**Final Project – Data Visualization**

**Topic – New York AirBnB Analysis**

**Puneet Kochar**

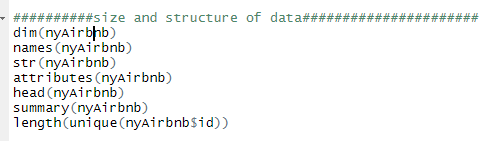
**University of Colorado Denver Business School**

New York City is one of the world's best-known cities. Each year it attracts millions of tourists which boosts our economy. Hence, NYC is one of Airbnb's hottest markets. New York City (NYC) has an active Airbnb market with over 48,000 rentals in the 2019 calendar year as of August. The management of the Airbnb wants to know the neighborhoods that are highly preferred in Airbnb bookings in New York city and Which neighborhoods and types of properties to be focused on. In order to answer queries of Management, I performed short analysis below on New York Airbnb listings data of 2019.

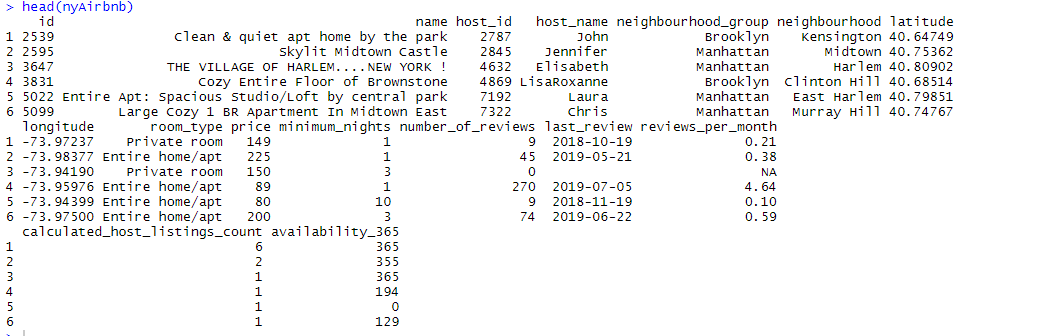
* **Description of Data**

The data is sourced from <https://www.kaggle.com/>. The dataset contains the Airbnb listings of New York City in different neighborhoods with the price of each property, type of property, latitude and longitude location, host id, host name, number of host counts, and various other features etc.

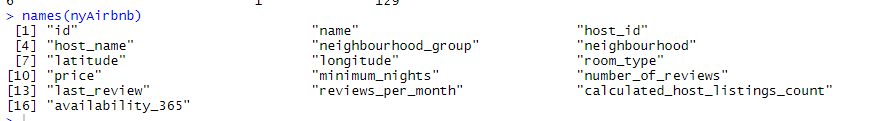
* **Getting to Know the Data**



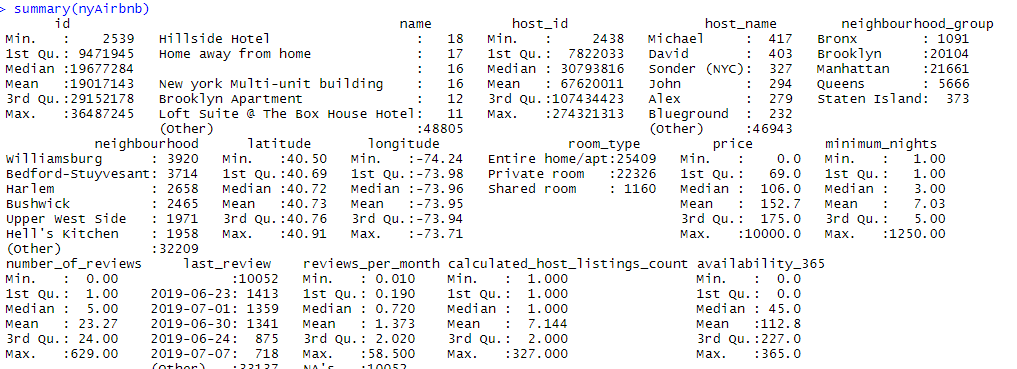
**The dataset has 48895 rows and 16 columns, below is the head of the data**



**Below are the column names of Data**



**Below is the Summary of Dataset**

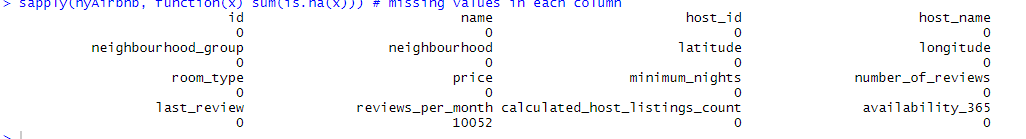


**All the rows in dataset are recorded with the unique host id and each row contains information about each Airbnb host listings in New York City**



* **Checking and Dealing with Missing Values**

On initial inspection through “sapply” function we found NA values in the dataset. On observing each column of the data for missing values, I identified that only “reviews\_per\_month” column had the missing values.

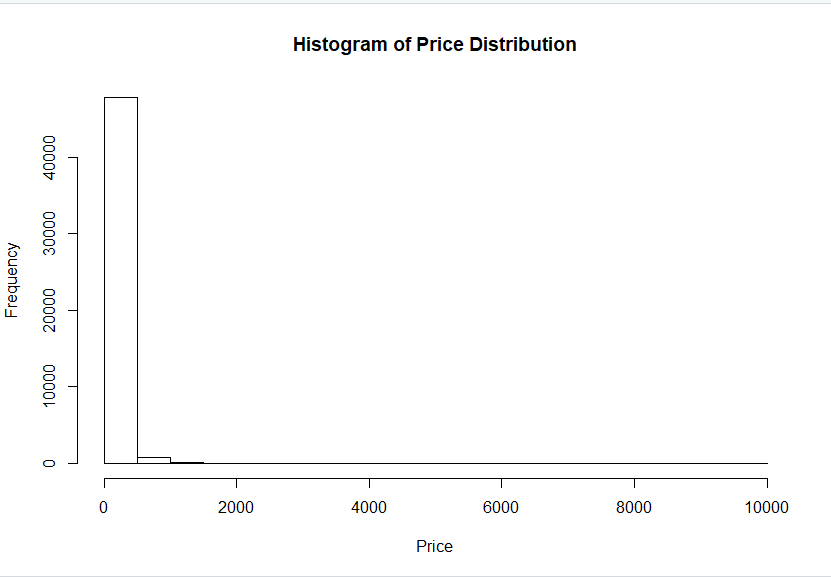


In order to deal with missing values, I replaced all the NA values in review\_per\_month column by 0 on the assumption that it is possible that property may have 0 reviews.



