

AIRBNB CASE STUDY

Problem Statement -Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.

Story Boarding:

- ✓ Went through data/records provided and get used to with it, noted down important fields and points.
- ✓ Designed various templates on the basis of mind map/Presentation.

Objective: is to improve understanding about market condition once covid situation improved or market is about to open.

Below are key things to be achieved via this analysis:

- ✓ Improve safety measures for customers and understand their needs.
- ✓ Suggest recommendations to team to gear up and improve existing services.

Data set are cleaned using python and Important graph and relations were interpreted via plotly/Seaborn & Tableau.

Step-1 Reading the file AB_NYC_2019.csv using pandas.

```
bnbair=pd.read_csv("AB_NYC_2019.csv")
bnbair.head()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149		1
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225		1
2	3647	THE VILLAGE OF HARLEM....NEW YORK!	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150		3
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89		1
4	5022	Entire Apt. Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80		10

✓ Check for shape and Size

```
In [3]: bnbair.shape
```

```
Out[3]: (48895, 16)
```

✓ Check for null values in each column

Checking null values Percentage -In each Column

```
In [4]: round(bnbair.isnull().sum().sort_values(ascending=False)/len(bnbair.index),2)*100
```

```
Out[4]: reviews_per_month      21.0
last_review                    21.0
host_name                      0.0
name                           0.0
availability_365               0.0
calculated_host_listings_count 0.0
number_of_reviews              0.0
minimum_nights                 0.0
price                          0.0
room_type                     0.0
longitude                      0.0
latitude                      0.0
neighbourhood                 0.0
neighbourhood_group           0.0
host_id                       0.0
id                             0.0
dtype: float64
```

There are two column reviews_per_month and last_review has maximum null values

There are two column reviews_per_month and last_review has maximum null values

✓ Dropping un-necessary column & replacing null values with zero.

Removing Columns that are not needed

```
In [5]:bnbair.drop(['name','id','host_name','last_review'], axis=1, inplace=True)
```

Replacing null values in column reviews_per_month to 0

```
In [6]:bnbair.fillna({'reviews_per_month':0},inplace=True)
bnbair.reviews_per_month.isnull().sum()
```

```
Out[6]: 0
```

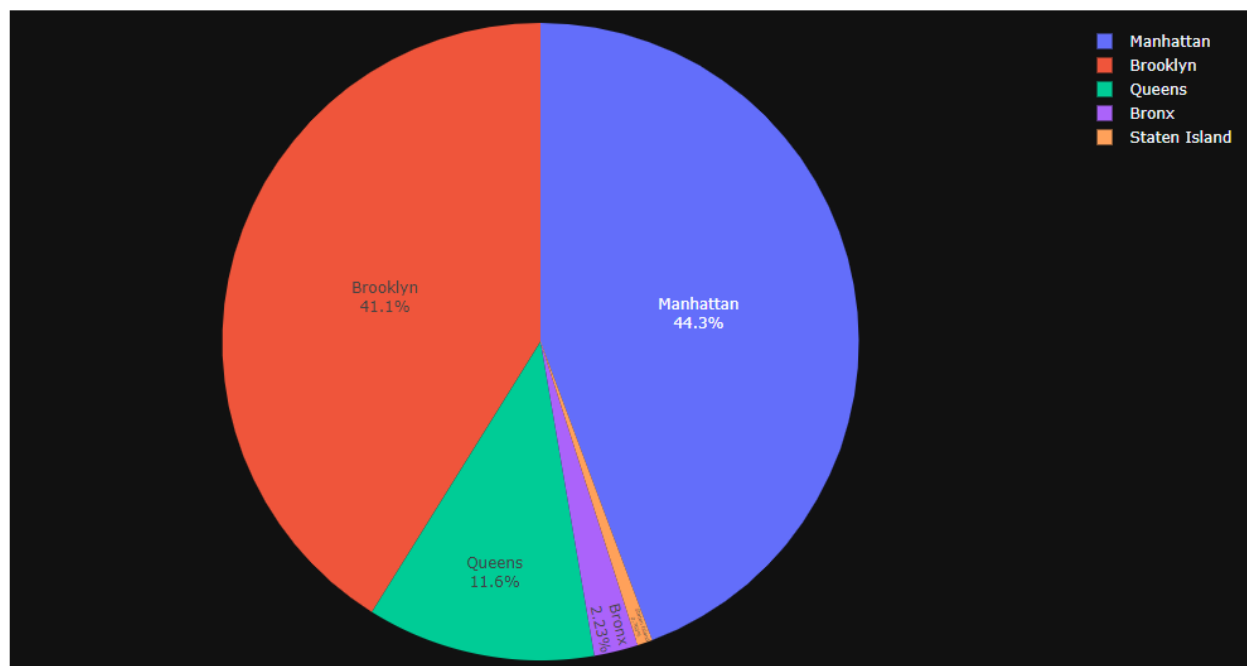
```
In [7]:bnbair.shape
```

