

# Edx Capstone - Airbnb Analysis Report

Eshna Airon

November 17,2020

## Contents

<b>1 Executive Summary</b>	<b>2</b>
<b>2 Dataset</b>	<b>2</b>
<b>3 Analysis</b>	<b>2</b>
3.1 Descriptive Analysis . . . . .	2
3.2 Exploratory Data Anaylsis . . . . .	12
3.3 Geographic analysis . . . . .	16
<b>4 Modeling</b>	<b>19</b>
4.1 Partial Least-Squares Regression (PLS) . . . . .	21
4.2 Boosted Generalized Linear Model . . . . .	21
4.3 Pruned Tree Models . . . . .	22
4.4 Bagged CART . . . . .	22
4.5 Random Forest . . . . .	22
4.6 Stochastic Gradient Boosting . . . . .	23
4.7 K-Nearest Neighbour . . . . .	23
4.8 Bagged Multivariate Adaptive Regression Spline . . . . .	23
4.9 Ridge Regression . . . . .	24
4.10 Lasso Regression . . . . .	24
4.11 Visualize the Ridge & Lasso Models . . . . .	25
4.12 Support Vector Machines with Linear Kernel . . . . .	25
4.13 Support Vector Machines Radial Basis Function Kernel . . . . .	25
4.14 Support Vector Tuned Model 1 . . . . .	25
4.15 Support Vector Tuned Model 2 . . . . .	25
<b>5 Result</b>	<b>26</b>
<b>6 Conclusion</b>	<b>27</b>
<b>7 References</b>	<b>27</b>

# 1 Executive Summary

Airbnb, Inc. is an online marketplace for offering lodging,homestays, or tourism experiences. The company does not own any of the real estate listings, nor does it host events; it acts as a broker, receiving commissions from each booking. This service has started in 2008 with 2 hosts and 3 travellers. Its number one concern is assuring safety of its customers while providing enjoyable experiences.

I performed a descriptive and exploratory analysis of the data, in order to understand how the phenomena of each variable behave individually and transversely. To perform this work, I will use statistical techniques very common in any type of analysis, simple or complex, such as classification of variables, frequency distribution tables, histograms, measures of central tendency and etc.

I have also done analysis on predicting the price of Airbnb listings as a suggestion for the hosts.

The models I have applied here are Linear Regression , Partial Least-Squares Regression (PLS), Boosted Generalized Linear Model ,Recursive Partitioning and Regression Trees, Pruned Tree Models, Bagged CART ,Random Forest ,Stochastic Gradient Boosting ,KNN ,Ridge Regression ,Lasso Regression and Support Vector Regression

## 2 Dataset

There are 48895 observations which has data in 16 columns. Both categorical and quantitative data types could be observed. Each row represents details about lodgings in NYC.

**Source** <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

## 3 Analysis

I will perform a descriptive ,geographic and exploratory analysis of the data, in order to understand how the phenomena of each variable behave individually and transversely, in addition to to generate hypotheses useful for future decision-making.

### 3.1 Descriptive Analysis

Descriptive analyzes are the first manipulations performed in a quantitative study and their main objective is to summarize and explore the behavior of the data involved in the study.

#### Dimensions

```
## [1] 48895     16
```

#### Summary

```
##      id          name        host_id       host_name
##  Min. : 2539  Length:48895   Min. : 2438  Length:48895
##  1st Qu.: 9471945 Class :character  1st Qu.: 7822033 Class :character
##  Median :19677284 Mode  :character  Median : 30793816 Mode  :character
##  Mean   :19017143                   Mean   : 67620011
##  3rd Qu.:29152178                   3rd Qu.:107434423
##  Max.   :36487245                   Max.   :274321313
##
##      neighbourhood_group neighbourhood      latitude      longitude
##  Length:48895      Length:48895   Min.   :40.50  Min.   :-74.24
##  Class :character  Class :character  1st Qu.:40.69  1st Qu.:-73.98
```

```

## Mode :character      Mode :character   Median :40.72    Median :-73.96
##                                         Mean  :40.73    Mean  :-73.95
##                                         3rd Qu.:40.76   3rd Qu.:-73.94
##                                         Max.  :40.91    Max.  :-73.71
##
##   room_type          price       minimum_nights   number_of_reviews
## Length:48895      Min.    : 0.0     Min.    : 1.00    Min.    : 0.00
## Class :character   1st Qu.: 69.0    1st Qu.: 1.00    1st Qu.: 1.00
## Mode  :character   Median : 106.0   Median : 3.00    Median : 5.00
##                                         Mean   : 152.7   Mean   : 7.03    Mean   : 23.27
##                                         3rd Qu.: 175.0   3rd Qu.: 5.00    3rd Qu.: 24.00
##                                         Max.   :10000.0   Max.   :1250.00   Max.   :629.00
##
##   last_review        reviews_per_month calculated_host_listings_count
## Length:48895      Min.    : 0.010   Min.    : 1.000
## Class :character   1st Qu.: 0.190   1st Qu.: 1.000
## Mode  :character   Median : 0.720   Median : 1.000
##                                         Mean   : 1.373   Mean   : 7.144
##                                         3rd Qu.: 2.020   3rd Qu.: 2.000
##                                         Max.   :58.500   Max.   :327.000
##                                         NA's    :10052
##
## availability_365
## Min.    : 0.0
## 1st Qu.: 0.0
## Median : 45.0
## Mean   :112.8
## 3rd Qu.:227.0
## Max.   :365.0
##

```

## Unique Neighbourhood Groups

```

## $neighbourhood_group
## [1] "Brooklyn"      "Manhattan"     "Queens"        "Staten Island"
## [5] "Bronx"

```

## Unique Neighbourhoods

```

## $neighbourhood
## [1] "Kensington"           "Midtown"
## [3] "Harlem"                "Clinton Hill"
## [5] "East Harlem"          "Murray Hill"
## [7] "Bedford-Stuyvesant"   "Hell's Kitchen"
## [9] "Upper West Side"       "Chinatown"
## [11] "South Slope"          "West Village"
## [13] "Williamsburg"         "Fort Greene"
## [15] "Chelsea"               "Crown Heights"
## [17] "Park Slope"           "Windsor Terrace"
## [19] "Inwood"                 "East Village"
## [21] "Greenpoint"            "Bushwick"
## [23] "Flatbush"              "Lower East Side"
## [25] "Prospect-Lefferts Gardens" "Long Island City"
## [27] "Kips Bay"                "SoHo"
## [29] "Upper East Side"        "Prospect Heights"
## [31] "Washington Heights"     "Woodside"
## [33] "Brooklyn Heights"       "Carroll Gardens"

```

```

## [35] "Gowanus"
## [37] "Cobble Hill"
## [39] "Boerum Hill"
## [41] "DUMBO"
## [43] "Highbridge"
## [45] "Ridgewood"
## [47] "Jamaica"
## [49] "NoHo"
## [51] "Flatiron District"
## [53] "Greenwich Village"
## [55] "East Flatbush"
## [57] "Astoria"
## [59] "Eastchester"
## [61] "Two Bridges"
## [63] "Rockaway Beach"
## [65] "Nolita"
## [67] "University Heights"
## [69] "Gramercy"
## [71] "East New York"
## [73] "Concourse Village"
## [75] "Emerson Hill"
## [77] "Bensonhurst"
## [79] "Shore Acres"
## [81] "Concourse"
## [83] "Brighton Beach"
## [85] "Cypress Hills"
## [87] "Arrochar"
## [89] "Wakefield"
## [91] "Bay Ridge"
## [93] "Spuyten Duyvil"
## [95] "Briarwood"
## [97] "Columbia St"
## [99] "Mott Haven"
## [101] "Canarsie"
## [103] "Civic Center"
## [105] "New Springville"
## [107] "Arverne"
## [109] "Tottenville"
## [111] "Concord"
## [113] "Bayside"
## [115] "Port Morris"
## [117] "Kew Gardens"
## [119] "College Point"
## [121] "City Island"
## [123] "Port Richmond"
## [125] "Richmond Hill"
## [127] "Maspeth"
## [129] "Soundview"
## [131] "Woodrow"
## [133] "Stuyvesant Town"
## [135] "North Riverdale"
## [137] "Bronxdale"
## [139] "Riverdale"
## [141] "Bay Terrace"
"Flatlands"
"Flushing"
"Sunnyside"
"St. George"
"Financial District"
"MorningSide Heights"
"Middle Village"
"Ditmars Steinway"
"Roosevelt Island"
"Little Italy"
"Tompkinsville"
"Clason Point"
"Kingsbridge"
"Queens Village"
"Forest Hills"
"Woodlawn"
"Gravesend"
"Allerton"
"Theater District"
"Sheepshead Bay"
"Fort Hamilton"
"TriBeCa"
"Sunset Park"
"Elmhurst"
"Jackson Heights"
"St. Albans"
"Rego Park"
"Clifton"
"Graniteville"
"Stapleton"
"Ozone Park"
"Vinegar Hill"
"Longwood"
"Battery Park City"
"East Elmhurst"
"Morris Heights"
"Cambria Heights"
"Mariners Harbor"
"Borough Park"
"Downtown Brooklyn"
"Fieldston"
"Midwood"
"Mount Eden"
"Glendale"
"Red Hook"
"Bellerose"
"Williamsbridge"
"Woodhaven"
"Co-op City"
"Parkchester"
"Dyker Heights"
"Sea Gate"
"Kew Gardens Hills"
"Norwood"

```

```

## [143] "Claremont Village"          "Whitestone"
## [145] "Fordham"                   "Bayswater"
## [147] "Navy Yard"                 "Brownsville"
## [149] "Eltingville"                "Fresh Meadows"
## [151] "Mount Hope"                 "Lighthouse Hill"
## [153] "Springfield Gardens"        "Howard Beach"
## [155] "Belle Harbor"               "Jamaica Estates"
## [157] "Van Nest"                  "Morris Park"
## [159] "West Brighton"              "Far Rockaway"
## [161] "South Ozone Park"           "Tremont"
## [163] "Corona"                     "Great Kills"
## [165] "Manhattan Beach"            "Marble Hill"
## [167] "Dongan Hills"              "Castleton Corners"
## [169] "East Morrisania"            "Hunts Point"
## [171] "Neponsit"                   "Pelham Bay"
## [173] "Randall Manor"              "Throgs Neck"
## [175] "Todt Hill"                  "West Farms"
## [177] "Silver Lake"                "Morrisania"
## [179] "Laurelton"                  "Grymes Hill"
## [181] "Holliswood"                 "Pelham Gardens"
## [183] "Belmont"                    "Rosedale"
## [185] "Edgemere"                   "New Brighton"
## [187] "Midland Beach"              "Baychester"
## [189] "Melrose"                    "Bergen Beach"
## [191] "Richmondtown"                "Howland Hook"
## [193] "Schuylerville"               "Coney Island"
## [195] "New Dorp Beach"              "Prince's Bay"
## [197] "South Beach"                 "Bath Beach"
## [199] "Jamaica Hills"              "Oakwood"
## [201] "Castle Hill"                "Hollis"
## [203] "Douglasston"                "Huguenot"
## [205] "Olinville"                  "Edenwald"
## [207] "Grant City"                 "Westerleigh"
## [209] "Bay Terrace, Staten Island" "Westchester Square"
## [211] "Little Neck"                 "Fort Wadsworth"
## [213] "Rosebank"                   "Unionport"
## [215] "Mill Basin"                  "Arden Heights"
## [217] "Bull's Head"                 "New Dorp"
## [219] "Rossville"                   "Breezy Point"
## [221] "Willowbrook"

```

### Unique Room types

```

## $room_type
## [1] "Private room"      "Entire home/apt" "Shared room"

```

### Range of Prices

```

## Minimum Price: 0 | Maximum Price: 10000

```

### Range of Longitude

```

## Minimum Longitude : -74.24442 | Maximum Longitude: -73.71299

```

### Range of Latitude

```

## Minimum Latitude : 40.49979 | Maximum Latitude: 40.91306

```

The first step in an analysis work is the knowledge of the behavior of the variables involved in the study. Using statistical techniques such as frequency distribution tables, histograms and bar graphs we can better understand how the phenomena under study are distributed.

