Capstone Project

Airbnb Bookings Analysis

Done by:

Rajni Shukla

CONTENTS

PROBLEM STATEMENT

- The customer can book a room through the Airbnb website. To explore more, the customer is moving to the Queens, Bronx, Staten Island, Brooklyn, and Manhattan neighborhoods.
- The customer requires assistance in identifying a number of bookings in each neighboring country and guiding that customer in selecting the best place to live at an affordable cost, analyzing the room's price so that the customer can afford it, and selecting the appropriate room type based on their needs.
- The quantity of reviews, the number of reviews per month, and the availability of the rooms will all meet these criteria.
- All of these insights will be obtained through EDA and a thorough comprehension of the data.

APPROACH

- We live in an era where data is produced and circulated in an enormous amount. Those data can be collected and allow us to infer meaningful results and make well-informed decisions. However, as the number of data increases, we need to visualize the data to help us in conducting data analysis. By using visualization tools, we can deliver a message to our audience and inform them about our findings. This project explores a dataset from a technology company, maps the result clearly through visualization tools, and gives new insight to the public and other relevant parties.

- We will explore and visualize the dataset from Airbnb in New York using basic exploratory data analysis techniques. We will find out the distribution of every Airbnb listing based on their location, including their price range, room type, listing name, and other related factors.

- Our goal here is to explore the data and find useful insights from the data and find out different relations between the columns.

INTRODUCTION

- Airbnb is an online marketplace for lodging, principally homestays for vacation rentals, and tourism activities based in the United States. Airbnb does not own any of the houses advertised; instead, it makes money from commissions on each booking.
- Airbnb is a database of accommodation bookings. This data is only available for New York City from 2011 to 2019.
- We can grasp all of the features by exploring the data. The objective is to look into the data and then evaluate it with all of the necessary information.
- We have extracted crucial insights from the data through data exploration and data analysis.
- Airbnb is an abbreviated version of AirBedandBreakfast.com company's original name.

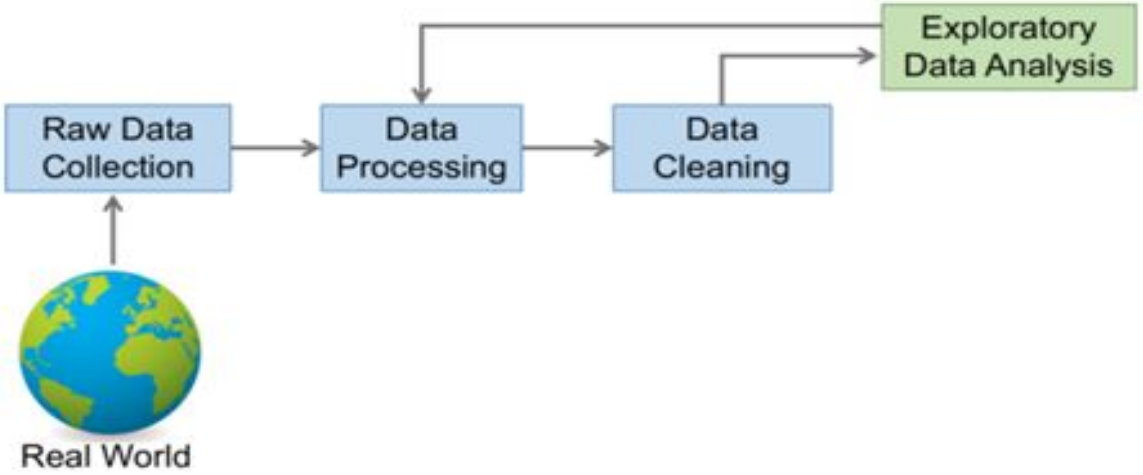
DATA SUMMARY

The dataset has 16 different features with more than 49,000 observations. The most important features of given dataset are

- **Host id:** It is the id given to specific host and there are 1270 host id available in give dataset.
- **Neighborhood Groups:** It represents location, given data set contain 5 different location.
- **Neighborhood:** It represent specific areas where the listing is located.
- **Room type:** It represent category of room type being listed.
- **Minimum nights:** It represents number of nights spend by customer in given listing.
- **Number of reviews:** It represents the number of reviews for listings.
- **Availability 365:** It represents number of days in year for which given property is available for rent.
- **Price**: It represent rate for given room type in given location for one night.

STEP INVOLVED

1. **Acquire and loading data:** For this project, we are using Google colab a web IDE with a python programming language to write our script. To get the data, we are using Airbnb data that is publicly shared on the internet under the Creative Commons License. Before we are able to load the data into our IDE, first we need to import various external libraries/modules that are needed for visualization and analysis.



