AIR BNB Analysis

BY – HARSHITA AND ISHAN

Objective

- To analyze the way forward for Air BNB after COVID – how can we increase the profits for Air BNB

- Checking and analyzing the univariant and bivariant analyses for the data to come to some valid pointers in line to the study conducted

BACKGROUND

- AIRBNB, Inc. is an American vacation rental online marketplace company based in San Francisco, California, United States. AIRBNB offers arrangement for lodging, primarily homestays, or tourism experiences.

- Due to COVID situation, AIRBNB has seen drop in revenue.

- Once COVID situation is uplifted, AIRBNB wants to be fully prepared with their precautionary measures & new changes.

Reading and Cleaning the Data

- The Data of Air Bnb was read in the python notebook wherein some points were observed
- The duplicates were removed and check was done for the null values in each column

Removing the Duplicates if any

```
airbnb.duplicated().sum()
airbnb.drop_duplicates(inplace=True)
```

Check for the null values in each column

```
airbnb.isnull().sum()
```

```
id                                  0
name                               16
host_id                             0
host_name                          21
neighbourhood_group                 0
neighbourhood                       0
latitude                            0
longitude                           0
room_type                           0
price                               0
minimum_nights                      0
number_of_reviews                   0
last_review                     10052
reviews_per_month               10052
calculated_host_listings_count      0
availability_365                    0
dtype: int64
```

Cleaning the Data

- Dropping unnecessary columns from the data for analysis and checking the shape of the new data
- Removing the NaN values from the DataSet and verifying if it has been removed from the data
- Examining the continuous variables and understanding where the majority of the data is present

Dropping unnecessary columns

```
airbnb.drop(['name','id','host_name','last_review'], axis=1, inplace=True)
```

```
airbnb.head()
```

	host_id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	reviews_per_month	c
0	2787	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	0.21	
1	2845	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	0.38	
2	4632	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	
3	4869	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	4.64	
4	7192	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	0.10	

```
airbnb.shape
```

```
(48895, 12)
```

Remove the NaN values from the dataset

```
airbnb.isnull().sum()
```

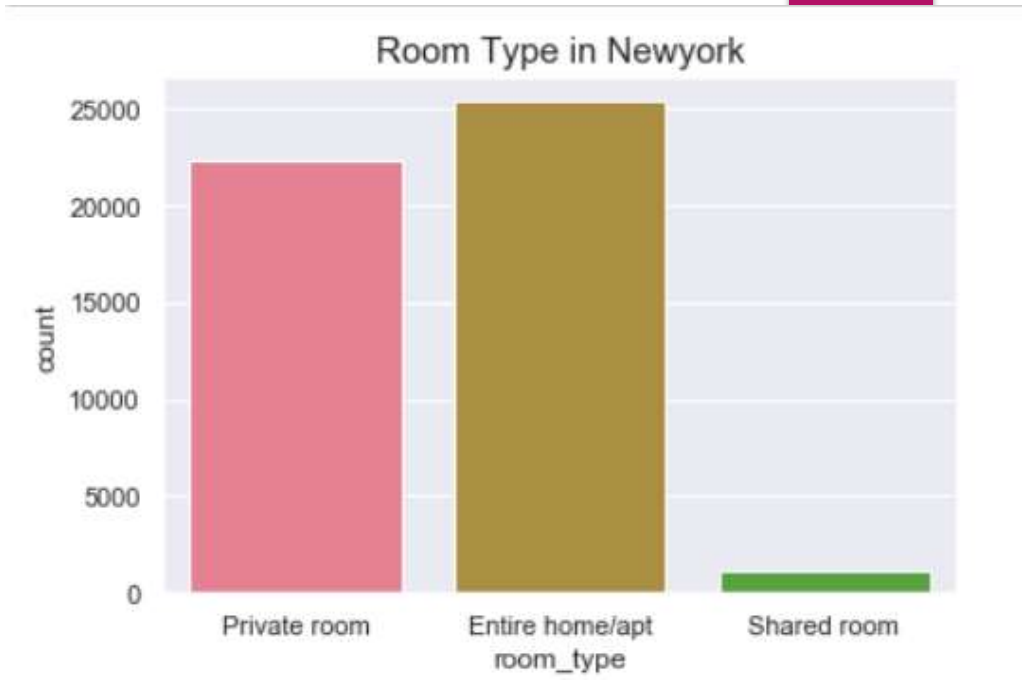
```
host_id                         0
neighbourhood_group             0
neighbourhood                   0
latitude                        0
longitude                       0
room_type                       0
price                           0
minimum_nights                  0
number_of_reviews               0
reviews_per_month               0
calculated_host_listings_count  0
availability_365                0
dtype: int64
```