#### **Neighbourhood\_group/Location Freq table**

	Frequency	Percent
Staten Island	373	0.7628592
Bronx	1091	2.2313120
Queens	5666	11.5880969
Brooklyn	20104	41.1166786
Manhattan	21661	44.3010533

#### **Neighbourhood/Area Freq table (Displaying only top 10 rows)**

	Frequency	Percent
Fort Wadsworth	1	0.0020452
New Dorp	1	0.0020452
Richmondtown	1	0.0020452
Rossville	1	0.0020452
Willowbrook	1	0.0020452
Woodrow	1	0.0020452
Bay Terrace, Staten Island	2	0.0040904
Co-op City	2	0.0040904
Howland Hook	2	0.0040904
Lighthouse Hill	2	0.0040904

#### **Room Type Freq table**

	Frequency	Percent
Shared room	1160	2.372431
Private room	22326	45.661110
Entire home/apt	25409	51.966459

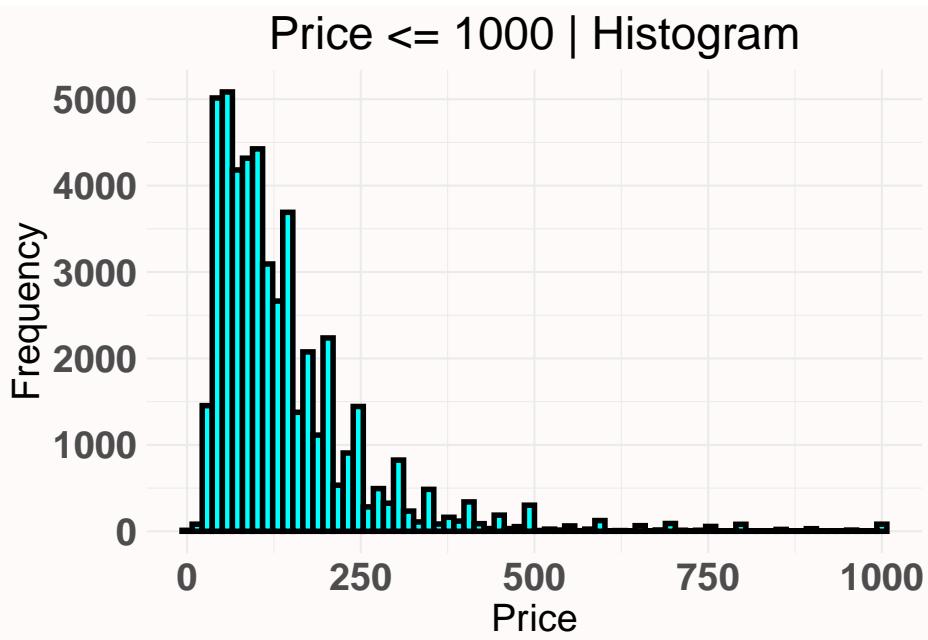
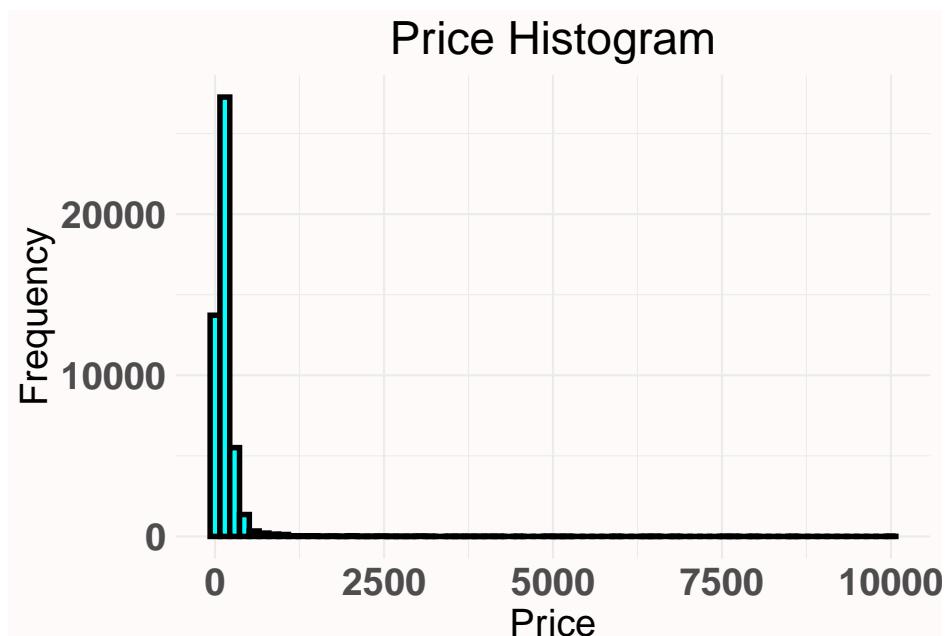
With the frequency tables created above, we can conclude that the frequency and the representative percentage of the most frequent categories of the categorical variables neighborhood\_group, neighborhood and room\_type are-

neighbourhood\_group - Manhattan -> 21661(44.30%)

neighborhood - Williamsburg -> 3920(8.01%)

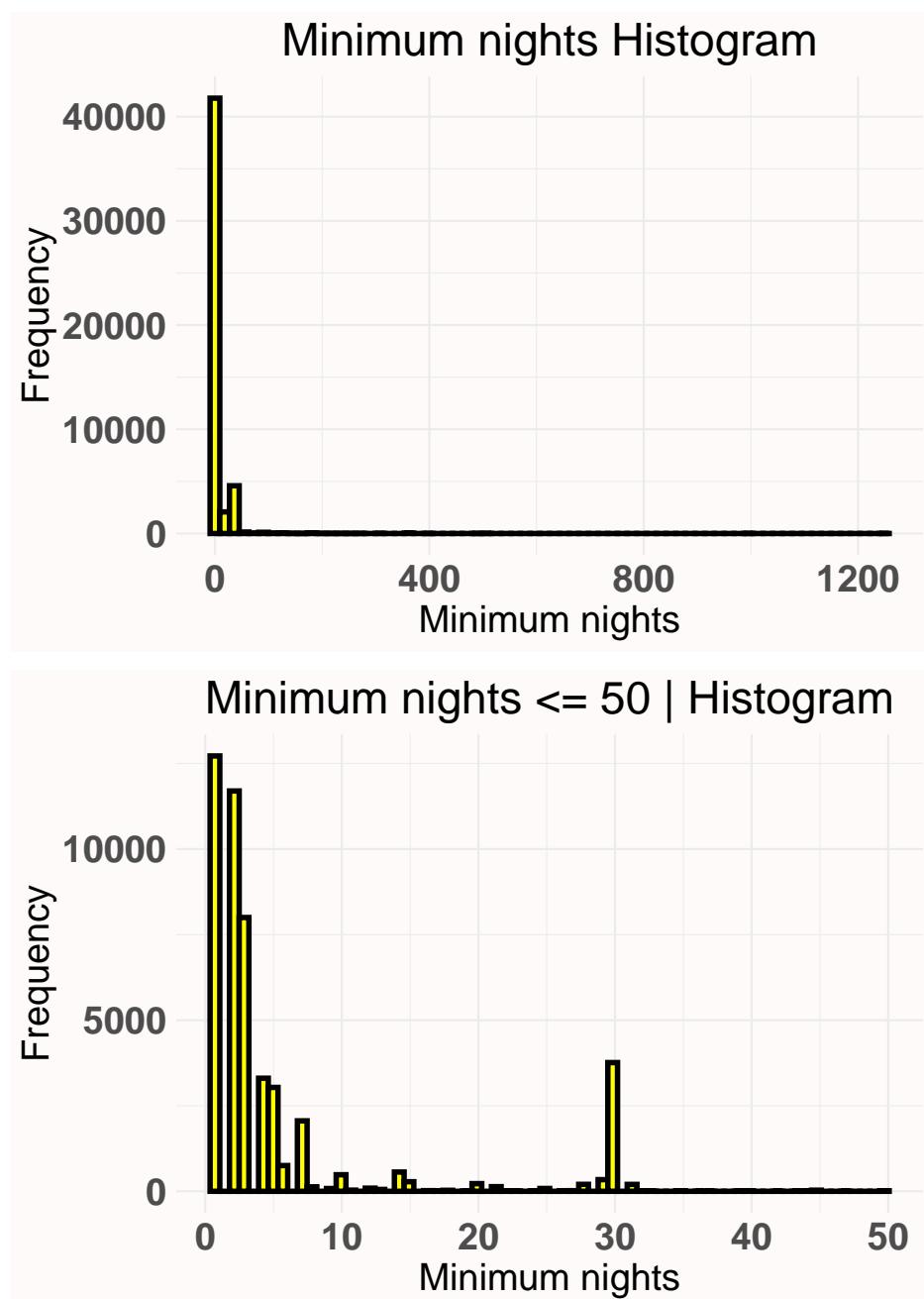
room\_type - Entire home/apt -> 25409(51.96%)