2. Cleaning Dataset:

The next step is cleaning up the data, oftentimes the data we load have various faults, such as missing value, incomplete data, etc. By cleaning up, the data quality will have better quality to be used for further analysis.

a. Checking column with missing values
b. Removing redundant variables
c. Replacing all the missing values

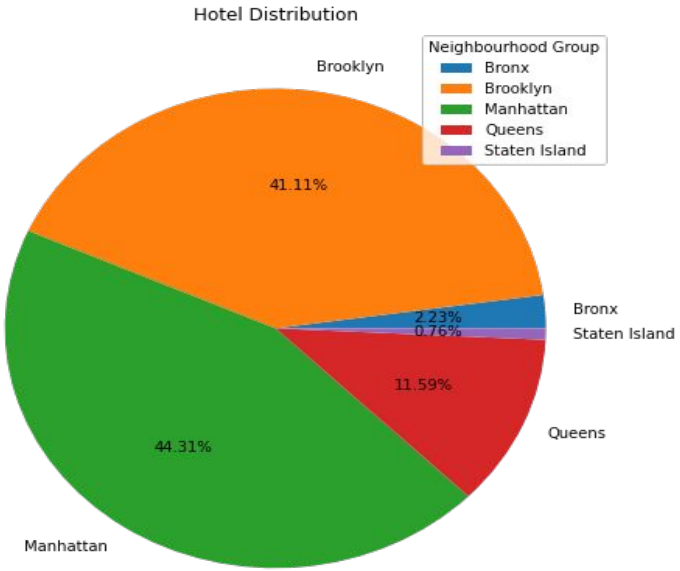
3. Exploring and Visualizing Data:

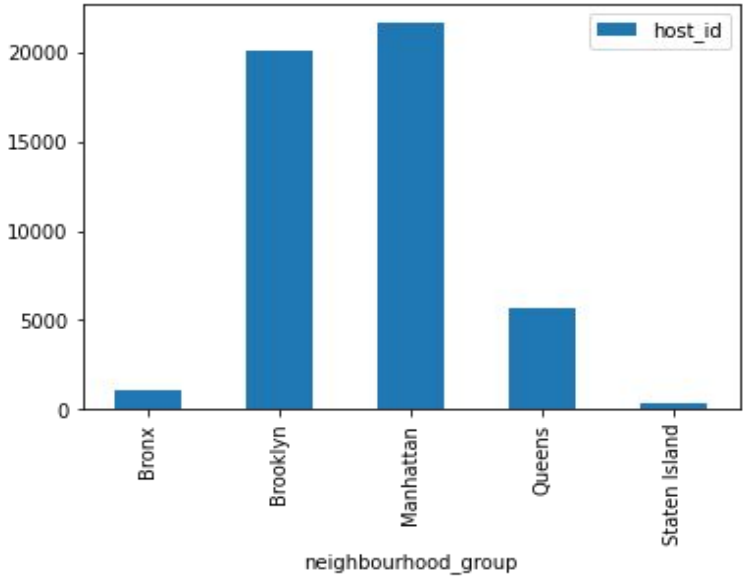
In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images.

EXPLORATORY DATA ANALYSIS

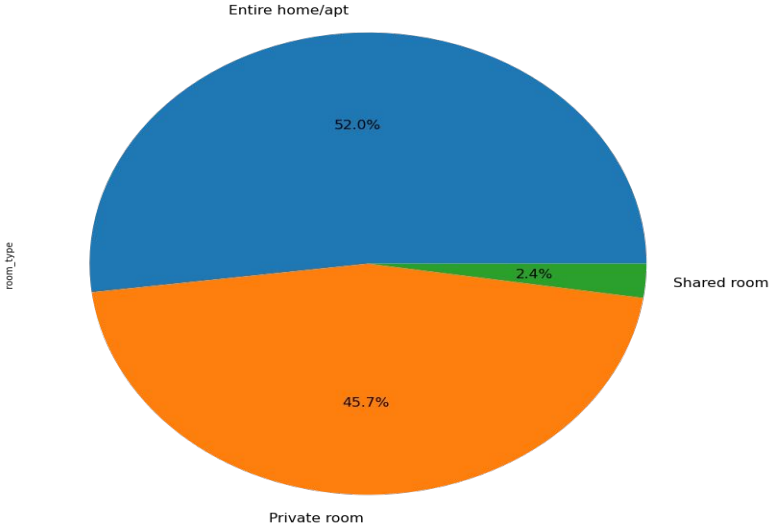
Airbnb Distribution in different Neighborhood groups.





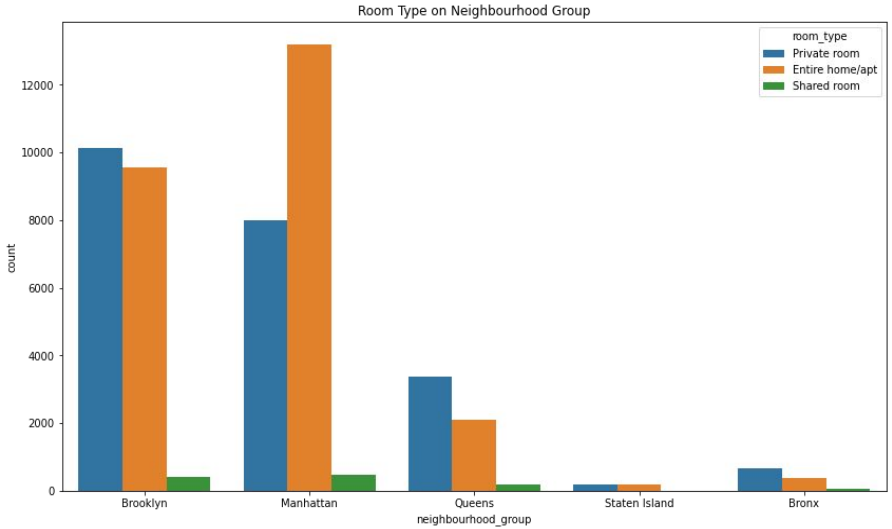
Majority of the hosts belong to the locations of Manhattan and Brooklyn hence making these the most popular destinations. These are favorite destinations for travelers.