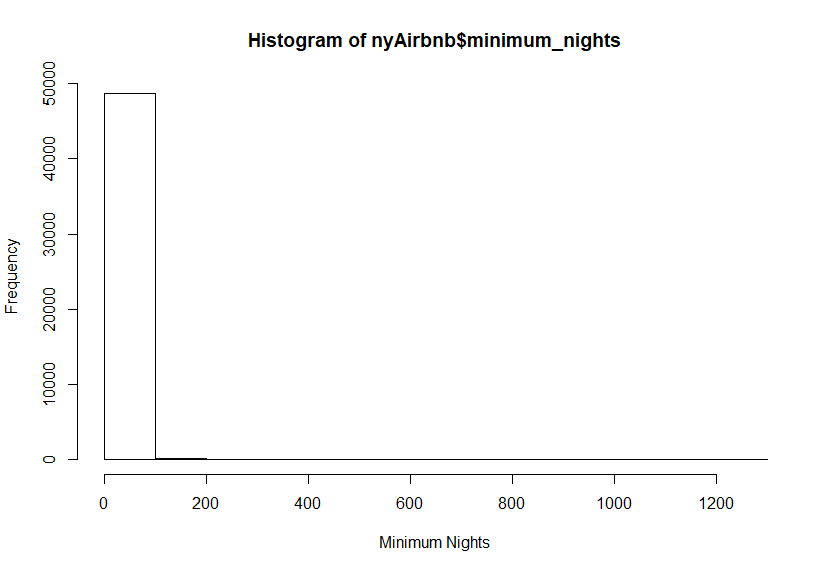
* **Distribution of Data**

The most important variable in this dataset in Price. So, I decided to check the distribution of Price data of Airbnb listings through histogram.

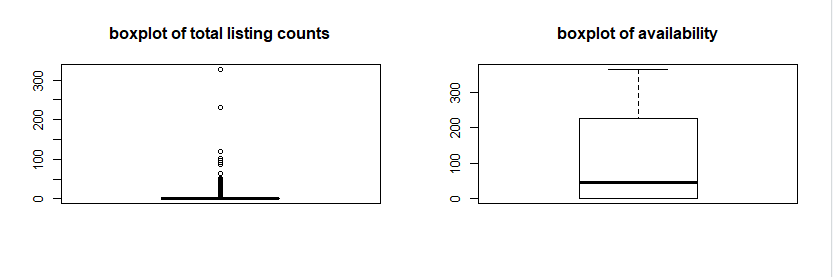


From the histogram, we can see that data of price is very much right skewed. Since original data distribution is skewed, we can take logarithmic transformation of price data for better insights

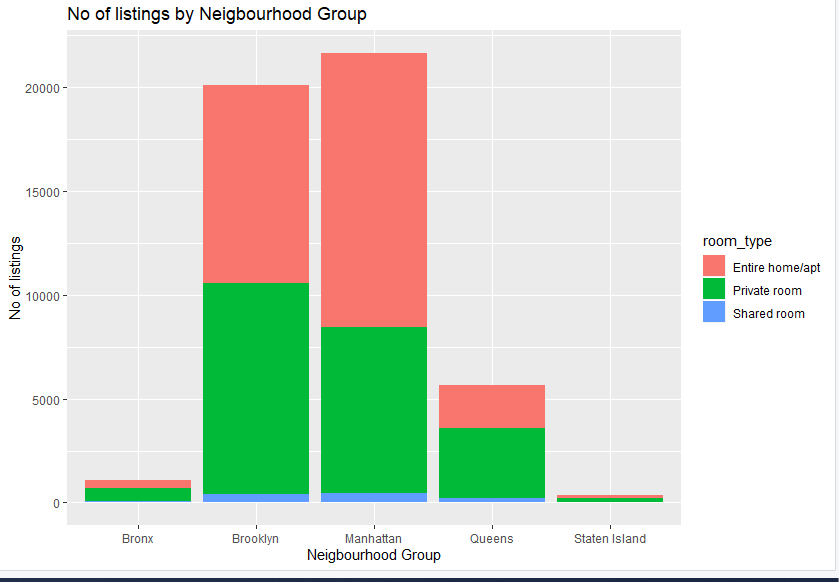
The data distribution of Minimum Nights is also right skewed.



Boxplot of Host counts for each Airbnb listings shows some extreme outliers which can be possible as a property can have large number of host counts, so we will ignore that on other hand data distribution of Availability of an listing shows no outliers and average availability of listings shows around 50 days.

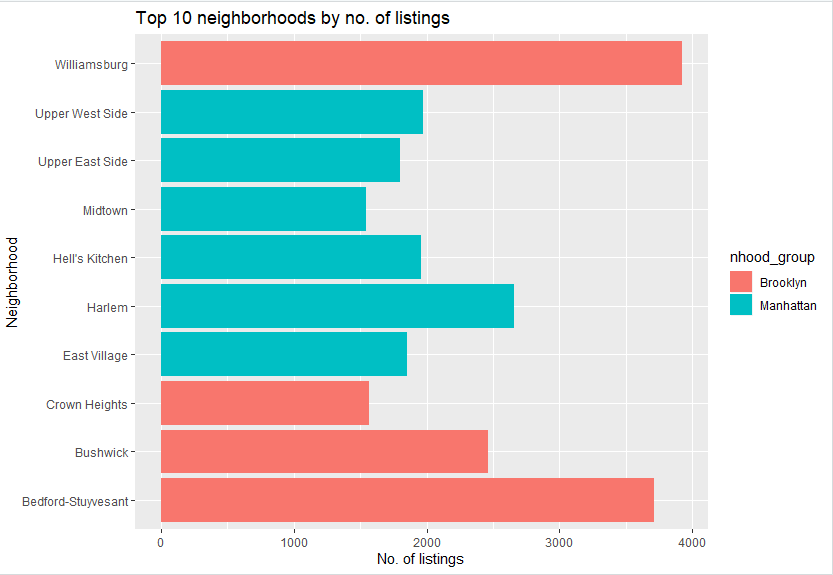


* **No of Listings by Neighborhood**



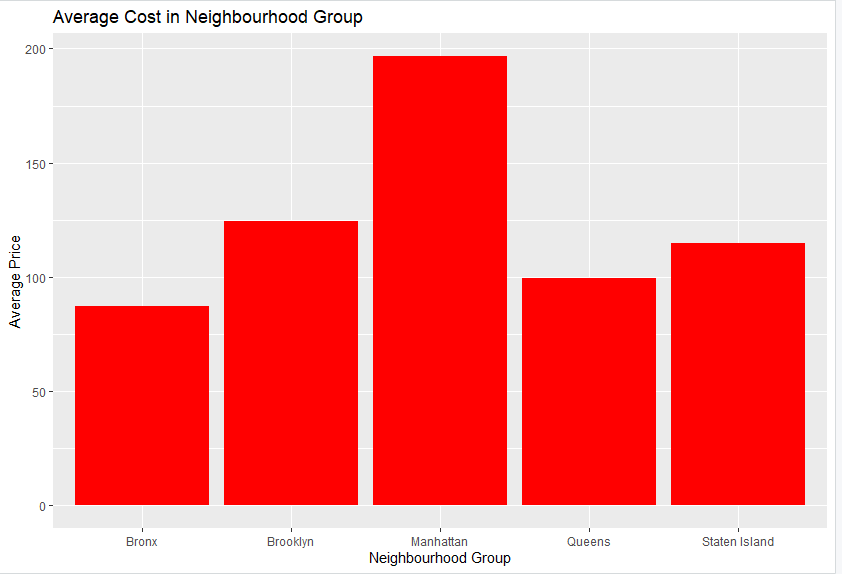
The above bar plot tells us that Manhattan has the most number of Airbnb listings in New York City followed by Brooklyn while the least listings can be identified in Staten Island. From the plot we can also identify that, Manhattan have more entire home apartment types of property than other neighborhood groups while the number of shared room listings are particularly low in all the neighborhoods.