Histogram for Price



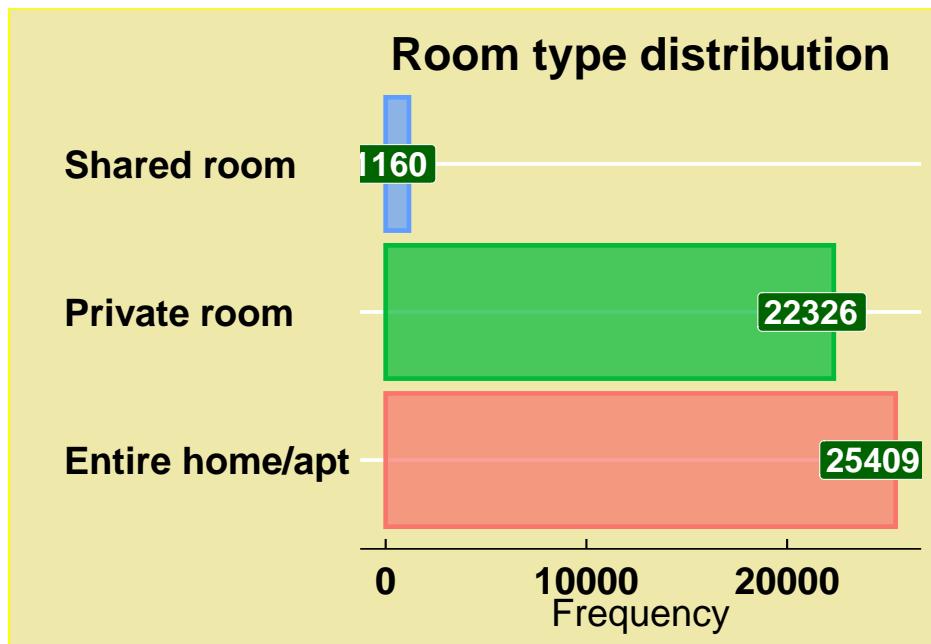
We can see most of prices our under 1000

### Histogram For Minimum\_Nights



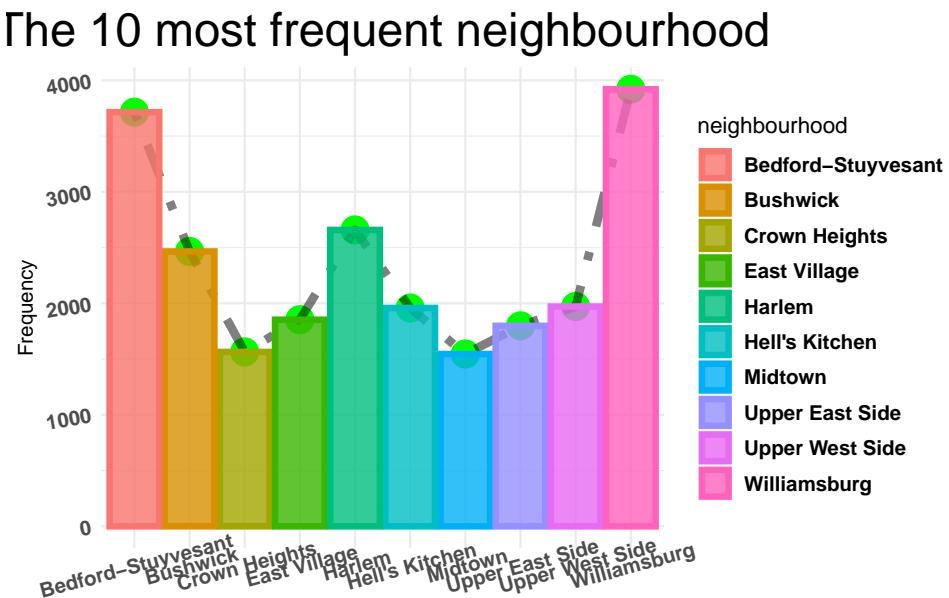
We can see that the minimum number of nights for all reservations made on Airbnb are concentrated below 10 with a small peak at 30.

Bar Graph for Room Type



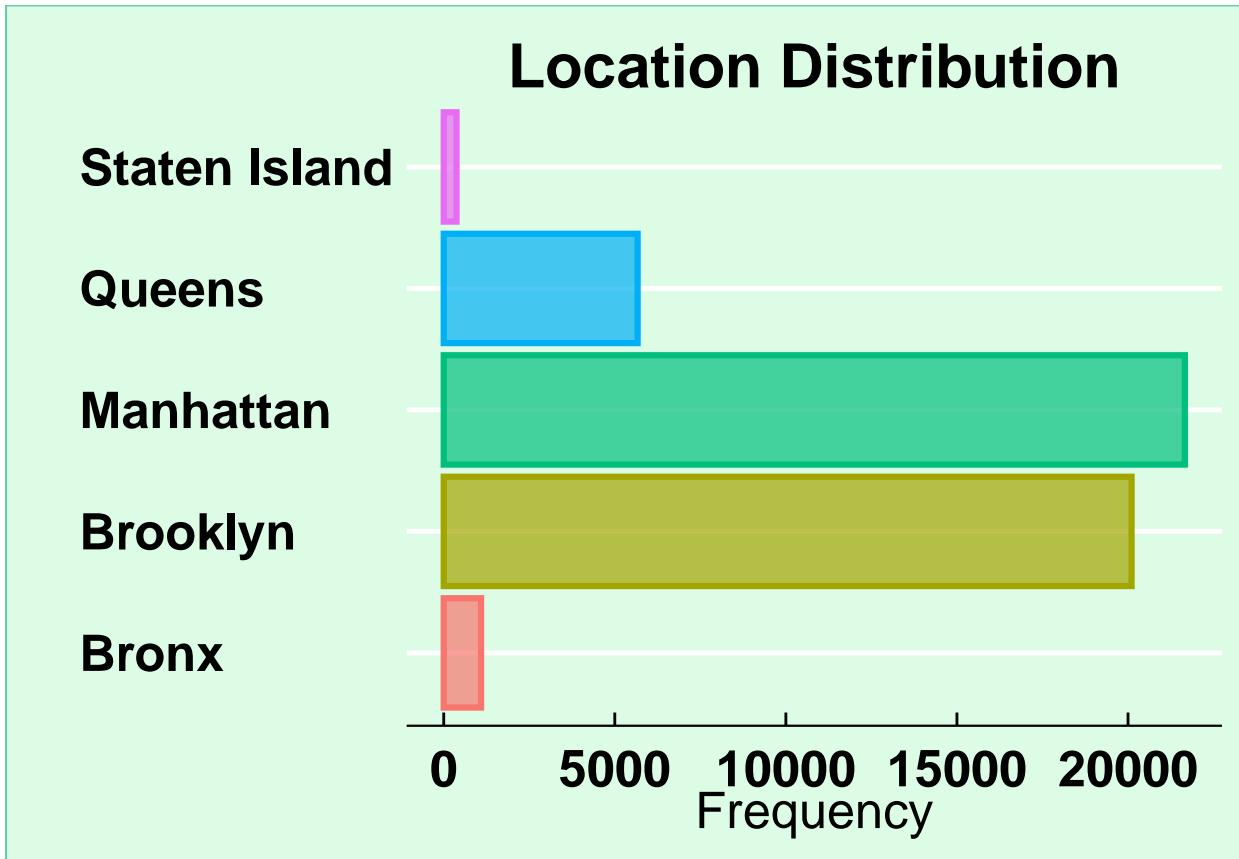
We can see most of the rooms belong private or entire apt/home category.

Bar Graph of The 10 most frequent neighbourhood



Of all 221 neighbourhoods ,these are the top 10 neighborhoods, which are most requested by customers for advertisements and accommodation reservations on the airbnb website.

Bar Graph for neighbourhood\_group



Manhattan is the most famous neighbouring group followed by Brooklyn, Queens, Bronx and Staten Island

#### Separating Measures

```
##      Price minimum_nights
## 1%      30              1
## 2%      35              1
## 3%      36              1
## 4%      39              1
## 5%      40              1
## 6%      42              1
## 7%      45              1
## 8%      45              1
## 9%      47              1
## 10%     49              1
## 11%     50              1
## 12%     50              1
## 13%     50              1
## 14%     53              1
## 15%     55              1
## 16%     55              1
## 17%     59              1
## 18%     60              1
## 19%     60              1
## 20%     60              1
```

## 21%	63	1
## 22%	65	1
## 23%	65	1
## 24%	67	1
## 25%	69	1
## 26%	70	1
## 27%	70	2
## 28%	72	2
## 29%	75	2
## 30%	75	2
## 31%	75	2
## 32%	79	2
## 33%	80	2
## 34%	80	2
## 35%	81	2
## 36%	85	2
## 37%	85	2
## 38%	89	2
## 39%	90	2
## 40%	90	2
## 41%	93	2
## 42%	95	2
## 43%	98	2
## 44%	99	2
## 45%	100	2
## 46%	100	2
## 47%	100	2
## 48%	100	2
## 49%	101	2
## 50%	106	3
## 51%	110	3
## 52%	110	3
## 53%	115	3
## 54%	119	3
## 55%	120	3
## 56%	120	3
## 57%	125	3
## 58%	125	3
## 59%	128	3
## 60%	130	3
## 61%	134	3
## 62%	137	3
## 63%	140	3
## 64%	145	3
## 65%	149	3
## 66%	150	3
## 67%	150	4
## 68%	150	4
## 69%	150	4
## 70%	155	4
## 71%	160	4
## 72%	165	4
## 73%	170	4
## 74%	175	5

```

## 75%    175      5
## 76%    180      5
## 77%    185      5
## 78%    190      5
## 79%    197      5
## 80%    200      6
## 81%    200      7
## 82%    200      7
## 83%    205      7
## 84%    220      7
## 85%    225      7
## 86%    233      10
## 87%    249      14
## 88%    250      15
## 89%    250      20
## 90%    269      28
## 91%    285      30
## 92%    300      30
## 93%    300      30
## 94%    340      30
## 95%    355      30
## 96%    400      30
## 97%    450      30
## 98%    550      30
## 99%    799      45

```

Some important points that you can describe in relation to the type of analysis are-

- 1) 25.00% of bookings made on airbnb are of values equal to or less than 69 dollars and 1 minimum night.
- 2) 50.00% of bookings made on airbnb are of values equal to or less than 106 dollars and 3 minimum nights.
- 3) 75.00% of bookings made on airbnb are of values equal to or less than 175 dollars and 5 minimum nights.
- 4) 99.00% of bookings made on airbnb are of values equal to or less than 799 dollars and 45 minimum nights, meaning only 1.0% of bookings are of values above 799 dollars with a maximum value of 10000 dollars and 1250 minimum nights.

