Decoding Airbnb in The Big Apple

Table of Contents

INTRODUCTION:

Airbnb has been one of the most successful companies since its inception in 2008. According to Airbnb Newsrooms, currently Airbnb has more than 7 million listings in more than 191 countries and regions and operating in more than 100,000 cities. As one of the most popular cities in the world, New York City has been one of the hottest markets for Airbnb. With close to 50,000 listings in the city, Airbnb has interwoven with the rental landscape within 10 years of its inception. Analyses on such a dataset would not only provide intuition about the rental metrics but also shed some light on the socio-economic setting of the city.

The aim of the project is to perform analyses on New York City Airbnb dataset and uncover insights about the sharing economy in one of the biggest cities of the world. The tasks involve developing business intelligence for both hosts who are listing their apartments and the guests who are using them to meet their accommodation requirements.

Following are the questions the project tries to answer which are split into three broad sections:

1) Insights into Airbnb

- How has Airbnb presence grown over the years?
- How costly are the Airbnb rates in the neighbourhoods across the five boroughs?
- How badly the Covid-19 crisis affect Airbnb?

2) Insights for Hosts

- What should be the rental value if you want to list your property with Airbnb?
- What are the pain points that a guest finds in Airbnb?

3) Insights for Customers

- What are the top 10 listing recommendations based on customer constraints?

DATA DESCRIPTION

The second-hand dataset is taken from Inside Airbnb which provides non-commercial set of tools and data that allows us to explore how Airbnb is really being used in cities around the world. The New York Airbnb dataset is compiled on 6 May 2020.

There are three data sets that we used for our analysis, namely –

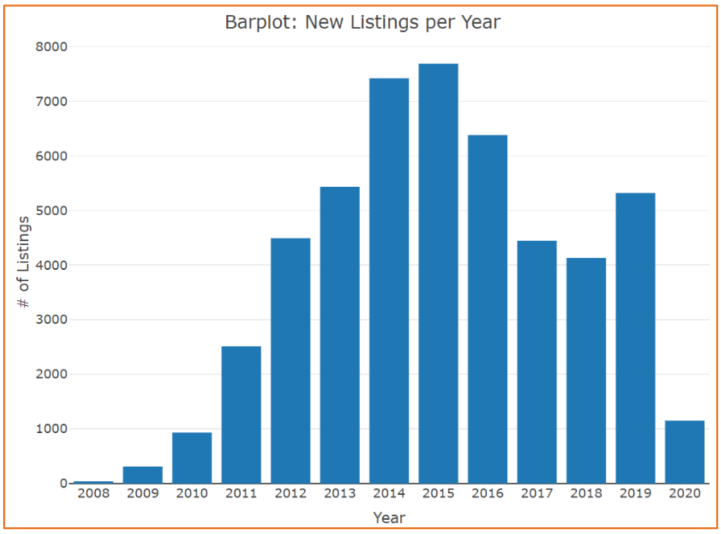
1. listings.csv – file contains 106 variables and 50,246 listing information. Details about the listings such as price, apartment details, ratings of the apartment, number of rooms, neighbourhood and host information are included in this file.
2. calendar.csv – file includes the daily rates of the listings up till a year. The data in the file was used to project the prices during the holiday season.
3. reviews.csv – file includes the reviews of each listing posted by guests. This file was primarily used for text analytics.

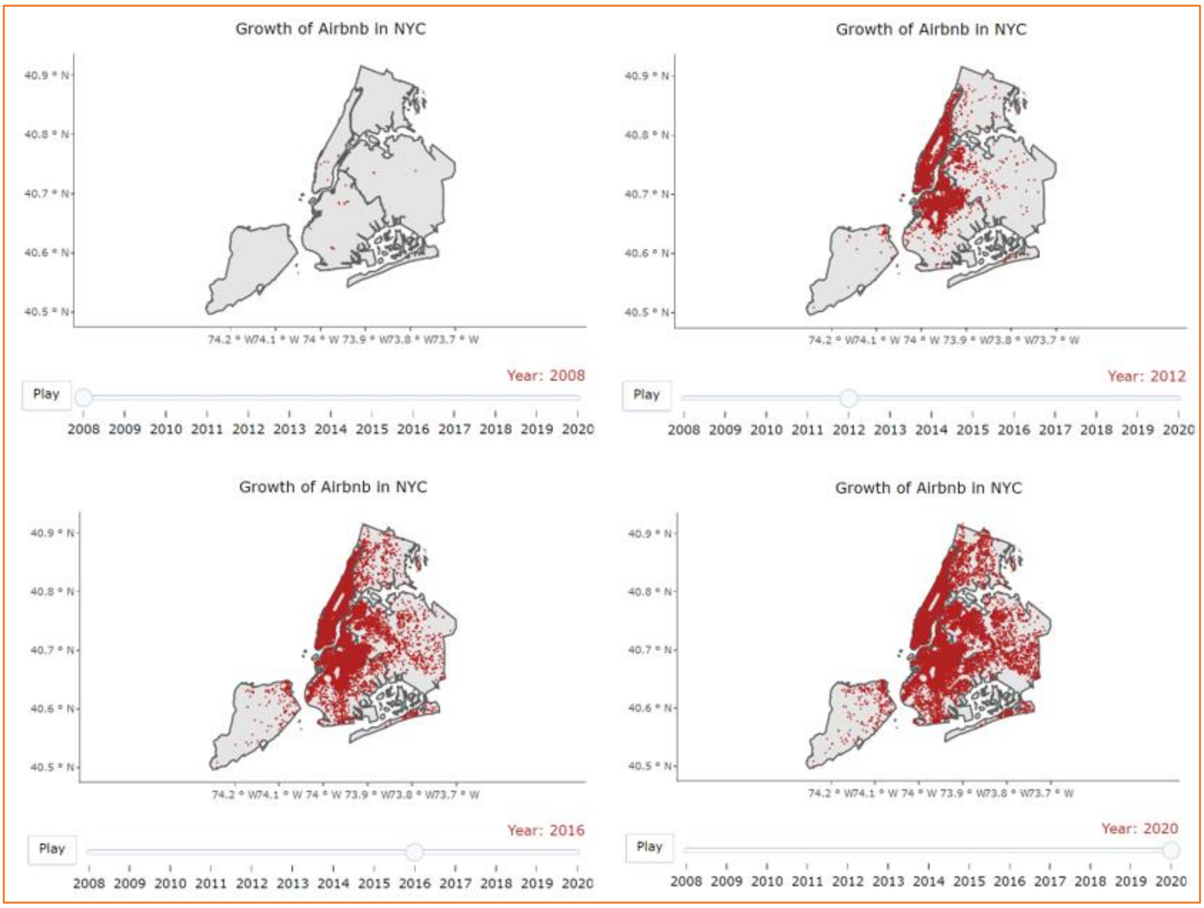
TASK (AIRBNB):

How has Airbnb presence grown over the years?

Analysis:

New York City being the most densely populous city in US, has over 50,000 Airbnb listings as of May 2020. Bar plot on the right shows that new listings in NYC increased steadily from 2008 to 2015. Post 2015 new listings started to go down and averaged around 4000 up till last year. Geo plot below shows the landscape of Airbnb listings over the years. A quick glance at the geo plot reveals that Manhattan and North Brooklyn around the East river are the most populated areas by Airbnb listings.



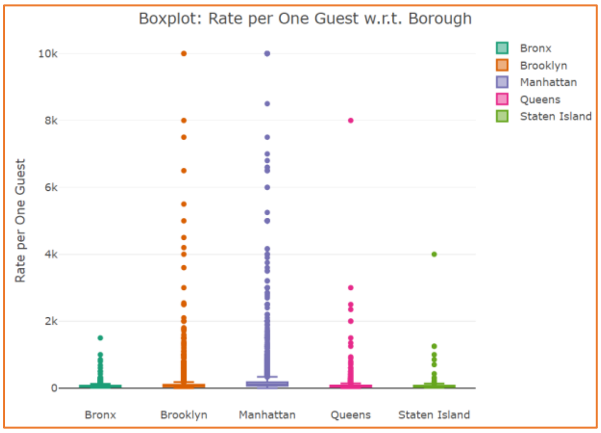


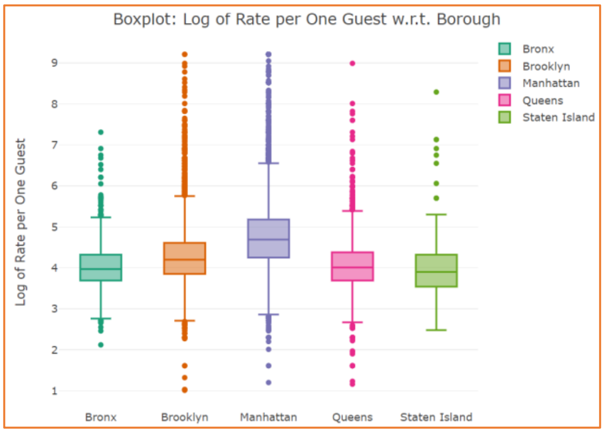
TASK (AIRBNB):

How costly are the Airbnb rates in the neighbourhoods across the five boroughs?

