Airbnb Booking Analysis

Team - Data Scientist Squad

Navjot khatri

Abstract:

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.

Key words:

Data analysis

Business decisions,

Understanding of customers and providers,

Performance,

Traveling possibilities

Introduction:

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.

Data and information:

There is information about the name of accommodation and first name of the accommodated person, accommodation area, longitude and latitude, room type, its price in dollars, minimum number of nights spent here, number of reviews the accommodation, number of days when listing is available for booking. Total is analysed 48895 accommodation spaces, 44% of them, located in Manhattan, Brooklyn 41%, 15% in other parts. It is about entire home/apt (52%), private room (46%) or shared room (2%).

Steps involved:

- **Manual data analysis**

  After loading the dataset we performed this method by comparing our target variable. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

- **Null values treatment**

  Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

- **Drop some columns**

  We have 4 columns contain a missing value. We dropped unnecessary columns

  Such as id, name, last review

```
id                                0
name                             16
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
last_review                   10052
reviews_per_month             10052
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

```
host_id                           0
host_name                        21
neighbourhood_group               0
neighbourhood                     0
latitude                          0
longitude                         0
room_type                         0
price                             0
minimum_nights                    0
number_of_reviews                 0
reviews_per_month                 0
calculated_host_listings_count    0
availability_365                  0
dtype: int64
```

- **Coding and analysing data**

  Use coding for get a desired output, establishes relations between columns and data we have

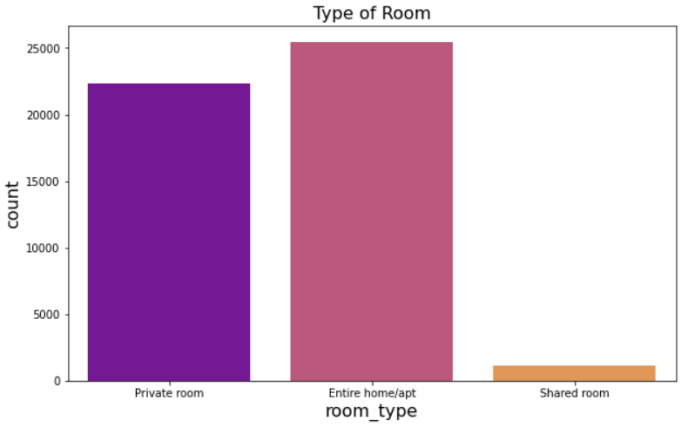
- **Plotting graph**

  Then we plot some graph for better understanding of data and visualizing the things.

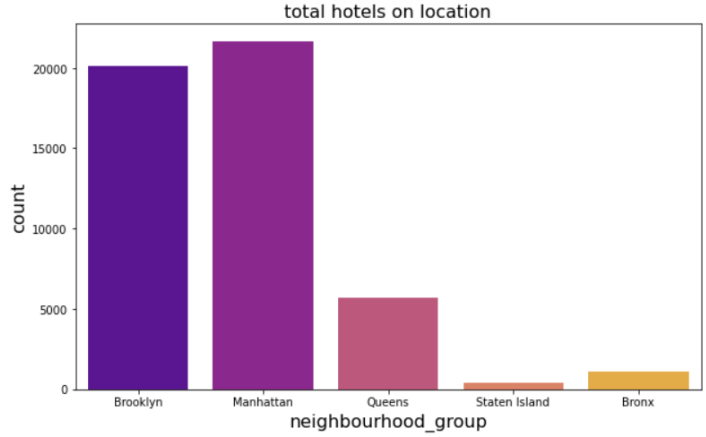
- **Conclusions**
  Finally write the conclusions for better understandings of business and what we get from the dataset

Graphical analysis of data set:

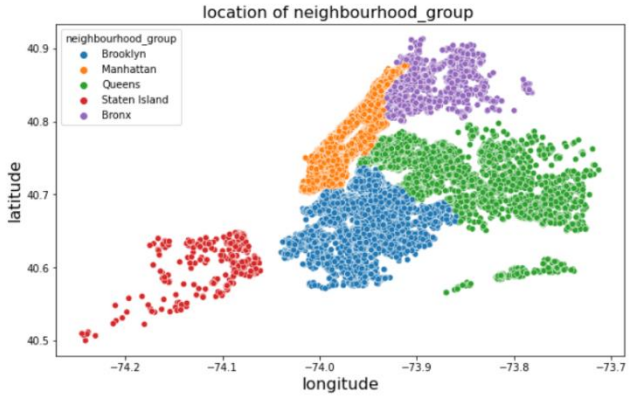
1- Figure shows the number of different types of accommodation. It is clear from the graph that most people rented whole houses or whole apartments through AIRBN in New York City, private rooms are also used quite a lot from the point of view of this chart, with only a small difference compared to whole houses and apartments. Finally, shared rooms, where it is possible to stay with strangers together, were significantly lower.



