International Business & Economics Studies
ISSN 2640-9852 (Print) ISSN 2640-9860 (Online)
Vol. 4, No. 1, 2022
www.scholink.org/ojs/index.php/ibes

Original Paper

What Factors drives the Airbnb Listing's Prices?

Mwigeka, Samwel[1*]

[1] Department of Humanities, Ruaha Catholic University, Iringa, Tanzania

[*] Mwigeka, Samwel, Department of Humanities, Ruaha Catholic University, Iringa, Tanzania

Received: January 10, 2022 Accepted: February 26, 2022 Online Published: February 28, 2022

doi:10.22158/ibes.v4n1p26 URL: http://dx.doi.org/10.22158/ibes.v4n1p26

Abstract

The main aim of this study is to assess the impact of Airbnb listing attributes on price. Pricing is widely acknowledged to be one of the most critical factors determining the long-term success of the accommodation industry (Hung et al., 2010). The study has applied OLS regression in assessing the relationship between price and its determinants.

The study has shown that there exist significant impacts of the variables on price. The variables such as number of beds, number reviews, room type dummies, property type dummies and neighborhood dummies have shown the impact of prices for the New York City, supported by Wang and Nicolau (2017), Portolan (2013), and Dogru and Pekin (2017).

In supplementing OLS results the study has employed variance-weighted least square and feasible generalized least square which they possess stronger estimation properties than OLS.

The results of this research provide instructions and direction to main tourism stakeholders and policymakers, and at the same time assist facility owners in shaping prices helping them to create instruments for future planning.

Keywords

Airbnb prices, ordinary least square and generalized least square

1. Introduction

Sharing economy has significant importance in the growth of the economy in the apartment sector; the world economy has witnessed increased demand for housing/accommodation due increased global business as well the growth of the tourism industry worldwide. This situation leads to renting and buying of residential apartments, which in turn derives the residential apartment's demand/need of housing for accommodation.

Consequently, it is has been indispensable to find out what are drivers of apartment prices. Pricing is widely acknowledged to be one of the most critical factors determining the long-term success of the accommodation industry (Hung, Shang, & Wang, 2010). The general economic theory claims that the price, in the absence of market failures should be a valid economic signal that helps customers make informed decisions, since it includes information regarding the quality of a good or a service (Leoni & Nilsson, 2020)

Airbnb is a home-sharing platform that permits home-owners and renters (“hosts”) to put their possessions (“listings”) online, so that lodgers can pay for lodging in. The owners set their own prices for their listings, and though Airbnb and other sites arrange for particular guidance, at present there is no free and precise services that can assist hosts to price their properties using a wide variety of data points.

Airbnb has developed a highly sophisticated algorithm, called Smart Pricing, which suggests the optimal price (Hill, 2015; Ye et al., 2018) but leaves the host free to decide whether to accept the recommendation.

The sticker price is the general nightly price that is advertised to vital lodgers, rather than the actual average amount paid per night by preceding lodgers. The prices could be usually set at any choice amount by the host, and hosts that are less experienced with Airbnb will often set these to very low or very high.

Airbnb was established in August 2008 by Brian Chesky, Joe Gebbia, and Nate Blecharczyk in San Francisco, California and it presented itself on its site as: “a trusted a community marketplace for people to list, discovers, and book unique accommodations around the world”

According to Airbnb (Airbnb, 2016), an average of two million travelers is staying in accommodations provided via the platform every night (Airbnb, 2019). The emergence of online platforms and the smartphone revolution have drastically changed its scale (Guttentag, 2015).

2. Method

2.1 Data

Toan Ngoc Bui (2020) expose the influence of the size of an apartment, presence of balcony, presence of swimming pool, presence of shopping malls, and periodic rental income or value, and proximity to the center which shows a negative impact on apartment prices while the rest factors affect apartment prices positively. Yang et al (2016) also, Yayar and Demir (2014) supports this argument that distance from the city center, access to public transportation, and proximity to attractions affects negatively prices.

