**Airbnb Booking Analysis**

**Fouziya Afreen, Sumit Giri, Rajkumar K**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Airbnb is an online community marketplace that connects people, who are looking to rent their homes with the people who are looking for accommodation

Founded in San Francisco in 2008 as a start-up, the company has become a worldwide booking platform

Today it contributes to the movement of more than 60 million people in 192 countries.

We are finding the data insights and come up with the conclusion that which neighborhood area has maximum hosts.

***Keywords: EDA, Data Visualization***

**1.Problem Statement**

Data provided by Airbnb is an online community marketplace. They are trying to find out

* whichneighborhood group has maximum number of hosts
* which neighborhood has minimum price
* which room type has maximum review
* Maximum nights preferred by the hosts

**2. Steps involved:**

* **Exploratory Data Analysis**

After loading the dataset, we performed this method by comparing our target variables that is Neighbourhood\_group, price, room\_type with other independent variables. 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**

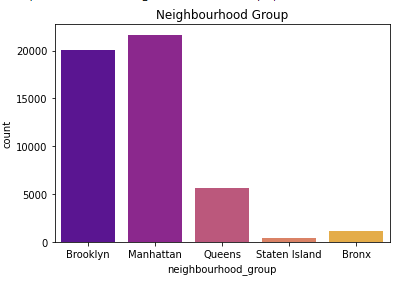
Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

* **Data Visualization**

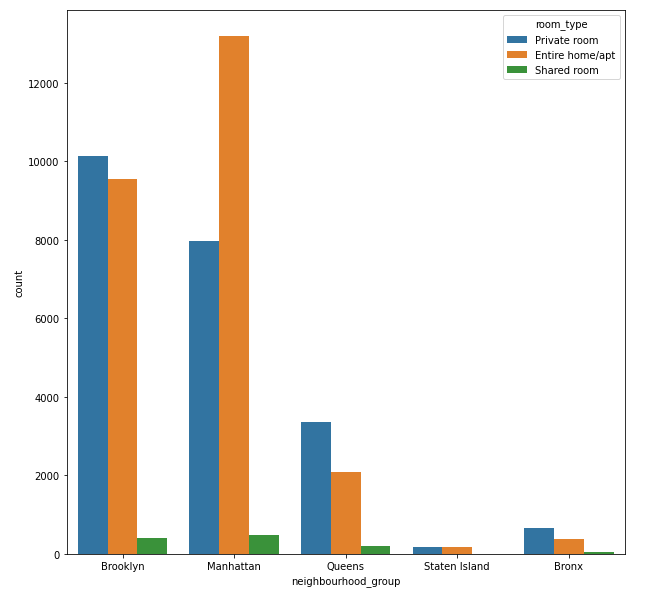
We have done different plotting to find the relationship between the data.

**Data Visualization**

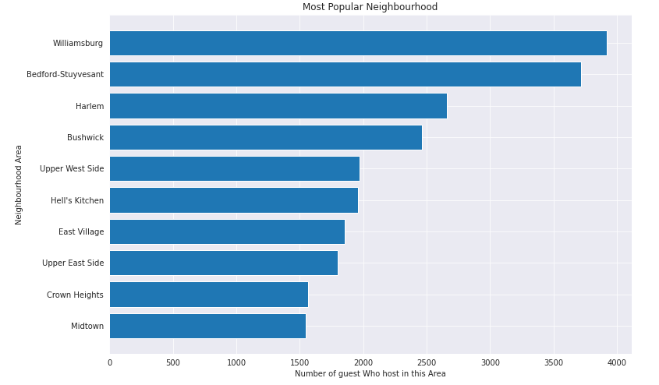
In the below bar plot the relationship between the Neighbourhood group is explained. Manhattan and Brooklyn has maximum hosts compared to other cities

****

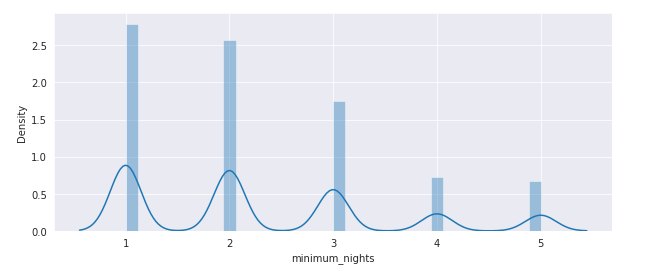
In the below graph the relationship between neighbourhood group and room types is explained where private room and entire home or apartment is mostly preferred compared to shared rooms.

****

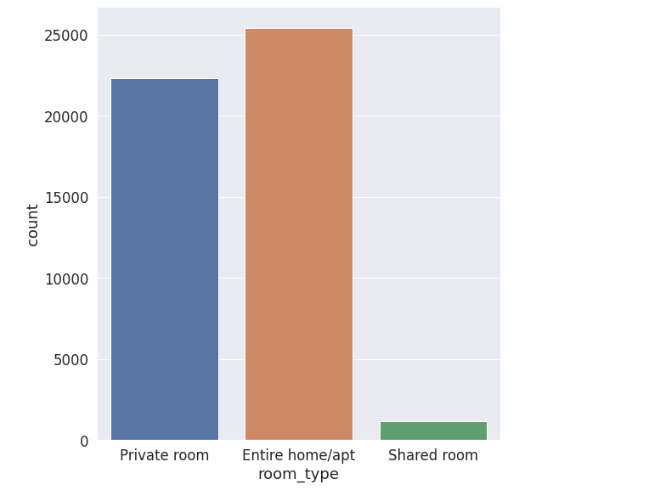
In the below graph Top 10 neghbourhood areas is extracted to find which neighbour has maximum number of hosts



