

Capstone Project

Airbnb booking



What is Airbnb?

An Airbnb is a residential property that hosts rent on a short term basis to travelers. It can be anything from a house, a single room, a boat or even a tree house. Think of it as pop-up accommodation – a market place where people rent out their properties.

With the above valuation and the staggering amount of money that Airbnb is making, you would think they have a huge property empire. But no, the company make all their money off property that WE own!



Airbnb Dataset

After deep diving into the data report we came to know about several things which helped Airbnb to succeed in the business.

There is 11452 hosts in the business working with their own strategy to expand their business with such a good profit.

The dataset which we have has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

Part-I Analysis

We started reading the data by analysing what the data set contains.

What type of informations are there in the dataset.

Data Exploration and variable Identification.

And then we got that there are some incomplete data or null data.

Single host containing multiple hotels at different locations.

Price varies according to locations and stay duration.

Some hotels have name repeated.

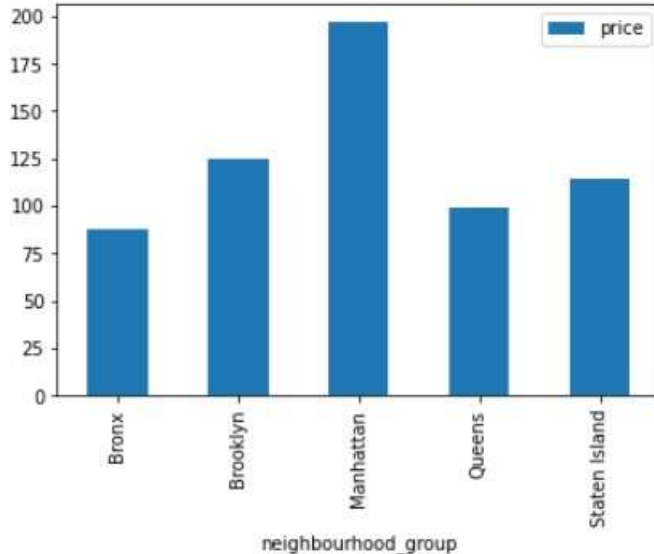
221 loctions are available for airbnb.

Same property with different room types & varied price ranges.

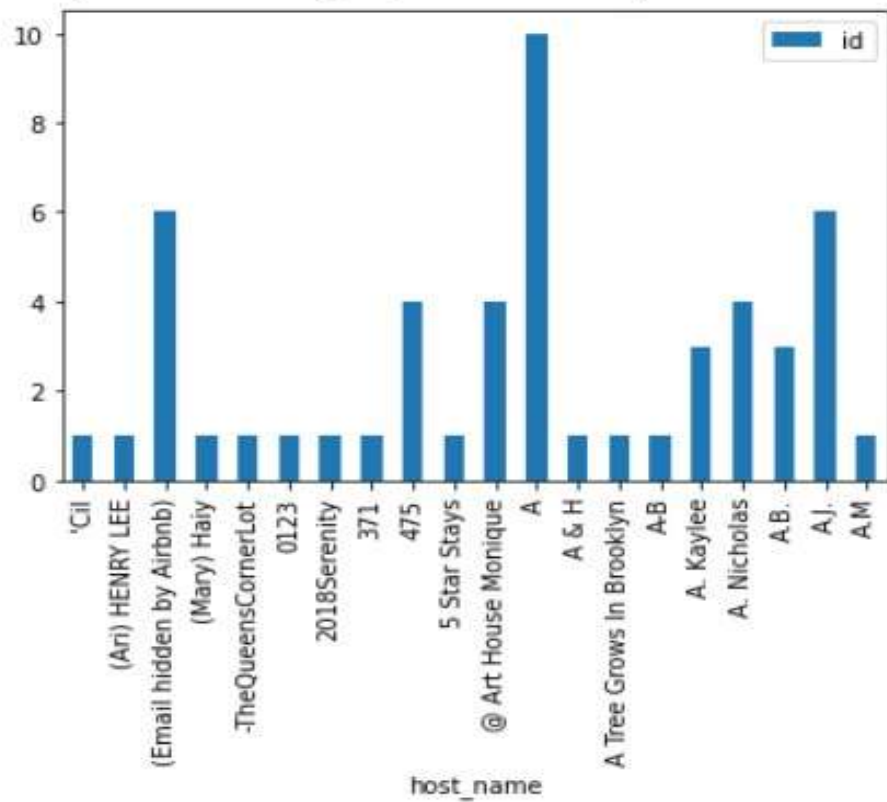
Part-II Analysis

- We did the analysis for most costly Airbnb and found that Manhattan is the most costly Airbnb by plotting the bar graph.