## 3.2 Exploratory Data Analysis

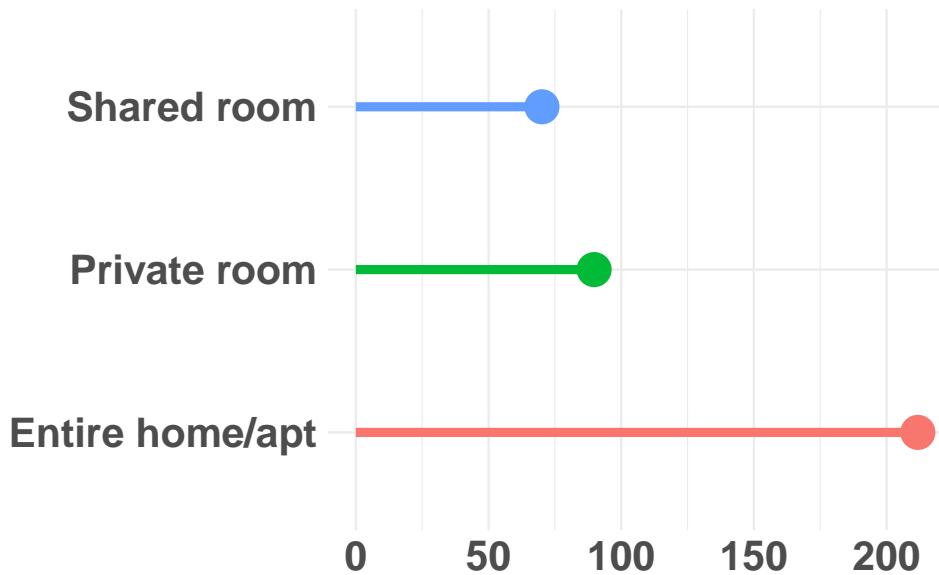
In statistics, exploratory data analysis is an approach to the analysis of data sets in order to summarize their main characteristics, often with visual methods. **Average price per room type**

```

##          room_type average_price Percent
## 1 Entire home/apt     211.79425 56.97946
## 2 Private room        89.78097 24.15397
## 3 Shared room         70.12759 18.86657

```

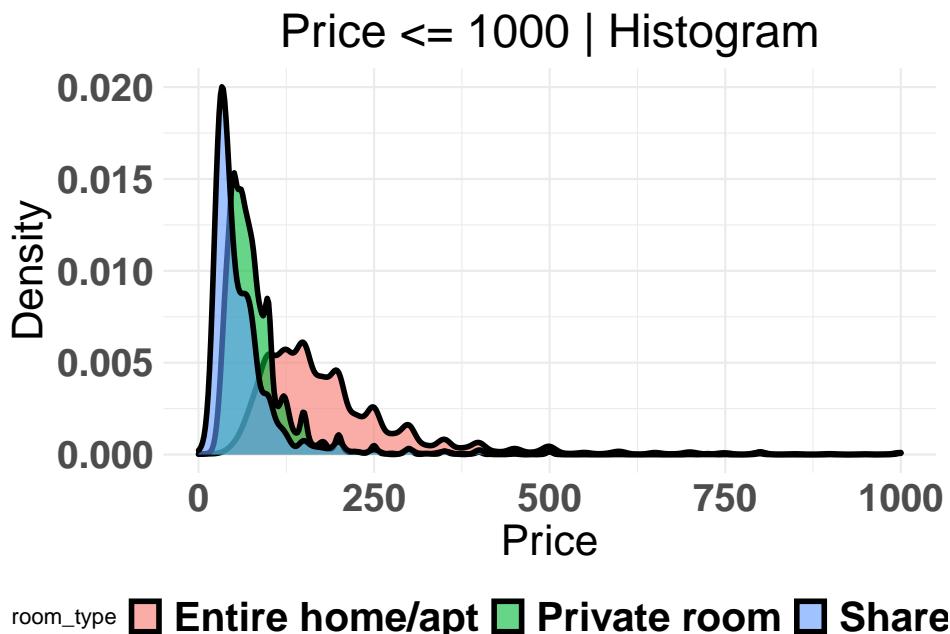
## Average price per room type



We can conclude that-

- 1) The Entire home / apt type has an average price for reservations around 211.79 dollars , which represents 56.97 % of all types of rooms . We have the Entire home / apt has an average price of 32.82% more expensive than the Private room and 38.11% more expensive than the Shared room .
- 2) The Private room which has an average booking price of around 89.78 dollar, which represents 24.15% of all types of rooms . We have that the Private room has an average price 32.82% less than Entire home / apt and 5.29% larger than the Shared room.
- 3) The Shared room which has an average booking price of around 70.12 dollars , which represents 18.86% of all types of rooms . We have that the Shared room has an average price 38.1% less than Entire home / apt and 5.29% smaller than the Private room.

Price behavior in relation to room types



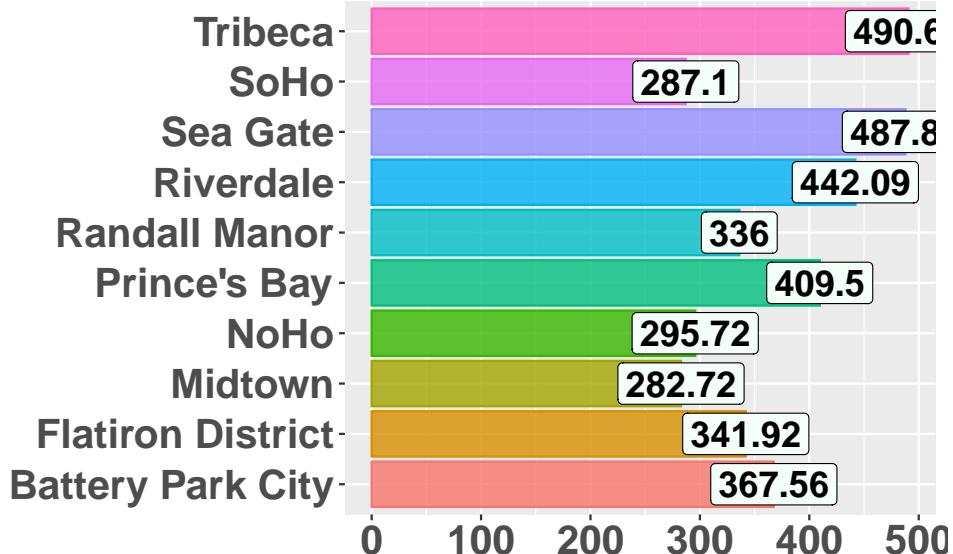
In addition to obtaining information such as the average price for reservations, it is interesting to know how these values that resulted in the average are distributed.

### Price in Relation to Neighborhood

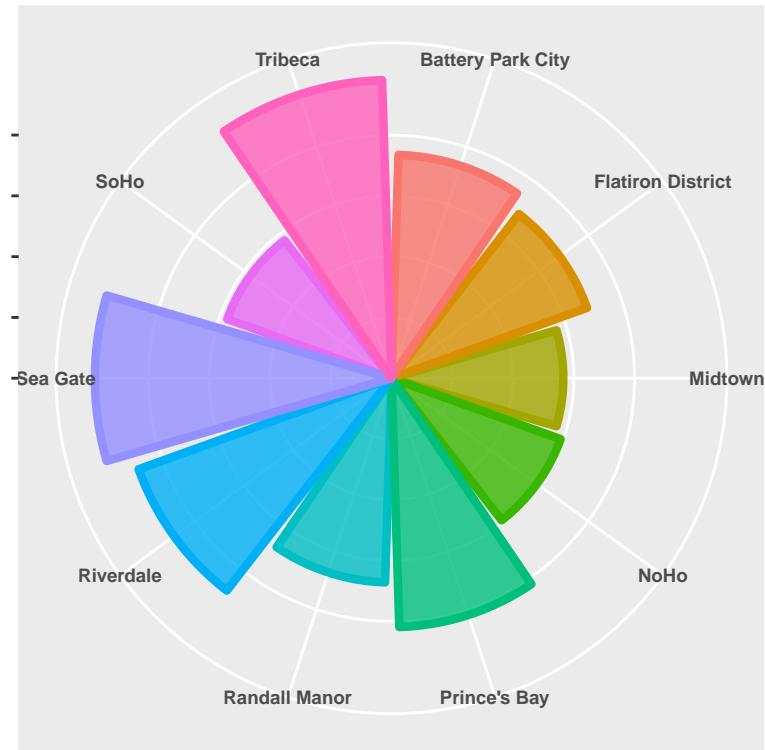
#### The 10 most expensive neighborhoods to book on airbnb

##	neighbourhood	Average_price_per_neighborhood
## 10	Midtown	282.7191
## 9	SoHo	287.1034
## 8	NoHo	295.7179
## 7	Randall Manor	336.0000
## 6	Flatiron District	341.9250
## 5	Battery Park City	367.5571
## 4	Prince's Bay	409.5000
## 3	Riverdale	442.0909
## 2	Sea Gate	487.8571
## 1	Tribeca	490.6384