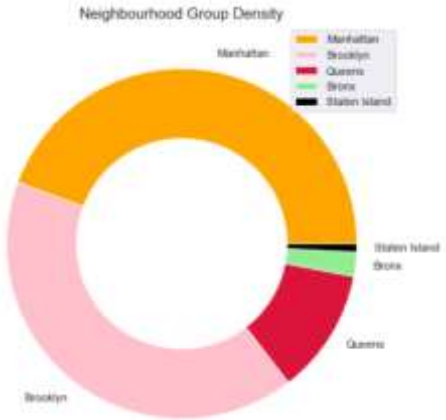
```
airbnb.dropna(how='any',inplace=True)
airbnb.info()
```

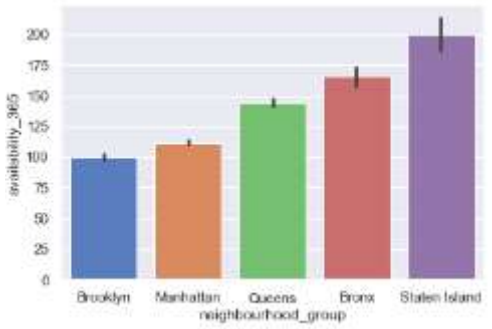
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 48895 entries, 0 to 48894
Data columns (total 12 columns):
host_id                         48895 non-null int64
neighbourhood_group             48895 non-null object
neighbourhood                   48895 non-null object
latitude                        48895 non-null float64
longitude                       48895 non-null float64
room_type                       48895 non-null object
price                           48895 non-null int64
minimum_nights                  48895 non-null int64
number_of_reviews               48895 non-null int64
reviews_per_month               48895 non-null float64
```

Univariate Analysis

- There are basically 3 types of rooms in which we see that Private rooms and Entire Apt / room are the ones which are preferred
