AIRBNB PROJECT

Application of Analytical Models on Airbnb Price Prediction

Jooho Jo, Anthony Choi, Sujaya Darke, Site Bai, Darshana Daga,
12.13.2020 Group_7

TABLE OF CONTENTS

INTRODUCTION **2**

RELATED WORK **3**

DATASET **3**

Cleaning the Listing Dataset: 4

EXPLORATORY ANALYSIS **5**

Linear regression 6

Lasso regression 6

Deep learning 6

OLS; specify model and fit with training data 7

ANALYSES **7**

CONCLUSION **10**

BIBLIOGRAPHY **11**

FIGURES AND TABLES **12**

INTRODUCTION

As the pandemic started to slowly take over the world, we watched as the tourism and travel industry completely shut down. Airlines, hotels and car companies struggled to stay afloat while government officials in the US and worldwide tried to help us stay in business. Amidst this struggle, many eyes were on Airbnb as they were hit hard by the stay in place orders. The company has gone through struggles and creative processes to try and stay afloat.

According to the Airbnb website, the company has 5.6 million listings worldwide, operates in 100,000 cities, and has more than four million hosts. They have listings in more than 220 countries and have served more than eight million guests. This company is massive and has spread fast and wide since they started. It's no wonder all eyes were on them to see how they will grow or decline during the pandemic.

While Airbnb works through the challenges that the lockdown has brought on, the rest of us are wondering when we can get back to booking through airbnb and traveling again. Lockdown drags on, and we are stuck at home reminiscing on our previous travel experiences. While we dream about the places we will vacation once the vaccine has rolled out, we also dream about the beautiful villas and condos we will rent. Airbnb has transformed the travel industry - for millennials like us, checking Airbnb before hotels.com or booking.com is the norm. The beauty of Airbnb's multi-sided platform (MSP) is that travelers are able to experience living like locals while home-owners generate extra income using their own assets. Hosts currently make an average of $7,900 in earnings. Supplemental income like this is the beauty of MSPs. There is a high reward for minimal effort.

Normally, when thinking of airbnb, we consider factors from the travelers side: location, proximity to public transportation, parking, number of beds, etc. At the same time, we are looking for the best deal. In the earlier days of Airbnb, there was less competition. Nowadays, everything on the listing matters. The way the photos are taken, the amenities, whether or not you are a superhost…. Travelers look at everything while making a decision on which place to book. We are interested not only from the travelers perspective, but also from the hosts perspective: which factors affect listing prices the most? And can we generate a model that will predict the listing price?

RELATED WORK

Airbnb has been a lifesaving platform for members of our group on multiple occasions. While traveling through Mexico City, flight delays caused Sujaya and her friends to be stranded for an extra day. Without Airbnb, finding affordable accomodation in the middle of a bustling city, would have been near impossible in July.

Multi-sided platforms such as Uber and Airbnb have completely dismantled the status quo. Uber removed the need for taxis by allowing drivers to use their own cars to take people places. Airbnb connects people that own homes or apartments with people that want to stay in homes or apartments. These types of transactions allow people to use their personal assets as a way to generate extra income and maximize the value of our personal assets. By studying which factors affect the prices of airbnb listings, we will be able to understand how to potentially maximize our earning potential in case we are ever in a position to rent our homes out. Furthermore, with the rise of MSPs in every industry - cars, accomodation, food - understanding how every side of the business will benefit us.

Amidst the pandemic, delivery businesses are thriving. While Airbnb is not *thriving* persay, the company is still looking to go public soon. The pandemic took a huge toll on Airbnb's business. In April, the company saw $2 billion of debt and laid off almost 2k employees in May. However, there is hope for profitability. Given the right conditions, Airbnb can thrive and investors are excited for the IPO. Despite the pandemic, Airbnb stock will likely be successful. Later on in the report, we will see Airbnb booked rate in summer of 2019 compared to summer of 2020, which proves the theory that Airbnb will have a successful IPO given the current world state.

DATASET

How did we get the data and why we chose this data

Details on how we cleaned data

Airbnb hosts their data on the Inside Airbnb website. Here we can publicly source information from the Airbnb site. On this site, they host data from several cities around the world separated into three categories: listings, calendar, and reviews. This data, since it's sourced directly from Airbnb, would give us the most accurate information to answer the questions we are curious about. We chose to focus on Los Angeles to understand better about where we live and what affects prices of Airbnb listings here.

We downloaded the listing datasets, one by one, for the year 2019 and concatenated those datasets with each other. Then, we did the same process for the calendar files. Once we had all listings and calendar files, we started to analyze them separately. We examined the compiled dataset to understand which types of variables we are working with (Figure 2).

Cleaning the Listing Dataset:

We took the following steps to ensure that we were working with clean data.

1. Remove any rows with locations outside of California
2. Dropped columns that were 100% null values
3. Dropped rows that had null values
4. Dropped irrelevant columns
    a. We decided which columns were irrelevant by examining each of our 108 columns one by one. Columns like URL were unnecessary to answer our questions. We also removed columns that were repetitive. For example, we did not need *zip code* if we have *neighborhood*.
5. Outlier analysis
    a. We checked the *price*, *bedrooms*, and *bathroom* variables for outliers
    b. We chose not to remove the outliers because they represent the market
6. Convert data types
    a. Changed *price* and *cleaning fee* from string to float
    b. Object type to categorical (ex: convert dates to categorical)
7. Filtered data

    a. We had 47 different property types (apartment, house, guest house, villa, condo, etc.) We kept the top 80% which were the top 4 types of properties (apartment, guesthouse, house, and condominium)
8. One hot encoding
    a. We created dummy variables for categorical variables
        i. Superhost, property type, room type, neighborhood cleansed
    b. We are left with 274 dummy variables and 7 non-dummy variables
9. Correlation matrix
    a. We want to see which variables are highly correlated and can be removed to reduce redundancy (Figure 1)
    b. We saw that the accommodates variable is highly correlated with both beds and bedrooms
10. Reformat amenities column
    a. There are hundreds of different types of amenities
        i. Kitchen access, washer and dryer, hair dryer, parking, etc.
    b. Instead of analyzing each amenity separately, we want to see if the number of amenities impacts price of the listing
        i. We used lambda function to turn the string of amenities into total number of amenities
            1. Ex: kitchen, washer, dryer → 2

After running the correlation matrix and data cleaning process, we have 10 columns with 9 features, and 1 target variable - price. Generating dummy variables for categorical variables (host is super host, neighborhood cleansed, etc) brought our dataset to 281 columns ready for model training.

