New York City Airbnb Analysis

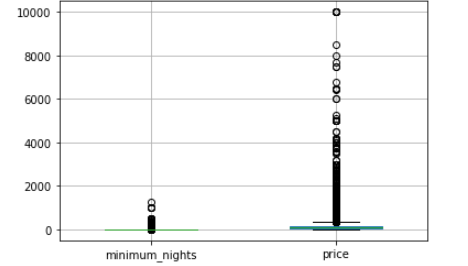
Airbnb entered the market in 2008, and ever since it is been one of the best options for people who plans vacation trip, business trip and much more.

New York City is considered to be the largest city for Airbnb guest with over 40 homes being rented per square Km. In this project I performed an exploratory data analysis of Airbnb dataset sourced from Kaggle.com website. I used Python, my code and visualizations can be found on my GitHub:

**Data Cleaning and Preparation:**

Datasets are usually unorganized and require a lot of cleaning and processing before diving into the analysis. For this purpose, I use Python programming language to do most of my work. I used most of Pandas and NumPy library to clean the data which consisted of:

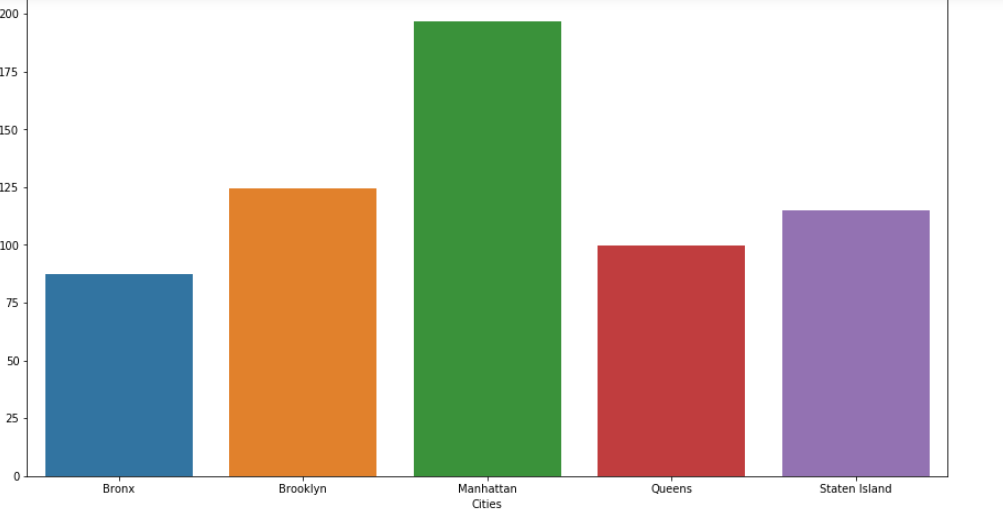
1. Performing R’s Summary by getting the 1st quartile, min, max, variance…
2. Checking outliers in which I found two major outliers in price specially having the price going from $0 to $10,000 per night. Also, we have a major outlier in minimum nights where people spent 1250 nights in an Airbnb house.



1. Missing values were handled by replacing NAN with 0 for numbers and ‘not given’ for names. Removing duplicates was done by using the popular function .duplicates() from python. Lastly, I remove unwanted columns such as ‘Id ‘and ‘host\_id’.

