Capstone Project-1

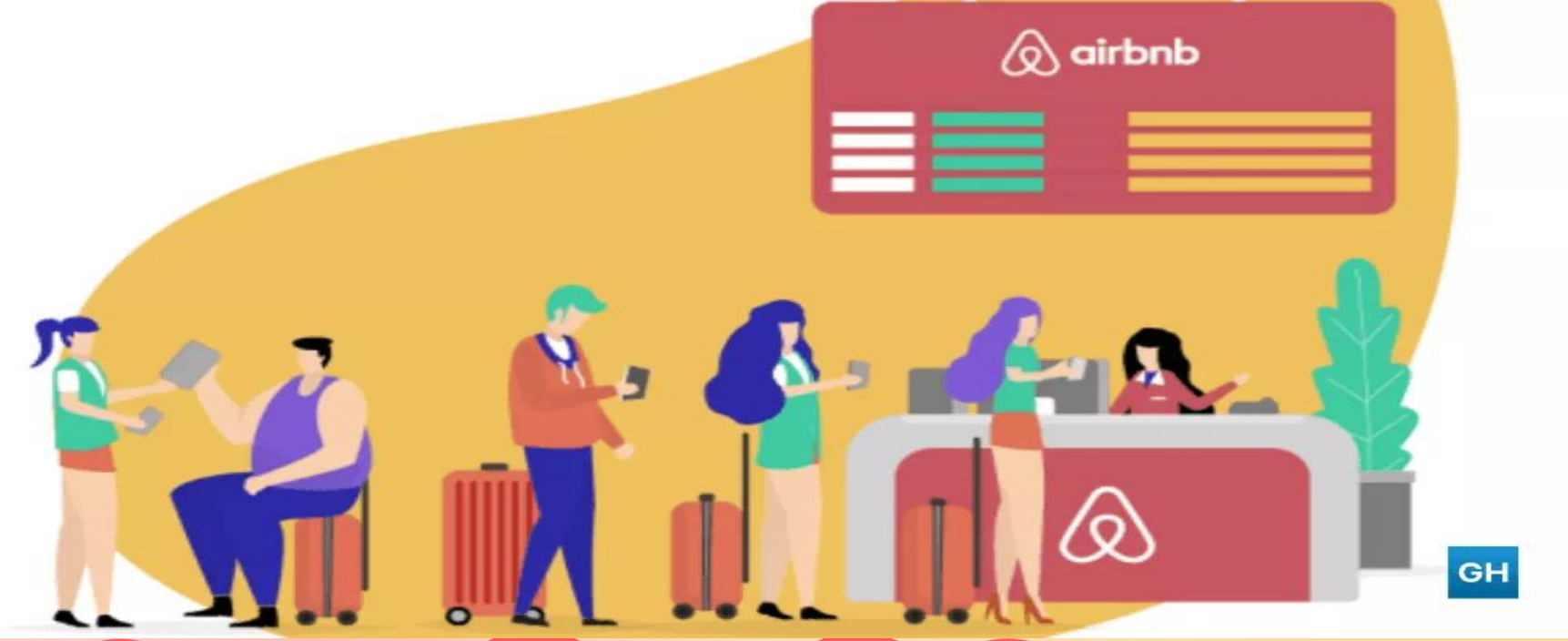
Project Title

Exploratory Data Analysis

Airbnb Bookings Analysis

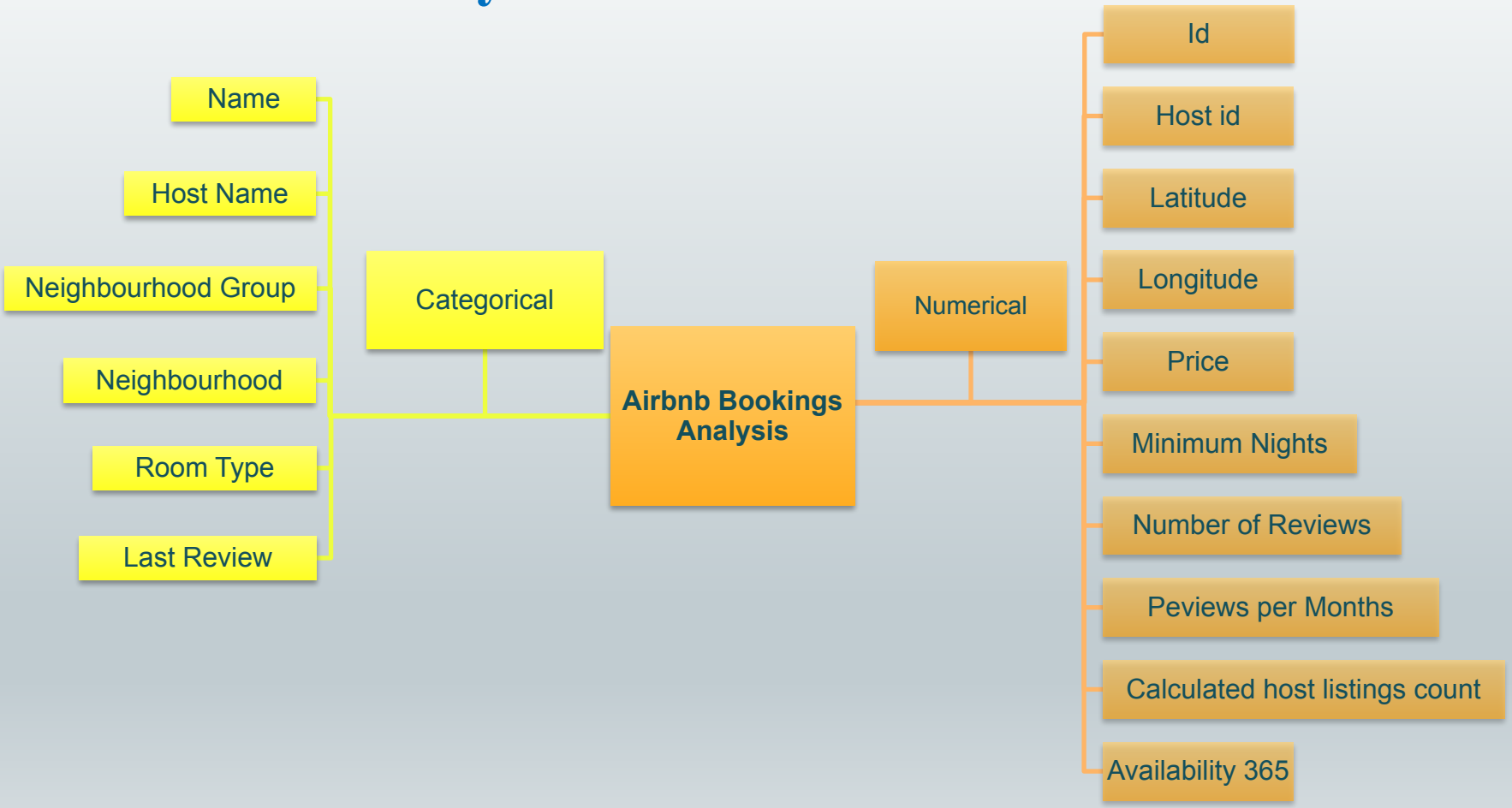
Team Members

Swapnil Zambare
Bhushan Patil
Shubham Patil



PEOPLE + PLACES + LOVE + AIRBNB + MOMENTS =

Data Summary



Definitions

Based on the information on the Airbnb website, the definition of each room type are:

Private room

Guests have exclusive access to the bedroom/sleeping area of the listing. Other parts area such as the living room, kitchen, and bathroom are likely open either to the host even to other guests.

Entire home/apt

Guests have the whole place for themselves. It usually includes a bedroom, bathroom, and kitchen.

Shared Room

Guest sleep in a bedroom or a common area that could be shared with others. bold text.

Dataset

- There are 48895 entries and 16 columns
- 10 out of 16 are numeric:

(Id, host_id, latitude, longitude, price, minimum_nights, number_of_reviews, reviews_per_month, calculated_host_listings_count, availability_365)

- 6 out of 16 are categorical:

(name, host_name, neighbourhood_group, neighbourhood, room_type, last_review)