EXPLORATORY ANALYSIS

To start our analysis, we decided to run 3 models that would give the best fit for our data. We chose Linear Regression, Lasso Regression, and Deep Learning models. To further our knowledge and insights, we ran Ordinary Least Squares (OLS). The following section will explain our methods and results from the model. We split the dataset into 20% testing data and 80% training data. Below you will find details on each model and what we found from the output.

1. **Linear regression**
    a. This model helped us to identify coefficients which have an effect on the listing price
    b. We checked the square root of the mean squared error between y test variable and predicted y variable = 71950
    c. This model did not help us identify specifically which features affected the listing price
    d. This procedure helps teach the model to predict listing price
    e. Linear regression is commonly used to predict numerical values. We felt that this model would give us an accurate prediction of our price variable.
2. **Lasso regression**
    a. This model helps with regularization, by adding penalty on the model
    b. This method helps us to reduce overfitting
    c. We predicted y based on x test data and calculated square root of mean squared error = 71963
    d. We had assumed that using Lasso regression technique would help us to remove irrelevant attributes and improve the model accuracy. This model did help us delete some attributes, but this model did not improve our accuracy levels.
3. **Deep learning**
    a. This method has more power and higher accuracy score
    b. Used sequential technique which is a commonly used model in deep learning
    c. Trained the model then fit the model - instead of using split data, we specified the validation split inside the model at 20%
    d. The square root of mean squared error for training data is 63986, and for validation is 72638
    e. Deep learning is a powerful tool that involves more advanced algorithms to train and test our models. We used a more simplistic version of this model and we found that the model accuracy was not higher than the linear regression model.

4. **OLS; specify model and fit with training data**
    a. We had to run this so that we can see the effect of features on price and by how much
    b. Earlier models were just to predict the listing prices, but this one goes into more detail on each feature
    c. We were able to identify which features are important in predicting price which we will elaborate on in the next section

In regression models we focus on the square root of the mean squared error (sqrt of MSE) as a way to compare the models. Our linear regression sqrt of MSE was the lowest which leads us to use the insights from this model and the OLS to gather our insights (Figure 6).

ANALYSES

Our linear regression model was useful to see whether we could predict the price variable based on our independent variables. The OLS method allows us to see the magnitude of the coefficients so that we can determine which factors affect listing price the most. Below we will describe our key takeaways from the OLS analysis.

- OLS regression results (Figure 3) show:
    - Our theory was that if a host was classified as a superhost, the price of their listing would be higher, but after checking the OLS model, we found that superhost classification does not have a significant effect on listing price. Whether a host is a super host or not does not have an important positive effect on price, and the p value supports this finding.
    - By running the ordinary least squares regression with months variables on price, which is the target variable, the summary shows that the seasonality of the listing price is significant as all of the p-value is less than the cutoff value of 0.05. The summary shows that July has the largest positive effect on the price.
    - In addition, we also fit the ordinary least squares regression with month dummy variables on the target variable available indicating whether the room is booked. If the is booked the value is 1 and if the room is available, the value is 0. Similarly, the summary (figure2)shows the seasonality on

the target variable is significant as all of the p-value is less than the cutoff value of 0.05. This result is consistent with the outcome when calculating the booked rate for each month.

    - Figure 7a shows the coefficient of magnitude by numerical variable.
        - Number of bathrooms in each listing is by far the most important attribute in determining the price of a listing. This is because the number of bathrooms likely indicates how many rooms the listing has. Our key takeaway here is that listings with a higher number of bathrooms can likely be priced higher due to having more space and privacy in the property.
    - Figure 7b shows the coefficient of magnitude by categorical variable.
        - From here we noticed that location has an effect on the listing price as well. This bubble plot shows the top 15 locations with a positive effect on listing price. The most expensive listings are located in Bel-Air and Malibu (Figure 8).
        - Through further investigation (Figure 9), we found that in Hollywood Hills, Beverly Hills, and Malibu (the 3 locations with highest positive effect on listing price) also have a large majority of property types listed as "entire home or apartment". Likely the reason these areas are expensive is not only because they are scenic and popular travel destinations, but also because homeowners are renting out their entire property rather than sharing the space with their tenants.
- We found that the timing has an effect on price as well.
    - Monthly Booked Rate compared to Monthly Average Price (Figure 4)
        - The booked rate spikes every 3 months and then declines for 2 months.
            - We assume this quarterly cycle is in line with work and school cycles. This booked rate cycle lines up with fall, summer, spring, and winter breaks for school. Families often plan their vacation around their children's vacations.

        - The average price is highest in July and August, and this is likely because of summer vacations.
        - The booked rate is 2nd highest in October - we think this is because travelers are gearing up to spend their PTO before the year-end.
    - Weekly Booked Rate compared to Weekly Average Price (Figure 5)
        - Prices are highest on Friday and Saturday, likely due to the weekend.
        - The booked rate is highest on Friday and Saturday as well
        - It seems that people try to leave early and have Friday night and Saturday night, and head back on Sundays since the price and booked rate drops on Sunday.
- Room type Figure 10
    - Hotel rooms has a large negative effect on price, but with further investigation we learned that there are only 2 listings with “hotel room”.
    - You can see in the output that Entire Home/apt and Private room have a large positive effect on the listing price.
    - For the top 15 areas, their room types are dominated by the entire house or entire apartment which lines up with the results of our OLS summary.

