Problem Statement

✓Airbnb is an American rental online company based in San Francisco , California in America . Airbnb offers services for lodging , stay and help tourist .

✓Airbnb has seen a major decline in revenue due to Pandemic Situation . 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.

Objectives

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.

Going through the Data and Cleaning

✓ Reading the file AB_NYC_2019.csv using pandas.

✓ Check for shape and Size

✓ Check for null values in each column

```
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_revie
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	

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

Going through the Data and Cleaning

✓Two column reviews_per_month and last_review has maximum null values .

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

✓Check for final shape , now it has 48895 rows and 12 columns .

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)
```

Data Analysis

✓There are three room type –home/apt , Private room & Shared room .

✓Highest share of home/apt -52% followed by Private room 45.7% and Shared room -2.37.

✓home/apt has highest share 51.97 followed by Private room 45.66 and shared room 2.37.

✓Manhattan has highest percentage in home/apt 60.93 which is around 14 percent higher compare to overall contribution of other home/apt.

✓Queens has highest percentage in Private rooms 59.51 which is around 23 percent higher compared to other Private rooms players.

Legend:
- Entire home/apt
- Private room
- Shared room

Private room 45.7%
Entire home/apt 52%
Shared room 2.37%

Pages
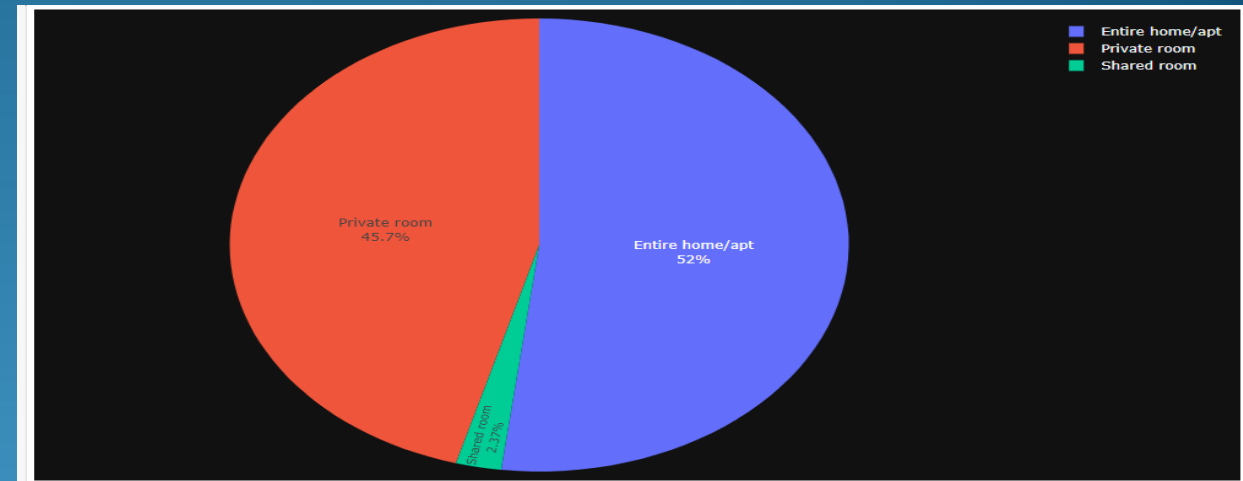
Columns

Rows — Room Type — Neighbourhood Group

Filters

%wise Room Share in neighbourhood

Marks

Automatic

Color | Size | Text

Detail | Tooltip

CNT(Room .. △

Room Type	Neighbourhoo..	
Entire home/apt	Bronx	34.74%
	Brooklyn	47.55%
	Manhattan	60.93%
	Queens	36.99%
	Staten Island	47.18%
Private room	Bronx	59.76%
	Brooklyn	50.40%
	Manhattan	36.85%
	Queens	59.51%
	Staten Island	50.40%
Shared room	Bronx	5.50%
	Brooklyn	2.05%
	Manhattan	2.22%
	Queens	3.49%
	Staten Island	2.41%

Data Analysis

✓Manhattan and Brooklyn are densely populated .