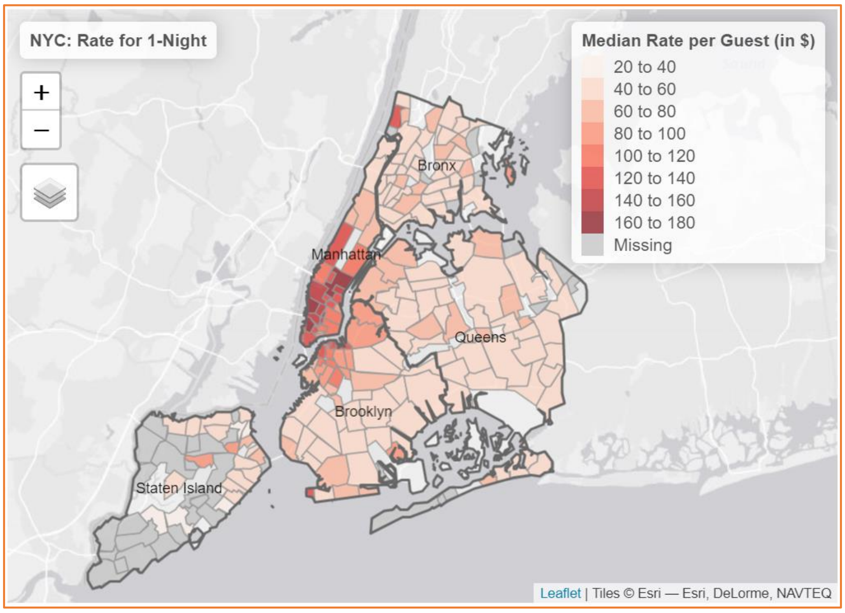
Analysis:

Since Airbnb rates are not necessarily per individual basis, it makes logical sense to standardize the rates with respect to an individual. Also, entities involving price or income generally tend to be right skewed (outliers on the higher end), median is considered to be the best measure of central tendency. To capture the data well, the logarithm of listing price per single guest is taken. The box plots with respect to the five boroughs in NYC as shown below illustrate the intuition. Coinciding with the reasoning of high cost of living in Manhattan, the Airbnb rates are similar to the expectations.

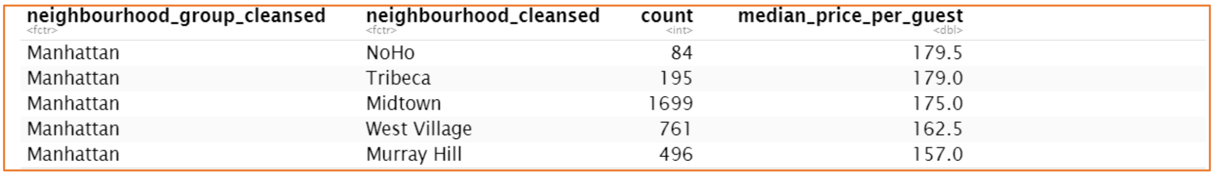




To visualize in depth pricing analysis of neighbourhoods in each borough, a heatmap of prices with respect to the neighbourhoods having minimum of 5 listings is plotted below. This provides crucial insights on the median price range of neighbourhoods. The grey area in the heat map shows neighbourhood with less than 5 listings. Most of the neighbourhoods in Staten Island have less than 5 listings probably due to its suburb nature. The region around East River including North Brooklyn and the entire Manhattan are the costliest places to rent an Airbnb in addition to the most number of listings in the region.

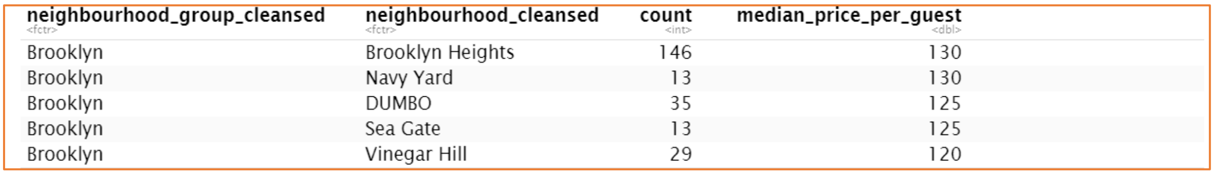


- Top 5 costly Airbnb neighbourhoods in Manhattan with median rate



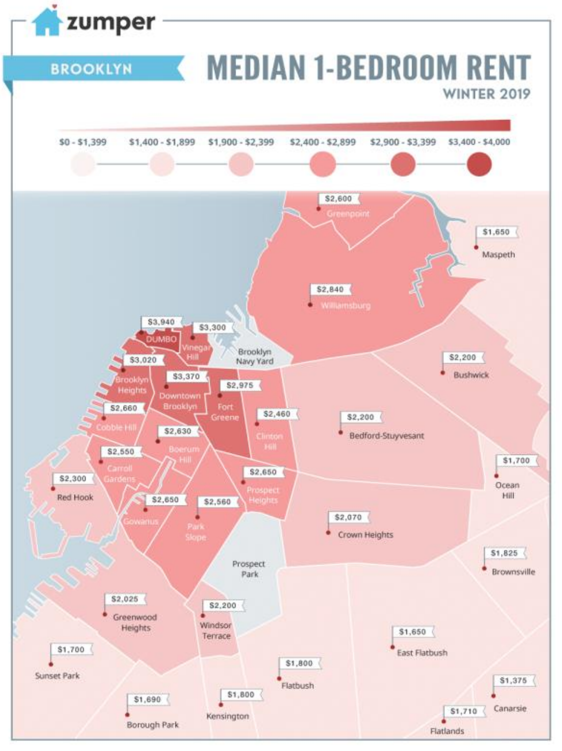
neighbourhood_group_cleansed	neighbourhood_cleansed	count	median_price_per_guest
<fctr>	<fctr>	<int>	<dbl>
Manhattan	NoHo	84	179.5
Manhattan	Tribeca	195	179.0
Manhattan	Midtown	1699	175.0
Manhattan	West Village	761	162.5
Manhattan	Murray Hill	496	157.0

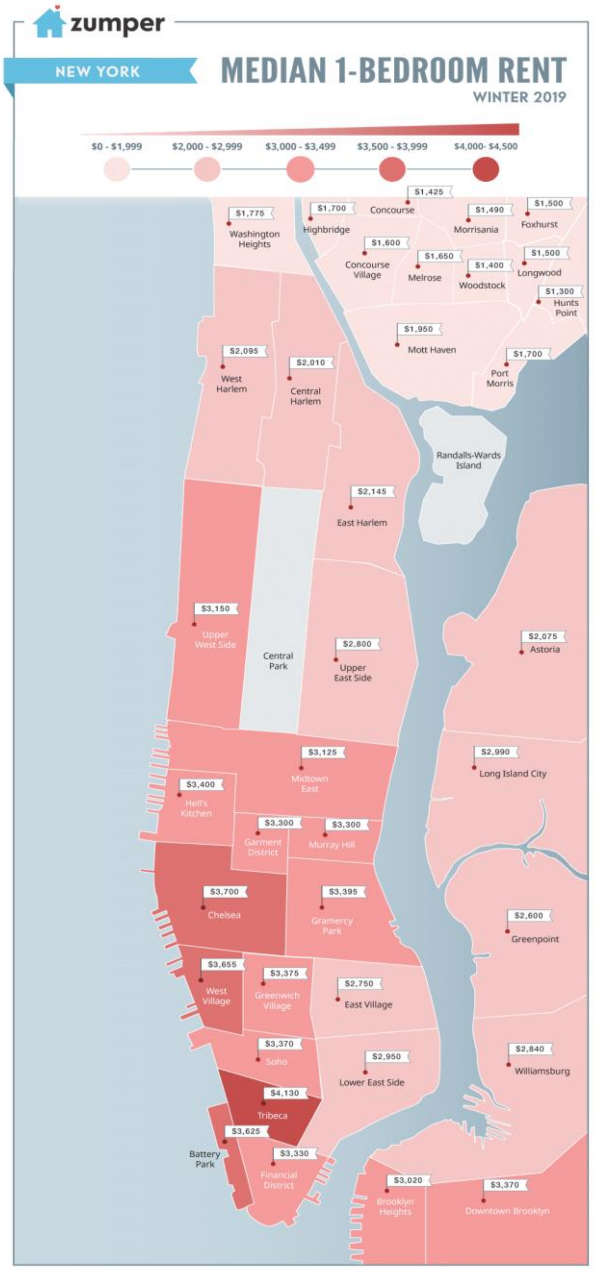
- Top 5 costly Airbnb neighbourhoods in Brooklyn with median rate



neighbourhood_group_cleansed	neighbourhood_cleansed	count	median_price_per_guest
<fctr>	<fctr>	<int>	<dbl>
Brooklyn	Brooklyn Heights	146	130
Brooklyn	Navy Yard	13	130
Brooklyn	DUMBO	35	125
Brooklyn	Sea Gate	13	125
Brooklyn	Vinegar Hill	29	120

Zumper has mapped NYC neighbourhood rents for winter 2019. and the maps show median 1-bedroom rents in Brooklyn and Manhattan. Places like Dumbo, Vinegar Hill, Brooklyn Heights, Downtown Brooklyn, and Fort Greene are costlier neighbourhoods. Similarly places like Tribeca, Battery Park, Soho, West Village, and Chelsea in Manhattan are costlier in Manhattan. This presents the real estate setting for the New York boroughs. These places form New York Skyline and is a hub for intercultural and financial activities. Both Zumper (real estate setting) and Airbnb (rental landscape) paint a similar picture.

