Capstone Project

Airbnb Bookings Analysis

Team Data Breakers

Shraddha Shandilya

Contents

What Is Airbnb?

- Airbnb, as in “Air Bed and Breakfast,” is a service that lets property owners rent out their spaces to travellers 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.
- The Airbnb Dataset Lists the number of spaces available with 16 columns describing various attributes.

Who books on Airbnb?

Knowing your demographic is imperative when deciding on a new booking site. The general traveler booking on Airbnb will be a young, tech-savvy millennial; it is estimated that millennials account for roughly 60% of all guests that have ever booked on Airbnb. These travelers are particularly attracted to authentic, local properties. They believe accommodations listed on Airbnb will provide them with more cultural experiences, which is a top priority for them when they travel. Its popularity stems from the fact that not everyone who is traveling can afford to stay in a hotel, much less find a hotel room in a busy, urban area. Airbnb provides them with a simple — and often less expensive — alternative.

Problem Statement

- 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 analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.
- This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

Explore and analyze the data to discover key understandings.

Data Summary

Id- It provides a unique id to each listing in the dataset.

Name-It describes the property in a concise manner.

Host_id-It is used to ascertain the ownership of a certain individual

Host_name-It is the name of the property owner

Neighbourhood_group-It basically categorizes neighbourhood into five groups Bronx,Brooklyn, Manhattan,Staten Islands and Queens

Neighbourhood-It designates the location of a property.

Lattitude-It Specifies the Lattitude of the property.

Longitude-It denotes the Longitude of the property.

Room_type-It classifies the type of property into Private_Room,Shared_Room and Entire_Apartment.

Data Summary

Price- It provides the price of each listed property.

Minimum_Nights-It gives the minimum night one needs to pay for in the property.

Number_of_reviews-It is used to ascertain the number of reviews received by a property.

Reviews_per_month-It denotes the number of reviews per month.

Calcuated_host_listings_count-It corresponds to the number of properties hosted by the unique host_id

Availability_365-The number of days the property is available in a year.

Data Collection

- We already have the data set required to perform our exploratory data analysis on. We just need to load it and perform the necessary cleaning and then start with our analysis.

Data Cleaning

Before starting with our data analysis, we need to perform the necessary cleaning . It mainly involves-

- Looking for and handling Nan/Null/missing values in the data set.
- Dropping any columns which aren’t needed or useful for our data analysis.

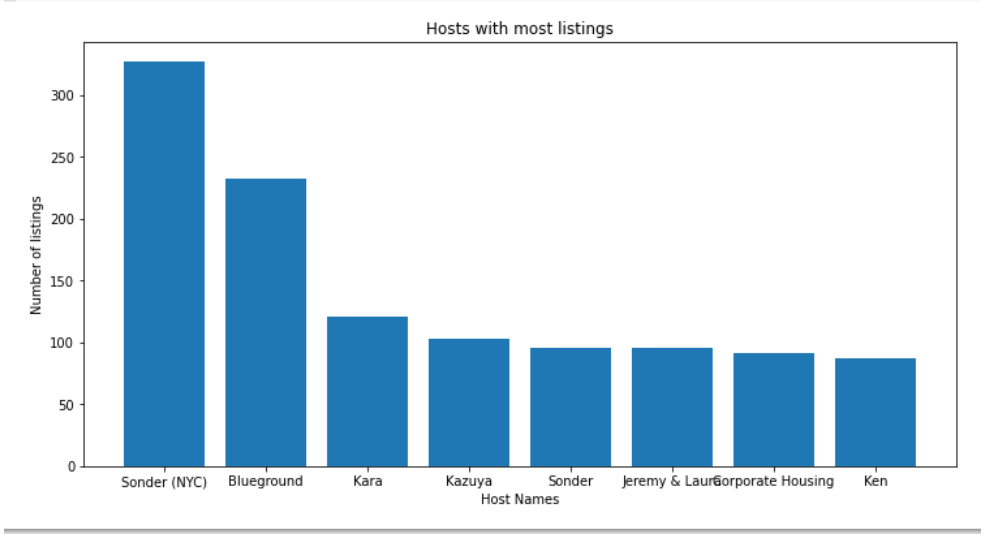
This is done to make our data analysis process easier.