* **Top Neighborhoods**



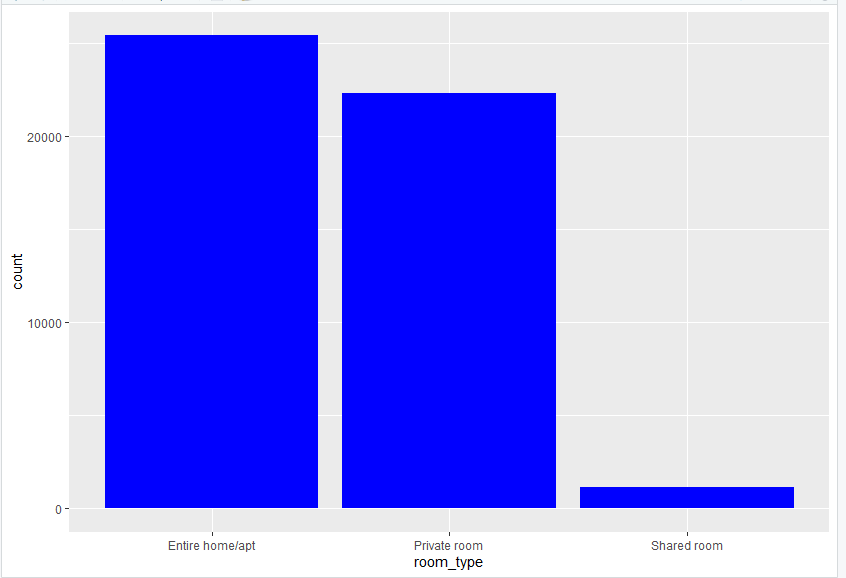
Through analysis of above plot, we can identify the top 10 neighborhoods based on listings in New York city. Williamsburg neighborhood has the most number of listings in New York City followed by Bedford-Stuyvesant. Also, all the top neighborhoods are either in Manhattan or Brooklyn.

* **Average Cost in Each Neighborhood Group**



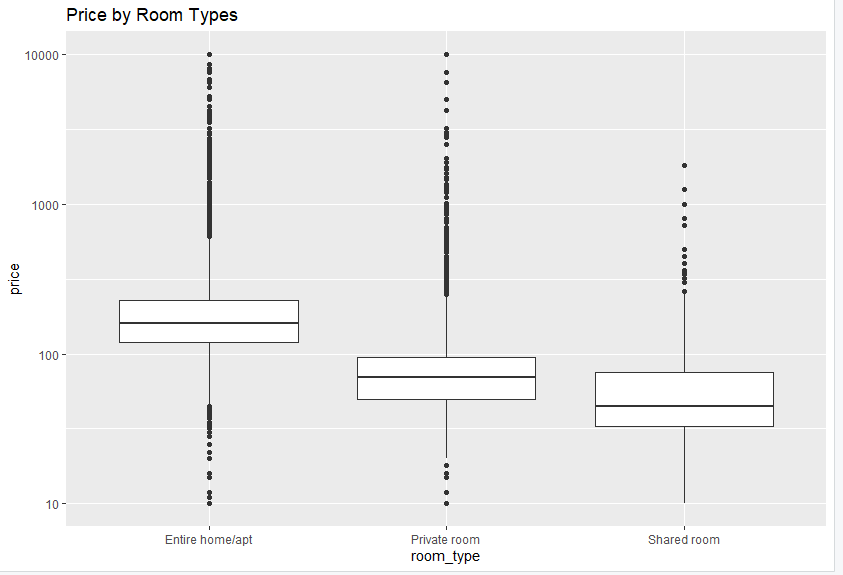
We can identify from the above plot that Average cost of each listing is higher in Manhattan followed by Brooklyn. The interesting analysis in this plot is that Average cost in the Staten Island listing is higher than Queens and Bronx even though the number of listings is less in Staten Island which we analyses before. This might be due to a greater number of footfall of tourists and Airbnb bookings in Staten Island than Queens and Bronx.

* **Analysis of Property Types**



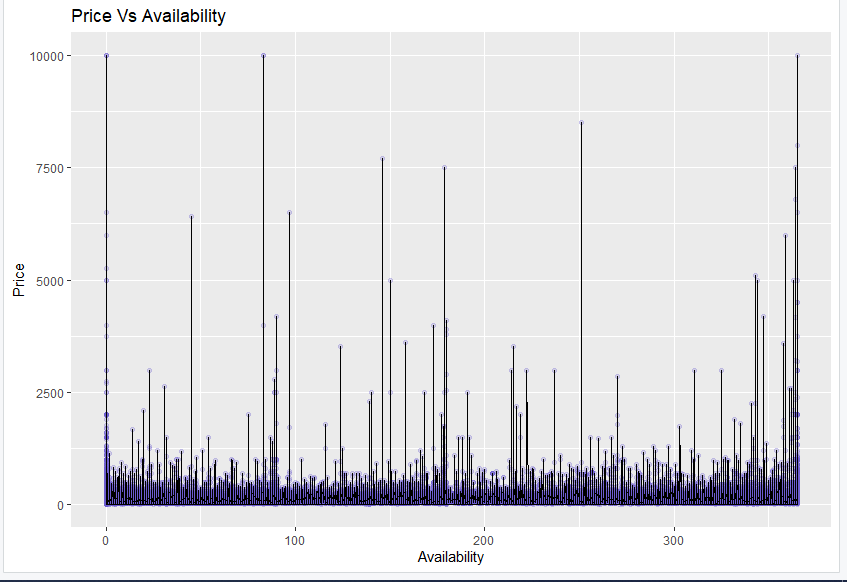
From the above plot, we can identify that Entire home\Apartment are mostly preferred choices in New York Airbnb listings followed by Private room. The more focus by management should be on Entire apartment listings.

* **Price Analysis based on Property Types**



The boxplot shows the price distribution of Airbnb listings in all three categories of property type. There seems to be much variation in price in each property type. Overall, it seems that “Entire apartment” listings are slightly priced than “Private room and “Shared Room”.

