EDA On Airbnb Booking Analysis

by
Shaik Ahmad Basha
Data science trainee,
AlmaBetter, Bangalore.

Abstract:

Airbnb, as in "Air Bed and Breakfast," is a service that lets property owners rent out their spaces to travelers looking for a place to stay. Travelers can rent a space for multiple people to share, a shared space with private rooms, or the entire property for themselves.

Airbnb is based on a peer-to-peer business model. This makes it simple, easy to use, and tends to be more profitable for both parties. The model also gives you the opportunity to customize and personalize your guests' experience the way you want.

1.Problem Statement

For this project, we will be analyzing Airbnb's New York City (NYC) data of 2019. This dataset contains listings information such as listing name, host name, room types, minimum night stays, availability, area, reviews etc.

Our main objective behind this project is to explore and analyze the data to discover the key understandings. For this, we will explore and visualize the dataset from Airbnb in NYC using basic exploratory data analysis techniques.

2.Data Description

The data has 48895 rows and 16 columns. Those 16 columns are:

1. **id:** Unique id of listing
2. **name:** Name of the listing
3. **host_id:** Unique id of host
4. **host_name:** Name of the host
5. **neighbourhood_group:** Location where listing belongs to.
6. **neighbourhood:** Area of the listing
7. **latitude:** Latitude of listing
8. **longitude:** Longitude of listing
9. **room_type:** Type of rooms (Entire home, private room, shared room)
10. **price**: Price of listing
11. **minimum_nights:** Minumum number of nights people stay
12. **number_of_reviews:** No. of reviews given for listing
13. **last_review:** Latest review given for listing
14. **reviews_per_month:** No.of review given per month
15. **calculated_host_listings_count:** Total No. of listings for host
16. **availability_365:** No. of days listing available

3. Introduction:

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analysed and used for security, business decisions, understanding of customers' and providers(hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

4. Data Understanding

After the loading and collecting the dataset, understanding the data is very important. I Understand the various features of the dataset and their meaning and what are the numerical and categorical features.

5. Data Cleaning

Our dataset contains large number of null values. Out of 16 features, only 4 features have null values. Two features named ‘last_review’, ‘reviews_per_month’ has 10052 null values which is more than 20% of the data. So I removed those features from dataset.

The features named ‘latitude’, ‘longitude’ has no effect for our analysis. So I removed those features also.

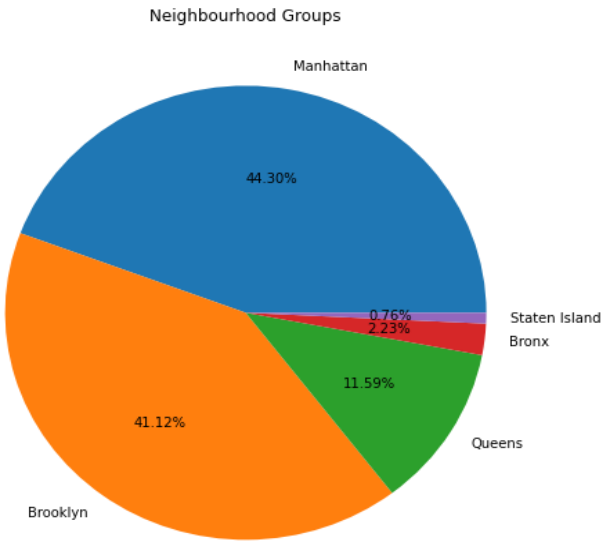
6. Exploratory Data Analysis

After removing the unnecessary features, the data is ready for Exploratory Data Analysis. I started the analysis by dividing the features into numerical and categorical. I checked the multicollinearity in the data and observed the relation between numerical features and dependent feature (price). And then I started to explore categorical features and how these categorical features vary with numerical features and drawn some conclusions from it. And I visualize those observations using bar chart, scatter plot, violin plot, pie chart, multiple bar chart.

Following are some analysis which I have done.