The booked rate and average price is valuable information for hosts and travelers alike. For hosts, offering discounts during the weekdays might attract more customers on days that would otherwise not be booked. We recommend increasing the price on the weekend since the demand is higher these days. Travelers might be able to find better deals on the days that demand is low and supply is high (Monday - Thursday). Airbnb can use this information to focus marketing efforts on weekend getaways. Destinations that are weekday friendly, such as properties with nice office spaces or properties near office districts in big cities, should be highlighted by Airbnb for travelers browsing weekday destinations. Hosts can use this information to adjust price by month as well. There will be more demand during the summer months where prices will be higher, but hosts should increase price during October and April as well since the booked rate is higher during these months.

CONCLUSION

Our original objective was to find which factors affect listing price on Airbnb. We aimed to use this information to help hosts adjust their listing price to maximize their own profit. Furthermore, since Airbnb makes 3% of revenue per booking, our findings will help the company improve their operations as well.

The first glaring recommendation is to look closely at location. The specific location of the listing has a high effect on listing price. Each neighborhood and zipcode has a different average listing price. The areas with more tourist attractions, the locations close to the ocean, and the more highly populated areas often have the most expensive listing prices. A tip for hosts is that if you live outside of these hotspots, you may be able to offer discounts or charge slightly less and still attract customers that do not want to splurge on the higher priced listings. Through further investigation, we found that room type is also an important factor when determining listing price.

There are many different room types included in Airbnb: guesthouse, condo, shared room in apartment, entire home, etc. The areas with the highest average listing price have an overwhelming majority of room type: entire home/apartment. Naturally, an entire property is listed at a much higher price than a listing where the tenant has to share the place with the owner. Airbnb should encourage such listings in more locations in order to drive prices up. Furthermore, tenants that rent out shared spaces should ensure that there is privacy and advertise this on their listing. If there is enough privacy, we believe that the listing could be higher than places where there is less privacy.

The next biggest recommendation is to pay close attention to timing. As mentioned previously, the month and day of week have significant effects on price. We believe that sharing this information about price fluctuations per month and per day with hosts would be beneficial for Airbnb. With this information, hosts can set the best possible price and maximize revenues both for themselves and for the company. Travel patterns are changing with the pandemic. Longer stays are becoming more popular instead of shorter weekend getaways because of the work from home flexibility. While most of us would expect that travel would have significantly declined in 2020 compared to 2019, the data shows that the decline was only marginal (Figure 11). At first this data is alarming, but we have to remember that people may have to arrange for alternate accommodation depending on the Covid-19 situation in their permanent homes.

Now that we know that location, room type, and timing are the most important factors to listing price, we want to give further information to improve Airbnb operations. If we had more time and resources, we would spend this time to help Airbnb incorporate a local area's safety score into the service available for travellers.

Last September, Site and her mom took a trip to Philadelphia to visit those historical attractions such as The Liberty Bell Center. They had booked an Airbnb room at night and expected a great living experience. Communication with the host was smooth, the price of the room was relatively high, and they had great experiences with Airbnb before. The assumption here was that the higher priced room would ensure safety and a good experience overall. Unfortunately, at this location, they had to endure noisy parties and lots of commotion all night in the street. After some quick research online, they realized the neighborhood they were living in has a high crime rate.

Site couldn't help but think what could've been done to avoid this type of dreadful experience in the future? If Airbnb incorporated local areas' crime rate into the platform and allowed the travellers to check the crime rate before booking the room, would Site and other travellers have a different experience? We think this service could be a business opportunity for Airbnb as it could offer membership for travellers who would like to get more information such as crime rate when booking the room. This membership can be applied to Experiences service as well. For instance, travellers who subscribed to membership could get a discount on booking Experiences services. In conclusion, we believe offering membership would be beneficial for both Airbnb and consumers. By providing a safety score to customers, we can avoid unpleasant and unexpected experiences. By showing hosts where on the safety score their properties are listed, they can properly price their listings. Finally, we want to leave you with some ways that our model can be improved in the future. We used a simplistic method of deep learning. With more advanced techniques, we believe that our model prediction might improve and give better results than the linear regression method.

BIBLIOGRAPHY

https://investmentu.com/airbnb-ipo-date-abnb-stock-going-public/

Airbnb.com

Insideairbnb.com

All other information here is from BANA 205 class, and general knowledge about Airbnb platform.