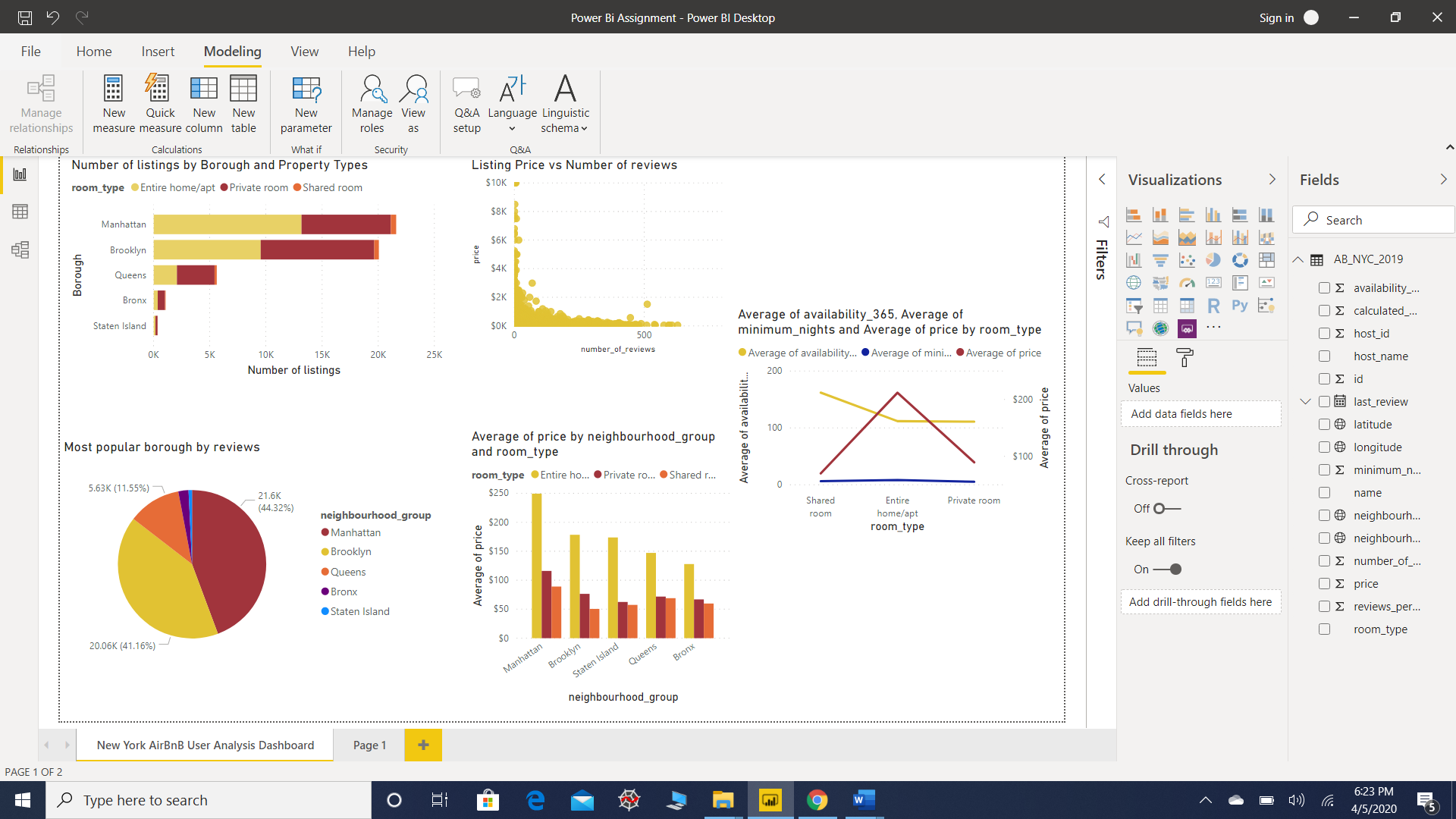
* **Price vs Availability**



**From the above scatterplot, we can identify no pattern which means that there is no linear or nonlinear relationship between price and availability. There is lot of expensive listings with few available days.**

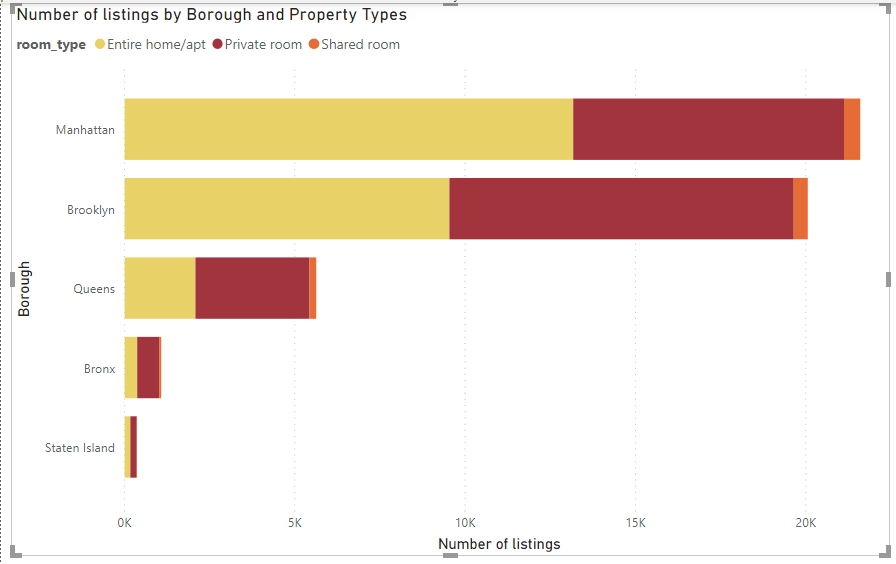
* **POWER BI**

New York City is one of the world's best-known cities. Each year it attracts millions of tourists which boosts our economy. Hence, NYC is one of Airbnb's hottest markets. New York City (NYC) has an active Airbnb market with over 48,000 rentals in the 2019 calendar year as of August. Below is the Power Bi Analysis of New York Airbnb data which can help user in determining the best Airbnb hospitality in New York based on Location, Prices, reviews, Availability in year and Property Type.



**Power BI Dashboard of New York Airbnb User Analysis**

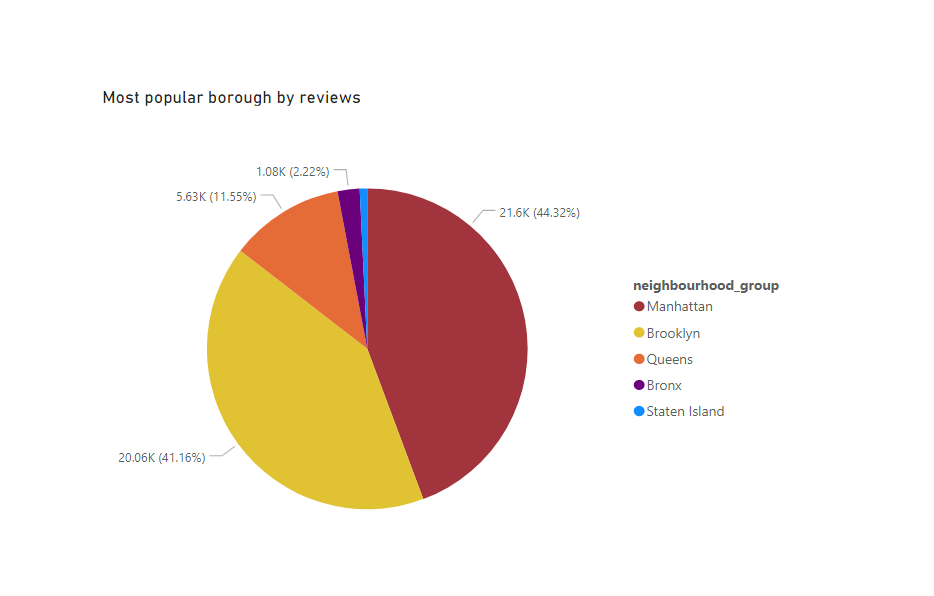
* **Number of Listings by Borough and Property Types**



**Fig 1**

This analysis can help user in determining where are the most number of Airbnb listings located in New York. From the analysis, Manhattan city has the greatest number of listings in New York with followed by Brooklyn. This analysis also reveals the user the distribution of property types in different borough. For example, the more listings of Private rooms are in Brooklyn than in Manhattan While the Manhattan host greater number of Entire Home/ Apt.