✓Manhattan has 44% followed by Brooklyn 41% has highest populated in neighbourhood i.e. categorized as tourist spot .

✓The mean of availability is 112 which indicates the probability is 112/365 =30 percents of finding room in Airbnb .
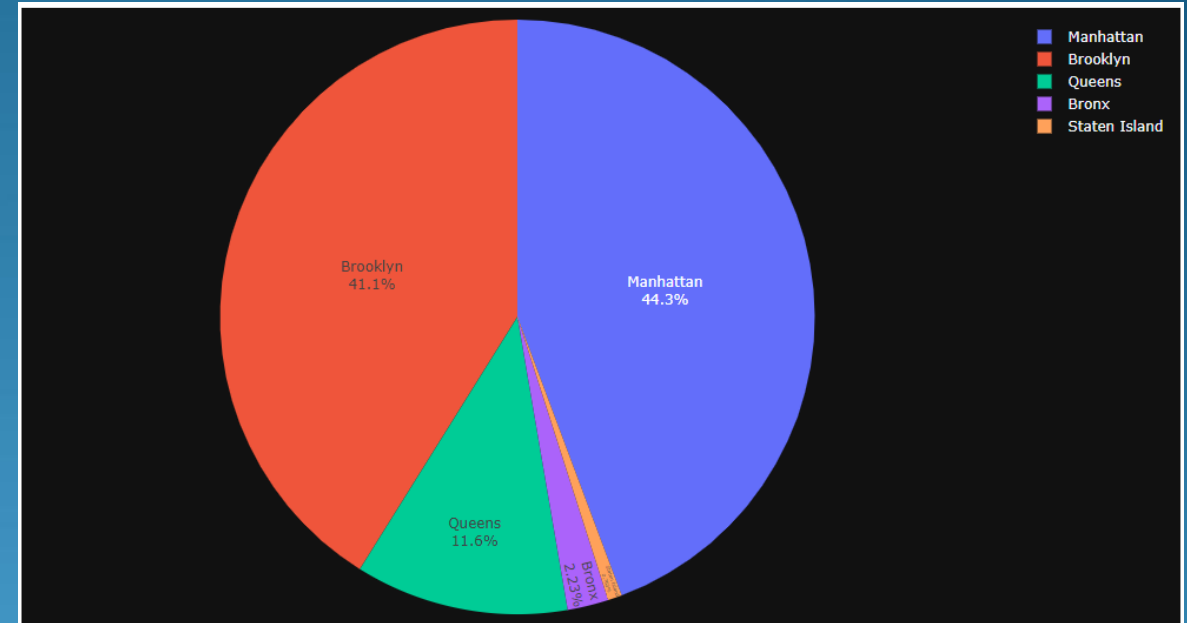

```
count    48895.000000
mean       112.781327
std        131.622289
min          0.000000
5%           0.000000
25%          0.000000
50%         45.000000
75%        227.000000
90%        337.000000
95%        359.000000
99%        365.000000
max        365.000000