```
id                                int64
name                             object
host_id                           int64
host_name                        object
neighbourhood_group              object
neighbourhood                    object
latitude                        float64
longitude                       float64
room_type                        object
price                             int64
minimum_nights                    int64
number_of_reviews                 int64
last_review                      object
reviews_per_month               float64
calculated_host_listings_count    int64
availability_365                  int64
dtype: object
```

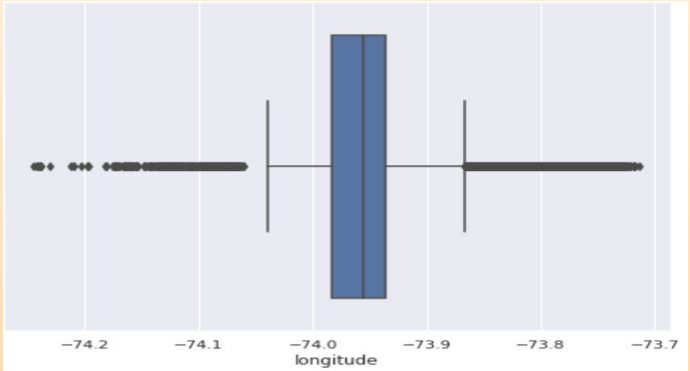
Descriptive Analysis

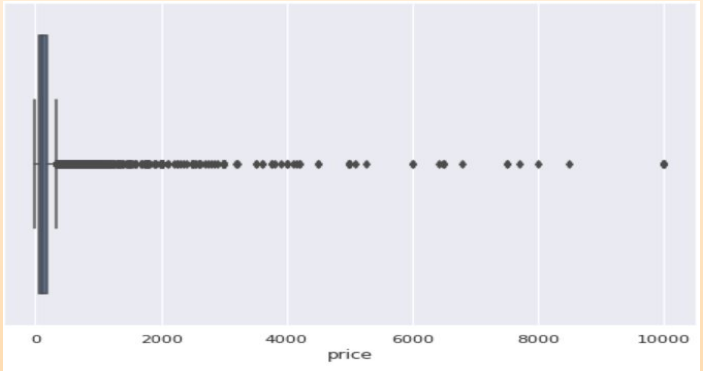
- All columns except latitude and longitude have a big gap between the third quartile and the max value.
- The min value of price, number_of_reviews, reviews_per_month and availability_365 is 0.
- calculated_host_listings_count max value is 327 is very far as compare to quartile value.
- Also max value of minimum_nights is very far as compare to quartile value.