Types of room preferred by the people



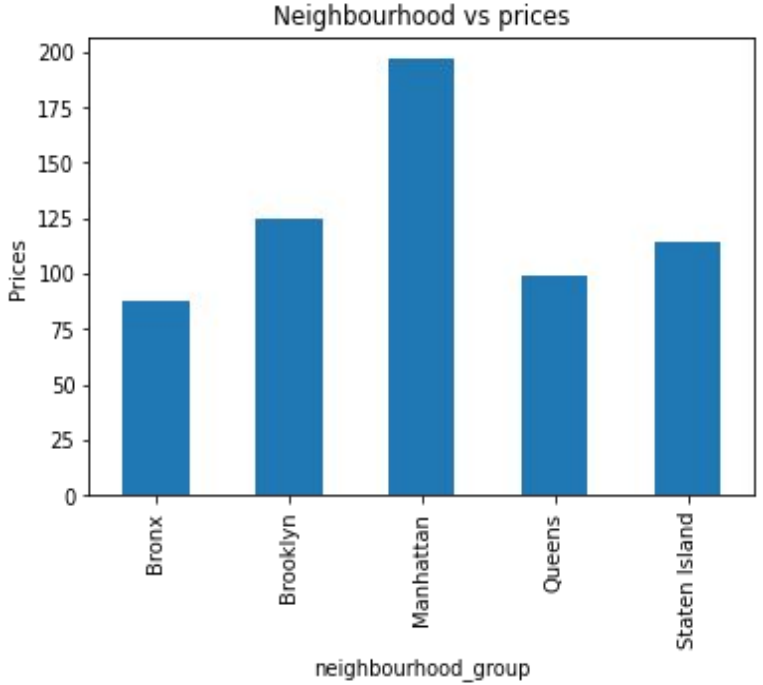
From this graph we realize that customers prefer living in Apartment or private rooms instead of shared rooms.

Distribution of Room types in various Neighbourhood groups



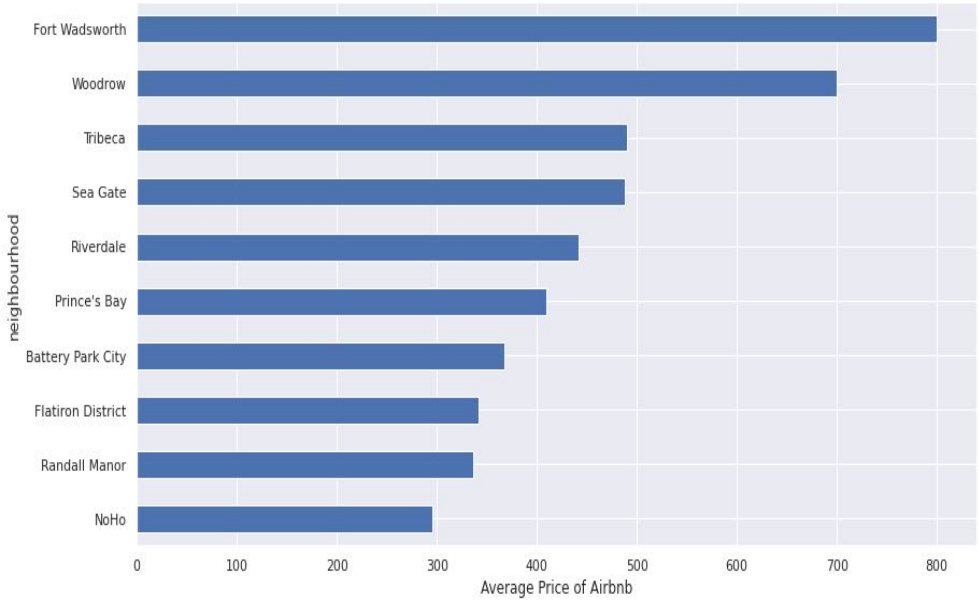
Majority hosts in Manhattan and Brooklyn provide either a Private room or Entire home/apt.

