Airbnb Listings Analysis

Priyanka Bijlani, Sharmeelee Bijlani, Harsha Koopanaraju, Hrishi Kulkarni, Lakshmi Venkatasubramanian

Abstract

Airbnb operates an online marketplace for hospitality services [1]. One such service they offer is lodging. Hosts can make a *listing* of their lodging on Airbnb's website, where *users* can find lodging for future travel plans and review listings they have visited. In this analysis, we assess the association between the price per night of a listing based upon a variety of factors and attempt to determine whether the impact of said factors changes based upon the city market using a Poisson regression model.

Introduction

Airbnb currently has over 7 million listings worldwide, over 100,000 cities with listings and over 220 countries and regions with listings [4]. These listings can cover a variety of property types (e.g., whole houses, apartments, private rooms, etc.) and types of stays (e.g., short-term, long-term, etc.). Hosts can fill out structured information for users to filter and evaluate listings (e.g., number of bedrooms, property type, etc.). They can add additional unstructured data on the listing as well, such as a listing description. Given all of this information and the user's own preferences, they can decide whether or not to reserve the booking at the advertised price per night.

There are many questions that could be asked about this process, but given the data and time restrictions we had during this project, we have focus our analysis on the two following questions:

- What factors have an effect on the advertised price per night of an Airbnb listing?
- Do the factors which have an effect on the advertised price per night of an Airbnb listing vary across different cities?

Data Set Description

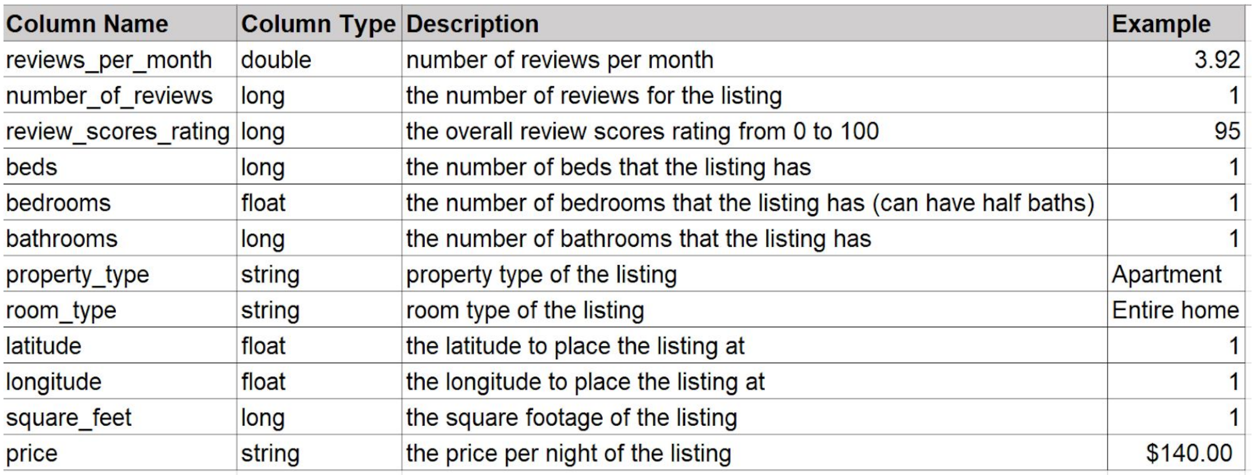
The data used for this analysis is information available to the public collected, cleaned and compiled by Inside Airbnb for the purpose of exploring how Airbnb is used in different cities [2]. The data is observational: the data is scraped from Airbnb's website at a particular snapshot in time. All the address information is anonymized by altering the exact location to be within 450 feet of the actual address. Spam reviews are not filtered out in the dataset as they are allowed by the Airbnb website, which is the original source of data.

The datasets used in this analysis were compiled by Inside Airbnb on the following dates:

- ❖ Seattle – November 21, 2019
- ❖ Boston – December 4, 2019

The datasets we used in this analysis were the listings files with 106 variables including our chosen response variable of price. We reduced the data for analysis to 11 predictor variables as seen in *Table 1: Variables of Interest* (see *Exploratory Data Analysis* for the city-specific exploration of these predictor variables).

Table 1: Variables of Interest



Column Name	Column Type	Description	Example
reviews_per_month	double	number of reviews per month	3.92
number_of_reviews	long	the number of reviews for the listing	1
review_scores_rating	long	the overall review scores rating from 0 to 100	95
beds	long	the number of beds that the listing has	1
bedrooms	float	the number of bedrooms that the listing has (can have half baths)	1
bathrooms	long	the number of bathrooms that the listing has	1
property_type	string	property type of the listing	Apartment
room_type	string	room type of the listing	Entire home
latitude	float	the latitude to place the listing at	1
longitude	float	the longitude to place the listing at	1
square_feet	long	the square footage of the listing	1
price	string	the price per night of the listing	$140.00

Exploratory Data Analysis

In this section, we will walk through a variety of issues in handling and analyzing the data set. We first use Seattle data to explain the problems, and then present a summary of Seattle and Boston data at the end. Note for the EDA charts (see Appendix): the Airbnb data for the city of Seattle contains 9,023 listings; the Airbnb data for the city of Boston contains 3,903 listings.

Cleaning the Data for Analysis

Issue #1: Unclean price data

A minor issue when using the data set is that the price has to be converted into a usable numerical quantity to analyze. In order to do this, we dropped any '$' characters from the price and converted them to float values in Python.

Issue #2: Missing values in other variables

A handful of variables reported missing values in rows, as seen in *Table 2: Missing Values.*

Table 2: Missing Values

Variable Name	Seattle		Boston	
	# Rows Missing Values	% Missing values	# Rows Missing Values	% Missing values
reviews_per_month	1261	~14%	684	~19%
review_scores_rating	1320	~15%	694	~20%
beds	3	<1%	5	~1%
bedrooms	7	<1%	3	<1%
bathrooms	2	<1%	2	<1%
square_feet	8620	~96%	3383	~96%

Looking across both cities, we can see that there are relatively few listings missing *beds*, *bedrooms*, or *bathrooms*, but there are quite a few missing values in *reviews_per_month* and *review_scores_rating*, and a substantial amount in *square_feet*.

Thus, we opted to remove *square_feet* from our analysis. For the remaining variables, we removed any rows where those values were missing for ease of analysis. However, as we discuss later in the *Discussion* section, this handling of missing values could be revisited in later analysis to handle this case more robustly (e.g., imputing the missing values, finding suitable proxies for them, etc.).

After removing the *square_feet* variable and removing any rows with missing values, our dataset for Seattle contained 7,697 listings and Boston contained 2,809 listings.

Issue #3: Outliers & Skewness

We plotted boxplots to visualize the distribution of our variables of interest, see *Figure 1A: Boxplots of Variables of Interest for Seattle* below (for Boston, see *Figure 1B: Boxplots of Variables of interest for Boston* in *Appendix*).

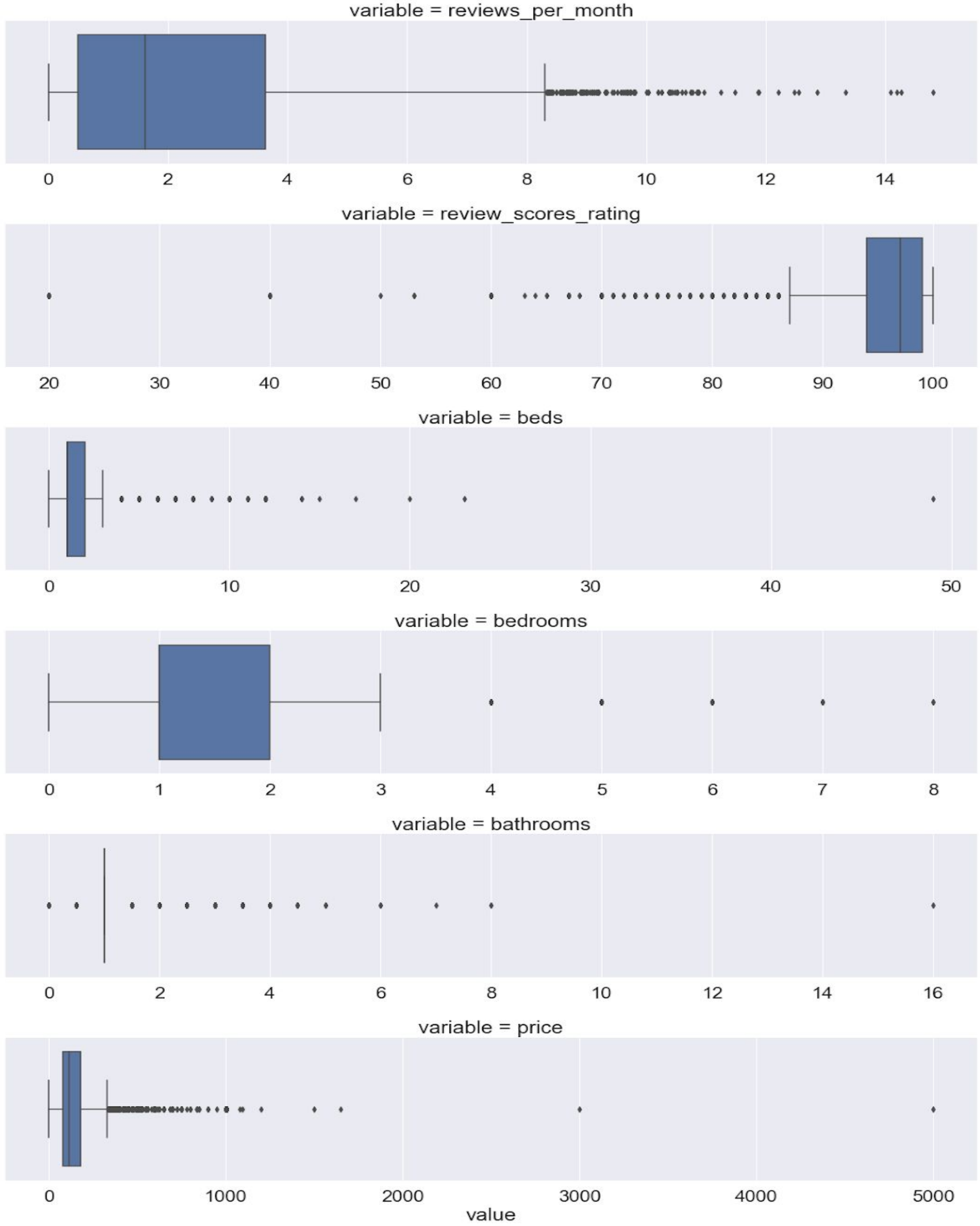
Using these boxplots, we can see a variety of the variables of interest have skewed distributions. Outliers in *review_scores_rating* and *reviews_per_month* may indicate users feel strongly about a particular listing -- and this may reflect in price. However, we made the judgement on intuition that the outliers in *beds*, *bedrooms* and *bathrooms* may not provide a meaningful signal towards the prediction of a listing's *price*. Unrealistically high values for these variables may bias the coefficient of a regression modeling how these predictors impact price. For example, having 50 beds in a listing may not impact the price any more than having 25 beds. To account for such outliers, we decided to cap the outliers at the 99th percentile value for these variables of interest. The capped values for these variables are presented below in *Table 3: Capped Values for Chosen Variables*. The capped value for the number of beds is 7. The capped value for the number of bedrooms is 5. The capped value for the number of bathrooms is 3.5. The interpretation for capping these predictors is that any value greater than the capped value will have the same weight in the regression for modeling price per night.