Performing EDA

Exploratory Data Analysis (EDA) is an approach to analyze the data to get useful insights. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

We have performed EDA on our given Airbnb dataset for seeing what the data can tell us by data visualisation which includes different types of graphs.

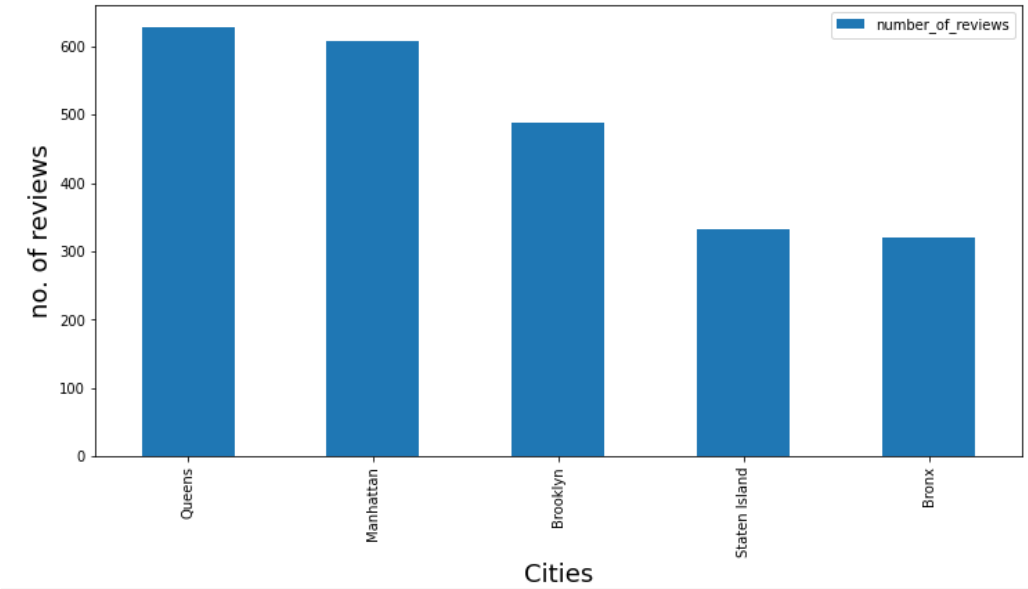
EDA



host_name	neighbourhood_group	calculated_host_listings_count
Sonder (NYC)	Manhattan	327
Blueground	Manhattan	232
Blueground	Brooklyn	232
Kara	Manhattan	121
Kazuya	Queens	103
Kazuya	Manhattan	103
Kazuya	Brooklyn	103
Sonder	Manhattan	96
Jeremy & Laura	Manhattan	96
Corporate Housing	Manhattan	91
Ken	Manhattan	87
Ken	Brooklyn	87

Our analysis gives us the list of top 10 hosts with the maximum number of listings in Airbnb.