Name: availability_365, dtype: float64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

The mean of availability is 112 which indicates the probabilty is 112/365 =30 percents
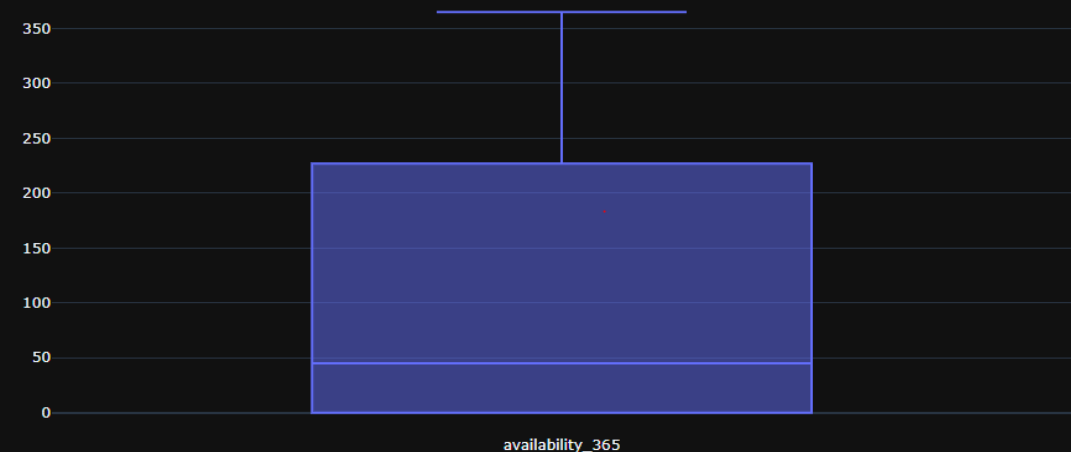
Data Analysis

✓There are listing which offers services from 2 night to 3 year . Most of distribution is in between 1 night to 1 year .

✓If you see column minimum night , it has average 7 days and almost 95 percent distribution is less than 1 month .

✓Most customer plan for month long vacation .
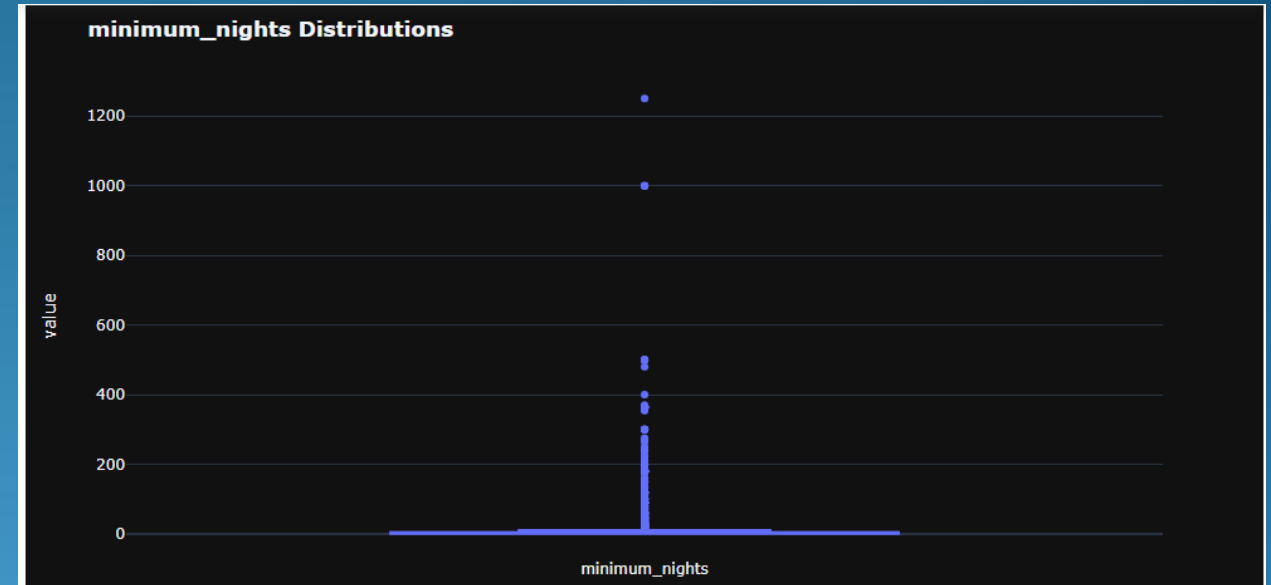

