Determinants of Boston Airbnb Listing Prices

Sun, Yitong

yitong.sun@tufts.edu

Tufts University

Instructed by Jeffrey Zabel

Abstract

Peer-to-peer rental service platforms such as Airbnb have become a crucial part of the short-term rental industry. In this paper, I examine the causal relationship between features such as the total number of reviews and the prices for listings in the city of Boston, MA. Since the pandemic has greatly impacted much of the industry, I also explore how these features have influenced the prices from October to November 2020. By running the Ordinary Least Square model, I discovered that the number of reviews negatively impacts the listing price. Given the results from the sentiment analysis, I conclude that the quality of the reviews does not significantly influence the listing price, as the range of review scores is relatively narrow and not diverse enough to have a significant effect on price. The feature importance of the variables included in the OLS model changed during the fall-to-winter transition, and the overall trend of the prices shifted as well.

Determinants of Boston Airbnb Listing Prices

The short-term rental business has become a very crucial player in the housing rental business in Boston, especially under the fast-growing sharing economy (Gutierrez et al., 2016, Zervas et al., 2014). Many asset-light platforms that connect the service provider and consumer have emerged. Among such platforms, Airbnb is the most well-known example of a peer-to-peer short-term rental service business. With the Airbnb dataset retrieved from InsideAirbnb[1]. we can explore the pattern of house listings and the correlation between price and other variables that could affect the price, such as location (e.g., downtown, Back Bay), the number of reviews per month, and minimum number of nights required by the house owner.

Therefore, in this study, I ran a multiple-regression model to explore the relationship between the quantity of reviews per Airbnb listing and the listing price. Additionally, I aim to validate my hypothesis of the determinants.

Many factors have proven to be determinants of the listing price. According to Kyle (2019), when Airbnb listings in a city increase, so do rent prices. Intuitively, the location is an important factor in pricing for both long-term rental for students or employees and short-term rental for visitors or travelers. Eyal and Aliza (2016) discovered that the review score in the Airbnb user interface, surprisingly, is not correlated with customers' decision making or the price range set by the house owner, which is likely due to low variance in the review price range. With the dataset, I examined if such a pattern exists in the given data and analyzed other factors such as location, quality of reviews, and number of reviews through a linear regression model and sentiment analysis.

Since the nature of P2P short-term rental service differs from traditional short-term rental services, the price indicators for the various aspects of Airbnb rentals are also quite distinct from those of the traditional hospitality industry. For example, customers would pay more for a more personalized experience, which is not easy to find in the traditional hospitality industry. Therefore, it is important to learn the determinants of Airbnb listing prices and gain insights from the study.

---

[1] http://insideairbnb.com/get-the-data.html. Complied in October 2020.

This not only helps us to understand the P2P hospitality market, but also enables new Airbnb hosts to set optimal prices for their listings. Another aspect that I want to explore is how the price and price determinants have changed during the pandemic. There are over 100,000 listings all over the world on Airbnb as of 2019[2]. I would like to see how the pandemic affects business and customers' spending patterns at the beginning of November.

Ever since the COVID-19 pandemic erupted around the world, the tourism and hospitality industries have been greatly impacted (Richard T. & Jinah P. 2020). The virus was initially detected in the Chinese city of Wuhan in early December (Shi Z., 2020): According to the World Hospitality Organization (2020), "COVID-19 spread rapidly and by May 2020 there were over 3.85 million reported cases of infected people and 270,000 reported deaths." As a result, governments around the world started to place strict restrictions on travel, while closing their borders to suspend international and domestic travels in some cities. This has no doubt been a hard hit on global tourism and the short-term rental business.

Airbnb had originally aimed for an IPO in 2020, but the pandemic postponed the plan. Before that, starting from January 1, 2019, short-term rentals in Boston need to register with the city of Boston (Boston gov). Under the city's new regulations, all home-sharing companies, including Airbnb, require hosts to register their listings with the city and suspend the listing unit that was designed for short-term rental and pushing for individual house listing exclusively, since the former would be an inflation factor in the rental market. This new change may lower the total number of listings and inflate the overall short-term rental prices, given the results of Kyle's study (2019) mentioned earlier.

The present study focuses on the review-related determinants (i.e., the number and quality of the reviews) that have a potential and significant impact on Airbnb listing prices. It then compares the regression results for October and November 2020 to find any trend changes in response to the COVID-19 pandemic spread throughout the winter season, given that researchers believe the pandemic will likely worsen in winter, especially in regions such as Boston, where the virus is not fully under control (Smriti Mallapaty, 2020). Thus, this paper tests the following

---

[2] https://news.airbnb.com/fast-facts/

hypotheses: (1) Both the number and the quality of the reviews have a positive impact on the Airbnb listing price. (2) The coefficient for some variables changed negatively from October to November 2020.

The rest of this paper is organized as follows: In the "Literature Review" section, I review four studies conducted based on determinants and short-term rental listing prices in addition to a study analyzing how trustworthiness plays a part in the market. In the "Data" section, I present the data summary and processing for both the listing and review data, as well as other techniques used in this study. In the "Model" section, I present the main model used to generate the primary results for the number of reviews, the revised model for the number of reviews, and the model to verify other binary features. I also explain the models and present my hypotheses. The subsequent section presents the empirical analysis of the regression modeling results and discuss the outcomes of the hypothesis tests. A summary of this paper and suggestions for future works are presented in the conclusion.