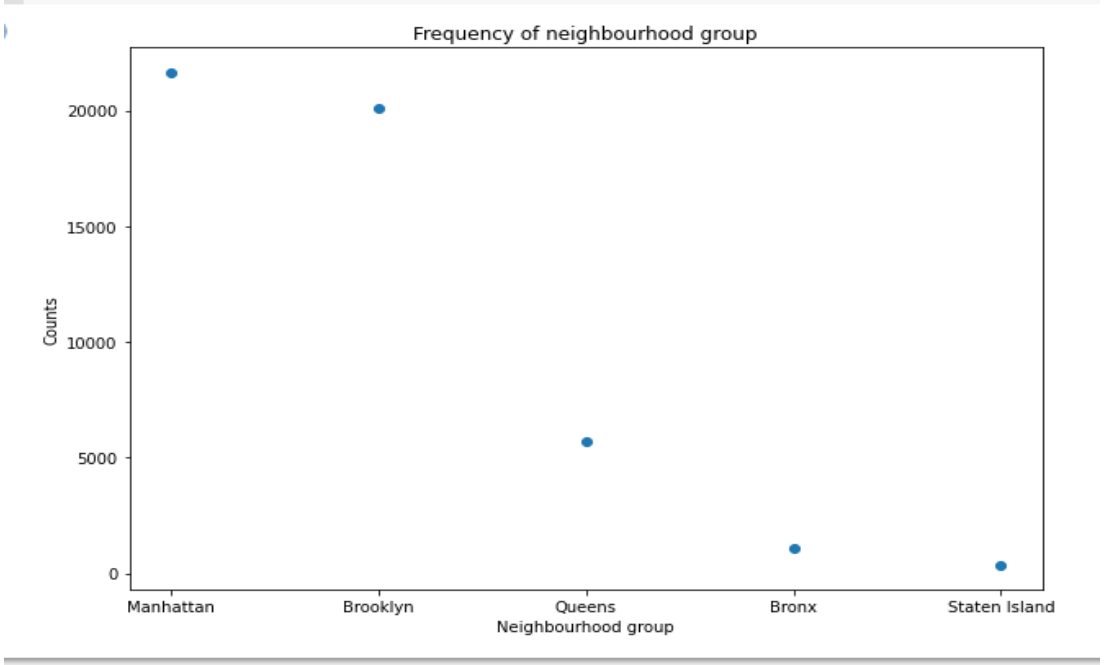
EDA



neighbourhood_group	number_of_reviews
Queens	629
Manhattan	607
Brooklyn	488
Staten Island	333
Bronx	321

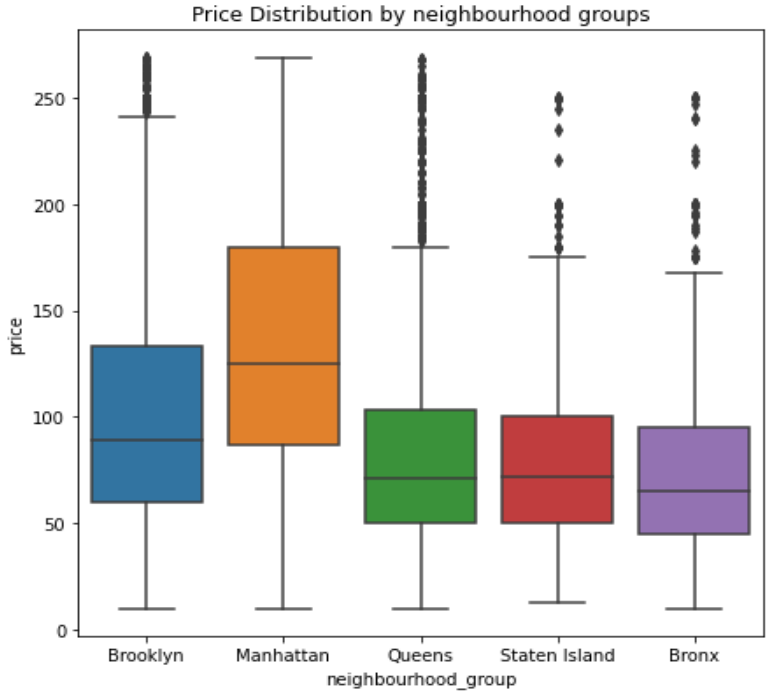
Our analysis gives us the neighborhood group with the most reviews.

EDA



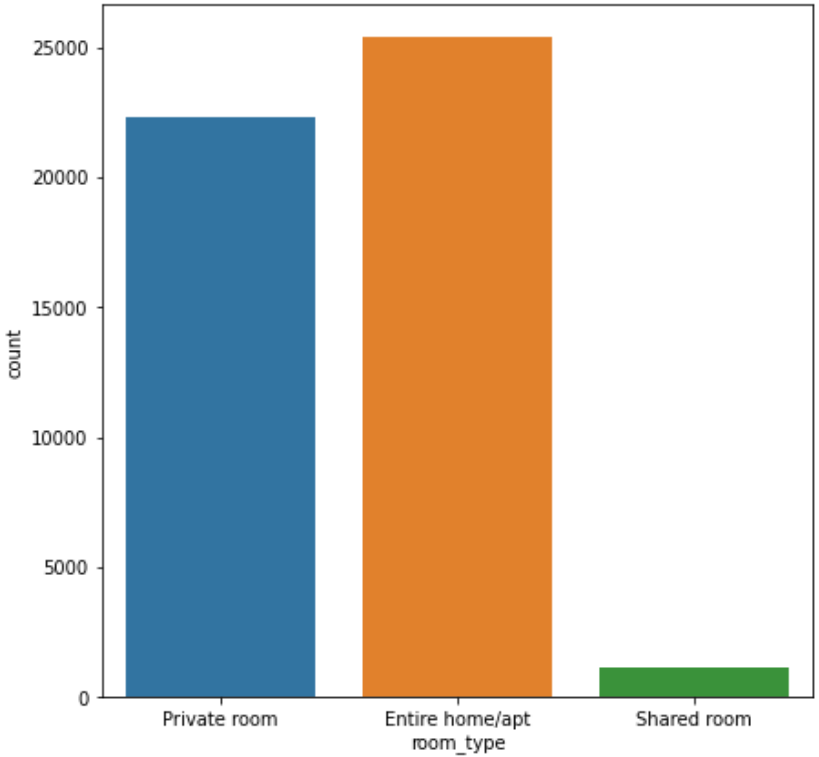
Graph shows the neighborhood group with maximum frequency of listings.