```
count    48895.000000
mean         7.029962
std         20.510550
min          1.000000
5%           1.000000
25%          1.000000
50%          3.000000
75%          5.000000
90%         28.000000
95%         30.000000
99%         45.000000
max       1250.000000
Name: minimum_nights, dtype: float64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

Average Price for room type

✓From the graph , we can see that Home/apt in Manhattan is most expensive one while Standard room in Brooklyn is cheapest .

✓Fort Words is having highest average price-800 among 200 plus neighbourhoods .

Average Price for Different Types rooms in neighbourhood group

Top 10 most expensive in neighbourhood
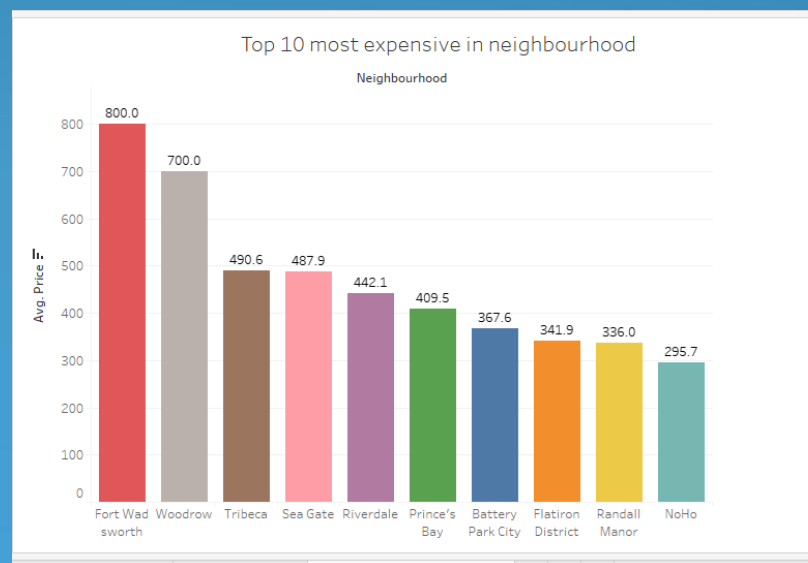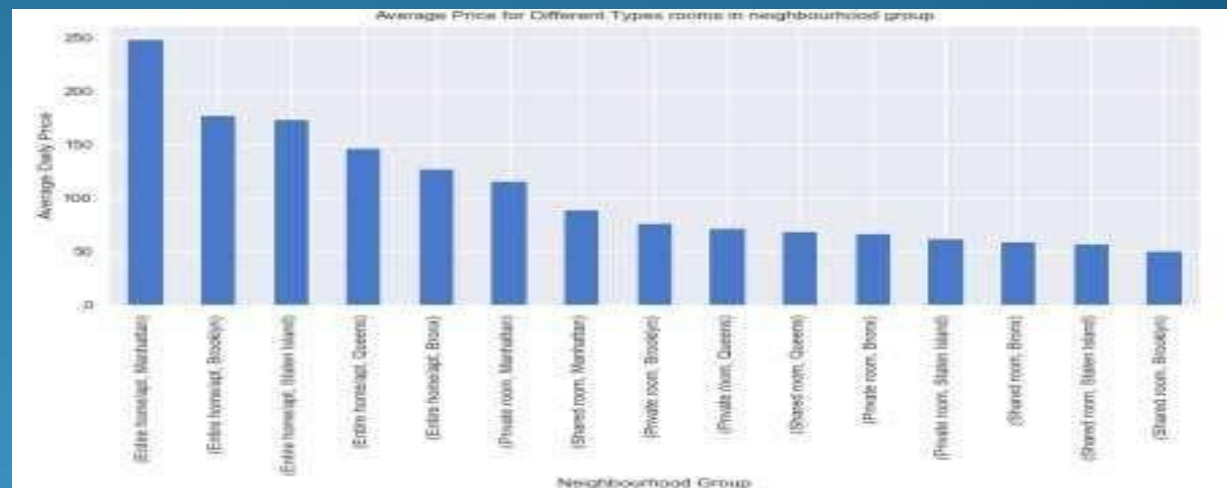
room_type	neighbourhood_group	price
Entire home/apt	Manhattan	249.239109
	Brooklyn	178.327545
	Staten Island	173.846591
	Queens	147.050573
	Bronx	127.506596
Private room	Manhattan	116.776622
Shared room	Manhattan	88.977083
Private room	Brooklyn	76.500099
	Queens	71.762456
Shared room	Queens	69.020202
Private room	Bronx	66.788344
	Staten Island	62.292553
	Bronx	59.800000
Shared room	Staten Island	57.444444
	Brooklyn	50.527845

% of room in neighbourhood Group

✓Manhattan has 27 percent room in home/apt followed by 20 % by Brooklyn .