Literature Review

The study that Dan Wang and Juan L. have conducted is mostly related to our investigations here. They built ordinary squares least model and conducted quantile regression analysis in order to investigate the determinants which are in five categories, host attributes, site and property attributes, amenities and services, rental rules, and online review ratings and the relationship between the determinants and listing price under short-term rental peer-to-peer service platform. From Dan's regression results, it is concluded that superhost1 status leads to higher listing price. Since the super host status is an indicator of reputation, we could expect a higher review score to cause higher score as well. And the QR estimate gives a similar result except the former gives a richer evidence across quantiles.

For our goal of giving the new hosts insights on setting optimal listing price, Chris and Daniel raised a claim about the list price suggesting algorithm might potentially cause issue for new hosts due to the lack of transparency in price setting algorithm for Airbnb, which make it confused for the new hosts to set the optimal price. Chris and Daniel thus applied a hedonic model on Airbnb listings for five metropolitan areas of Canada. From their modeling results, they confirmed the theory and further found out that more reviews are associated with a drop in price. Also, from their hedonic model run on five areas in Canada, they conclude that the traditional factors affecting the listing most significant even though Airbnb is promoting their innovative and unique features.

There is another study that is closely related to the one I'm conducting here is by Eyal and Aliza (2016). In the study, the authors analyzed the impact trustworthiness factors, such as owner's profile picture, have on the price and customer's decision making. In Eyal and Aliza's study, it is also suggested that the review score on Airbnb listing does not have significant impact on customer's decision making or the determination of listing price. Eyal explained in the article that it might be caused by the generally high review score and low variance in the review score, since there are 97% of the hosts receiving review scores between 4.5 and 5 stars (out of 5), which further provides less information for the customer in decision-making.

The study results revealed two main results. First, an increase in one unit of the visual-based trust score (scale of ten) leads to an increase of approximately seven percent in the price of the listing. Second, price increases if the listing features an entire apartment (rather than a private or shared room). Price also increases with number of rooms in each listing. I will reproduce the latter result in the regression model later. In the study conducted by Edelman and Luca (2014), it is found that personal photos might worsen racial discrimination: Non-black hosts in New York City charge higher prices than black host and is likely caused by their awareness that the profile pictures reveal the race of host. Given the study results, hosts who are perceived from their photos as more trustworthy charge higher prices than their counterparts who are perceived as less trustworthy. Even though it is stated that review score has no significant impact on either customer's decision or listing price in the study, I would like to conduct the model to test if the number of reviews per month will be a determinant on rental price. From the intuition, the two factors are correlated since higher review rate implies higher trustworthy rate. And the number of reviews does variate largely from listing to listing from the dataset, which will diminish the negative effect caused by low variance.

In Anna and Cristina's study, they examined the impact of COVID-19 on peer-to-peer accommodation platforms. As mentioned earlier, the peer-to-peer accommodation platforms is generically different from conventional accommodation service platform like hotels due to the distinct nature of P2P accommodation. The pandemic is a huge strike to the overall tourism industry and is inevitably be a disruptive impact on Airbnb. Authors conducted interviews with short-term rental hosts selected via purposive sampling technique. It is confirmed that 100% of the interviewees states that their business is negatively impact by the pandemic and directly impacted by the measures undertaken by government under pandemic. Some interviewees choose to lower the listing price to attract new guests and some choose to raise the price in order to cover the lost profit, but overall, the listing price as a whole would be fluctuating during the pandemic. And the majority of interviewees believes that they didn't receive the enough support from the p2p platform leading 'host resentment' (Johnson and Davis, 2020). On the other hand, there are certain percentage of hosts who are identified as 'pessimistic host' considering exiting the short-term rental market. Therefore, we would expect a decrease (not drastically) loss of listings from October to November 2020.

Data

From InsideAirbnb I obtained two data sets compiled in October and in November 2020. Each data set contains the listing data and reviews data. The listing data contains features that could be related to the listing price and thus have great potential on helping build an regression model. The list data is covering all the active listings in October 2020 on a daily basis.

The definitions of independent variables are included in Table 1. By generating summary statistics and plotting the distribution of variables, I was able to identify and removed some outliners such as listing that is marked as zero USD per day and as 9999 USD per stay. After removing the outliners, a descriptive statistics summary is generated and included in Table 2.

After plotting the latitude and longitude with respect to the listing price, it is showing that the price for longitude and latitude does not follow a linear relationship, but the prices are clustered. Thus, the latitude and longitude features should not be included in the regression model. As substitutes, and to further take account for location variable, there is neighborhood feature which I further convert into multiple binary dummy variables[3] and include in the model. The next step is to check whether there is any multicollinearity issue. To check that I plot a heatmap (figure.1). From the figure, we see that number of reviews and reviews per month are drawing concerns as these two variables are shown to be moderately correlated for a correlation coefficient magnitude of 0.61. Additionally, there are 890 null values in the 'review per month' column, which is mostly likely because there are zero review for October for the listings. Therefore, I dropped reviews per month and last review from the dataset. For the last part of feature engineering, in order to make our regressor more normally distributed, I take logarithm of the listing price. Figure 2 shows that the number of reviews and log of price are basically normally distributed.

I include essential numerical independent variables such as neighborhood, room type, price and number of reviews in the regression model. There is total 3254 listing entries and 122 review entries. In listing data, there are 893 null values in reviews per month variable, thus I replace the

---

[3] 'Allston', 'Back Bay', 'Bay Village', 'Beacon Hill', 'Brighton', 'Charlestown', 'Chinatown', 'Dorchester', 'Downtown', 'East Boston', 'Fenway', 'Hyde Park', 'Jamaica Plain', 'Leather District', 'Longwood Medical Area', 'Mattapan', 'Mission Hill', 'North End', 'Roslindale', 'Roxbury', 'South Boston', 'South Boston Waterfront', 'South End', 'West End', 'West Roxbury'

null value with mean to achieve better accuracy while keeping the data integrate. All information is compiled from Airbnb website and has been done primary data cleaning and processing.