<matplotlib.axes._subplots.AxesSubplot at 0x7f252a2479d0>



<matplotlib.axes._subplots.AxesSubplot at 0x7fb2bd138850>



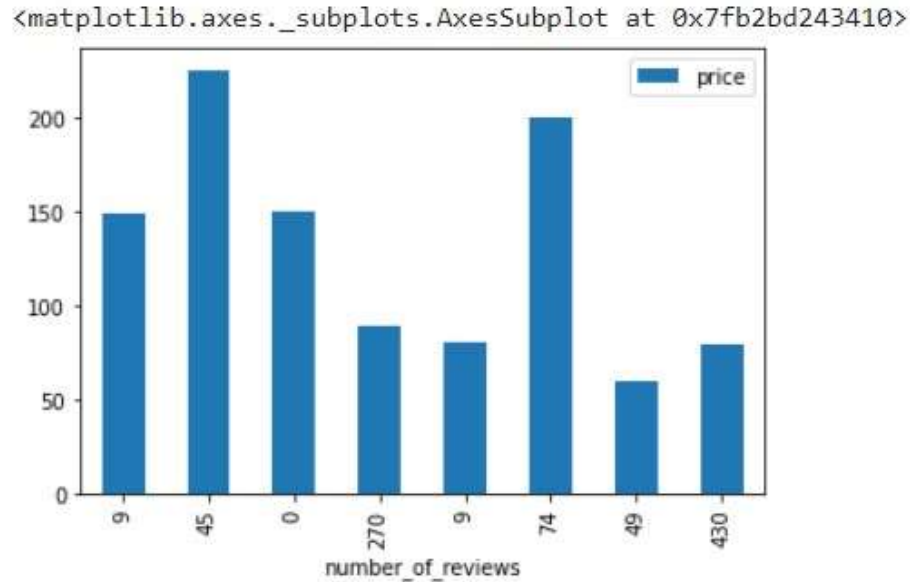
Part-III Analysis

- We did the analysis to know the airbnb reviews.

	neighbourhood_group	number_of_reviews
0	Bronx	1091
1	Brooklyn	20104
2	Manhattan	21661
3	Queens	5666
4	Staten Island	373

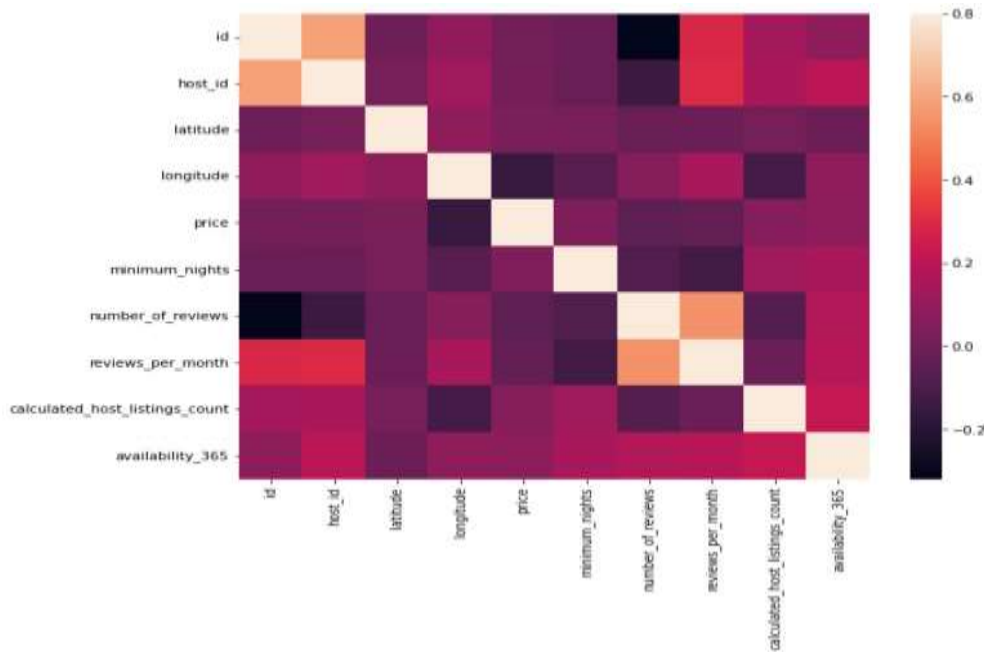
Part-IV Analysis

- We did analysis to know which price category has maximum no. of reviews.