EDA



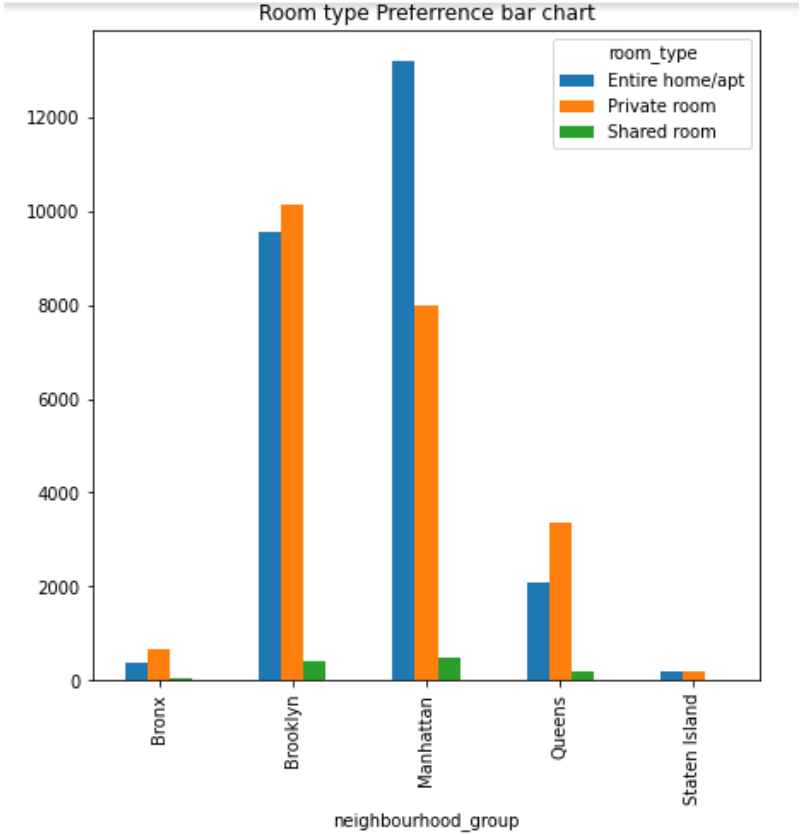
We observe that Manhattan has the highest range of prices with average price $150, followed by Brooklyn with an average price $90 per night. Queens and Staten Island have similar distributions, Bronx is the cheapest of them all.

EDA



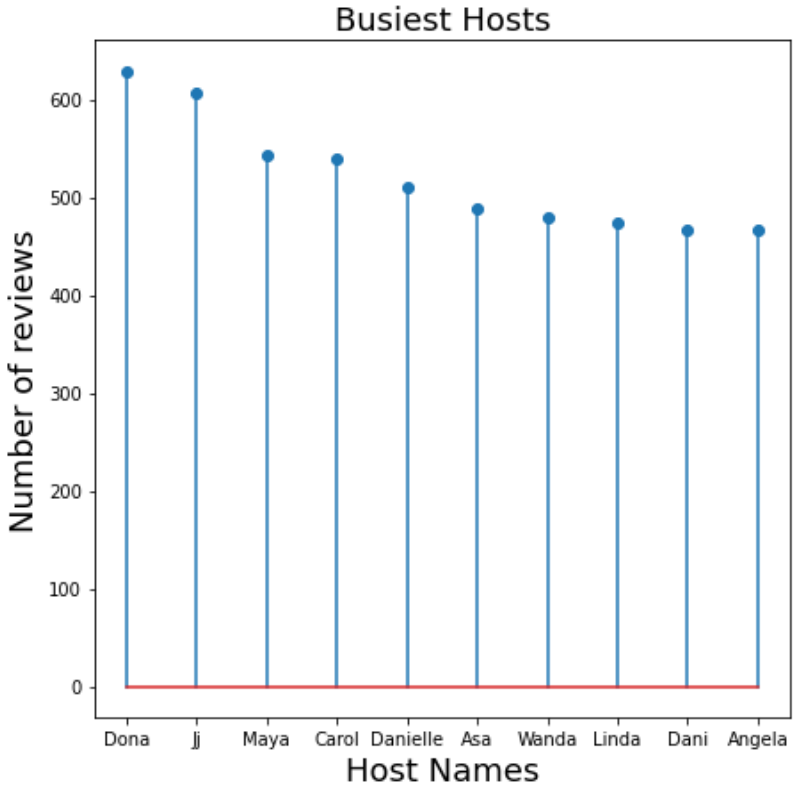
Our analysis gives us the most available room type.