We chose to observe other right-skewness that we observed in variables since the sample size was large enough to ignore this.

Table 3: Capped Values for Chosen Variables

Variable Name	**Capped Value for Seattle**	**Capped Value for Boston**
beds	7	7
bedrooms	5	4
bathrooms	3.5	3

Figure 1A: Boxplots for Variables of Interest for Seattle



Issue #4: Categorical variables

For the categorical variables, we used one hot encoding to prepare the variables for further analysis. In the property_type variable, there are 26 different types of properties. With one hot encoding, each type of property becomes its own binary variable. Similarly, there are four different types of rooms, each of which becomes its own variable after one hot encoding. The purpose of transforming these categorical variables was to make them numeric so they can be fitted in a regression model.

After, performing the steps to resolve Issues #1-#3, we see the distribution of the *property_type*'s below in *Figure 2A: Distribution of Top 10 Most Common property_type's After Cleaning for Seattle* (for Boston, see *Figure 2B: Distribution of Top 10 Most Common property_type's After Cleaning for Boston* in the *Appendix*). We note here that the *property_type*'s of *Apartment* and *Home* are much more common than the rest. Similarly, in *Figure 3A: Distribution of room_type's After Cleaning for Seattle* (see *Figure 3B* in the *Appendix* for Boston data) we see that the distribution of *room_type*'s is skewed: there are far more "Entire home/apt"-values than other categories of *room_type*.

Figure 2A: Distribution of Top 10 Most Common property_type's After Cleaning for Seattle

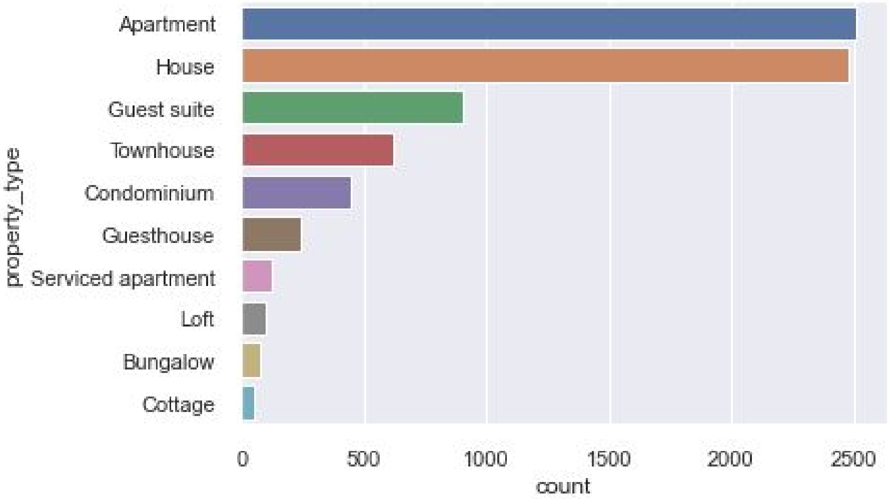
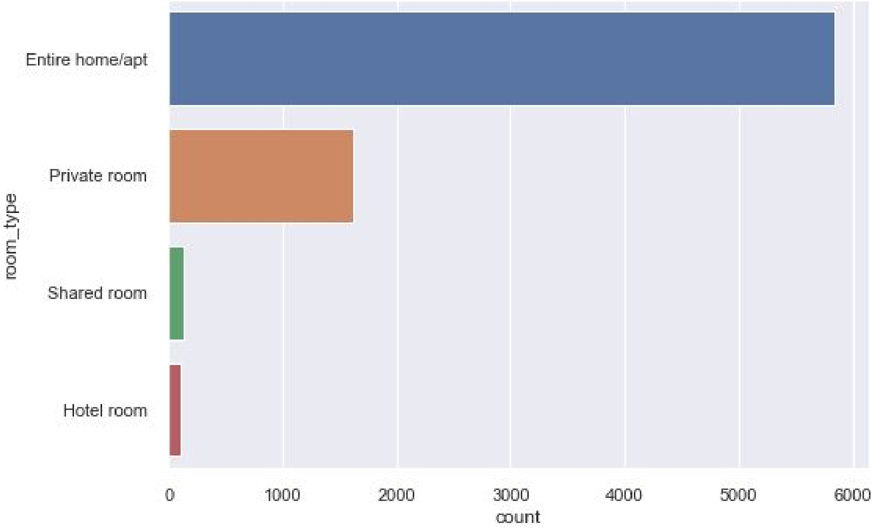


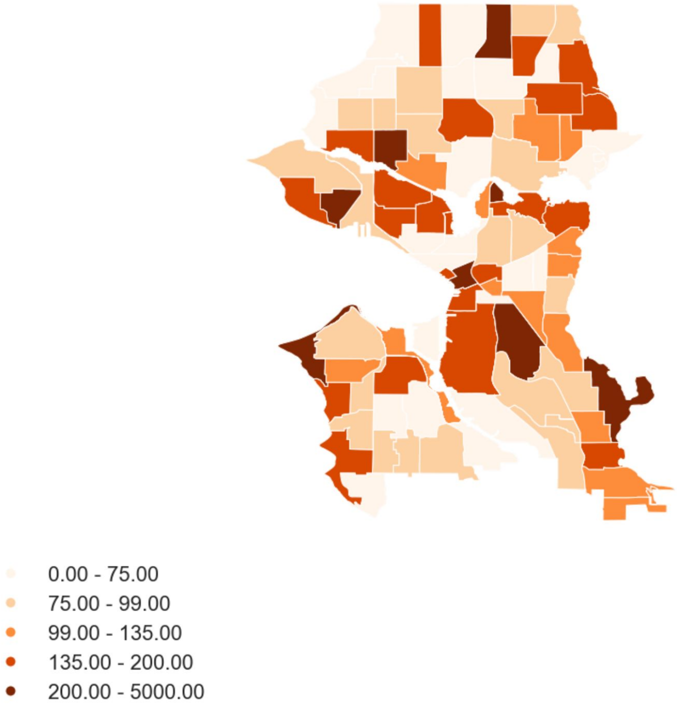
Figure 3A: Distribution of room_type's After Cleaning for Seattle

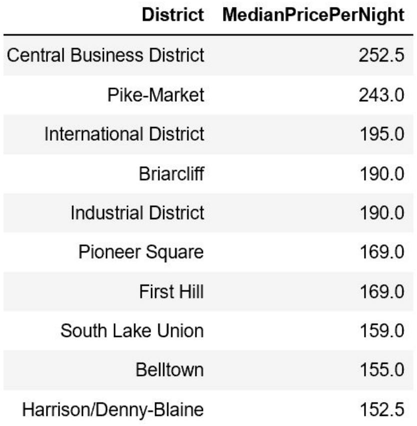


Geospatial Analysis

Another way we attempted to explore the data was by plotting out the quantile-price for neighborhoods in the city. As we see in *Figure 4A* below, some of the districts with the highest median price per night are intuitive: the more centrally located districts appear as the top 3 districts by median price per night. However, we notice that “Briarcliff” has the fourth highest median price per night of districts in Seattle. When we dive into the data, we also find that the median Briarcliff listing has more bedrooms than other districts, possibly indicative that home size contributes towards price per night.

Figure 4A: Median Price per Night for Listing by District - Seattle

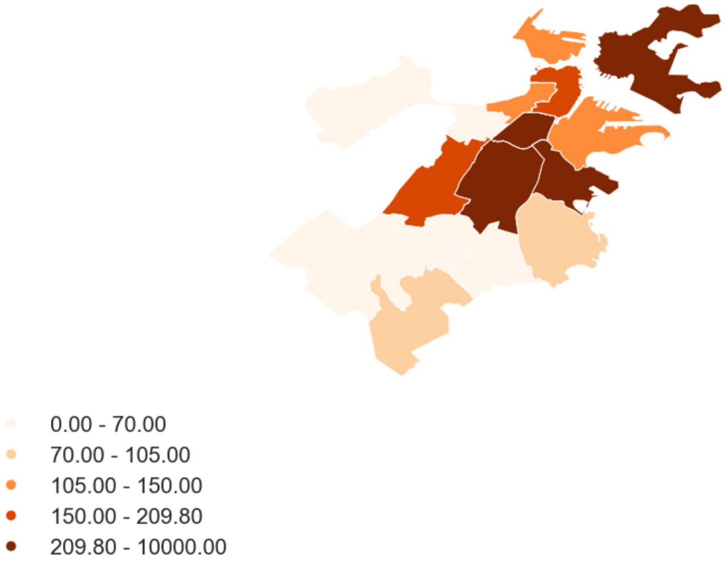


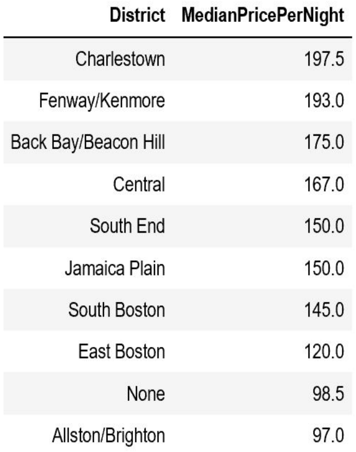


District	MedianPricePerNight
Central Business District	252.5
Pike-Market	243.0
International District	195.0
Briarcliff	190.0
Industrial District	190.0
Pioneer Square	169.0
First Hill	169.0
South Lake Union	159.0
Belltown	155.0
Harrison/Denny-Blaine	152.5

Similarly, for Boston, we find that more centrally located districts typically have a higher median price per night.