For the review's dataset, there are 122,624 entries ranged from 2015 through 2020. Since there are only 2,003 listing entries after processing as of October 2020, there are multiple reviews for each listing under estimation. The dataset includes each listing id and the comments in the review section (string variable). The problem here is that the data does not provide stars or review score directly for us to include in the model. In order to understand the quality of reviews and the impact the review quality could impose on listing price; I did the sentiment analysis on the review's dataset.

Sentiment analysis is also known as opinion mining, which are conducted based on Natural Language Processing techniques in order to understand and quantify the emotions behind literal strings (Xing F. & Justin Z., 2015). Xing and Justin conducted a sentiment analysis on the reviews for a set of Amazon products with the similar technique that I will conduct in my study. To start with, I clean the review data by removing rows containing null values and non-English values. And then, the comments for each entry are unified with same lower-case format. By applying tokenization method and filtering the tokens with stopwords[4] to remove tokens that does not have sentimental meaning by themselves but are of high frequency, such as 'the', 'my', 'in'. After the processing, the comments strings are converted to meaningful tokens and are ready for the sentiment score computation.

[4] https://www.nltk.org/_modules/nltk/corpus.html

Model

The base model I estimate is as follows:

$$\ln(PRICE_{it}) = \beta_0 + \beta_j X_{it} + u_i$$

For the first part of the hypothesis, I included the independent variables in Table 3 as X_{it}. In the model, $\ln(PRICE_{it})$ is the neperian logarithm of price for the listing "i". β_0 is the fixed component which is independent. β_j is the parameter to be estimated, related to the feature "j". u_i is the error term associate with each entry. t is the different time period. We have the data compiled in October and in November 2020, thus I will apply the model to both time period separately and in aggregate for the further analysis of pattern change from October to November 2020. Since there are locations and room type binary variables included in the dataset, the model can also be modified by having some binary variable as left out group and then run the regression model. The results should be helpful for the inexperienced hosts to set a relatively optimal price given the location where the property is or the room type. Additionally, since we are unsure about the relationship between price and number of reviews, that they can be reversely causational. To prepare for that, here is OLS model (2) I estimated in order to identify the impact of price change on number of reviews:

$$\text{Number\_of\_Rev}_i = \beta_0 + \beta_j X_i + u_i$$

Where the number of reviews is set as dependent variable and logarithm of price is set as one of the regressors along with other regressors remaining the same as in the previous model.

For the sentiment analysis part, by using the python nltk package, the review scores, which are positive score, negative score and compounded score, can be generated for each comment. The higher the scores are, the stronger emotions are included in each comment for each category.

Empirical analysis

The first OLS model result is shown in Table 4 where column (1) is for data compiled in October and column (2) is for data compiled in November. As shown in Table 5, all four variables are statistically significant at 5% level. From Table 4, the results show that number of reviews, minimum nights requirement and calculated host listing counts all have negative impact on listing price, while availability has a positive impact on the listing price, while the variables are not economically significant. The results do not concur with my first hypothesis that the number of reviews has a positive impact on listing price. One possible reason is that more reviews indicates that reviews tend to be more negative when there are more reviews compared to when there are fewer reviews (Judith B. & Camila V. 2016). Another possible explanation is that there might exist a reverse causality between number of reviews and listing price. Thus, I run the OLS model (2) to validate my hypothesis. The results as shown in Table 5 is showing that the impact price imposes on number of reviews is both statistically significant and economically significant, thus the reverse causality exists.

It's also worth noting that the variable calculated host listings count is negatively correlated with the listing price. The management of host listings count has always been a contentious topic. Hosts with more listings can be identified as experienced Airbnb hosts and should be able to set a relatively optimal price compared to new host and host with less listings (Chris G., 2017). But Karen and Zhenxin argued with their study result that there exists a trade-off between host quality and total number of listings under host, which accounts for the negative coefficient for calculated host listings count and listing price (2017).

Robbin has shown in his study that the location factors such as transit accessibility to jobs and neighborhood variation have a large impact on listing price (2019). Thus, we assume that the neighborhood factors will have a large impact on Airbnb listing price in Boston. Intuitively, we assume price might have a negative impact on people's inventiveness of leaving comments of each stay or other decision making. But Mingming Cheng and Xin Jin found out that price is not a key influencer on decision making (2019), instead, their study suggests a positivity bias in Airbnb users' comments while negative sentiments are mostly caused by 'noise', which concurs with our regression results.

To further look at the sentiment analysis, I generate the neutral, positive and negative scores for all the comments via sentiment analyzer python package and plot the distribution of level of scores in Fig. 2, Fig. 3 and Fig. 4 correspondingly. The only score matters to our study is negative score since we don't need to know how unsatisfied customers are to stays, thus we will focus on the negative score distribution. However, the negative score is highly skewed to the right, indicating that there is very few people leave comments with disappointment for their stays. In Table 7, I generate the summary statistics for all four scores, and we can tell from the table that the standard deviation (and thus variance) for negative score is too low to be an indicator of the price variation.

