**Assignment 10**

**Puneet Kochar**

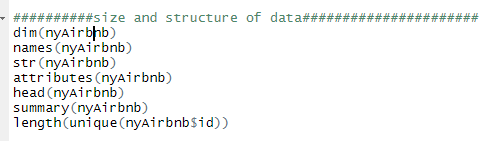
**University of Colorado Denver Business School**

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.

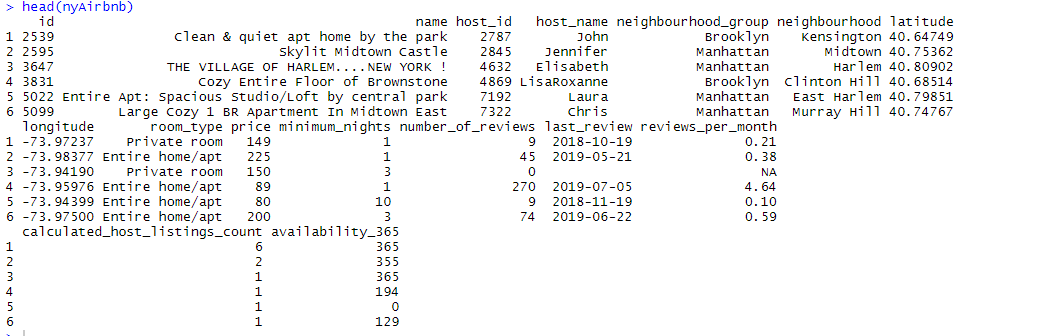
* **Description of Data**

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.

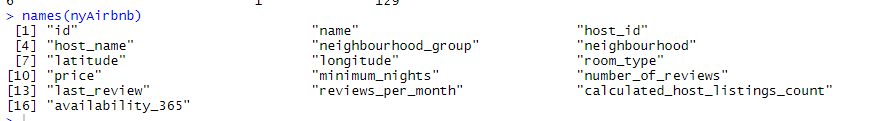
* **Getting to Know the Data**



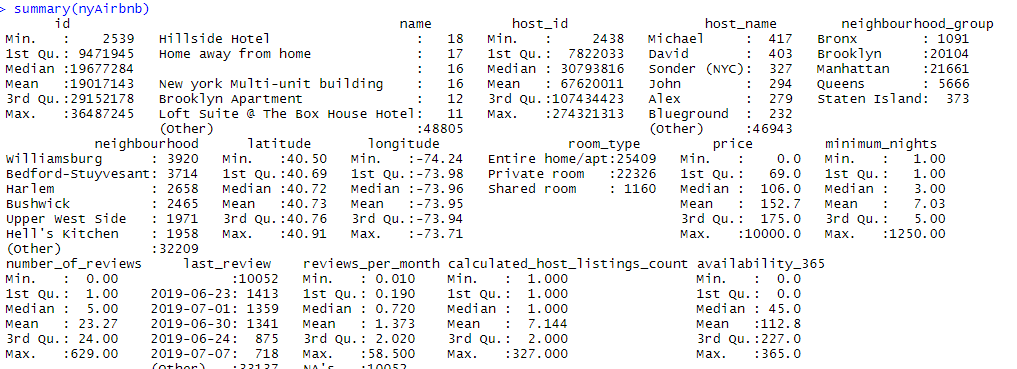
**The dataset has 48895 rows and 16 columns, below is the head of the data**



**Below are the column names of Data**



**Below is the Summary of Dataset**

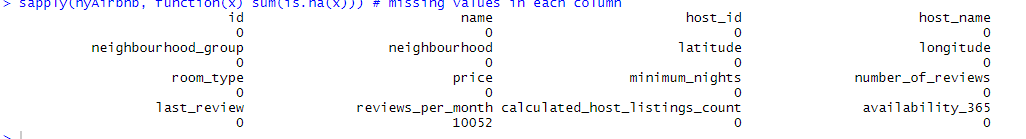


**All the rows in dataset are recorded with the unique host id and each row contains information about each Airbnb host listings in New York City**



* **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.