Figure 4B: Median Price per Night for Listing by District - Boston





District	MedianPricePerNight
Charlestown	197.5
Fenway/Kenmore	193.0
Back Bay/Beacon Hill	175.0
Central	167.0
South End	150.0
Jamaica Plain	150.0
South Boston	145.0
East Boston	120.0
None	98.5
Allston/Brighton	97.0

Summary statistics

Beyond the boxplots we observed, in *Figure 1A* and *Figure 1B*, we also printed a table of summary statistics for the variables of interest. To get a preliminary understanding of the distributions of the chosen variables of interest, we plotted histograms of the numerical variables. These can be found in tables *C1-C4* in the *Appendix*.

Statistical Methods

In this section, we attempt to understand the association between the aforementioned variables of interests and the price per night of Airbnb listings in Seattle and Boston using a regression model.

Model Selection and Assumptions

We chose a Poisson Regression as our model of choice because:

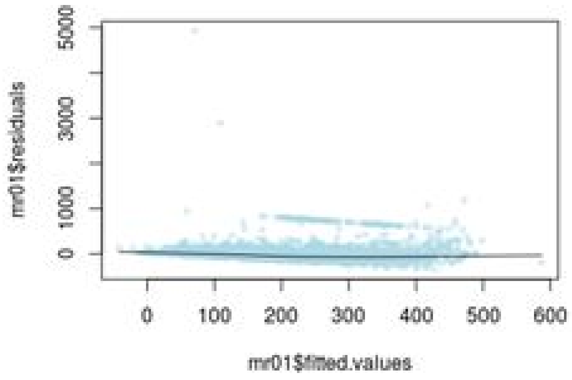
1. Price takes non-negative “integer-like” values.
2. Sample size is reasonably large in both cities.

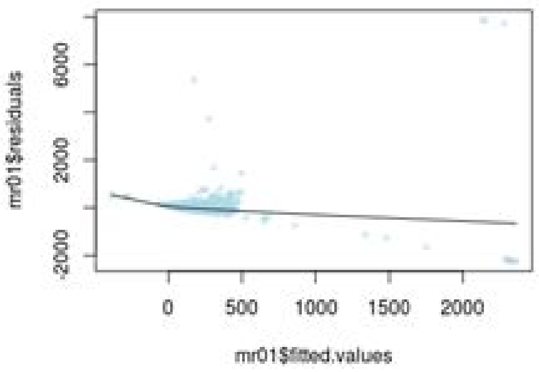
We address additional assumptions of other models we thought of below.

Assumption: Constant Variance

The data fails to meet the constant variance assumption required for regular linear regression, as seen in *Figures 5A* and *5B*.

Figure 5A/B: Non-Constant Variance of Residuals - Seattle and Boston





Independence of Observations

We note however, that since the data we are using is observational, it may not meet the assumption of independent observations. Here are two examples of how independence of observations may not hold in this data:

1. Hosts and users may be incentivized to modify or change their behavior over time to meet market requirements. For example, hosts may change the price of their listing in order to raise demand.
2. One host may submit multiple listings. These are two of the many ways in which the samples may not be independent from one another.

Poisson Regression

We tested if there exists a linear association between each numerical predictor variable and the response variable, price per night. Each of the tests resulted in a statistically significant coefficient estimate which is interpreted as the percent increase or decrease in price per day with a unit change of the predictor variable keeping the remaining predictors constant. For the categorical variables, room type and property type, we tested the null hypotheses about linear associations using an LRT test or analysis of deviance, which each yielded statistically significant results.

Interaction Terms Involving *beds, bathrooms, or bedrooms*

We provided interaction terms for variables which may serve as a proxy for the number of potential users who may stay in the listing (i.e.: beds, bedrooms, and bathrooms may all serve as proxy measures for the size of the property). We thought that from the data generating process, a larger property may result in more reviews (as more users may stay at the listing at any given time).

Interaction Terms Involving *room_type* or *property_type*

Interaction terms which involve *room_type* or *property_type* were chosen because we thought that these terms may provide information about the types of people who are interested in a listing. For example, a group of college students looking for a cheap abode to stay for summer break may be more inclined to choose a *hostel* type property, and their standards for the listing may be less stringent, than those for other *property_type’s* which may lead to differences in *review_score_rating*’s.

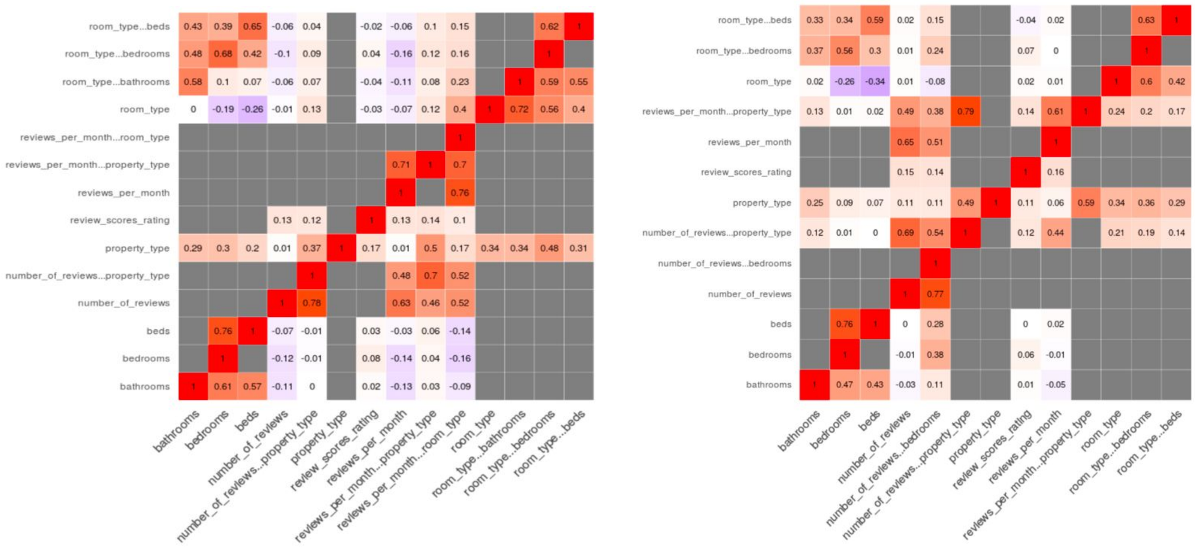
Other Reasons for Including Interaction Terms

Adding interaction terms to a regression model can greatly expand understanding of the relationships among the variables in the model and allows more hypotheses to be tested. Adding interaction terms makes the coefficients of the lower order terms conditional effects, not main effects. For example, for Seattle data, if we wanted to test the hypothesis that the relationship between the number of beds and price per night is different for ‘Hotel room’, ‘Shared room’,’Private room’ and ’Entire home/apt’, then adding the interaction term beds * factor(*room_type*) will be useful in this context. The presence of statistically significant interaction indicates that the effect of number of beds on *price* is different for different levels of room_type.

Similarly, for Boston data, if we wanted to test the hypothesis that the relationship between *number_of_reviews* and *price* is different for each level of *property_type*, then adding the interaction term *number_of_reviews* * factor(*property_type*) will help with this interpretation. The presence of statistically significant interaction indicates that the effect of *number_of_reviews* on *price* is different for different levels of *property_type*.

After eliminating all the highly correlated interaction terms, the resultant matrix looks like Figure 6 below:

Figure 6A/B: Correlation Matrix for Seattle and Boston Data



The results of this analysis will provide the inputs for the full model poisson regression that is our next method of analysis.

Finally, we conducted a poisson or log-linear regression with all of the predictor variables along with the statistically significant interactions found above under the null hypothesis that none of the factors are associated with price per day or all the coefficients are zero. While this null hypothesis was rejected in agreement with all of our methods of analysis discussed so far, we plotted the residuals against the fitted values to test the assumptions. We found that the assumption of non-constant variance is met since for a Poisson regression the variance is proportional to the mean. The results of the residuals vs. predictors plot in *Figure 5A* and *5B* indicated the same. The Poisson regression fits our data and interpretation best since we meet the non-variance assumption, the large sample size assumption, and do not assume normality.