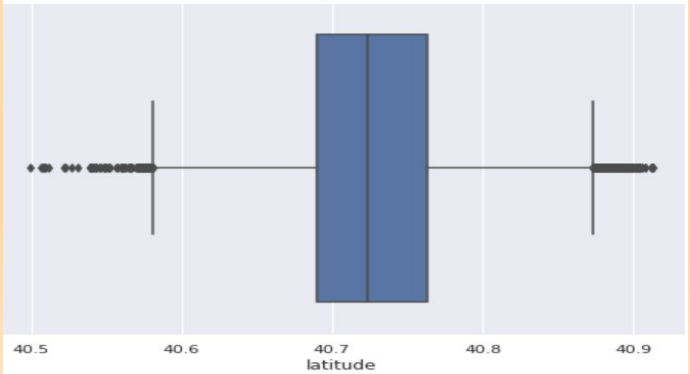
	host_id	latitude	longitude	price	minimum_nights	number_of_reviews	reviews_per_month	calculated_host_listings_count	availability_365
count	4.889500e+04	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000	48895.000000
mean	6.762001e+07	40.728949	-73.952170	152.720687	7.029962	23.274466	1.090910	7.143982	112.781327
std	7.861097e+07	0.054530	0.046157	240.154170	20.510550	44.550582	1.597283	32.952519	131.622289
min	2.438000e+03	40.499790	-74.244420	0.000000	1.000000	0.000000	0.000000	1.000000	0.000000
25%	7.822033e+06	40.690100	-73.983070	69.000000	1.000000	1.000000	0.040000	1.000000	0.000000
50%	3.079382e+07	40.723070	-73.955680	106.000000	3.000000	5.000000	0.370000	1.000000	45.000000
75%	1.074344e+08	40.763115	-73.936275	175.000000	5.000000	24.000000	1.580000	2.000000	227.000000
max	2.743213e+08	40.913060	-73.712990	10000.000000	1250.000000	629.000000	58.500000	327.000000	365.000000

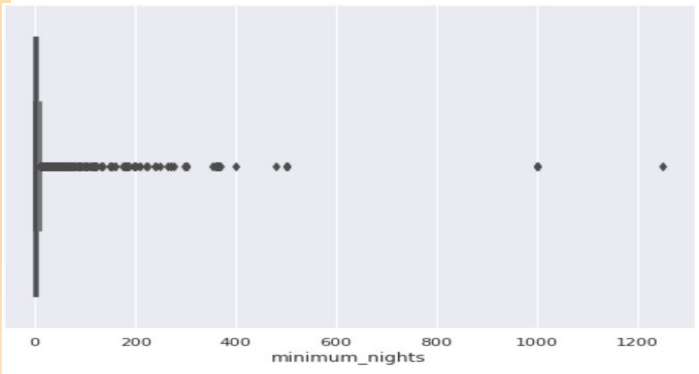
Outliers Data

In price and minimum_nights outliers are present. Also outliers are present in longitude and latitude but we don't take action on them.

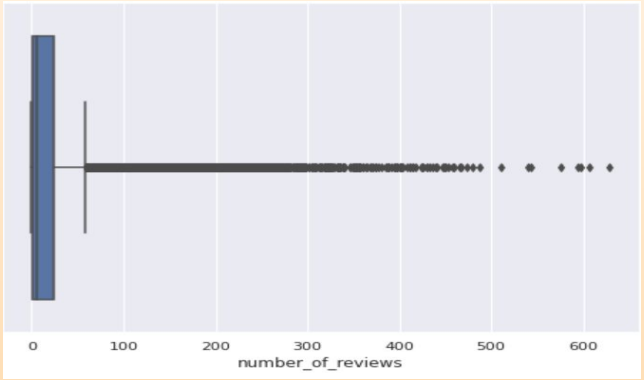


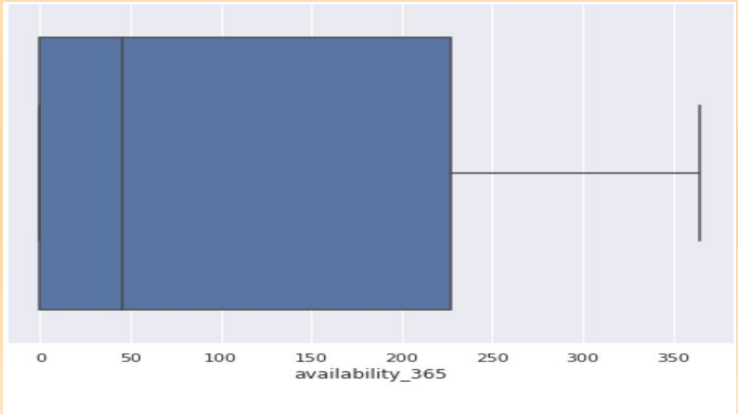






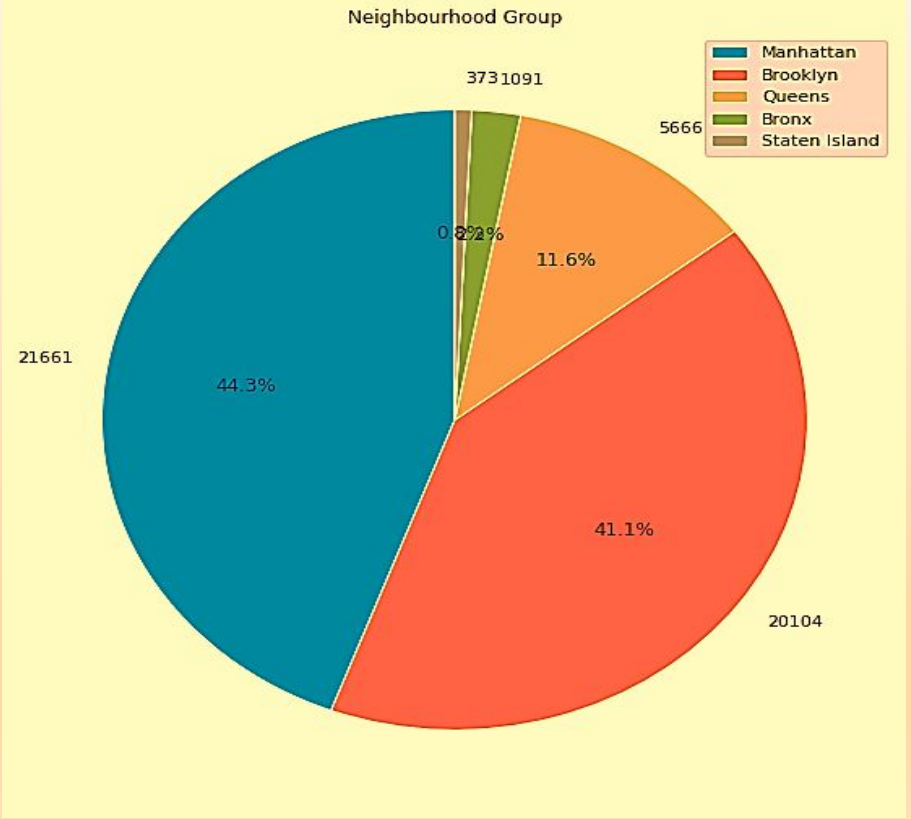
Outliers are present in number_of_reviews, reviews_per_month, calculated_host_listings_count
But for availability_365 there is no single outlier.