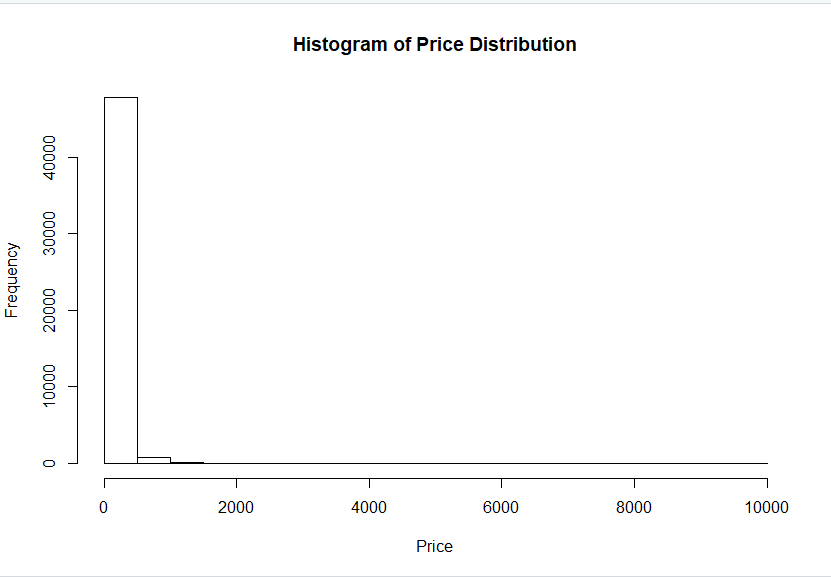


In order to deal with missing values, I replaced all the NA values in review\_per\_month column by 0 on the assumption that it is possible that property may have 0 reviews.



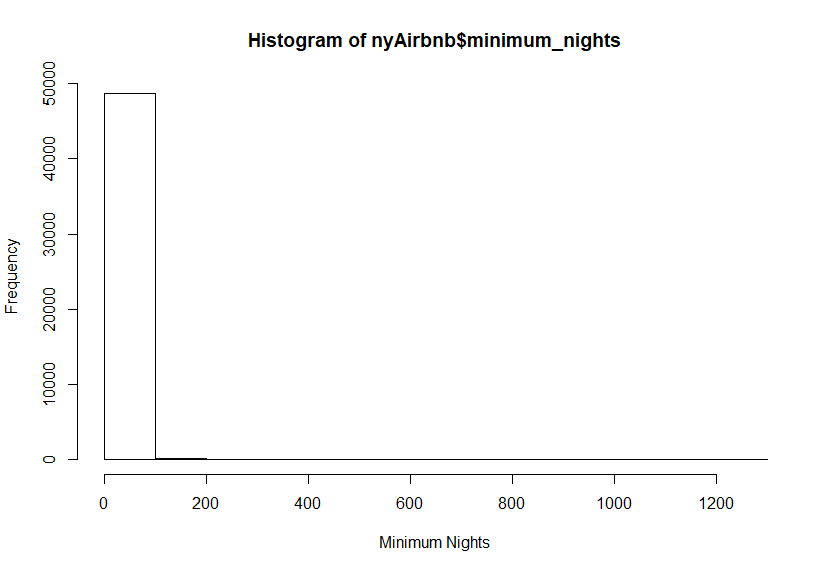
* **Distribution of Data**

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

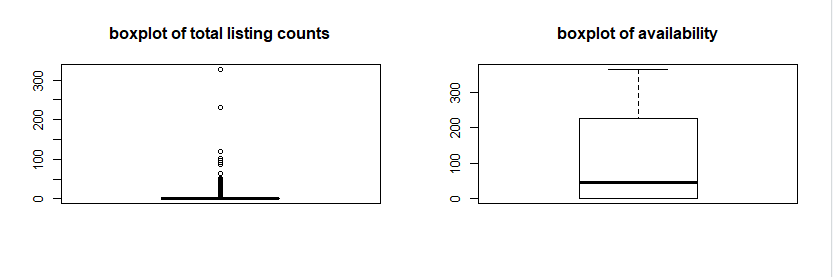


From the histogram, we can see that data of price is very much right skewed. Since original data distribution is skewed, we can take logarithmic transformation of price data for better insights

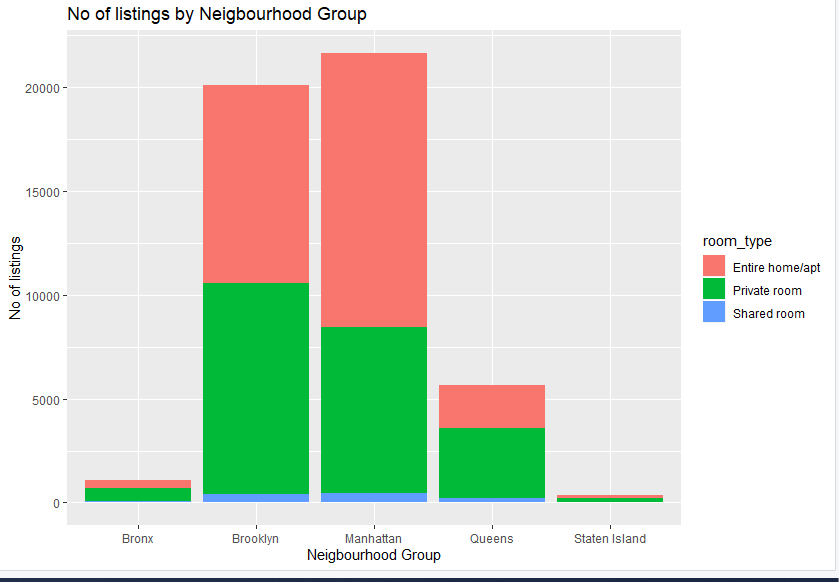
The data distribution of Minimum Nights is also right skewed.