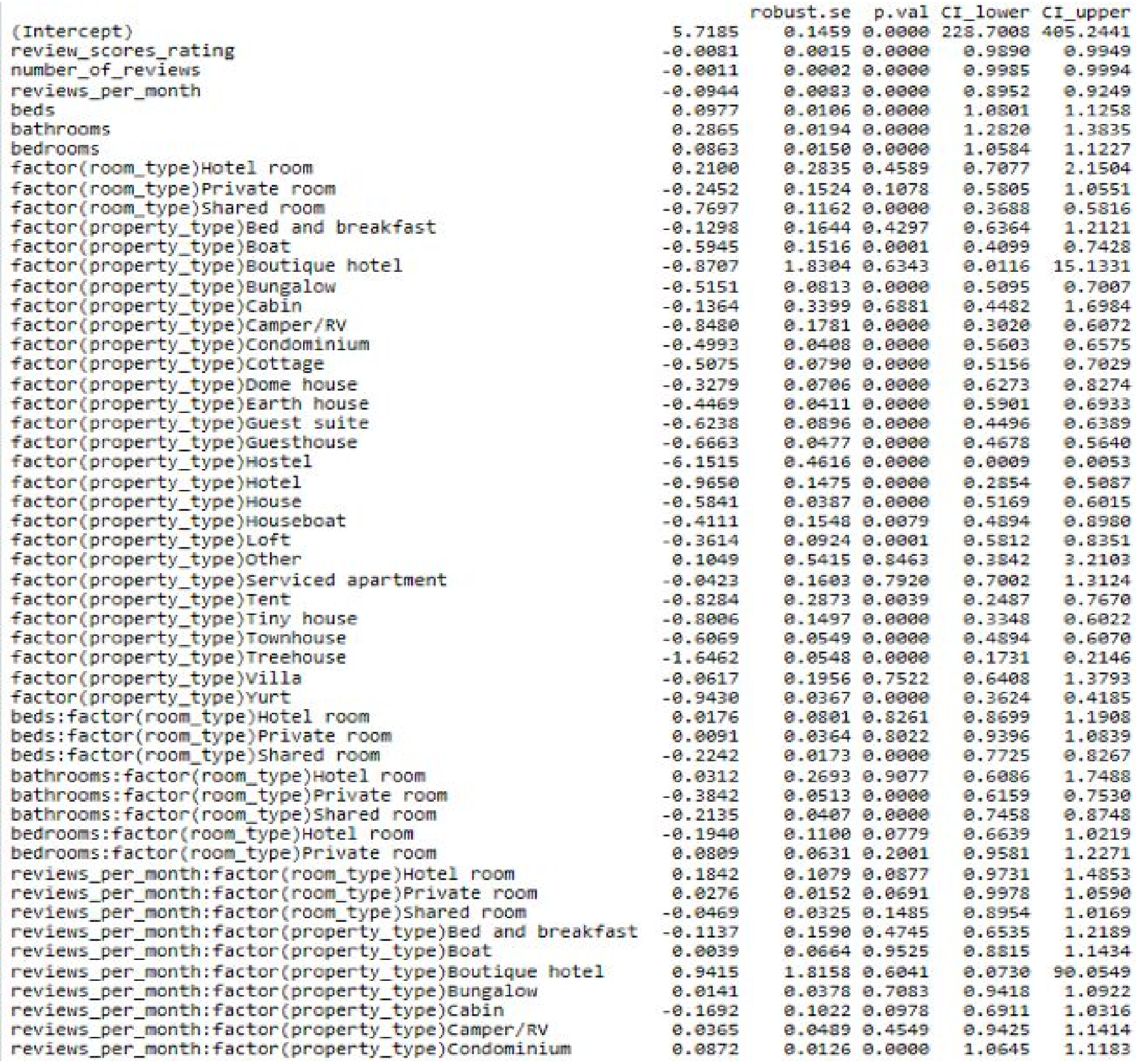
The same methods described in this section were used for both the Seattle and Boston dataset.

Results

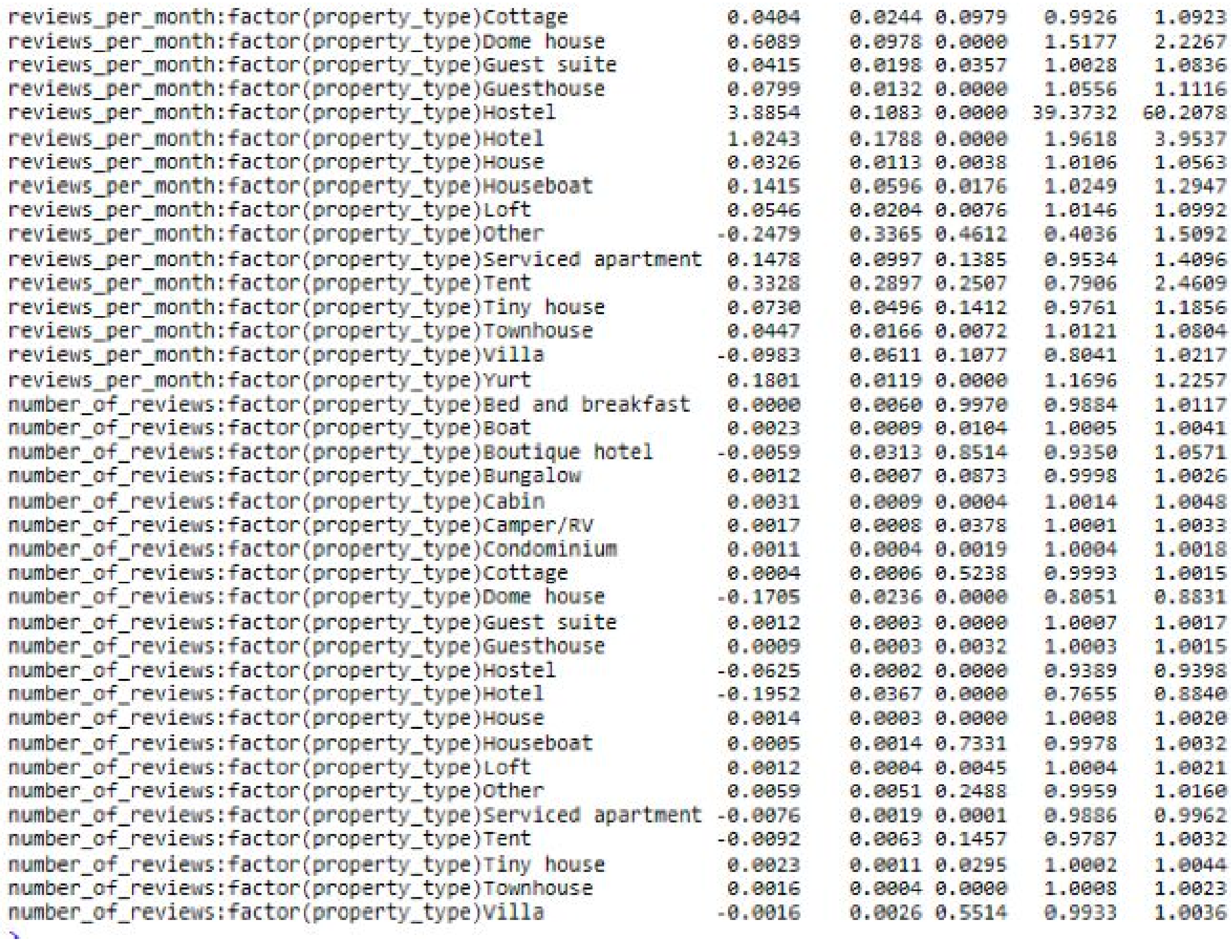
Seattle Poisson Regression

A summary of the tests for association between each of the predictors (including interaction terms) and price is shown below. Each null hypothesis was rejected since the tests were statistically significant.