The 10 most expensive neighborhoods to book on airbnb



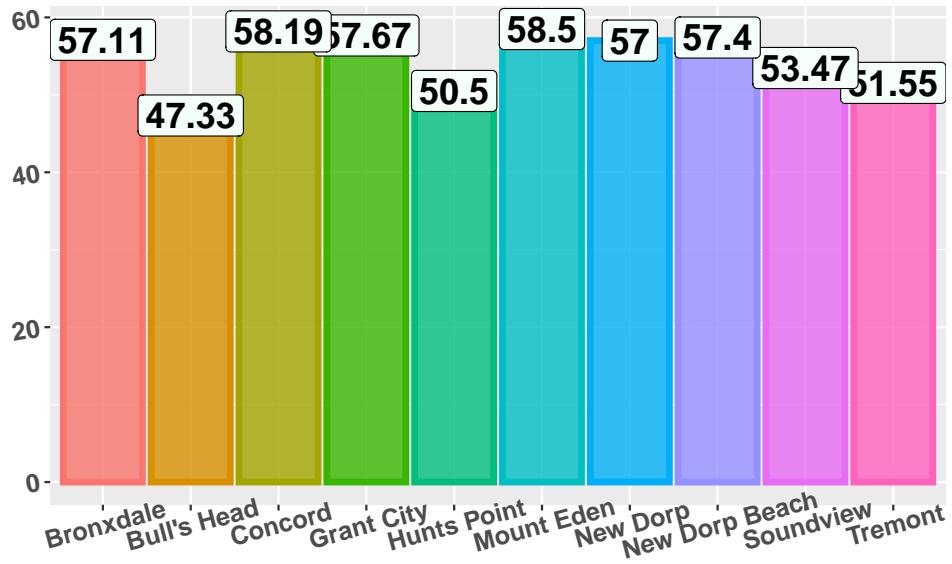
## The 10 most expensive neighborhoods



## The 10 cheapest neighborhoods to book on airbnb

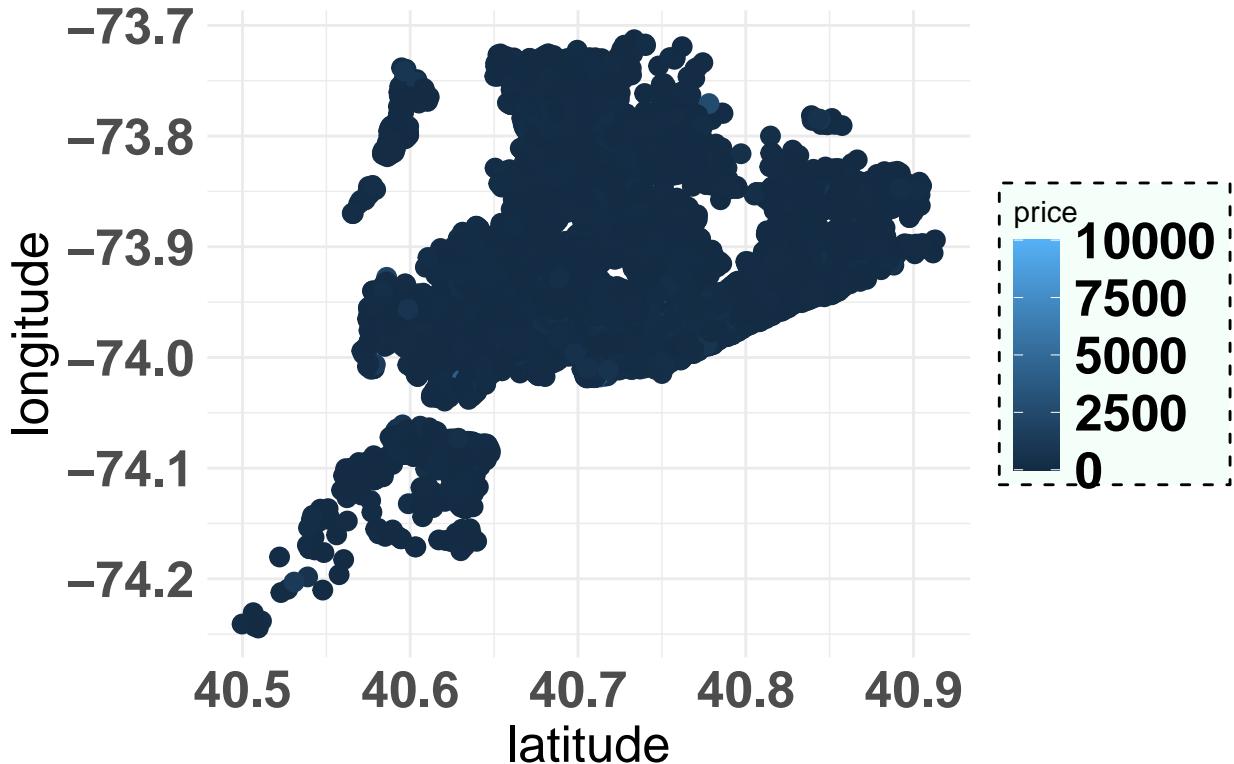
##	neighbourhood	Average_price_per_neighborhood
## 1	Bull's Head	47.33333
## 2	Hunts Point	50.50000
## 3	Tremont	51.54545
## 4	Soundview	53.46667
## 5	New Dorp	57.00000
## 6	Bronxdale	57.10526
## 7	New Dorp Beach	57.40000
## 8	Grant City	57.66667
## 9	Concord	58.19231
## 10	Mount Eden	58.50000

### The 10 cheapest neighborhoods

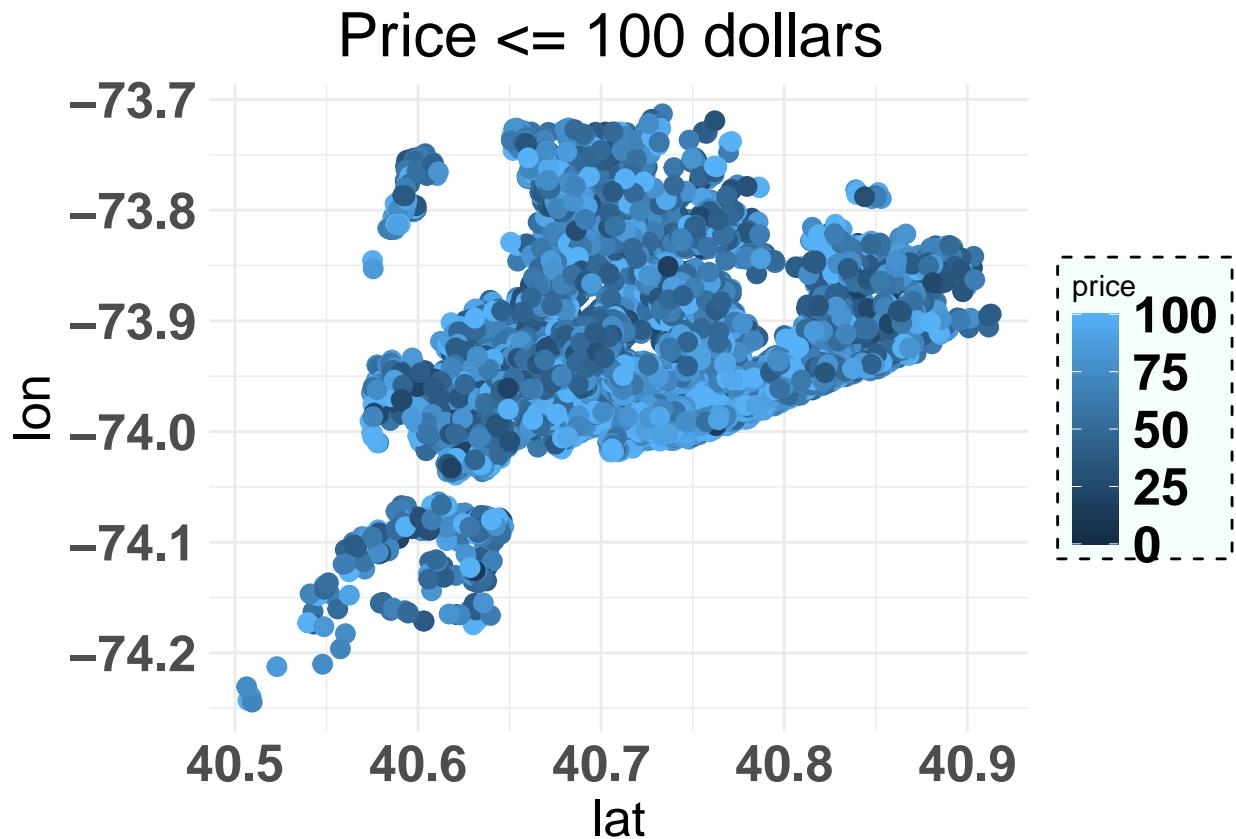


### 3.3 Geographic analysis

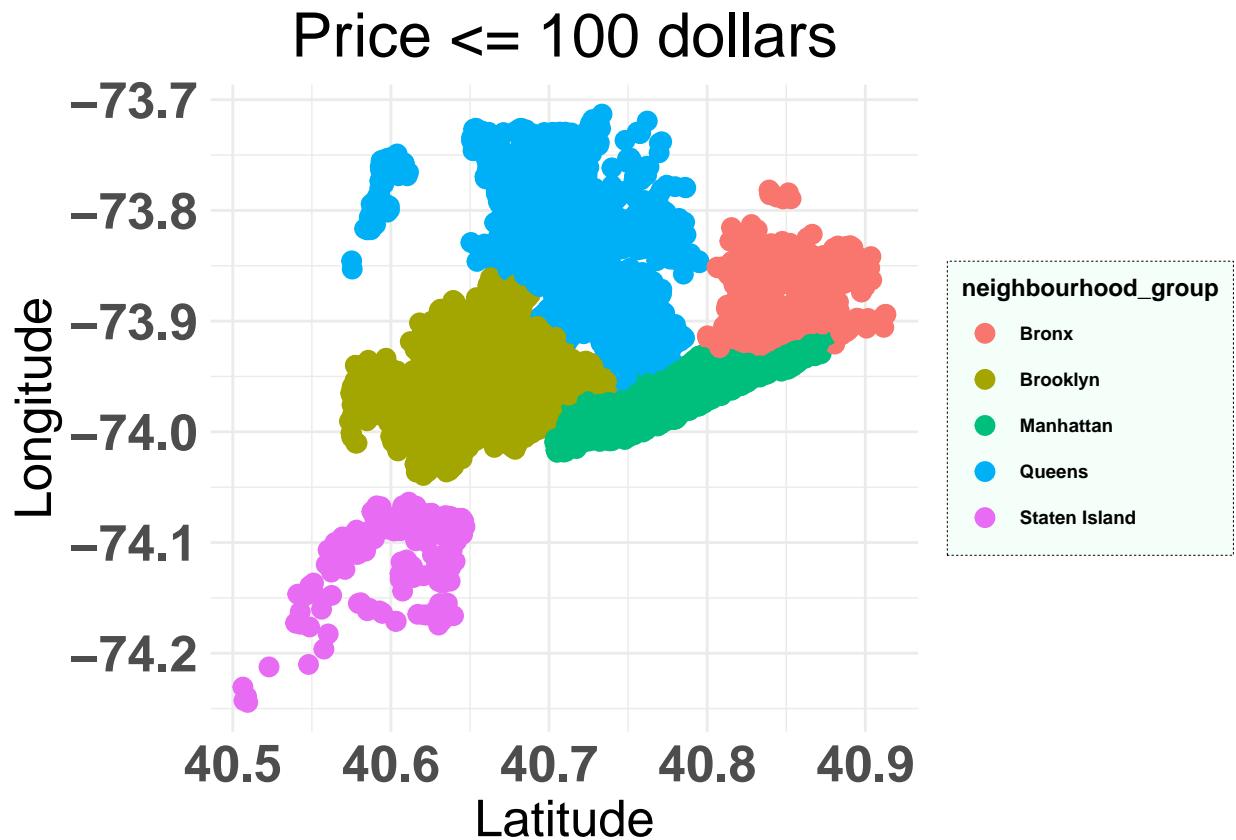
In this analysis I will explore the behavior of the price, neighborhood\_group, minimum\_nights and room\_type through the coordinated latitude and longitude available in the airbnb data. This type of exploitation is extremely useful for understanding the behavior of the data on a geographic scale, consequently helping in decision making.



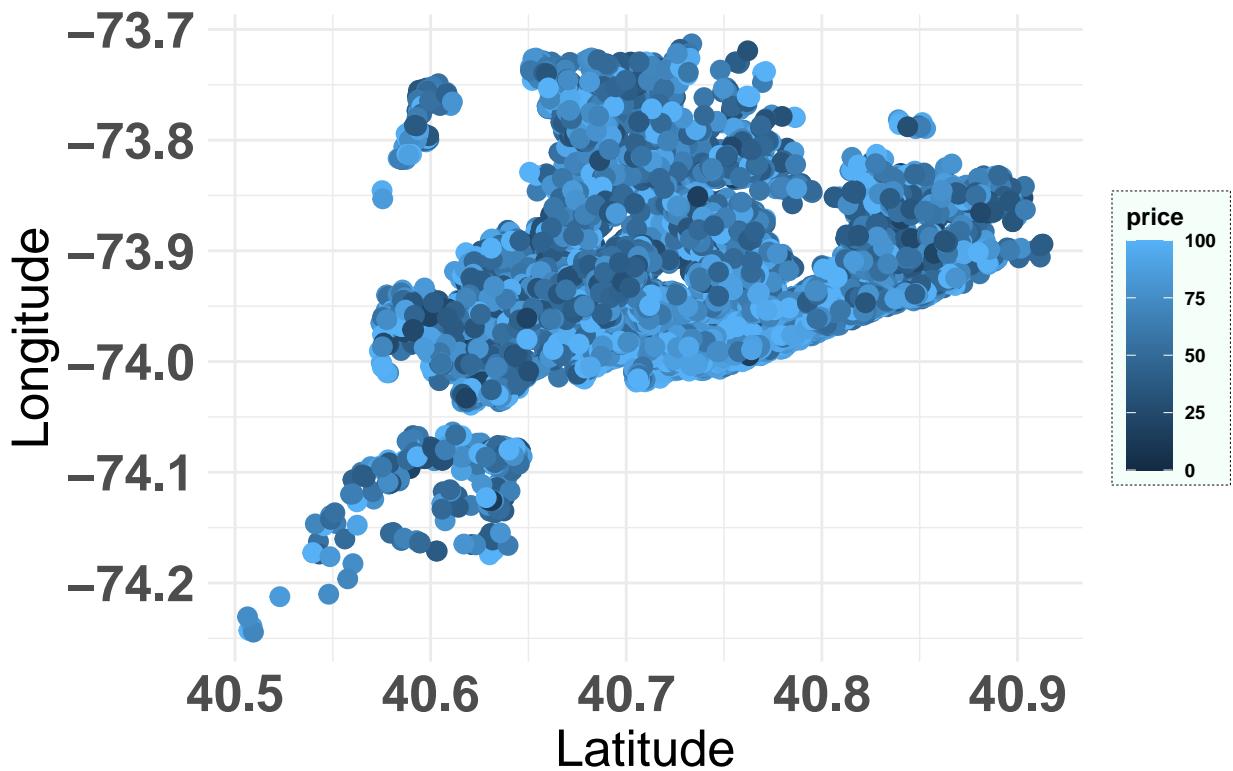
As our database prices are mostly below 100 dollars, we will filter the data to obtain only bookings below 100 dollars.