The regression results for data compiled in November is also included in the Table 6. In Fig. 6, it also gives us a general idea of how price pattern changes from October to November. First, the curve is shrinks to the mean which means that the price is shifting to the mean. The expensive stays are getting cheaper and hosts owning cheap stays are charging more. This finding concurs with the Anna's interview study (2020). From the regression results, we can also tell that the minimum nights is losing impact over price compared with other variables, while calculated hosting listing count and availability 365 are having more impact over price during the pandemic. This also makes sense, because Anna also finds that visitors are more likely to choose a 'professional host' for those hosts hosting multiple listings rather than other factors when making decisions (2020). Host listing count is also an indicator of 'trustworthiness' for customer when making decisions according to Eyal's study on Airbnb host trustworthiness (2016).

Conclusion

In this study, I used OLS models, sentiment analysis, and data visualization methods to show the impact of determinants on listing prices and the impact change from October to November 2020. We can conclude the following: (1) The number of reviews does not have an economically significant impact on price, but price has an economically significant impact on the number of reviews. (2) Both the minimum number of nights required, and the calculated host listing count have a negative impact on listing price. (3) During the fall-to-winter season change, customers were more likely to choose stays with hosts who have more listings, as these owners may be more professional than those with fewer listings. (4) The reviews on Airbnb are generally positive and neutral. Thus, my study here is not completely fulfilled due to the variance of the negative review scores, which is not large enough to relate with the changes in listing price. In addition, the capability of sentiment analysis is limited, as over 122,000 entries in the review's dataset are beyond the processing capability supported by the Python package. In future studies, I will explore other Python packages to conduct further sentiment analyses, and I expect to integrate the sentiment analysis results with the existing OLS models to produce better results.

References

Anna F., Cristina M. & Maria H. (2020). Impacts of Covid-19 on peer-to-peer accommodation platforms: Host perceptions and responses. *International Journal of Hospitality Management Volume 91,* 102663. https://doi.org/10.1016/j.ijhm.2020.102663

Chris G. & Daniel G. (2017). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing Volume 35, 2018 - Issue 1: Shareable Tourism: Tourism Marketing in the Sharing Economy.* https://doi.org/10.1080/10548408.2017.1308292

Edelman, B., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. *Harvard Business School NOM Unit Working Paper*, (*14*-054). http://www.west-info.eu/files/airbnb_research.pdf.

Eyal E., Aliza F. & Nathan M. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management, 55*, 62-73. https://doi.org/10.1016/j.tourman.2016.01.013.

Gutierrez et al., Gutierrez, J., Garcia-Palomares, J.C., Romanillos, G., & Salas-Olmedo, M.H. (2016). Airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation. arXiv preprint arXiv: 1606.07138.

Judith B. & Camila V. (2016). If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism Volume 21, 2018 - Issue 18.* https://doi.org/10.1080/13683500.2016.1267113

Karen X. & Zhenxing M. (2017). The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management, Vol. 29 No. 9,* pp. 2240-2260. https://doi.org/10.1108/IJCHM-07-2016-0345

Kyle B., Edward K. & Davide P. (2019). Research: When Airbnb Listings in a City Increase, So Do Rent Prices. *Harvard Business Reviews.* https://hbr.org/2019/04/research-when-airbnb-listings-in-a-city-increase-so-do-rent-prices.

Mingming C. & Jin X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management Volume 76, Part A*. https://doi.org/10.1016/j.ijhm.2018.04.004

Raul V. & Leticia S. (2018). The What, Where, and Why of Airbnb Price Determinants. *Sustainability 2018, 10(*12*),* 4596. https://doi.org/10.3390/su10124596

Richard T. & Jinah P. (2020). Social costs of tourism during the COVID-19 pandemic. *Annals of Tourism Research Volume 84*, 102994. https://doi.org/10.1016/j.annals.2020.102994

Robbin D., Danielle J., David W. &Ahmed E. (2019). Listen ReadSpeaker webReader: Listen Focus Article Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. *Regional Studies, Regional Science, 6:1,* 143-156. https://doi.org/10.1080/21681376.2019.1592699

Smriti M. (2020). Why COVID outbreaks look set to worsen this winter. *Nature* 586, 653. https://doi.org/10.1038/d41586-020-02972-4 https://www.nature.com/articles/d41586-020-02972-4

Shi Z., Qianying L., Jinjun R., & Salihu S. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases Volume 92*, Pages 214-217. https://doi.org/10.1016/j.ijid.2020.01.050

Xing F. & Justin Z. (2015). Sentiment analysis using product review data. *Journal of Big Data 2, 5*. https://doi.org/10.1186/s40537-015-0015-2

Table 1. Definition of Independent Variables

Variable	Definition
price	The price of listing. Per night.
number of reviews	The total count number of reviews for each listing.
reviews per month	The total count number of reviews for each listing divided by the numbers of days in corresponding month.
calculated host listings count	The total count number of listings that each host owns.

Table 2. Summary Statistics for Numerical Variables

	count	mean	std	min	25%	50%	75%	max
price	2003.0	160.204194	162.836832	20.0	72.0	115.0	184.0	2514.0
minimum_nights	2003.0	8.719920	11.307921	1.0	1.0	2.0	14.0	30.0
number_of_reviews	2003.0	52.261608	77.793744	0.0	2.0	18.0	68.0	569.0
calculated_host_listings_count	2003.0	15.968547	22.417983	1.0	2.0	5.0	22.0	84.0
availability_365	2003.0	187.717923	129.346752	0.0	77.5	175.0	325.0	365.0