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.

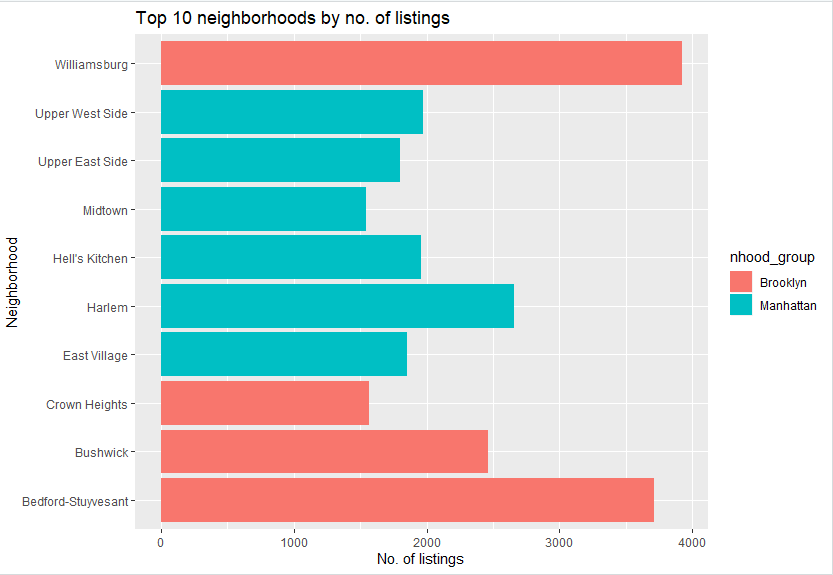


* **No of Listings by Neighborhood**



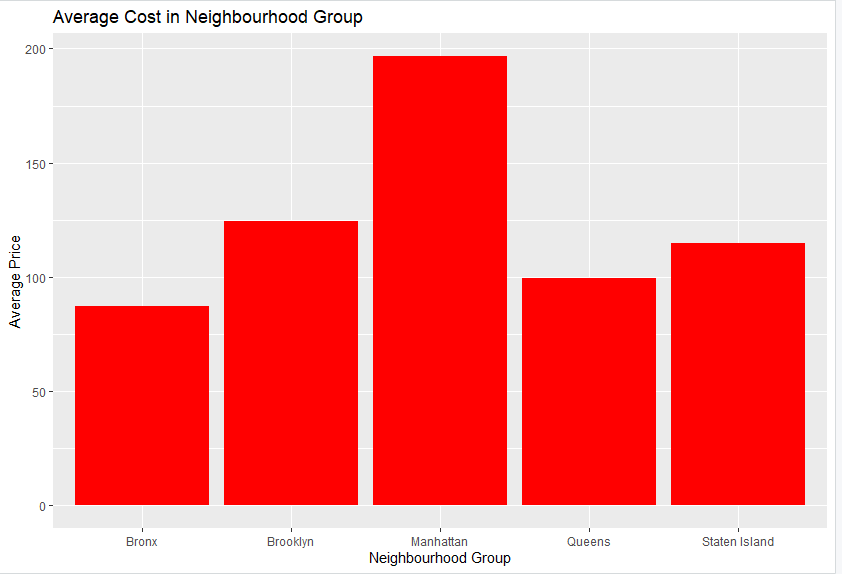
The above bar plot tells us that Manhattan has the most number of Airbnb listings in New York City followed by Brooklyn while the least listings can be identified in Staten Island. From the plot we can also identify that, Manhattan have more entire home apartment types of property than other neighborhood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