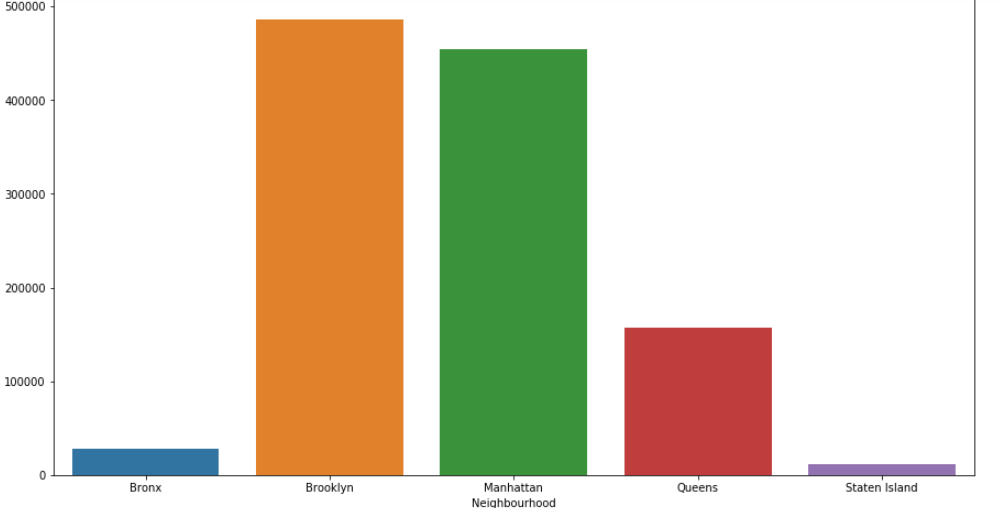
**Exploratory Data Analysis (EDA):**

In this process I was able to find the five most neighborhood group we were dealing with which are Manhattan, Brooklyn, Bronx, Staten Island and Queens.

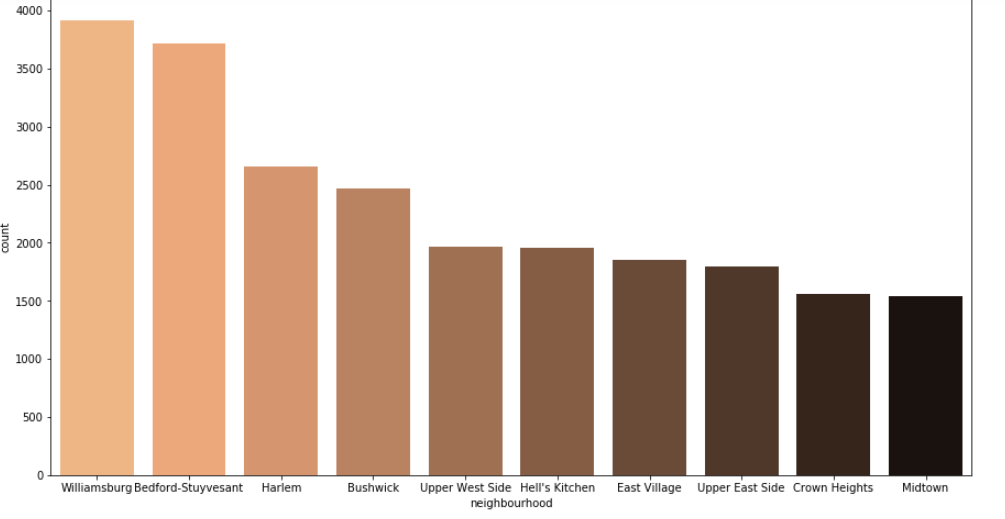
Plotting the average price per neighborhood group we can see that; Manhattan and Brooklyn are the most expensive cities in New York, with an average price per night of $249.24 for an

Entire home / apt, $116.77 for a private room and $88.97 for s shared room.

Second, I was able to get the reviews average per neighborhood as well, we can see that people leave a lot of reviews about Brooklyn and Manhattan.