Kain and Quigley (1970) indicate that housing prices are significantly affected by the total number of bedrooms, bathrooms, and size of the house. Cebula (2009) and Selim (2009) augment the results given by Kain and Quigley (1970); they accord on the influence of the number of bedrooms (size of the house), the number of bathrooms and presence of swimming pools as influential factors on housing prices.

Amenyah and Afenyi (2013) concluded that rental house prices are greatly influenced by the location and size of the house, as well as its nearby facilities. The number and proximity of competitors have been shown to influence hotel prices. Low market accessibility, indicated by high flight costs, is also associated with low hotel prices.

Reviews and ratings are similarly significant aspect in Airbnb listing prices. Usually, mutually higher ratings and more frequent reviews can result to a higher price. Therefore, the number of reviews and the ratings of Airbnb listings are considered as important explanatory variables in this particular study.

The study attempts to answer the following questions: How the price is impacted by listing metrics such as neighborhood and room type, number of bathrooms, swimming pool, and other factors. Also, how the availability of these listings was impacted by the same metrics. The data will be transformed into logarithm due to significant differences in terms of size; it minimizes the influence due to the size of the variables (data).

2.2 Choice Variable

The study uses cross-section data employed from the dataset of Airbnb. Cross-sectional data created at the same point in time may be impacted by an exogenous stochastic common shock that impacts all data units, maybe, to an unpredictable degree.

Given the review on the previous literature on price determination and related literature, the study has to examine the following variables: price, number of beds, review score rating, number of reviews, room type, and neighborhood. The reviewed studies have initiated efforts to examine the factors determining the price.

Description of the variables for the study:

Table 1. Variable Names and Their Generated Dummy Variables

variable	dummies
room type	private_room shared_room entire_home
neighbourhood	bronx brooklyn manhattan queens island
property type	apartment house other_property

Table 2. Listings Attributes

Variable names	meaning
ln_price	natural logarithm of the price set per night (by default)

private_room	dummy of private room
shared_room	dummy of shared room
entire_home	dummy of entire home
bronx	dummy of Bronx
brooklyn	dummy of Brooklyn
manhattan	dummy of Manhattan
queens	dummy of Queens
island	dummy of Island state
apartment	dudummy of apartment
house	dummy of house
other_property	dummy of other_property
beds	number of beds
numberofreviews	number of reviews
reviewscoresratingbin	review scores rating bin

3. Methodology

In analyzing the price of listings, the variable of price per person per night (in logarithmic form) is selected as the dependent variable. As the resulting expression is a semi-logarithmic specification, the coefficient values represent semi-elasticities, namely the percentage change in price when an explanatory variable varies by 1 (Wang, 2017).

The coefficients of non-transformed dummy variables will be interpreted either in the percentage term calculated by $(e^{\beta}-1)*100$, where β is the coefficient and e is the base of the natural logarithm, or in the dollar units (Monty & Skidmore, 2003).

3.1 Ordinary Least Square Regression

The initial step will involve describing the data (descriptive analysis) and followed by linear OLS regression model which is employed to detect linear relationships between a dependent variable and a set of explanatory variables. It makes use of the t-statistics to evaluate the significance of the obtained coefficients of parameters from the OLS results. The model for ordinary least square is as follows:

$$logP_i = \beta_1 + \beta_2 B_i + \beta_3 R_i + \beta_4 SR_i + \beta_5 RT_i + \beta_6 NBH_i + u_i \ldots \qquad \ldots.(1)$$

Where: P=price, B=number of beds, R=number of reviews, SR=review of score rating, RT=room type, NBH=neighborhood

With the use of dummy variables, the OLS regression has to use the following equation:

$$\text{logprice} = \beta_1 + \beta_2\text{beds} + \beta_3\text{entire\_home} + \beta_4\text{private\_room} + \beta_5\text{numberofreviews} + \beta_6\text{manhattan} + \beta_7\text{bronx} + \beta_8\text{queens} + \beta_9\text{brooklyn} + \beta_{10}\text{house} + \beta_{11}\text{other\_property} + u_i \qquad \ldots.(2)$$

3.2 Variance-Weighted Least Square

In improving the ordinary least square regression results the study has made use of the variance weighted least square regression. When the assumption of constant variance does not hold, the ordinary least squares estimates of the regression coefficients tend to have variances that are too high (incorrect t- and F-tests).

