Airbnb Booking Analysis

Names

Data Science Trainees

AlmaBetter

**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. From analysis we get huge information with help of Exploratory Data Analysis, Data Wrangling, Visualization etc.

**1.Problem Statement:**

The Customer is able to book a rooms in Airbnb website .The Customer name is **xyz** and also a traveller. He/She moving to Queens, Brooklyn and Manhattan Neighbourhoods country to explore it. **Primary goal** of customer is to identify a number of bookings in each borough country and guide that customer to choose a best place to live at affordable cost. Then Customer decide to live in Queen’s country for 20 days. The Customer **Secondary problems** is to analyse whether the room price is cheaper than other neighbourhood’s and type of room to live at affordable price. There also other major goals that customer need to know number of reviews for each month at year of 2019 and relationship between each features. Since customer need to find the host name are busiest and why. Then He/She also want to give feedback to the host of that room where customer stayed .

* **Id -**Id of Each row
* **Name -**Name of Each rooms 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. We can able to understand the data of each features. This analysis is focused only on borough country.

The objective of this data is to analyse all required information with help of exploratory data analysis, Data Cleaning, treating null values, outliers and Visualization etc. The main goal is to analyse the given problem statement by customer to satisfy all information needed. The entire analysis is focused mainly on price feature with respect to other feature.

**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. Understanding the relationship of all numerical and categorical columns. Also done the Bivariate Analysis with respect to numerical data and categorical data.We also plot a distribution of numerical column to verify whether it follows a Gaussian Distribution.

**Visual Analysis:**

Visualization tells thousands of meaningful data in to single plot. The visualization is done with help of seaborn, plotly, matplotlib etc. We performed various plot like bar chart, pie chart map chart, box plot, violin plot and waterfall chart etc. Data Analysis is performed before cleaning the data.

**Treating Null Values:**

Null values is present in the given data. This null values occur due to some sensitive information which user may not like to specify those data. It can be visualized using heat map plot.

**Imputation techniques for each columns**

1. Removing **host\_name** column datas from Dataframe
2. Replacing **name** column null values by **'Missing'** Labels
3. **Random Sample Imputation** Techniques used for **Last review** column
4. **Median Imputation** Techniques used for **reviews\_per\_month** column

**How Random Sample Imputation works?**

This counts number of null values present in the columns. Then it remove all null values from that columns. Once removed it takes random values from that columns based on the count of null values. Then it replace all the null values by this random data respect to its index in the original dataframe.

**Performing Outliers Removal:**

Outlier is a data which differ from actual distribution of data. It also cause 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)