2-Figure shows the bar graph of accommodation in Airbnb according to the district in which the accommodation is offered. This chart clearly shows that Manhattan and Brooklyn are the most popular destinations at Airbnb. Accommodation in the Bronx and Staten Island are rather marginal



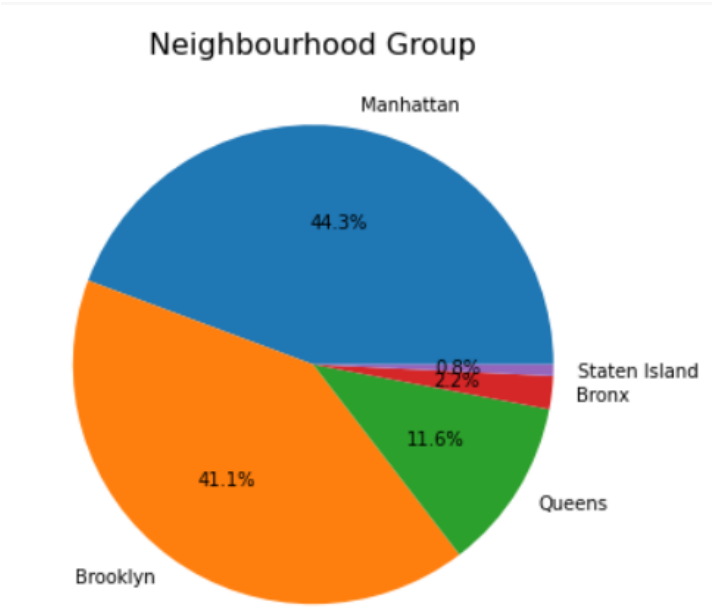
3- Location of hotels and price-The information we got from the graph is red color dots are the rooms with a higher price. Also, we can see that the Manhattan region has a more expensive room price



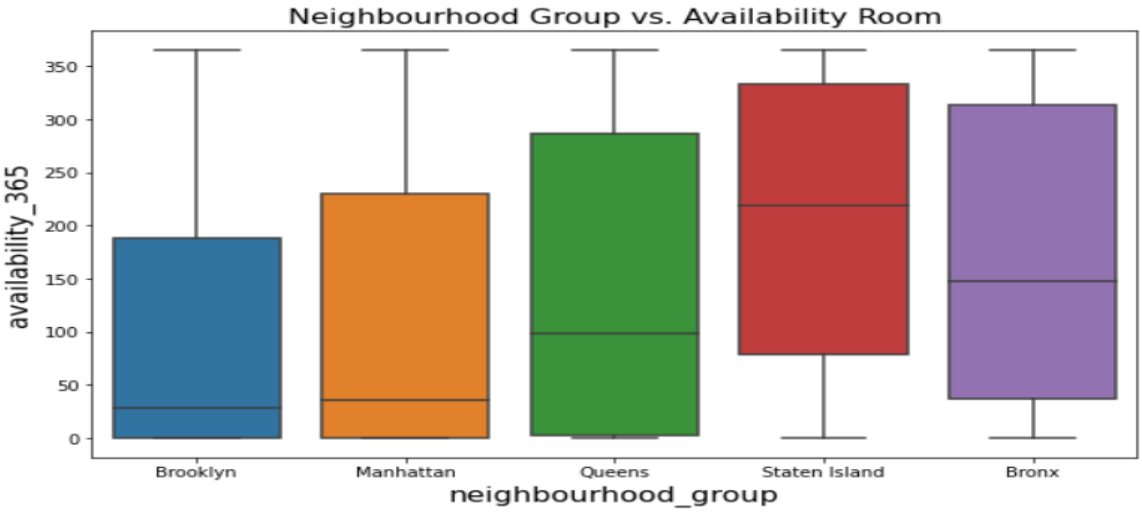
4-Top listed hosts - As we can see most listings are from Manhattan created by Sonder (NYC) . Followed by Blue ground.

	host_name	neighbourhood_group	calculated_host_listings_count
13217	Sonder (NYC)	Manhattan	327
1834	Blueground	Manhattan	232
1833	Blueground	Brooklyn	232
7275	Kara	Manhattan	121
7480	Kazuya	Queens	103

5-shares of hotels in gives data-The pie chart shows that Airbnb Listings in Manhattan and Brooklyn have the highest share of hotels



6-Availability 365-The graph shows the relationship between the availability of rooms and neighborhood groups.

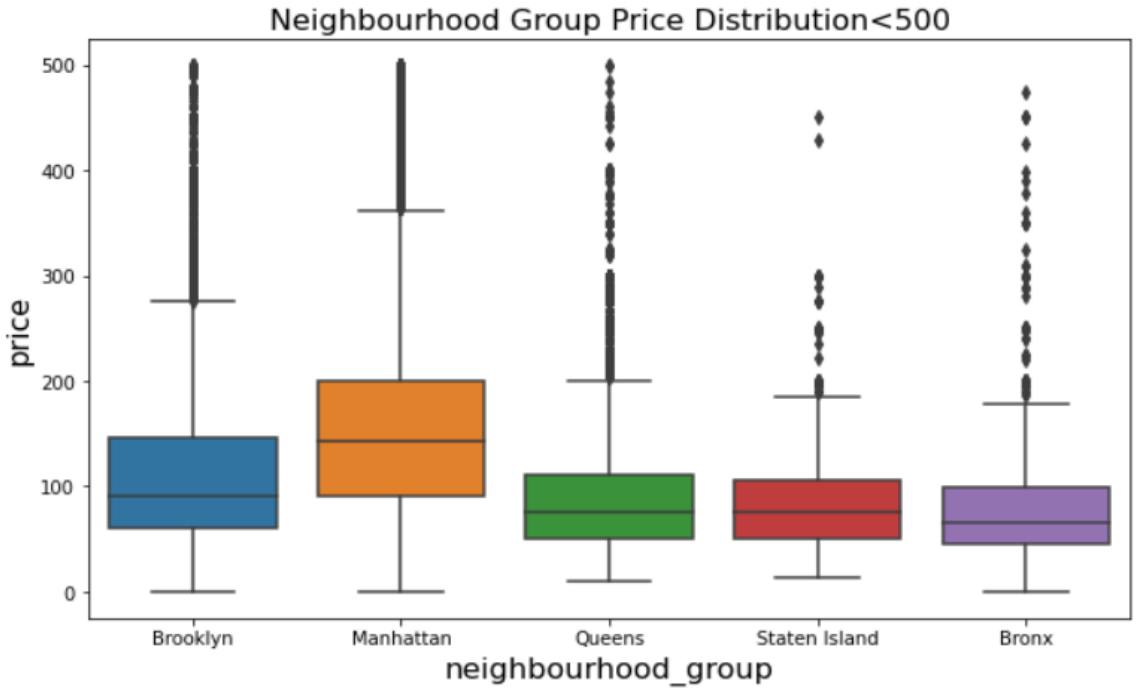


7-price in different locations-

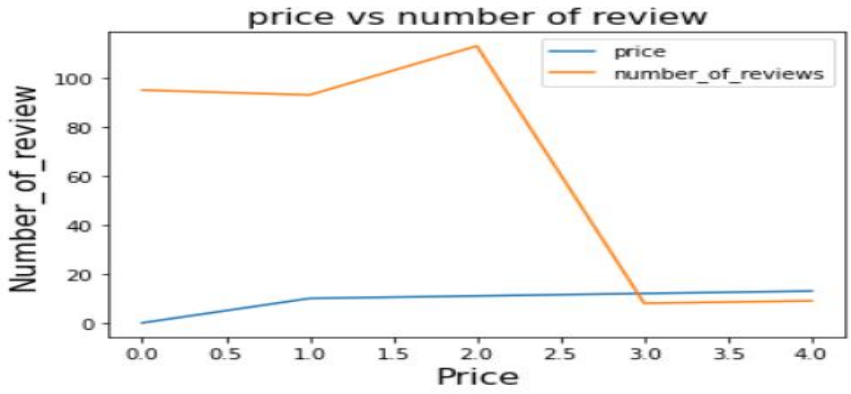
1-We can say that Manhattan has the highest price range for the listings, followed by Brooklyn

2-Queens and Staten Island seem to have a very similar distribution,

3-The Bronx is the cheapest.



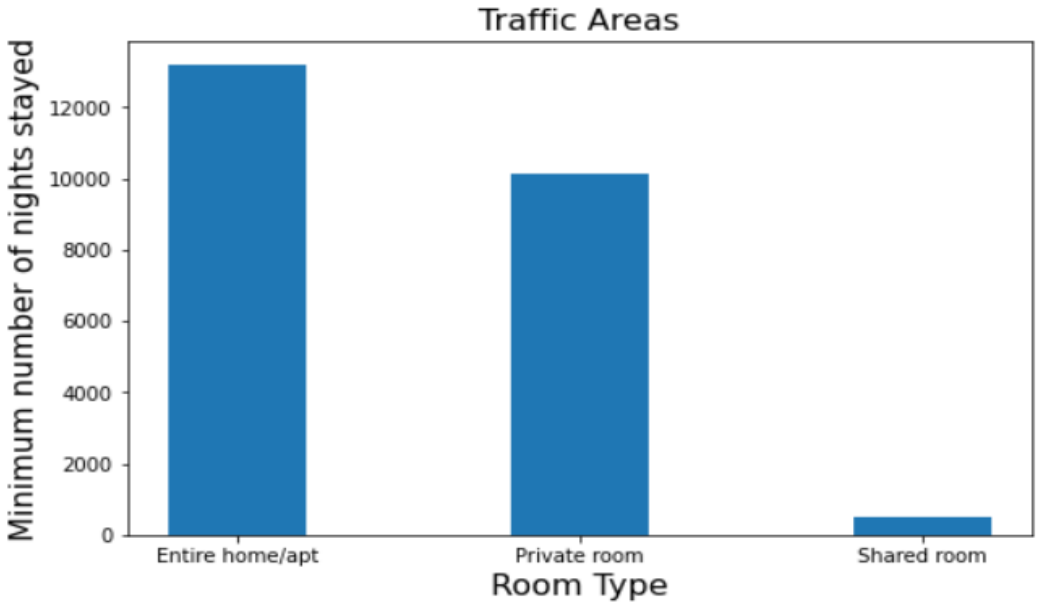
8- Price predictions-From the Analysis, we can say that most people prefer to stay in a place where the price is less



9-top busiest hosts-Dona, Jj, Maya, Carol, Danielle