Part-V Analysis

- We perform different coding to find, Which hosts are the busiest and why by making some corelations and representing them into graph.

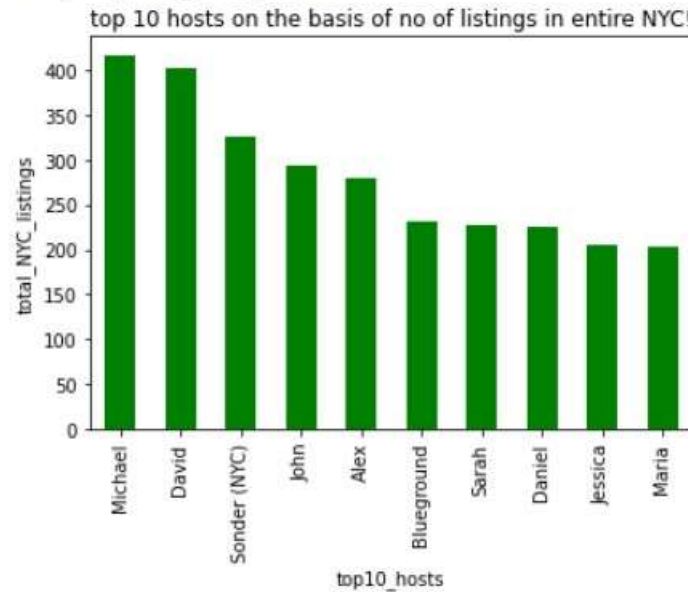


Analysis continued

	Listings on Airbnb	Total_listings
0	Hillside Hotel	18
1	Home away from home	17
2	New york Multi-unit building	16
3	Brooklyn Apartment	12
4	Loft Suite @ The Box House Hotel	11

Analysis continued

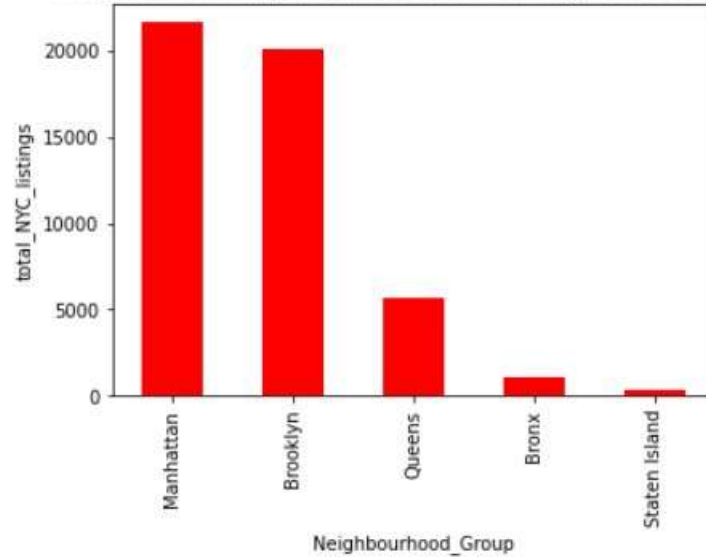
Text(0.5, 1.0, 'top 10 hosts on the basis of no of listings in entire NYC!')