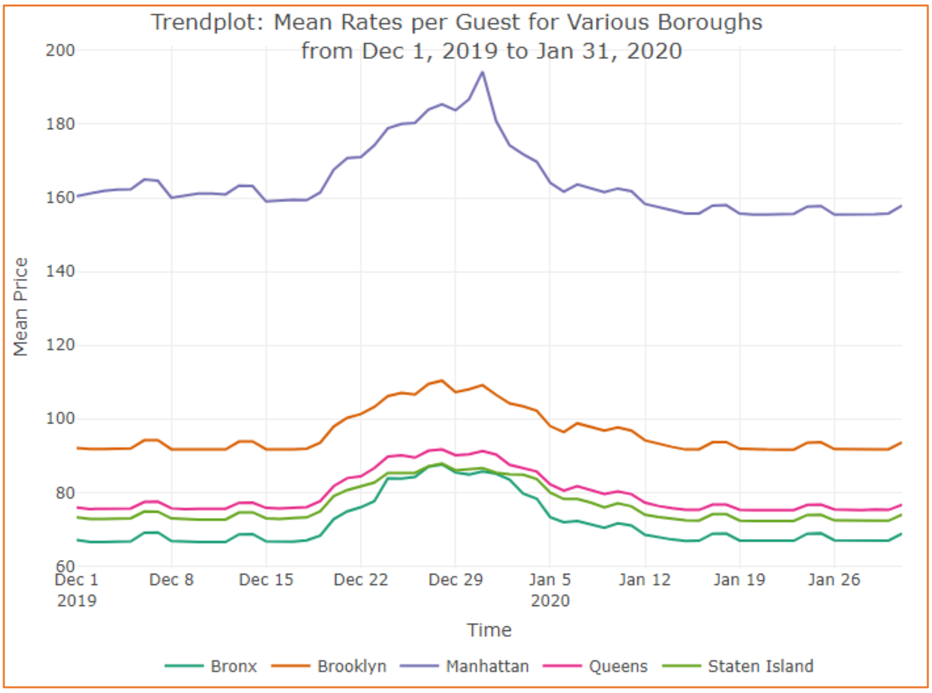


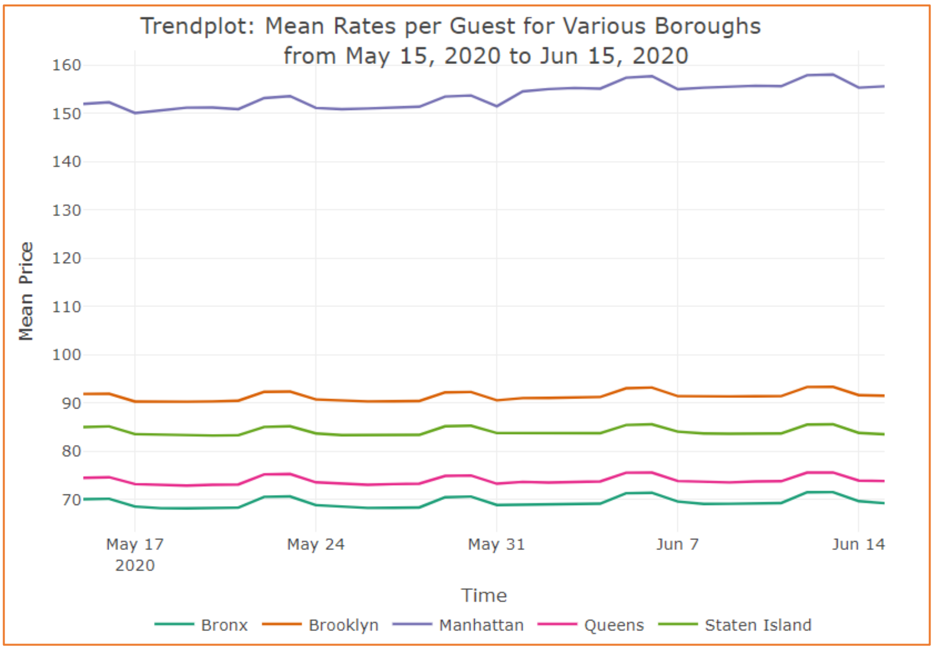


TASK (AIRBNB):

How badly the Covid-19 crisis affect Airbnb?

Analysis:





From the graphs one can grasp contrasting scenarios. As of September 12, 2019, an average person had to shell out extra 13-17% on accommodation during New Years' Week – booking almost three months in advance. Fast forward 5 months to May 06, 2020, the situation has changed dramatically. What was considered to be a peak summer season for Airbnb Rentals, the projections have changed for the worst. Covid-19 has halted most of the economic functions and recreational activities and isolation has become a new norm. Travel and hospitality industries are the worst affected due to this. As of May 06, 2020, the hosts have reduced the rents by more than 20% of what was charged during New Year's Week – that too with immediate availability.

TASK (HOSTS):

What should be the rental value if you want to list your property with Airbnb?

Analysis:

As a new host, one would like to how much his/her property can be listed with Airbnb. The analysis gives a crucial information for new hosts to estimate their listing price based on certain attributes.

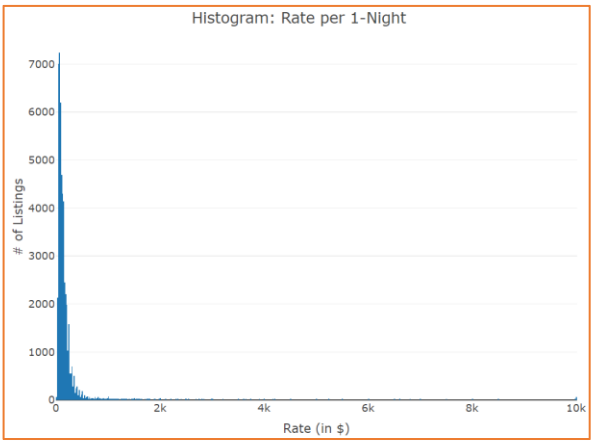
The parameters chosen are of:

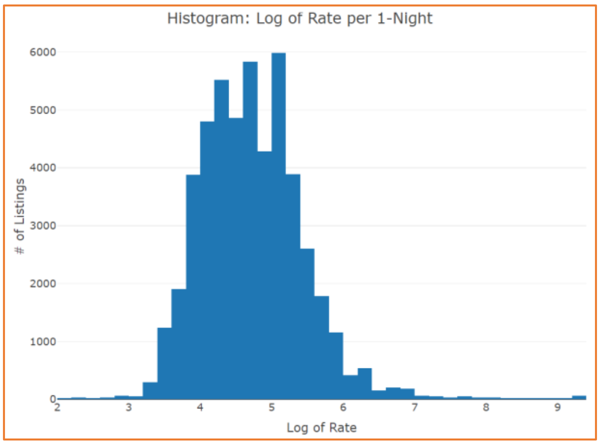
- Geographical importance: 1) borough, 2) neighbourhood
- Listing attributes: 1) property type, 2) room type, 3) number of bedrooms, 4) number of bathrooms, 5) number of guests included

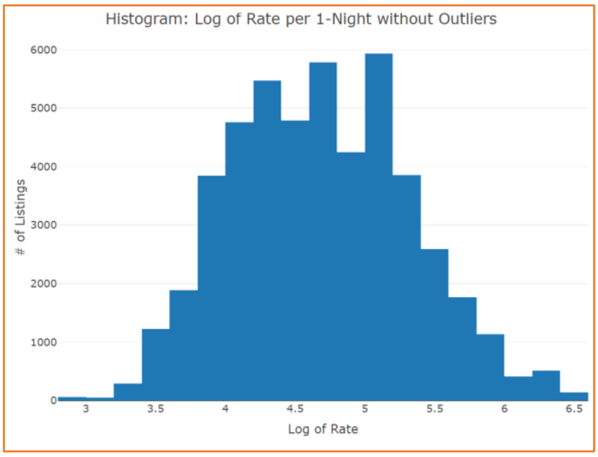
Since many of the parameters are categorical variables such as borough name, neighbourhood name, property type, and room type, we proceed with multilevel linear regression model to predict the price.