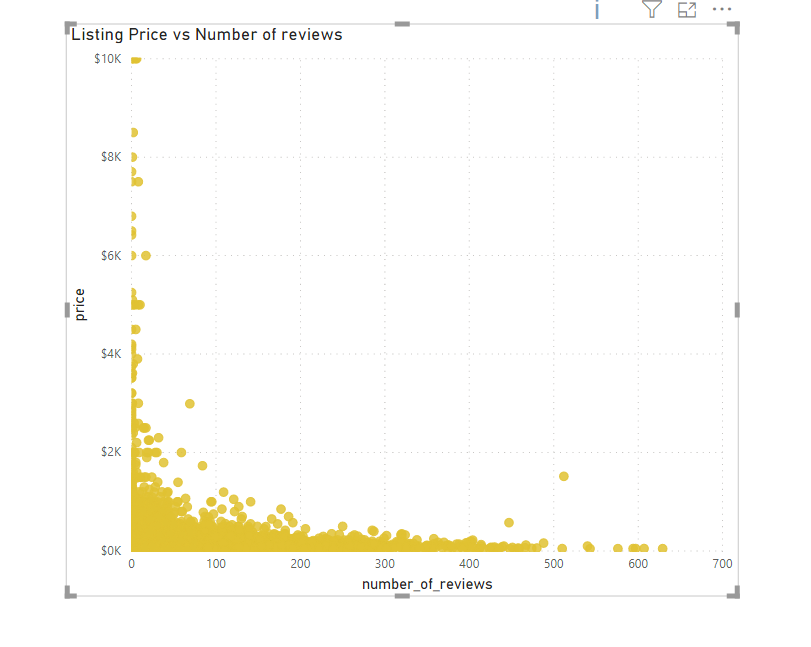
* **Most Popular Borough based on reviews.**



**Fig 2**

This Pie Chart that represents the most popular borough in New York for Airbnb listings. This pie chart can help user in determining which is the most popular borough which has received the most number of bookings based on reviews. This represents the listings of Airbnb in different borough which has more than 100 reviews. From the visualization, we can say that Manhattan seems to be popular neighborhood with based on reviews followed by Brooklyn while Staten Island seems to be least popular.

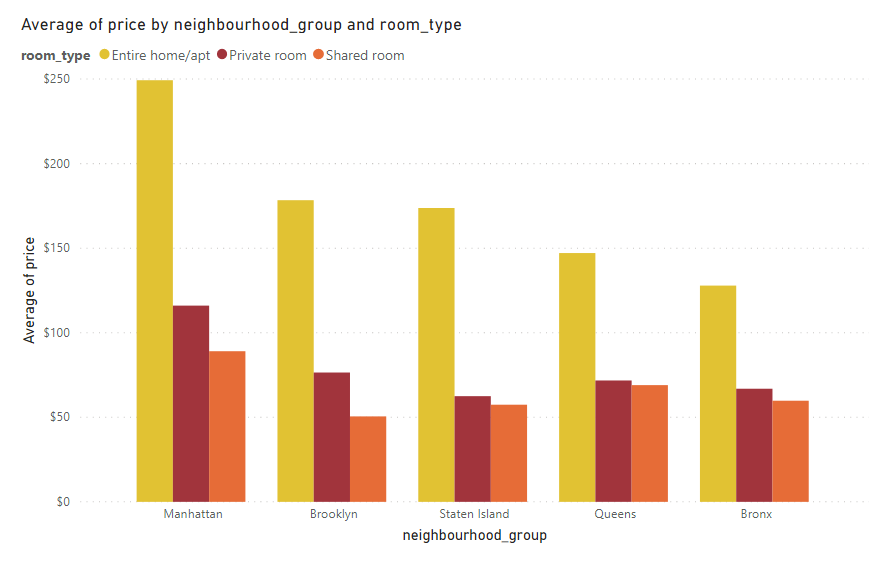
* **Relationship between Price of Listing and Reviews**



**Fig 3**

This is a scatter chart depicting the relationship between the number of reviews a property received and the price of Property. Although, not a clear relationship can be determined from the visualization, but we can identify from the visualization that high priced property received less reviews while the low-priced property received more number of reviews. We can infer from this that low-priced property might have received more booking than high priced properties.

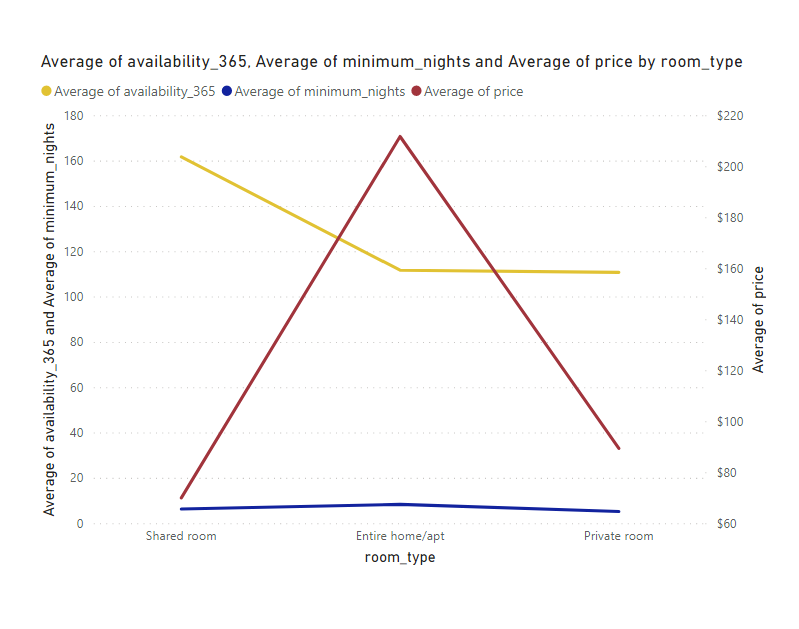
* **Average Price by Borough**



**Fig 4**

After analyzing the number of listings and popularity of listings in different borough of New York, the user might look for distribution of prices for different property types in different borough and can select a location according to his budget. This stacked bar chart represents Average Price of different property types of Airbnb in all the 5 boroughs. From the visualization, user can identify that Entire home or apartment is much costlier than Private room or shared room in all the 5 boroughs and Manhattan being the most expensive for Entire Home or Apartment property types. Also, we can identify that though Staten Island have a smaller number of listings and popularity form previous analysis, but the listings of Entire Home or Apartment in Staten Island are costlier than Queens and Bronx. Also, average price of private rooms is highest in Manhattan, followed by Brooklyn while it is cheapest in Staten Island, and coming to shared room, it also the most expensive in Manhattan followed by Queens while cheapest in Brooklyn.

This analysis can really help user in choosing the best Airbnb listing based on his property type preference and Budget.