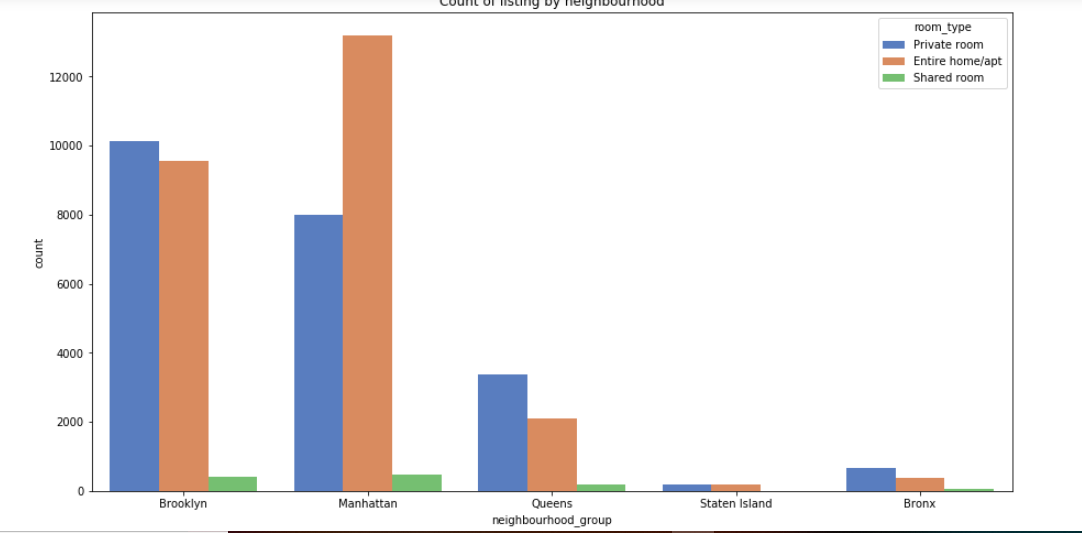


We cannot really whether or not these reviews are good reviews or bad reviews since the dataset is not really explicit about that. However, while doing the mean of reviews we can see that in some places Staten Island having the lowest price but the more reviews.

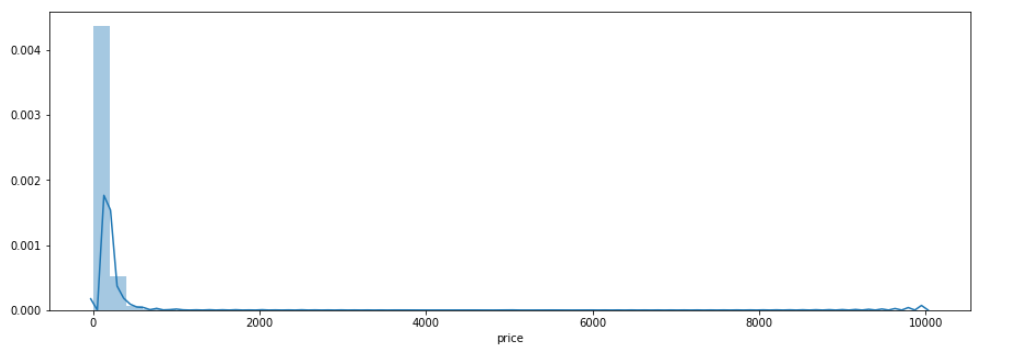


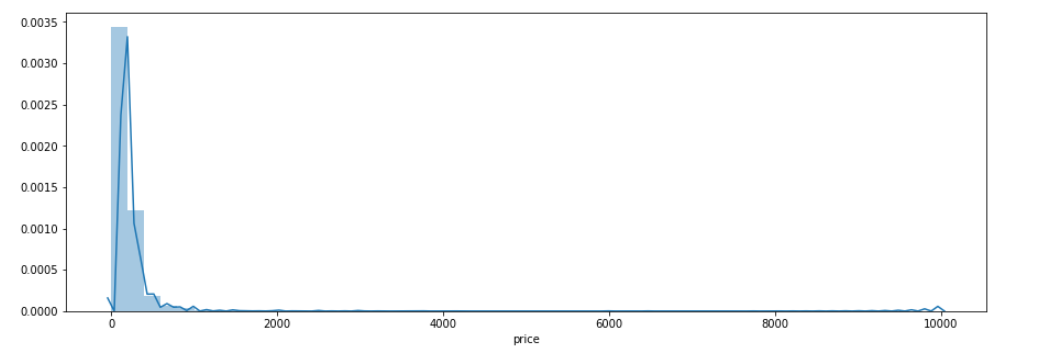
Clearly, we can see that a lot of people love to hand out in Brooklyn, cause after pulling more data on the neighborhood in these big cities we can see that people love staying at Williamsburg which is one of the biggest spots in Brooklyn known for their fancy boutiques and brewery.

I was able to see that many people love to get an entire room in these Airbnb that makes it a little expensive but comfortable. 51.97% of people are getting Entire Room or apartment, 45.66% rents only private room and 2.37% get shared room. Below there is a chart getting all the count by neighborhood.



To finish, I was able to get the price distribution per neighborhood, where we can see that most of the price in New York rank between $0 and $ 1000 for an entire room and of course these numbers change depending on the neighborhood.

 Brooklyn / Manhattan price distribution



**Conclusion**

To finish, we can tell that Manhattan is one of the most beautiful cities in New York which makes the price really expensive. And yet, a bit underrated compare to Brooklyn which is expensive and has a lot of reviews. The lack of time and more data kept me from getting specific reviews about what people really say about Manhattan and Brooklyn. We have some outliers that makes the average price a little higher but yet in Queens there some Airbnb with a price of $10000 per night. At the end of the project I was able to use some powerful Python libraries such as Matplotlib, Seaborn for visualizations and NumPy, Pandas for cleaning and creating data frames.