Table 3. Definition of Independent Variables Included in OLS Model

Variable	Definition
Minimum_nights	The minimum night that host requires to stay per lilsting
Number of reviews	Total number of reviews for each listing
Calculated host listings count	The total number of listings under a host
Availability_365	The available days in the given year
Allston	Rental property located in Allston
Back Bay	Rental property located in Back Bay
Bay Village	Rental property located in Bay Village
Beacon Hill	Rental property located in Beacon Hill
Brighton	Rental property located in Brighton
Charlestown	Rental property located in Charlestown
Chinatown	Rental property located in Chinatown
Dorchester	Rental property located in Dorchester
Downtown	Rental property located in Downtown
East Boston	Rental property located in East Boston
Fenway	Rental property located in Fenway
Hyde Park	Rental property located in Hyde Park
Jamaica Plain	Rental property located in Jamaica Plain
Leather District	Rental property located in Leather District
Longwood Medical Area	Rental property located in Longwood Medical Area
Mattapan	Rental property located in Mattapan
Mission Hill	Rental property located in Mission Hill
North End	Rental property located in North End
Roslindale	Rental property located in Roslindale
Roxbury	Rental property located in Roxbury
South Boston	Rental property located in South Boston
South Boston Waterfront	Rental property located in South Boston Waterfront
South End	Rental property located in South End
West End	Rental property located in West End
West Roxbury	Rental property located in West Roxbury
Entire home/apt	The room type which is entire home or apartment
Hotel room	The room type which is hotel room
Private room	The room type which is private room
Shared room	The room type which is shared room
Price_log	Price per night (including 20% of cleaning fee), log transformed

Table 4. Regression Results for October and November

	Dependent variable:	
	price_log	
	(1)	(2)
minimum_nights	-0.009***	-0.008***
	(0.001)	(0.001)
number_of_reviews	-0.001***	-0.001***
	(0.0002)	(0.0002)
calculated_host_listings_count	-0.003***	-0.005***
	(0.001)	(0.001)
availability_365	0.0002**	0.0004***
	(0.0001)	(0.0001)
Allston	-0.061	-0.034
	(0.126)	(0.130)
\`Back Bay\`	0.075	0.168
	(0.125)	(0.127)
\`Bay Village\`	0.054	0.071
	(0.141)	(0.142)
\`Beacon Hill\`	0.049	0.031
	(0.129)	(0.131)
Brighton	-0.155	-0.122
	(0.127)	(0.129)
Charlestown	0.431***	0.454***
	(0.135)	(0.139)
Chinatown	0.251*	0.285*
	(0.143)	(0.146)
Dorchester	0.112	0.128
	(0.117)	(0.120)
Downtown	0.098	0.127
	(0.120)	(0.123)

	(1)	(2)
`East Boston`	0.137	0.095
	(0.123)	(0.127)
Fenway	0.960***	0.942***
	(0.130)	(0.133)
`Hyde Park`	-0.337**	-0.301*
	(0.157)	(0.158)
`Jamaica Plain`	0.089	0.135
	(0.120)	(0.123)
`Leather District`	0.236	0.286
	(0.550)	(0.550)
`Longwood Medical Area`	0.377*	0.740**
	(0.222)	(0.332)
Mattapan	-0.213	-0.190
	(0.148)	(0.151)
`Mission Hill`	0.188	0.255*
	(0.143)	(0.151)
`North End`	0.176	0.151
	(0.141)	(0.143)
Roslindale	-0.090	-0.077
	(0.137)	(0.139)
Roxbury	0.094	0.093
	(0.122)	(0.124)
`South Boston`	0.399***	0.464***
	(0.125)	(0.128)
`South Boston Waterfront`	0.567**	0.723**
	(0.267)	(0.295)
`South End`	0.154	0.207*
	(0.122)	(0.124)
`West End`	0.585***	0.509***
	(0.159)	(0.169)
`West Roxbury`		
`Entire home/apt`	1.777***	1.215***
	(0.195)	(0.315)
`Hotel room`	2.088***	1.338***
	(0.224)	(0.333)
`Private room`	0.991***	0.473
	(0.195)	(0.315)
`Shared room`		
Constant	3.280***	3.747***
	(0.219)	(0.322)
Observations	2,003	1,990
R2	0.436	0.407
Adjusted R2	0.427	0.397
Residual Std. Error	0.538 (df = 1971)	0.538 (df = 1958)
F Statistic	49.185*** (df = 31; 1971)	43.264*** (df = 31; 1958)
Note:		*p<0.1; **p<0.05; ***p<0.01

*Column (1) is for data compiled in October and column (2) is for data compiled in November.

Table 5. T-Statistics Table for Variables (October)

Variables	t-statistics
minimum_nights	7.72
number_of_reviews	4.108
calculated_host_listing	4.672
availability_365	2.049

*all t-statistics values are absolute value.