According to Suarez et al. (2017) accords that in designed experiments with large numbers of replicates, weights can be estimated directly from sample variances of the response variable for each combination of the predictor variables. Variance-weighted least-squares estimates of the coefficients will usually be nearly the same as the OLS estimates. The the study uses the following equation for estimation of variance weighted least square:

$$\text{logprice} = \beta_1 + \beta_2\text{beds} + \beta_3\text{entire\_home} + \beta_4\text{private\_room} + \beta_5\text{numberofreviews} + \beta_6\text{manhattan} + \beta_7\text{bronx} + \beta_8\text{queens} + \beta_9\text{brooklyn} + \beta_{10}\text{house} + \beta_{11}\text{other\_property} + \boldsymbol{u_i} \quad \ldots.(3)$$

3.3 Feasible Generalized Least Square Regression

Usually, cross section data possess the problem of nonconstant variance, thus the study has employed the feasible generalized least square as a remedy to the problem. FGLS estimates the structure of heteroskedasticity from OLS as an alternative to the assumption of assuming the structure of heteroskedasticity.

Feasible generalized least square yields more precise estimates (smaller standard errors and bigger t-statistics). Feasible generalized least squares test is crucial in tackling the problem of outliers, heteroskedasticity, and bias in data. It is capable of producing estimators that are "Best Linear Unbiased Estimates". The equation to be estimated takes the following functional form:

$$\text{logprice} = \beta_1 + \beta_2\text{beds} + \beta_3\text{entire\_home} + \beta_4\text{private\_room} + \beta_5\text{numberofreviews} + \beta_6\text{manhattan} + \beta_7\text{bronx} + \beta_8\text{queens} + \beta_9\text{brooklyn} + \beta_{10}\text{house} + \beta_{11}\text{other\_property} + [\text{aw}=1/\text{hhat}] \quad \ldots.(4)$$

3.4 F-Test

The study makes use of the F-test to check the precision of the t-test. The F value in regression is the result of a test where the null hypothesis is that all of the regression coefficients are equal to zero.

The F-test of overall significance measures if the linear regression model gives a better fit to the data than a model that encompasses no independent variables. It compares the p-value for the F-test to the significance level. If the p-value is less than the significance level, an alternative hypothesis is concluded and vice versa.

3.5 Heteroscedasticity Test

However, the study has employed a heteroscedasticity test for the model. Linear regression models estimated via Ordinary Least Squares (OLS) rest on several assumptions, one of which is that the variance of the residual from the model is constant and unrelated to the independent variable(s). Constant variance is called homoscedasticity, while non-constant variance is called heteroscedasticity.

4. Results

The number of recording data was removed from our analysis because of high collinearity with explanatory variables included in the model.

4.1 Summary Statistics of the Dependent and Independent Variables

Table 3. Presents the Summary Statistics of the Dependent and Independent Variables Used in this Study, Along with Minimum and Maximum Values of These Variables where Applicable