Association between Price and the following variables:



```
                                                                 robust.se  p.val CI_lower CI_upper
(Intercept)                                               5.7185    0.1459 0.0000 228.7008 405.2441
review_scores_rating                                     -0.0081    0.0015 0.0000   0.9890   0.9949
number_of_reviews                                        -0.0011    0.0002 0.0000   0.9985   0.9994
reviews_per_month                                        -0.0944    0.0083 0.0000   0.8952   0.9249
beds                                                      0.0977    0.0106 0.0000   1.0801   1.1258
bathrooms                                                 0.2865    0.0194 0.0000   1.2820   1.3835
bedrooms                                                  0.0863    0.0150 0.0000   1.0584   1.1227
factor(room_type)Hotel room                               0.2100    0.2835 0.4589   0.7077   2.1504
factor(room_type)Private room                            -0.2452    0.1524 0.1078   0.5805   1.0551
factor(room_type)Shared room                             -0.7697    0.1162 0.0000   0.3688   0.5816
factor(property_type)Bed and breakfast                   -0.1298    0.1644 0.4297   0.6364   1.2121
factor(property_type)Boat                                -0.5945    0.1516 0.0001   0.4099   0.7428
factor(property_type)Boutique hotel                      -0.8707    1.8304 0.6343   0.0116  15.1331
factor(property_type)Bungalow                            -0.5151    0.0813 0.0000   0.5095   0.7007
factor(property_type)Cabin                               -0.1364    0.3399 0.6881   0.4482   1.6984
factor(property_type)Camper/RV                           -0.8480    0.1781 0.0000   0.3020   0.6072
factor(property_type)Condominium                         -0.4993    0.0408 0.0000   0.5603   0.6575
factor(property_type)Cottage                             -0.5075    0.0790 0.0000   0.5156   0.7029
factor(property_type)Dome house                          -0.3279    0.0706 0.0000   0.6273   0.8274
factor(property_type)Earth house                         -0.4469    0.0411 0.0000   0.5901   0.6933
factor(property_type)Guest suite                         -0.6238    0.0896 0.0000   0.4496   0.6389
factor(property_type)Guesthouse                          -0.6663    0.0477 0.0000   0.4678   0.5640
factor(property_type)Hostel                              -6.1515    0.4616 0.0000   0.0009   0.0053
factor(property_type)Hotel                               -0.9650    0.1475 0.0000   0.2854   0.5087
factor(property_type)House                               -0.5841    0.0387 0.0000   0.5169   0.6015
factor(property_type)Houseboat                           -0.4111    0.1548 0.0079   0.4894   0.8980
factor(property_type)Loft                                -0.3614    0.0924 0.0001   0.5812   0.8351
factor(property_type)Other                                0.1049    0.5415 0.8463   0.3842   3.2103
factor(property_type)Serviced apartment                  -0.0423    0.1603 0.7920   0.7002   1.3124
factor(property_type)Tent                                -0.8284    0.2873 0.0039   0.2487   0.7670
factor(property_type)Tiny house                          -0.8006    0.1497 0.0000   0.3348   0.6022
factor(property_type)Townhouse                           -0.6069    0.0549 0.0000   0.4894   0.6070
factor(property_type)Treehouse                           -1.6462    0.0548 0.0000   0.1731   0.2146
factor(property_type)Villa                               -0.0617    0.1956 0.7522   0.6408   1.3793
factor(property_type)Yurt                                -0.9430    0.0367 0.0000   0.3624   0.4185
beds:factor(room_type)Hotel room                          0.0176    0.0801 0.8261   0.8699   1.1908
beds:factor(room_type)Private room                        0.0091    0.0364 0.8022   0.9396   1.0839
beds:factor(room_type)Shared room                        -0.2242    0.0173 0.0000   0.7725   0.8267
bathrooms:factor(room_type)Hotel room                     0.0312    0.2693 0.9077   0.6086   1.7488
bathrooms:factor(room_type)Private room                  -0.3842    0.0513 0.0000   0.6159   0.7530
bathrooms:factor(room_type)Shared room                   -0.2135    0.0407 0.0000   0.7458   0.8748
bedrooms:factor(room_type)Hotel room                     -0.1940    0.1100 0.0779   0.6639   1.0219
bedrooms:factor(room_type)Private room                    0.0809    0.0631 0.2001   0.9581   1.2271
reviews_per_month:factor(room_type)Hotel room             0.1842    0.1079 0.0877   0.9731   1.4853
reviews_per_month:factor(room_type)Private room           0.0276    0.0152 0.0691   0.9978   1.0590
reviews_per_month:factor(room_type)Shared room           -0.0469    0.0325 0.1485   0.8954   1.0169
reviews_per_month:factor(property_type)Bed and breakfast -0.1137    0.1590 0.4745   0.6535   1.2189
reviews_per_month:factor(property_type)Boat               0.0039    0.0664 0.9525   0.8815   1.1434
reviews_per_month:factor(property_type)Boutique hotel     0.9415    1.8158 0.6041   0.0730  90.0549
reviews_per_month:factor(property_type)Bungalow           0.0141    0.0378 0.7083   0.9418   1.0922
reviews_per_month:factor(property_type)Cabin             -0.1692    0.1022 0.0978   0.6911   1.0316
reviews_per_month:factor(property_type)Camper/RV          0.0365    0.0489 0.4549   0.9425   1.1414
reviews_per_month:factor(property_type)Condominium        0.0872    0.0126 0.0000   1.0645   1.1183
```



```
reviews_per_month:factor(property_type)Cottage             0.0404    0.0244 0.0979    0.9926    1.0923
reviews_per_month:factor(property_type)Dome house          0.6089    0.0978 0.0000    1.5177    2.2267
reviews_per_month:factor(property_type)Guest suite         0.0415    0.0198 0.0357    1.0028    1.0836
reviews_per_month:factor(property_type)Guesthouse          0.0799    0.0132 0.0000    1.0556    1.1116
reviews_per_month:factor(property_type)Hostel              3.8854    0.1083 0.0000   39.3732   60.2078
reviews_per_month:factor(property_type)Hotel               1.0243    0.1788 0.0000    1.9618    3.9537
reviews_per_month:factor(property_type)House               0.0326    0.0113 0.0038    1.0106    1.0563
reviews_per_month:factor(property_type)Houseboat           0.1415    0.0596 0.0176    1.0249    1.2947
reviews_per_month:factor(property_type)Loft                0.0546    0.0204 0.0076    1.0146    1.0992
reviews_per_month:factor(property_type)Other              -0.2479    0.3365 0.4612    0.4036    1.5092
reviews_per_month:factor(property_type)Serviced apartment  0.1478    0.0997 0.1385    0.9534    1.4096
reviews_per_month:factor(property_type)Tent                0.3328    0.2897 0.2507    0.7906    2.4609
reviews_per_month:factor(property_type)Tiny house          0.0730    0.0496 0.1412    0.9761    1.1856
reviews_per_month:factor(property_type)Townhouse           0.0447    0.0166 0.0072    1.0121    1.0804
reviews_per_month:factor(property_type)Villa              -0.0983    0.0611 0.1077    0.8041    1.0217
reviews_per_month:factor(property_type)Yurt                0.1801    0.0119 0.0000    1.1696    1.2257
number_of_reviews:factor(property_type)Bed and breakfast   0.0000    0.0060 0.9970    0.9884    1.0117
number_of_reviews:factor(property_type)Boat                0.0023    0.0009 0.0104    1.0005    1.0041
number_of_reviews:factor(property_type)Boutique hotel     -0.0059    0.0313 0.8514    0.9350    1.0571
number_of_reviews:factor(property_type)Bungalow            0.0012    0.0007 0.0873    0.9998    1.0026
number_of_reviews:factor(property_type)Cabin               0.0031    0.0009 0.0004    1.0014    1.0048
number_of_reviews:factor(property_type)Camper/RV           0.0017    0.0008 0.0378    1.0001    1.0033
number_of_reviews:factor(property_type)Condominium         0.0011    0.0004 0.0019    1.0004    1.0018
number_of_reviews:factor(property_type)Cottage             0.0004    0.0006 0.5238    0.9993    1.0015
number_of_reviews:factor(property_type)Dome house         -0.1705    0.0236 0.0000    0.8051    0.8831
number_of_reviews:factor(property_type)Guest suite         0.0012    0.0003 0.0000    1.0007    1.0017
number_of_reviews:factor(property_type)Guesthouse          0.0009    0.0003 0.0032    1.0003    1.0015
number_of_reviews:factor(property_type)Hostel             -0.0625    0.0002 0.0000    0.9389    0.9398
number_of_reviews:factor(property_type)Hotel              -0.1952    0.0367 0.0000    0.7655    0.8840
number_of_reviews:factor(property_type)House               0.0014    0.0003 0.0000    1.0008    1.0020
number_of_reviews:factor(property_type)Houseboat           0.0005    0.0014 0.7331    0.9978    1.0032
number_of_reviews:factor(property_type)Loft                0.0012    0.0004 0.0045    1.0004    1.0021
number_of_reviews:factor(property_type)Other               0.0059    0.0051 0.2488    0.9959    1.0160
number_of_reviews:factor(property_type)Serviced apartment -0.0076    0.0019 0.0001    0.9886    0.9962
number_of_reviews:factor(property_type)Tent               -0.0092    0.0063 0.1457    0.9787    1.0032
number_of_reviews:factor(property_type)Tiny house          0.0023    0.0011 0.0295    1.0002    1.0044
number_of_reviews:factor(property_type)Townhouse           0.0016    0.0004 0.0000    1.0008    1.0023
number_of_reviews:factor(property_type)Villa              -0.0016    0.0026 0.5514    0.9933    1.0036
```

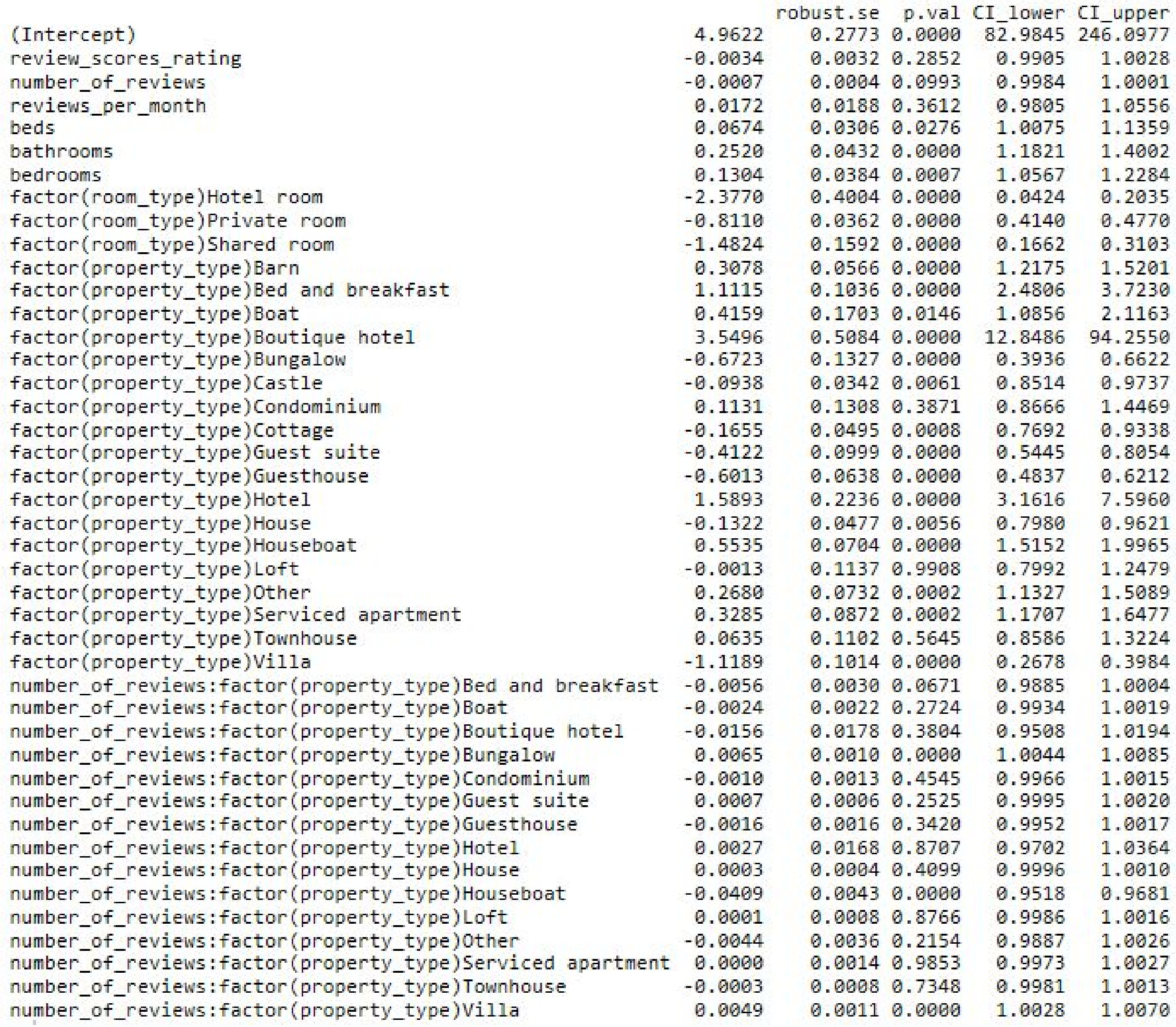
The interpretation for the results described in the table above are as follows. The coefficient estimates are exponentiated to describe a percent effect on mean price per night.

- Mean price decreases by 1% for each increase of one-unit review_scores_rating keeping all other variables and their interaction terms constant.
- Mean price decreases by 1% for each increase of one-unit number_of_reviews for property type ‘Bed and Breakfast’ keeping all other variables and their interaction terms constant.
- Mean price decreases by 3% for each increase of one-unit reviews_per_month for property type ‘Bread and Breakfast’ and room_type ‘Hotel room’ keeping all other variables and their interaction terms constant.
- Mean price increases by 12% for each increase of one-unit beds for room type ‘Hotel room’ keeping all other variables and their interaction terms constant.
- Mean price increases by 18% for each increase of one-unit bedrooms for room type ‘private room’ keeping all other variables and their interaction terms constant.
- Mean price increases by 37% for each increase of one-unit bathrooms for room type ‘Hotel room’ keeping all other variables and their interaction terms constant.
- Mean price decreases by 21% for each increase of one-unit property type ‘Bed and Breakfast’ for one-unit number_of_reviews and one-unit reviews_per_month keeping all other variables and their interaction terms constant.

