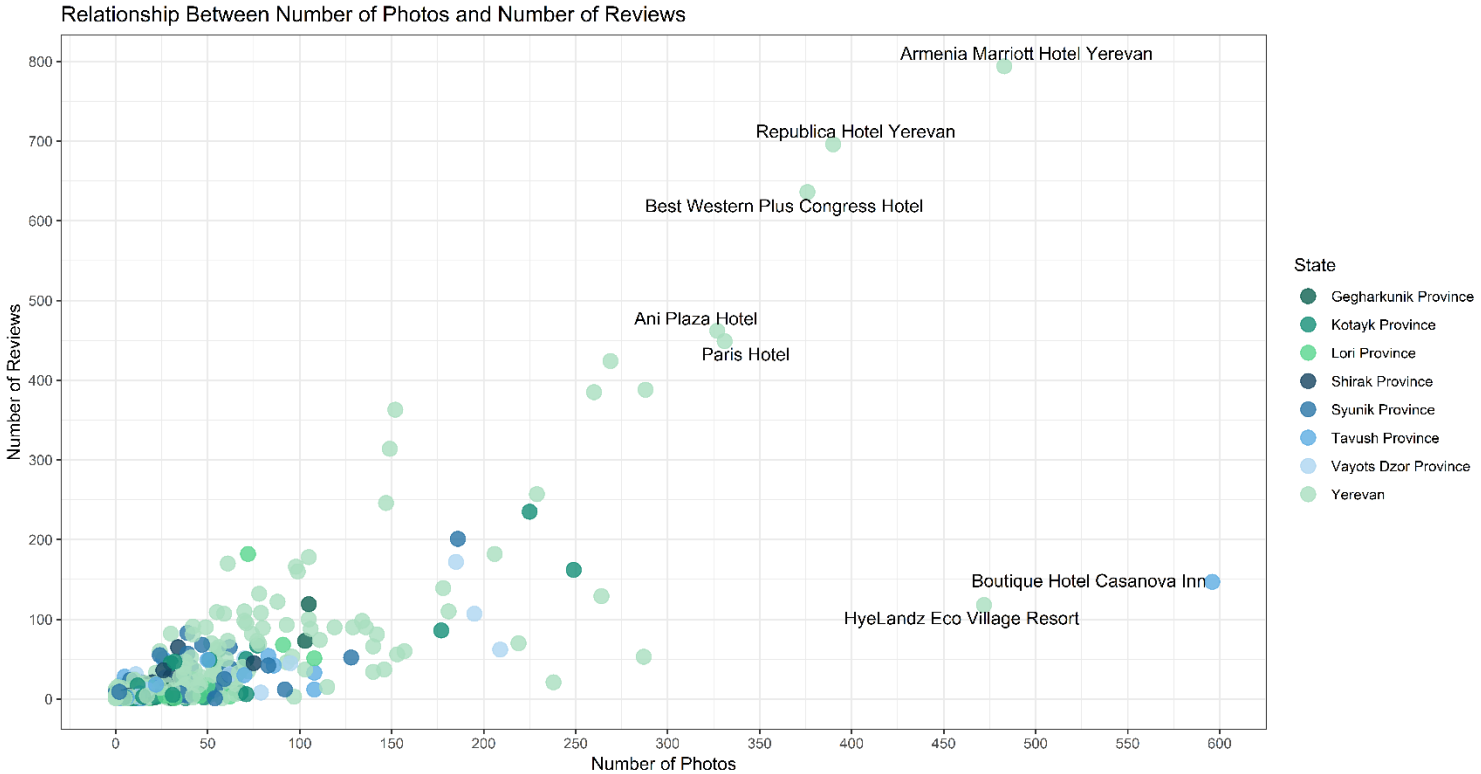
7. Appendices

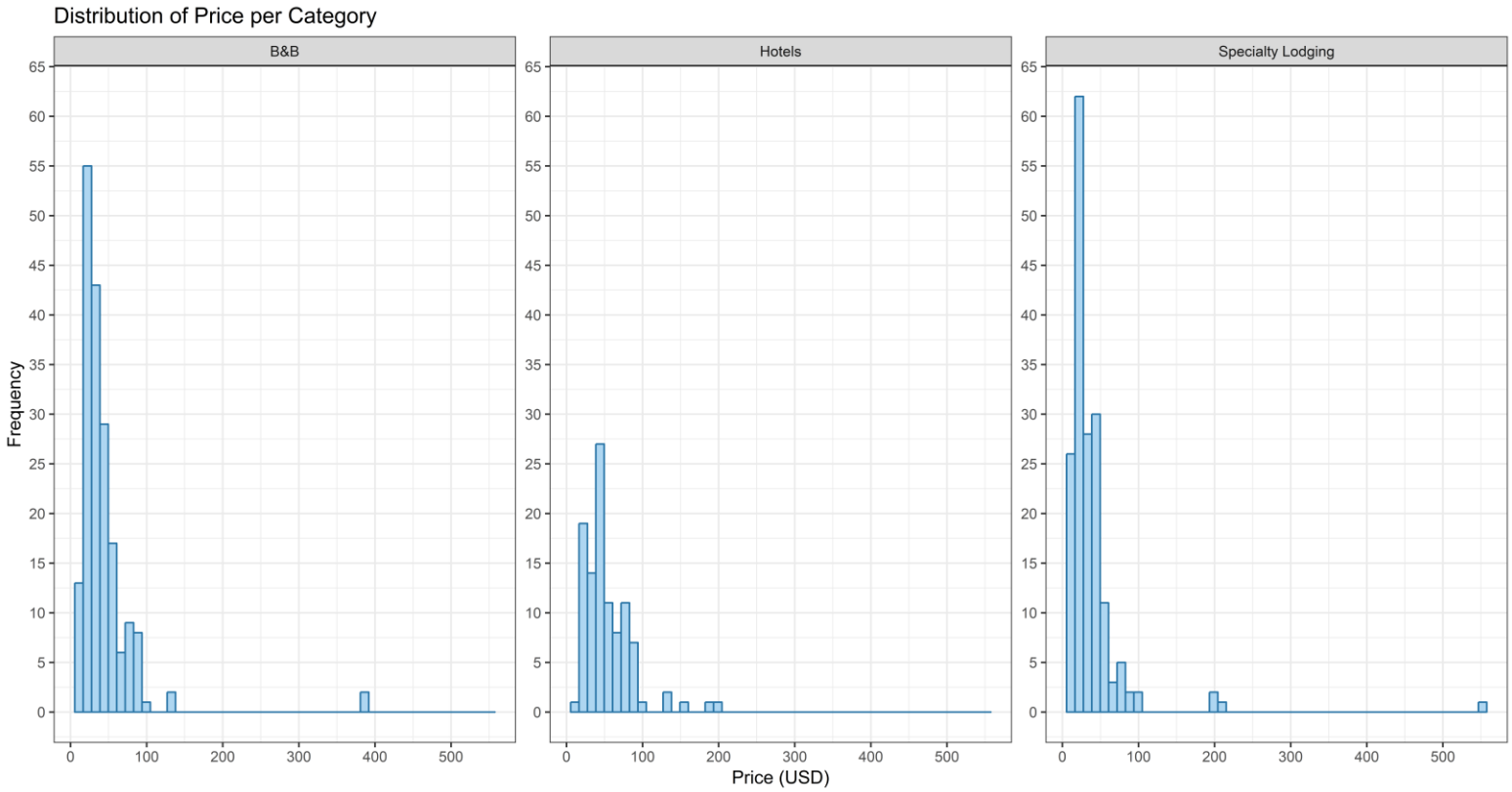
7.1. Data

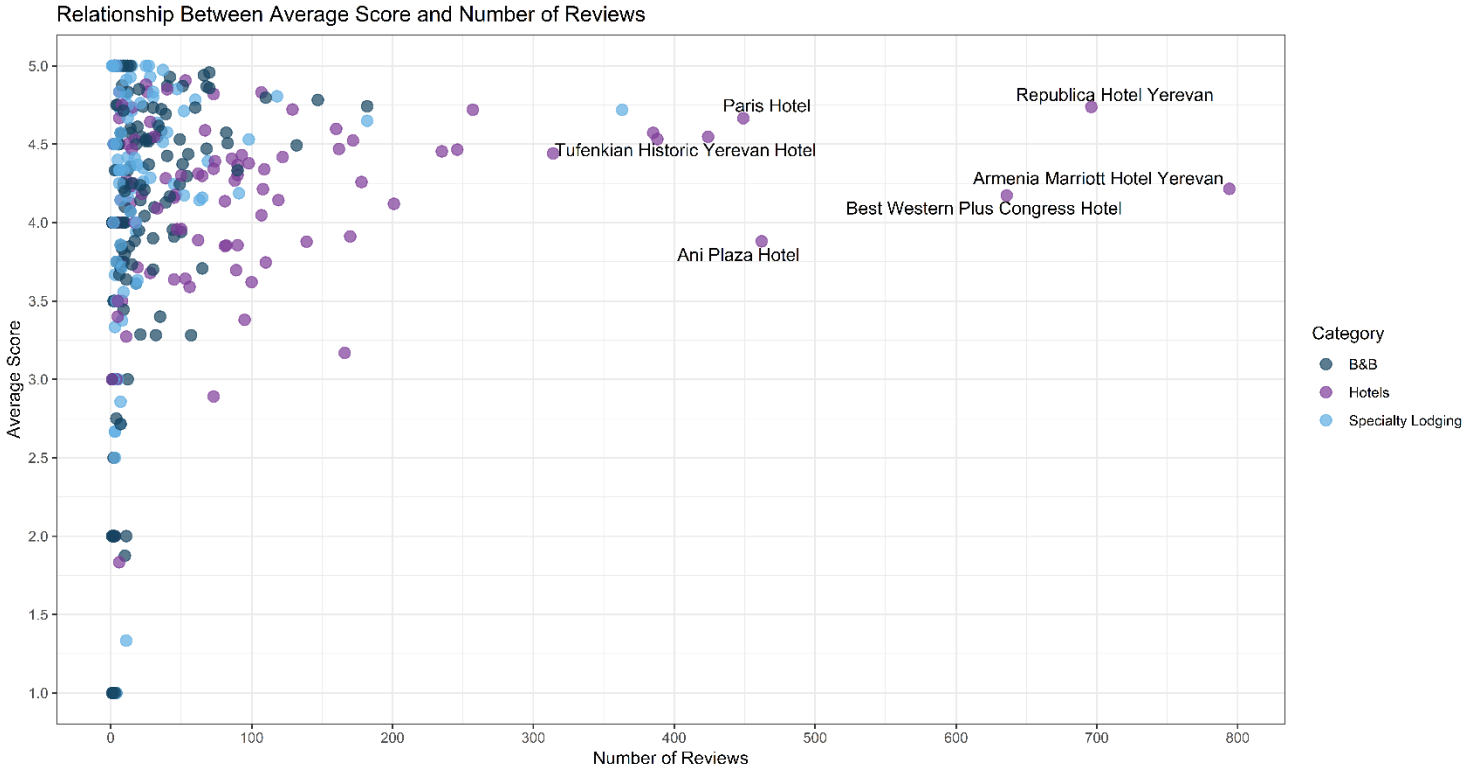
price	*photos*	*review count*	*average score*	*free wifi*	*free parking*	*yerevan*	*B&B*	*specialty lodging*	*hotels*	*high score*	*medium score*	*low score*