✓Fort Words is having highest average price-800 among 200 plus neighbourhoods .

✓Brooklyn has highest share of Private room followed by 16% Manhattan .

✓Shared room is not having much share in any of neighbourhood group.
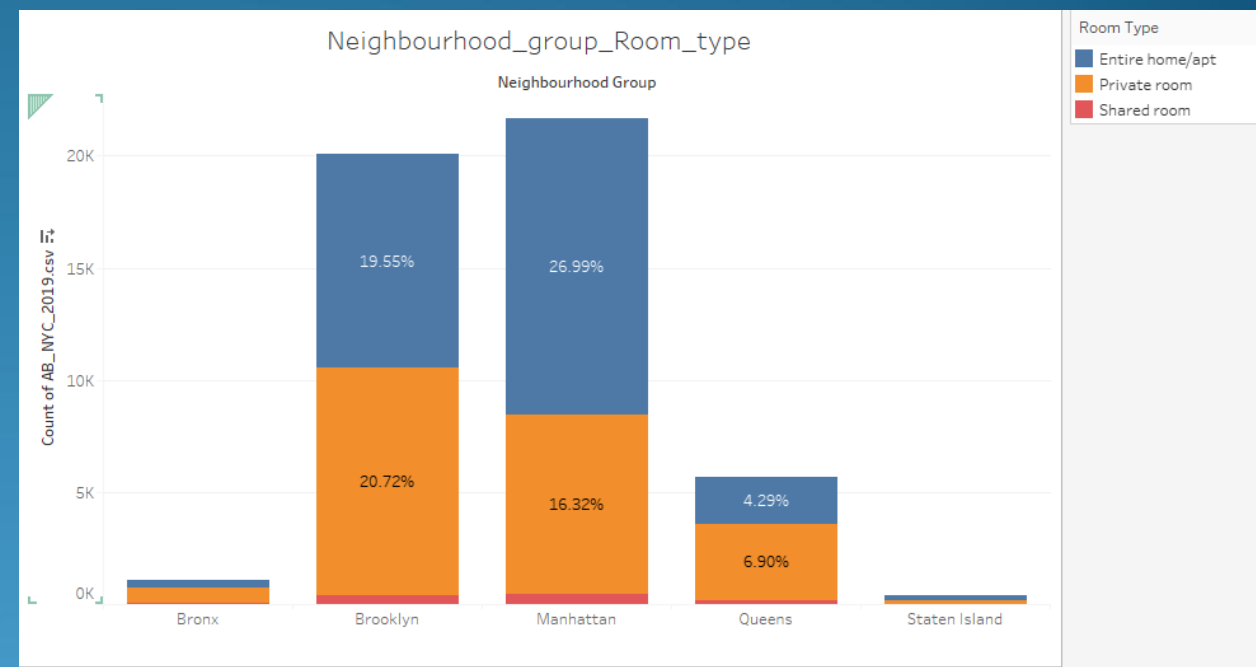
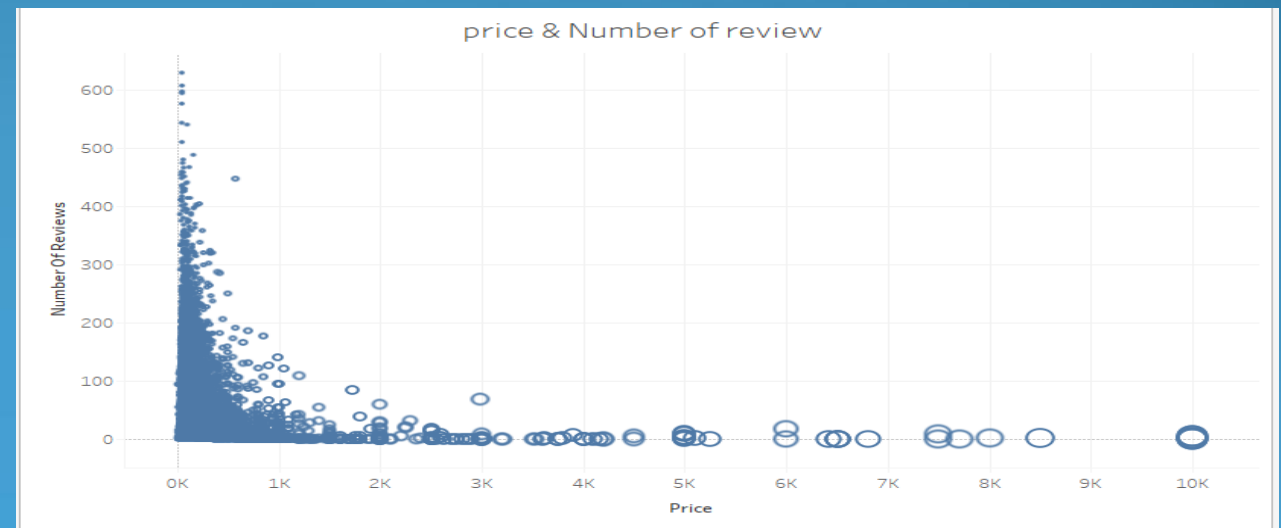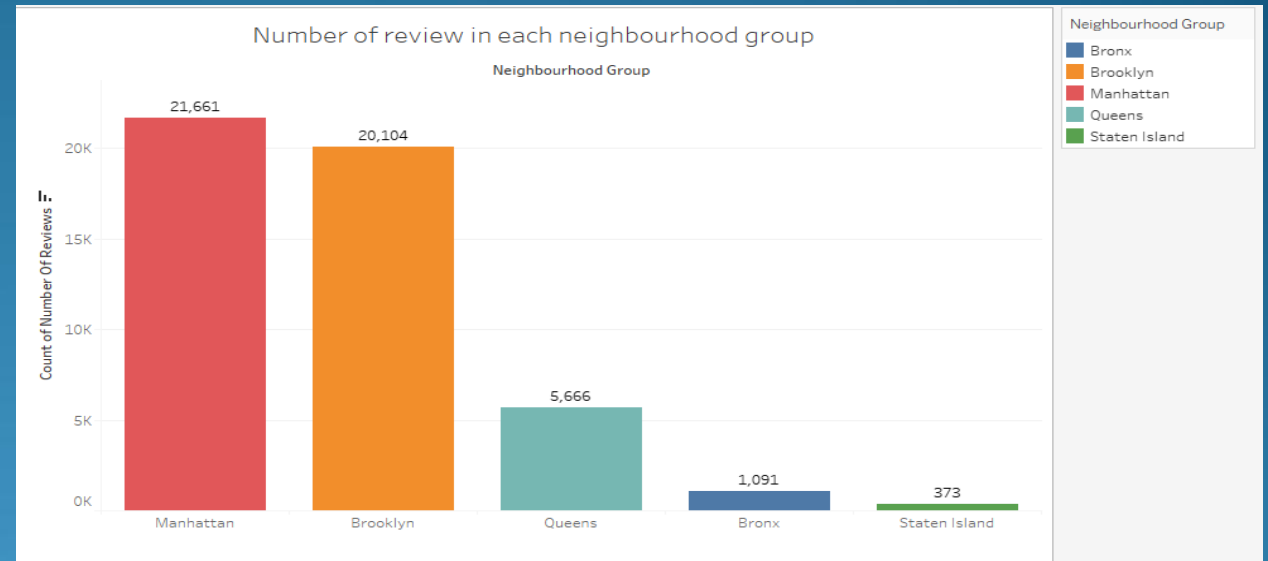# Reviews