Analysis continued

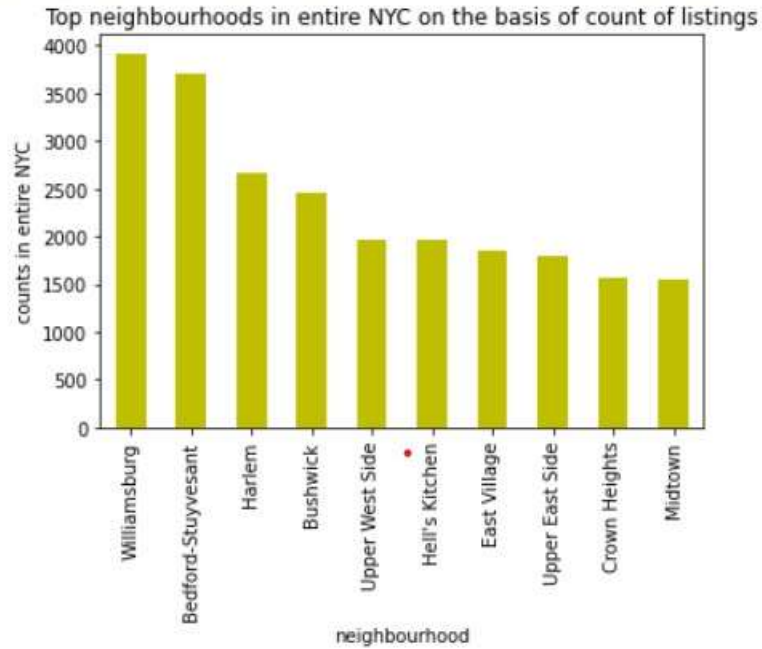
Text(0.5, 1.0, 'Count of no of listings in entire NYC of each neighbourhood group!')

Count of no of listings in entire NYC of each neighbourhood group!

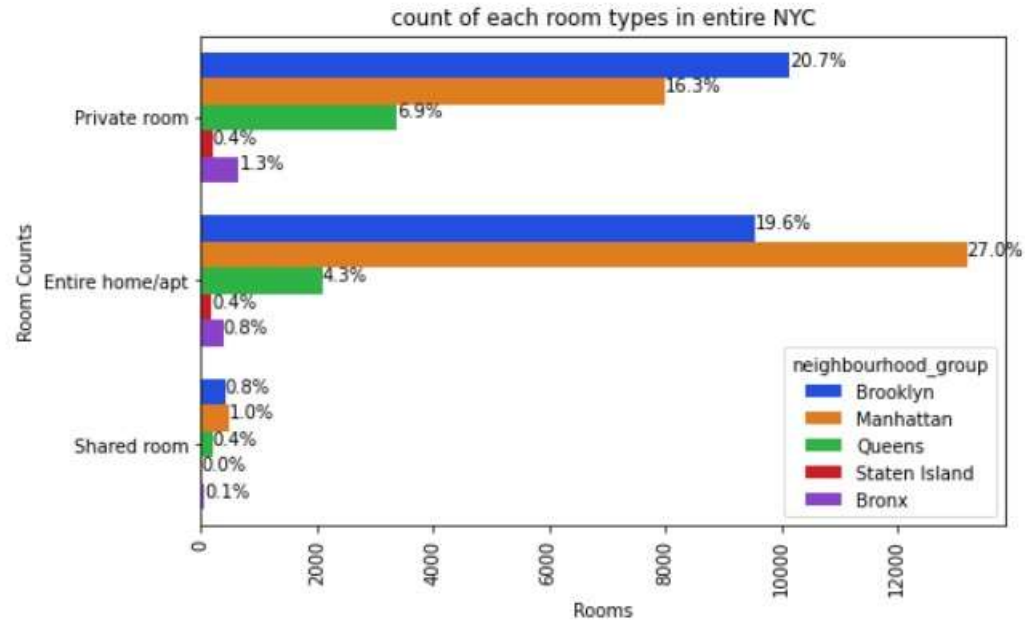


Analysis continued

Text(0.5, 1.0, 'Top neighbourhoods in entire NYC on the basis of count of listings')