69	119	90	4.37	1	1	1	0	0	1	1	0	0
30	0	2	1.00	0	0	1	0	1	0	0	0	1
549	15	11	4.73	1	1	1	0	1	0	1	0	0
35	8	3	5.00	1	1	1	0	1	0	1	0	0
46	6	11	4.82	1	1	1	0	1	0	1	0	0
60	23	3	5.00	1	1	0	0	1	0	1	0	0
92	33	7	5.00	0	0	1	0	0	1	1	0	0
21	63	14	4.12	0	0	0	0	0	1	1	0	0
26	56	62	3.89	1	1	0	0	0	1	0	1	0
52	33	11	3.27	1	1	0	0	0	1	0	1	0
38	25	7	4.14	1	1	1	0	0	1	1	0	0
21	34	22	4.55	1	1	1	1	0	0	1	0	0
16	1	3	3.50	1	1	1	1	0	0	0	1	0
31	57	30	3.90	1	1	0	1	0	0	0	1	0
19	0	2	3.50	0	0	0	1	0	0	0	1	0
72	6	8	3.50	1	0	0	0	0	1	0	1	0
23	11	2	5.00	0	0	1	0	1	0	1	0	0
51	57	39	4.28	1	1	0	0	0	1	1	0	0
22	10	2	5.00	0	0	0	0	1	0	1	0	0
35	26	11	4.27	1	1	1	0	0	1	1	0	0