Table 6. OLS model (2) Results Summary

```
Coefficients: (2 not defined because of singularities)
                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   68.09887   31.43085   2.167  0.03038 *  
minimum_nights                -1.20510    0.16463  -7.320 3.59e-13 ***
calculated_host_listings_count  -0.58666    0.08646  -6.786 1.52e-11 ***
availability_365               0.01398    0.01357   1.030  0.30334    
Allston                      -16.16614   17.19827  -0.940  0.34734    
`Back Bay`                    -6.63931   16.98628  -0.391  0.69594    
`Bay Village`                -22.22006   19.23149  -1.155  0.24807    
`Beacon Hill`                 20.99567   17.50955   1.199  0.23063    
Brighton                      -4.71373   17.25792  -0.273  0.78478    
Charlestown                  -16.09583   18.38054  -0.876  0.38130    
Chinatown                    -23.86496   19.46596  -1.226  0.22035    
Dorchester                   -14.37078   15.94551  -0.901  0.36757    
Downtown                     -27.79913   16.30775  -1.705  0.08842 .  
`East Boston`                 35.26901   16.75795   2.105  0.03545 *  
Fenway                        -1.46629   17.93910  -0.082  0.93486    
`Hyde Park`                  -23.64125   21.42506  -1.103  0.26997    
`Jamaica Plain`               -4.98140   16.33361  -0.305  0.76041    
`Leather District`           -51.66877   74.78660  -0.691  0.48972    
`Longwood Medical Area`      -47.00314   30.17389  -1.558  0.11945    
Mattapan                     -25.54138   20.18702  -1.265  0.20594    
`Mission Hill`               -21.00514   19.46517  -1.079  0.28067    
`North End`                   44.81988   19.22844   2.331  0.01986 *  
Roslindale                   -20.13896   18.58422  -1.084  0.27865    
Roxbury                        2.61730   16.54858   0.158  0.87435    
`South Boston`                 9.39013   17.10493   0.549  0.58309    
`South Boston Waterfront`     87.11491   36.37158   2.395  0.01671 *  
`South End`                    8.71458   16.57538   0.526  0.59912    
`West End`                   -12.41000   21.67714  -0.572  0.56705    
`West Roxbury`                      NA         NA      NA       NA    
`Entire home/apt`             70.99124   27.08035   2.622  0.00882 ** 
`Hotel room`                  40.61008   31.05886   1.308  0.19119    
`Private room`                58.78376   26.71612   2.200  0.02790 *  
`Shared room`                       NA         NA      NA       NA    
price_log                    -12.53539    3.05112  -4.108 4.15e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 73.14 on 1971 degrees of freedom
Multiple R-squared:  0.1297,	Adjusted R-squared:  0.116 
F-statistic: 9.478 on 31 and 1971 DF,  p-value: < 2.2e-16
```

Table 7. Review Score Summary Statistics

	count	mean	std	min	25%	50%	75%	max
compound	110752.0	0.825133	0.249616	-0.9976	0.8016	0.9136	0.9606	0.9994
negativity	110752.0	0.013045	0.034852	0.0000	0.0000	0.0000	0.0000	1.0000
neutrality	110752.0	0.651825	0.151399	0.0000	0.5630	0.6720	0.7570	1.0000
positivity	110752.0	0.335112	0.157926	0.0000	0.2250	0.3160	0.4300	0.9280

Figure 1. Heatmap for Numerical Variables

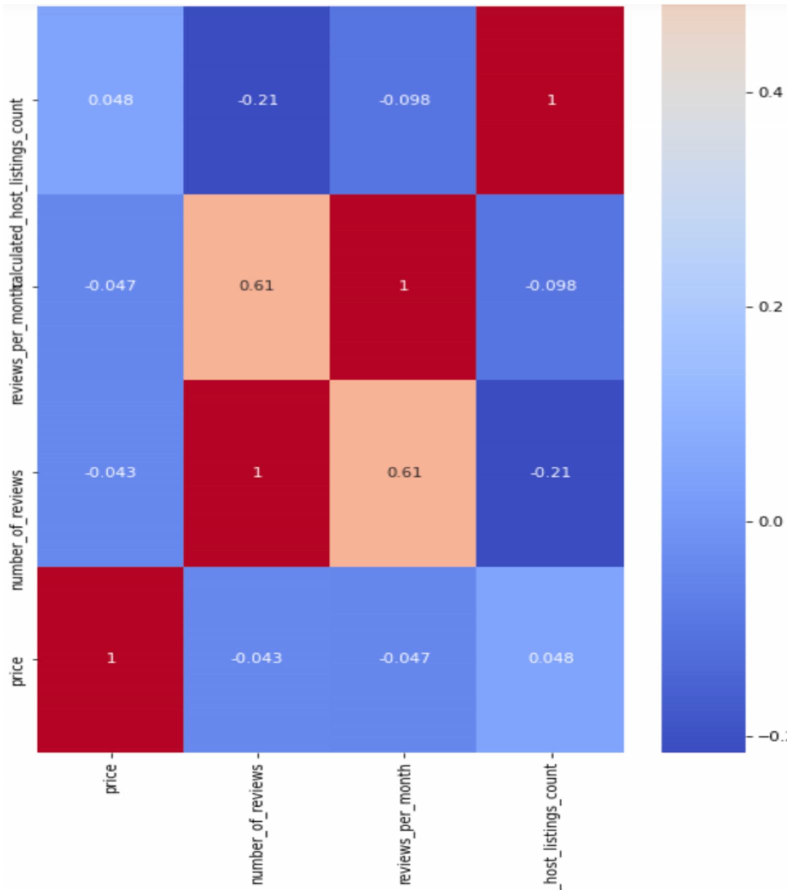
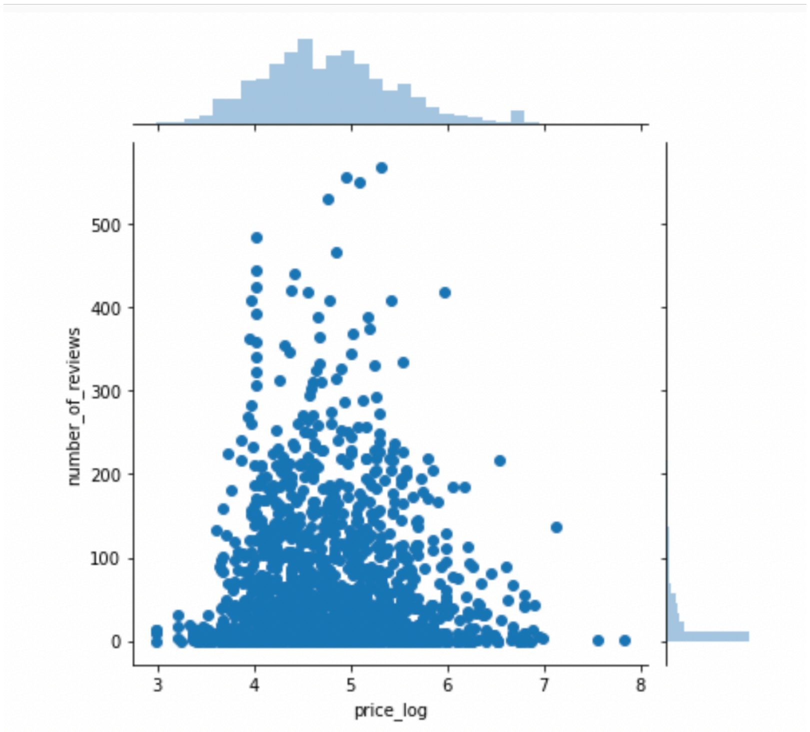
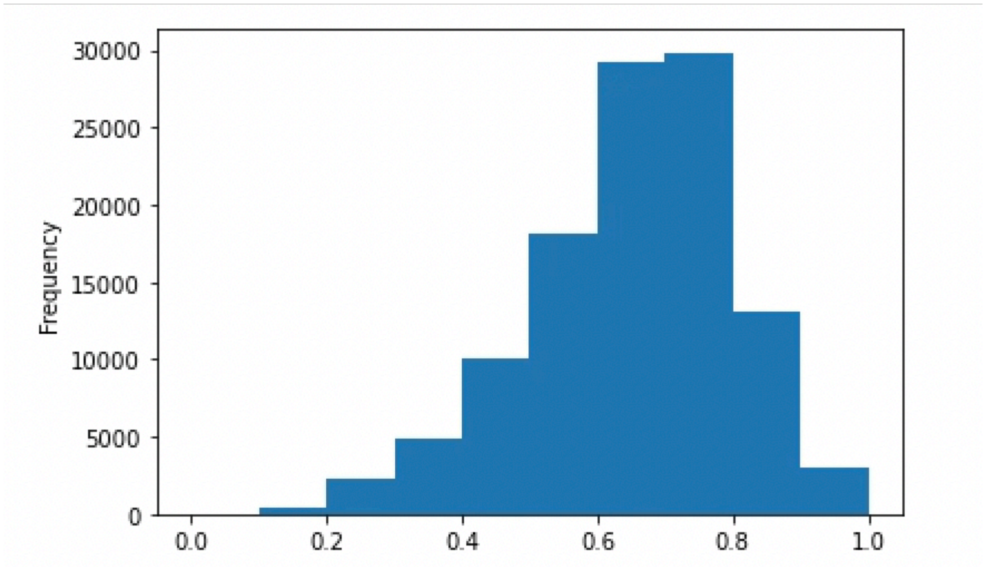