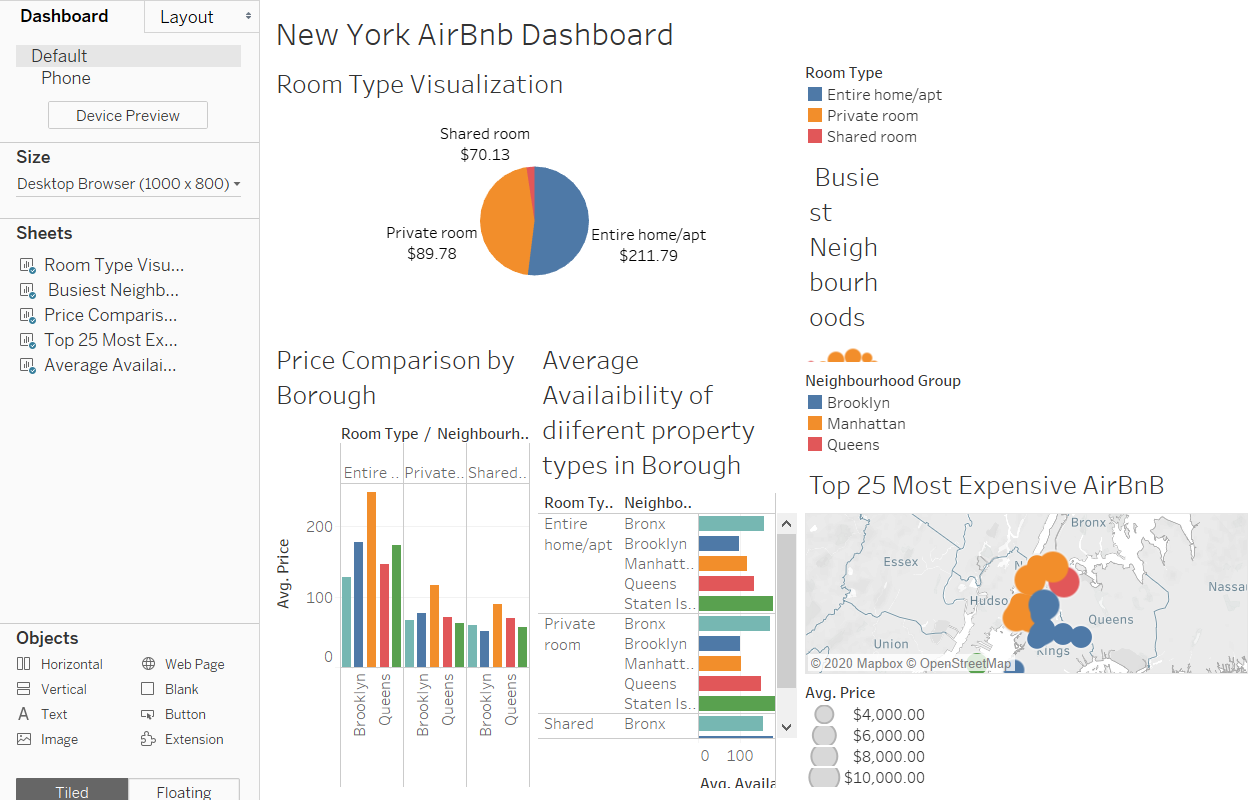


**Fig 5**

This visualization represents the average price, availability of property and average minimum nights of different types of properties. This visualization can help user in determining the type of property that he can prefer based on availability, price and minimum nights. From the visualization, we can identify that in context of availability, Shared rooms are available for the most number of days in year with an average of around 161 days followed by Entire home and Apartment while the least available are Private room properties. In the context of Price, the Entire home/Apt is much costlier than Private rooms and Shared Rooms while in the context of minimum nights, Entire home have higher range of minimum nights followed by Shared room and least for Private rooms.

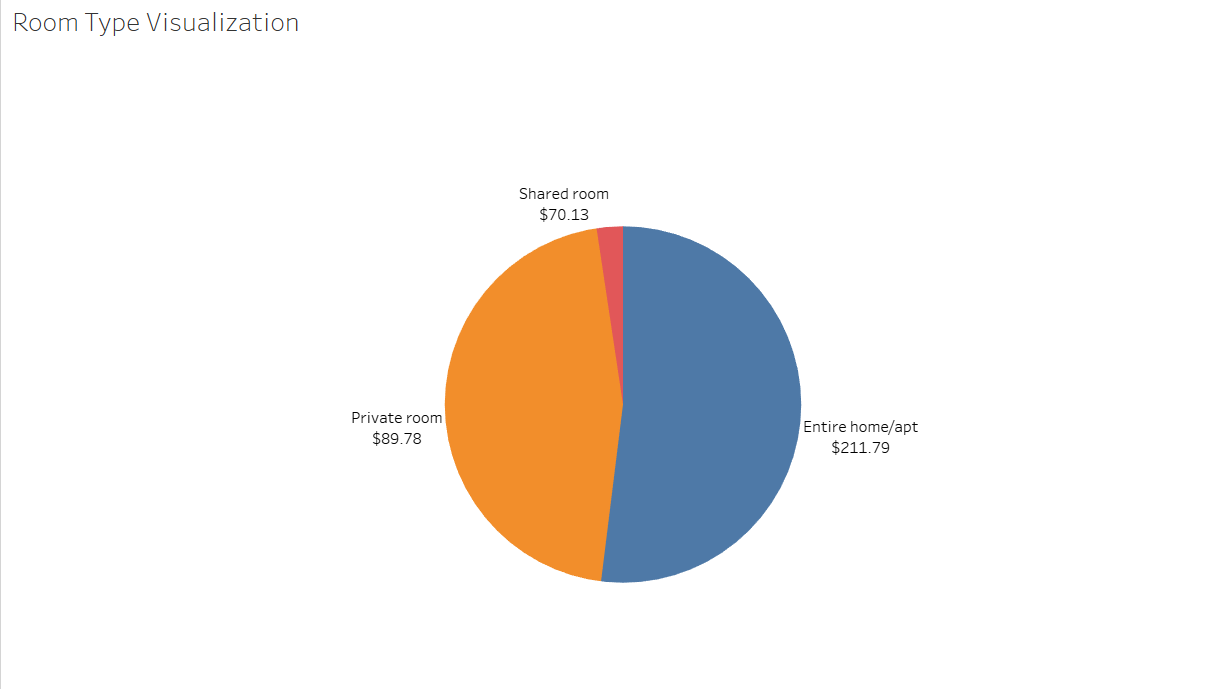
* **TABLEAU**

New York City is one of the world's best-known cities. Each year it attracts millions of tourists which boosts our economy. Hence, NYC is one of Airbnb's hottest markets. New York City (NYC) has an active Airbnb market with over 48,000 rentals in the 2019 calendar year as of August. Below is the Analysis of New York Airbnb data in Tableau which can help user in determining the best Airbnb hospitality in New York.