- ❖ Mean price increases by 28% for each increase of one-unit room type ‘Hotel room’ for one-unit beds,one-unit bedrooms, one-unit bathrooms and one-unit reviews_per_month keeping all other variables and their interaction terms constant.

Boston Poisson Regression

Association between Price and the following variables:



```
                                                                  robust.se  p.val CI_lower CI_upper
(Intercept)                                                4.9622    0.2773 0.0000  82.9845 246.0977
review_scores_rating                                      -0.0034    0.0032 0.2852   0.9905   1.0028
number_of_reviews                                         -0.0007    0.0004 0.0993   0.9984   1.0001
reviews_per_month                                          0.0172    0.0188 0.3612   0.9805   1.0556
beds                                                       0.0674    0.0306 0.0276   1.0075   1.1359
bathrooms                                                  0.2520    0.0432 0.0000   1.1821   1.4002
bedrooms                                                   0.1304    0.0384 0.0007   1.0567   1.2284
factor(room_type)Hotel room                               -2.3770    0.4004 0.0000   0.0424   0.2035
factor(room_type)Private room                             -0.8110    0.0362 0.0000   0.4140   0.4770
factor(room_type)Shared room                              -1.4824    0.1592 0.0000   0.1662   0.3103
factor(property_type)Barn                                  0.3078    0.0566 0.0000   1.2175   1.5201
factor(property_type)Bed and breakfast                     1.1115    0.1036 0.0000   2.4806   3.7230
factor(property_type)Boat                                  0.4159    0.1703 0.0146   1.0856   2.1163
factor(property_type)Boutique hotel                        3.5496    0.5084 0.0000  12.8486  94.2550
factor(property_type)Bungalow                             -0.6723    0.1327 0.0000   0.3936   0.6622
factor(property_type)Castle                               -0.0938    0.0342 0.0061   0.8514   0.9737
factor(property_type)Condominium                           0.1131    0.1308 0.3871   0.8666   1.4469
factor(property_type)Cottage                              -0.1655    0.0495 0.0008   0.7692   0.9338
factor(property_type)Guest suite                          -0.4122    0.0999 0.0000   0.5445   0.8054
factor(property_type)Guesthouse                           -0.6013    0.0638 0.0000   0.4837   0.6212
factor(property_type)Hotel                                 1.5893    0.2236 0.0000   3.1616   7.5960
factor(property_type)House                                -0.1322    0.0477 0.0056   0.7980   0.9621
factor(property_type)Houseboat                             0.5535    0.0704 0.0000   1.5152   1.9965
factor(property_type)Loft                                 -0.0013    0.1137 0.9908   0.7992   1.2479
factor(property_type)Other                                 0.2680    0.0732 0.0002   1.1327   1.5089
factor(property_type)Serviced apartment                    0.3285    0.0872 0.0002   1.1707   1.6477
factor(property_type)Townhouse                             0.0635    0.1102 0.5645   0.8586   1.3224
factor(property_type)Villa                                -1.1189    0.1014 0.0000   0.2678   0.3984
number_of_reviews:factor(property_type)Bed and breakfast  -0.0056    0.0030 0.0671   0.9885   1.0004
number_of_reviews:factor(property_type)Boat               -0.0024    0.0022 0.2724   0.9934   1.0019
number_of_reviews:factor(property_type)Boutique hotel     -0.0156    0.0178 0.3804   0.9508   1.0194
number_of_reviews:factor(property_type)Bungalow            0.0065    0.0010 0.0000   1.0044   1.0085
number_of_reviews:factor(property_type)Condominium        -0.0010    0.0013 0.4545   0.9966   1.0015
number_of_reviews:factor(property_type)Guest suite         0.0007    0.0006 0.2525   0.9995   1.0020
number_of_reviews:factor(property_type)Guesthouse         -0.0016    0.0016 0.3420   0.9952   1.0017
number_of_reviews:factor(property_type)Hotel               0.0027    0.0168 0.8707   0.9702   1.0364
number_of_reviews:factor(property_type)House               0.0003    0.0004 0.4099   0.9996   1.0010
number_of_reviews:factor(property_type)Houseboat          -0.0409    0.0043 0.0000   0.9518   0.9681
number_of_reviews:factor(property_type)Loft                0.0001    0.0008 0.8766   0.9986   1.0016
number_of_reviews:factor(property_type)Other              -0.0044    0.0036 0.2154   0.9887   1.0026
number_of_reviews:factor(property_type)Serviced apartment  0.0000    0.0014 0.9853   0.9973   1.0027
number_of_reviews:factor(property_type)Townhouse          -0.0003    0.0008 0.7348   0.9981   1.0013
number_of_reviews:factor(property_type)Villa               0.0049    0.0011 0.0000   1.0028   1.0070
```

- ❖ Mean price decreases by 1% for each increase of one-unit review_scores_rating keeping all other variables and their interaction terms constant.
- ❖ Mean price decreases by 1% for each increase of one-unit number_of_reviews for property type ‘Bed and Breakfast’ keeping all other variables and their interaction terms constant.
- ❖ Mean price increases by 1.7% for each increase of one-unit reviews_per_month keeping all other variables and their interaction terms constant.

- Mean price increases by 6.9% for each increase of one-unit beds keeping all other variables and their interaction terms constant.
- Mean price increases by 13% for each increase of one-unit bedrooms keeping all other variables and their interaction terms constant.
- Mean price increases by 28% for each increase of one-unit bathrooms keeping all other variables and their interaction terms constant.
- Mean price increases by 200% for each increase of one-unit property type ‘Bed and Breakfast’ for one-unit number_of_reviews keeping all other variables and their interaction terms constant.
- Mean price decreases by approx 90% for each increase of one-unit room type ‘Hotel room’ keeping all other variables and their interaction terms constant.

Question 1: Most Influential Factors

Based on our analysis, the most influential factors that affect price per night in Seattle Airbnb listings are bathrooms, bedrooms, and beds. The mean price per night increases for each increase in one unit of these variables. This is intuitive because a listing with more amenities would be of higher value. The reason we chose not to include the interaction terms as the most influential is because they do not apply to all property or room types. Their interpretation is very specific to a scenario.

In the Boston analysis, the most influential factors that affect price per night are bathrooms, bedrooms, and beds. This is consistent with the analysis from Seattle and also follows the same order of influence. For the same reasons as in the Seattle analysis, the interaction terms, although statistically significant, cannot be interpreted as influential to every scenario. That is why we have included them in the linear regression to get accurate results for each property and room type but not as an overall influential factor.

Question 2: Comparison of Listing Prices in Seattle and Boston

To answer our second question about how the factors affecting price per night vary across different cities, we can use the results of our analysis of Seattle and Boston Airbnb listings. In both cities, the bathrooms, bedrooms and beds were found to be the most influential factors on price per night. However, for Seattle, each of the most influential predictors had higher coefficient estimates than for Boston. Another major difference between the results of analysis of Seattle and Boston listings lies in the interaction. For Boston, the different property types and number of reviews were found to have a significant interaction, whereas for Seattle, property types and room types had significant interactions with many variables.

Discussion

Intrinsic characteristics of the Airbnb listings, bedrooms, beds and bathrooms were found to be the most influential factors in price per night. In this section, we expound upon some different ways in which our work could be extended or improved.

Considering more predictor variables & missing values

In this analysis we dropped all rows with missing values. However, we found that the missing values resulted in a different distribution of *property_type*'s as some *property_type*'s had missing values more often than others. Furthermore, future analysis may want to consider different predictor variables since the dataset is fairly rich with data.

More robust regional analysis

Firstly, the analysis of Seattle and Boston does not take different local factors into account. In fact the different distribution of *property_type*'s in *Figure 2A and 2B (Appendix)* suggest that the data generating process in the two cities may not result in the same underlying distribution for the same variables. As such, future work may want to leverage a fixed effects model to incorporate city-specific factors into the regression model. Secondly, future analysis could draw upon additional regional markets for Airbnb to draw a more robust geographic analysis of the impact of various predictor variables on price.

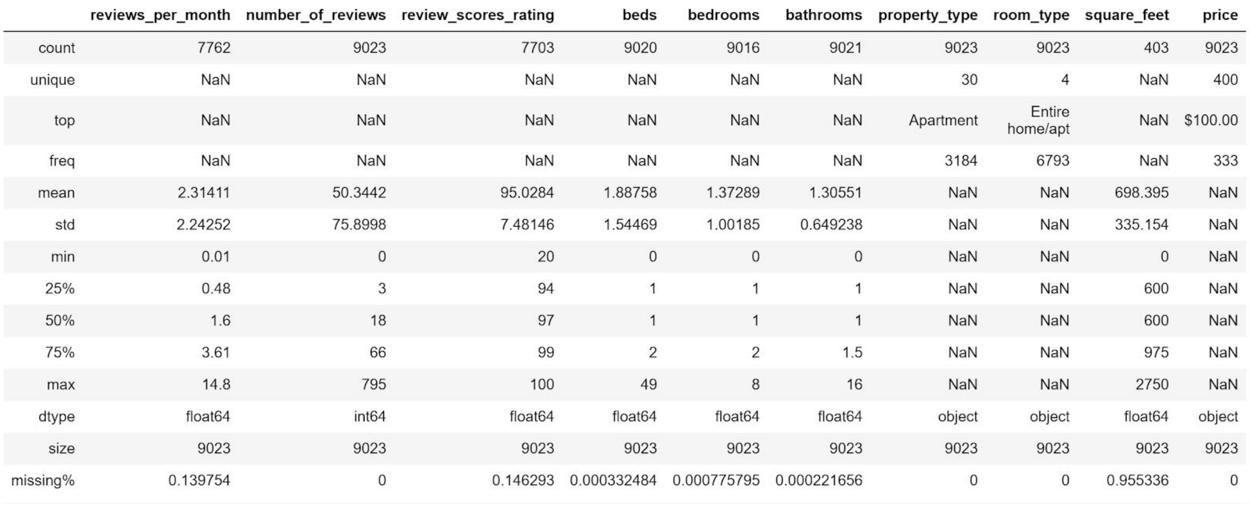
Consider local policy factors

Various cities have moved to enforce restrictions in local housing markets on short-term rentals. In fact the dataset we utilized from Inside Airbnb is focused on providing transparent data to help support the formulation of policy towards Airbnb. As such, attempting to analyze the impact of different Airbnb regulations on price would be interesting.