- New york have five neighbourhood. All the listings in dataset belong to either one of them. We can visualize which neighbourhood has dense Airbnb listings
- Brooklyn and Manhattan has a largest chunk in this donut of neighbourhood
- The graph shows us **the availability** of the neighbourhood group in a year – it mentions that Staten Island is the most available – which might also imply that there can be a focus deviation required to tap that potential

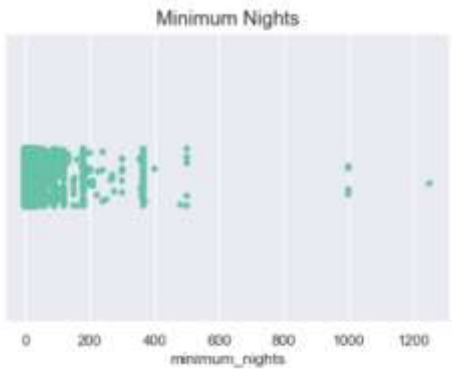


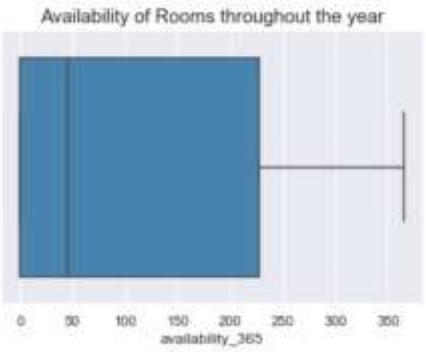




Univariate Analysis

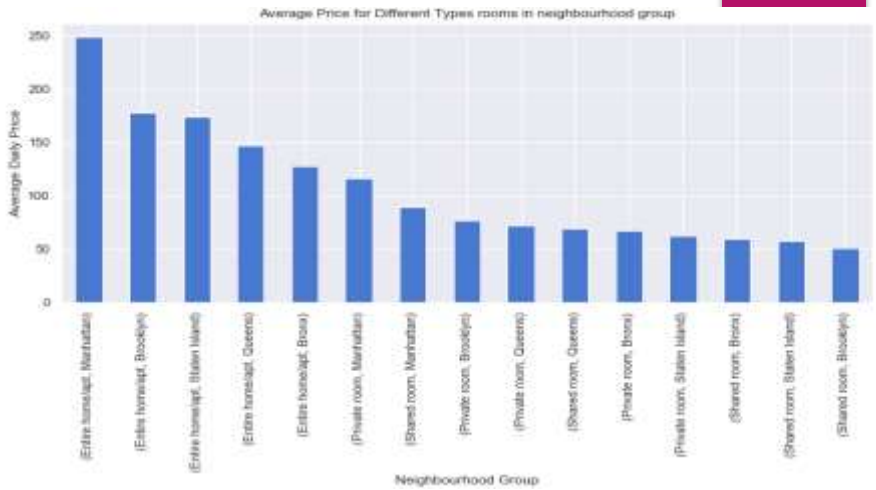
- We can observe that prices of listing start from less than 100 and maximum price reaches around 10000. The distribution curve shows that most of listings prices ranges below 500