host_name	host_id	room_type	number_of_reviews
Dona	47621202	Private room	629
Jj	4734398	Private room	607
Maya	37312959	Private room	543
Carol	2369681	Private room	540
Danielle	26432133	Private room	510

10-traffic areas-From the Analysis, We can Say that People are preferring Entire homes/apt or Private rooms which are present in Manhattan, Brooklyn, and Queens, peoples are preferring listings that are less in price



Conclusions:

The given dataset appears to be a very rich dataset with a variety of columns that allowed us to do deep exploration on each significant column presented.

- The people who prefer to stay in an Entire home or Apartment are going to stay a bit longer in that particular Neighborhood only.
- The people who prefer to stay in a Private room won't stay longer as compared to a Home or Apartment.
- Most people prefer to pay less price.
- If there are more reviews for a particular neighborhood group that means that a place is a tourist place.
- If people are not staying more than one night means they are travelers.

- For the given data set I found that there are a total of 221 different areas out of which “Williamsburg” has a maximum number of listings.
- There are a total of 37457 hosts and the host with host id-219517861 “Sonder” is the top host with 327 listings.
- No strong correlation was observed between price, reviews, and location.
- Out of 5 different locations in the dataset Manhattan is the most crowded location with 44.3% of listings.

Thank you

-Navjot khatri