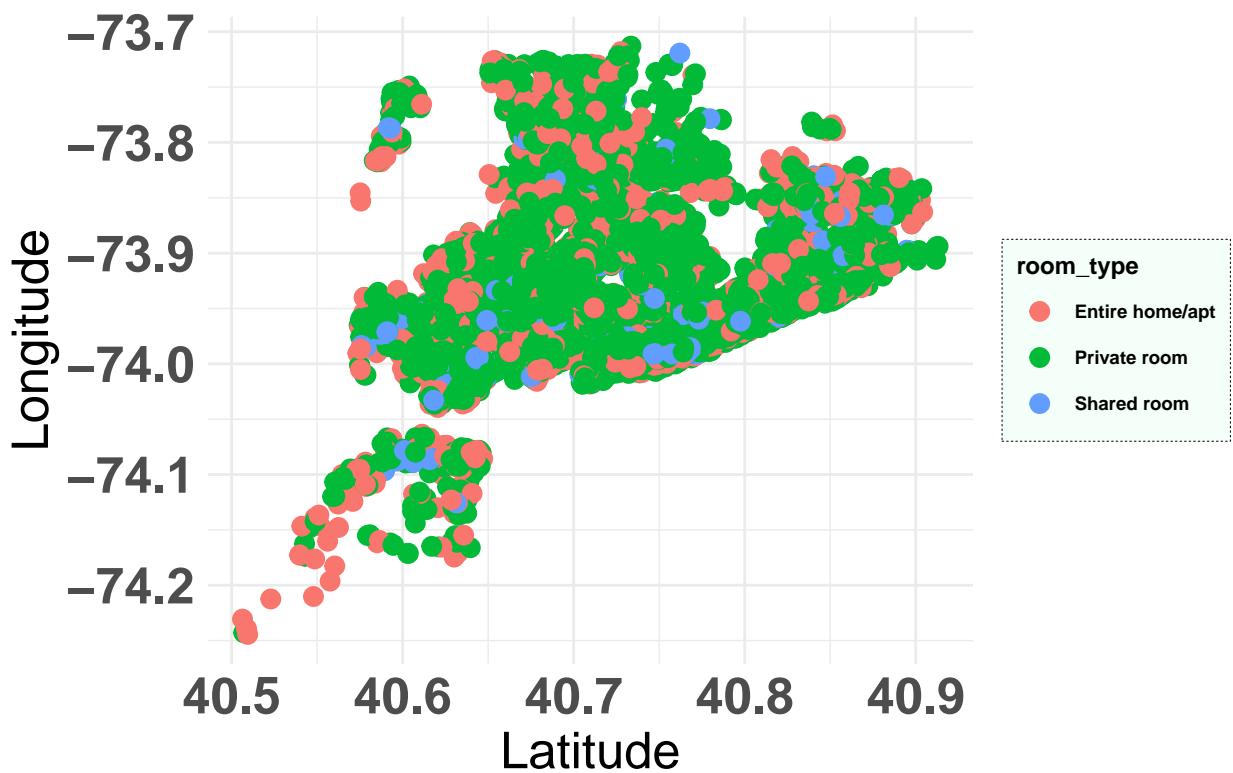
Understanding how this grouping behaves in relation to the coordinate.

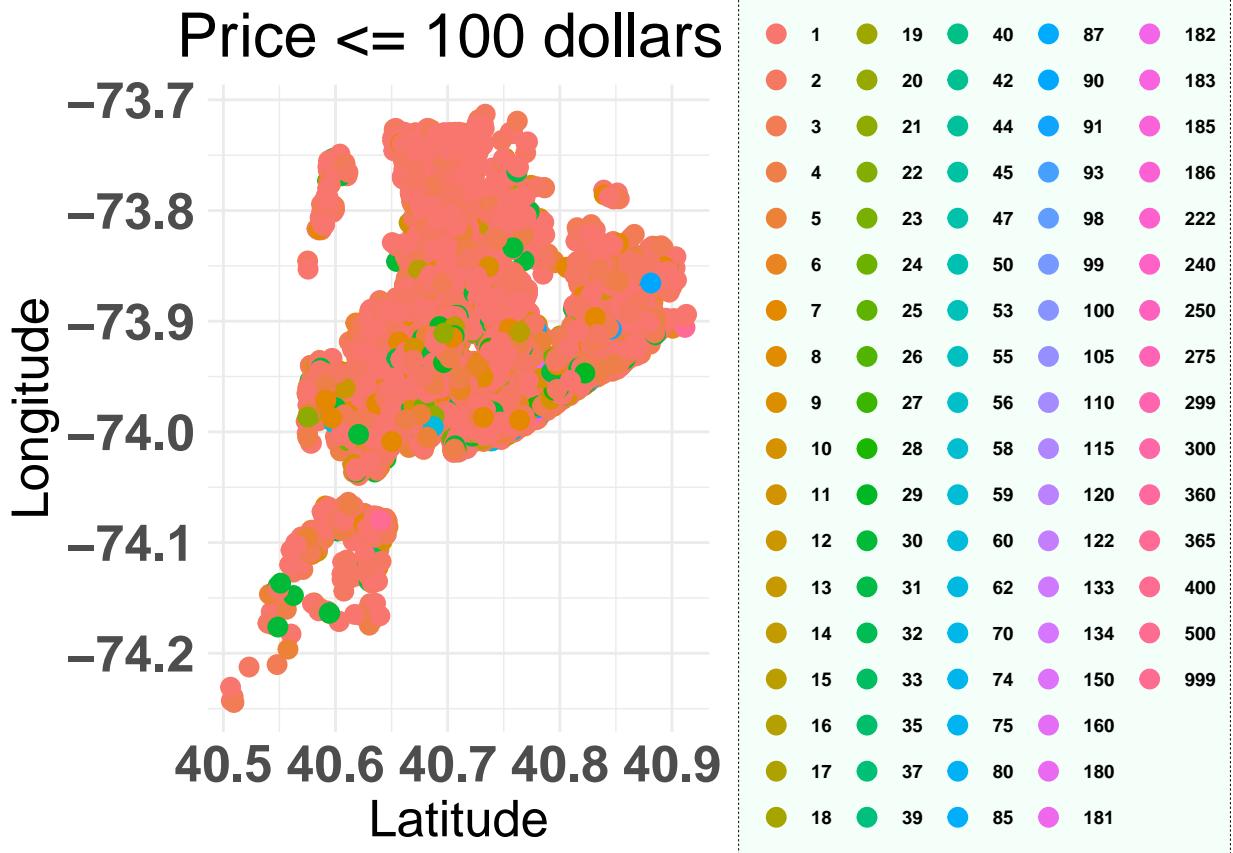


Price  $\leq$  100 dollars



Price  $\leq$  100 dollars





## 4 Modeling

In this i have tried different methods to predict the price of Airbnb listings.

The models I have applied here are Linear Regression , Partial Least-Squares Regression (PLS), Boosted Generalized Linear Model ,Recursive Partitioning and Regression Trees, Pruned Tree Models, Bagged CART ,Random Forest ,Stochastic Gradient Boosting ,KNN ,Ridge Regression ,Lasso Regression and Support Vector Regression. ## Data prepartion

### Removing Unwanted Columns

Some of the variables have been removed from the dataset due to several reasons.

- 1) id,name,host\_id : We know that logically, we don't look at these variables when we select a place to stay during vacation.
- 2) neighbourhood : This variable consisted of a large number of factor levels and this complicates the models and also some neighbourhoods have very less information and eventually will result in misleading results.
- 3) last\_review : This was a date variable and we have extracted sufficient information from this variable and these information is recorded in the variable year\_cat.

### Converting string to Factors

```
## 'data.frame': 48895 obs. of 2 variables:
## $ neighbourhood_group: Factor w/ 5 levels "Bronx","Brooklyn",...: 2 3 3 2 3 3 2 3 3 3 ...
## $ room_type          : Factor w/ 3 levels "Entire home/apt",...: 2 1 2 1 1 2 2 2 1 ...
```

```

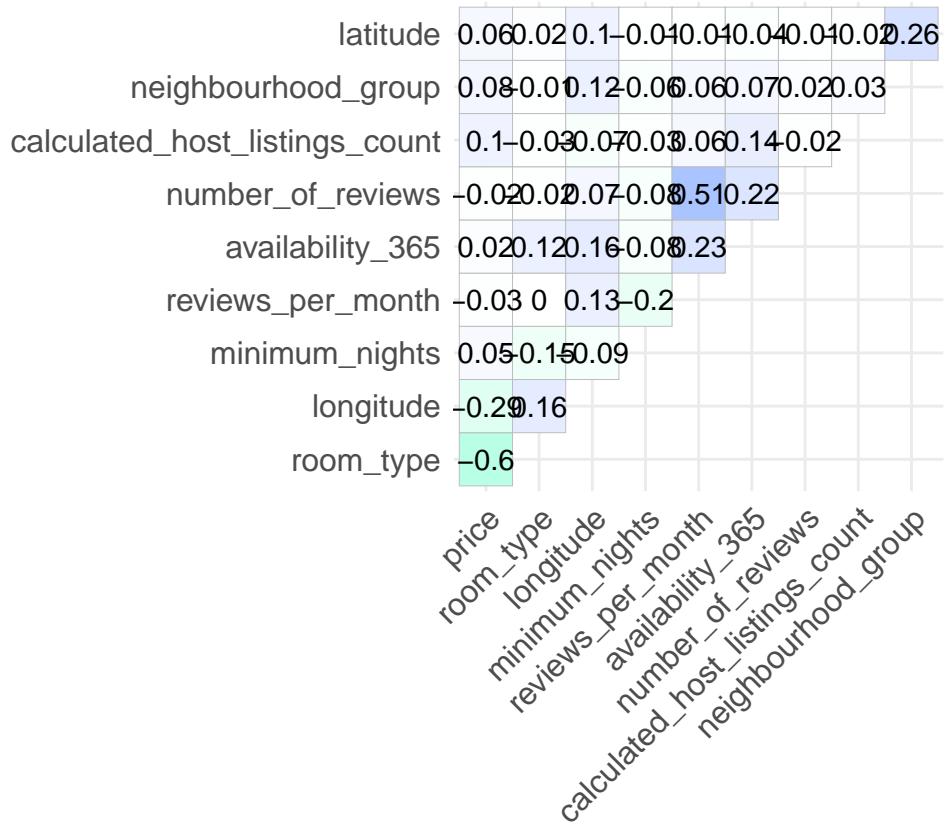
##   price neighbourhood_group latitude longitude      room_type minimum_nights
## 1    149           Brooklyn 40.64749 -73.97237 Private room                 1
## 2    225          Manhattan 40.75362 -73.98377 Entire home/apt                1
## 3    150          Manhattan 40.80902 -73.94190 Private room                 3
## 4     89           Brooklyn 40.68514 -73.95976 Entire home/apt                1
## 5     80          Manhattan 40.79851 -73.94399 Entire home/apt                10
##   number_of_reviews reviews_per_month calculated_host_listings_count
## 1                  9             0.21                               6
## 2                 45             0.38                               2
## 3                  0             0.00                               1
## 4                270             4.64                               1
## 5                  9             0.10                               1
##   availability_365
## 1            365
## 2            355
## 3            365
## 4            194
## 5              0