Price Distribution Vs Neighbourhood



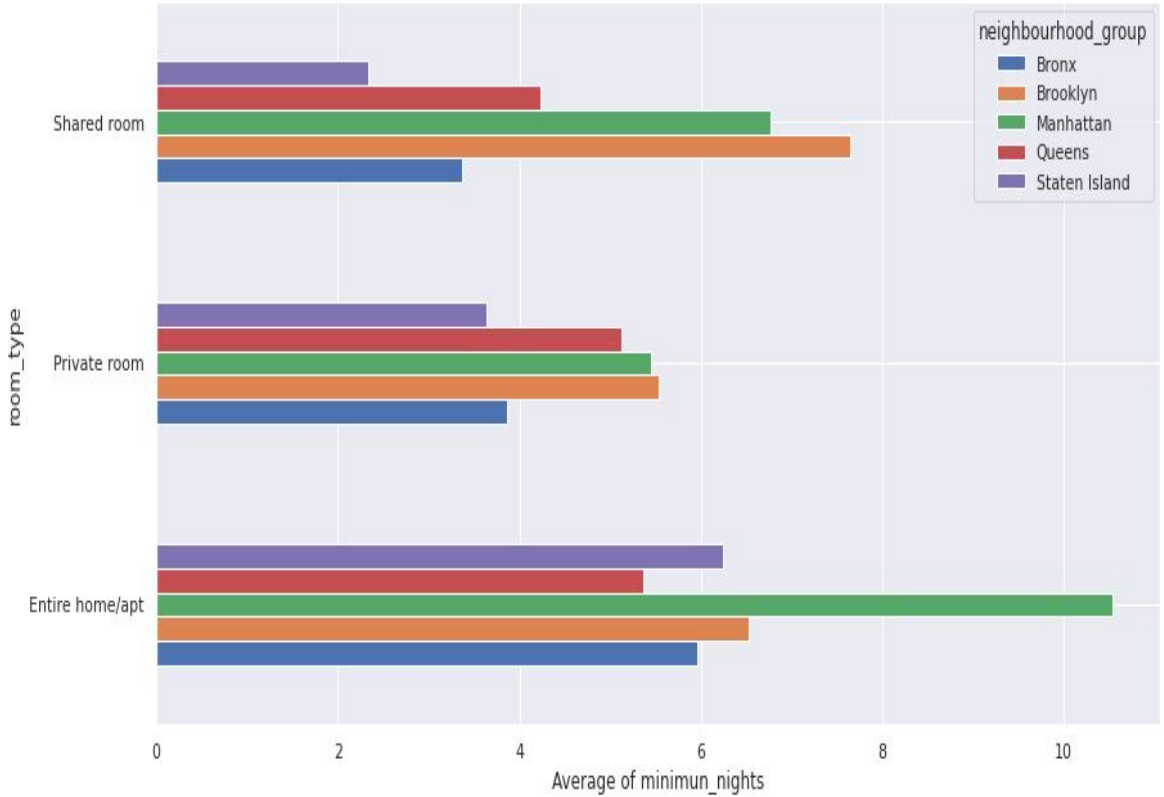
Manhattan followed by Brooklyn have most expensive properties.

Top 10 Highest Average price distribution of neighbourhood



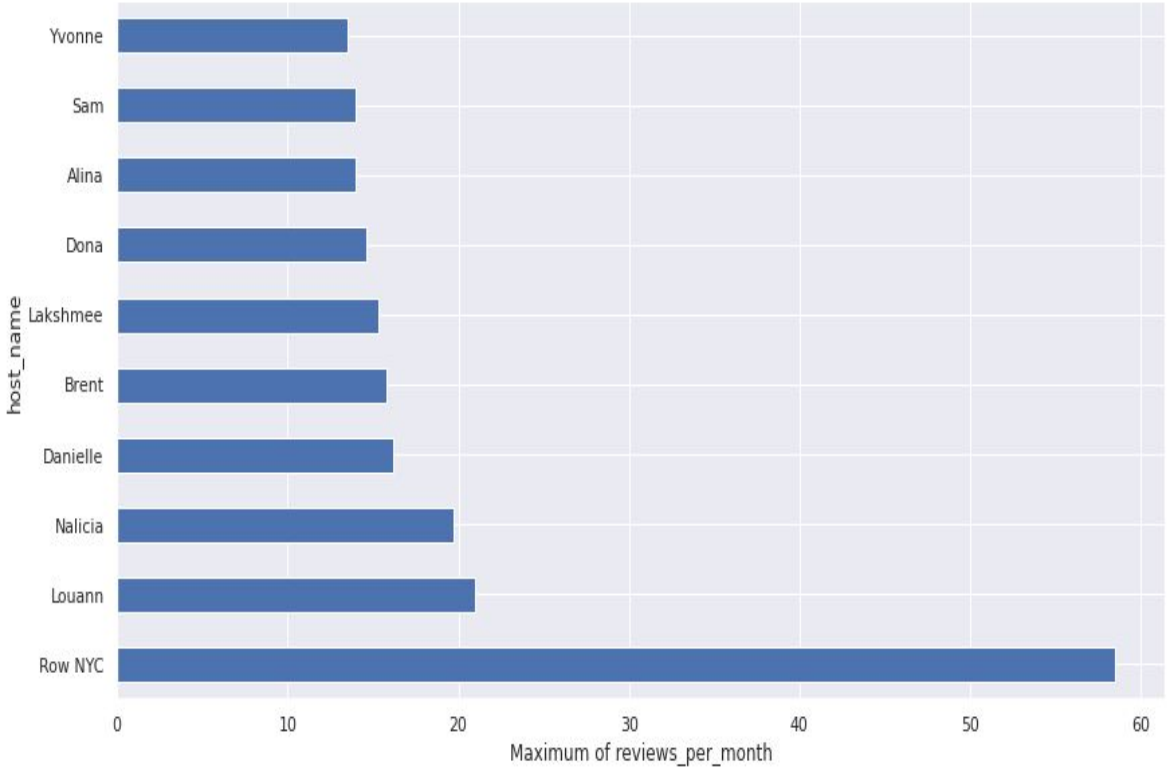
From the graph it is clear that the Fort Wadsworth of Staten Island group has the maximum average price of Airbnb.