**Tableau Dashboard of New York Airbnb Analysis**

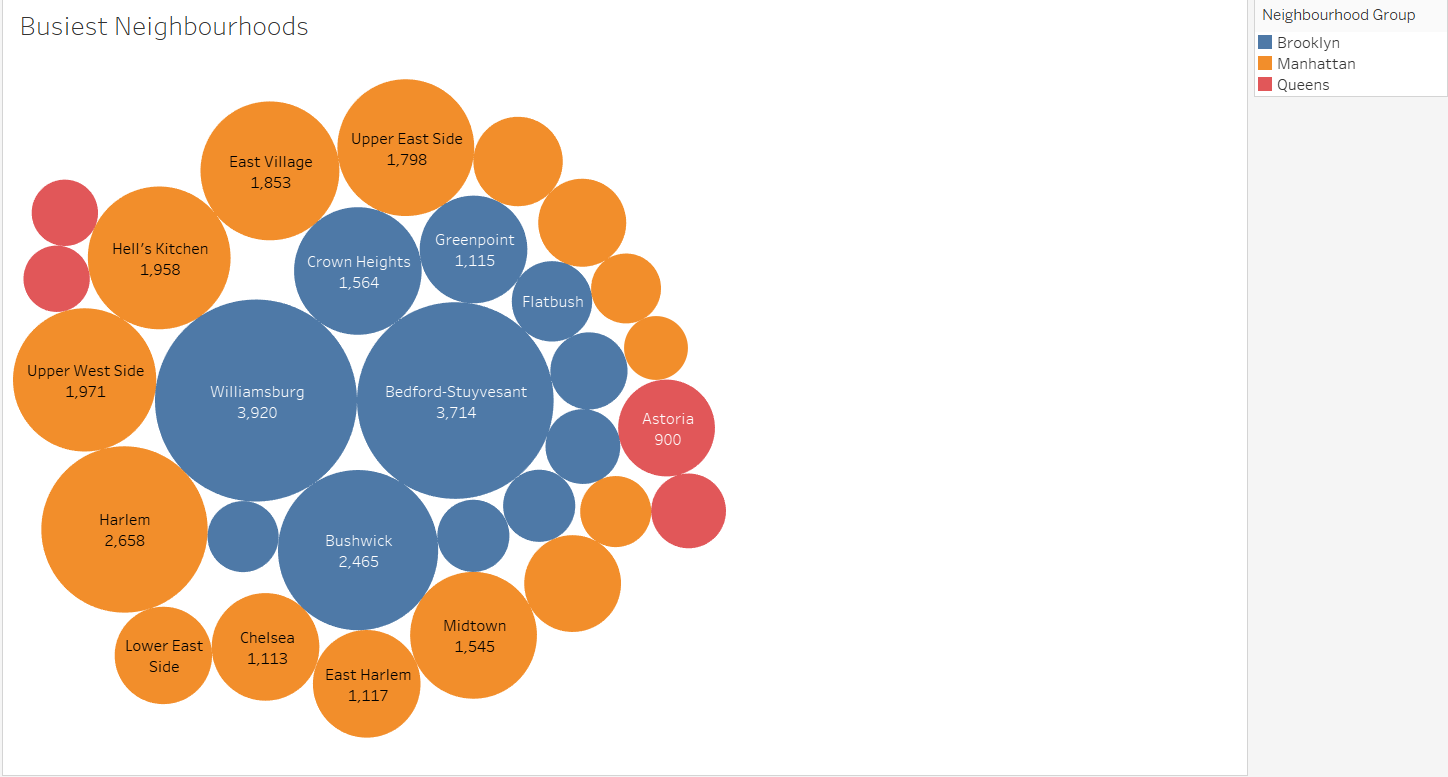
* **Room Type Statistics**



**Fig 1**

This analysis of Pie Chart shows the distribution of Types of Airbnb properties in New York, The color of the pie chart depicts the type of Property and the size of Pie slice denotes the number of listings of a property in New York. On visualizing the pie chart, we can identify that there are more of an entire home/apt type of listings of Airbnb property in New York than Private room and least is shared room property type. While the entire home/apt is much costlier on an average that Private or shared room.

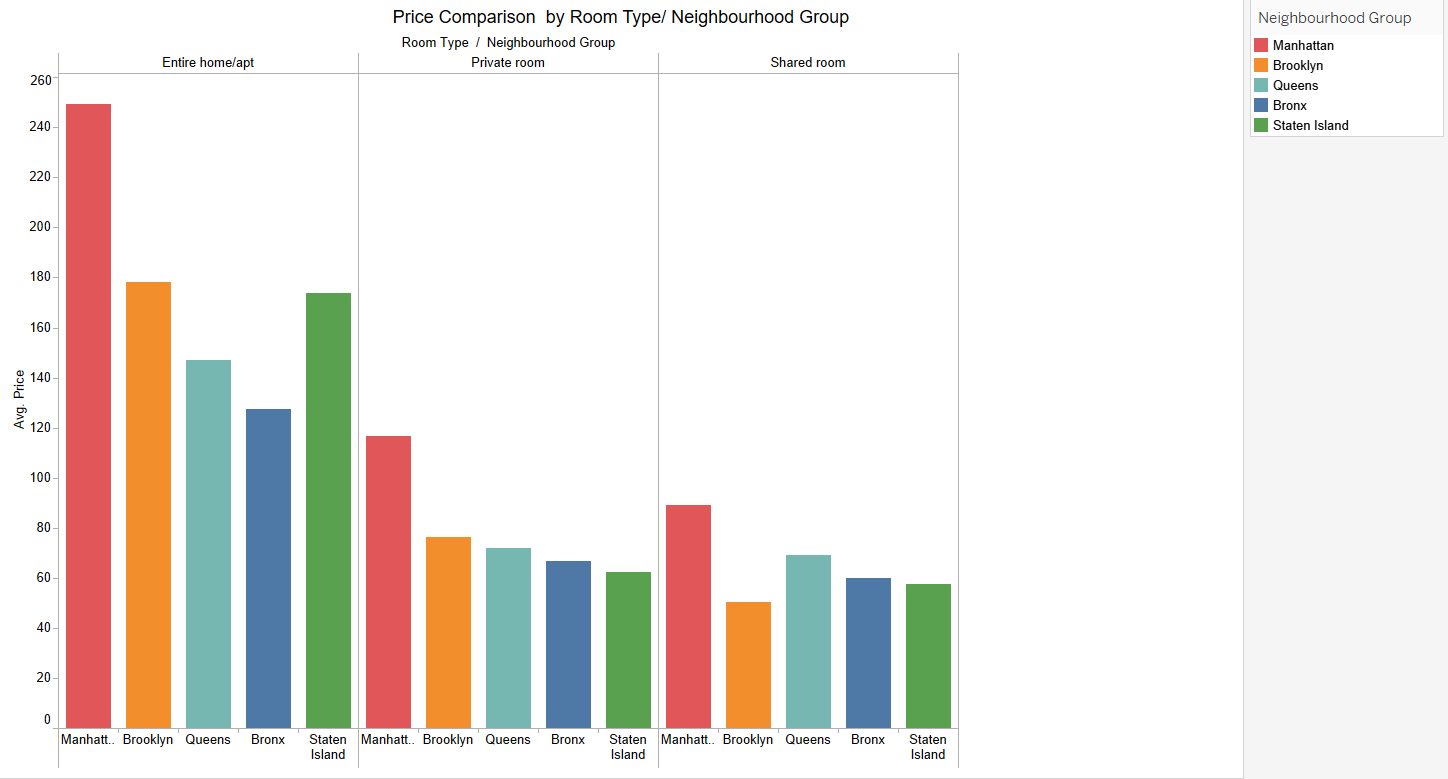
* **Busiest Neighborhoods**



**Fig 2**

This Bubble Chart represents the busiest Neighborhoods in New York. The color represents the Borough the neighborhood belongs to and the size represents the calculated host listing counts. Brooklyn and Manhattan seem to be the busiest Borough as top 10 busiest neighborhood mostly belongs to Manhattan and Brooklyn. Williamsburg is the busiest neighborhood followed by Bedford-Stuyvesant.