```

Other Alterations-.

- 1)All the missing values of reviews\_per\_month were replaced by 0.
- 2)Removing Outliers for Price,Minimum nights and number of reviews.

### Finding Correlations



```

##               neighbourhood_group      latitude

```

```

##          1.093851          1.083564
##      longitude          room_type
##          1.087609          1.065210
## minimum_nights      number_of_reviews
##          1.073376          1.378366
## reviews_per_month calculated_host_listings_count
##          1.434202          1.039998
## availability_365
##          1.141387

```

**Splitting into training and test sets**

Training Set=75%

Testing Set=25%

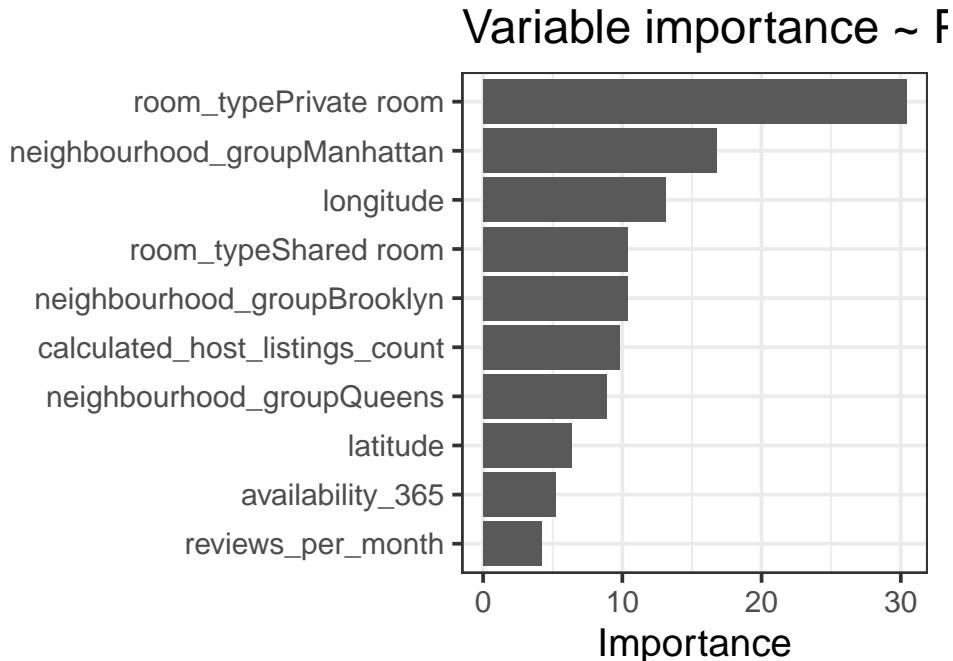
**Trying Different Methods to Predict the price of Airbnb listings.**