Before fitting a linear model, a careful examination of dependent variable and explanatory variables is necessary to see if the variables meet linear model assumptions such as normality. The below plots show that a log transformation reduced the skewness to a great extent but removing outliers were necessary to meet the normality assumption. Couple of filters are also applied to the dataset as part of cleaning, so that there are enough observations for each of the combination of categorical variables. Therefore, we assume following filters on the unclean dataset.

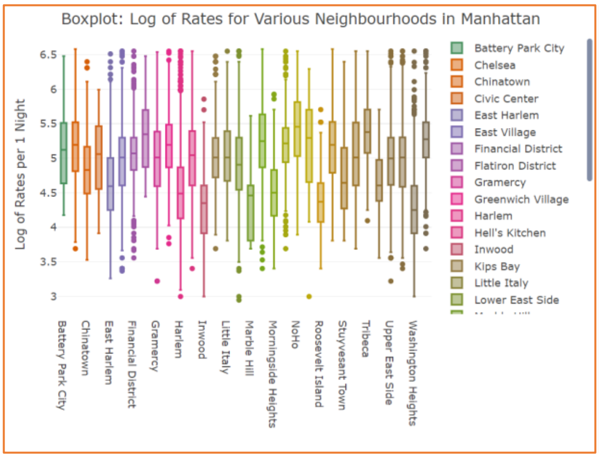
- Neighbourhood: >= 5 listings
- Property Type: >= 100 listings

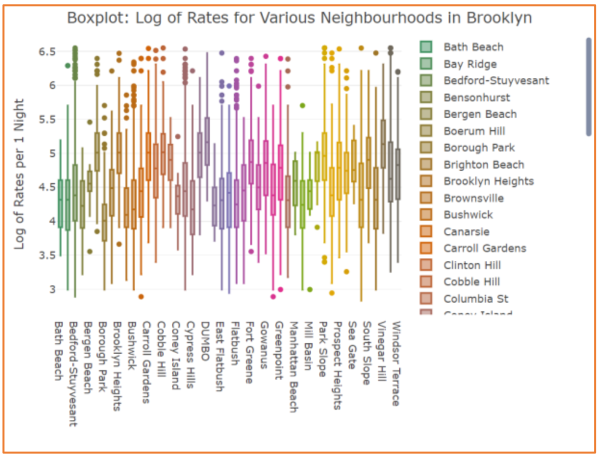


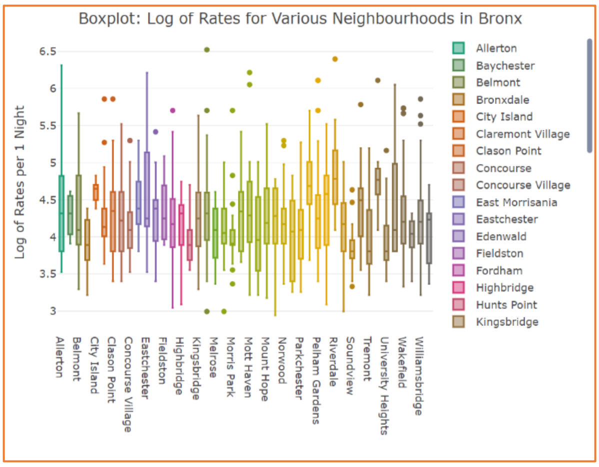


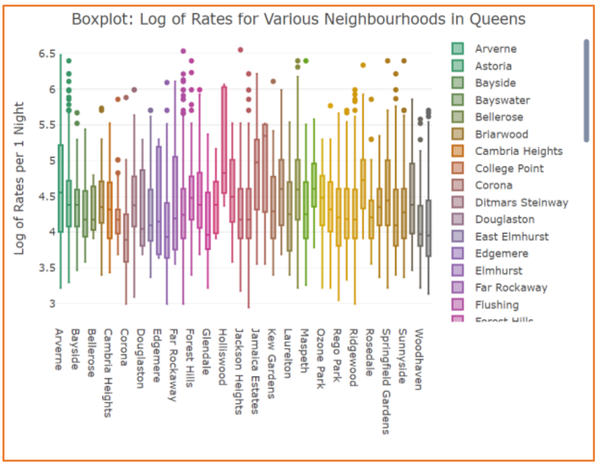


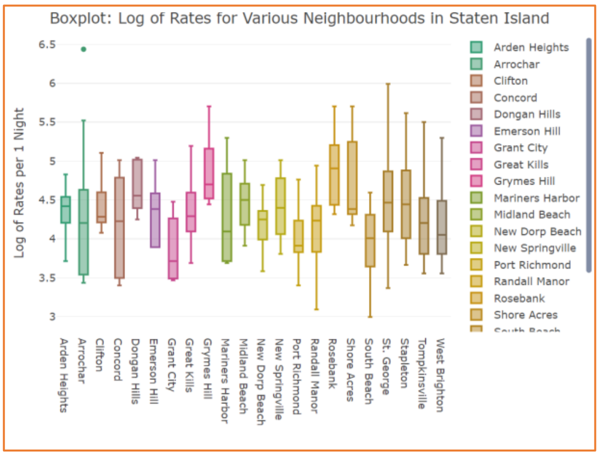
As seen earlier, the median rates for various boroughs are different and their effects need to be considered. Similarly, neighbourhoods in each of the boroughs differ in terms of median rate per one guest per one night. The plots below show how rates vary across the five boroughs namely – Manhattan, Brooklyn, Bronx, Queens, and Staten Island respectively. Although the rates in various neighbourhoods vary identically around borough averages, but it is important to see whether the neighbourhood effects are stronger than the borough effects.


Boxplot: Log of Rates for Various Neighbourhoods in Manhattan

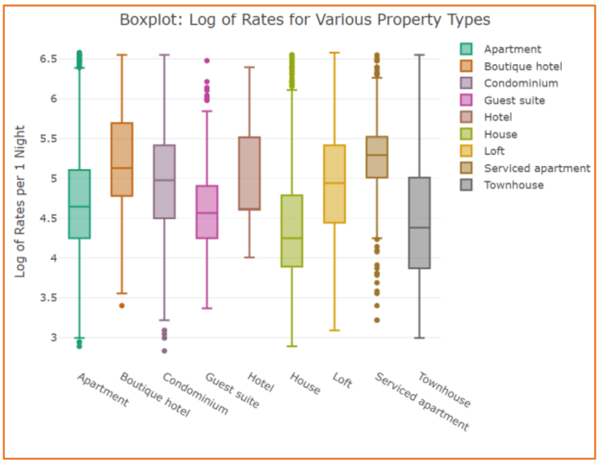

Boxplot: Log of Rates for Various Neighbourhoods in Brooklyn

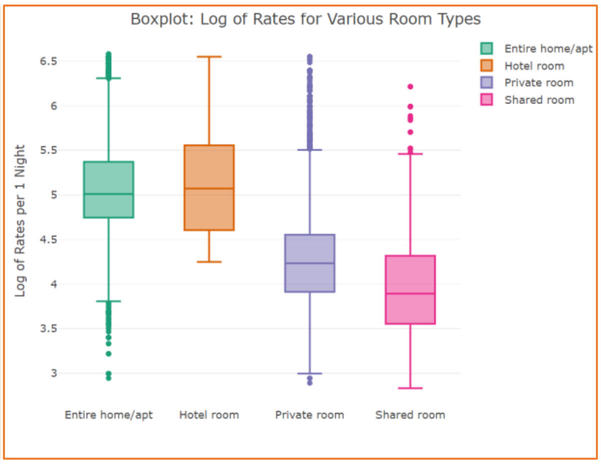

Boxplot: Log of Rates for Various Neighbourhoods in Bronx


Boxplot: Log of Rates for Various Neighbourhoods in Queens

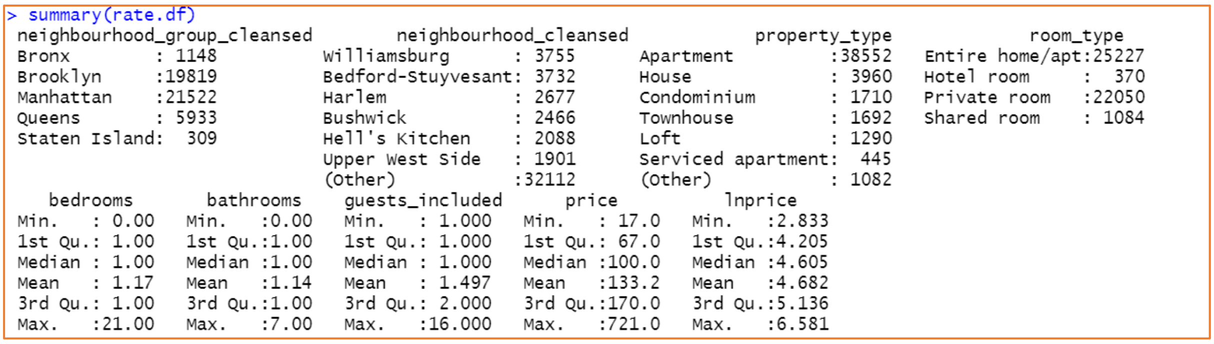

Boxplot: Log of Rates for Various Neighbourhoods in Staten Island

Around 80% of all the properties listed with Airbnb are apartments, aligning to the company's main idea of lodgings and homestays. Hotels and Serviced Apartments tend to be costlier, adhering to the general notion. Airbnb also provides private and shared rooms for cheaper accommodation options with shared rooms only accounting for 2% of all the registered listings.