```
Out[7]: (48895, 12)
```

Step-2 –Data Analysis

- ✓ Check for higher presence of Airbnb

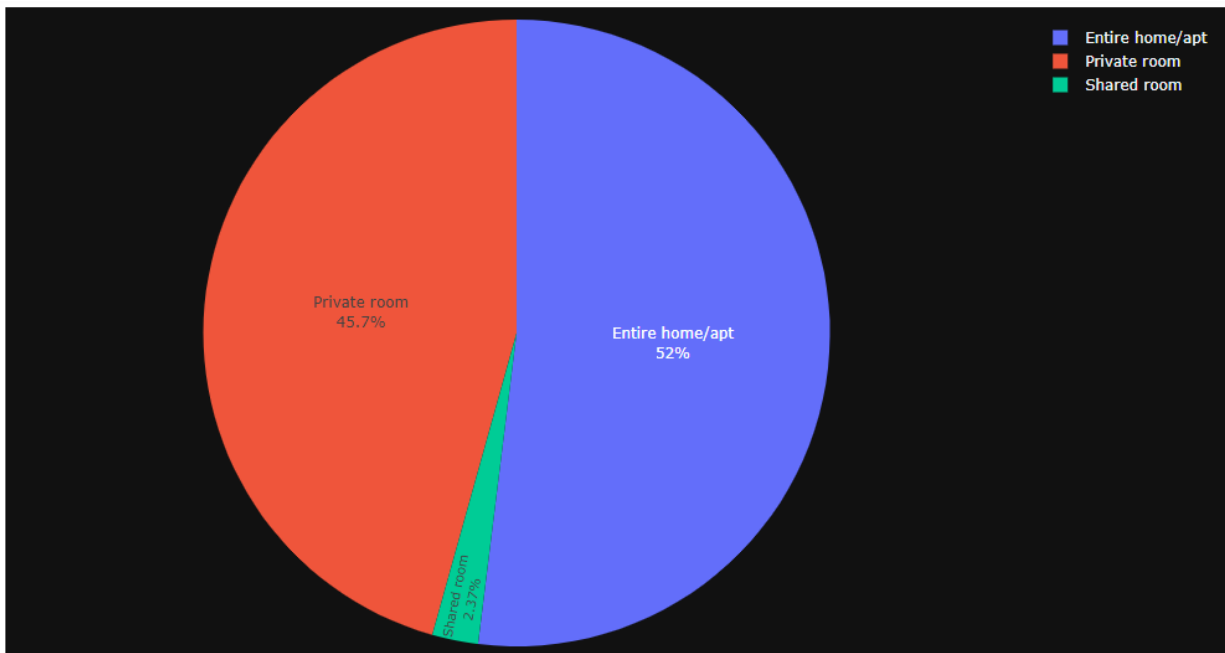
```
grp_neighbour=bnbair['neighbourhood_group']
fig = px.pie(values=grp_neighbour.value_counts(),names=grp_neighbour.value_counts().index)
fig.update_traces(textposition='inside',textinfo='percent+label')
fig.update_layout(
    margin=dict(l=10, r=10, t=10, b=10),
    template="plotly_dark",
)
fig.show()
```



From above, we can see that Manhattan and Brooklyn are having higher density.

- ✓ Type of Room –Check for % of room in Airbnb