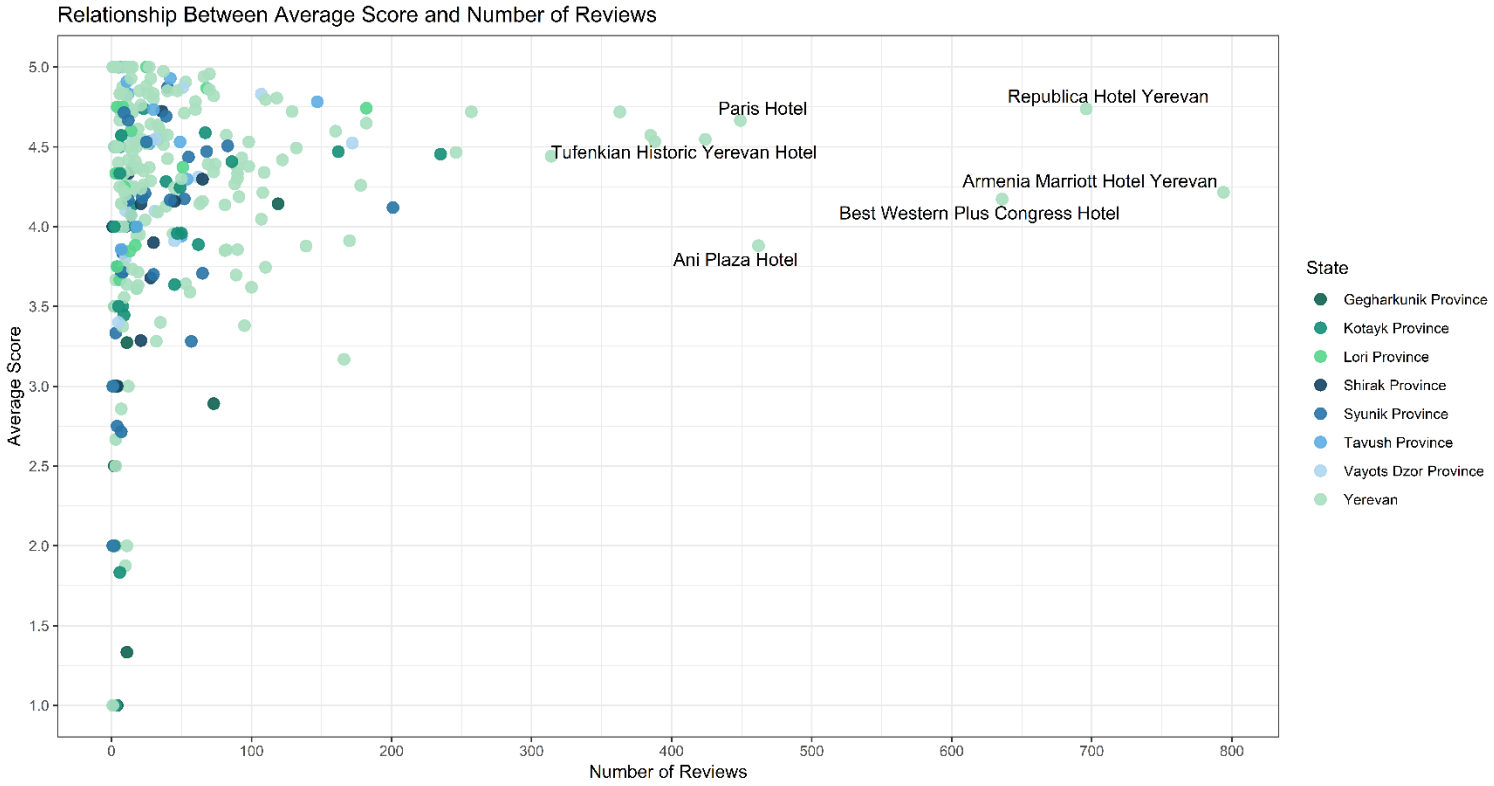
7.2. Graphs

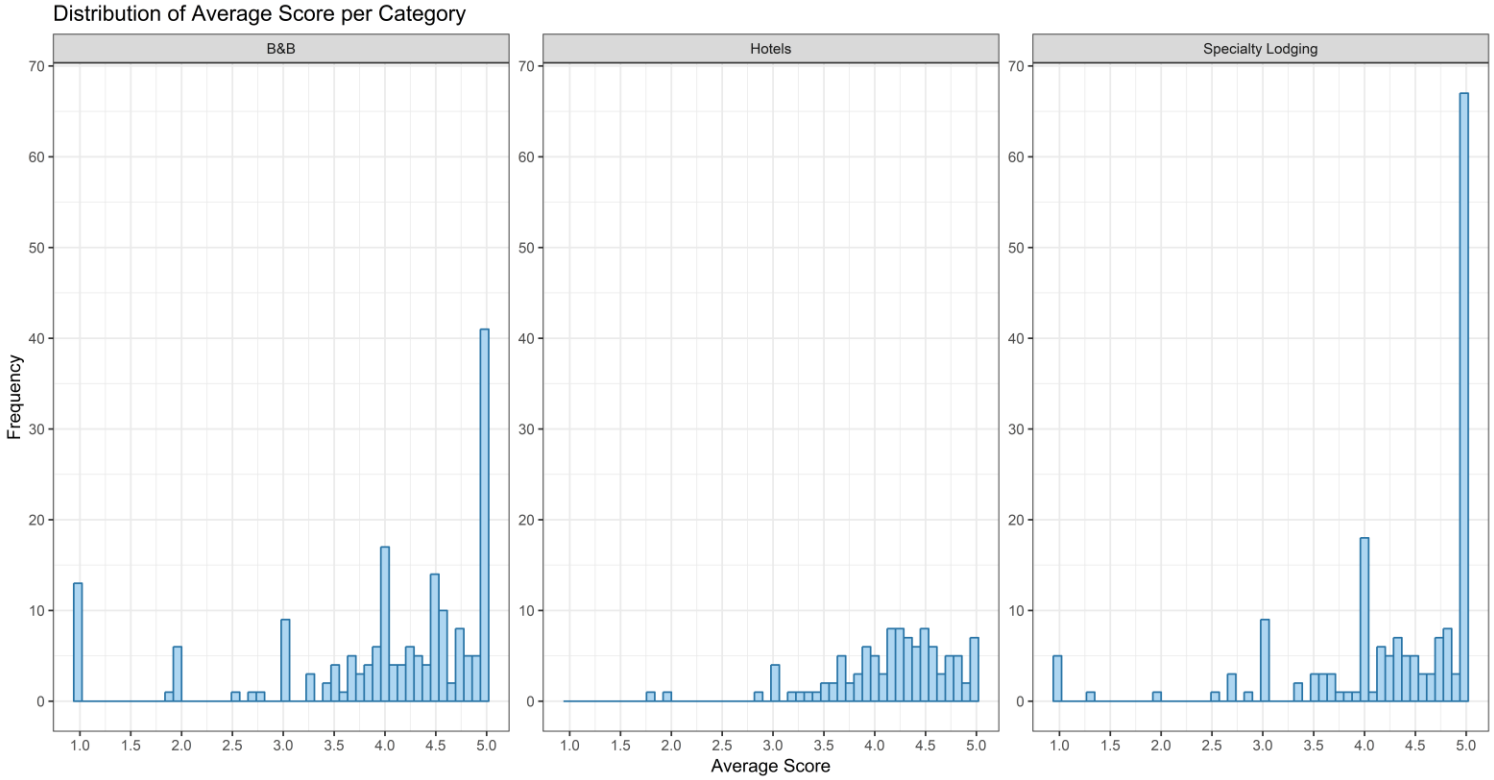

Relationship Between Price and Number of Reviews
Price (USD)
Number of Reviews
0
50
100
150
200
250
300
350
400
450
500
550
0
100
200
300
400
500
600
700
800