Variable	Obs	Mean	Std. Dev.	Min	Max
reviewscor~n	22,155	90.73866	9.059519	20	100
beds	30,393	1.530089	1.015359	0	16
numberofr~ds	30,478	1	0	1	1
numberofr~ws	30,478	12.01873	21.9807	0	257
entire_home	30,478	.5585668	.4965662	0	1
private_room	30,478	.4137082	.4925055	0	1
shared_room	30,478	.0277249	.1641863	0	1
bronx	30,478	.0113196	.1057917	0	1
brooklyn	30,478	.3830632	.4861415	0	1
manhattan	30,478	.5260516	.499329	0	1
queens	30,478	.0747424	.2629796	0	1
island	30,478	.0048232	.0692824	0	1
apartment	30,478	.8892316	.3138503	0	1
house	30,478	.0685741	.2527326	0	1
other_prop~y	30,478	.0421944	.2010356	0	1
logprice	30,478	4.860495	.627704	2.302585	9.21034

4.2 Ordinary Least Square Regression Results

Table 4. Gives a Summary Result from the Ordinary Least Square Regression

	Number of obs = 30,393 F(10, 30382) = 3420.23 Prob> F = 0.0000 R-squaed = 0.5474 Root MSE = .42208
logprice	robust Coefficients. Std. Error t-statistics P>|t|
Beds	.1477817 .004127 35.81 0.000
entire_home	.9466141 .0185869 50.93 0.000
private_room	.3058343 .0184215 16.60 0.000
numberofreviews	-.0013299 .0000968 -13.73 0.000

manhattan	.5460121 .0430679 12.68 0.000
Bronx	-.0076206 .0484621 -0.16 0.875
queens	.0915817.0434136 2.11 0.035
Brooklyn	.2226393 .0430134 5.18 0.000
House	.0184209 .011222 1.64 0.101
other_property	.1966551 .0158924 12.37 0.000
Constant	3.605702 .0464503 77.62 0.000

Explanatory power of the model is medium, explaining 54.7% of the variations in prices as measured by the unadjusted R^2. This implies that the variation of variables in the model does account for the variation of the price (dependent variable) by 54.7 percent.

The output from OLS indicates that the variables are statistically significant at 1%, 5%, and 10%. As expected, the higher number of beds leads to higher prices (in particular, according to the semi-elasticity estimated, the price increases 15.9%. Host profile image/picture has a substantial negative parameter, related with lower prices (the semi-elasticity displays a reduction of 10.89%).

The farther the accommodation from the city center, the lower its price; in particular, the estimated semi-elasticity presents a reduction of 0.75% per kilometer for Bronx. The results are similar with the study by Wang (2019), while Manhattan, Queens, and Brooklyn increase by 72.6%, 9.6%, and 24.9% respectively from city centre. These findings accord with previous studies results by Yang et al (2016) also Yayar and Demir (2014).

The reference base for the room types “entire home” and “private room” is the private room. Private is statistically significant and positive parameters (semi-elasticity equal to 35.8%), with a units increase in rooms prices increases by 35.8%, and the entire home with semi-elasticity 157.7% which implies that with units increase in homes price will increase by 157.7%. The results are supported by the study by Wang and Nicolau (2017). However, the “private room” variable has smaller parameters than the “entire apartment,” and thus has a less positive effect on prices.

The variable “number of reviews” has a negative effect on the price, as indicated by both OLS (as an example, each additional review leads to a price decrease of 0.13% (semi-elasticity equal to 0.13%). Previous researchers have reported that most tourists choose to rent sharing economy based accommodation to reduce costs (Balck & Cracau, 2015; Guttentag, 2015; Quinby & Gasdia, 2014).