```
grp_room_type=bnbair['room_type']  
fig = px.pie(values=grp_room_type.value_counts(),names=grp_room_type.value_counts().index)  
fig.update_traces(textposition='inside',textinfo='percent+label')  
fig.update_layout(  
    margin=dict(l=10, r=10, t=10, b=10),  
    template="plotly_dark",  
)  
fig.show()
```

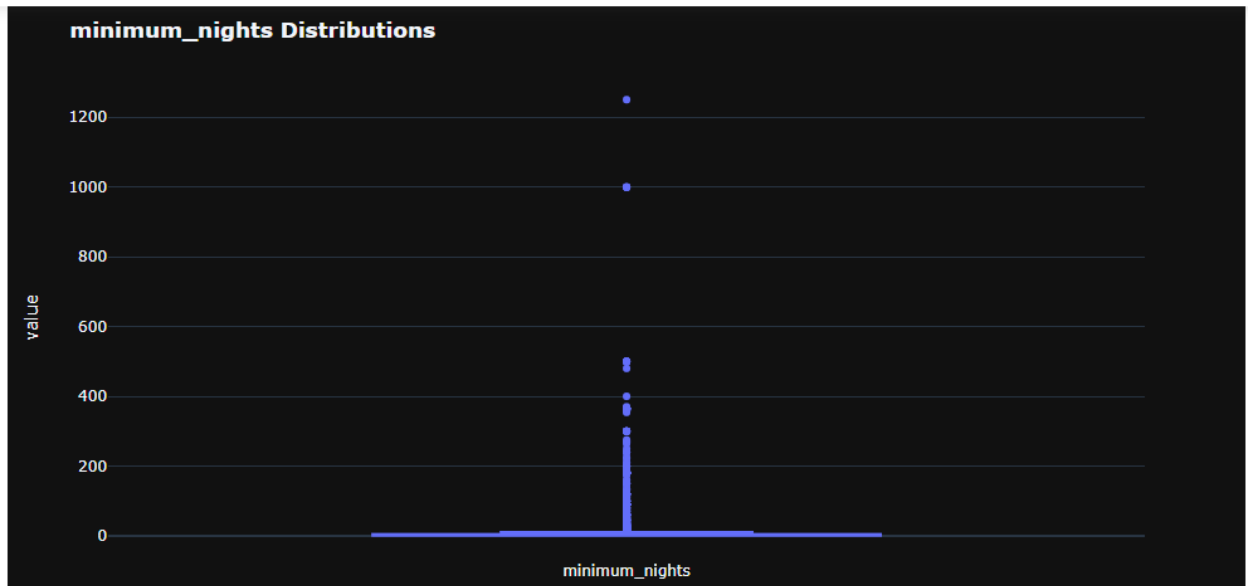


Entire home/apt has highest share followed by private room and shared room has very low share compared to other two category –means people do not mind spending extra bucks while travel or using services of Airbnb.

- ✓ Column Minimum Nights

Minimum Nights ¶

```
import plotly.express as px  
fig = px.box(bnbair['minimum_nights'],template="plotly_dark")  
fig.update_layout(title_text="<b>minimum_nights Distributions ">  
fig.show()  
bnbair['minimum_nights'].describe(percentiles=[.05, .25,.5, .75,.90, .95,.99])
```



```
count    48895.000000
mean       7.029962
std       20.510550
min        1.000000
5%         1.000000
25%        1.000000
50%         3.000000
75%         5.000000
90%        28.000000
95%        30.000000
99%        45.000000
max       1250.000000
Name: minimum_nights, dtype: float64
```

Average of column minimum_night is 7 days at least for week.

If you see the data carefully, it can be observed most of customer around 95 percent ; plan it for month .

✓ Column Availability 365 :

```
]]: import plotly.express as px
fig = px.box(bnbair['availability_365'],template="plotly_dark")
fig.update_layout(title_text="<b>Throughout year room availability")
fig.show()
bnbair['availability_365'].describe(percentiles=[.05, .25,.5, .75,.90, .95,.99])
```