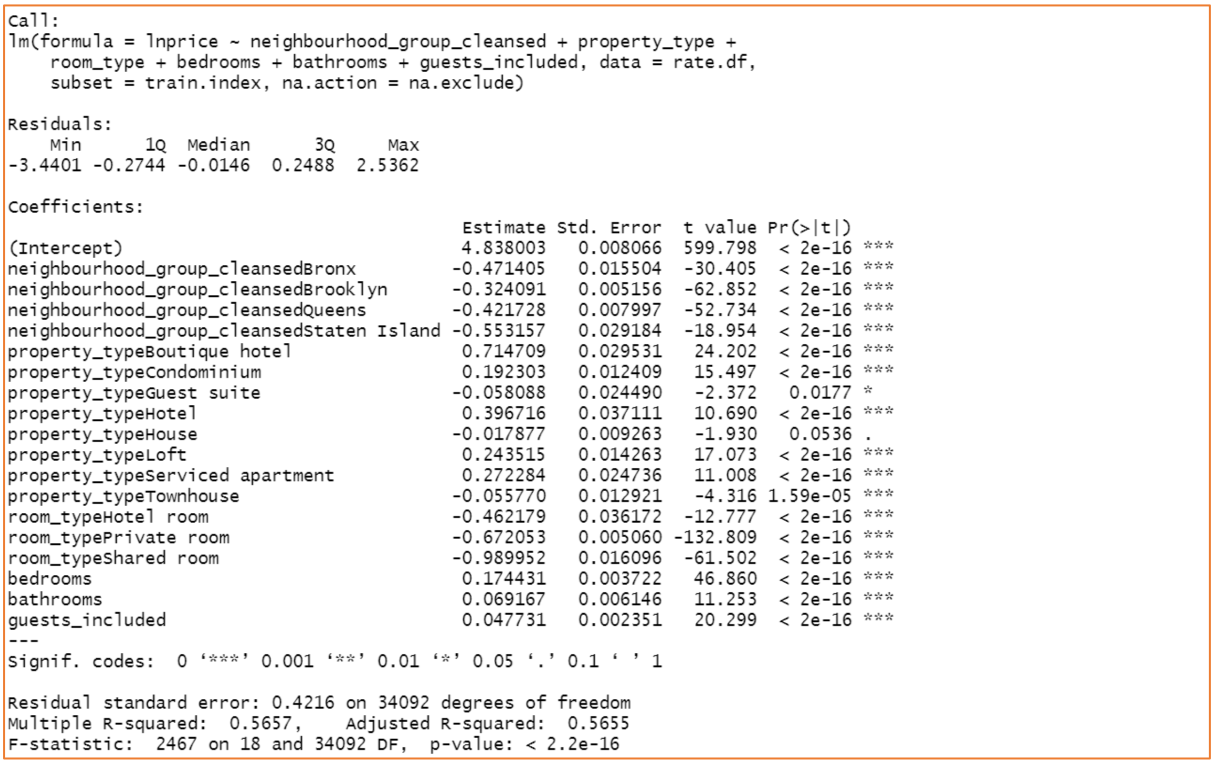


As a result of cleaning and filtering, records in the dataset are reduced by around 1500. The base levels are: 1) Borough (Manhattan), 2) Neighbourhood (Harlem), 3) Property Type (Apartment), 4) Room Type (Entire home/Apartment).

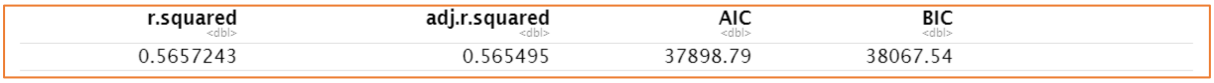


```
> summary(rate.df)
 neighbourhood_group_cleansed         neighbourhood_cleansed             property_type                room_type
 Bronx         : 1148          Williamsburg       : 3755      Apartment         :38552     Entire home/apt:25227
 Brooklyn      :19819          Bedford-Stuyvesant: 3732      House             : 3960     Hotel room     :  370
 Manhattan     :21522          Harlem             : 2677      Condominium       : 1710     Private room   :22050
 Queens        : 5933          Bushwick           : 2466      Townhouse         : 1692     Shared room    : 1084
 Staten Island:  309          Hell's Kitchen     : 2088      Loft              : 1290
                               Upper West Side    : 1901      Serviced apartment:  445
                               (Other)            :32112      (Other)           : 1082
    bedrooms        bathrooms      guests_included      price          lnprice
 Min.   : 0.00   Min.   :0.00   Min.   : 1.000   Min.   : 17.0   Min.   :2.833
 1st Qu.: 1.00   1st Qu.:1.00   1st Qu.: 1.000   1st Qu.: 67.0   1st Qu.:4.205
 Median : 1.00   Median :1.00   Median : 1.000   Median :100.0   Median :4.605
 Mean   : 1.17   Mean   :1.14   Mean   : 1.497   Mean   :133.2   Mean   :4.682
 3rd Qu.: 1.00   3rd Qu.:1.00   3rd Qu.: 2.000   3rd Qu.:170.0   3rd Qu.:5.136
 Max.   :21.00   Max.   :7.00   Max.   :16.000   Max.   :721.0   Max.   :6.581
```

One of the important factors in choosing regressors is to explain the model in a simpler way. Running a multilevel linear regression on the dataset with boroughs results in adjusted R^2 of 56.55%.



