AIR BnB

Price Prediction

***Project Description:***

**1.Context:**

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in NYC, NY for 2019*.*

**2.Content:**

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions*.Although airbnb and other sites are provides some general guidelines,there are currently paided and accurate service which helps the host price their property using wide range of data points.*

**3.Acknowledgements:**

*This public dataset is part of Airbnb, and the original source can be found on this website kaggle.com.*

4.Problem statements:

* What can we learn about different hosts and location?
* What can we learn from predictions? (ex: locations, prices, reviews, etc)
* Which hosts are the busiest and why?
* Is there any noticeable difference of traffic among different location and what could be the reason for it?

5.Introduction:

In this **project**, I will perform a **descriptive** and **exploratory** analysis of the data, in order to understand how the phenomena of each variable behave individually and transversely, in addition to generate **hypothesis** useful for future **decision-making**.

The whole analysis will follow a simple and direct structure, well detailed in all topics, aiming at the same time, to create an intuitive and simple **guide**of which steps must be followed to carry out a good analysis, to in order to understand the data involved in any study.

6. Steps involved:

* **Libraries and data loading**

In this first we import some useful libraries like numpy, pandas, seaborn, matplotlib.pyplot, matplotlib.image.

After we load our data which was in the form of csv file.

We used pd.read\_csv command to load data in our notebook

* **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. Then we removed NAN values from our data set.

* **Duplicates values Treatment**

Since our data set contain duplicate values so we removed duplicate values from our data set.

* **Removal of unwanted columns**

We dropped some unwanted columns from our dataset since these were not use full for analysis.

### Exploring and Visualizing Data

##### Exploring the data by analyzing its statistics and visualizing the values of features and correlations between different features. Explaining the process and the results.

##### Now I take an exploration on my Airbnb data, then I take a decision that am going to be working from left to right. The reason some may prefer to do this is due to its set approach - some datasets have a big number of attributes, plus this way I will remember to explore each column individually to make sure am learn as much as I can about my dataset.

1. After that made my first bar chart on host id and their count of listing base that are showing the variation in hosts with the most listing in NYC. From this chart i have saw a good distribution between top 10 hosts with the most listing.first host has more than 300+ listing.
2. And am trying to filter price from zero to price value.

Then check for an outliers in price column as it’s the most important column for my analysis for which I need to be caution for our future analysis. And reaches to the minimum,maximum and mean value of each room type with its price. And which is shows in an violineplot .

1. Then move to check which room type is more in AIRBNB NYC. And we analyse the result through the pie chart.It is found that 52% of entire room/apartement ,46% private room,2% of shared room.

And also this distribution is reperesented in bar plot also.

**4.** After that am try to find the relation between location and price.

And made graph for it showing the relation between location and price counting.

i) Brooklyn, manhattan,queen has the maximum price .

ii) While bronx and staten island has the lowest price.

Then am create a bar plot according to the variation.

**5**.After that created the relation between location and mean price that show how much the place was costly,and it is represented in bar plot.

Then am concluded that maximum price is in manhattan and least price is in Bronx.

**6**. Then create relation between area, room type and price. From this relation we can calculate the price of each rooms .

### And it is plotted in scatter figure.

**7.** After that created a relation between price and number of reviews .from this I under stood that how much affected the price for number of reviews , and it analyze through a scatter graph.

**8.** And find out a relation between host name and number of reviews .which show how much the place was busy, it was checked by no. of reviews , the more no. of reviews means more customer which implies that more the host will be busy in the location.

**9.** Also I have got a relation between room type and minimum nights,and proven that peoples are prefered more the entire room/ apartement.

**7.** Here find the least and most expensive room of every type of each region.

**6.** Manhattan has the highest number of listing of about 44.3 % followed by Brooklyn of 41%. State Island stands the least number of listing less the 1 %.

**7.** And find the avg. room price in each neighborhood group.

As usual Manhattan being the costliest place to live in NYC, having average price more than 140 USD followed by Brooklyn with around 80 USD on an average for the listings. Queens and Staten Island has nearly the same of 75 USD. The highest price range could go just above 360 USD.

**8.** Also calculated the avg. cost of each room type:

Entire Home may averagely cost 150 USD .

Private room average cost around 65 USD.

Shared rooms cost the least price of less than 50 USD.

So, lets check what people prefer to live in with review counts.

**9.** And from all of the above find out the availability of rooms at neighbourhood group:

Staten Island has the most avability of rooms all the year.

Brooklyn having the least availability of rooms through a year with the second most count listing gives an opportunity to have an increase in number of rooms.

Manhattan, Bronx and Queens and nearly an equal avability of rooms throughout the year.

**10.** Find minimum nights spend in each room type:

Here we can state that in which room type customers prefer for night stay. From bar chart conclude that most of the customers loved to stay in apartment/entire room for maximum durations.

Only less customers spend night in shared room.

## Conclusion:

## 1. The people who prefer to stay in Entire home or Apartment they are going to stay bit longer in that particular Neighborhood only.

## 2. The people who prefer to stay in Private room they won't stay longer as compared to Home or Apartment.

## 3. Most people prefer to pay less price.

### 4. If there are more number of Reviews for particular Neighborhood group that means that place is a tourist place.

### 5. If people are not staying more than one night means they are travellers.

THANK YOU