```

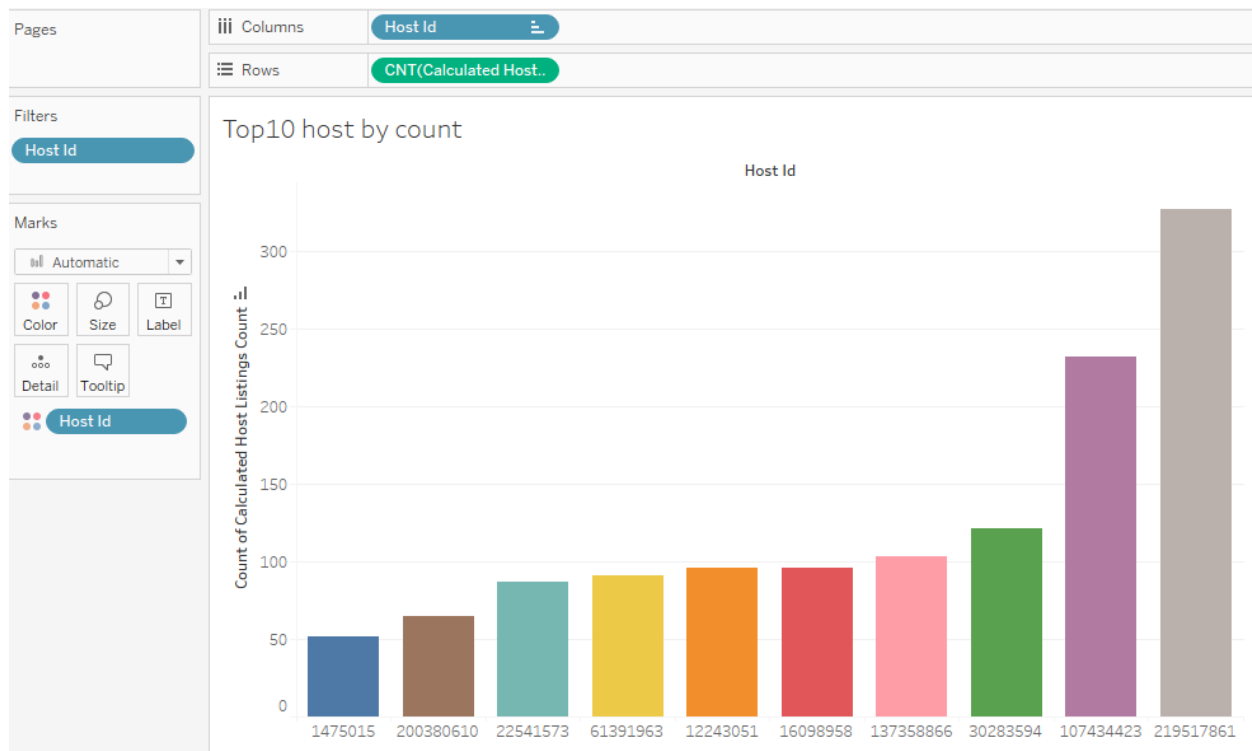
count    48895.000000
mean      112.781327
std       131.622289
min        0.000000
5%         0.000000
25%        0.000000
50%        45.000000
75%       227.000000
90%       337.000000
95%       359.000000
99%       365.000000
max        365.000000
Name: availability_365, dtype: float64

```

The mean of availability is 112 which indicates the probability is $112/365 \approx 30$ percents

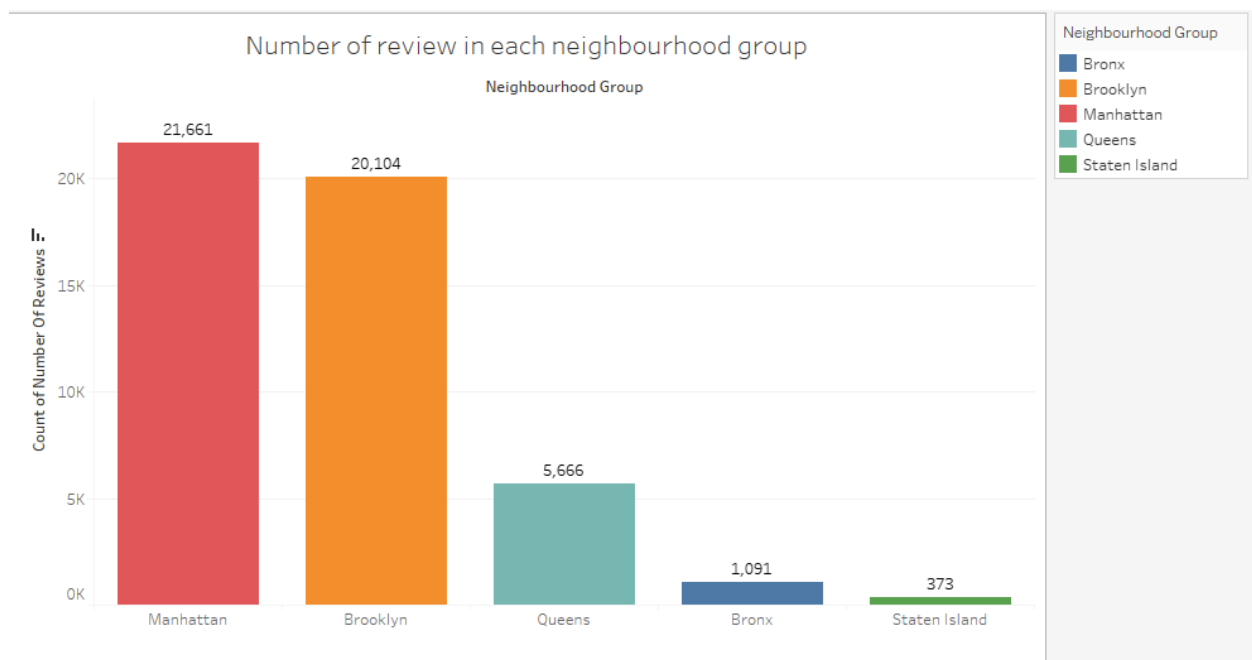