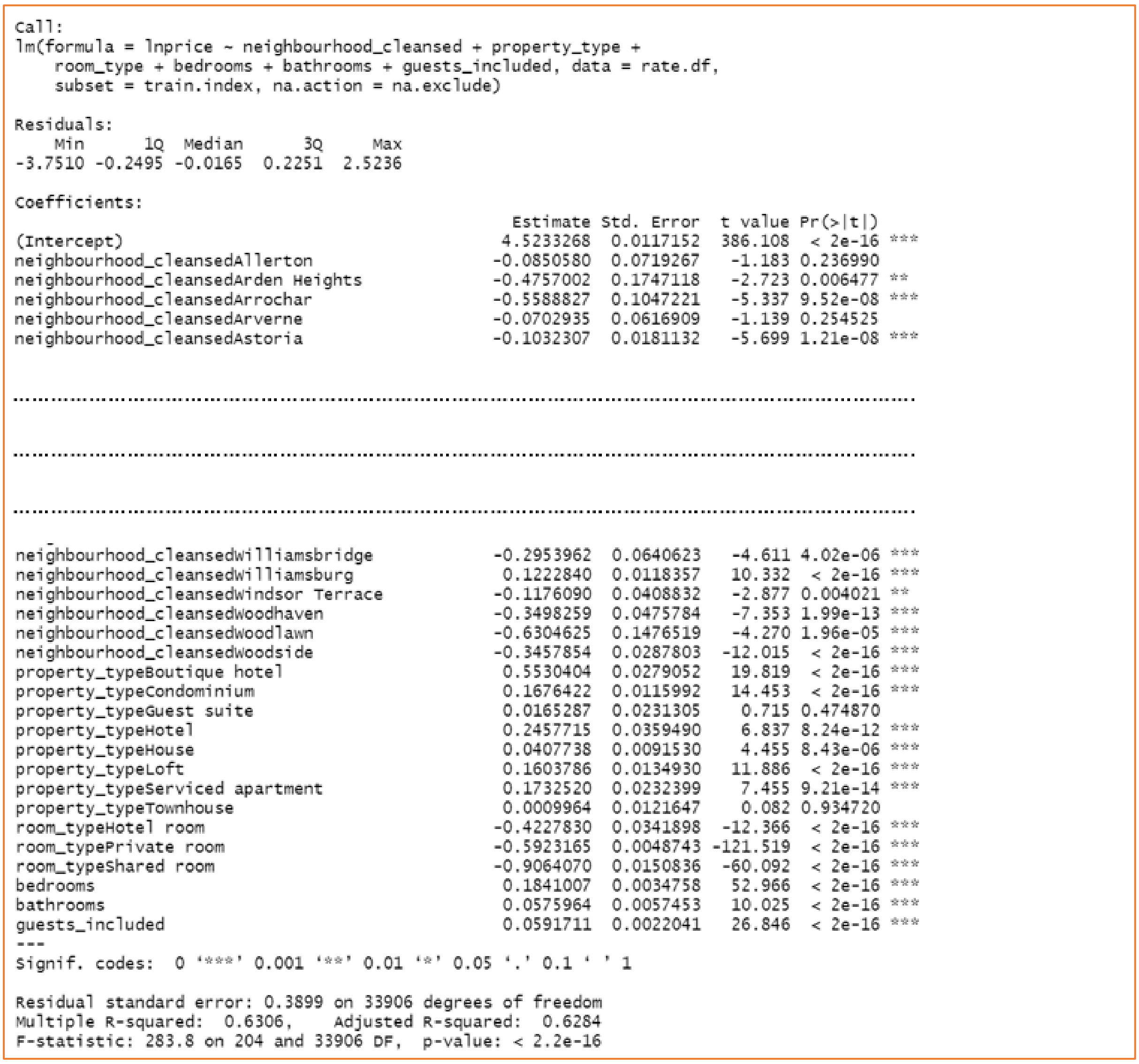
```
Call:
lm(formula = lnprice ~ neighbourhood_group_cleansed + property_type + 
    room_type + bedrooms + bathrooms + guests_included, data = rate.df, 
    subset = train.index, na.action = na.exclude)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.4401 -0.2744 -0.0146  0.2488  2.5362 

Coefficients:
                                           Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                                4.838003   0.008066  599.798  < 2e-16 ***
neighbourhood_group_cleansedBronx         -0.471405   0.015504  -30.405  < 2e-16 ***
neighbourhood_group_cleansedBrooklyn      -0.324091   0.005156  -62.852  < 2e-16 ***
neighbourhood_group_cleansedQueens        -0.421728   0.007997  -52.734  < 2e-16 ***
neighbourhood_group_cleansedStaten Island -0.553157   0.029184  -18.954  < 2e-16 ***
property_typeBoutique hotel                0.714709   0.029531   24.202  < 2e-16 ***
property_typeCondominium                   0.192303   0.012409   15.497  < 2e-16 ***
property_typeGuest suite                  -0.058088   0.024490   -2.372   0.0177 *  
property_typeHotel                         0.396716   0.037111   10.690  < 2e-16 ***
property_typeHouse                        -0.017877   0.009263   -1.930   0.0536 .  
property_typeLoft                          0.243515   0.014263   17.073  < 2e-16 ***
property_typeServiced apartment            0.272284   0.024736   11.008  < 2e-16 ***
property_typeTownhouse                    -0.055770   0.012921   -4.316 1.59e-05 ***
room_typeHotel room                       -0.462179   0.036172  -12.777  < 2e-16 ***
room_typePrivate room                     -0.672053   0.005060 -132.809  < 2e-16 ***
room_typeShared room                      -0.989952   0.016096  -61.502  < 2e-16 ***
bedrooms                                   0.174431   0.003722   46.860  < 2e-16 ***
bathrooms                                  0.069167   0.006146   11.253  < 2e-16 ***
guests_included                            0.047731   0.002351   20.299  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4216 on 34092 degrees of freedom
Multiple R-squared:  0.5657,	Adjusted R-squared:  0.5655 
F-statistic:  2467 on 18 and 34092 DF,  p-value: < 2.2e-16
```



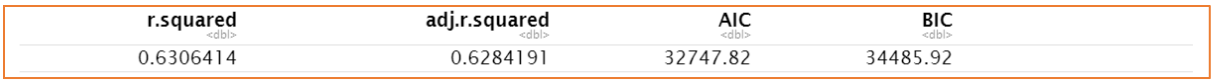
r.squared <dbl>	adj.r.squared <dbl>	AIC <dbl>	BIC <dbl>
0.5657243	0.565495	37898.79	38067.54

The signs of the coefficients on boroughs, room types, # of bedrooms, # of bathrooms and # of guests are as expected and are significant. Except for guest suite and house, the coefficients of other property types are positive and significant.

Although the previous model explains the variation fairly decent, we also need to see whether there is any significant improvement in model fit if neighbourhood effects are considered over borough effects. By including neighbourhood effects, the adjusted R^2 increased to 62.84%. Similarly, AIC and BIC values have also reduced. The signs and magnitudes of the coefficients on boroughs, room types, # of bedrooms and # of bathrooms remain almost same. Coefficients for guest suite and townhouse are similar to that of apartment.



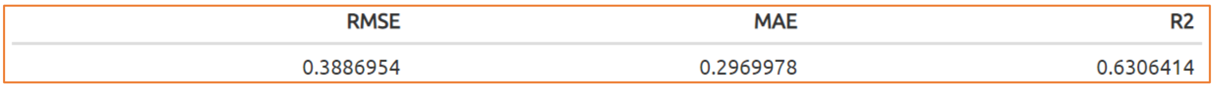
```
Call:
lm(formula = lnprice ~ neighbourhood_cleansed + property_type + 
    room_type + bedrooms + bathrooms + guests_included, data = rate.df, 
    subset = train.index, na.action = na.exclude)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.7510 -0.2495 -0.0165  0.2251  2.5236 

Coefficients:
                                               Estimate Std. Error  t value Pr(>|t|)    
(Intercept)                                   4.5233268  0.0117152  386.108  < 2e-16 ***
neighbourhood_cleansedAllerton               -0.0850580  0.0719267   -1.183 0.236990    
neighbourhood_cleansedArden Heights          -0.4757002  0.1747118   -2.723 0.006477 ** 
neighbourhood_cleansedArrochar               -0.5588827  0.1047221   -5.337 9.52e-08 ***
neighbourhood_cleansedArverne                -0.0702935  0.0616909   -1.139 0.254525    
neighbourhood_cleansedAstoria                -0.1032307  0.0181132   -5.699 1.21e-08 ***


.........................................................................................


.........................................................................................


.........................................................................................

neighbourhood_cleansedWilliamsbridge         -0.2953962  0.0640623   -4.611 4.02e-06 ***
neighbourhood_cleansedWilliamsburg            0.1222840  0.0118357   10.332  < 2e-16 ***
neighbourhood_cleansedWindsor Terrace        -0.1176090  0.0408832   -2.877 0.004021 ** 
neighbourhood_cleansedWoodhaven              -0.3498259  0.0475784   -7.353 1.99e-13 ***
neighbourhood_cleansedWoodlawn               -0.6304625  0.1476519   -4.270 1.96e-05 ***
neighbourhood_cleansedWoodside               -0.3457854  0.0287803  -12.015  < 2e-16 ***
property_typeBoutique hotel                   0.5530404  0.0279052   19.819  < 2e-16 ***
property_typeCondominium                      0.1676422  0.0115992   14.453  < 2e-16 ***
property_typeGuest suite                      0.0165287  0.0231305    0.715 0.474870    
property_typeHotel                            0.2457715  0.0359490    6.837 8.24e-12 ***
property_typeHouse                            0.0407738  0.0091530    4.455 8.43e-06 ***
property_typeLoft                             0.1603786  0.0134930   11.886  < 2e-16 ***
property_typeServiced apartment               0.1732520  0.0232399    7.455 9.21e-14 ***
property_typeTownhouse                        0.0009964  0.0121647    0.082 0.934720    
room_typeHotel room                          -0.4227830  0.0341898  -12.366  < 2e-16 ***
room_typePrivate room                        -0.5923165  0.0048743 -121.519  < 2e-16 ***
room_typeShared room                         -0.9064070  0.0150836  -60.092  < 2e-16 ***
bedrooms                                      0.1841007  0.0034758   52.966  < 2e-16 ***
bathrooms                                     0.0575964  0.0057453   10.025  < 2e-16 ***
guests_included                               0.0591711  0.0022041   26.846  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3899 on 33906 degrees of freedom
Multiple R-squared:  0.6306,	Adjusted R-squared:  0.6284 
F-statistic: 283.8 on 204 and 33906 DF,  p-value: < 2.2e-16
```



r.squared <dbl>	adj.r.squared <dbl>	AIC <dbl>	BIC <dbl>
0.6306414	0.6284191	32747.82	34485.92

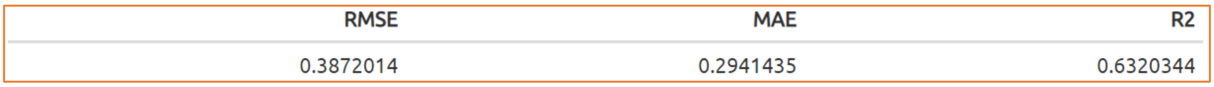
Both Forward and Backward Selection have chosen the previous model with neighbourhood effects as the best and hence this full model is chosen for prediction. Even with a low R-squared, statistically significant p-values continue to identify relationships and coefficients have the same interpretation.