Appendix

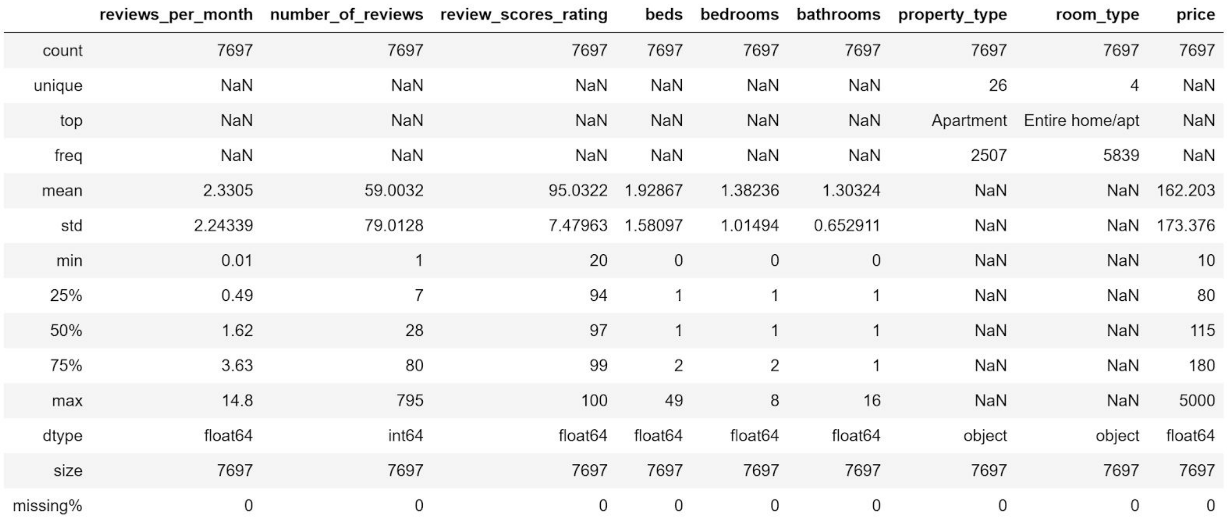
Tables

Table C1: Statistical Summary - Seattle



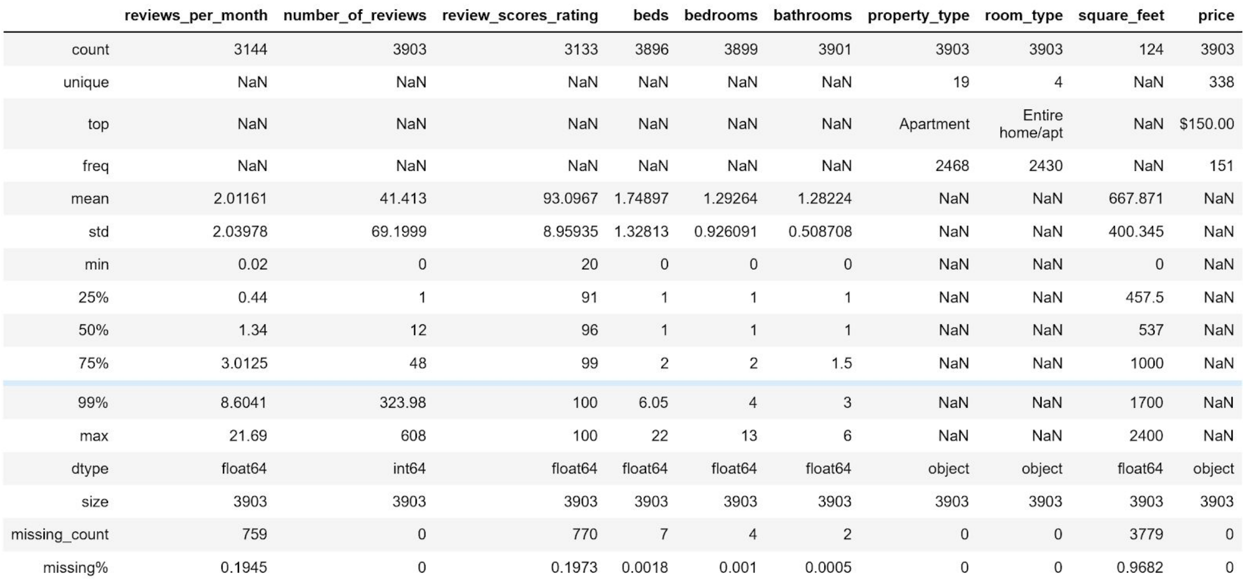
	reviews_per_month	number_of_reviews	review_scores_rating	beds	bedrooms	bathrooms	property_type	room_type	square_feet	price
count	7762	9023	7703	9020	9016	9021	9023	9023	403	9023
unique	NaN	NaN	NaN	NaN	NaN	NaN	30	4	NaN	400
top	NaN	NaN	NaN	NaN	NaN	NaN	Apartment	Entire home/apt	NaN	$100.00
freq	NaN	NaN	NaN	NaN	NaN	NaN	3184	6793	NaN	333
mean	2.31411	50.3442	95.0284	1.88758	1.37289	1.30551	NaN	NaN	698.395	NaN
std	2.24252	75.8998	7.48146	1.54469	1.00185	0.649238	NaN	NaN	335.154	NaN
min	0.01	0	20	0	0	0	NaN	NaN	0	NaN
25%	0.48	3	94	1	1	1	NaN	NaN	600	NaN
50%	1.6	18	97	1	1	1	NaN	NaN	600	NaN
75%	3.61	66	99	2	2	1.5	NaN	NaN	975	NaN
max	14.8	795	100	49	8	16	NaN	NaN	2750	NaN
dtype	float64	int64	float64	float64	float64	float64	object	object	float64	object
size	9023	9023	9023	9023	9023	9023	9023	9023	9023	9023
missing%	0.139754	0	0.146293	0.000332484	0.000775795	0.000221656	0	0	0.955336	0

Table C2: Statistical Summary - Seattle Cleaned Data



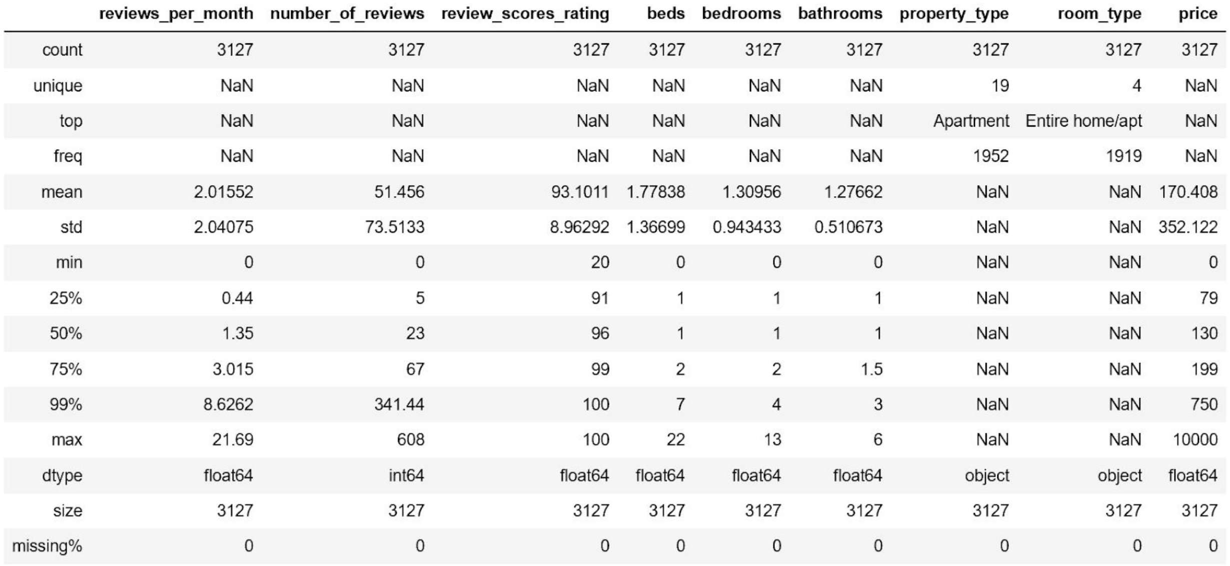
	reviews_per_month	number_of_reviews	review_scores_rating	beds	bedrooms	bathrooms	property_type	room_type	price
count	7697	7697	7697	7697	7697	7697	7697	7697	7697
unique	NaN	NaN	NaN	NaN	NaN	NaN	26	4	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN	Apartment	Entire home/apt	NaN
freq	NaN	NaN	NaN	NaN	NaN	NaN	2507	5839	NaN
mean	2.3305	59.0032	95.0322	1.92867	1.38236	1.30324	NaN	NaN	162.203
std	2.24339	79.0128	7.47963	1.58097	1.01494	0.652911	NaN	NaN	173.376
min	0.01	1	20	0	0	0	NaN	NaN	10
25%	0.49	7	94	1	1	1	NaN	NaN	80
50%	1.62	28	97	1	1	1	NaN	NaN	115
75%	3.63	80	99	2	2	1	NaN	NaN	180
max	14.8	795	100	49	8	16	NaN	NaN	5000
dtype	float64	int64	float64	float64	float64	float64	object	object	float64
size	7697	7697	7697	7697	7697	7697	7697	7697	7697
missing%	0	0	0	0	0	0	0	0	0