- There are listing which are providing service ranging from 1 night to 3 years. Most of the distribution is between 1night to 1 year
- The mean of availability is around 110 which indicates probability of finding a room is 1/3 through out a year



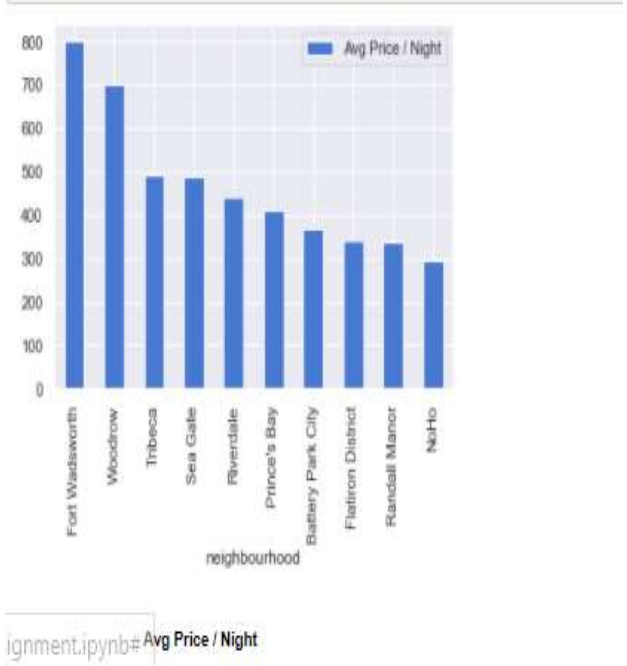


Average Price of Different types of room

- The adjacent graph shows that which room is the most priced compared to the other neighborhood areas. It shows that Entire Home/apt of Manhattan is the most expensive while Standard room of Brooklyn is the least expensive
- Fort Wadsworth seems to be the most Expensive neighborhood on avg. out of 221 total neighborhoods.



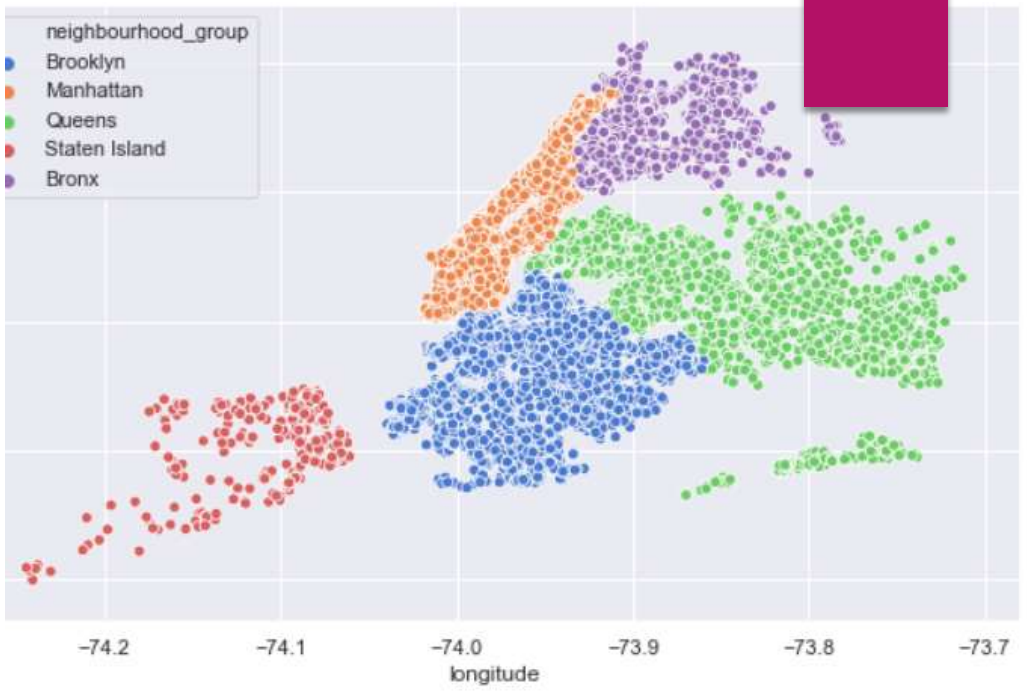
		price
room_type	neighbourhood_group	
Entire home/apt	Manhattan	249.239109
	Brooklyn	178.327545
	Staten Island	173.846591
	Queens	147.050573
	Bronx	127.506596
Private room	Manhattan	116.776622
Shared room	Manhattan	88.977083