* **Top Neighborhoods**



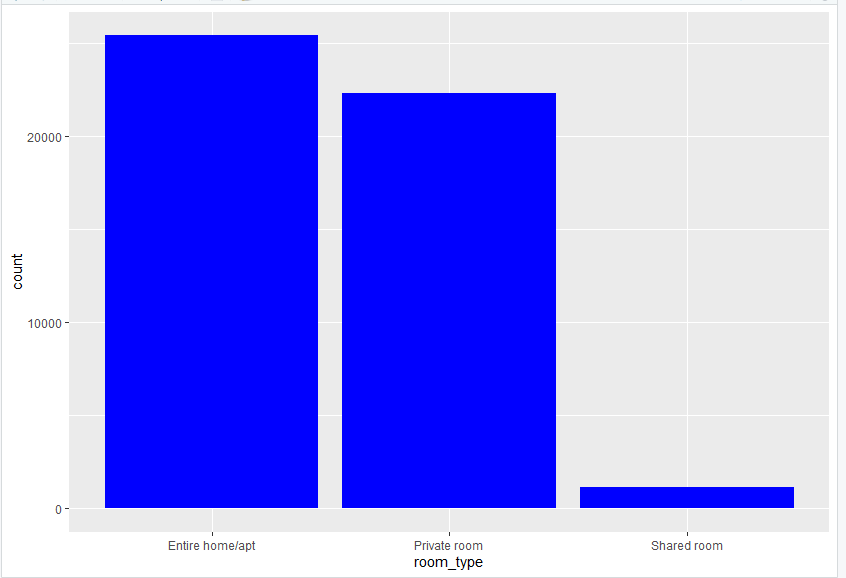
Through analysis of above plot, we can identify the top 10 neighborhoods based on listings in New York city. Williamsburg neighborhood has the most number of listings in New York City followed by Bedford-Stuyvesant. Also, all the top neighborhoods are either in Manhattan or Brooklyn.

* **Average Cost in Each Neighborhood Group**



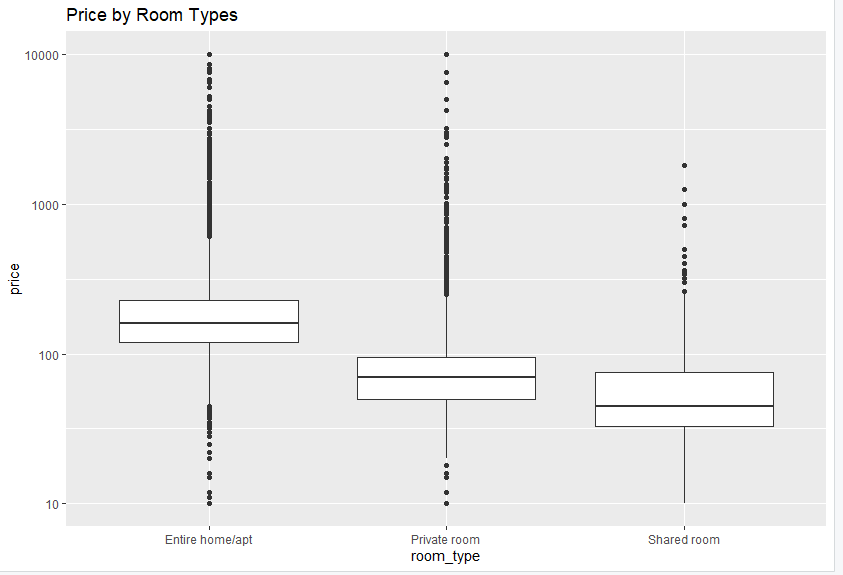
We can identify from the above plot that Average cost of each listing is higher in Manhattan followed by Brooklyn. The interesting analysis in this plot is that Average cost in the Staten Island listing is higher than Queens and Bronx even though the number of listings is less in Staten Island which 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**



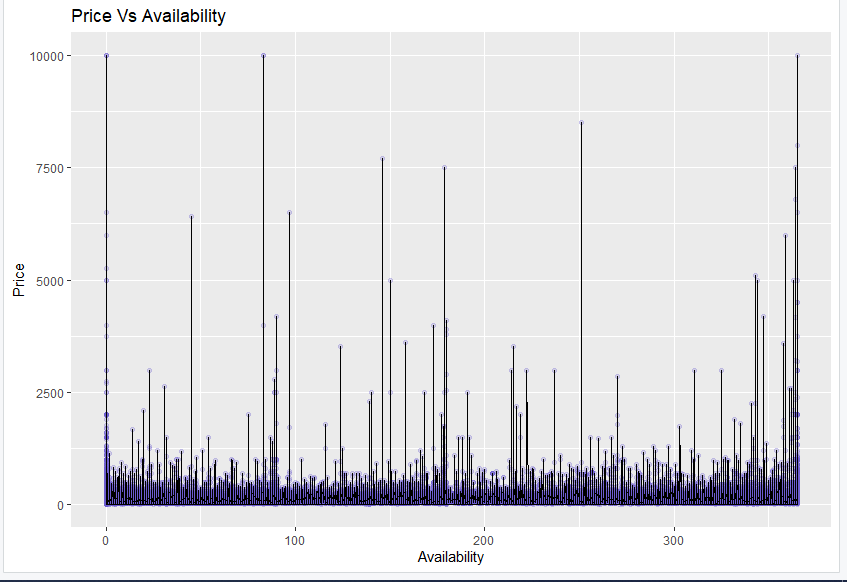
From the above plot, we can identify that Entire home\Apartment are mostly preferred choices in New York Airbnb listings followed by Private room. The more focus by management should be on Entire apartment listings.

* **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.**