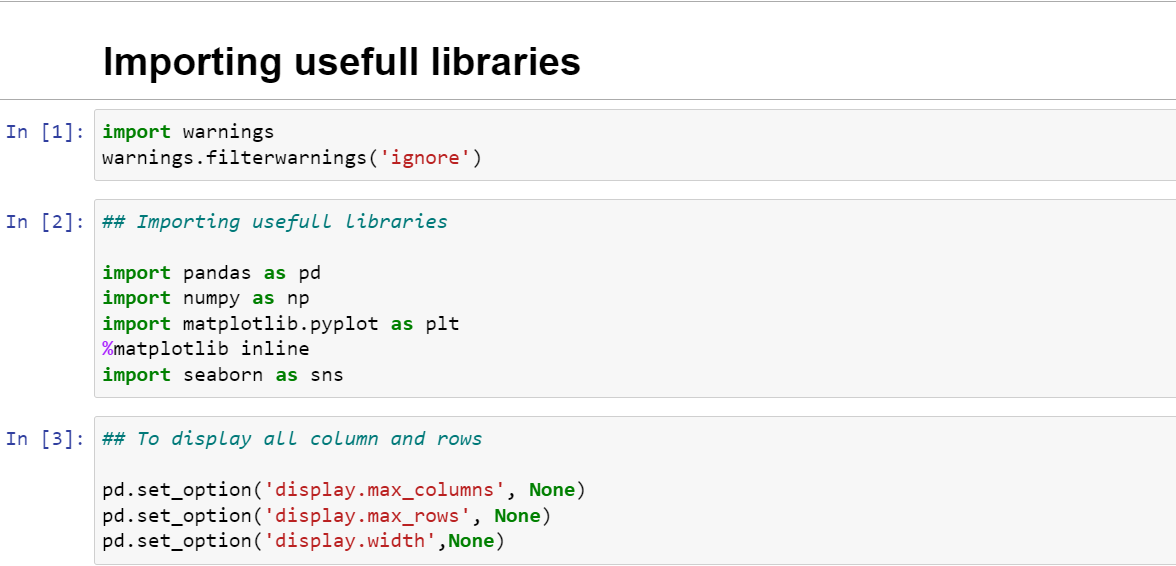
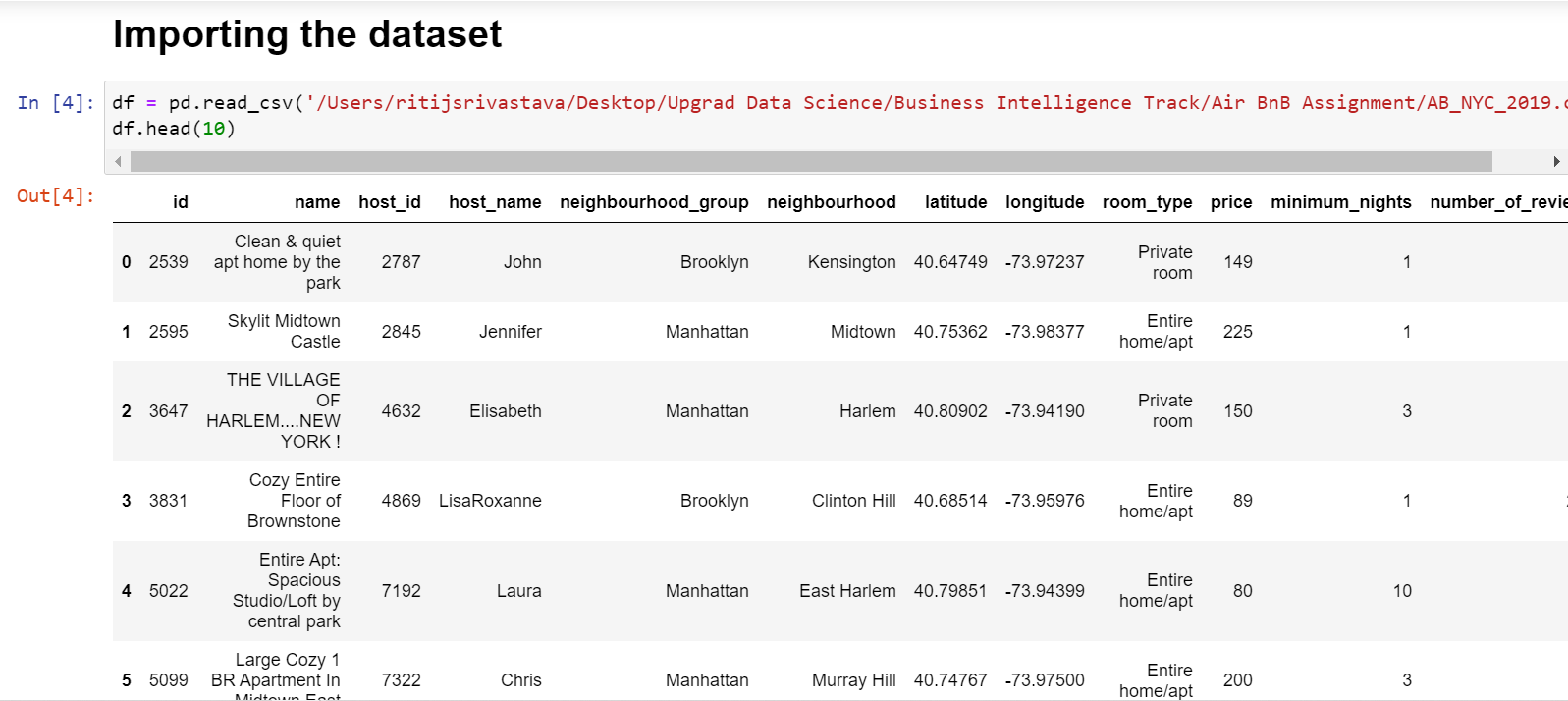
**Airbnb Case Study**

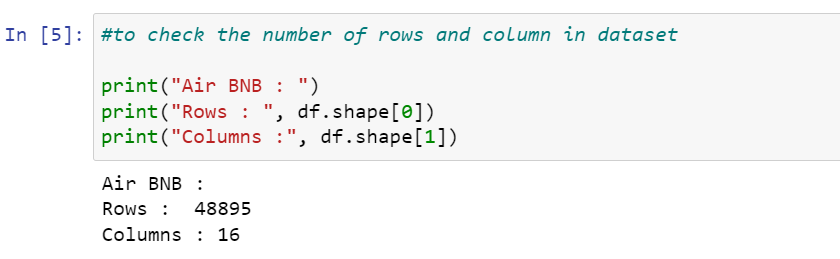
1. Importing the data into Pandas Data frame



1. Importing and Reading the Dataset:

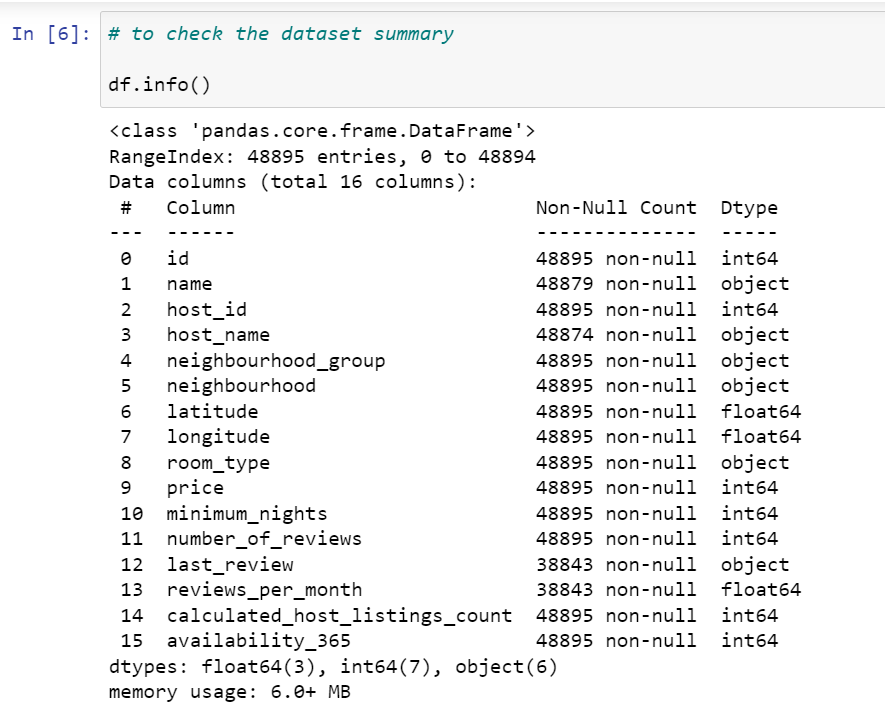


To check the number of rows and columns present in the data.



We can see there are 48895 rows and 16 columns present in the dataset.