4.3 Variance-Weighted Least Square Regression

Table 5. Presents the Results from Variance-Weighted Least Square Regression

Goodness-of-fit chi2(1587) = 55515.88	Number of obs = 28,630
Prob> chi2 = 0.0000	Model chi2(10) = 1370425.51
	Prob> chi2 = 0.0000

logprice	Coefficients. Std. Error t-statistics P>|t|
Beds	.2403133 .0008111 296.26 0.000
entire_home	1.071521 .0045411 235.96 0.000
private_room	.3887257 .0043825 88.70 0.000
numberofreviews	-.0043356 .0000291 -148.82 0.000
manhattan	.26721 .0095624 27.94 0.000
Bronx	-.4187497 .015946 -26.26 0.000
queens	-.3024116 .0089951 -33.62 0.000
Brooklyn	-.3447763 .0092347 -37.33 0.000
House	-.1534615 .0040767 -37.64 0.000
other_property	-.0087894 .0046435 -1.89 0.058
Constant	3.948592 .0090625 435.71 0.000

The variables are statistically significant at 1% and 5%. Variance-weighted least-square provides more robust coefficients compared to OLS as well test statistics. In the difference, the major difference has occurred on the signs of some variables which might be associated with ordering variables during regression. Furthermore, the study has employed the feasible generalized least square regression which takes into the problem of heterogeneity in which the OLS cannot resolve. It has superior properties over ordinary least squares; the results in table 6 substantiate the argument.

4.4 Feasible Generalized Least Square Regression

Table 6. Presents the Results from Feasible Generalized Least Square Regression

Source	SS df MS	Number of obs = 30,393 F(10, 30382) = 3565.82
Model Residual	5691.01862 10 569.101862 4848.94056 30,382 .159599123	Prob> F = 0.0000 R-squared = 0.5399
Total	10539.9592 30,392 .346800447	Adj R-squared = 0.5398 Root MSE = .3995

logprice	Coefficients. Std. Error t-statistics P>|t|
Beds	.1432262 .0033927 42.22 0.000
entire_home	.9275557 .0110489 83.95 0.000
private_room	.294603 .0106561 27.65 0.000
numberofreviews	-.0009172 .0000988 -9.28 0.000
manhattan	.5301429 .0282583 18.76 0.000
Bronx	.0071345 .0321937 0.22 0.825
queens	.1110801 .0285751 3.89 0.000

Brooklyn	.2180231 .0280821 7.76 0.000
House	-.0183823 .0086775 -2.12 0.034
other_property	.1584673.0119094 13.31 0.000
Constant	3.633216 .0302117 120.26 0.000

Lastly, the study presents the results from F test statistics. F statistics show that the independent variables were statistically highly significant (Independent variables are jointly significant) since p-value associated with the F-statistic is less than the significance level (0.0000<0.05). Sample data provide sufficient evidence to conclude that the regression model fits the data better than the model with no independent variables. It implies that the independent variables in the model improve the fit.