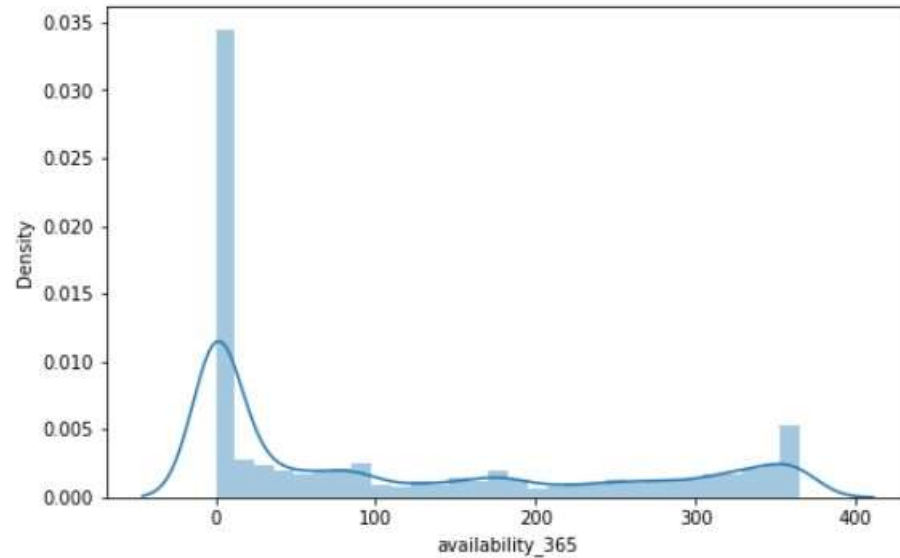


Analysis continued

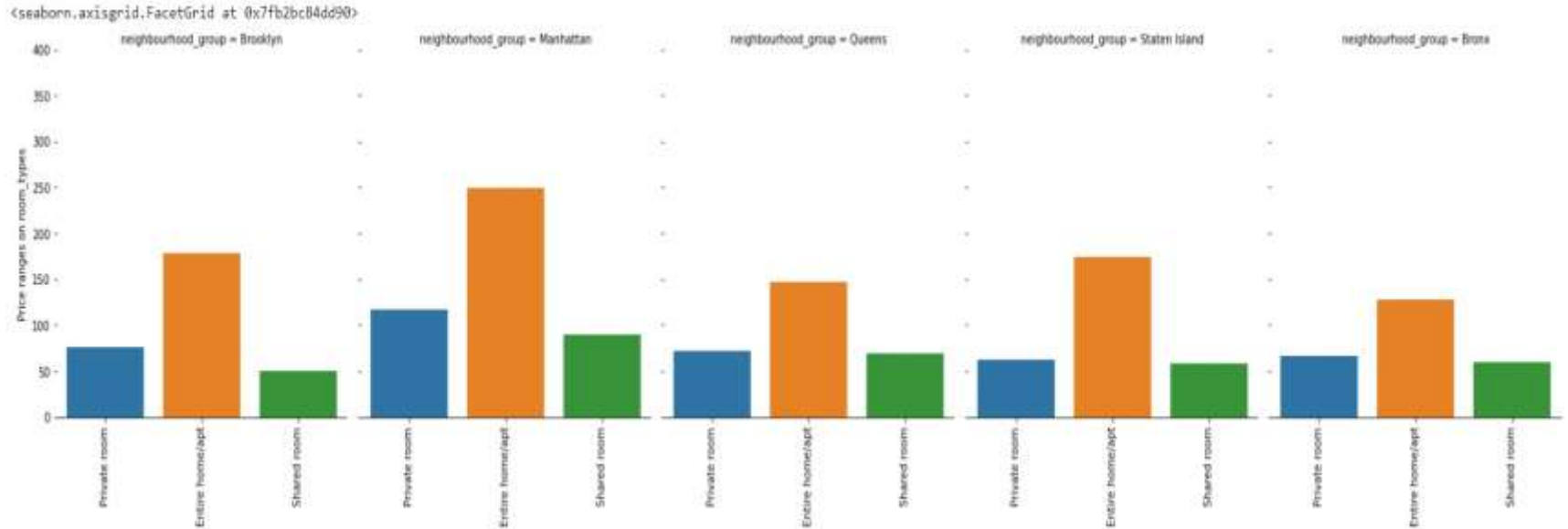


Analysis continued

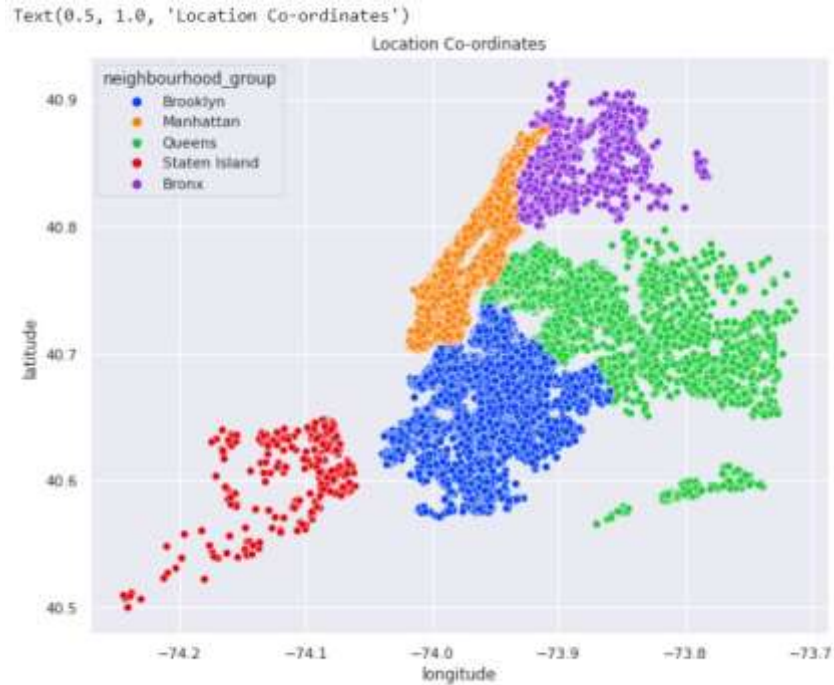
```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619  
warnings.warn(msg, FutureWarning)  
<matplotlib.axes._subplots.AxesSubplot at 0x7fb2bc8ede10>
```



Analysis continued



Analysis continued

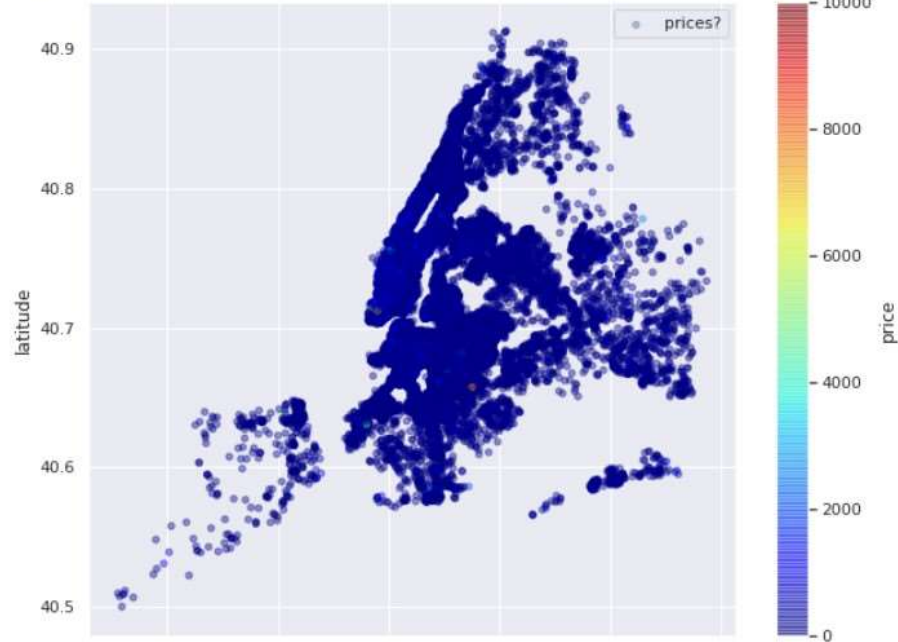


Analysis continued



Analysis continued

<matplotlib.legend.Legend at 0x7fb2bc673c50>



Conclusions

Use new customer basic important features to predict the final destinations where new customers will choose and then make personalized recommendation to them.

Use other models to compare the accuracy.

Focus more on hill areas to attract customers to increase the sales as these areas are the busiest.

Focus on areas other than Manhattan to improve the sales for other areas as these areas are less busy unlike Manhattan.

Challenges

- Finding the busiest host and why was quite challenging as dataset has many types of data related to the customers, prices, reviews, hosts, etc.
- Huge chunk of data was to be handled keeping in mind not to miss anything which is even of little relevance.
- Huge chunk of data takes too much to understand and analyse the dataset.