Minimum No. of Nights people stayed in each room types



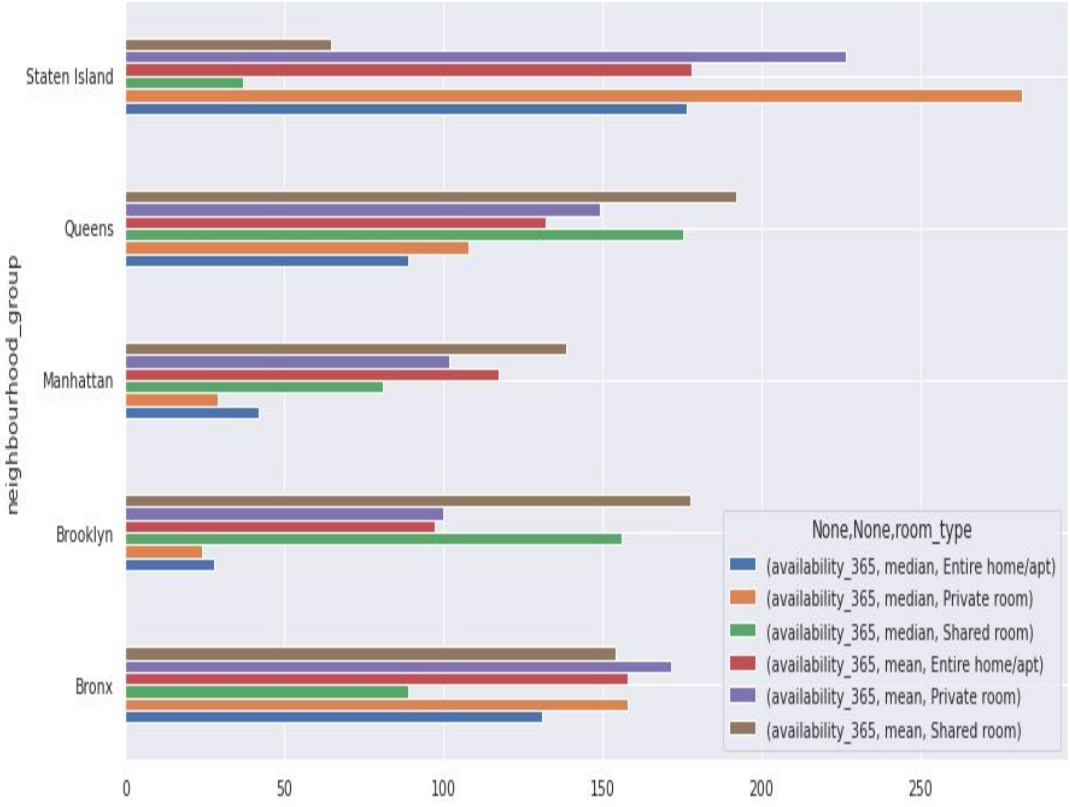
It seems like half the listings renting out Entire houses and apartments are from Manhattan only.

Top 10 reviewed hosts per month



Host id 244361589 (Row NYC) listed or booked was 58.50 which is highest followed by the others .

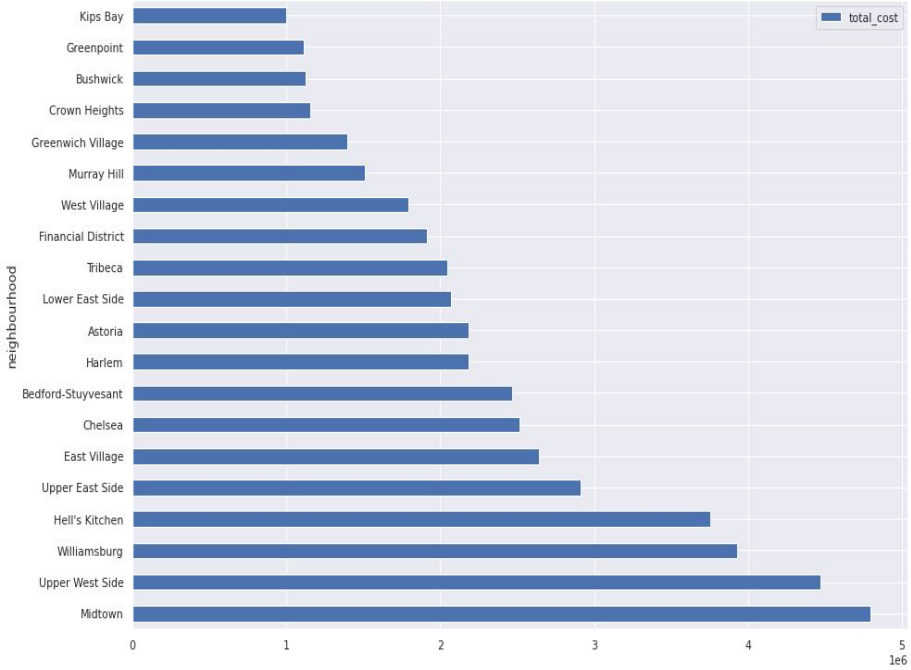
Room types and their relation with availability in different Neighborhood group



- Property available for 365 days have price maximum of 1250. Average price is 250. no of reviews on an average is 10.