✓Manhattan has highest count of reviews followed by Brooklyn .

✓It is observed that where people are staying they do not mind sharing their though and giving reviews .

✓From Graph Average price/reviews we can see that lesser the price and higher the review it will get.

Number of review in each neighbourhood group

price & Number of review

Host id & Avg. Count of reviews & NG

✓On the basis of host-id also, it can be seen that Manhattan has dominated the chart has highest number of avg. reviews .

✓First four host id is from Manhattan and then Brooklyn - it shows dominance of Manhattan as tourist hub and need focus to increase revenue and services .
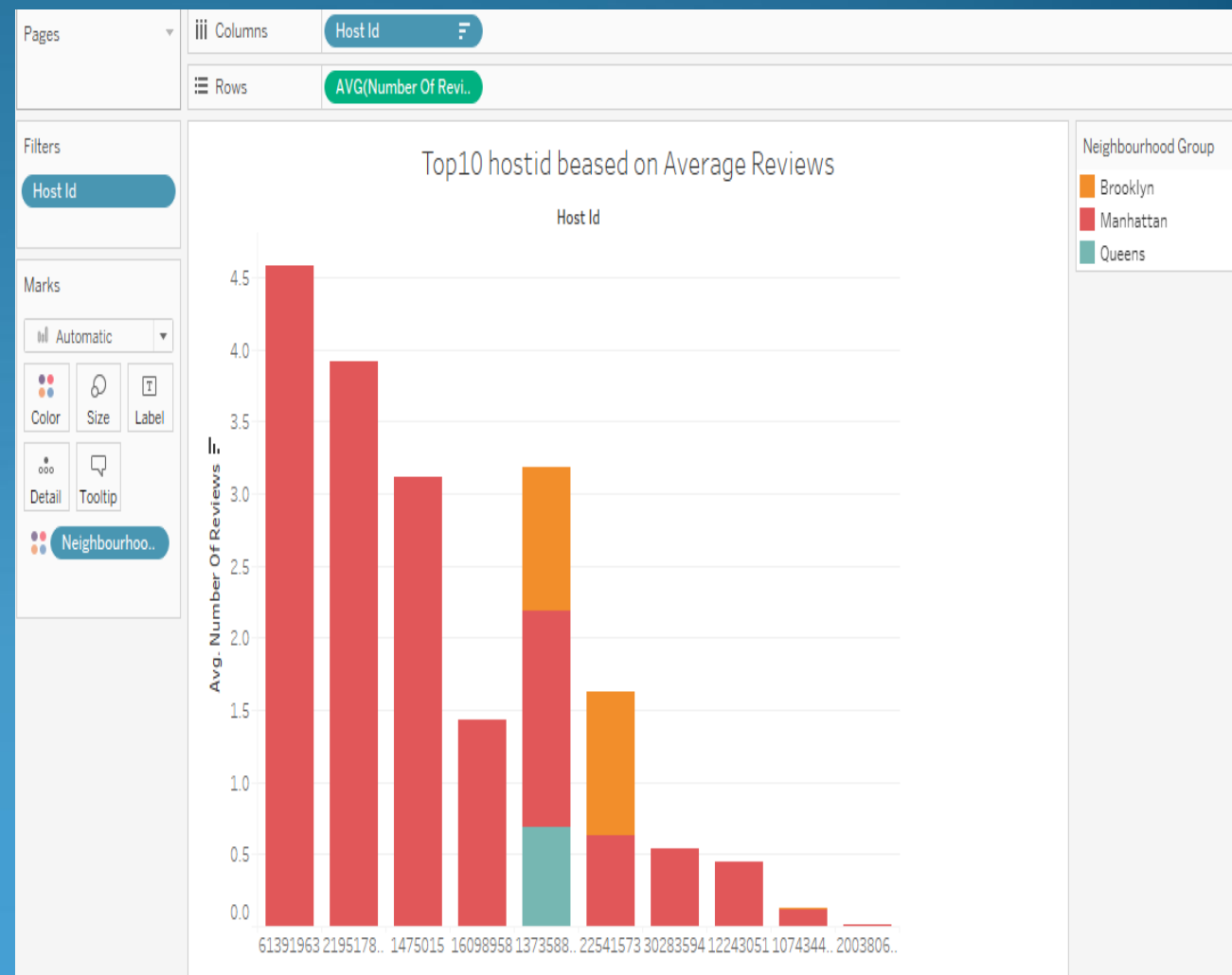
Summary –Recommendations