1. **Which location has more number of listings?**

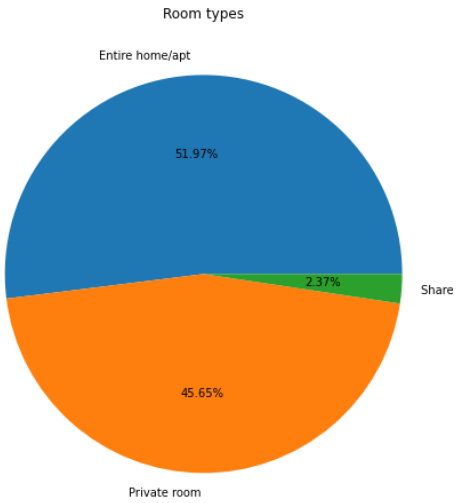
   There are 5 locations namely Manhattan, Brooklyn, Queens, Bronx, Staten Island



Here Manhattan has high number of listings

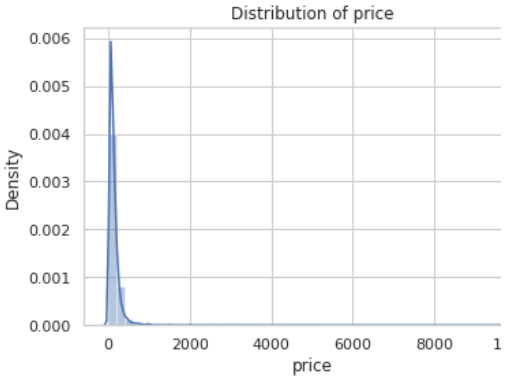
2. **Which room type has more number of listings?**

There are 3 unique room types namely Entire home/apt, Private room, Shared room



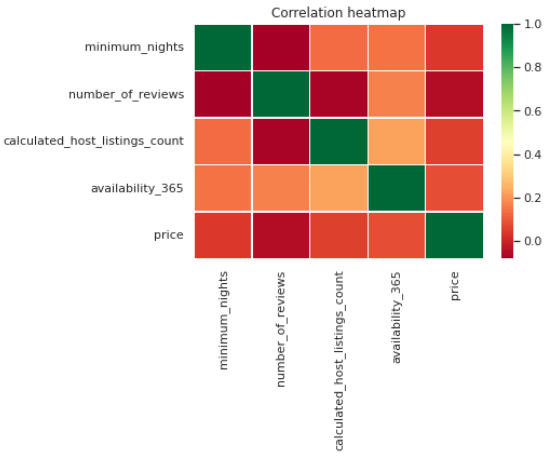
Here, Entire room category has more listings followed by private room and the least is Shared room.

3. **How price of listing varies?**



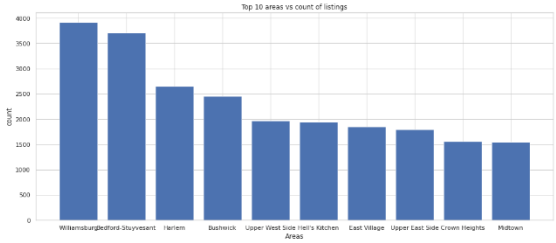
Here we can observe that price is mostly in the range of 0 to 1000 and a few are in range of 1000 to 10000

4. **Is any multicollinearity exist in the features?**



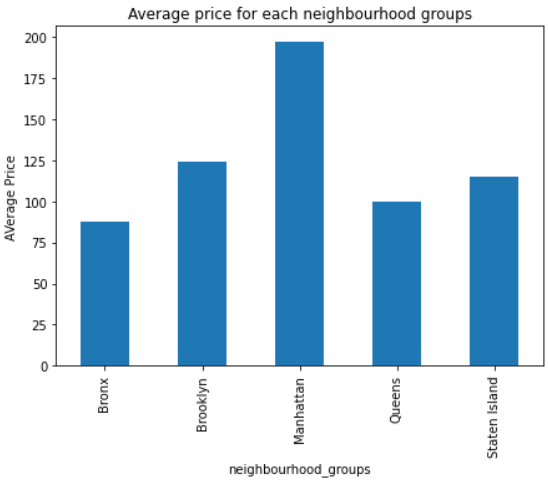
From the above heatmap we can see there is no multicollinearity in our features

5. **What are the top 10 areas with most number of listings?**



From the above bar chart, the names on X-axis are the areas with most number of listings.