- Entire houses and apartments are expensive in every neighborhood group.

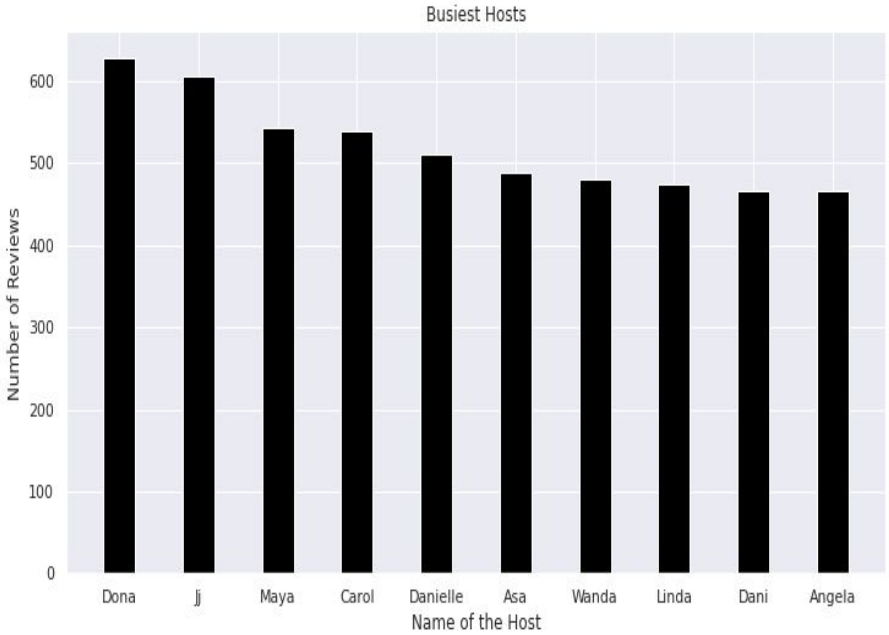
Different hosts and Areas

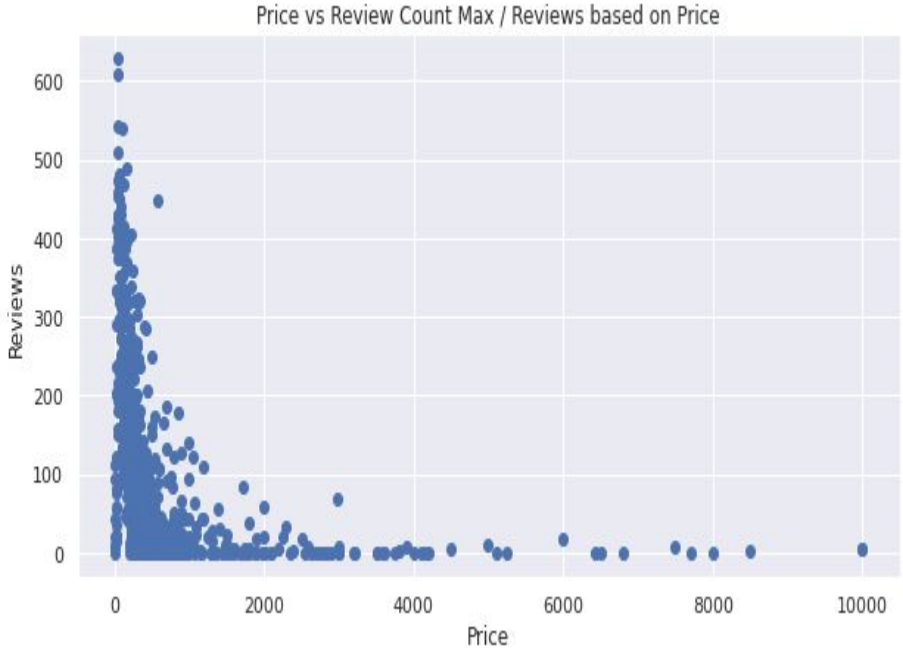


Midtown is the most popular and expensive neighborhood which makes more money.

Among all Host id 107434423 earn more money.