✓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.
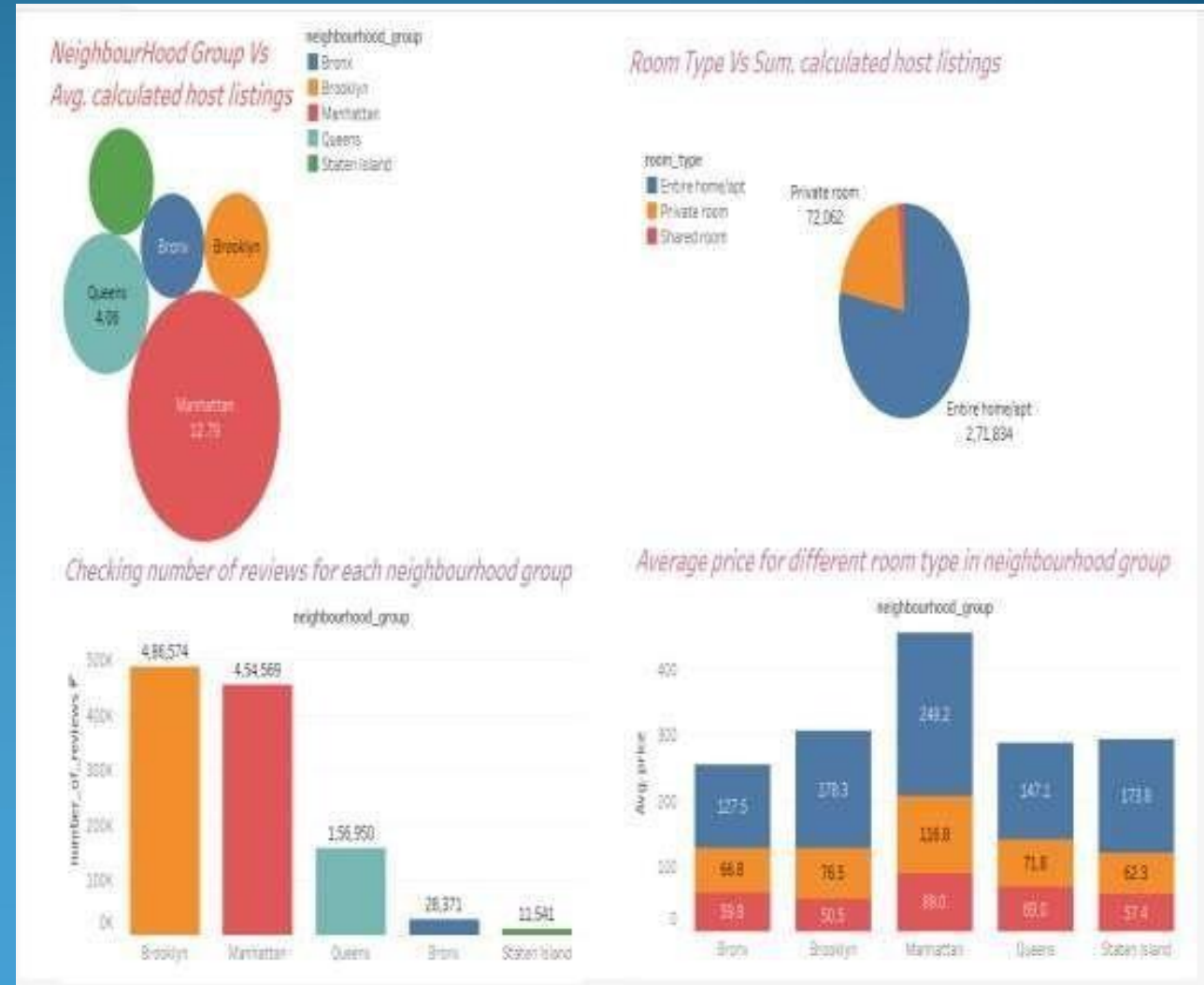
Data Methodology

Data Analysis was done using in below steps :

1. Reading the file using pandas and check for null values in each columns

2. Remove not necessary columns from data set .

3. Replace null with suitable values

4. Performed Univariate and Bi-variate Analysis on existing columns in Data Set .

5. Visualize the relations/data via tool Tableau/Plotly .

6. On the basis of Analysis suggest Plan and identify the hub .