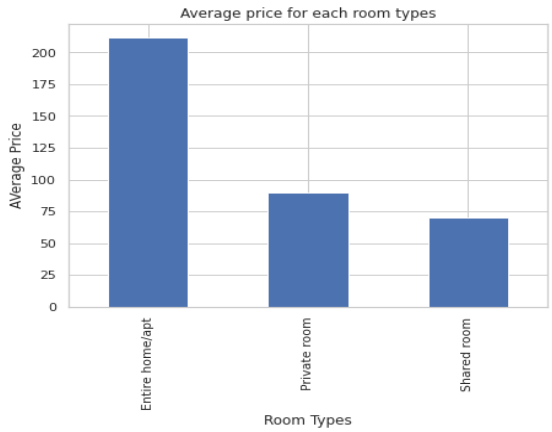
6. Which location has highest average price?



From the above visualization, we can observe that Manhattan has highest

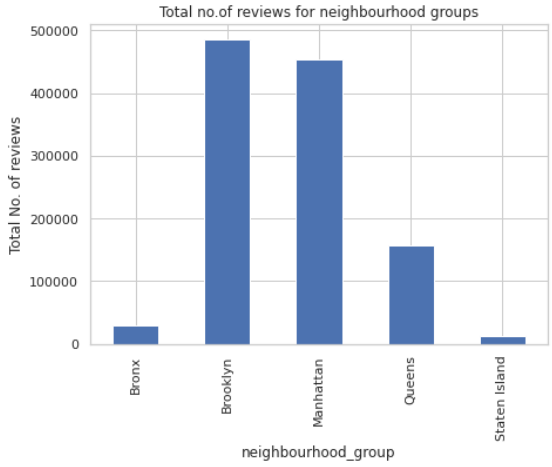
average price followed by Brooklyn and least is Bronx.

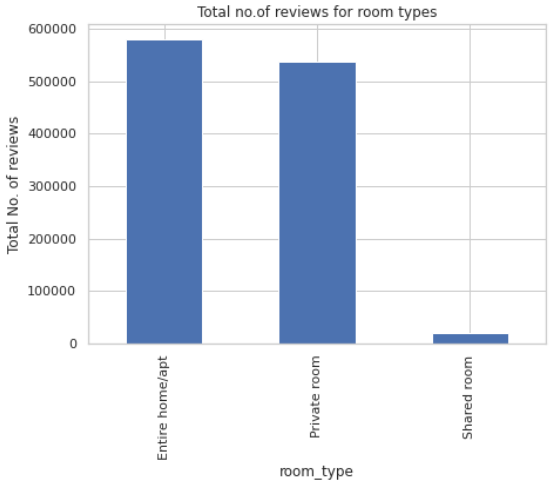
7. What type rooms have highest average price ?



From the above visualization, we can observe that Entire home/apt has highest average price followed by Private room and the least is Shared room.

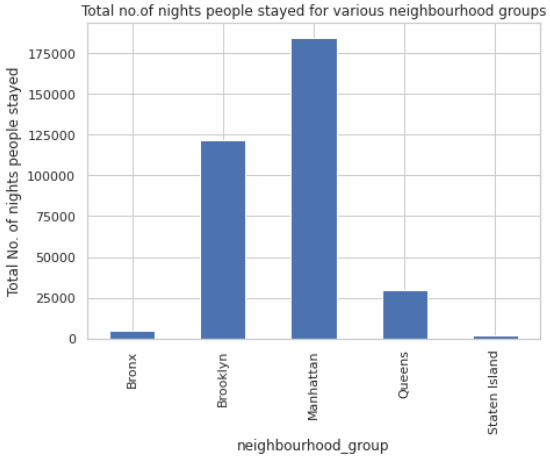
8. **Which location and room type have more reviews?**

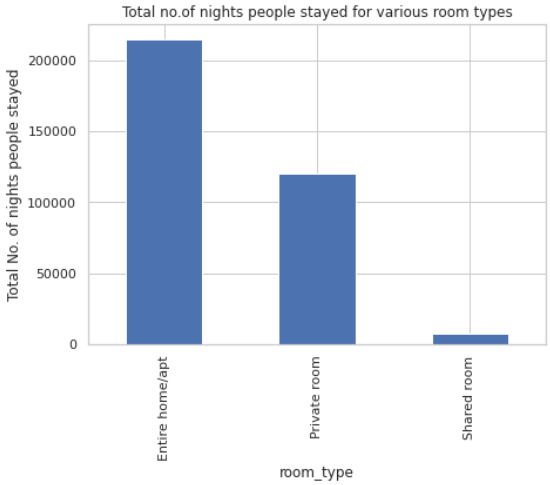




From above visualizations, Brooklyn and Entire home/apt have more reviews.