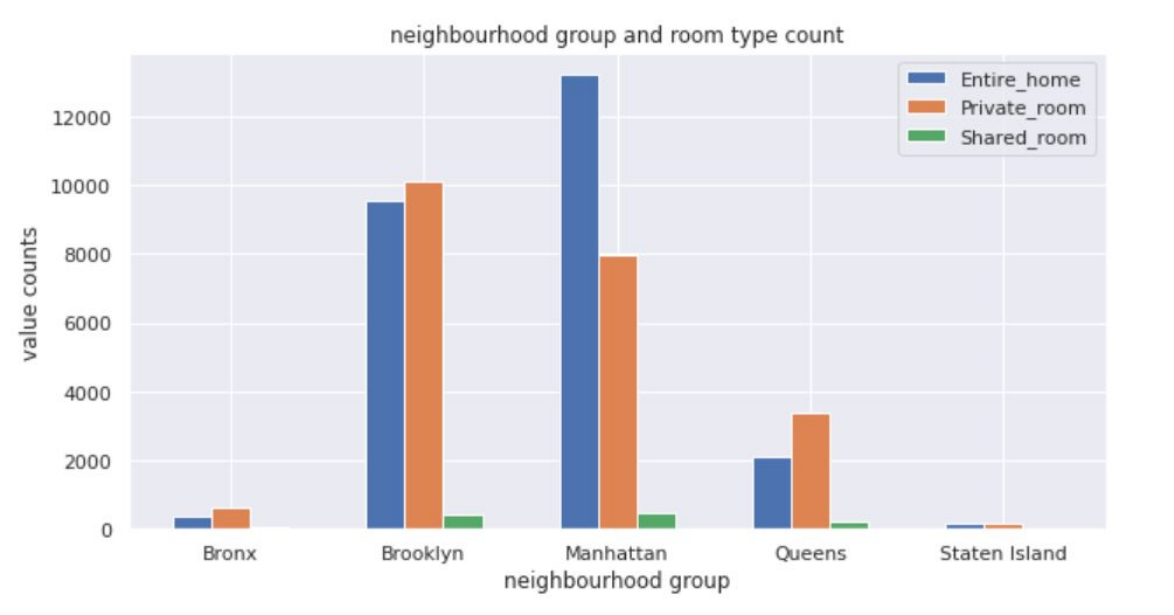
Neighbourhood Group

We have to try to check out the percentage of each neighbourhood group. It turns out that many rooms are rented out in two areas: Manhattan 44.3% and Brooklyn 41.1%. Staten island cover very low area as compare to other four neighbourhood group.



Neighbourhood Group and Room Type

Average price per neighbourhood group and room type shows that Manhattan is the neighbourhood group with the highest room type when compare to other neighbourhood group.

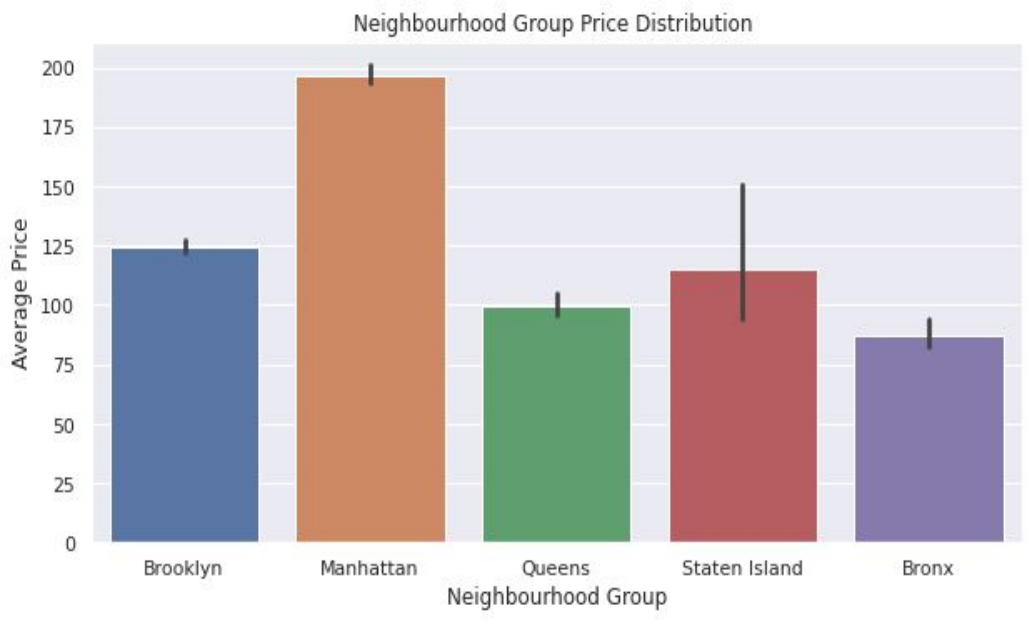


Neighbourhood Group	Room Type	Count
Bronx	Private room	652
	Entire home/apt	379
	Shared room	60
Brooklyn	Private room	10132
	Entire home/apt	9559
	Shared room	413
Manhattan	Entire home/apt	13199
	Private room	7982
	Shared room	480
Queens	Private room	3372
	Entire home/apt	2096
	Shared room	198
Staten Island	Private room	188
	Entire home/apt	176
	Shared room	9