Figure 2. Distribution of Logarithm of Price over Number of Reviews



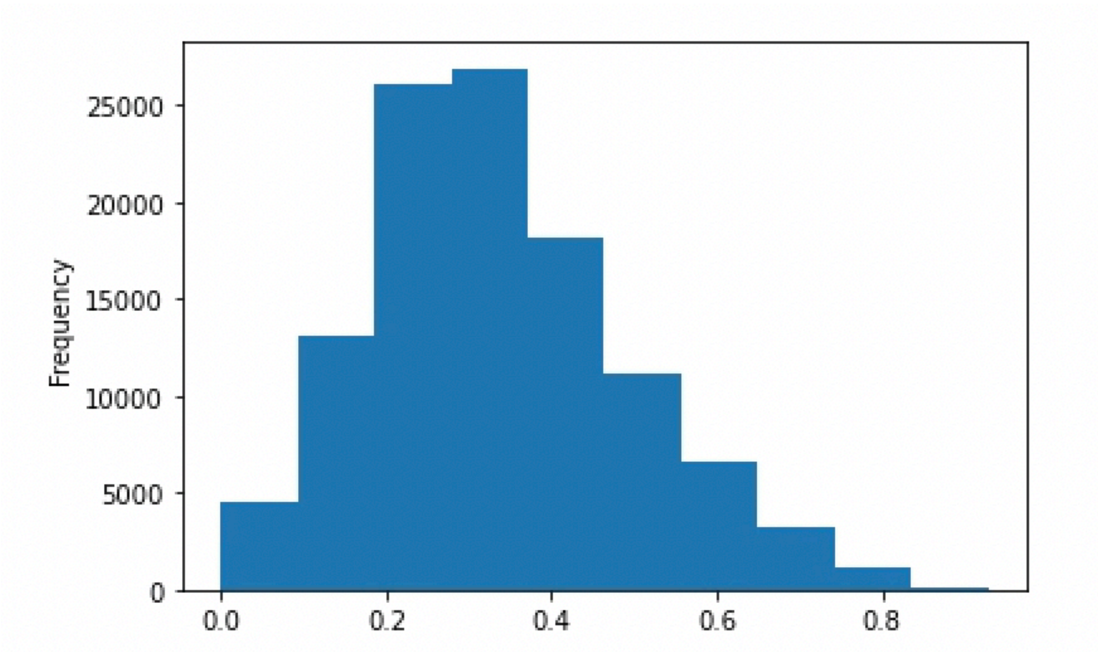
*unit for price_log is the logarithm of price in USD.