Private room	Brooklyn	76.500099
	Queens	71.762456
Shared room	Queens	69.020202
Private room	Bronx	66.788344
	Staten Island	62.292553
	Bronx	59.800000
Shared room	Staten Island	57.444444
	Brooklyn	50.527845

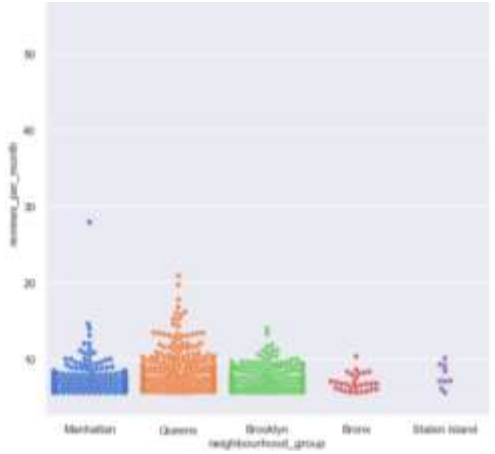


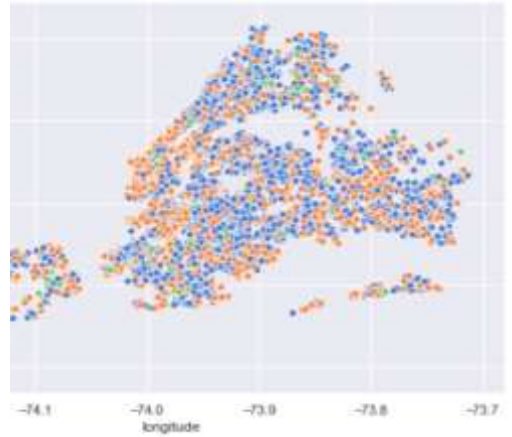
ignment.ipynb# Avg Price / Night

Bivariate Analysis Using MAP Distribution

- The map shows us the distribution of the room types as per the lat long of the place
- Also it shows the distribution of the neighbourhood as well in the beside depiction

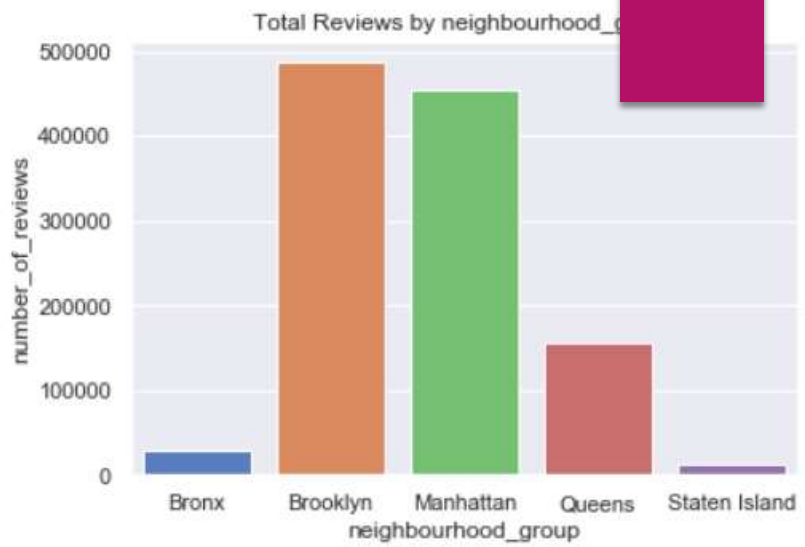


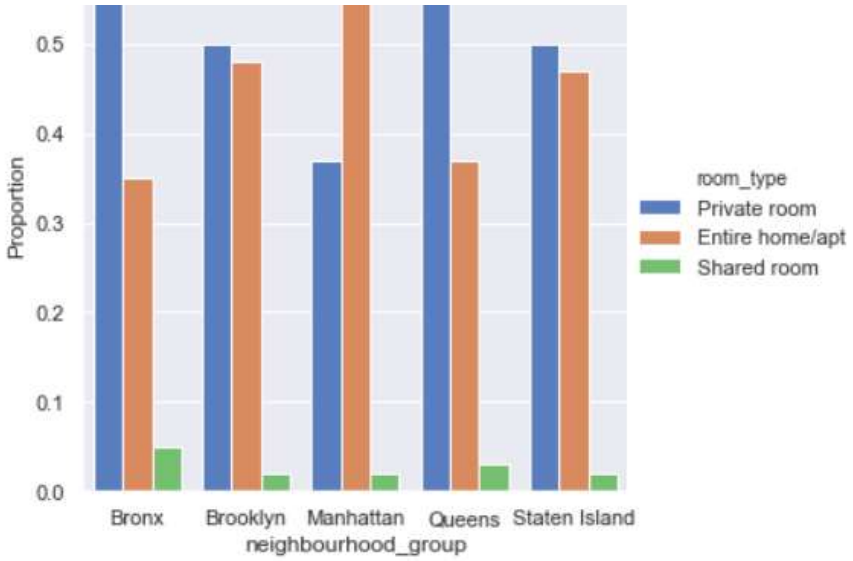




Bivariate Analysis

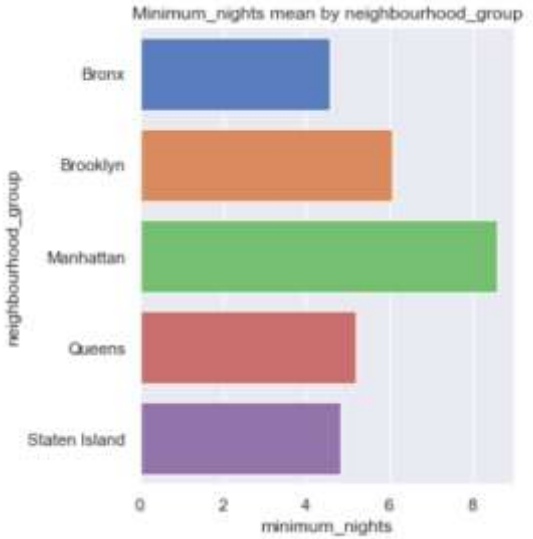
- The graph shows us the total reviews by neighbourhood group and no of reviews
- Also the below graph depicts the **proportion of "private room", "entire apt" and "shared room" for each "neighbourhood group“**