Average price per night

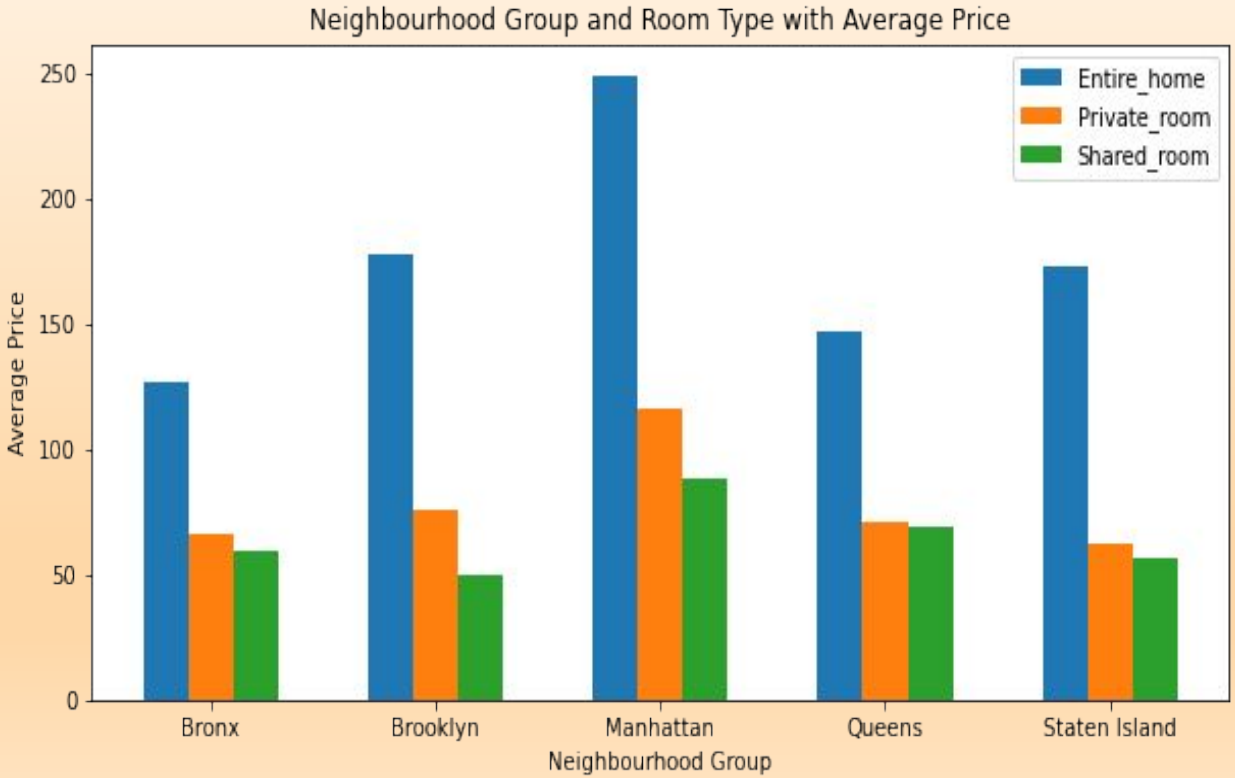
- Average price per night of top 100 reviewed listings is 95.53
- Whole average price per night in New York is 152.72

- **Average price per night in each neighbourhood group**



Neighbourhood Group	Average Price
Bronx	87.496792
Brooklyn	124.383207
Manhattan	196.875814
Queens	99.517649
Staten Island	114.812332

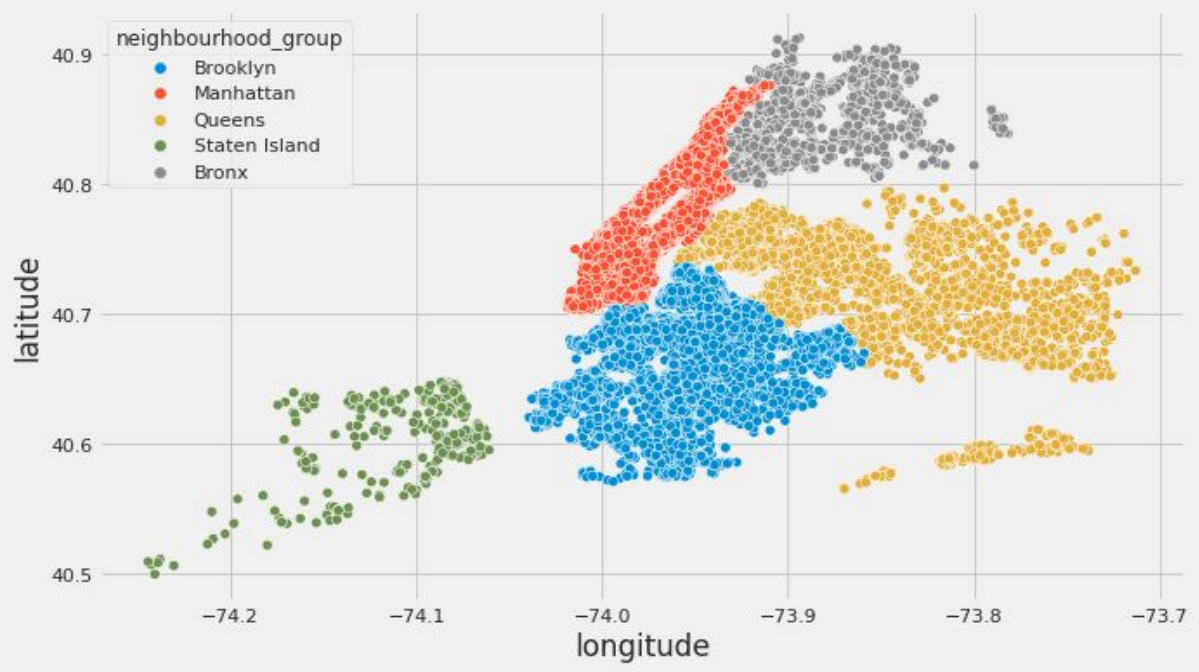
Average Price of Each Room Type in Each Neighbourhood Group