The mean of availability is 112 which indicates the probability is $112/365 \approx 30$ percents of finding room in Airbnb.

✓ Top 10 popular hosts for Airbnb :

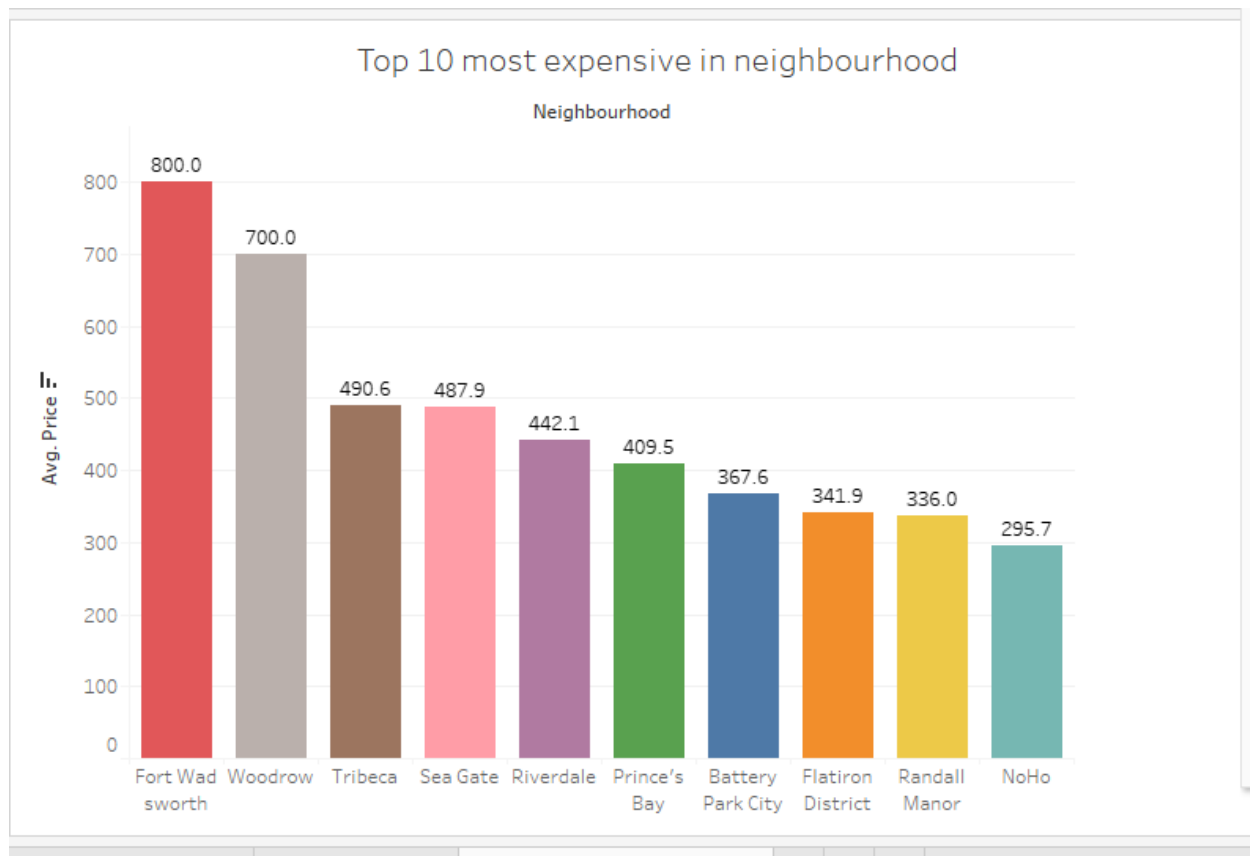


✓ Number of review in neighborhood Group

Manhattan and Brooklyn has most number of reviews.

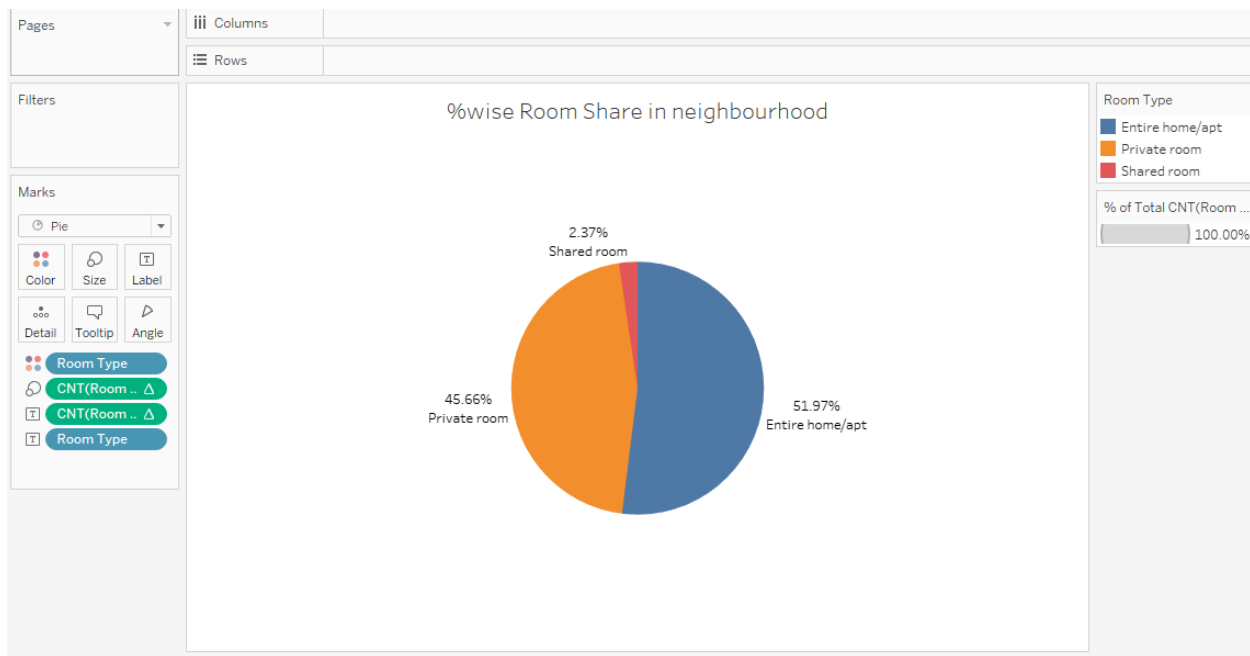
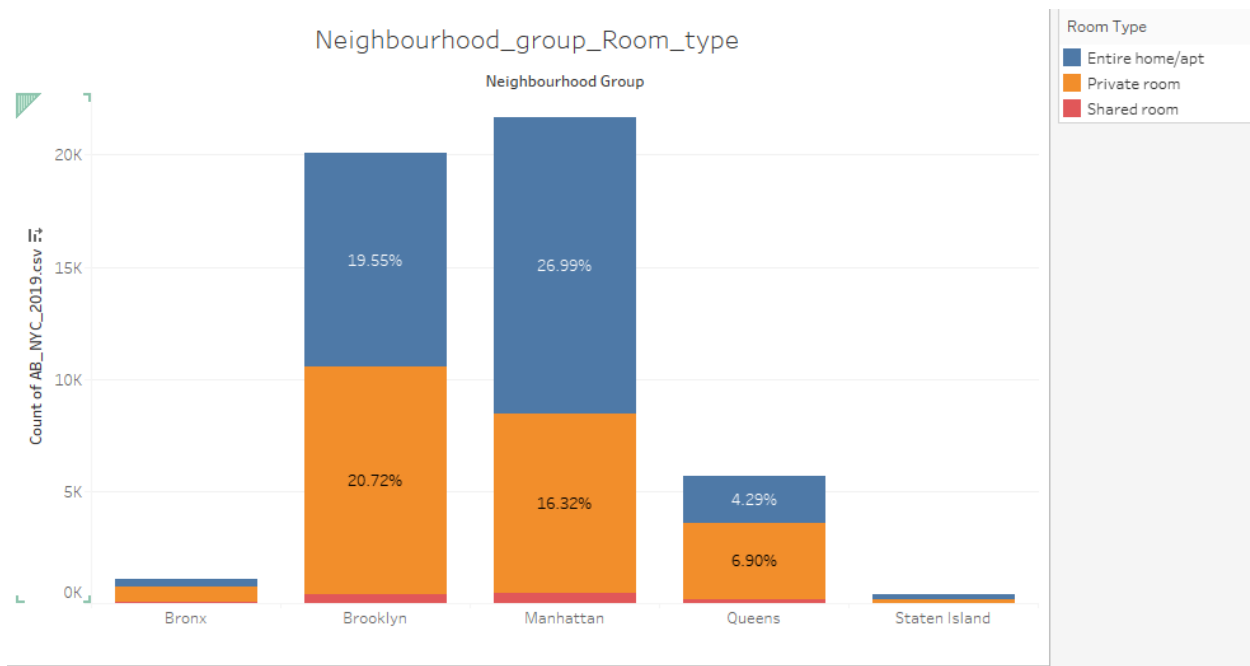


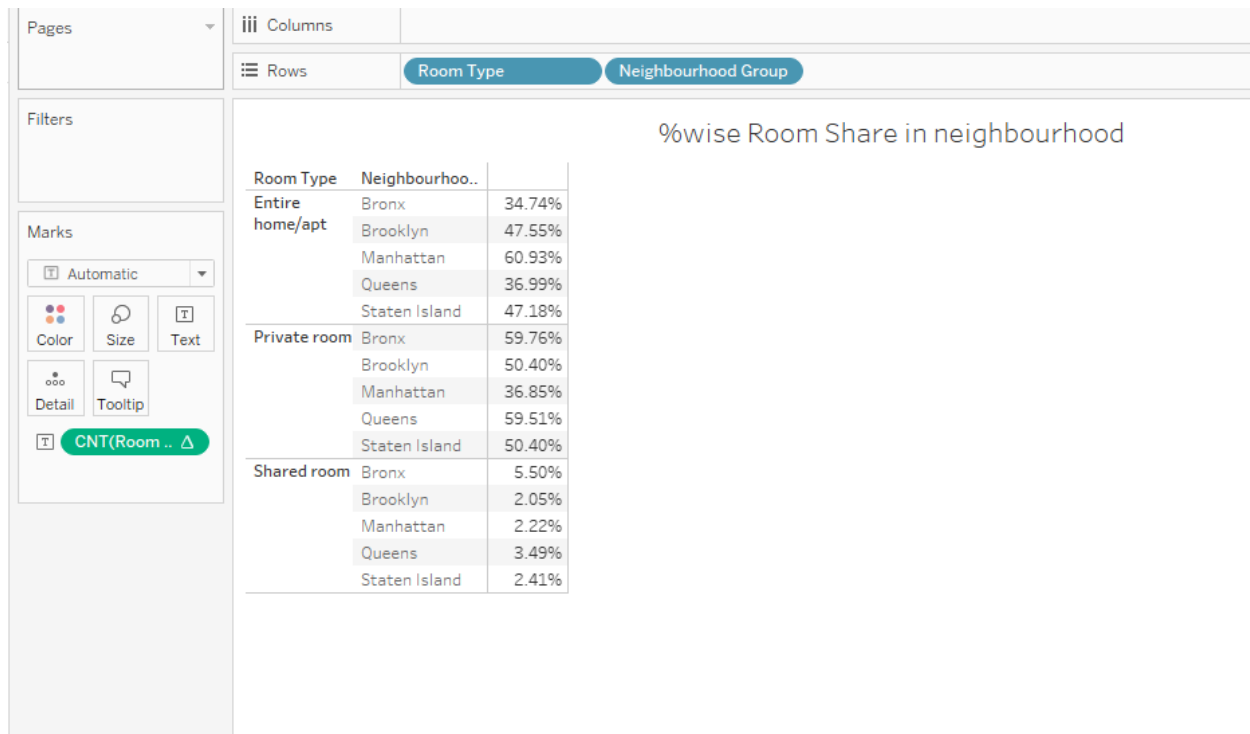
✓ Top 10 most expensive in neighborhood by Average Price