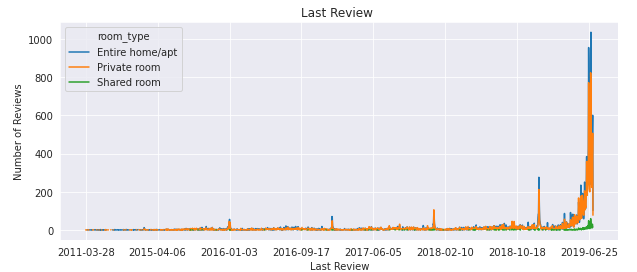
In the below distplot minimum number of nights preferred by the hosts are explained i.e, maximum hosts prefer to stay for 1 to 3 nights

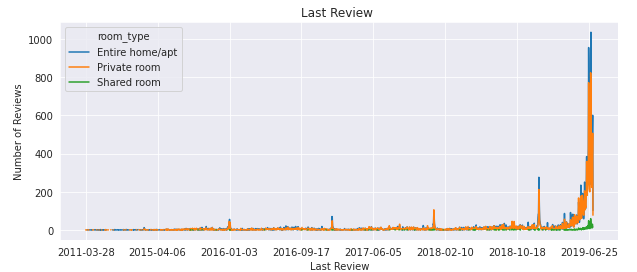
****

In the below distplot room type preferred by the hosts are explained. Here we can find out Private and Entire home or apartment is most preferred when compared to shared rooms.

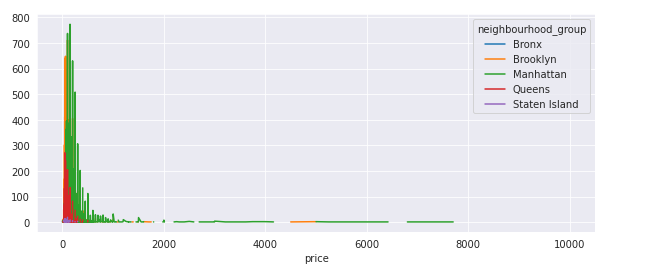


In the below plot review dates are displayed. By this we can understand that in the year 2019 the number of reviews has increased compared to initial year.

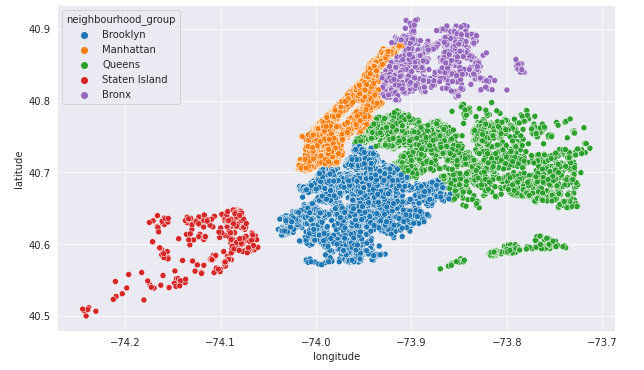


****

In the below graph average prices preferred by the hostiles are explained. Here we can find out that maximum hosts prefer to stay for minimum price around less than 300

****

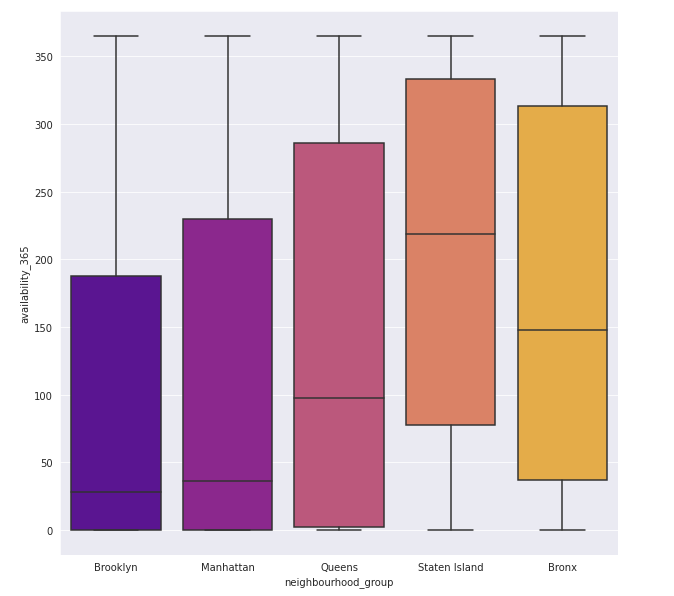
In the below scatter plot which will make us understand the geographical location of the major New York cities.

****

In the below correlation graph correlation between the two different variables are explained we have maximum correlation of 1.00 and minimum correlation of -0.75 which is explained in the graph

****

In the below box plot maximum availability of the hotel rooms are explained. Here we can see that Brooklyn and Manhattan has very less availability when compared to Queens, Staten Island and Bronx.

****

**Conclusion:**

That's it! We reached the end of our exercise.

Starting with loading the data so far, we have done EDA, null values treatment, Data Visualization using bar plot, distplot, scatterplot, boxplot which helped us to understand the insights of data.

**References-**

1. GeeksforGeeks
2. Stack Overflow
3. Pandas Documentary