- Manhattan and Brooklyn are probably well received in terms of reviews because they are the centers of attraction .Every listing in those locales are mostly built in a way that makes a Tourist feel at home .





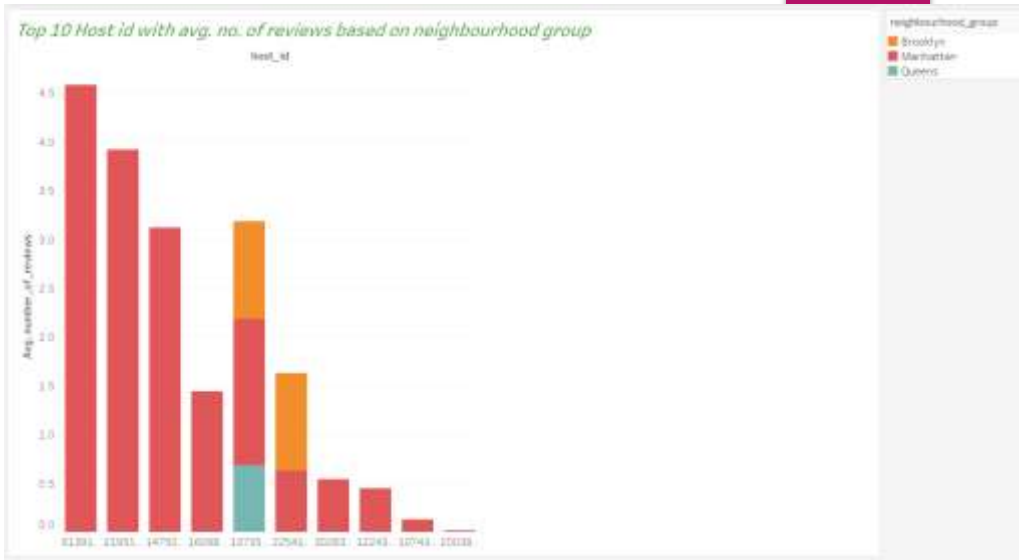
Bivariate Analysis

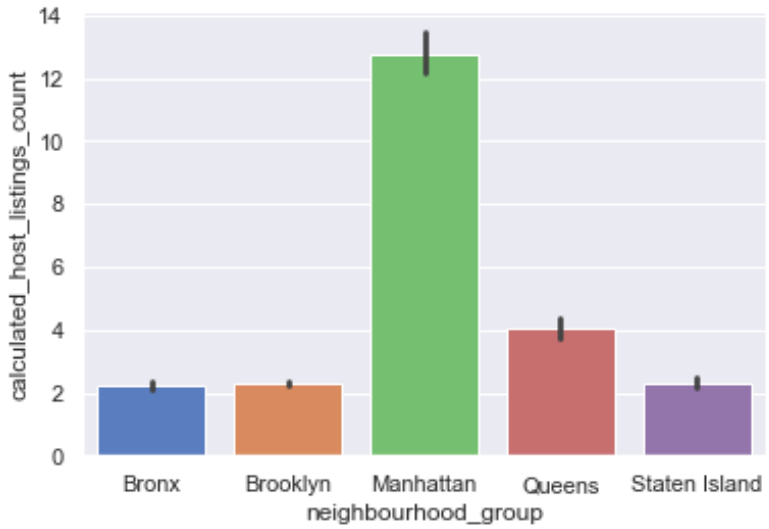
- The graph shows the minimum nights mean by the neighbourhood group
- Also shows the map for the prices and the room per neighbourhood group
- It shows that Manhattan and Brooklyn are the key neighbourhood groups which are in demand



Host Listing

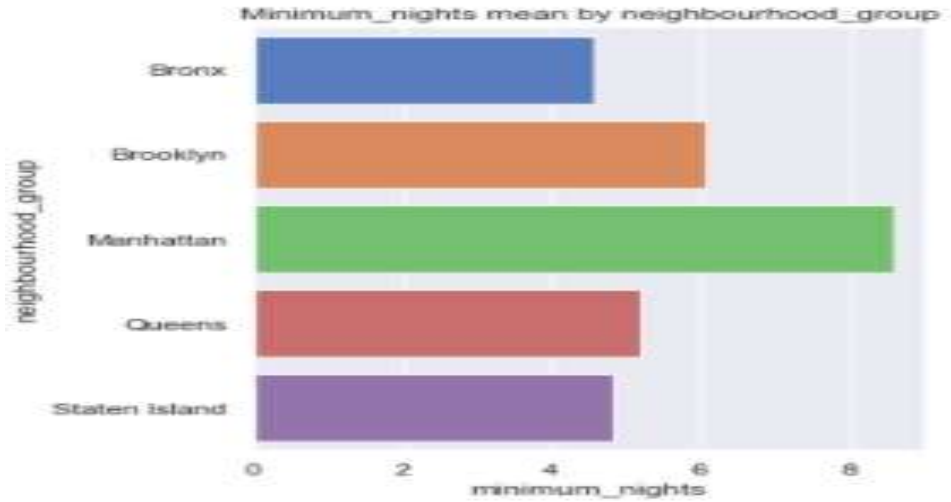
- Manhattan and Brooklyn have most number of reviews in top 10 host id list.
- Top Hosts make their money from Brooklyn and Manhattan .
- Also, Here we see that Manhattan is the most listed neighbourhood amongst the 5 which are being considered as the count of its host listing tops the charts.