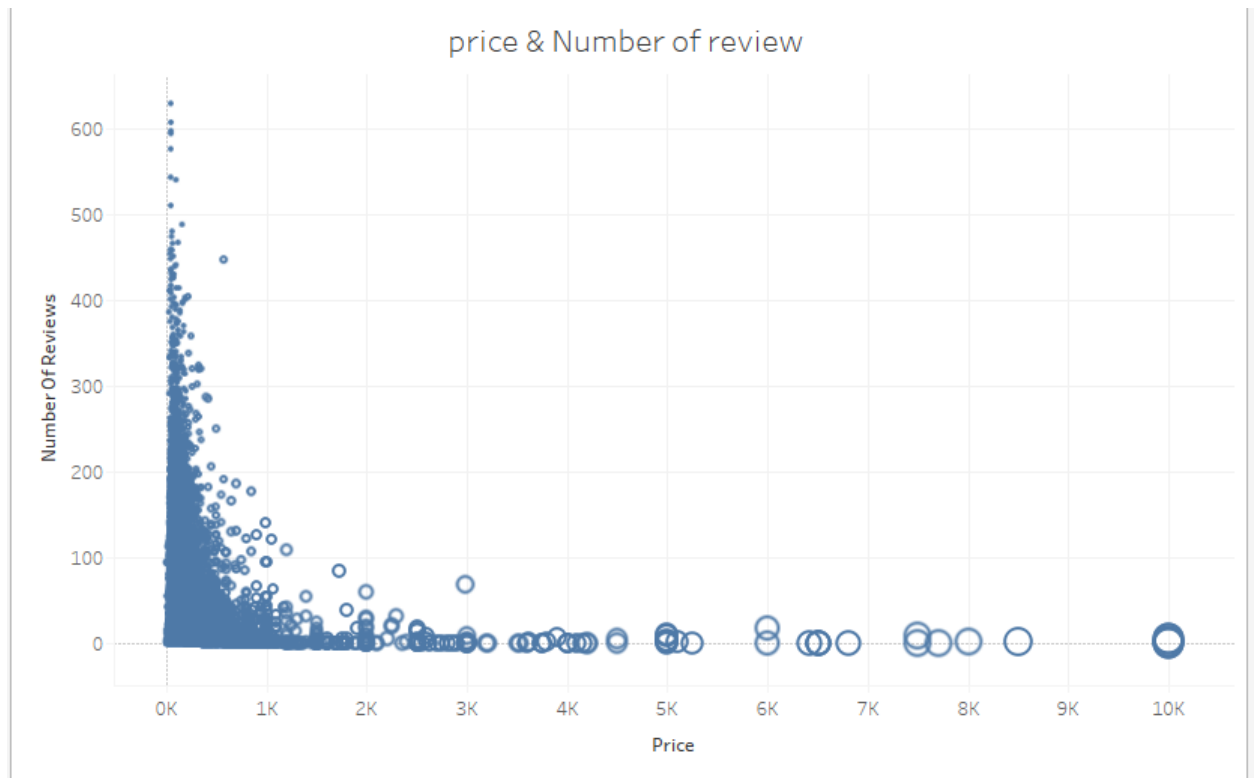
1. Fort Wadsworth is the most expensive; average price around 800.

✓ Percentage of room type in neighborhood Group

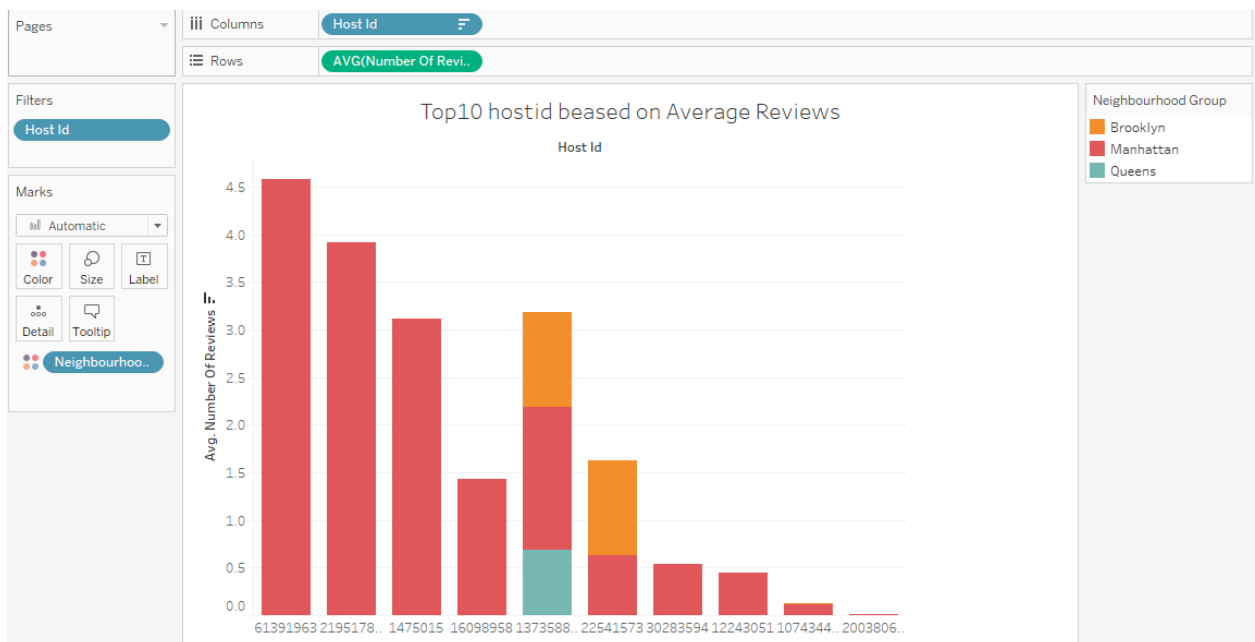




1. As you see from above pie chart home/apt has highest share 51.97 followed by Private room 45.66 and shared room 2.37.
 2. Manhattan has highest percentage in home/apt 60.93 which is around 14 percent higher compare to overall contribution of other home/apt.
 3. Queens has highest percentage in Private rooms 59.51 which is around 23 percent higher compared to other Private rooms players.
- ✓ Price & Number of reviews
1. From below, you can see that graph is left side aligned i.e lesser the price and higher the review it will get.



Top 10 host id based on Average number of reviews:



1. In Top 10 hostid , Manhattan and Brooklyn has maximum number of Share .

Summary: (Key Observation)

- Manhattan and Brooklyn are major hub and attracted maximum number of tourists.
- Percentage of shared room is very less compared to others which denotes people do not mind spending extra money for their comfort.
- In Top 10 hostid , Manhattan and Brooklyn has maximum number of Share .
- Availability is 112 which indicates the probability is $112/365 = 30$ percents of finding room in Airbnb.
- Lesser the price and higher the review it will get.
- Fort Wadsworth is the most expensive; average price around 800.