Table C3: Statistical Summary - Boston



	reviews_per_month	number_of_reviews	review_scores_rating	beds	bedrooms	bathrooms	property_type	room_type	square_feet	price
count	3144	3903	3133	3896	3899	3901	3903	3903	124	3903
unique	NaN	NaN	NaN	NaN	NaN	NaN	19	4	NaN	338
top	NaN	NaN	NaN	NaN	NaN	NaN	Apartment	Entire home/apt	NaN	$150.00
freq	NaN	NaN	NaN	NaN	NaN	NaN	2468	2430	NaN	151
mean	2.01161	41.413	93.0967	1.74897	1.29264	1.28224	NaN	NaN	667.871	NaN
std	2.03978	69.1999	8.95935	1.32813	0.926091	0.508708	NaN	NaN	400.345	NaN
min	0.02	0	20	0	0	0	NaN	NaN	0	NaN
25%	0.44	1	91	1	1	1	NaN	NaN	457.5	NaN
50%	1.34	12	96	1	1	1	NaN	NaN	537	NaN
75%	3.0125	48	99	2	2	1.5	NaN	NaN	1000	NaN
99%	8.6041	323.98	100	6.05	4	3	NaN	NaN	1700	NaN
max	21.69	608	100	22	13	6	NaN	NaN	2400	NaN
dtype	float64	int64	float64	float64	float64	float64	object	object	float64	object
size	3903	3903	3903	3903	3903	3903	3903	3903	3903	3903
missing_count	759	0	770	7	4	2	0	0	3779	0
missing%	0.1945	0	0.1973	0.0018	0.001	0.0005	0	0	0.9682	0

Table C4: Statistical Summary - Boston Data Cleaned Data



	reviews_per_month	number_of_reviews	review_scores_rating	beds	bedrooms	bathrooms	property_type	room_type	price
count	3127	3127	3127	3127	3127	3127	3127	3127	3127
unique	NaN	NaN	NaN	NaN	NaN	NaN	19	4	NaN
top	NaN	NaN	NaN	NaN	NaN	NaN	Apartment	Entire home/apt	NaN
freq	NaN	NaN	NaN	NaN	NaN	NaN	1952	1919	NaN
mean	2.01552	51.456	93.1011	1.77838	1.30956	1.27662	NaN	NaN	170.408
std	2.04075	73.5133	8.96292	1.36699	0.943433	0.510673	NaN	NaN	352.122
min	0	0	20	0	0	0	NaN	NaN	0
25%	0.44	5	91	1	1	1	NaN	NaN	79
50%	1.35	23	96	1	1	1	NaN	NaN	130
75%	3.015	67	99	2	2	1.5	NaN	NaN	199
99%	8.6262	341.44	100	7	4	3	NaN	NaN	750
max	21.69	608	100	22	13	6	NaN	NaN	10000
dtype	float64	int64	float64	float64	float64	float64	object	object	float64
size	3127	3127	3127	3127	3127	3127	3127	3127	3127
missing%	0	0	0	0	0	0	0	0	0

Figures

Figure 1B: Box Plots - Boston

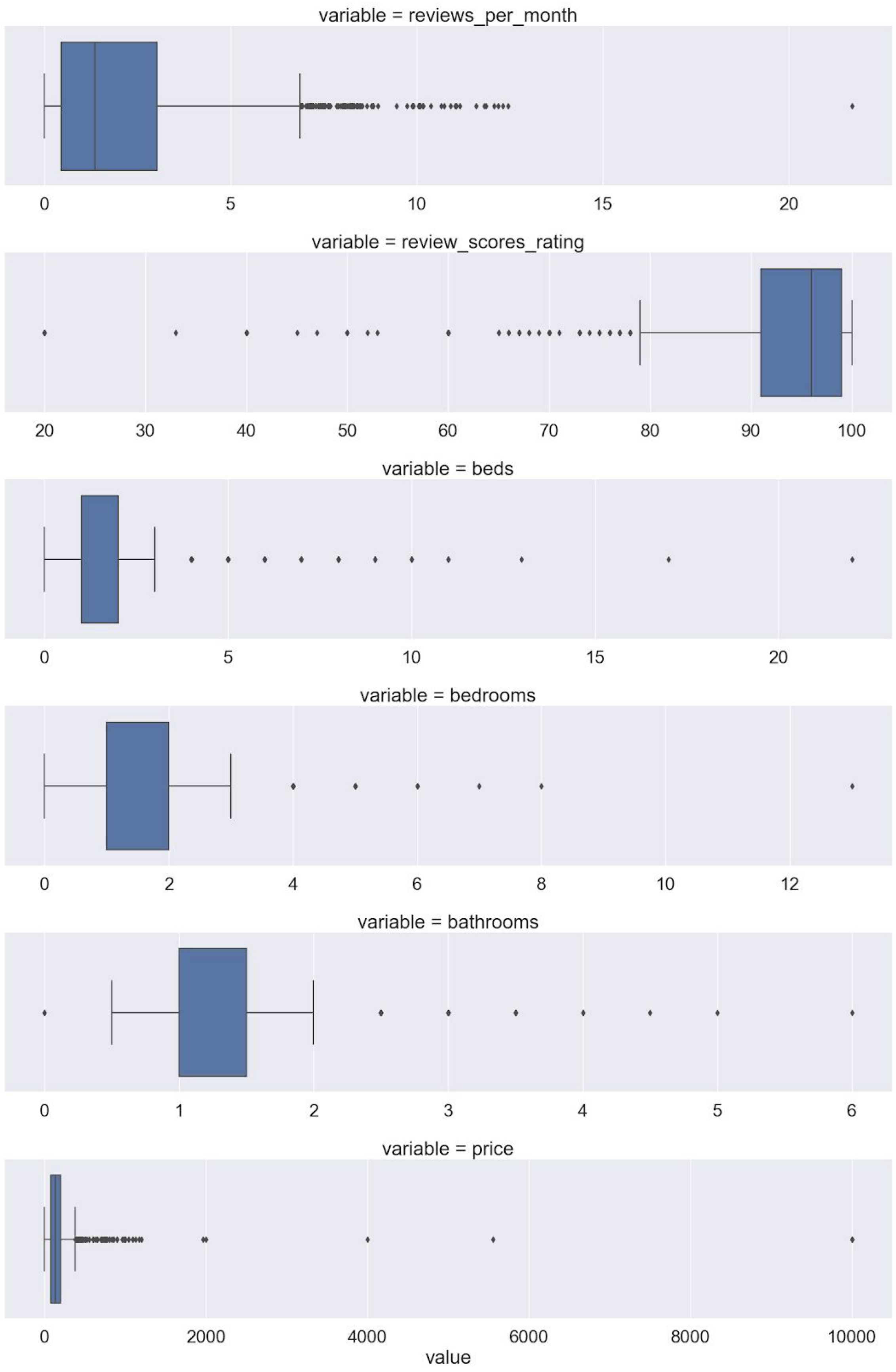

variable = reviews_per_month
0
5
10
15
20
variable = review_scores_rating
20
30
40
50
60
70
80
90
100
variable = beds
0
5
10
15
20
variable = bedrooms
0
2
4
6
8
10
12
variable = bathrooms
0
1
2
3
4
5
6
variable = price
0
2000
4000
6000
8000
10000
value

Figure 2B: Distribution of Top 10 Most Common property_type's After Cleaning for Boston

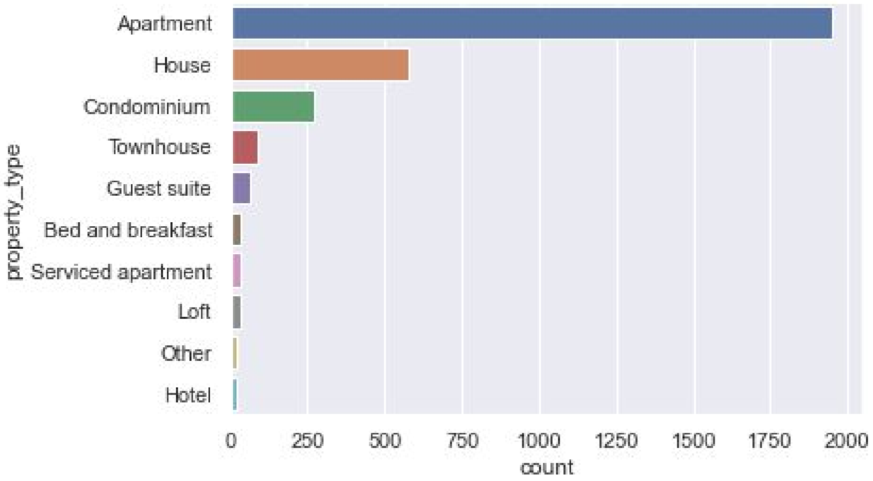
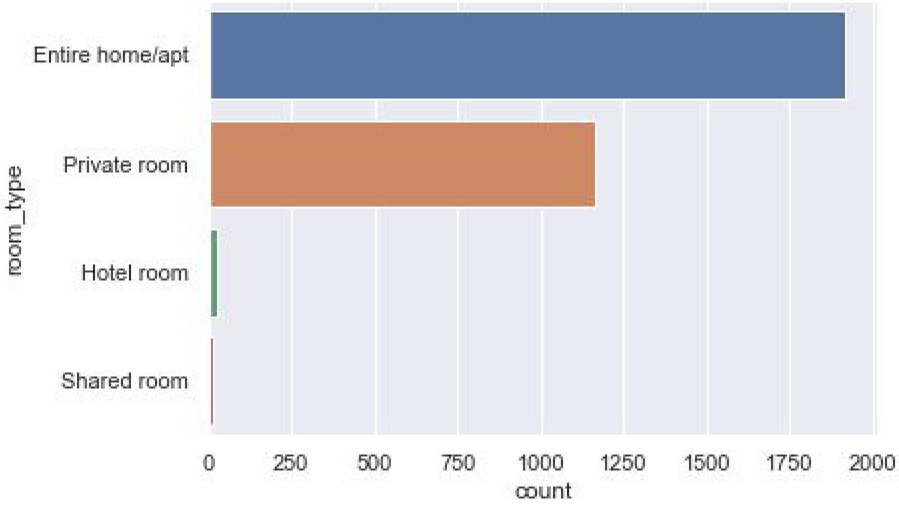


Figure 3B: Distribution of room_type's After Cleaning for Boston



References

[1]https://www.bloomberg.com/profile/company/9865065Z:US
[2]http://insideairbnb.com/about.html#disclaimers
[3]https://stackoverflow.com/questions/22649536/model-matrix-with-all-pairwise-interactions-between-variables
[4]https://news.airbnb.com/fast-facts/