Airbnb Booking Analysis

Names

Data Science Trainees

Alma Better

**Abstract:**

Our team has a deep knowledge about Airbnb Booking Analysis. We aim to focus only on New York city. The analysis is majorly done on borough country they are Brooklyn. Manhattan, Queen, Staten Island, Bronx. This data describes all host details, price of rooms, reviews etc.

**1.Problem Statement:**

The Customer is able to book a room in Airbnb. Customer is moving to Queens, Bronx, Staten Island, Brooklyn and Manhattan neighbourhoods to explore it.

The customer is in need to identify a number of bookings in each borough country & guide that customer to choose a best place to live at affordable cost, analyse the price of the room so that customer can afford it, type of room according to their needs. **Id -**Id of Each row

* **Name -**Name of Each room given by host
* **Host\_id –**Id of each Host
* **Host\_name –**Name of each Host
* **Neighbourhood group –** Borough country Name
* **Neighbourhood –**All cities of Each Borough Country
* **Latitude and Longitude –**Geographical Data of Borough Country
* **Room type** –Name of each room type
* **Price –**price of each rooms
* **Minimum Nights-** price for Minimum nights to stay
* **Number of review-** Total reviews given by customer
* **Last reviews-** Date of last reviews given by customer
* **Reviews per month –**Average review rate per month
* **Calculated host listing count** –Count of Host listing in Airbnb
* **Availabilty\_365 –** Number of days availability

**Introduction:**

The Airbnb is a room bookings data. This data is given only for New York City from the year of 2011 to 2019.By exploring the data we are able to understand all the features. The objective is to explore the data and then analyse the data with all the required information. With the help of data exploration, data cleaning and analysis we have to discover key understandings from it. So that the understandings can be used for guiding marketing, understanding of customers’ and hosts’ behaviours.

**Steps Involved:**

**Exploratory Data Analysis:**

Once our data loaded using pandas library. To understand the data, we printed top 5 rows and performed various operation like casting the columns into their respective data types. Creating new column **price\_per\_night** for fast analysis. numerical column to verify whether it follows a Gaussian Distribution.

**Performing Outliers Removal:**

Outlier is a data which differ from actual distribution of data. It also causes the Original distribution of data. The Outlier can be visualized using box plot and violin plot. The technique used here is to perform removal of outlier using Inter Quantile Range (IQR).

Q1=25th percentile

Q3=75th percentile

IQR=Q3-Q1

Lower\_bound= Q1 - (1.5 \* IQR)

Upper\_bound= Q3 + (1.5 \* IQR)

**Conclusion**