Performance Metrics for Training: 70% split



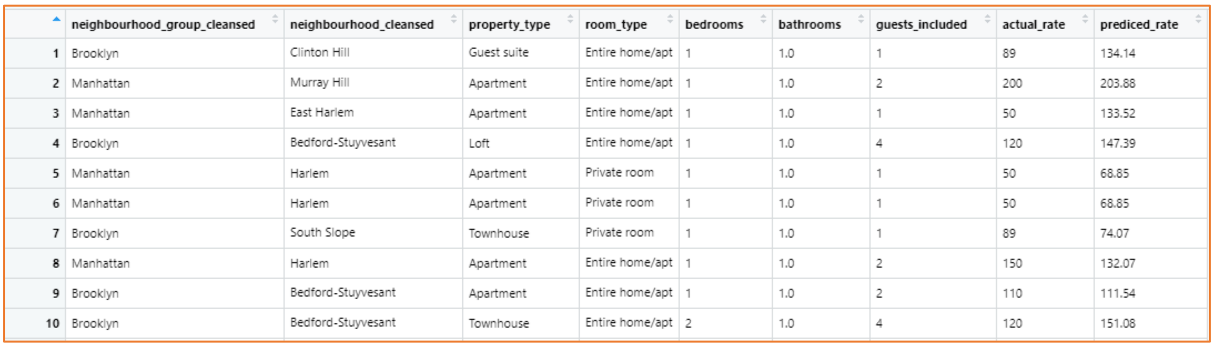
RMSE	MAE	R2
0.3886954	0.2969978	0.6306414

Performance Metrics for Test: 30% split



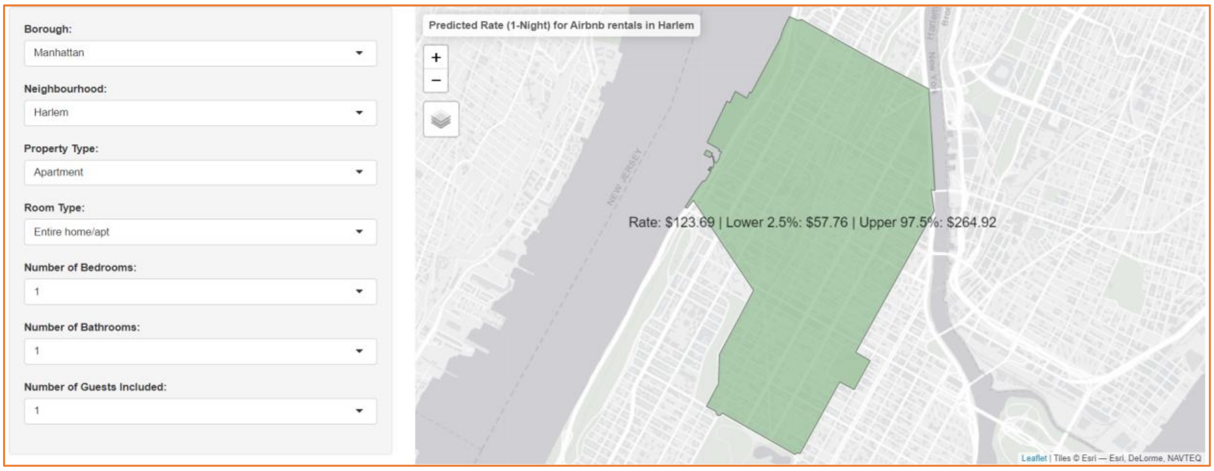
RMSE	MAE	R2
0.3872014	0.2941435	0.6320344

Prediction:



	neighbourhood_group_cleansed	neighbourhood_cleansed	property_type	room_type	bedrooms	bathrooms	guests_included	actual_rate	prediced_rate
1	Brooklyn	Clinton Hill	Guest suite	Entire home/apt	1	1.0	1	89	134.14
2	Manhattan	Murray Hill	Apartment	Entire home/apt	1	1.0	2	200	203.88
3	Manhattan	East Harlem	Apartment	Entire home/apt	1	1.0	1	50	133.52
4	Brooklyn	Bedford-Stuyvesant	Loft	Entire home/apt	1	1.0	4	120	147.39
5	Manhattan	Harlem	Apartment	Private room	1	1.0	1	50	68.85
6	Manhattan	Harlem	Apartment	Private room	1	1.0	1	50	68.85
7	Brooklyn	South Slope	Townhouse	Private room	1	1.0	1	89	74.07
8	Manhattan	Harlem	Apartment	Entire home/apt	1	1.0	2	150	132.07
9	Brooklyn	Bedford-Stuyvesant	Apartment	Entire home/apt	1	1.0	2	110	111.54
10	Brooklyn	Bedford-Stuyvesant	Townhouse	Entire home/apt	2	1.0	4	120	151.08

Below is the user interface for hosts to input parameters for suggesting the price and 95% prediction interval at which they can register their listing.



Borough: Manhattan

Neighbourhood: Harlem

Property Type: Apartment

Room Type: Entire home/apt

Number of Bedrooms: 1

Number of Bathrooms: 1

Number of Guests Included: 1

Predicted Rate (1-Night) for Airbnb rentals in Harlem

Rate: $123.69 | Lower 2.5%: $57.76 | Upper 97.5%: $264.92

Choosing the best explainable model:

The parameters chosen in the linear regression are run on 1) Tree-based models like Decision Tree, Random Forest, AdaBoosted Decision Tree, and XGBoost (Gradient Boosting Framework) and 2) Neural Network model to predict the logarithm of price variable. Grid Search is done to choose the hyper-parameters that lead to the best model with lower RMSE (or greater negative RMSE) on 5-fold Cross Validation Set. The training set of 70% is used to fit the model and test of 30% is utilized to evaluate the performance of the model. The following tables contain the information regarding best hyper-parameters of the model and its performance metrics on test set such as RMSE, MAE and R^2.

- Tree-based Models:

Model	**Hyper-Parameters**	**Negative CV RMSE**	**Test RMSE**	**Test MAE**	**Test R^2**
Decision Tree	max_depth=15, min_samples_leaf=8	-0.3947	0.3905	0.2981	0.6268
Random Forest	n_estimators=100	-0.3959	0.3866	0.2914	0.6343
AdaBoosted Decision Tree	n_estimators=50 , learning_rate=0.1	-0.4412	0.4400	0.3424	0.5262
XGBoost (GBM Framework)	objective=reg:squarederror, learning_rate=0.1, n_estimators=1000, max_depth=5, min_child_weight = 1, gamma=0, subsample=0.8, colsample_bytree=0.8, scale_pos_weight=1	–	0.3806	0.2896	0.6455

- Neural Network Model:

Hyper-Parameters	**Test RMSE**	**Test MAE**	**Test R^2**
batch_size=100, epochs=10 **Hidden Layer 1:** neurons=10, kernel_initializer=normal, activation=ReLU **Hidden Layer 2:** neurons=5, kernel_initializer=normal, activation=ReLU **Output Layer:** neurons=1, kernel_initializer=normal, activation=ReLU **Compiler:** optimizer=Adam (learning_rate=0.1), loss=MeanSquaredError, metrics=MeanSquaredError	0.3890	0.2969	0.6327

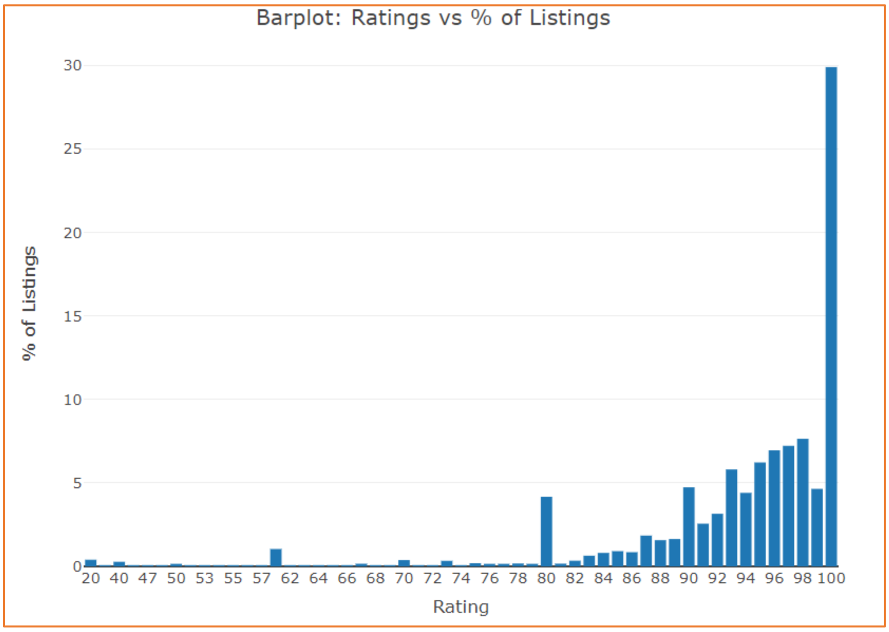
All the models performed similar to that of linear regression on the test set. XGBoost with a tree-based booster has the best test metric. However, XGBoost model did not significantly perform better than the multilevel linear regression model. In terms of explainability and interpretability, linear regression pips tree-based and neural network models. Linear model's Test R^2 is similar to Train R^2 suggesting the generalization of model. Therefore, linear regression model is chosen as the final model to predict listing rates.