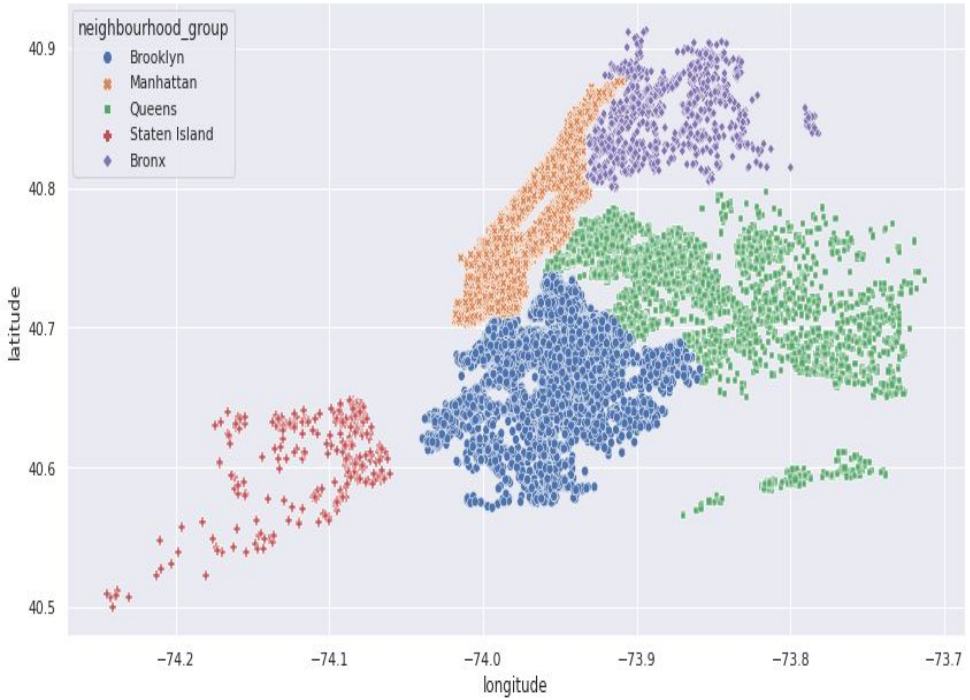
Busiest Hosts



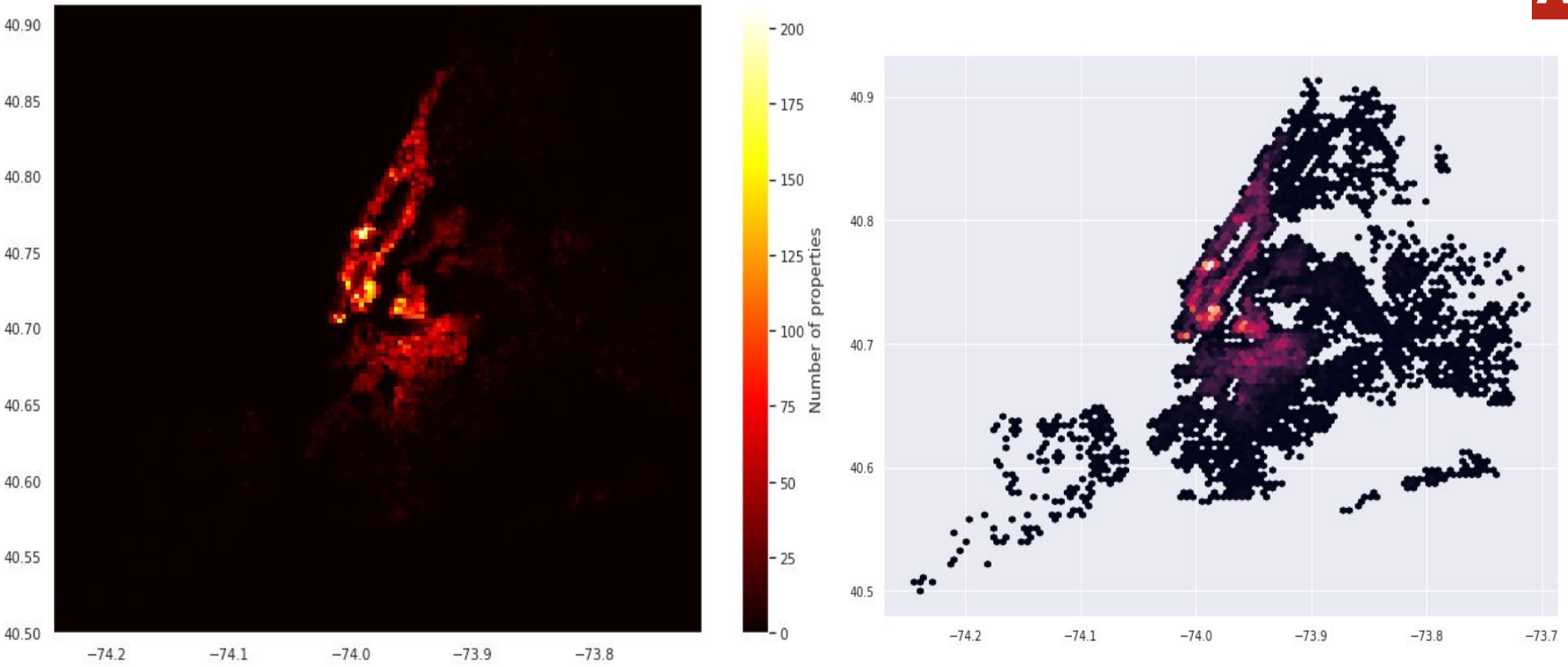


- Neighborhood groups Manhattan and Brooklyn receive the most number of reviews making the hosts from these areas the most busiest among all others.
- Lower the price higher is the no of reviews.