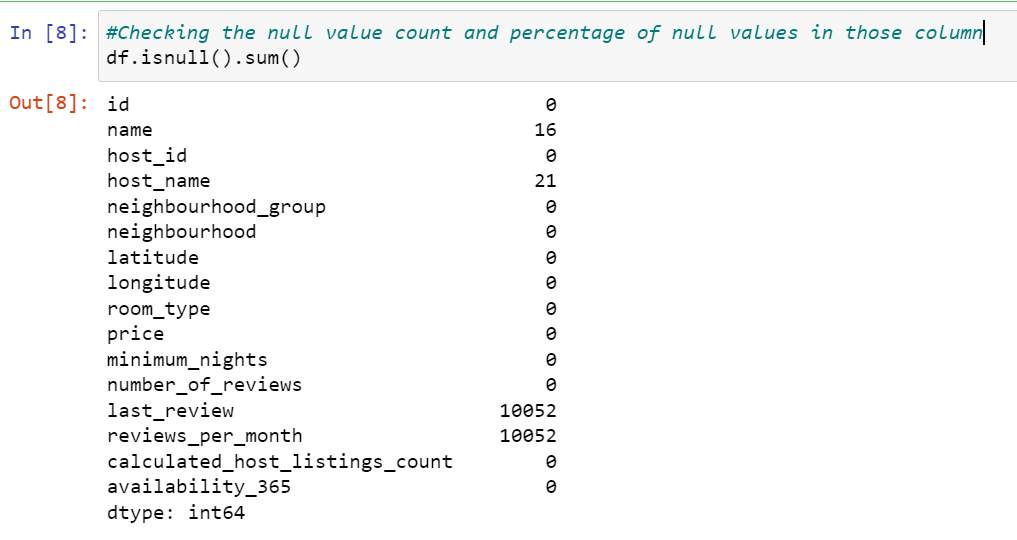
Now checking datatypes and columns names using info().

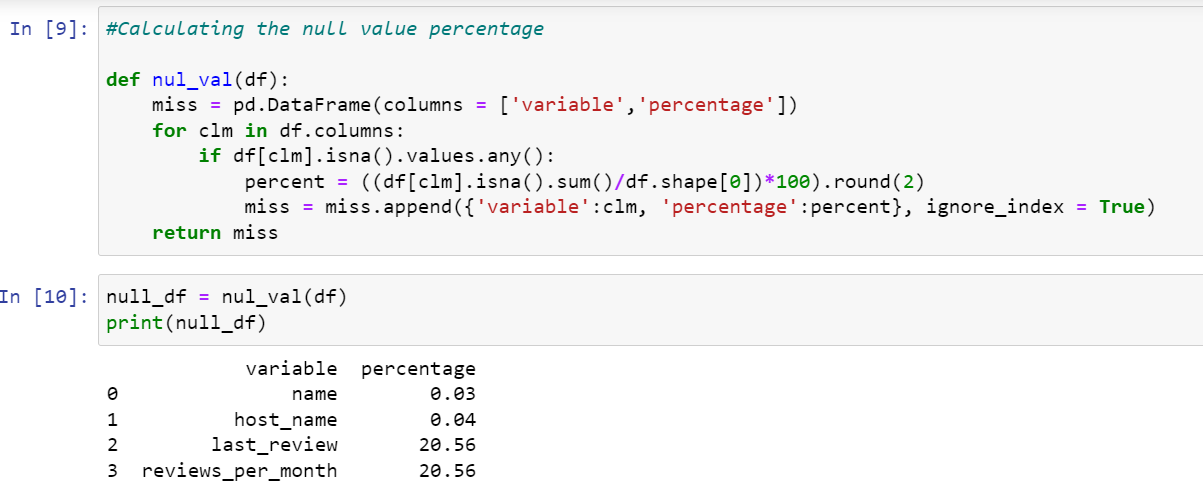


2. Data Cleaning:

- Missing values

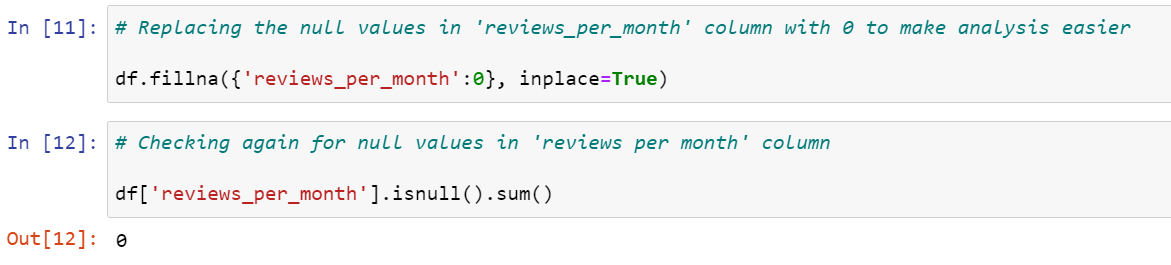
Next checking the null value count and percentage of null values in those column





Missing value treatment:

As "last\_review" and "reviews\_per\_month" column has 20.56% missing values. Since we are only performing data analysis & trying to gain insight hence we will keep them intact.



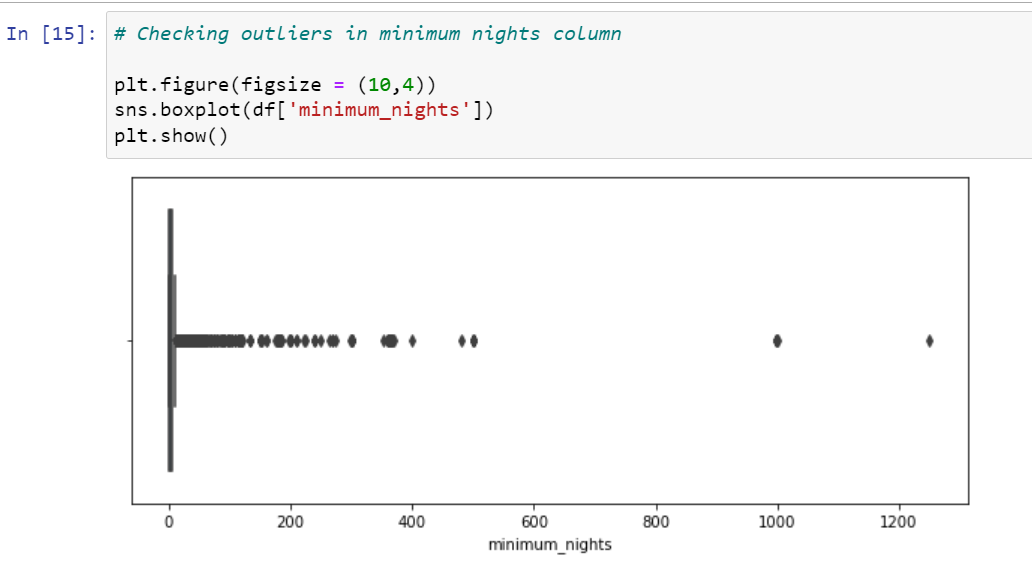
Outlier analysis and treatment:

We perform Outlier Analysis for all the numerical columns and check if required to exclude any outliers.

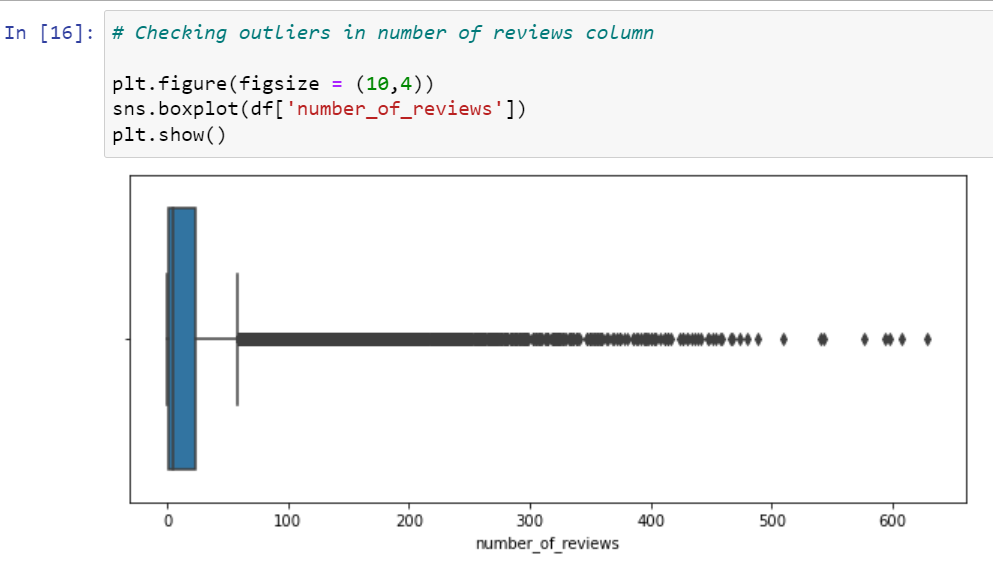
price column:



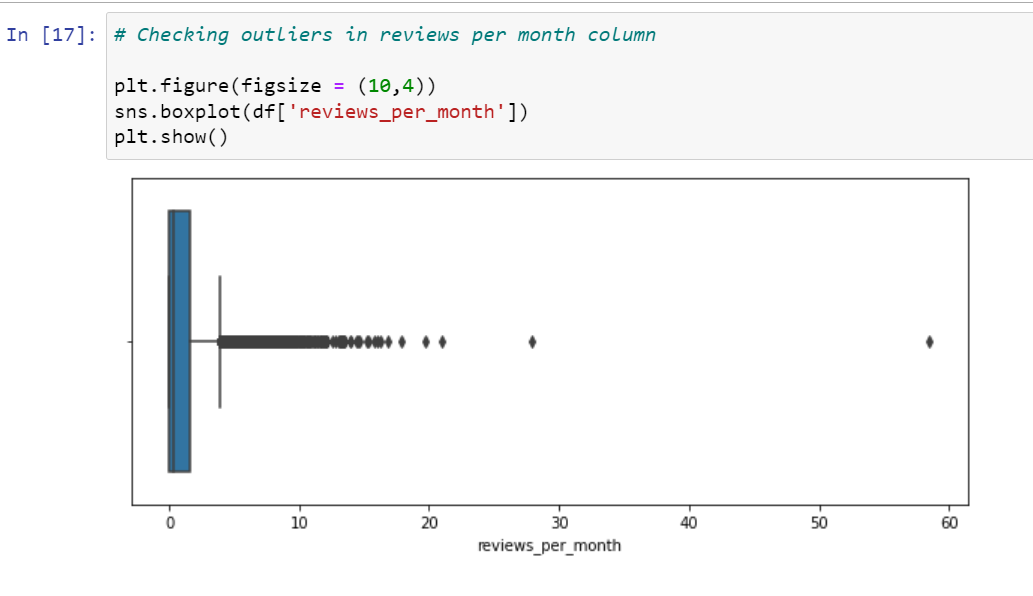
minimum nights:



reviews:



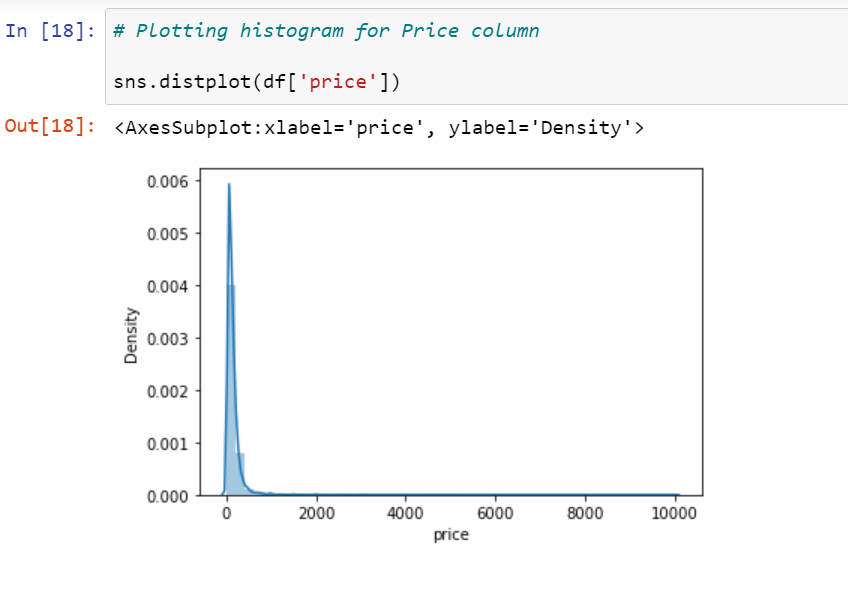
reviews per month:



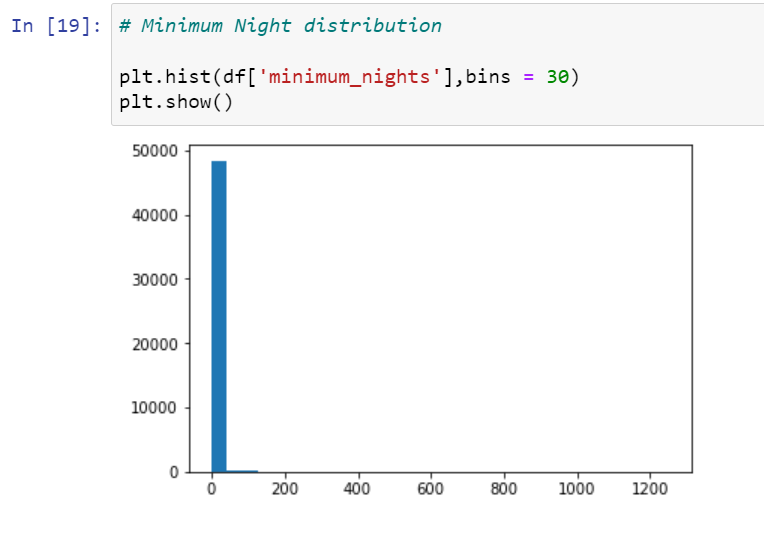
The presence of outliers was expected in number of reviews & reviews per month column as there would be some properties which is quite popular among the visitors and more people prefer to stay at those. Hence, those properties have received more reviews than other. It will be helpful in our data analysis.

3. Exploratory Data Analysis:

Univariate Analysis:



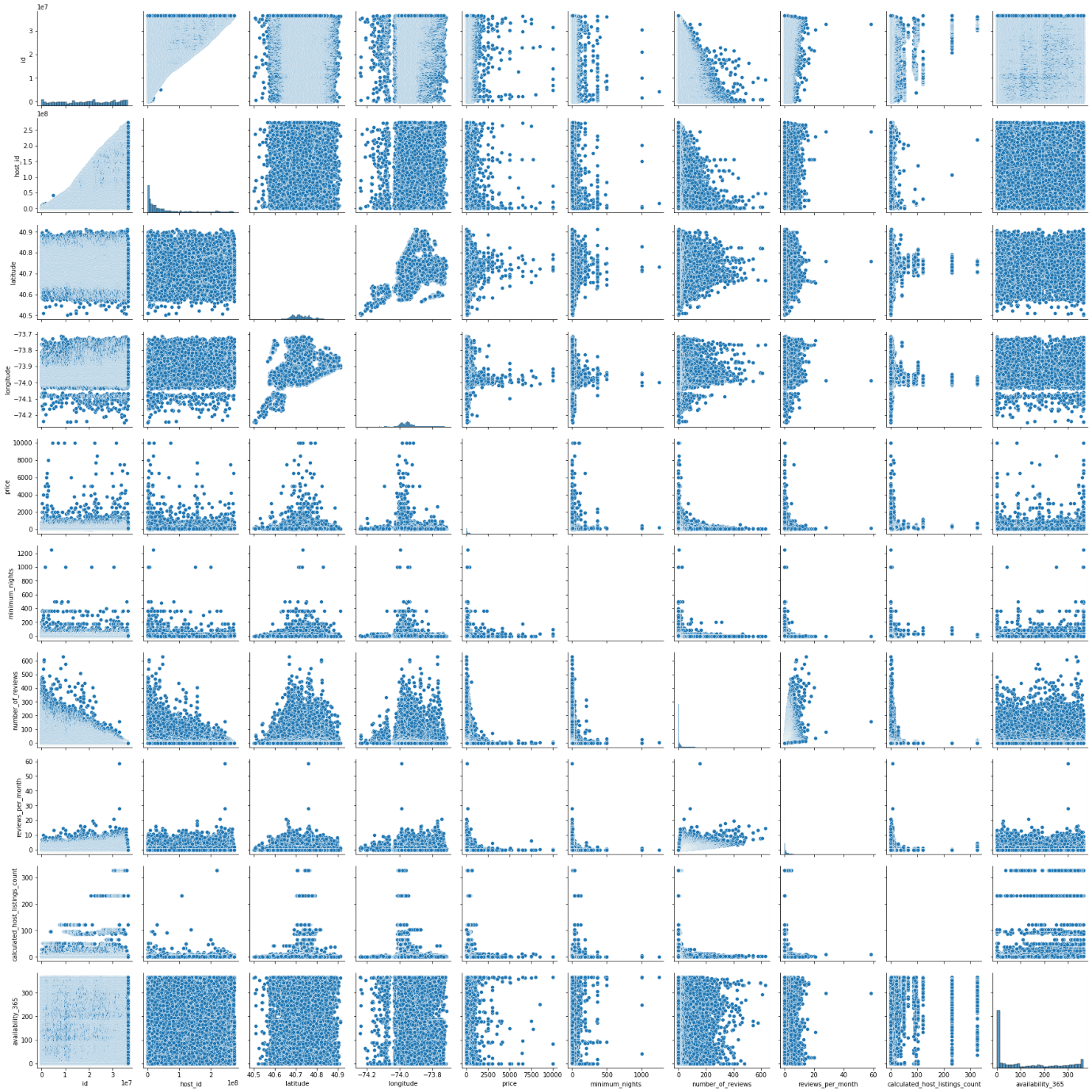
From the above plot we can see most of the property price falls between 0 to 1000, with few properties ranging from 2000 to 10000 which are an outlier.