TASK (HOSTS):

What are the pain points that a guest finds in Airbnb listings based on the basis of reviews?

Analysis:

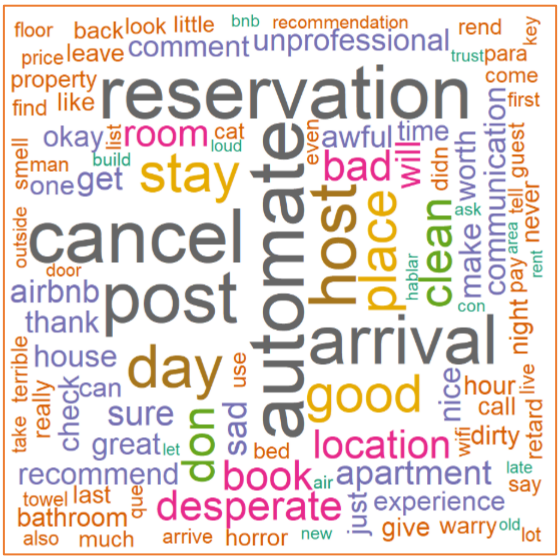
It is imperative for hosts to understand the customer expectations. Since most of Airbnb hosts are a part of informal sector in hospitality industry, it is important for them to provide service which is on par with those of formal sector. Reviews provide a feedback to the hosts on how the stay was and what can be improved, if necessary. Text analytics on the reviews of listings with poor ratings (i.e., ratings less than 50%) would provide crucial insights about bad customer experience.



The plot on the right shows that customers tend to give high ratings because people generally like to say good things. But bad rating means that there are some major issues with the Airbnb rental. Around 300 listings have net ratings less than 50%.

For this task, reviews are tokenized, lemmatized, and void of stop words as part of data cleaning. A TF-IDF matrix is constructed on the processed reviews. An interpretation of the word cloud reveals that the word ‘host’ appears possibly hinting a disconnect between the customer and the host. ‘Reservation’ and ‘cancel’ suggest that hosts do not honour their commitment.

A better designed word cloud with sentiment factor can give superior insights and help create guidelines for onboarding new hosts to warn them of potential do’s and don’ts.



TASK (CUSTOMERS):

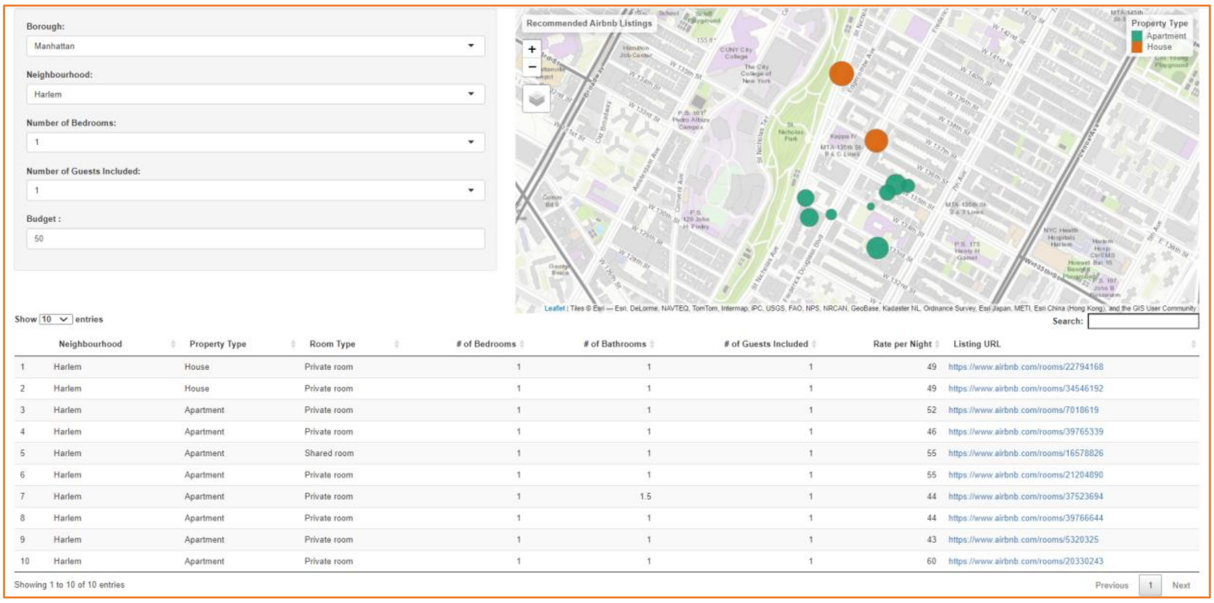
What are the top 10 listing recommendations based on customer constraints?

Analysis:

As a customer, one would like to get recommendations for their given budget and other constraints such as number of bedrooms and number of guests included.

To proceed with this analysis, top 100 locations are selected which are in close proximity to the neighbourhood centres that the user has selected. Then they are ranked according to the Euclidean distances calculated on the three scaled parameters, namely – # of bedrooms, # of guests included and rate of the listing per day. Standardization is done to make sure that the data is internally consistent i.e., each variable has equal dominating effect in recommending the output. Caution is observed while using the rate variable. Euclidean distances are calculated on the log-transformed rate that are per single guest i.e., entire rate is divided with number of guests that were included in the listing record and then it is log-transformed. Top 10 records are then recommended to the customer.

Below is the user interface for the customer to choose parameters and to see the suggestions graphically on a map. The larger the size of the bubble, the better the match. The populated dataset contains the detailed information about the recommended listings with decreasing order of priority.



Borough: Manhattan

Neighbourhood: Harlem

Number of Bedrooms: 1

Number of Guests Included: 1

Budget : 50

Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

Show 10 entries

Search:

	Neighbourhood	Property Type	Room Type	# of Bedrooms	# of Bathrooms	# of Guests Included	Rate per Night	Listing URL
1	Harlem	House	Private room	1	1	1	49	https://www.airbnb.com/rooms/22794168
2	Harlem	House	Private room	1	1	1	49	https://www.airbnb.com/rooms/34546192
3	Harlem	Apartment	Private room	1	1	1	52	https://www.airbnb.com/rooms/7018619
4	Harlem	Apartment	Private room	1	1	1	46	https://www.airbnb.com/rooms/39765339
5	Harlem	Apartment	Shared room	1	1	1	55	https://www.airbnb.com/rooms/16578826
6	Harlem	Apartment	Private room	1	1	1	55	https://www.airbnb.com/rooms/21204890
7	Harlem	Apartment	Private room	1	1.5	1	44	https://www.airbnb.com/rooms/37523694
8	Harlem	Apartment	Private room	1	1	1	44	https://www.airbnb.com/rooms/39766644
9	Harlem	Apartment	Private room	1	1	1	43	https://www.airbnb.com/rooms/5320325
10	Harlem	Apartment	Private room	1	1	1	60	https://www.airbnb.com/rooms/20330243

Showing 1 to 10 of 10 entries

Previous 1 Next

CONCLUSION

The massive dataset has a lot of insights to offer. What has been presented in this report is tip of an iceberg. The dataset provided key insights into how Airbnb grew in New York City, especially in the boroughs of Manhattan and Brooklyn. The rental landscape painted the same picture as the real estate setting of New York. Insights into various listing attributes led to the development of a multilevel linear regression model that help hosts to list their new properties for a suitable price range. Text analytics on the reviews of low rated listings has suggested that customers hate when the hosts do not honour their commitment and cancel reservations. For customers, top 10 Airbnb rental recommendations were suggested based on their constraints. However, during these testing times the hospitality sector is badly hit. For hosts who occasionally rent out their spare room in the style of a real bed & breakfast, the lost Airbnb income due the Covid-19 is a frustration.

REFERENCES

- *Get the Data*. http://insideairbnb.com/get-the-data.html
- *About Us*. https://press.airbnb.com/en-us/about-us/
- *About Inside Airbnb*. http://insideairbnb.com/about.html
- *Mapped: New York City Neighborhood Rent Prices (Winter 2019).* https://www.zumper.com/blog/2019/01/mapped-new-york-city-neighborhood-rent-prices-winter-2019/
- *Visualizing Geospatial Data in R*. https://www.datacamp.com/courses/working-with-geospatial-data-in-r