EDA



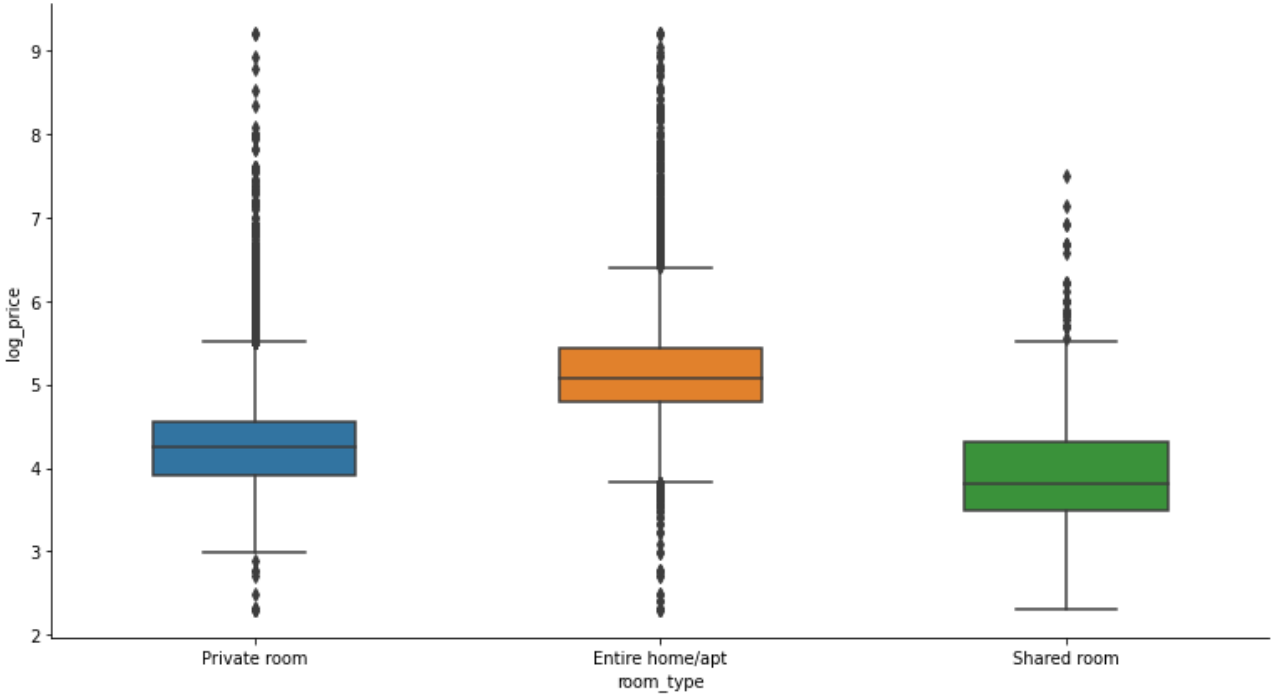
The most preferred room type in different neighborhood groups.