FIGURES AND TABLES

Figure 1

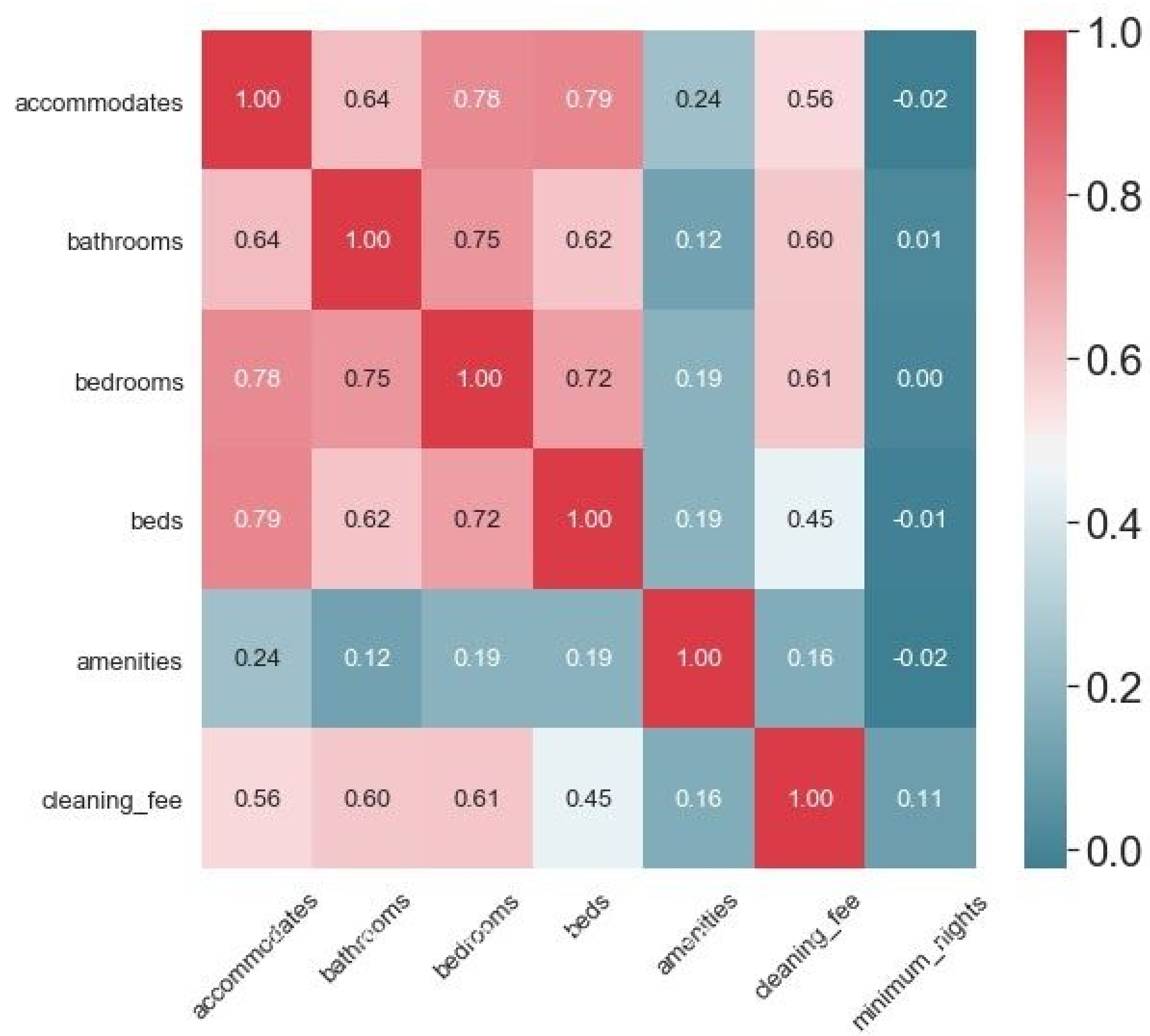


Figure 2

```
airbnb_ca.dtypes.value_counts()
```

```
]: object     63
   float64    22
   int64      21
   dtype: int64
```

```
In [8]: airbnb_df.shape
Out[8]: (606496, 106)
```

Figure 3

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.000
Model:	OLS	**Adj. R-squared:**	0.000
Method:	Least Squares	**F-statistic:**	1415.
Date:	Fri, 11 Dec 2020	**Prob (F-statistic):**	0.00
Time:	09:20:21	**Log-Likelihood:**	-4.0732e+08
No. Observations:	52784878	**AIC:**	8.146e+08
Df Residuals:	52784866	**BIC:**	8.146e+08
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
1	202.1873	0.234	863.641	0.000	201.728	202.646
2	197.5009	0.245	805.296	0.000	197.020	197.982
3	195.4801	0.235	831.599	0.000	195.019	195.941
4	207.3177	0.239	867.997	0.000	206.850	207.786
5	205.9274	0.234	880.173	0.000	205.469	206.386
6	209.8406	0.238	882.143	0.000	209.374	210.307
7	224.1127	0.232	964.706	0.000	223.657	224.568
8	223.1473	0.231	964.420	0.000	222.694	223.601
9	216.2037	0.235	919.049	0.000	215.743	216.665
10	209.2649	0.421	497.141	0.000	208.440	210.090
11	208.7173	0.421	495.838	0.000	207.892	209.542
12	213.0776	0.414	514.564	0.000	212.266	213.889