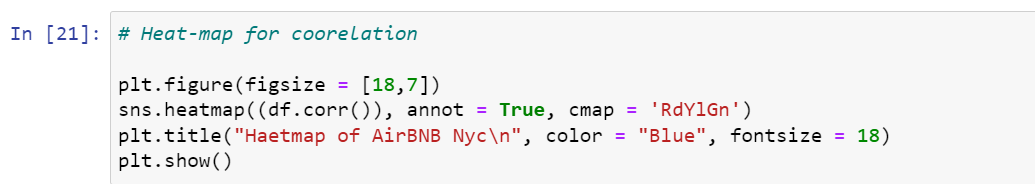
As observed from above plot most of the property offers minimum stay between 0 to 5 days.

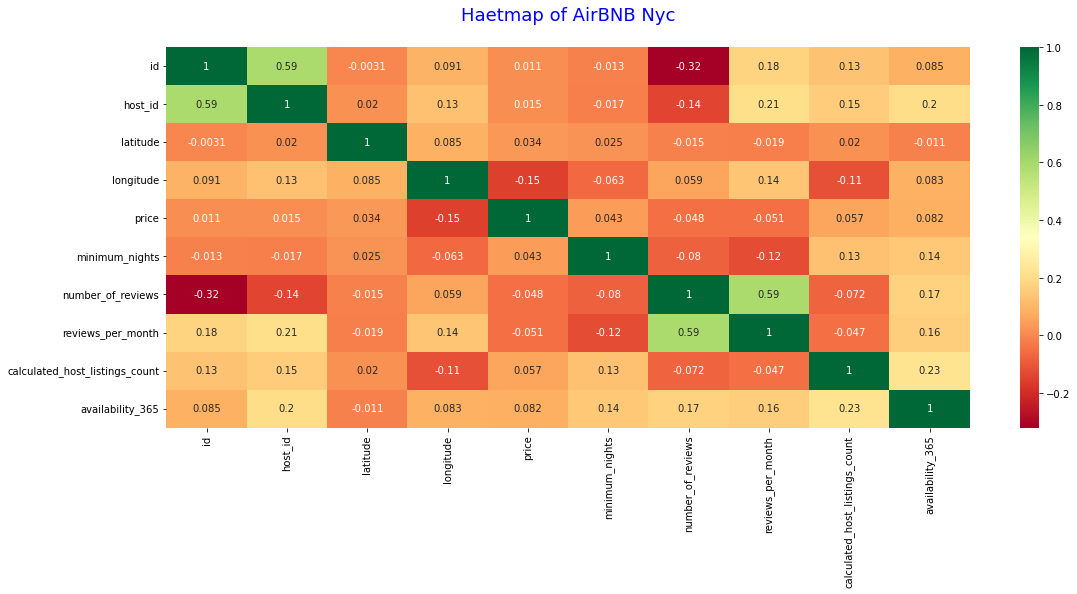
**Bivariate Analysis:**

By plotting the pair plot to understand the correlation between the numerical columns.



**Heatmap of AirBnB data:**

****

****

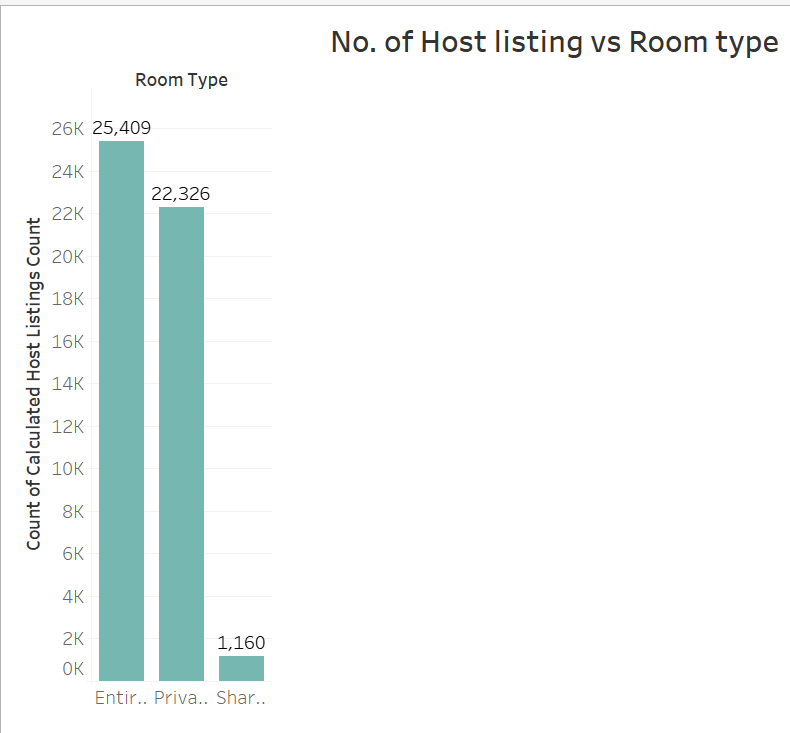
From the pair plot and correlation matrix we can know there is a negative correlation between price, minimum nights and number of reviews. And on the other hand, we can see there is a positive correlation between Calculated\_host\_listings\_count and minimum\_nights & availability \_365 columns.

**Data Visualization and Analysis using Tableau:**

We have used tableau to visualize the data for the assignment.

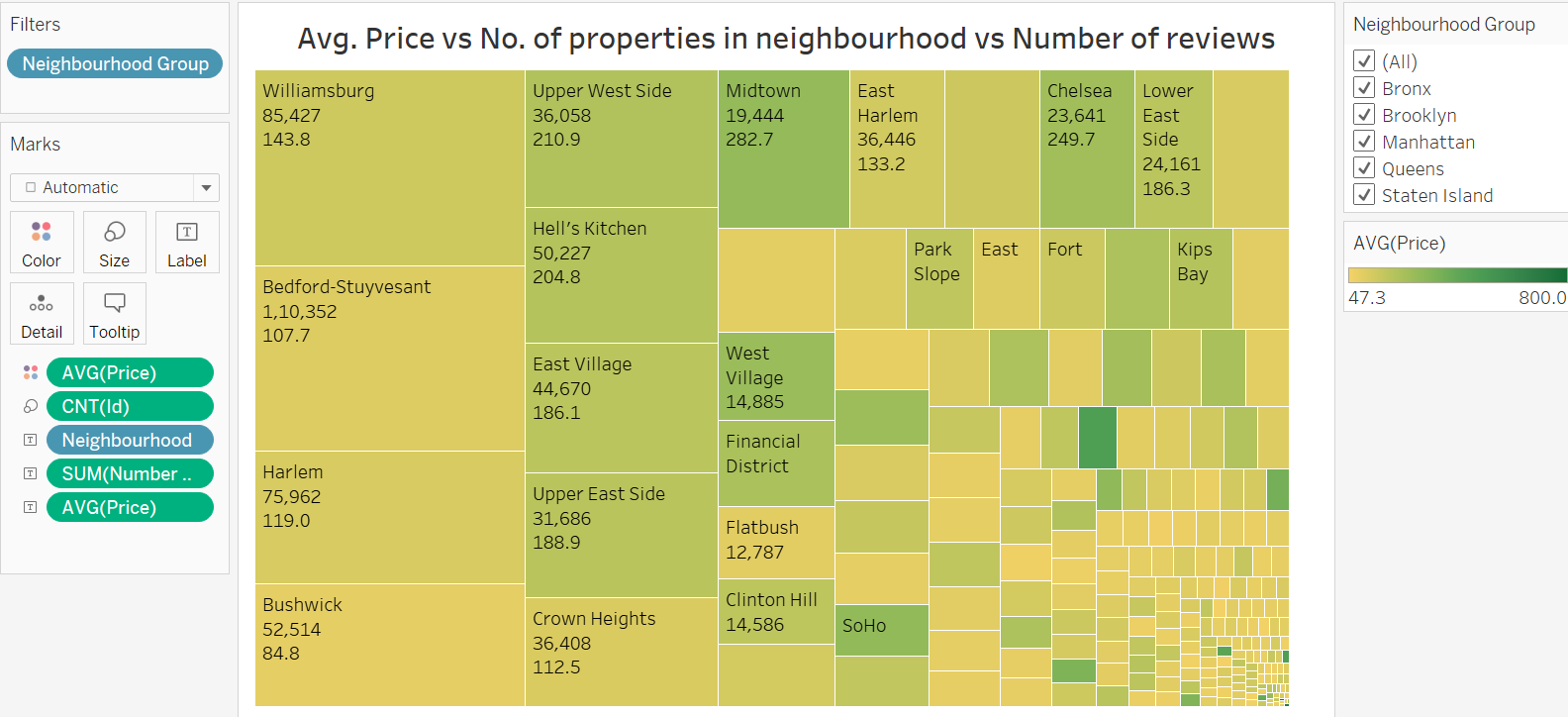
We will use Tableau for Data Visualization and Analysis to come up with Insights and observations. Recommendations are made from the insights and observations drawn from the Analysis.