Neighbourhood Group	Room Type	Average Price
Bronx	Private room	127
	Entire home/apt	66
	Shared room	59
Brooklyn	Private room	178
	Entire home/apt	76
	Shared room	50
Manhattan	Entire home/apt	249
	Private room	116
	Shared room	88
Queens	Private room	147
	Entire home/apt	71
	Shared room	69
Staten Island	Private room	173
	Entire home/apt	62
	Shared room	57

Mapping Neighbourhood Group with Lishtings

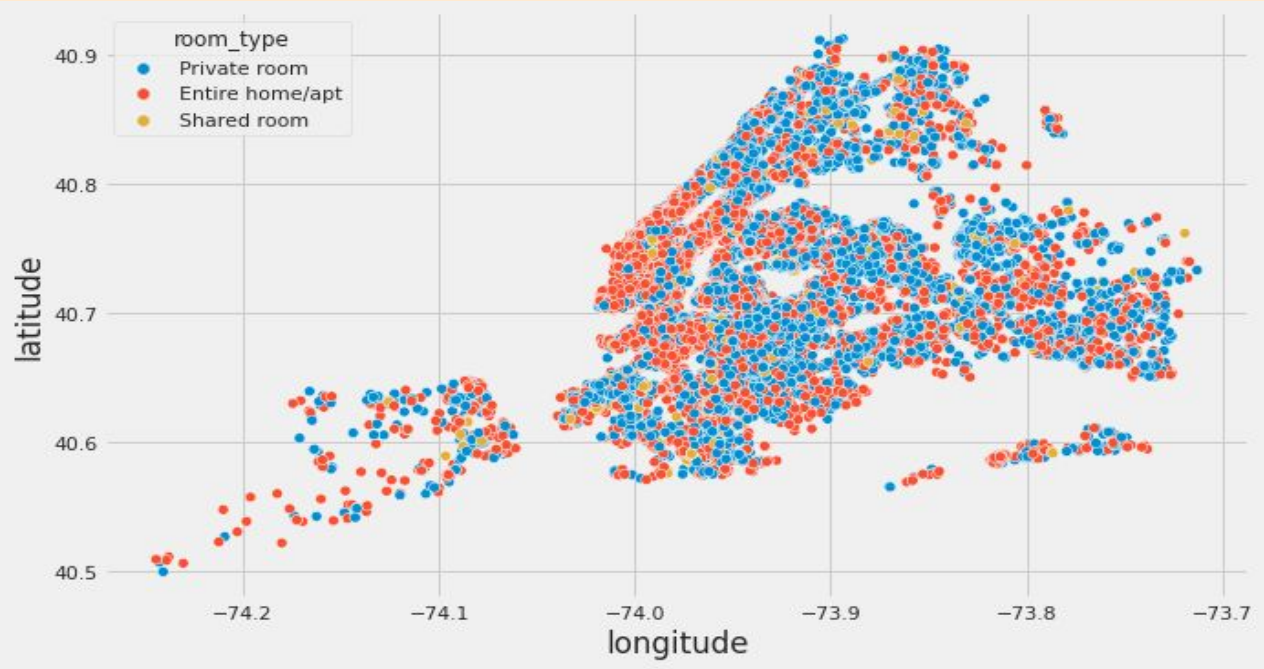
- The below pie chart shows that Manhattan and Brooklyn has highest share of listing as compare to other three.
- Manhattan and Brooklyn are most traveled destinations.



Neighbourhood Group	Listings
Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373

Distribution of Room Type in NYC

- More area covered by Entire home/apt and number of rooms are 25409.
- Least area covered by Shared room and number of rooms are 1160.

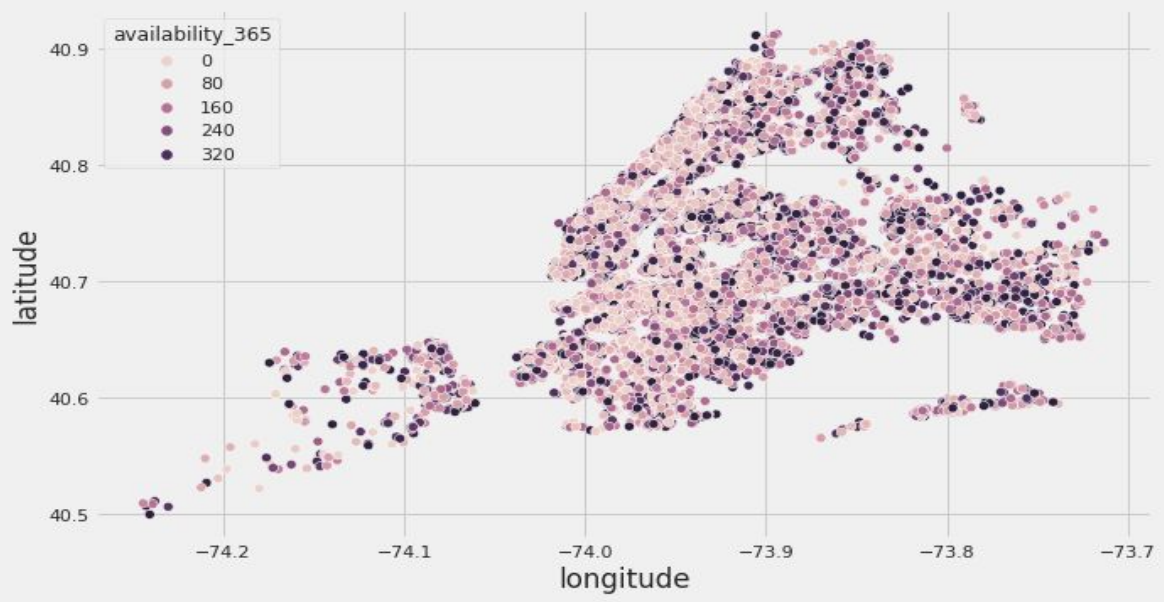


Room Type	Count
Entire home/apt	25409
Private room	22326
Shared room	1160

Room Availability

- More number of rooms are available 0 days of year but less number of rooms are available 202 days of year which are 20. Whole year available rooms are 1295.
- Manhattan has less room availability during 365 days.