#### 4.0.1 Linear Regression - Math Model

Mean Absolute Error

```
## [1] 69.481
```

### 4.1 Partial Least-Squares Regression (PLS)



Mean Absolute Error

```
## [1] 65.298
```

### 4.2 Boosted Generalized Linear Model

Mean Absolute Error

```
## [1] 58.023
```

### 4.3 Pruned Tree Models

Mean Absolute Error Before Pruning

```
## [1] 65.34
```

Mean Absolute Error After Pruning

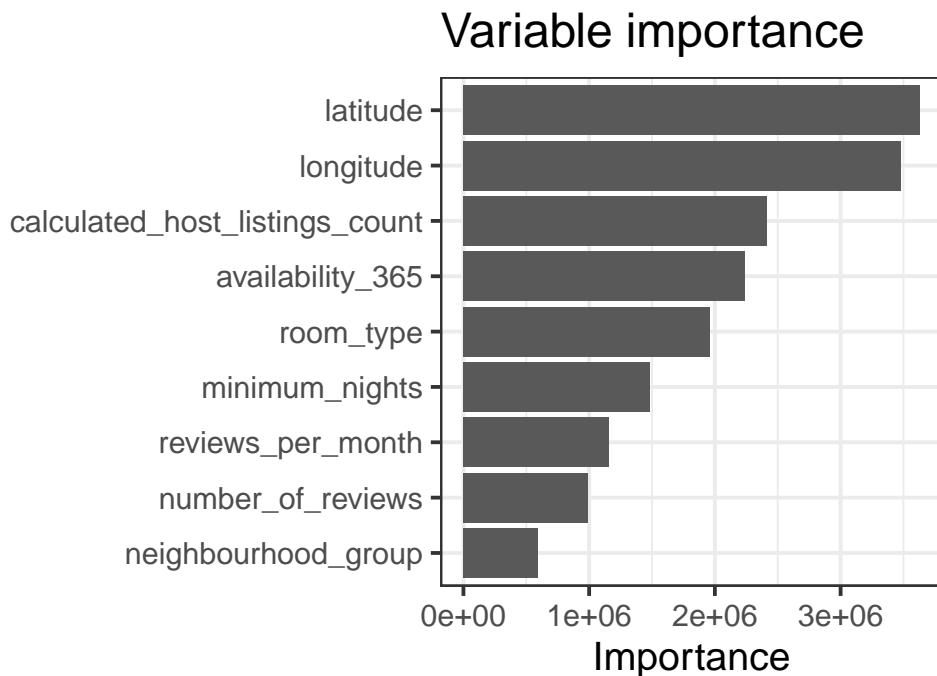
```
## [1] 56.691
```

### 4.4 Bagged CART

Mean Absolute Error

```
## [1] 59.445
```

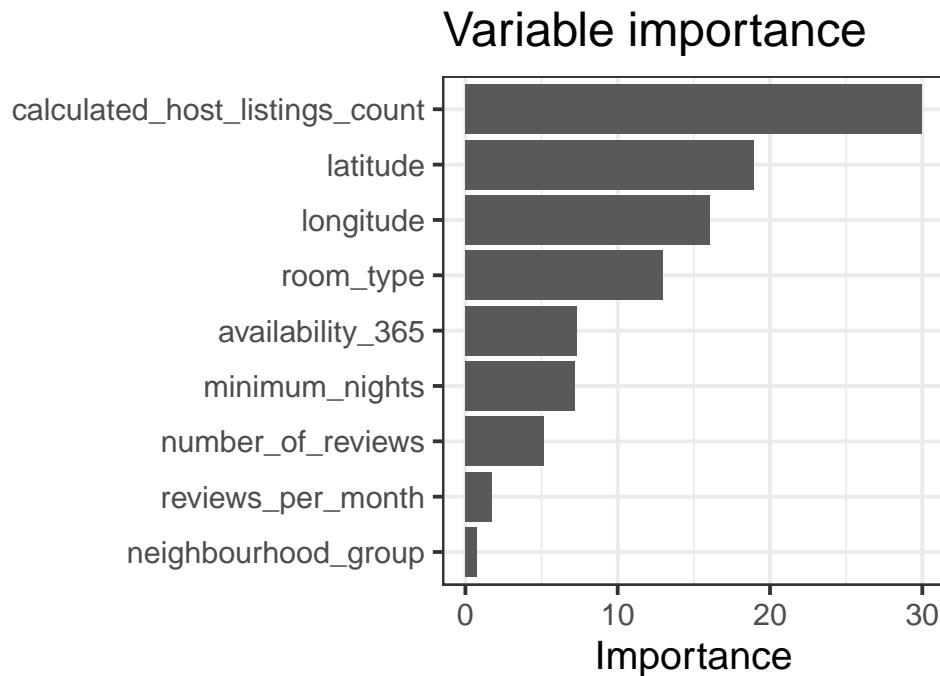
### 4.5 Random Forest



Mean Absolute Error

```
## [1] 45.086
```

## 4.6 Stochastic Gradient Boosting



Mean Absolute Error

```
## [1] 57.603
```

## 4.7 K-Nearest Neighbour

Mean Absolute Error

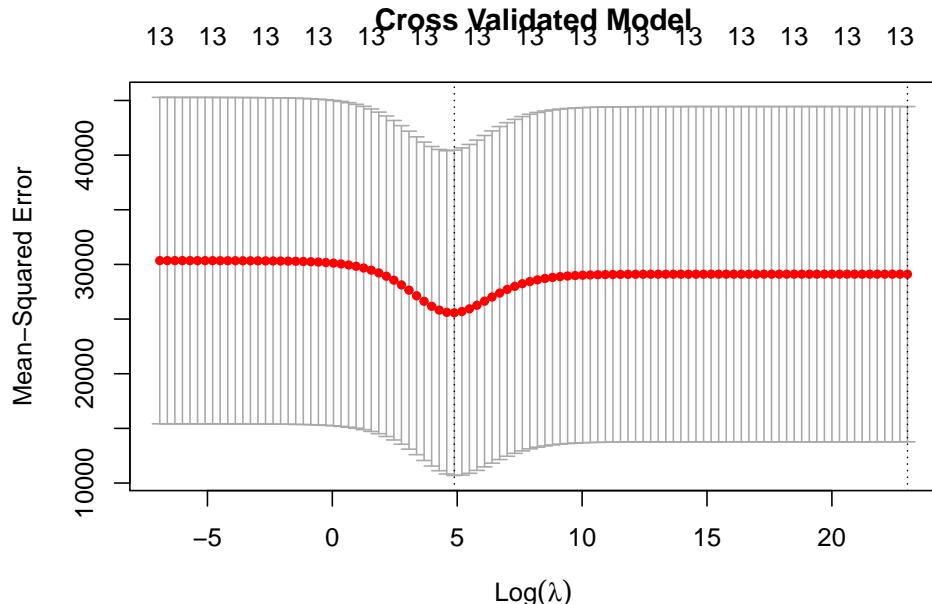
```
## [1] 56.341
```

## 4.8 Bagged Multivariate Adaptive Regression Spline

Mean Absolute Error

```
## [1] 66.044
```

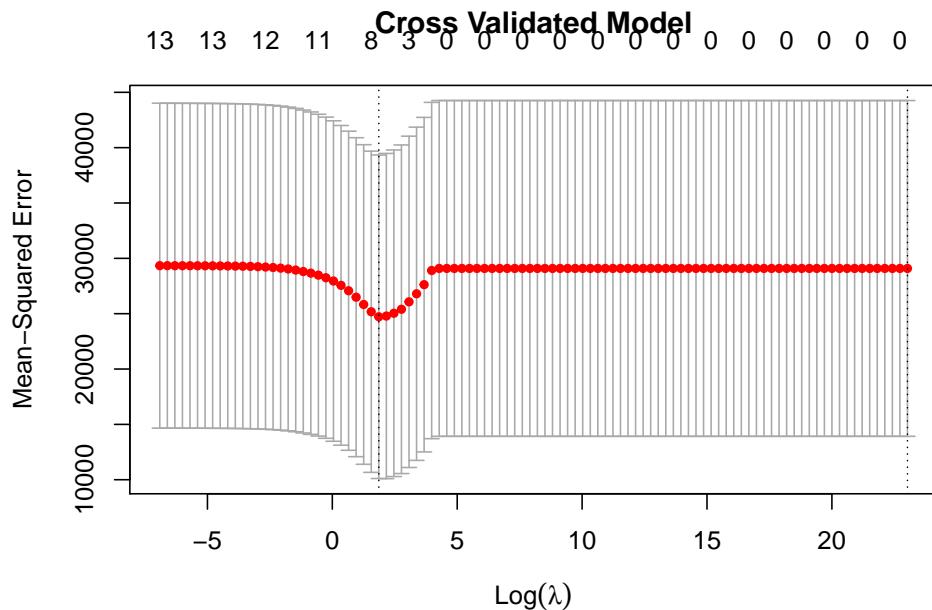
## 4.9 Ridge Regression



Mean Absolute Error

```
## [1] 69.576
```

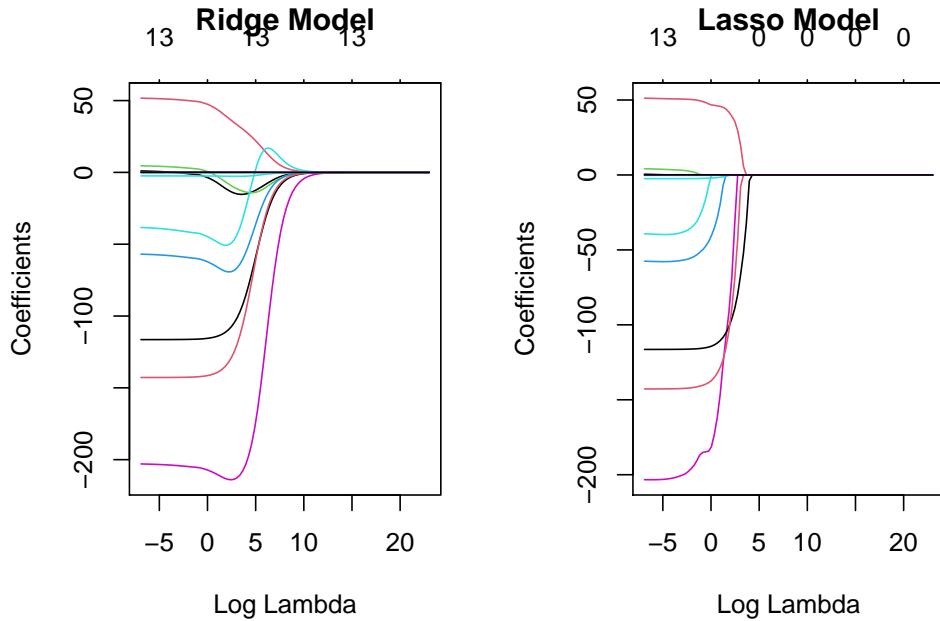
## 4.10 Lasso Regression



Mean Absolute Error

```
## [1] 66.562
```

## 4.11 Visualize the Ridge & Lasso Models



## 4.12 Support Vector Machines with Linear Kernel

Mean Absolute Error

```
## [1] 58.685
```

## 4.13 Support Vector Machines Radial Basis Function Kernel

Mean Absolute Error

```
## [1] 53.413
```

## 4.14 Support Vector Tuned Model 1

Mean Absolute Error

```
## [1] 50.889
```

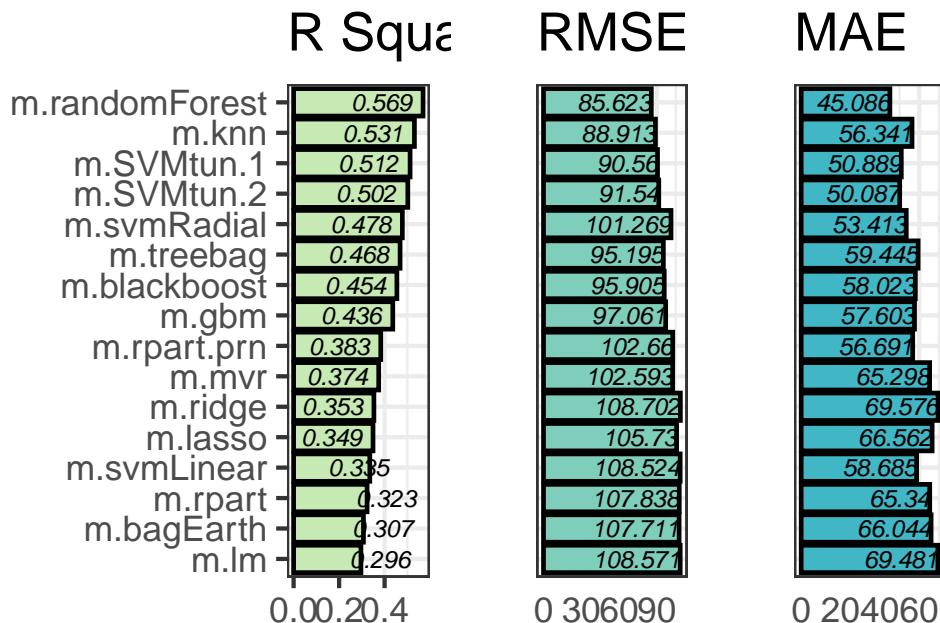
## 4.15 Support Vector Tuned Model 2

Mean Absolute Error

```
## [1] 50.087
```

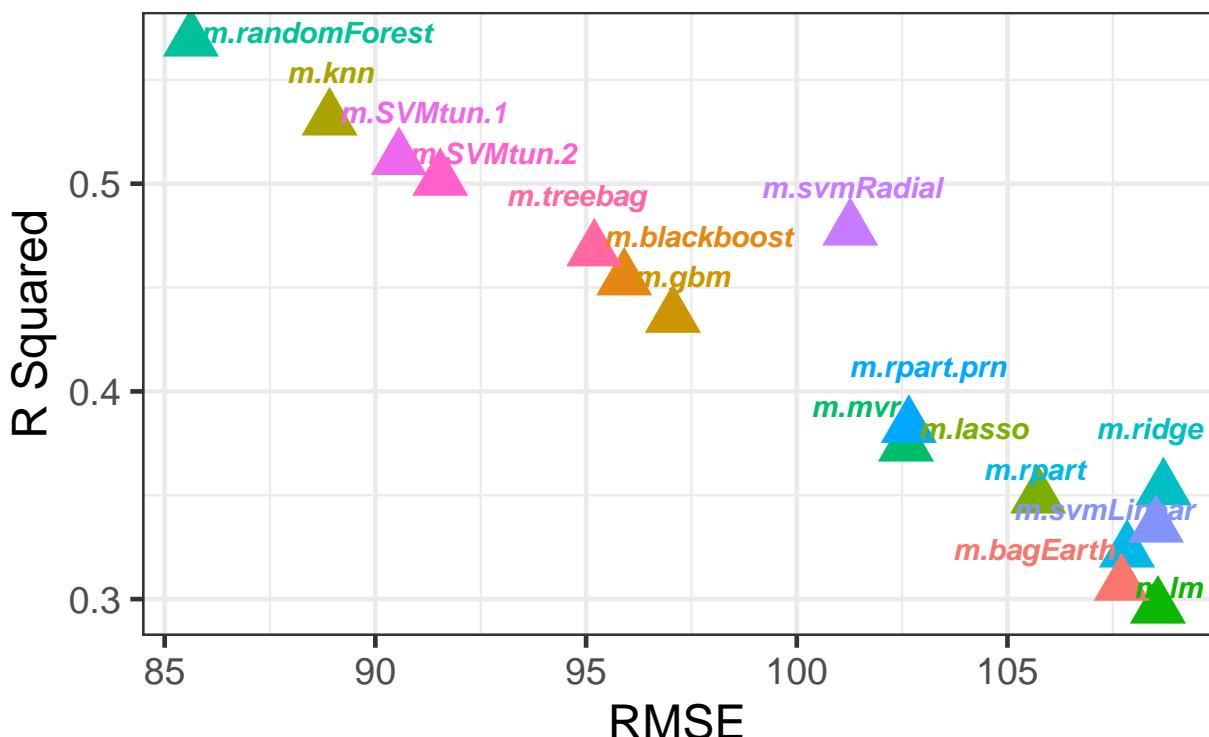
## 5 Result

### Final Model Comparisons



Scatter plot of Model Performance

### Model Performance Comparison



By comparing the values obtained above, it is clear that the best model is Random Forest.

## Model Performance

MAE	RMSE	R.Squared
45.086	85.623	0.569

## 6 Conclusion

The objective of this project was to use machine learning models to predict the Airbnb Prices and understand what features were important to explain the transit. In this report, we tried many models, created some new features, we selected Random Forest model because it has the best performance. The final model obtained an RMSE of 85.623 and MAE of 45.086 . The most important feature to explain price is neighbourhood group .

## 7 References

<https://towardsdatascience.com/predicting-airbnb-prices-with-machine-learning-and-deep-learning-f46d44afb8a6>

<https://medium.com/airbnb-engineering/categorizing-listing-photos-at-airbnb-f9483f3ab7e3>

<https://www.kaggle.com/chamodiperera/price-suggestion-for-airbnb-hosts-nyc-i>

[http://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/Airbnb\\_Pricing\\_TeamR\\_MASTER.HTML](http://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/Airbnb_Pricing_TeamR_MASTER.HTML)