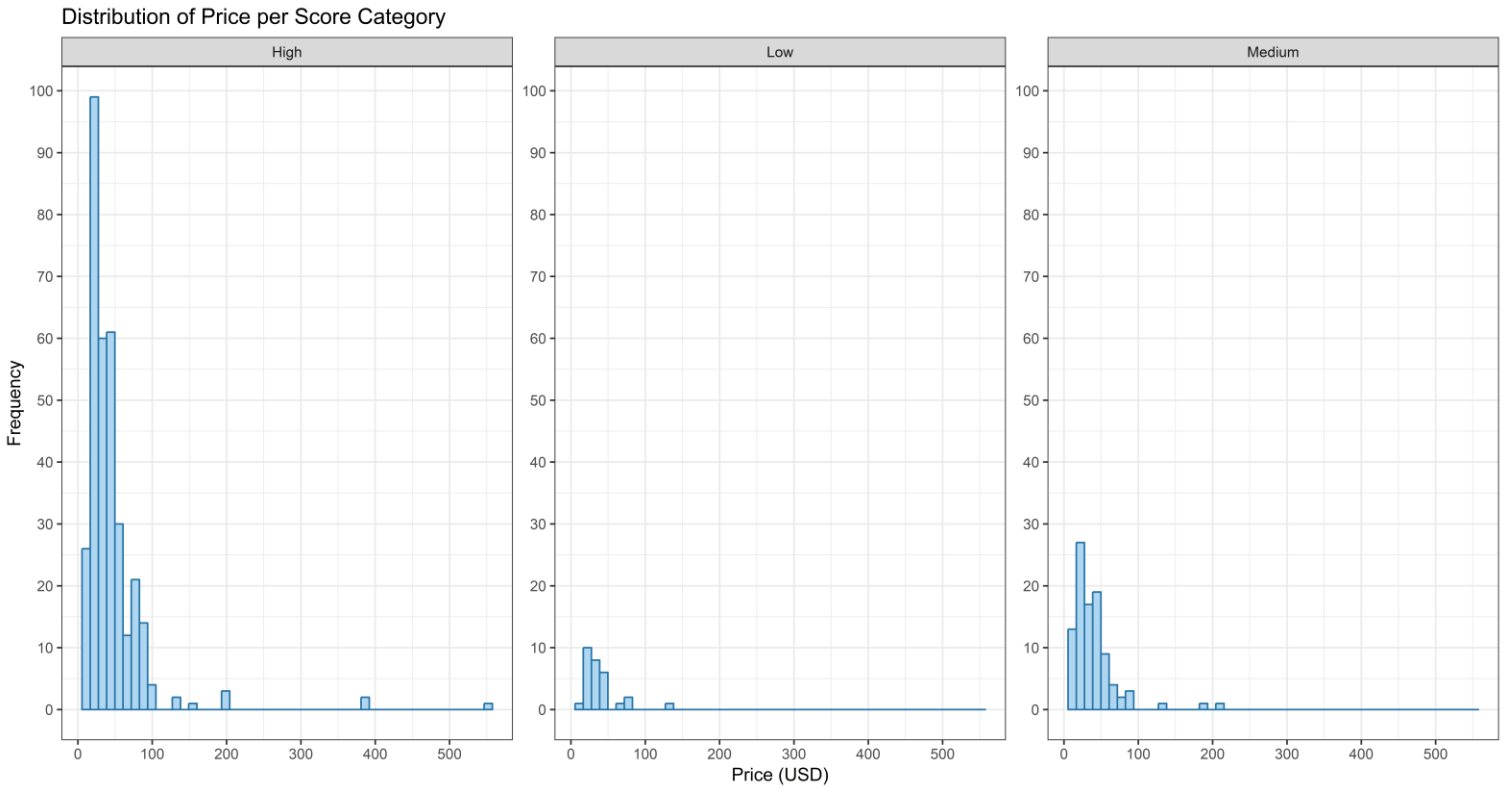

Relationship Between Number of Photos and Number of Reviews
Number of Reviews
Number of Photos
Armenia Marriott Hotel Yerevan
Republica Hotel Yerevan
Best Western Plus Congress Hotel
Ani Plaza Hotel
Paris Hotel
Boutique Hotel Casanova Inn
HyeLandz Eco Village Resort
State
Gegharkunik Province
Kotayk Province
Lori Province
Shirak Province
Syunik Province
Tavush Province
Vayots Dzor Province
Yerevan