Table 7. Presents F-Test Statistics Results

F(10, 30382) = 3565.82
Prob> F = 0.0000

4.5 Test for Heteroscedasticity

Scatter Plot

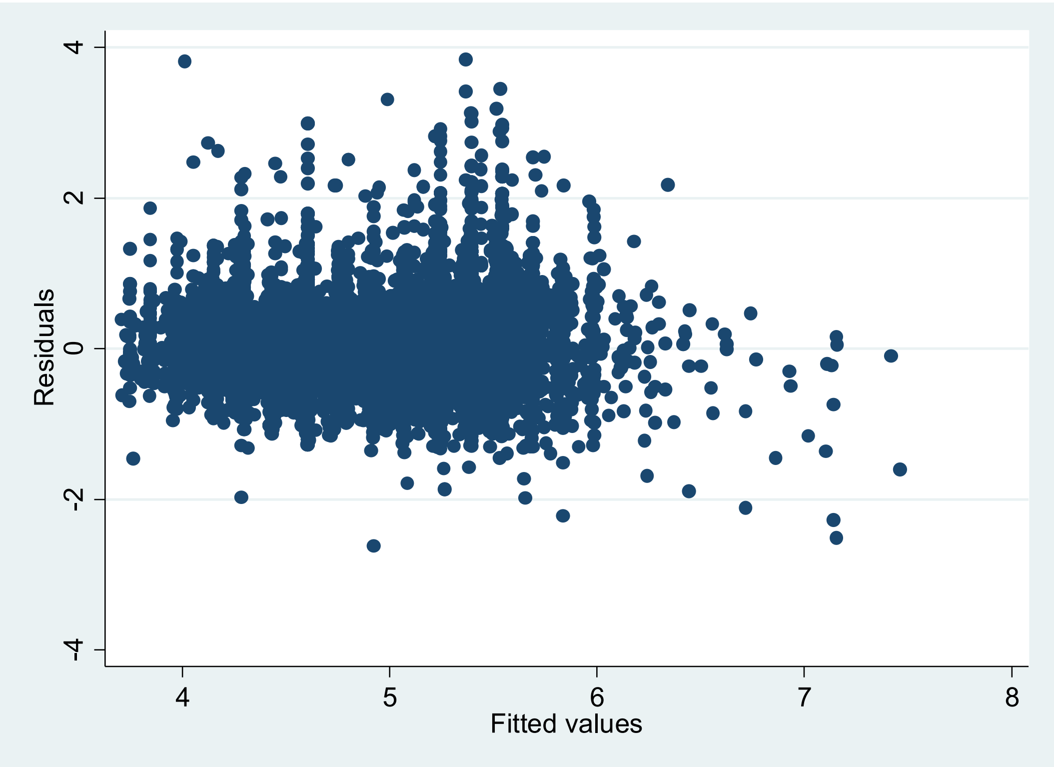

Figure 1. Presents the Results the Scatter Plot for Testing the Presence of Heteroscedasticity

The Figure above shows that the vertical spread of the residuals is relatively low among students with lower predicted reading scores. However, as we move left to right and the predicted reading scores increase, we see the spread of the residuals also increasing. The resulting image appears like a cone or fan that is spreading out as we move from left to right in the figure. This means that the variance of the residuals is not constant and, thus, we appear to have evidence of heteroscedasticity.

Table 8. Presents the Results of the Breusch-Pagan Test for Heteroscedasticity, with a Test Statistic

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance Variables: fitted values of logprice chi2(1) = 1139.01 Prob> chi2 = 0.0000

Table 8 presents the results of the Breusch–Pagan test for heteroscedasticity, with a the test statistic of 1139.01. When compared to a Chi-Squared distribution with one degree of freedom, the resulting P-value falls well below the standard .05 level. Thus, we have clear evidence to reject the null hypothesis of homoscedasticity and accept the alternative hypothesis that we do in fact have heteroscedasticity in the residual of this regression model.

5. Conclusion

This study confirms that the factors related to the number of beds, property type, number of reviews, room type, neighborhoods, and customer reviews also significantly influence the prices of sharing economy-based accommodation. However, entire homes/apartments are more expensive than private rooms. The study has also indicated with the increase in the number of beds the prices of accommodation tend to increase.

Practically, the study gives an imperative hint to the stakeholders such as accommodation rental suppliers to evaluate their market condition and improve their proceeds. Furthermore, the study enlightens the sharing economy-based accommodation rental platforms such as Airbnb to strategize tools to monitor suppliers for pricing grounded on the up-to-date price determinants. The information is useful for urban planners to avoid congestion, that accommodation nearby/closer city center tend to attract more people than those in the vicinity it, therefore, calls them to improve eco-social infrastructure so as to avoid.

Acknowledgement

I would like to extend my thanks to Prof Makram El-Shagi from Henan University for giving me the access to the data also my thanks to Dr Cornel Joseph from University of Dar es Salaam for his support in running the data. Lastly, mythanks goes to my family their moral and financial support in accomplishing this article.

References

Amenyah, I. D., & Afenyi, E. F. (2013). Factors determining residential rental prices. *Asian Economic and Financial Review*, *3*(1), 39-50.