Properties based on Neighbourhood Group and Room Type.



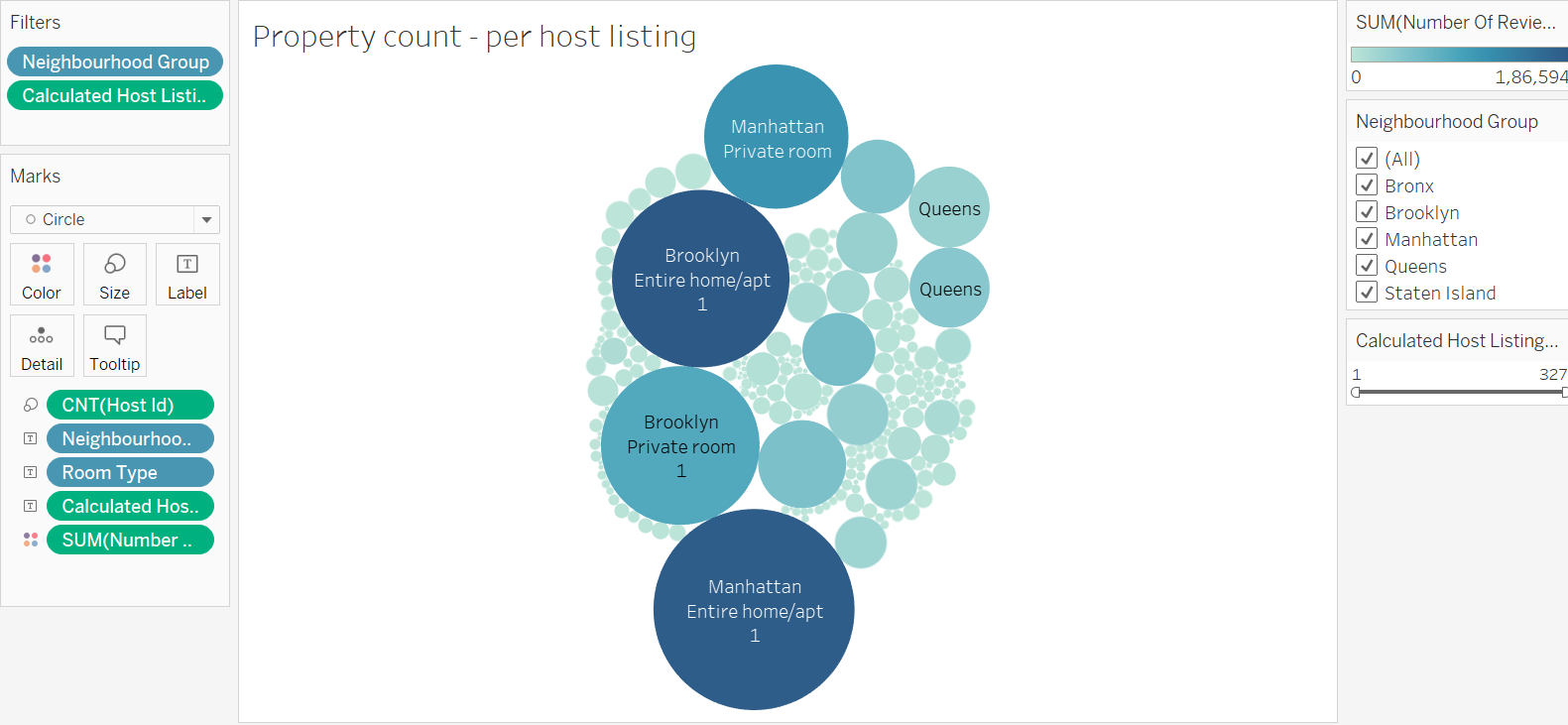
Observation: About 96% of the properties are based on Entire Home/apt. Compared to other apartments few people use Shared room across all the Neighbourhood. Private rooms are

Neighbourhood with Median Price and Number of Reviews



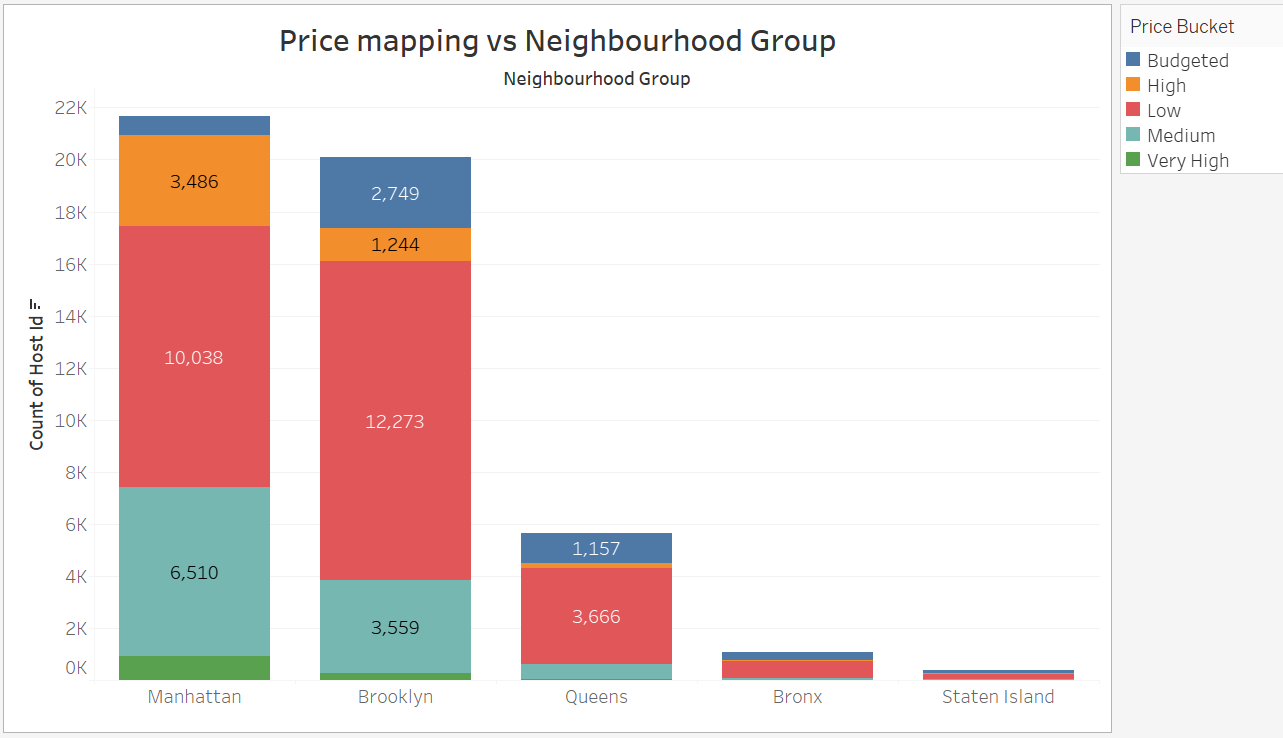
Observation: The general trend is booking is majorly focused on Price. Here number of reviews (booking) is more for lower priced properties. We can see a contrary trend also for some properties with high price got more bookings too. So some people do not mind price if good facilities may be offered.

Property count in neighbourhood



Observation:

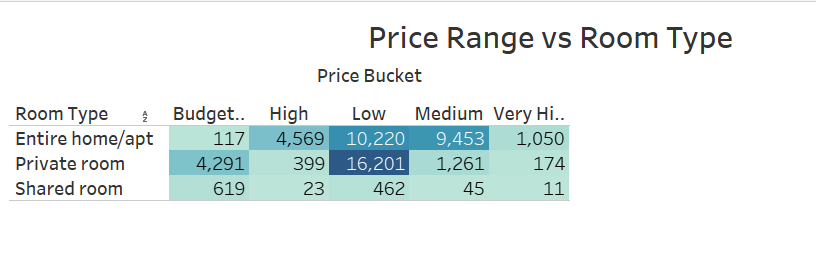
Popular Price Range in Neighbourhood Groups Vs Room Type



Observations: The popular price ranges are from 50 to 250 range. Above 250 , not many booking are seen.