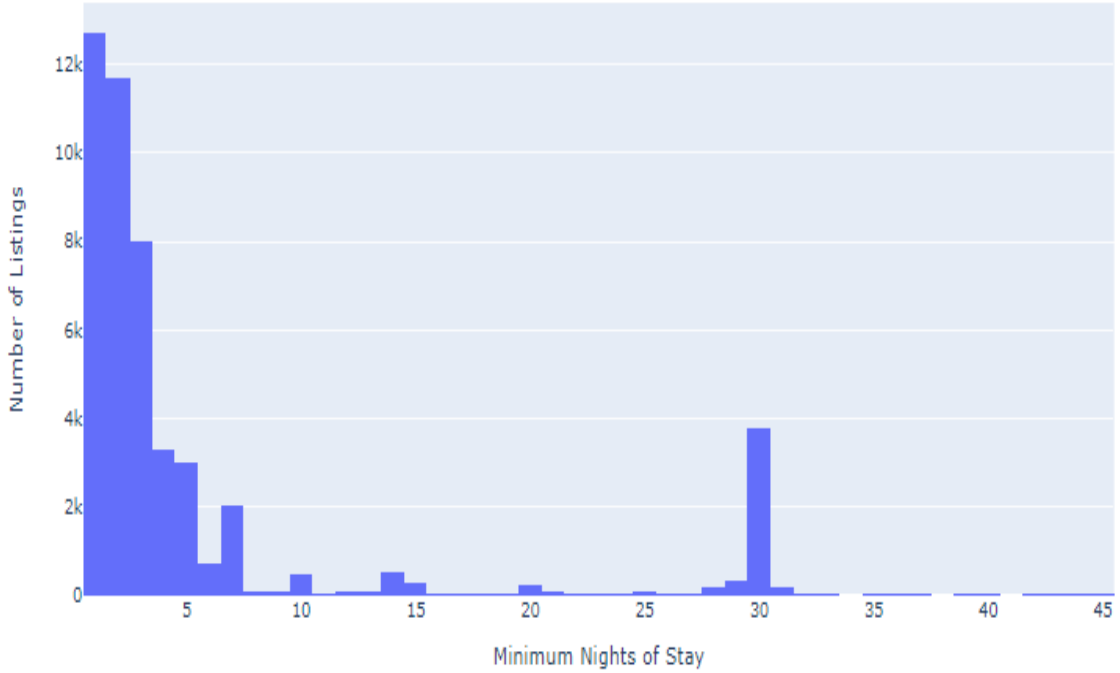




Minimum Nights of Stay

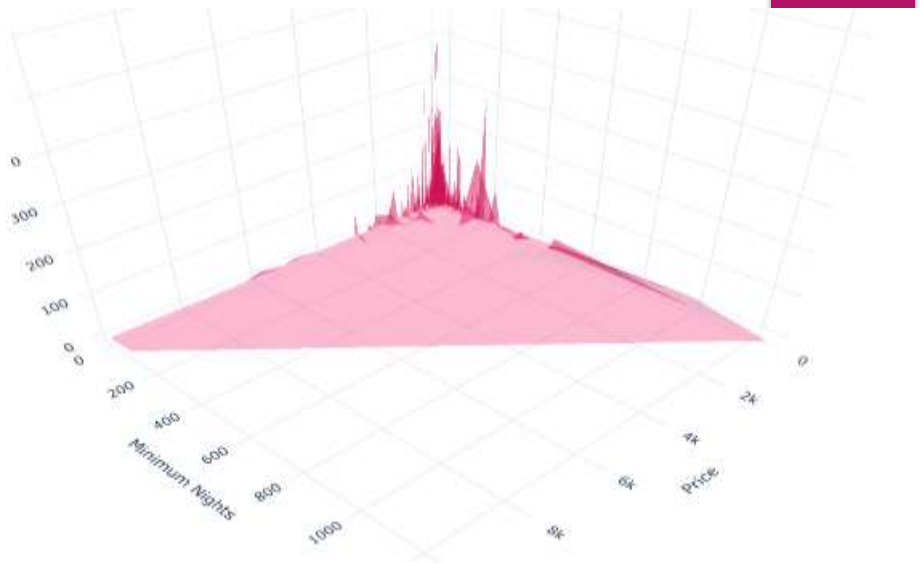
- Manhattan and Brooklyn are mostly preferred for minimum night for stay.
- Distribution plot (Histogram) on 'minimum nights' show the variation at different points
- - Distribution of the 'minimum nights' feature
- - Here in the Distribution plot shows also shows at '30 Day' spike in the number of listing properties

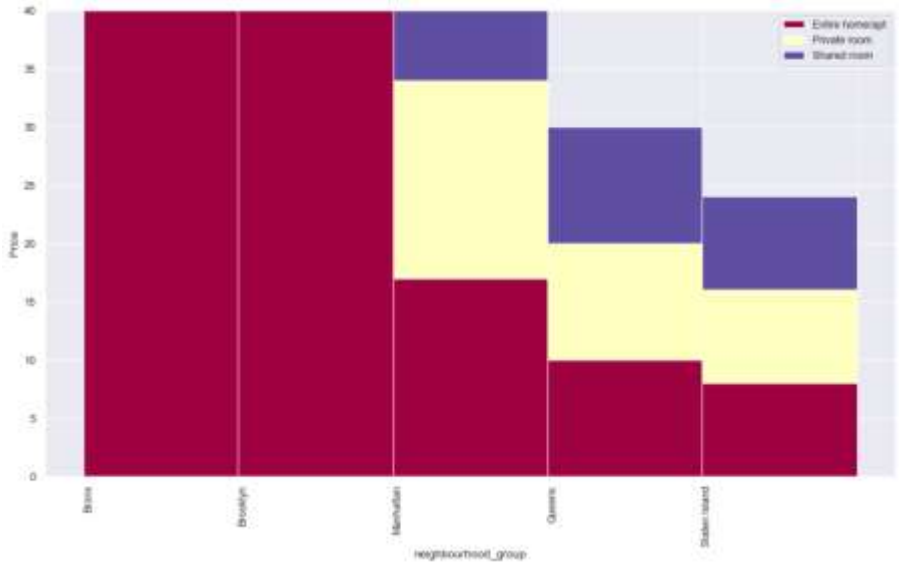




Use of Heat Map /Scatter Plot Analysis and Price Analysis

- 3D plot explains, properties with higher number of minimum nights of stay and higher price has no reviews and vice versa.
- Here we can note that Brooklyn and Manhattan tend to have more listings with price > 150. Also, most listings above price > 100 are entire home type followed by private room and shared room which is the cheapest.





Recommendations

- Manhattan and Brooklyn are the key neighbourhood groups which are in demand based on host listings.
- Major booking which happen is for the Entire Home/Apt and least is for shared – therefore focus should be on these two.
- Brooklyn and Manhattan has more demand of Entire Home/Apt booking based on average price.
- It's pretty obvious that both Manhattan and Brooklyn are famous for tourism.

Appendix

- We have conducted this analysis using following process –
- Cleaned dataset by removing duplicates, treating outliers/missing values & removing null values and unnecessary columns etc.
- Performed uni-variate & bi-variate analysis with all affecting features to draw some insights.
- Started visualizing the dataset using python , tableau & plotly. Based on that provided outcome/results & recommendations.

- **Attached Detailed Document -**

Thank you