AIR BnB

Price Prediction

**Project Description:**

**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*.*

**Content:**

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions*.*

**Acknowledgements:**

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

**1.Problem statements:**

* What can we learn about different hosts and areas?
* 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 areas and what could be the reason for it?

**2.Introduction:**

In this **project**, we 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 to generate **hypotheses** 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.

**3. Steps involved:**

* **Libraries and data loading**

In this first we import some useful pandas 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 that we are ready for an exploration of our data, we can make a rule that we are 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 we will remember to explore each column individually to make sure we learn as much as we can about our dataset.

**1.** After that we made our first pie chart on room type and their count on percent base that are showing the percentage distribution of room types.

We also made bar graph for room types distribution for even better understanding.

**2.** After that we try to find the relation between different hosts and area.

And made graph for it showing the relation between area and host counting.

i) Dona, Ji, Maya, Carol, Danielle has the maximum number of reviews.

ii) While Micheal, David, Sonder(NYC), John has the highest number of listing counts.

The being the top faciltors and busiest host.

**3.** After that we crated relation between location and reviews that 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 area will be busy.

**4**. Then we create relation between area, room type and minimum nights. From this relation we can calculate the traffic of people among different areas with the room type booked.

### From the Above Analysis We can Stay that People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.

**5.** Here we 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 we 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.** We 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.** We 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.** We find total no. of nights spend in each room type:

Here we can state that in which room type customers prefer for night stay. From pie chart we can conclude that 63.2% customers loved to stay in apartment for maximum durations.

Only 1.6% 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.