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              available   R-squared:                       0.020
Model:                            OLS   Adj. R-squared:                  0.020
Method:                 Least Squares   F-statistic:                 9.760e+04
Date:                Fri, 11 Dec 2020   Prob (F-statistic):               0.00
Time:                        09:34:39   Log-Likelihood:            -3.7541e+07
No. Observations:            52784878   AIC:                         7.508e+07
Df Residuals:                52784866   BIC:                         7.508e+07
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
1              0.5936      0.000   2795.711      0.000       0.593       0.594
2              0.5195      0.000   2335.344      0.000       0.519       0.520
3              0.4358      0.000   2043.861      0.000       0.435       0.436
4              0.6177      0.000   2851.510      0.000       0.617       0.618
5              0.5383      0.000   2536.872      0.000       0.538       0.539
6              0.4938      0.000   2288.855      0.000       0.493       0.494
7              0.6606      0.000   3134.956      0.000       0.660       0.661
8              0.5938      0.000   2829.517      0.000       0.593       0.594
9              0.4792      0.000   2245.884      0.000       0.479       0.480
10             0.6342      0.000   1661.032      0.000       0.633       0.635
11             0.5276      0.000   1381.868      0.000       0.527       0.528
12             0.4455      0.000   1186.201      0.000       0.445       0.446
==============================================================================
Omnibus:                193317225.867   Durbin-Watson:                   0.202
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          8113760.672
Skew:                          -0.184   Prob(JB):                         0.00
Kurtosis:                       1.115   Cond. No.                         1.82
==============================================================================
```

Figure 4

Monthly Booked Rate Trend

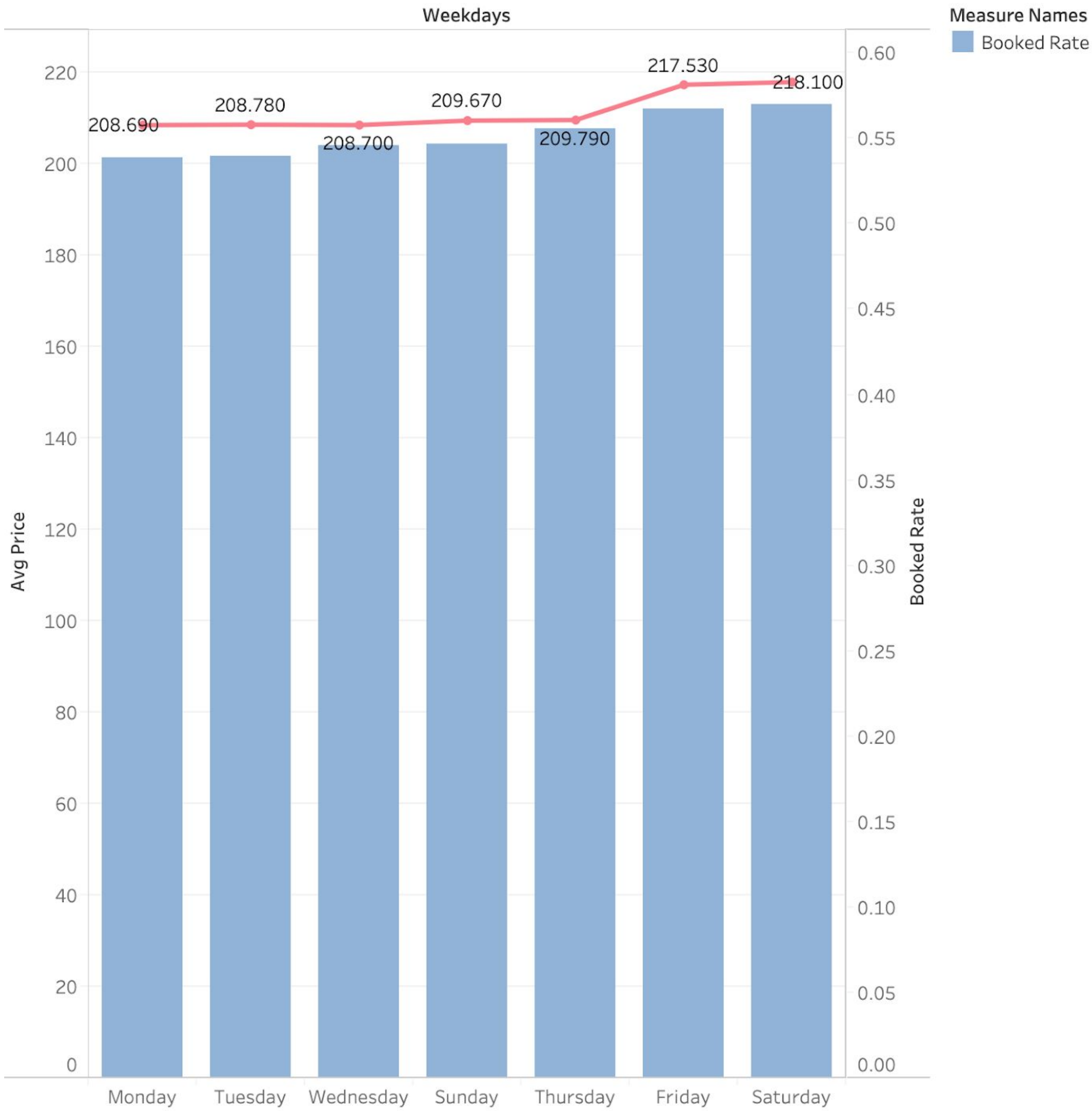
The trends of Avg Price and Booked Rate for Date. Color shows details about Avg Price and Booked Rate.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              available   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     941.6
Date:                Fri, 11 Dec 2020   Prob (F-statistic):               0.00
Time:                        14:19:04   Log-Likelihood:            -8.9381e+06
No. Observations:            12364401   AIC:                         1.788e+07
Df Residuals:                12364394   BIC:                         1.788e+07
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Friday         0.5506      0.000   1478.141      0.000       0.550       0.551
Monday         0.5283      0.000   1400.580      0.000       0.528       0.529
Saturday       0.5484      0.000   1465.928      0.000       0.548       0.549
Sunday         0.5383      0.000   1433.420      0.000       0.538       0.539
Thursday       0.5428      0.000   1452.613      0.000       0.542       0.544
Tuesday        0.5240      0.000   1392.068      0.000       0.523       0.525
Wednesday      0.5227      0.000   1388.137      0.000       0.522       0.523
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              available   R-squared:                       0.014
Model:                            OLS   Adj. R-squared:                  0.014
Method:                 Least Squares   F-statistic:                 8.793e+04
Date:                Fri, 11 Dec 2020   Prob (F-statistic):               0.00
Time:                        14:20:11   Log-Likelihood:            -8.8536e+06
No. Observations:            12364401   AIC:                         1.771e+07
Df Residuals:                12364398   BIC:                         1.771e+07
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
7              0.6100      0.000   2514.715      0.000       0.610       0.611
8              0.5320      0.000   2193.125      0.000       0.532       0.533
9              0.4652      0.000   1886.453      0.000       0.465       0.466
==============================================================================
```