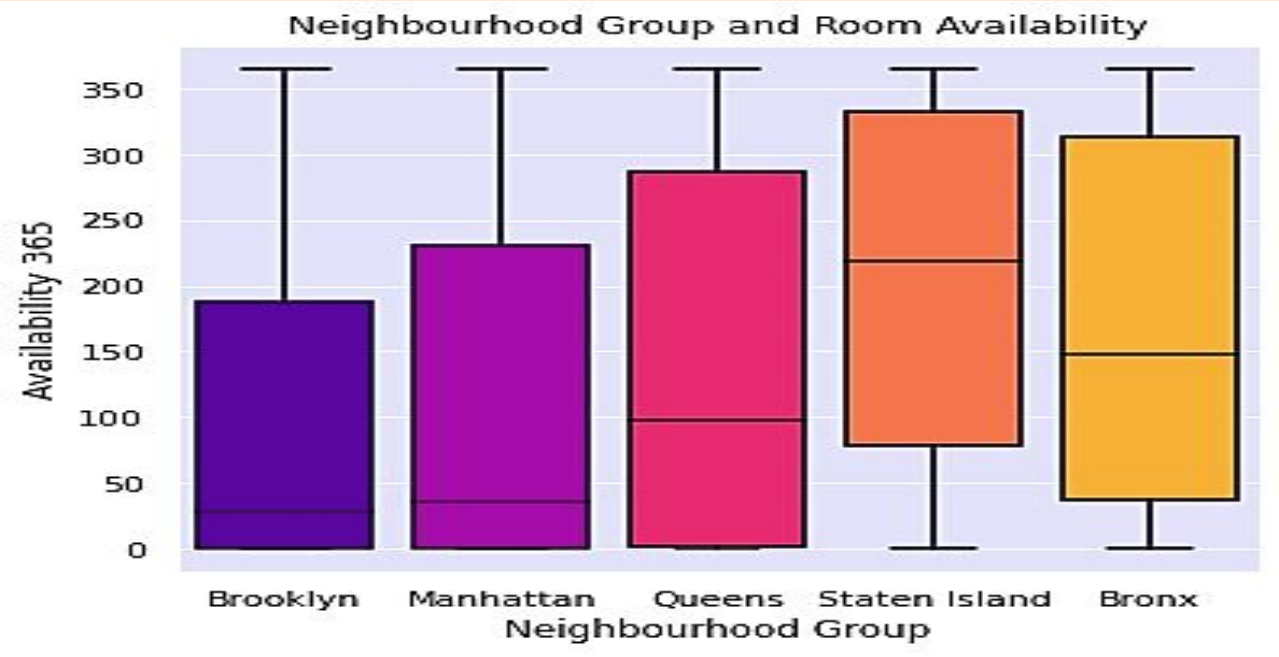
  As people are mostly prefer these areas for bookings.