EDA



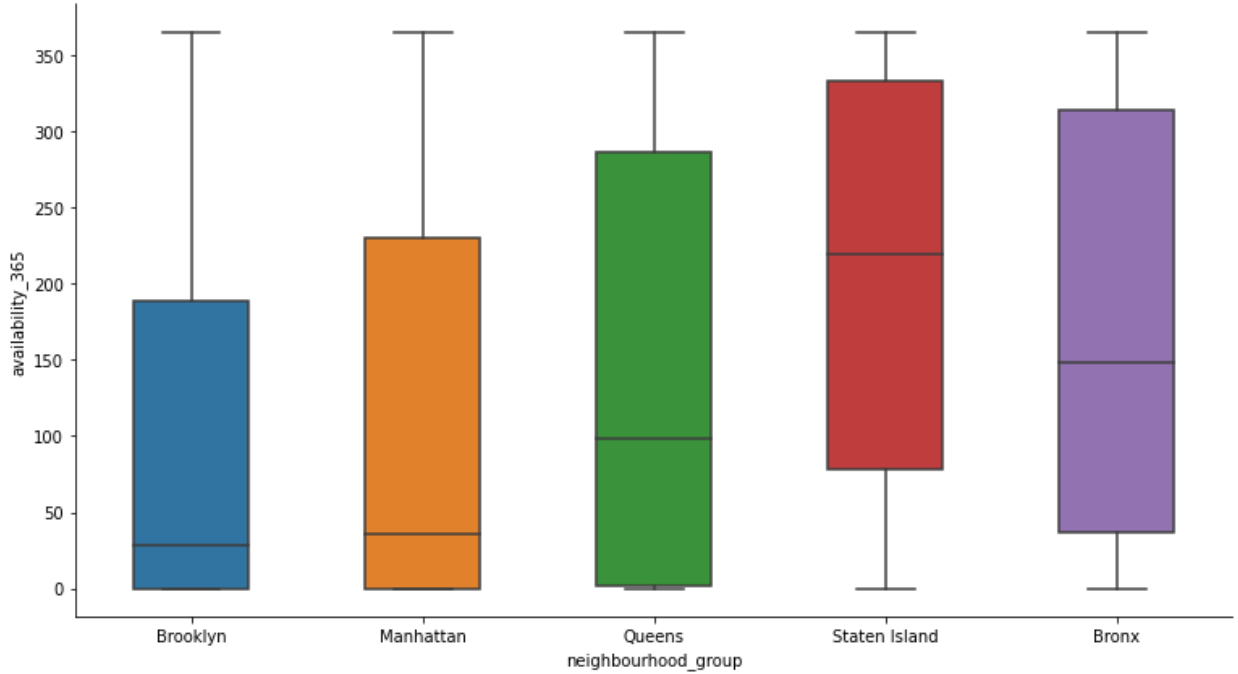
Our analysis gives us the 10 busiest hosts out of all the hosts.

EDA



Relationship between price of listings and type of room.

EDA



Relationship between neighborhood groups and their availability for no. of days.

Observations

- ❑The host named Sonder (NYC) has the most no. of listings.
- ❑Queens has majority of reviews followed by Manhattan.
- ❑The frequency of Manhattan is the maximum.
- ❑Manhattan has the maximum number of busy hosts.
- ❑Manhattan has the highest range of prices with average price $150, followed by Brooklyn with an average price $90 per night. Queens and Staten Island have similar distributions, Bronx is the cheapest of them all.
- ❑The most available room type is entire home/apt room.

* The most preferred room type in Manhattan is entire home/apt.
* The most preferred room type in Brooklyn is Private room.
* The most preferred room type in Bronx is Private room.
* The most preferred room type in Queens is Private room.

Observations

* The most preferred room type in Staten Island is Private room.
❑The busiest hosts would be those who have maximum number of reviews as people are booking frequently at those hosts.
❑Top 10 busiest host are:
❑Dona, JJ, Maya, carol, Danielle, Asa, Wanda, Linda, Dani and Angela.
❑Using the box plot, we checked how price changes according to the type of the room. Entire home/apt room type is costlier than private room and shared room type category.
❑People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.

.

Conclusion

In the given Airbnb dataset, firstly we did data cleaning in order to make the data suitable for our data analysis. Then, we explored the Airbnb dataset focusing on different features of the dataset. We tried to find out the relationship between different features and how they impact the bookings in different areas.

The data analysis would have been better if we had other features in the dataset like the amenities offered, the number of people that can be accommodated in the place, the time taken by the host to respond once a booking is done, reviews on the basis of stars scale like 5 star, etc.

We have performed the exploratory data analysis on the given dataset and hope that the insights prove out to be helpful.