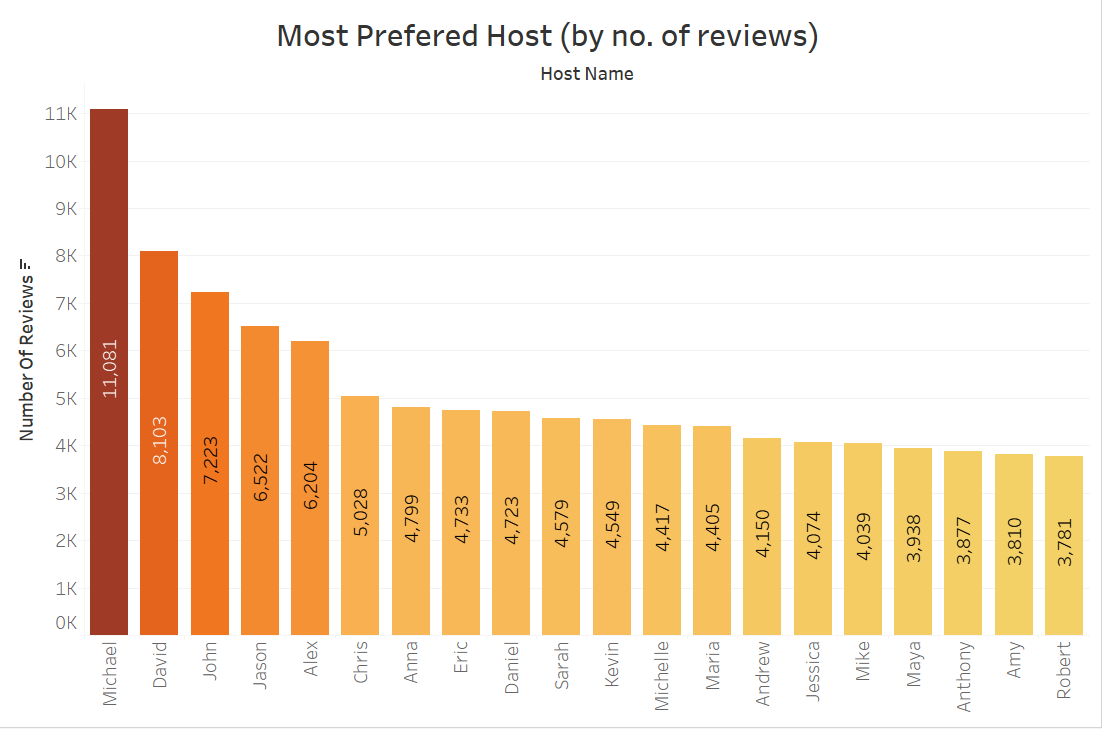
Insight: More booking are towards less priced properties and popular Room Type is Entire Home/Apt and Private Rooms. We can see there are less number of shared rooms only. But people are booking even in shared rooms in relatively lower price. So leveraging Shared rooms also may contribute to higher revenue.

Price spread for different Room Type



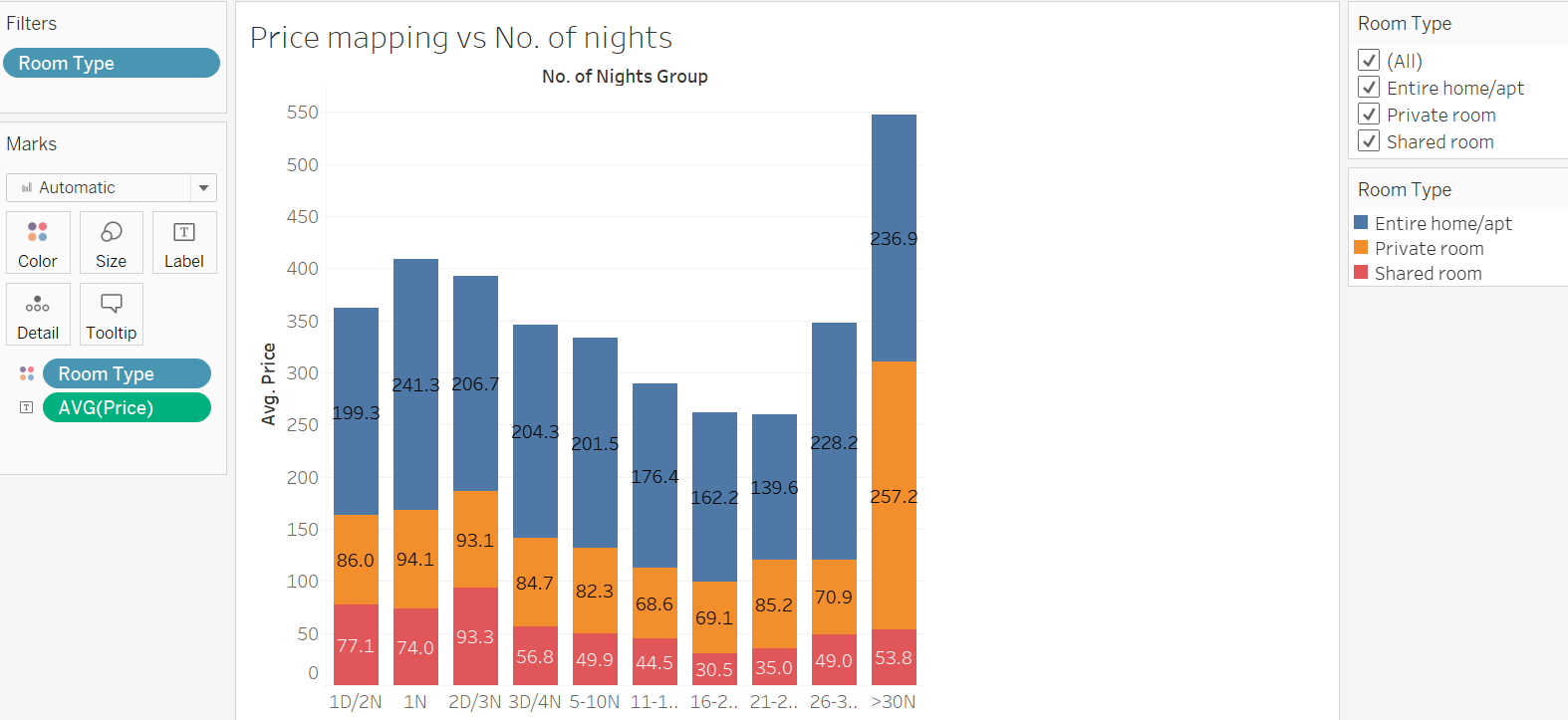
Observation: Entire Home/Apt is costly compared to Private rooms and shared rooms. People prefer mostly Entire Home or Private rooms based on the availability.

Most preferred Host( by no: of reviews)



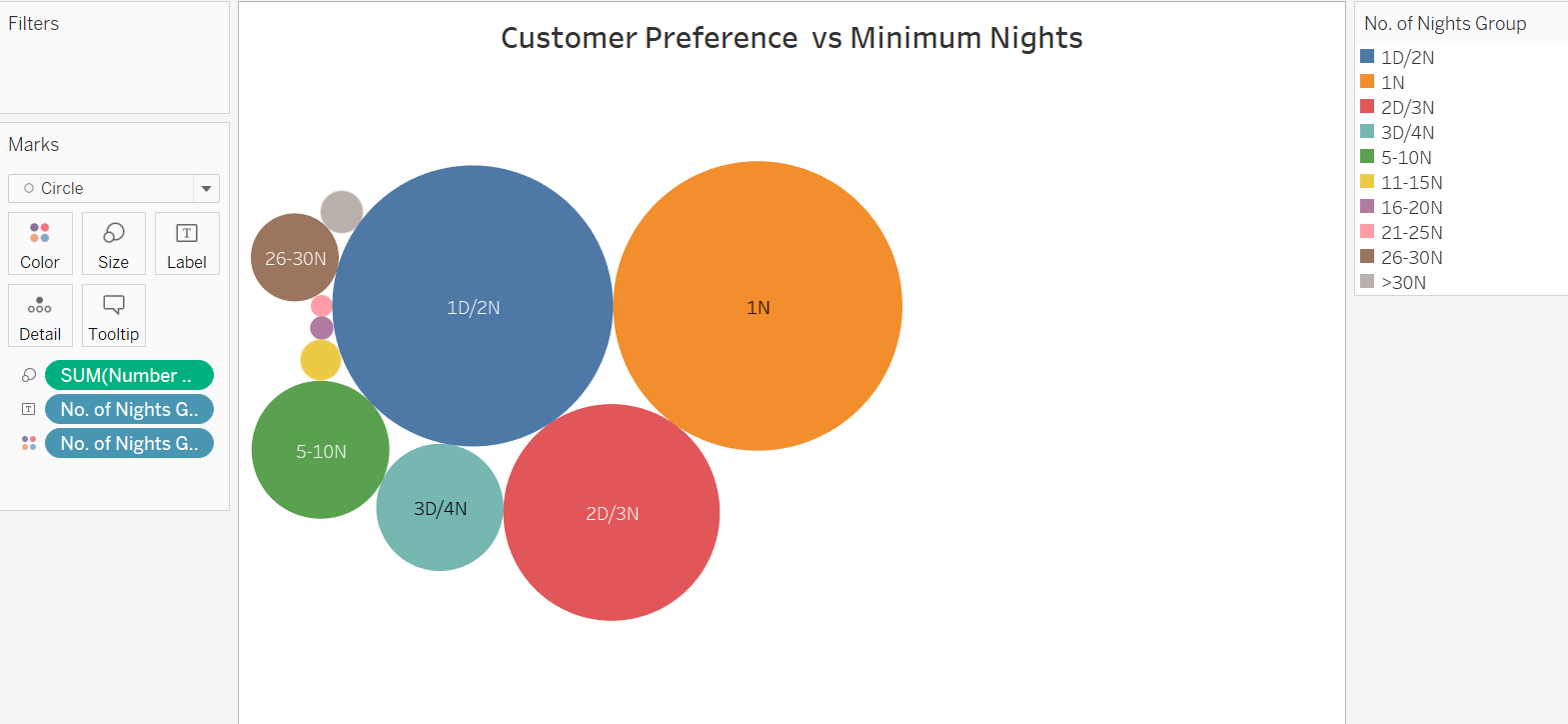
Observation/Insights:Maximum

Minimum Nights vs Revenue.