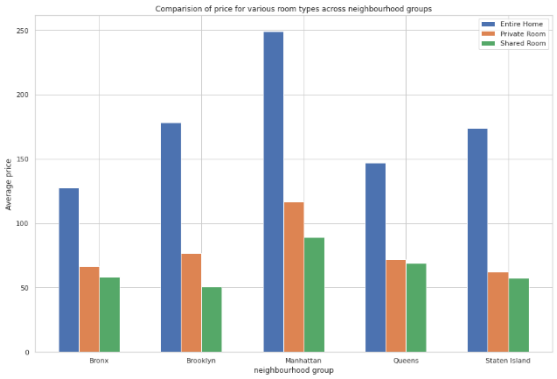
9. **What locations and room type have more night stays?**



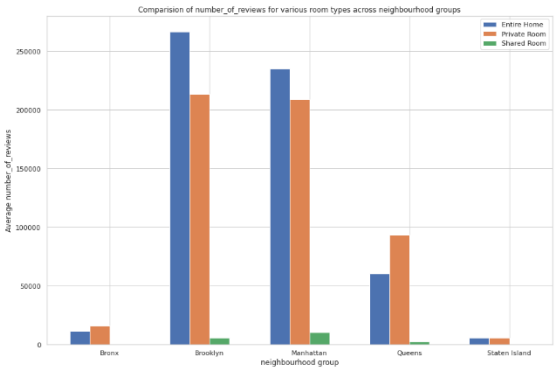


From the above visualizations, we can say that Manhattan and Entire room/apt has more night stays.

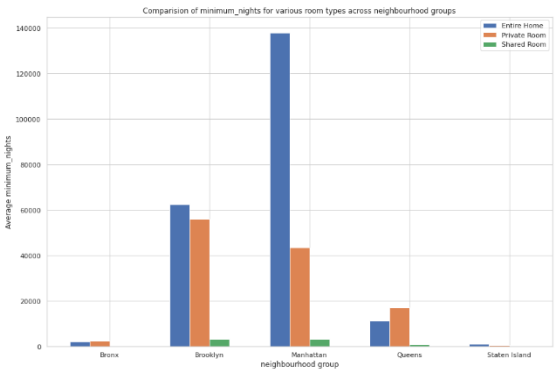
10. How can you visualize the variation of price of room types for each location?



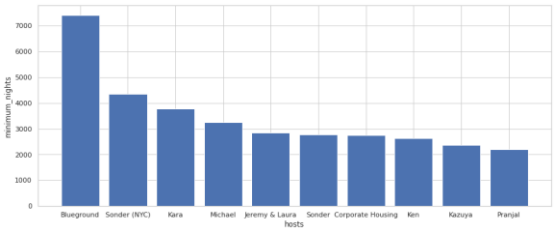
11. How can you visualize the variation of number of reviews of room types for each location?



12. How can you visualize the variation of night stays of room types for each location?

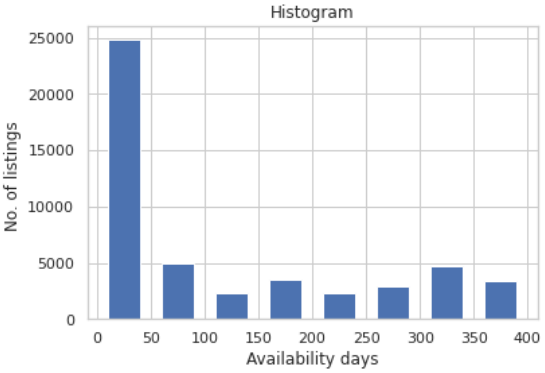


13. Who are the busiest hosts?



The busiest hosts are Blueground, Sonder (NYC), Kara, Michael, Jeremy & Laura, Sonder, Corporate Housing, Ken, Kazuya, Pranjal.

14. How many listings are available for 365 days?



From the above visualization we can say that only 1294 listings are available for 365 days.