Balck, B., & Cracau, D. (2015). *Empirical analysis of customer motives in the shareconomy: A cross-sector alcomparison* (Working Paper No. 2/2015). Otto-von-Guericke University Magdeburg. Retrieved from http://www.fww.ovgu.de/fww_media/femm/femm_2015/2015_02-EGOTEC

Cebula, R. J. (2009). The hedonic pricing model applied to the housing market of the city of Savannah and its Savannah historic landmark district. *The Review of Regional Studies*, *39*(1), 9-22. https://doi.org/10.52324/001c.8197

Daniel, G. (2015). *Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector, Current Issues in Tourism*, *18*(12), 1192-1217. https://doi.org/10.1080/13683500.2013.827159

Dogru, T., & Pekin, O. (2017). *What do guests value most in Airbnbaccommodations*? An application of the hedonic pricing approach, Boston Hospitality Review, AHospitality Industry Journey, Spring 2017, Volume 5, Issue 2

Guttentag, D. (2015). Airbnb: Sisruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192-1217. https://doi.org/10.1080/13683500.2013.827159

Hill, D. (2015). *The secret of Airbnb's pricing algorithm*. *IEEE Spectrum*, *20*.

Hung, W. T., Shang, J. K., & Wang, F. C. (2010). Pricing determinants in the hotel industry: Quantile regression analysis. *International Journal of Hospitality Management*, 29. https://doi.org/10.1016/j.ijhm.2009.09.001

Kain, J. F., & Quigley, J. M. (197*0).* Measuring the Value of Housing Quality. *Journal of the American Statistical Association*, *65*(330), 532-548. https://doi.org/10.1080/01621459.1970.10481102

Leoni, V., & Nilsson, J. O. W. (2020). *Dynamic pricing and revenues of Airbnb listings: Estimating heterogeneous causal effects* DEA WP no. 92 Working Paper Series. https://doi.org/10.1016/j.ijhm.2021.102914

Monty, B., & Skidmore, M. (2003). Hedonic pricing and willingness to pay for bed and breakfastamenities in Southeast Wisconsin. *Journal of Travel Research*, *42*. https://doi.org/10.1177/0047287503257500

Portolan, A. (2013). Impact of the attributes of private tourist accommodation facilities onto prices: A hedonic price approach. *European Journal of Tourism Research*, *6*(1), 74. http://www.mdpi.com/journal/sustainability ttps://doi.org/10.54055/ejtr.v6i1.117

Quinby, D., & Gasdia, M. (2014). Share this! Private accommodation and the rise of the new gen renters. Retrieved from http://www.phocuswright.com/Travel-Research/Consumer-Trends/Share-This-Private-Accommodation-amp-the-Rise-of-the-New-Gen-Renter.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, *36*, 2843-2852. https://doi.org/10.1016/j.eswa.2008.01.044

Suárez, C., M.Pérez, R. R., & Meliss, N. M. (2017). A*pplications of Regression Models in Epidemiology*. First Edition John Wiley & Sons, Inc. https://doi.org/10.1002/9781119212515

Toan, N. B. (2020). *A study of factors influencing the price of apartments: Evidence from Vietnam*. Management Science Letters, Canada.

Yang, Y., Mueller, N. J., & Croes, R. R. (2016). Market accessibility and hotel prices in the Caribbean: The moderating effect of quality-signaling factors. *Tourism Management Elsevier*, *56*(C), 40-51. https://ideas.repec.org/a/eee/touman/v56y2016icp40-51.html https://doi.org/10.1016/j.tourman.2016.03.021

Yayar, R., & Demir, D. (2014). Hedonic Estimation Of Housing Market Prices In Turkey. *Makaleninkabultarihi: Temmuz*, *43*, 67-82. https://doi.org/10.18070/euiibfd.42448

Ye, P., Qian, J., Chen, J., Wu, C. H., Zhou, Y., De Mars, S., & Zhang, L. (2018, July). *Customized Regression Model for Airbnb Dynamic Pricing.In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 932-940). ACM. https://doi.org/10.1145/3219819.3219830*