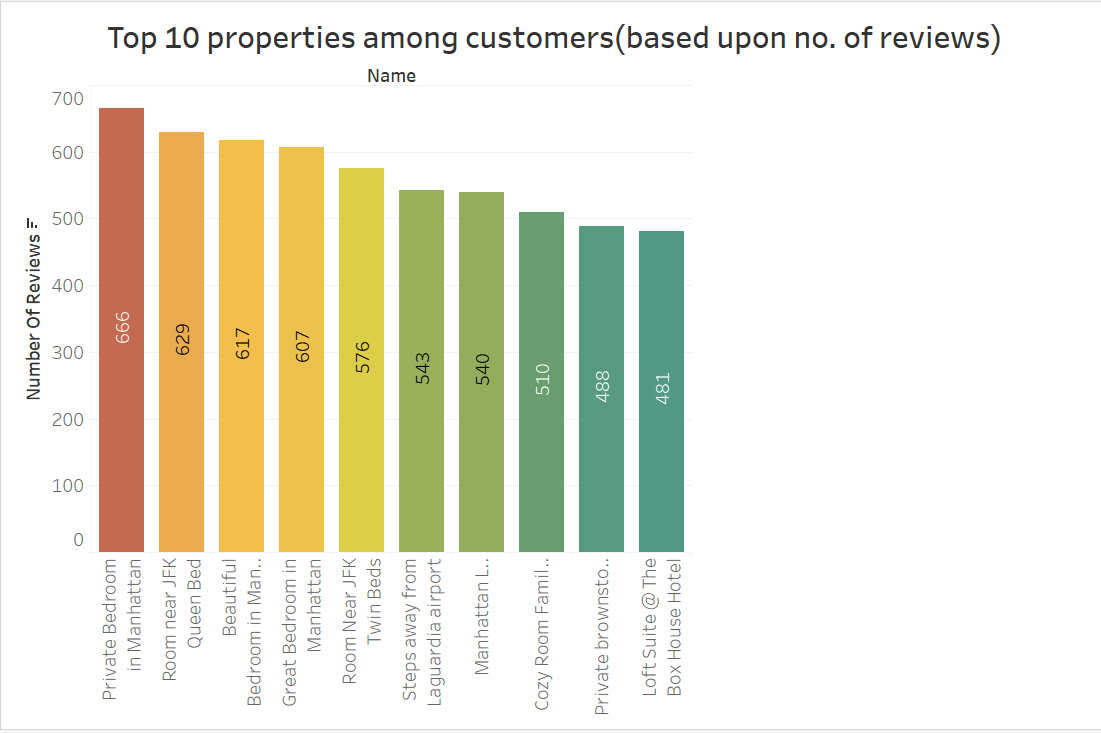


Observation: The hike in the revenue is for Minimum stays 1 to 5 days , mainly 1, 2 and 3 night stay and 30 days. Business can think about the properties having min night stays in these categories across all locations.

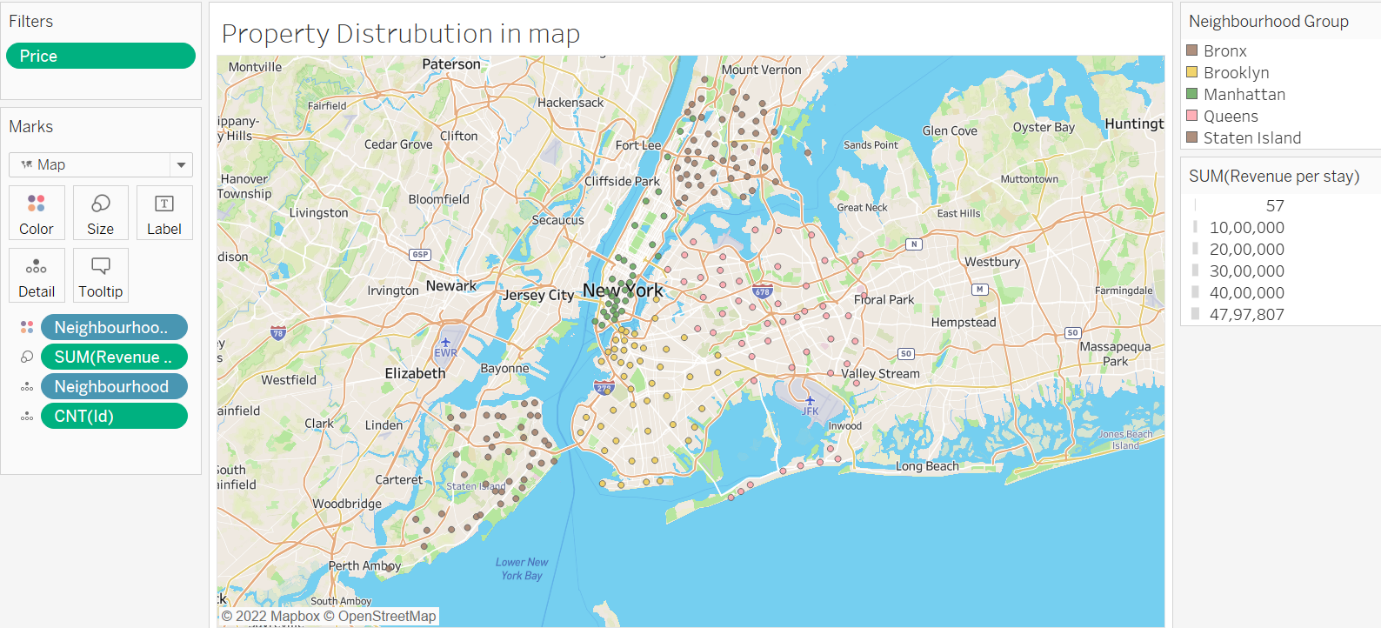
Minimum Nights Distribution



Observation: Maximum Properties are for Minimum night stay 1,2 and 3 days and 30 days for all the neighbourhood groups and Room Type.

