**Airbnb**

**Soniya Kumawat**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Introduction**

I present here my exploratory data analysis, visualizations, interactive plots, animations and lots of other interesting insights into the Airbnb data. I wish to perform an in-depth analysis on one of the most densely populated cities in the world.

Airbnb is an online marketplace that connects people who want to rent out their homes with people looking for accommodations in that locale. NYC is the most populous city in the United States, and one of the most popular tourism and business places globally.

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Nowadays, Airbnb became one of a kind service that is used by the whole world. Data analysts become a crucial factor for the company that provided millions of listings through Airbnb. These listings generate a lot of data that can be analyzed and used for security, business decisions, understanding of customers’ and providers’ behavior on the platform, implementing innovative additional services, guiding marketing initiatives, and much more.

**Problem Statement**

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind services that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.

**Steps to perform following analysis**

**Step 1-**Mounting the google drive

**Step 2 -**Loading the data.

The dataset we use is “[Airbnb Data](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data) we’ll load the CSV file with the help of a function read csv

**Step 3** -Exploring the data.

We will look into the data for better understanding of data . we will use function like head and tail to print five rows of data from head and bottom. With the help of function like shape and info we will dig into data more and get to know about all the columns.

* **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

**Step 4-** Checking the null values**.**

1. We will check the null values the with the isnull function andin columns likereviews\_per\_month, last\_review name host name we will find null values .
2. We have number of reviews so we dnt need last\_review reviews\_per\_month so we can drop this values as it have lot of null values
3. we can replace host name with mode to handel the data and drop name column because  as per question we dnt need to plot anything with respect to name

Now data have no null values its clean so we can go for EDA part

**Step 5-**Solving all the questions individually

**Question 1 What can we learn about different hosts and areas?**

1. To solve this question, we will use the groupby function and group host name Neighborhood group calculated host listings count.
2. To get top 5 neighborhood and count we will sort values in ascending order with respect to host listing counts with the help of sortvalues function of python
3. **As we can see most number of listings are from Sonder(NYC), Blueground, Michael, David and area is Manhattan**

**Question 2 What can we learn from predictions? (ex: locations, prices, reviews, etc)**

1. First we will look into neighborhood group and number of reviews we will group them together by groupby function and get to know that **Maximum reviews are from Brooklyn, Manhattan,Queens**
2. To get the information about the price we have applied scatter plot with respect to price and number of reviews from the Analysis we can say that most people prefer to stay in place where **price is less.**

**Question 3 Which hosts are the busiest and why?**

1. To solve this problem statement we have applied same approach as above we will group by host name host id room type and number of reviews and after then sort values in ascending order to get the most number reviews with respect to room type host name and host id and applied bar graph also for better understanding
2. Busiest hosts are: **Dona. Maya, Carol, Ji**
3. Because these hosts listed room type as **Entire home and Private room** as per data given which is preferred by the greatest number of people

**Question 4 Is there any noticeable difference of the traffic among different areas and what could be the reason for it**

1. To solve this problem we will groupby neighbourhood grouproom\_type minimum\_nightsAnd sort in decending values with respect to minimum nights and apply bargraph and countplot also with hue room type
2. From the Above Analysis We can Say 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.

**Conclusion**

1. he people who prefer to stay in Entire home or Apartment they are going to stay bit longer in that particular Neighbourhood 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 Neighbourhood group that means that place is a tourist place.
5. If people are not staying more then one night means they are travellers.

|  |  |  |  |
| --- | --- | --- | --- |