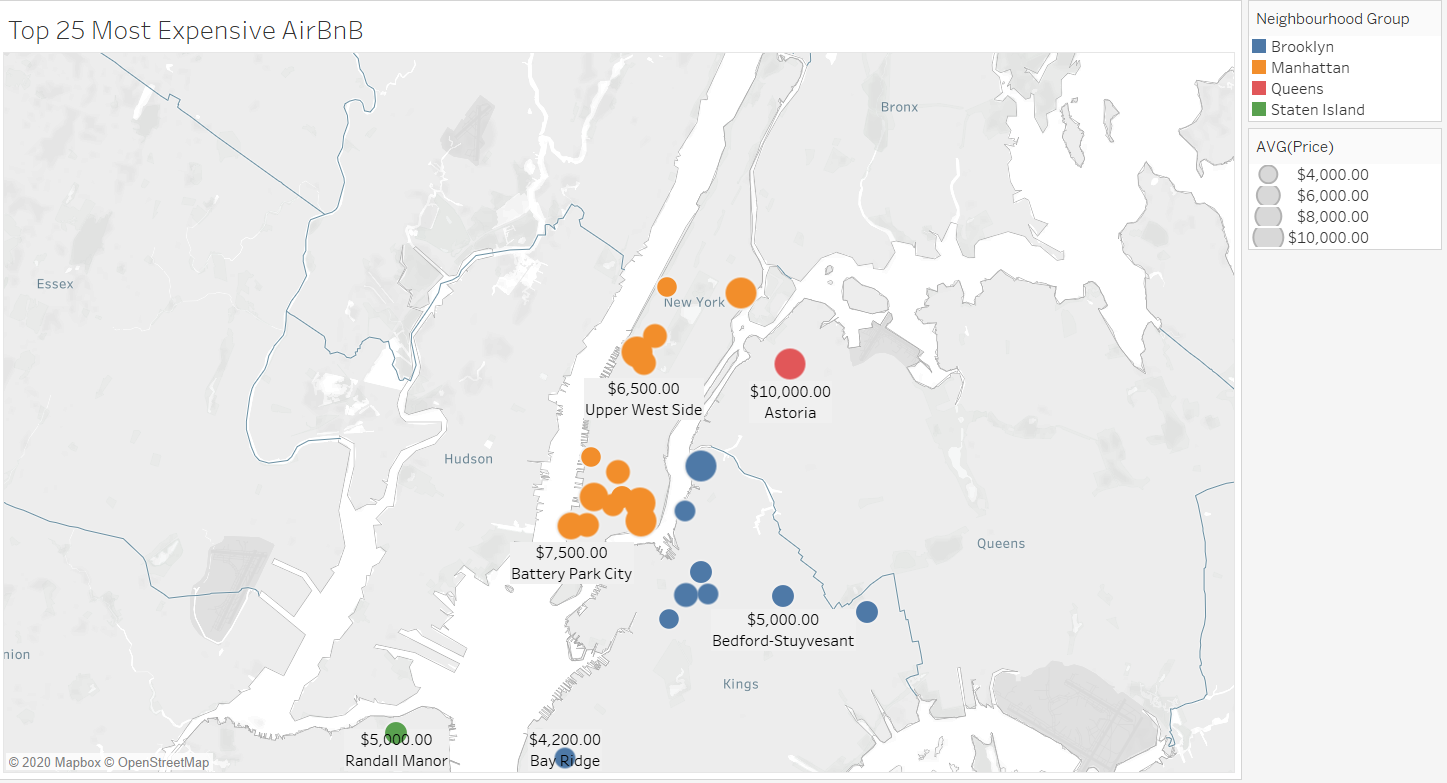
* **Price Comparison**



**Fig 3**

This is a bar chart that represents the average price of different property types in different borough of New York. The color represents different borough of New York. From the visualization, we can identify that Manhattan is costlier in all the property types in different borough. There is not much difference in the average price of entire home/apt in Brooklyn and Staten Island While the least priced entire Homes are in Bronx. Similarly, the least priced private rooms are in Staten island and the least priced shared rooms are in Brooklyn.

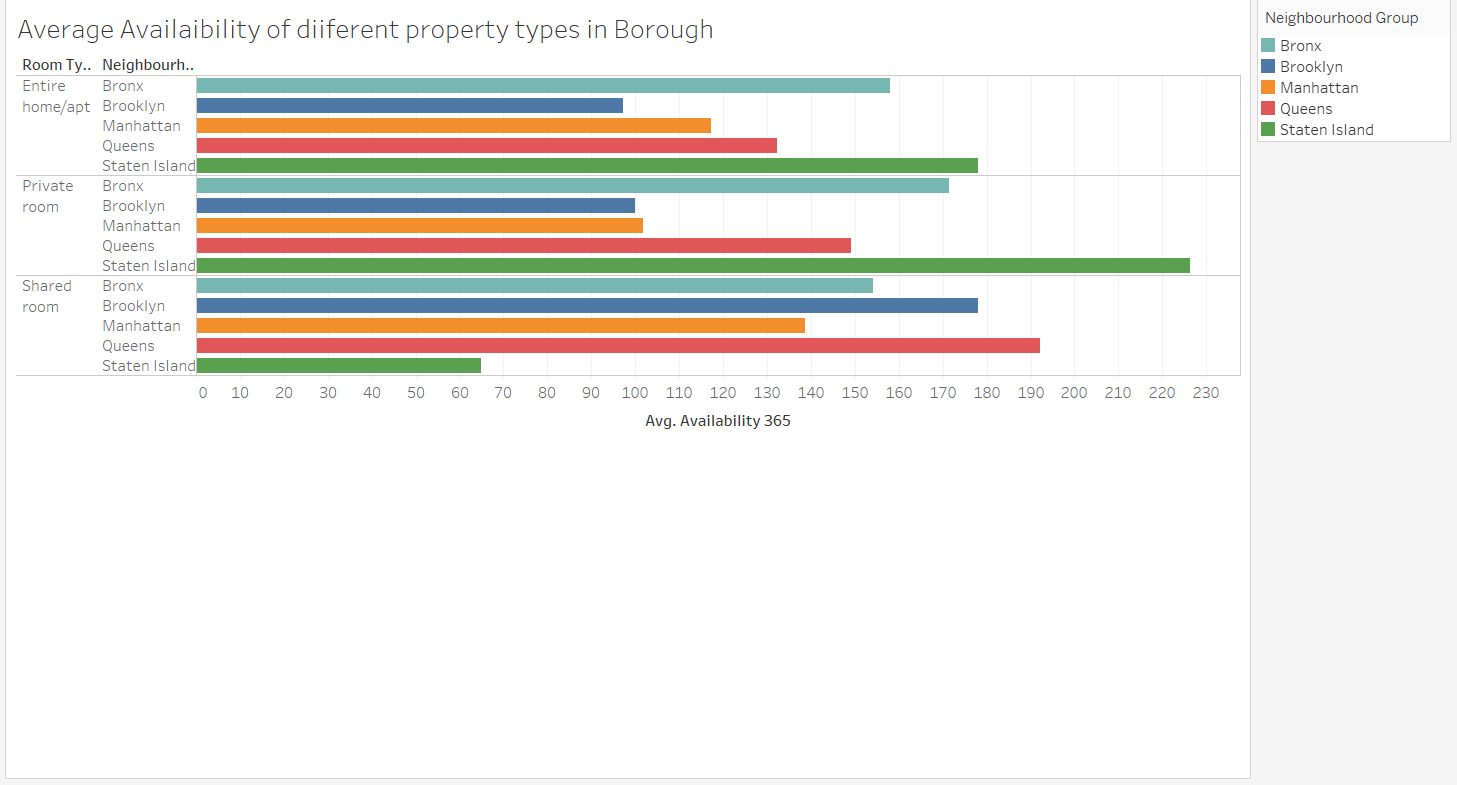
* **Top Expensive Neighborhood**



**Fig 4**

This Map represents the top 25 most expensive neighborhood in New York. The size denotes the price of an Airbnb Listings while the color denotes to the borough the listing belongs to. Most of the expensive Airbnb listings are in Manhattan and Brooklyn and One in Staten Island among Top 25 expensive neighborhood. The most expensive listing is in Manhattan called 1- Berlin center with an average price of $10000 in Upper west side neighborhood Of New York

* **Analysis of Availability for the year.**



**Fig 5**

This visualization represents the average availability of different property in 5 boroughs of New York. The color denotes the 5 boroughs of New York. The entire home or apartment are available for most number of days in Staten Island followed by Bronx while the last availability is in Brooklyn. The Private rooms are available for most number of days in Staten Island followed by Bronx while the least in Brooklyn and the shared rooms are available for most number of days in Queens followed by Brooklyn while the least in Staten Island. This we can infer that most people prefer Brooklyn and Manhattan for Private room and Entire Apartment while people mostly prefer shared rooms in Staten Island.