Figure 5

Weekly Avg_Price and Booked Rate



The trends of Avg Price and Booked Rate for Weekdays. For pane Sum of Avg Price: Details are shown for Avg Price and Booked Rate. For pane Sum of Booked Rate: Color shows details about Avg Price and Booked Rate.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     501.3
Date:                Fri, 11 Dec 2020   Prob (F-statistic):               0.00
Time:                        10:11:24   Log-Likelihood:            -4.0732e+08
No. Observations:            52784878   AIC:                         8.146e+08
Df Residuals:                52784871   BIC:                         8.146e+08
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Friday       215.4033      0.197   1095.471      0.000     215.018     215.789
Monday       206.2049      0.201   1025.794      0.000     205.811     206.599
Saturday     216.0133      0.195   1107.910      0.000     215.631     216.395
Sunday       207.3212      0.197   1054.606      0.000     206.936     207.706
Thursday     207.5816      0.196   1057.891      0.000     207.197     207.966
Tuesday      206.0460      0.203   1016.949      0.000     205.649     206.443
Wednesday    206.6038      0.197   1046.764      0.000     206.217     206.991
==============================================================================
Omnibus:                105064581.384   Durbin-Watson:                   0.059
Prob(Omnibus):                  0.000   Jarque-Bera (JB):      395278389801.514
Skew:                          16.177   Prob(JB):                         0.00
Kurtosis:                     425.702   Cond. No.                         1.04
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:              available   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     4882.
Date:                Fri, 11 Dec 2020   Prob (F-statistic):               0.00
Time:                        10:14:22   Log-Likelihood:            -3.8058e+07
No. Observations:            52784878   AIC:                         7.612e+07
Df Residuals:                52784871   BIC:                         7.612e+07
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Friday         0.5621      0.000   3121.613      0.000       0.562       0.562