Figure 3. Neutral Score Distribution



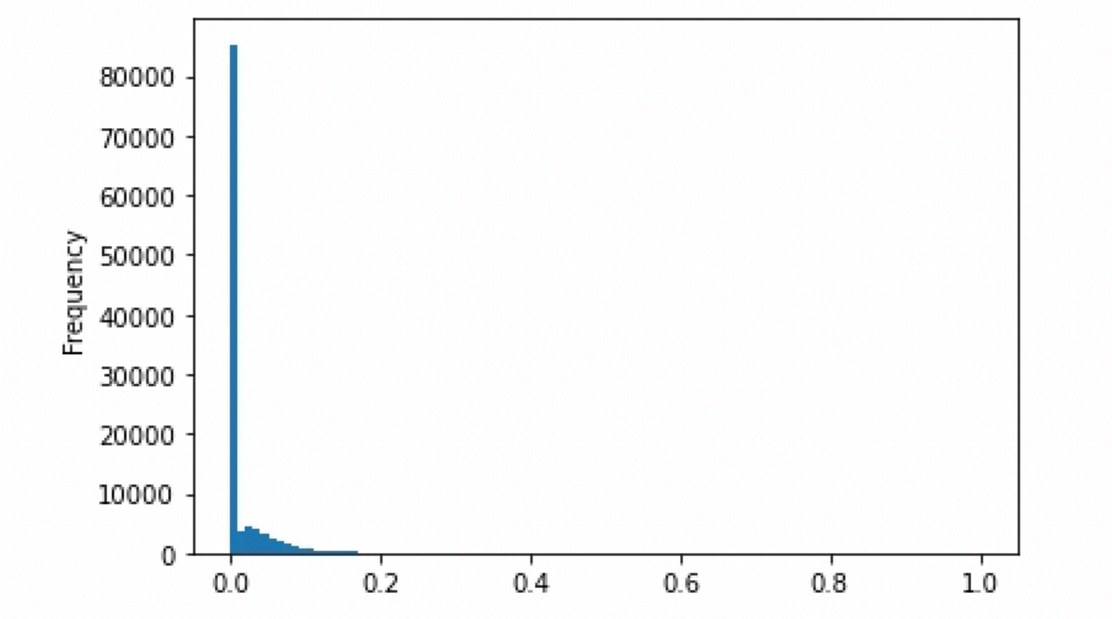
*x-axis is the neutral score.

Figure 4. Positive Score Distribution



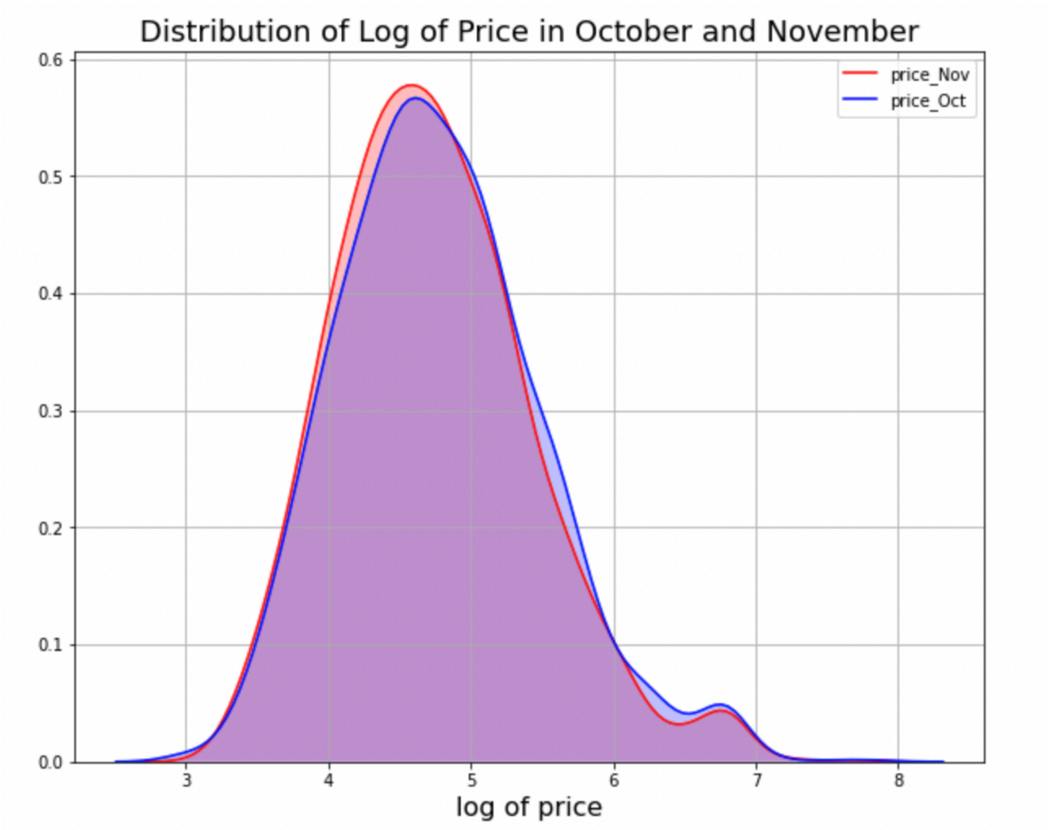
*x-axis is the positive score.

Figure 5. Negative Score Distribution



*x-axis is the negative score.

Figure 6. Distribution of Logarithm of Price in October vs. in November



*y-axis is the frequency of listings with respective to listing price.