Relationship between the locations and Neighborhoods

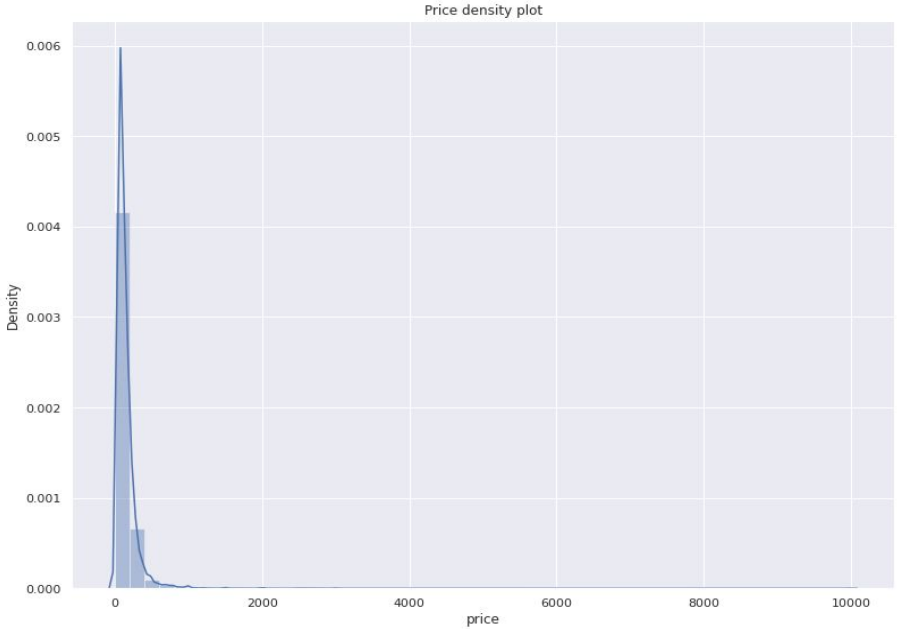


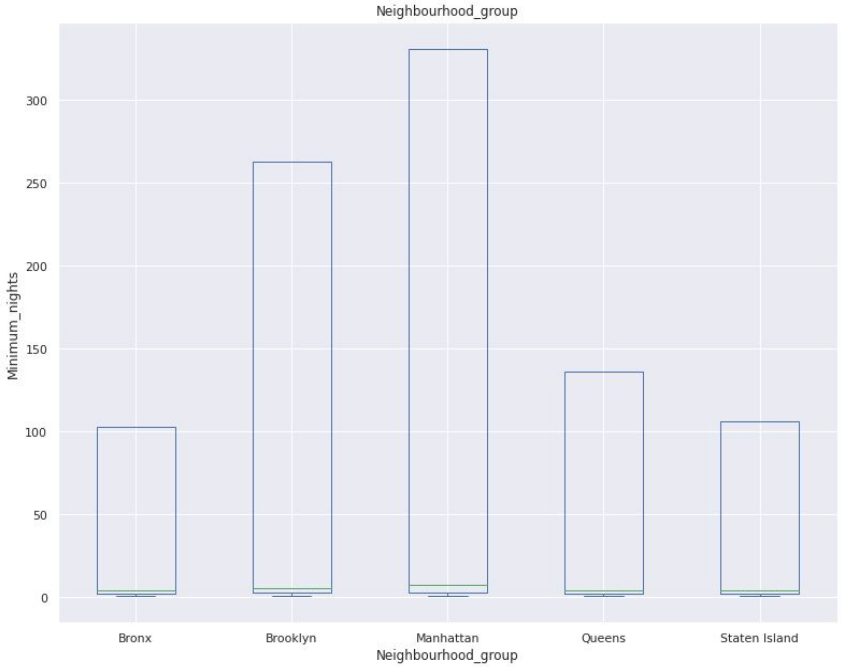
Density of hotels is large in manhattan and places in close contact of manhattan financial district.



Most of their listing in lower manhattan or are near airport and probably focus on tourist rentals. All of these leads to the places to be looked at throughout the year, which implies people who want these places may have to look months ahead, which leads to traffic in these locations.

Price range of rooms are more





Manhattan followed by Brooklyn have the most expensive properties on average. Density of hotels is large in Manhattan and places in close contact with Manhattan.

CONCLUSION

Starting with loading the data, So far we have done identification of variables and data types, Missing value treatment, Non-Graphical and Graphical Univariate Analysis, outlier treatment, and learned that-

- There are 36 percent of inactive hosts. Which are directly proportional to the density of listings in the neighborhoods.
- Manhattan and Brooklyn are the most crowded boroughs of New York City as compared to Staten Island, Queens and Bronx so increasing the number of hosts in Manhattan and Brooklyn will help in generating more revenue.
- Availability of the room in a year does not affect the reviews of the data set.
- Manhattan's listings are most expensive so offering special discounts in the peak season might help in attracting more customers.
- Customers prefer to stay in Entire home/apt or private room types as compared to shared room types so Airbnb should strategize on making more entire and private rooms available.

CONT…

- Density of hotels is large in Manhattan and places in close contact with Manhattan.
- By pushing the popular (most reviewed) hosts listings forward Airbnb can increase its user's experience.
- Ask for suggestions from most reviewed hosts and as well as from customers to enhance their experience.
- Lower the price higher is the no of reviews.
- Average price of Airbnb in New York is 152 and the maximum price is 10K.
- Advertising more about the famous sites in Queens, Bronx and Staten Island to tourists can boost revenue in these regions as well.

THANK YOU