Monday         0.5340      0.000   2900.646      0.000       0.534       0.534
Saturday       0.5652      0.000   3165.197      0.000       0.565       0.566
Sunday         0.5417      0.000   3008.962      0.000       0.541       0.542
Thursday       0.5511      0.000   3066.801      0.000       0.551       0.551
Tuesday        0.5346      0.000   2880.818      0.000       0.534       0.535
Wednesday      0.5412      0.000   2994.073      0.000       0.541       0.542
==============================================================================
Omnibus:                184062056.465   Durbin-Watson:                   0.199
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          8781040.020
Skew:                          -0.190   Prob(JB):                         0.00
Kurtosis:                       1.038   Cond. No.                         1.04
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     61.71
Date:                Fri, 11 Dec 2020   Prob (F-statistic):           6.92e-77
Time:                        14:15:41   Log-Likelihood:            -9.7654e+07
No. Observations:            12364401   AIC:                         1.953e+08
Df Residuals:                12364394   BIC:                         1.953e+08
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Friday       238.5583      0.487    490.221      0.000     237.605     239.512
Monday       230.7415      0.493    468.234      0.000     229.776     231.707
Saturday     239.1646      0.489    489.360      0.000     238.207     240.123
Sunday       232.4434      0.491    473.768      0.000     231.482     233.405
Thursday     232.1806      0.488    475.567      0.000     231.224     233.137
Tuesday      230.3354      0.492    468.336      0.000     229.371     231.299
Wednesday    230.2574      0.492    468.005      0.000     229.293     231.222
==============================================================================
```

Figure 6

LINEAR REGRESSION		LASSO REGRESSION		DEEP LEARNING		
Attributes	Sqrt MSE	Attributes	Sqrt MSE	Attributes	Training MSE	Testing MSE
281	268	173	268	281	253	270

Figure 7a

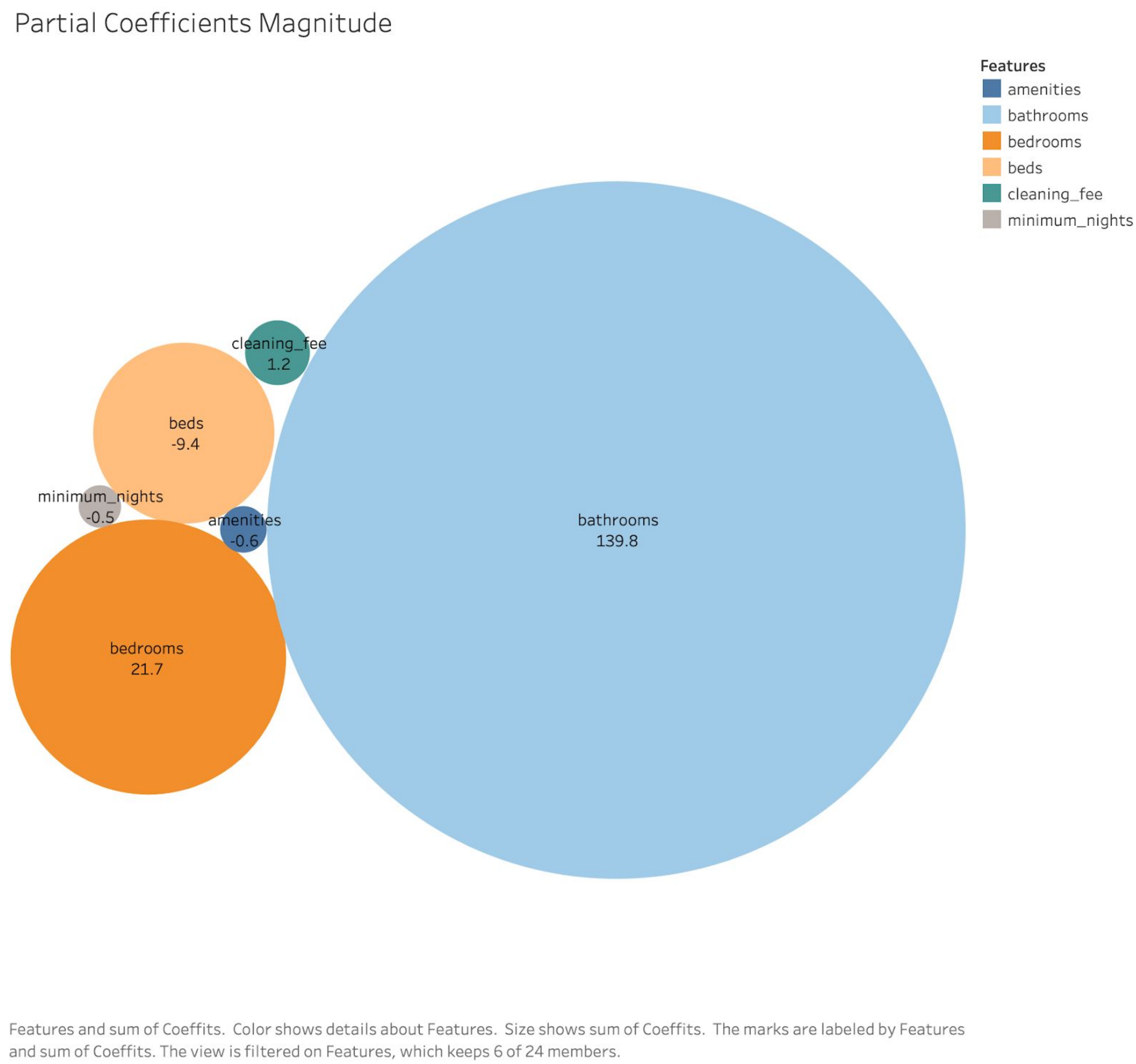

Partial Coefficients Magnitude
Features
amenities
bathrooms
bedrooms
beds
cleaning_fee
minimum_nights
cleaning_fee
1.2
beds
-9.4