0	17533
365	1295
364	491
1	408
89	361

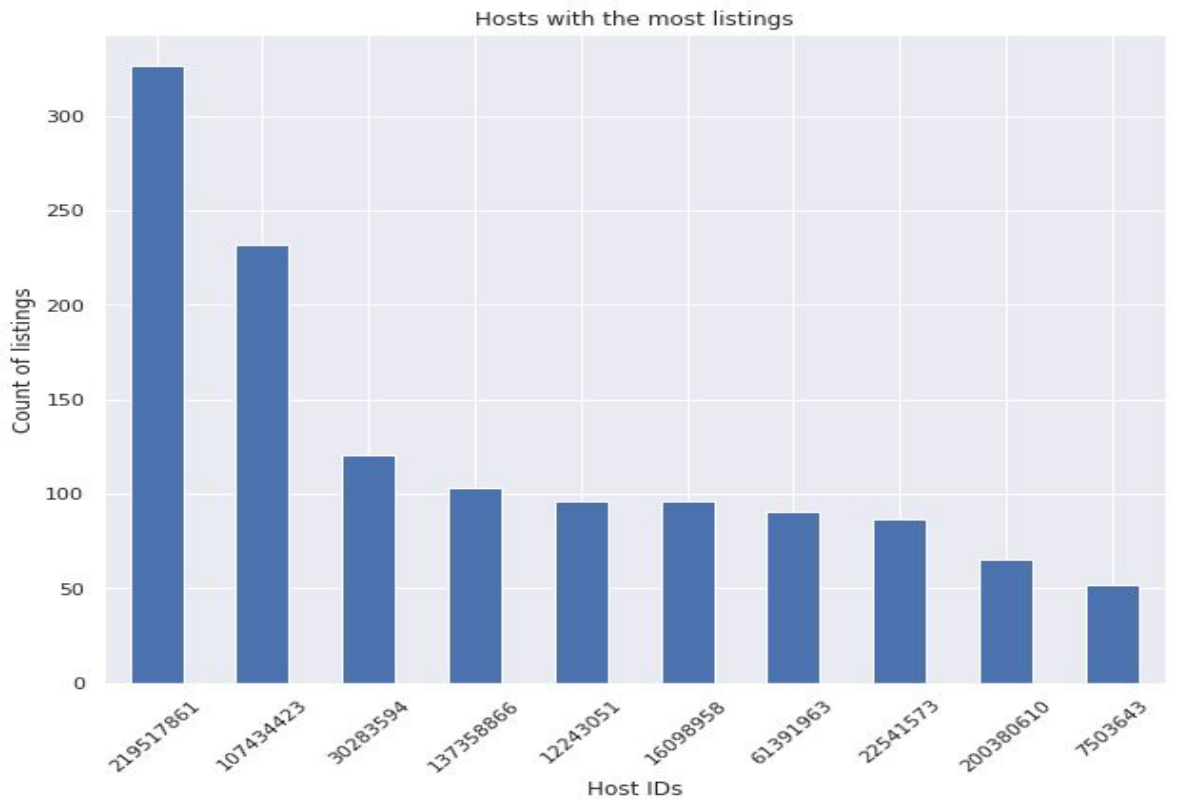
Availability of Room in Entire Year

Brooklyn and Manhattan region is most busy region as there is very less availability of room while in Staten Island there is more than average 200 days rooms are available as people are less prefer than other region.



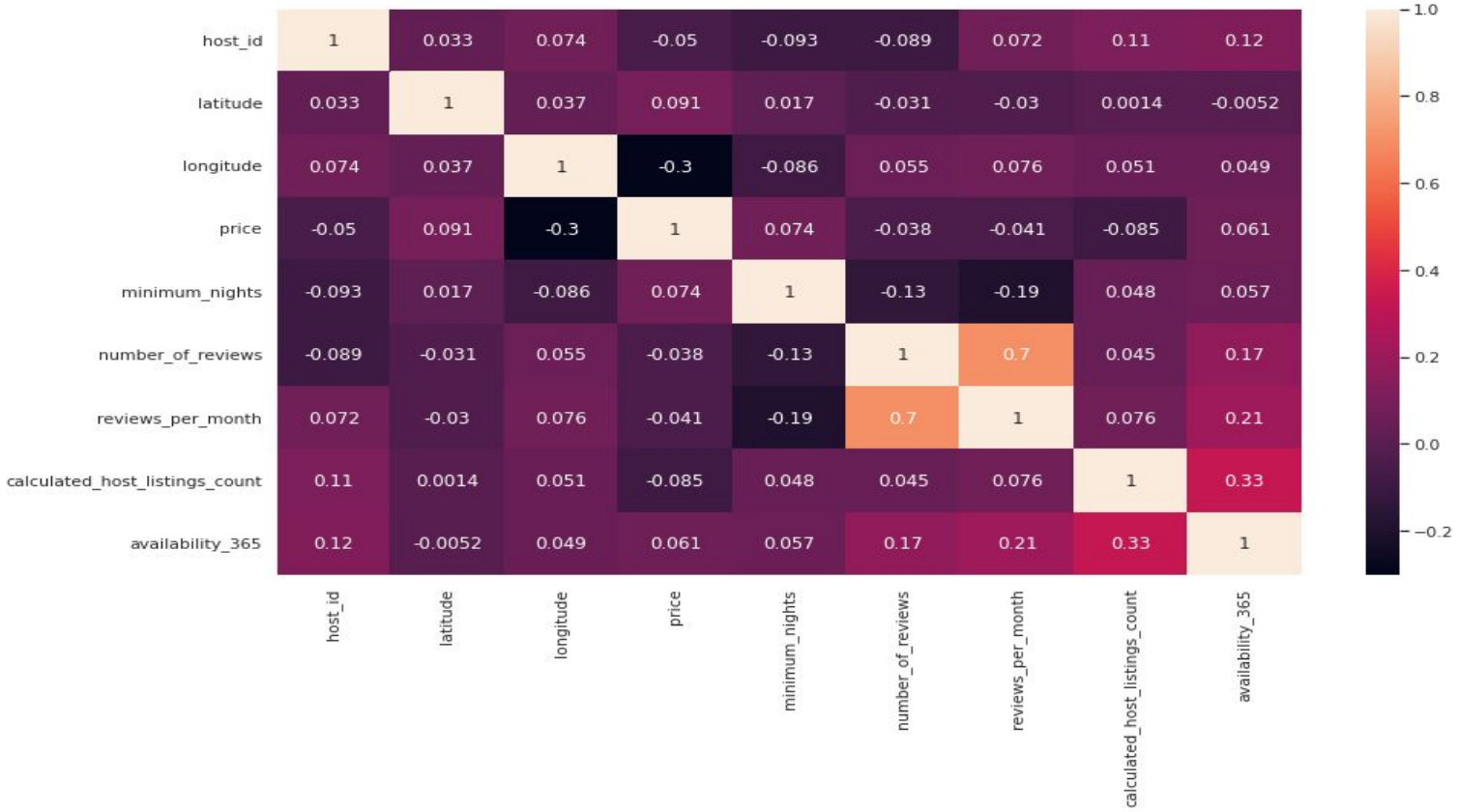
Top busiest Host

Total of top 10 hosts is almost 2%(1270 listings) of the whole dataset.



Host id	Listings
219517861	327
107434423	232
30283594	121
137358866	103
12243051	96
16098958	96
61391963	91
22541573	87
200380610	65
7503643	52

Correlation matrix



Conclusion

- First we have found that host that take good use of Airbnb platform and provide the most listings; we found that our top host has 327 listings.
- Simply by performing EDA on the dataset, we've identified various new insight on how the Airbnb listings distributed on New York, we know where are the listings located, found out the Manhattan is dominating the listing number and have the highest price range
- After we found that which neighbourhood group has more number of boroughs and higher listing densities and which area is more popular than other. Also we found which neighbourhood has higher number of listings.
- Next, we do the good use of longitude and latitude to create color coded geographical map of neighbourhood group, room type and room availability.
- We also found neighbourhood group price distribution and room availability.In which we boxplot the data.
- Overall, we found good relationships between features and explained each step of the process.


THANK
YOU