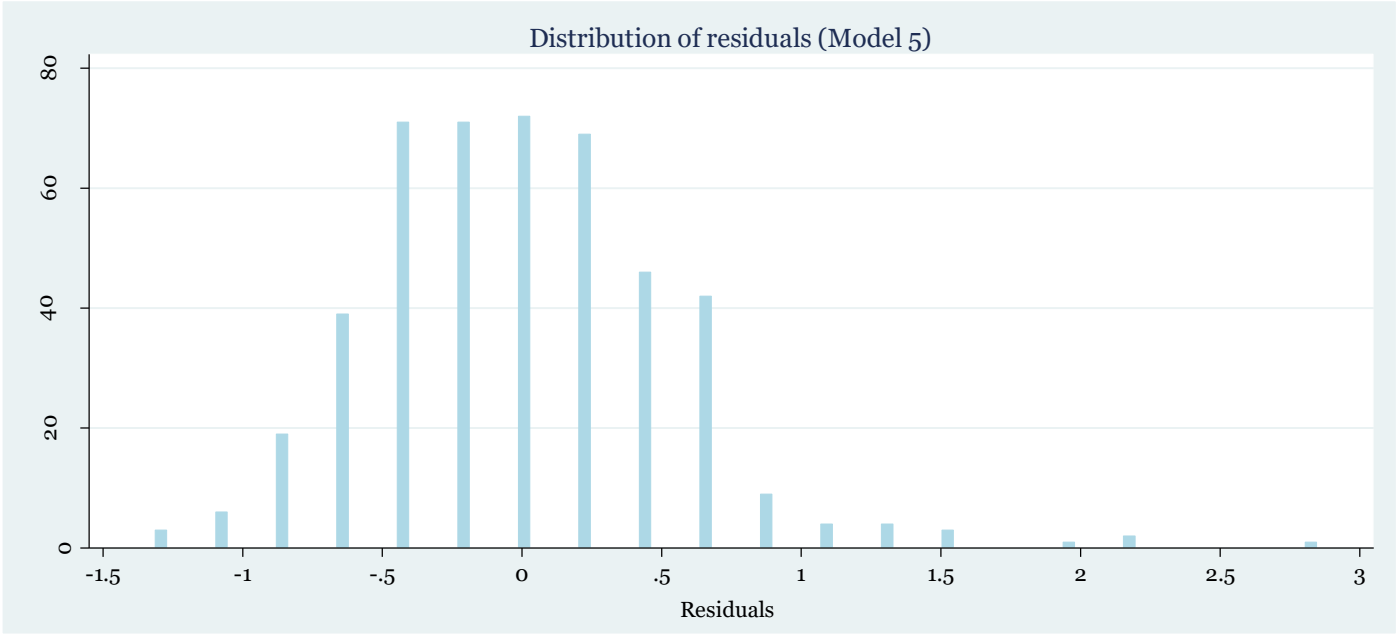

Distribution of Price per Category
B&B
Hotels
Specialty Lodging
Frequency
Price (USD)


Relationship Between Average Score and Number of Reviews
Paris Hotel
Republica Hotel Yerevan
Tufenkian Historic Yerevan Hotel
Armenia Marriott Hotel Yerevan
Best Western Plus Congress Hotel
Ani Plaza Hotel
Average Score
Number of Reviews
Category
B&B
Hotels
Specialty Lodging


Relationship Between Average Score and Number of Reviews
Average Score
Number of Reviews
Paris Hotel
Republica Hotel Yerevan
Tufenkian Historic Yerevan Hotel
Armenia Marriott Hotel Yerevan
Best Western Plus Congress Hotel
Ani Plaza Hotel
State
Gegharkunik Province
Kotayk Province
Lori Province
Shirak Province
Syunik Province
Tavush Province
Vayots Dzor Province
Yerevan


Distribution of Average Score per Category
B&B
Hotels
Specialty Lodging
Frequency
Average Score


Distribution of Price per Score Category
High
Low
Medium
Frequency
Price (USD)


Distribution of residuals (Model 5)
Residuals

8. References

Ferreira, F. A. (2018). Online hotel ratings and its influence on hotel room rates: the case of Lisbon, Portugal.

Gregory Lewis, G. Z. (2016). The Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry.

Hu, B. K. (2012). Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments.

Ph.D., C. A. (2012). The Impact of Social Media on Lodging Perforamnce.

Sparks, B. A. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behaviour.

UNWTO. (2015). United Nations World Tourism Organization Study on Online Guest Reviews and Hotel Classification Systems: An Integrated Approach.

Walker, S. A. (2012). Identifying Influential and Susceptible.