minimum_nights
-0.5
amenities
-0.6
bathrooms
139.8
bedrooms
21.7
Features and sum of Coeffits. Color shows details about Features. Size shows sum of Coeffits. The marks are labeled by Features and sum of Coeffits. The view is filtered on Features, which keeps 6 of 24 members.

Figure 7b

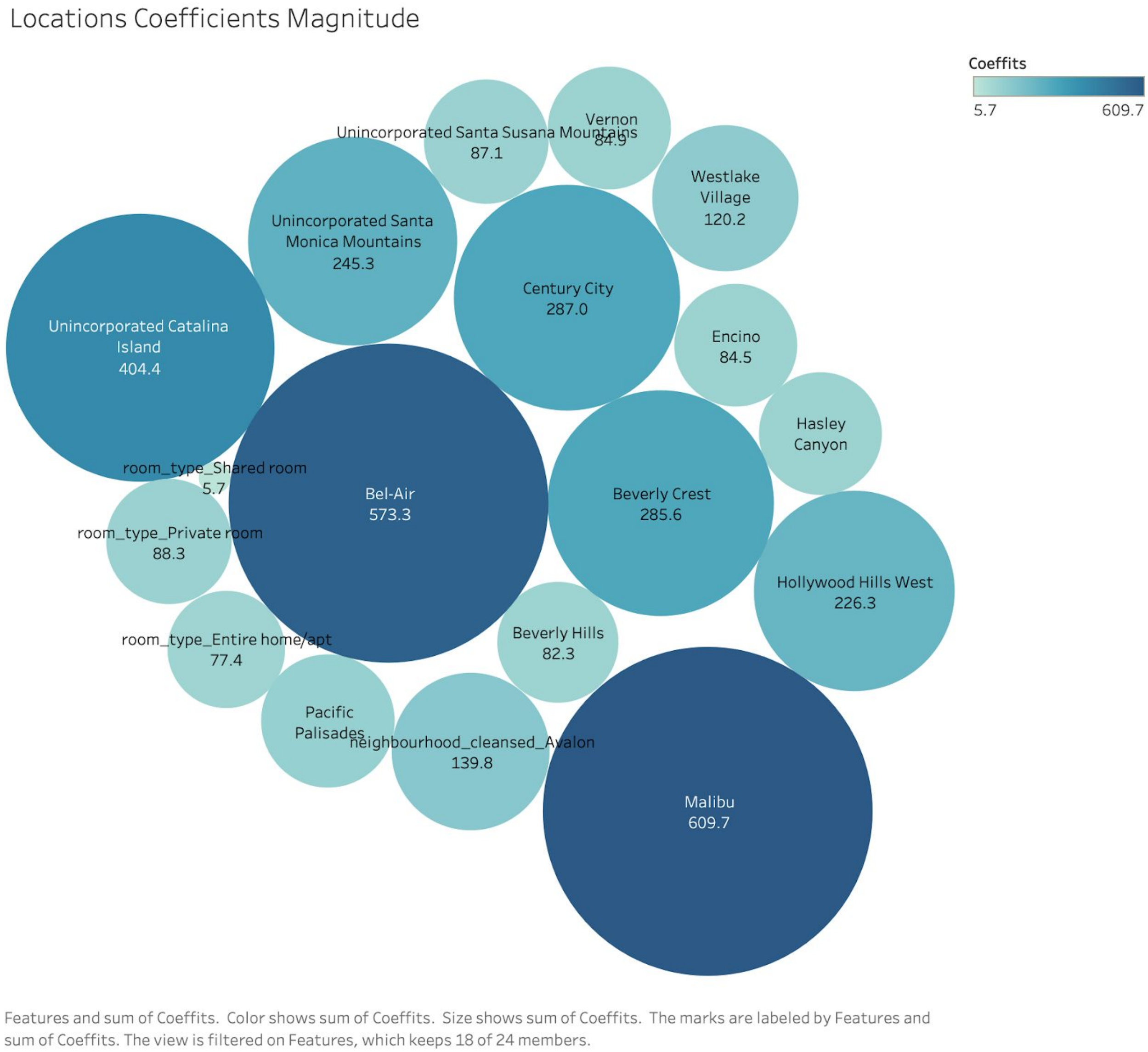

Locations Coefficients Magnitude
Coeffits
5.7
609.7
Unincorporated Santa Susana Mountains
87.1
Vernon
84.9
Westlake
Village
120.2
Unincorporated Santa
Monica Mountains
245.3
Century City
287.0
Unincorporated Catalina
Island
404.4
Encino
84.5
Hasley
Canyon
room_type_Shared room
5.7
Bel-Air
573.3
Beverly Crest
285.6
room_type_Private room
88.3
Hollywood Hills West
226.3
room_type_Entire home/apt
77.4
Beverly Hills
82.3
Pacific
Palisades
neighbourhood_cleansed_Avalon
139.8
Malibu
609.7
Features and sum of Coeffits. Color shows sum of Coeffits. Size shows sum of Coeffits. The marks are labeled by Features and sum of Coeffits. The view is filtered on Features, which keeps 18 of 24 members.

Figure 8

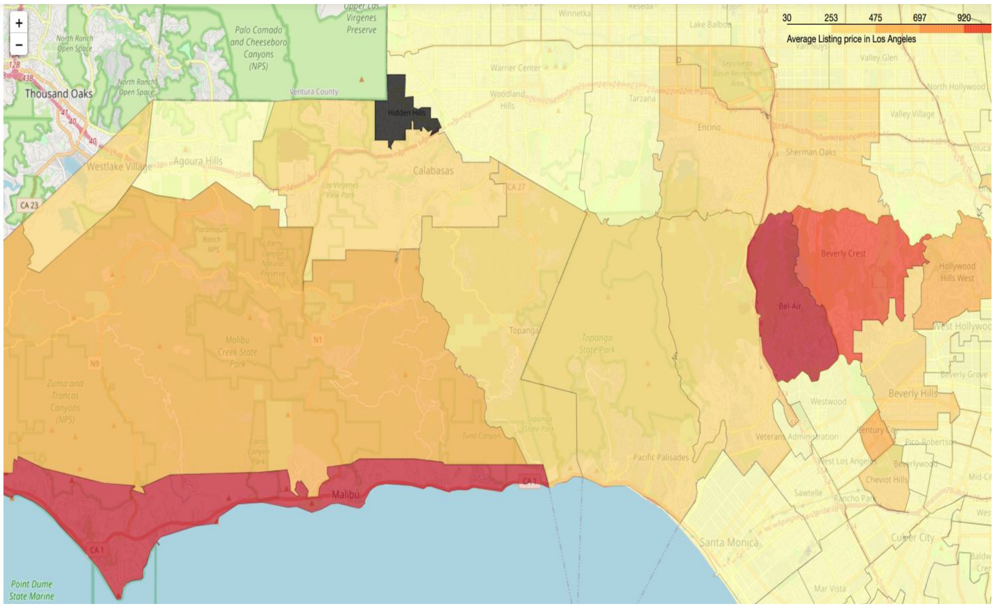


Figure 9

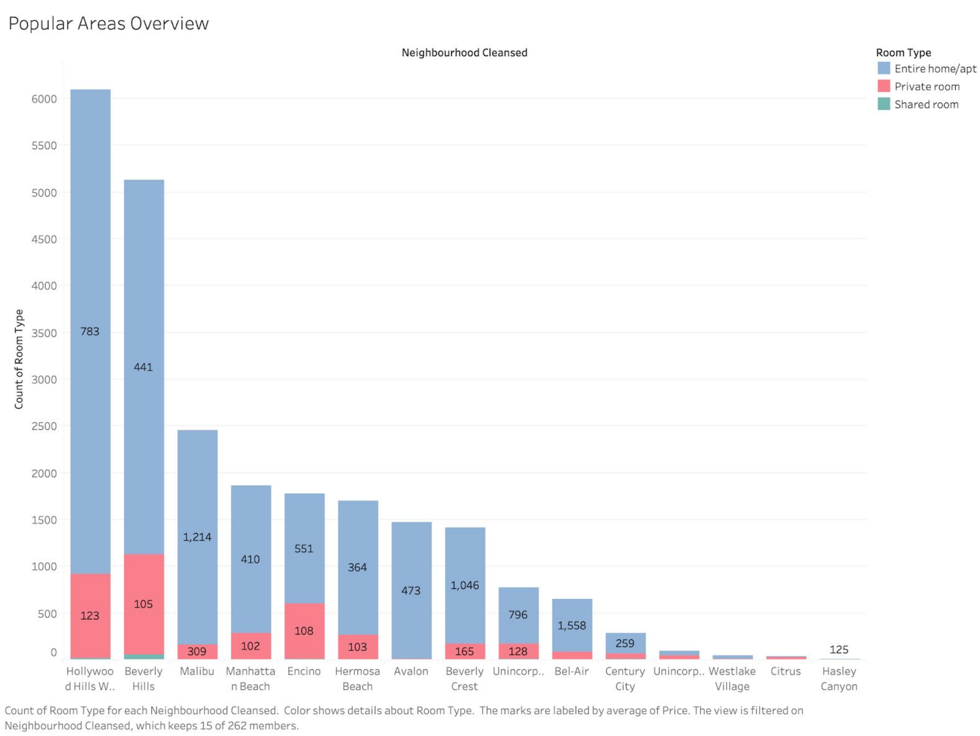
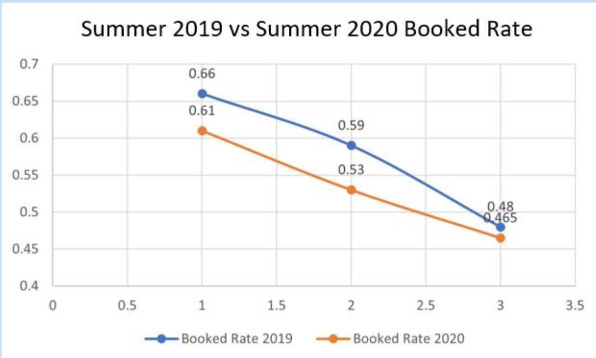
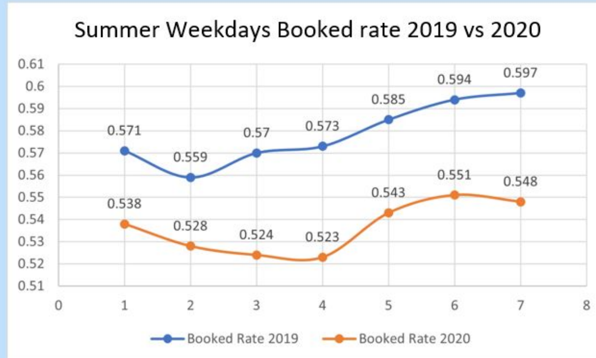


Figure 10

room_type_Entire home/apt	77.4344	36.941	2.096	0.036	5.031	149.838
room_type_Hotel room	-446.8670	159.747	-2.797	0.005	-759.966	-133.768
room_type_Private room	88.3273	36.938	2.391	0.017	15.929	160.725
room_type_Shared room	5.6587	36.988	0.153	0.878	-66.837	78.154

Figure 11






```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     290.6
Date:                Fri, 11 Dec 2020   Prob (F-statistic):          6.46e-127
Time:                        14:12:39   Log-Likelihood:            -9.7654e+07
No. Observations:            12364401   AIC:                         1.953e+08
Df Residuals:                12364398   BIC:                         1.953e+08
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
7            237.7640      0.319    745.068      0.000     237.139     238.389
8            235.0709      0.319    736.632      0.000     234.445     235.696
9            227.1881      0.324    700.361      0.000     226.552     227.824
==============================================================================
```