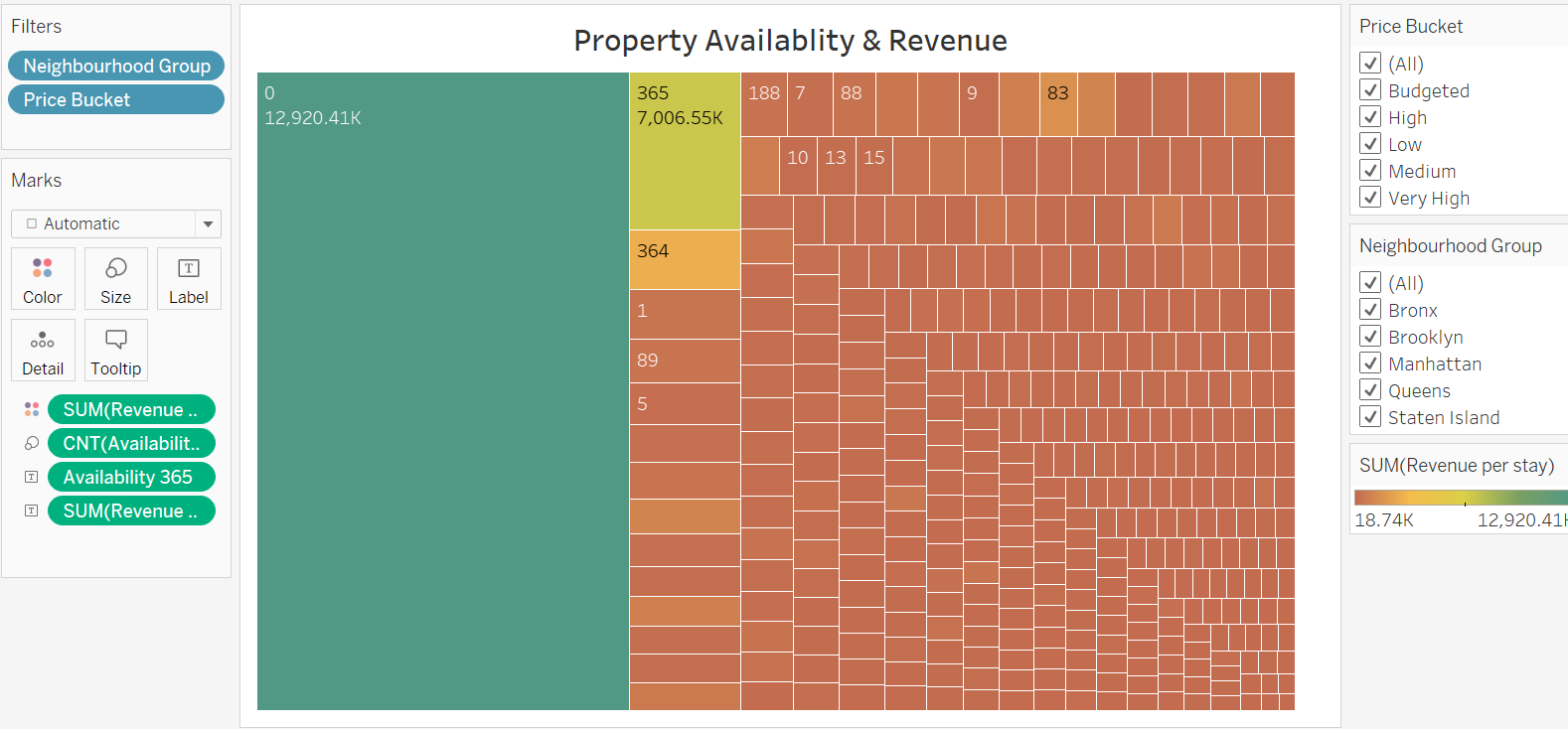


Property distribution in Map



Observation: It is clearly seen that the median price for properties are affordable except few properties. Manhattan and Brooklyn at an average, price is medium to high ranges compared to Bronx and Queens. In Staten Island we see the median price is higher in some areas like Woodrow and FortWadsworth than other regions like Manhattan and Brooklyn.

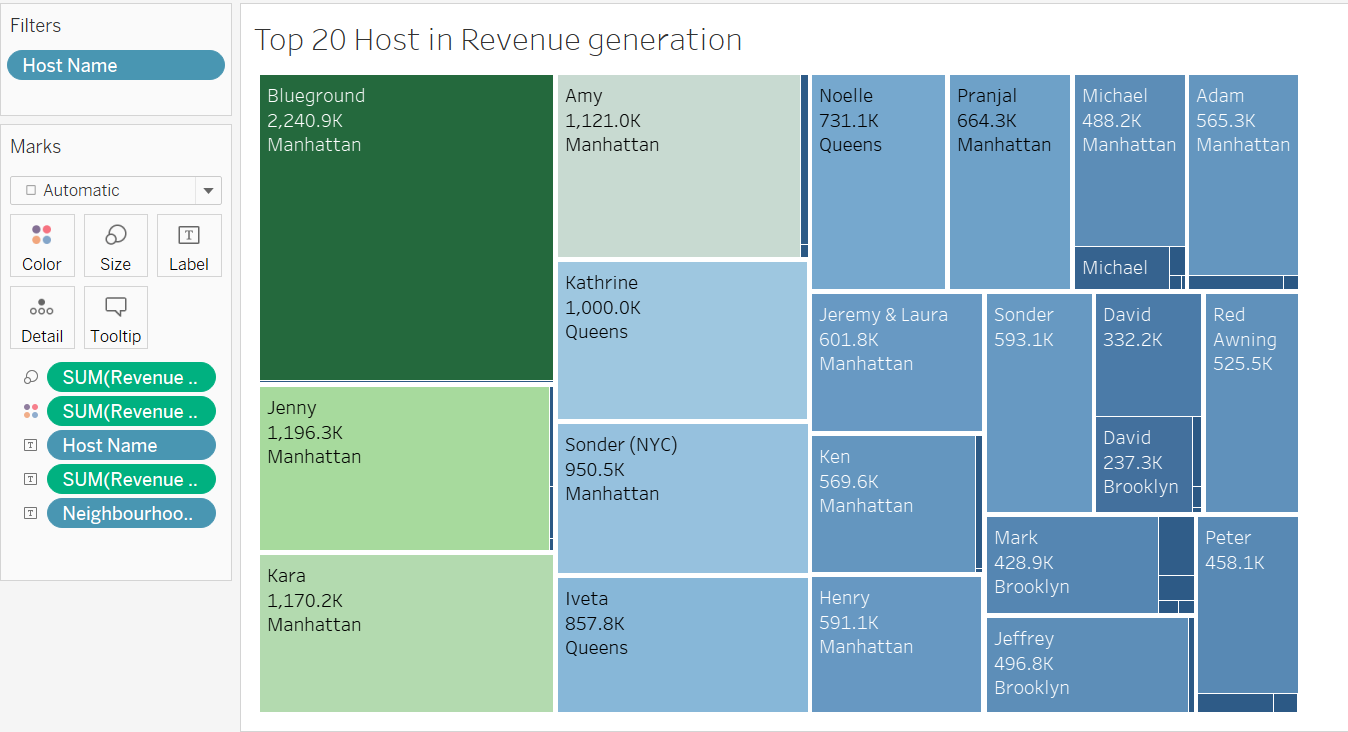
 Property Availibility distribution with revenue



Observation: The figure 1 represents the highest review received is for Availability as “0” compared to other Availability\_365 entries.

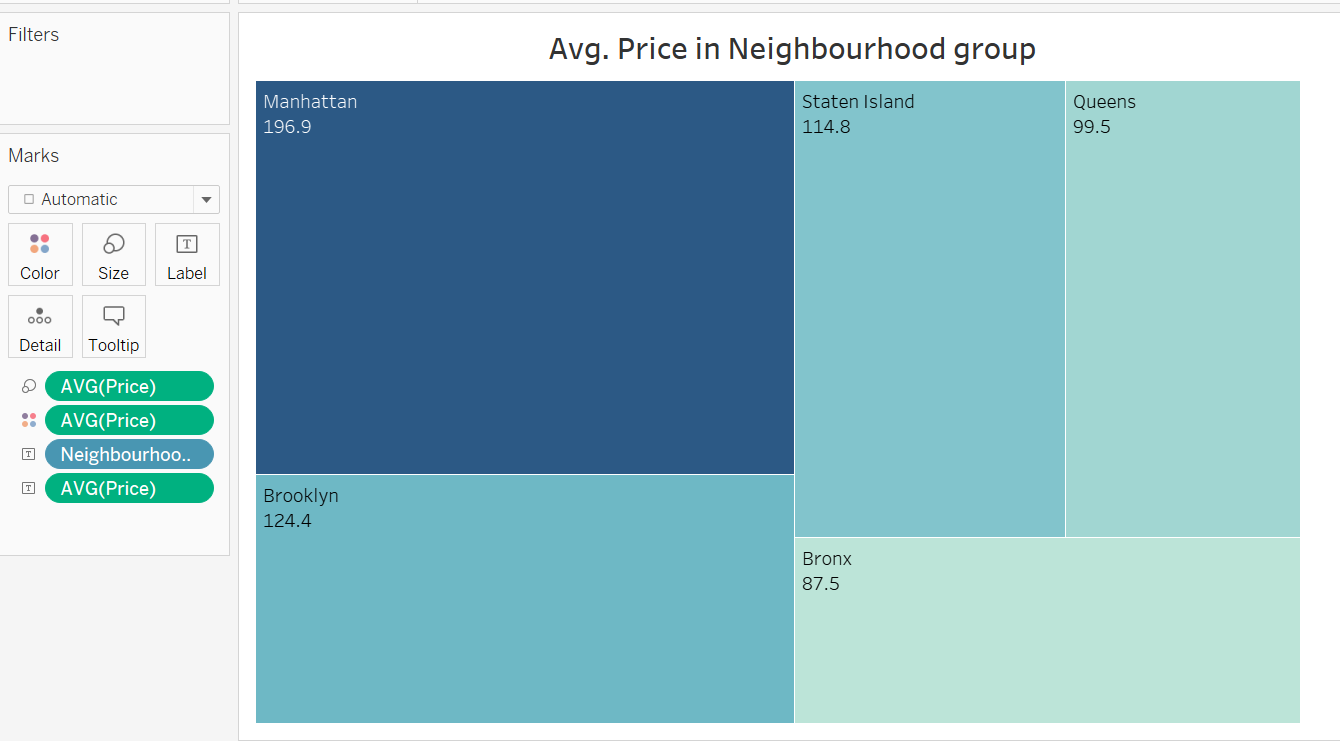
Insight: It is possible that many reviewers or people are interested in these properties and like to visit. Assuming that, we can restart the service in order to increase the revenue.

Top 20 Earning Hosts in revenue generation:



Observations: Above chart shows the top 20 host who earn highest revenues.

NeighbourHood group Vs Avg Price



Observations: We can observe that the avg rate of properties in Manhattan is higher compared to other properties among all other neighbourhoods.