**Airbnb Booking Analysis**

**Viral Bhatu Shewale**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract: -**

Although networked hospitality businesses such as Airbnb are a recent phenomenon, rapid growth has made them a leading competitor in the hospitality industry. The purpose of this analysis is to analyze the nature of the phenomenon, its potential further development in the next few years and the impact these developments will have on tourism, hotels and city destinations.

**Problem Statement: -**

* Since 2008, guests and hosts have used Airbnb to expand travelling possibilities. Today, Airbnb has become a one-of-a-kind service that is used and recognized by the whole world and generates a lot of data.
* Data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services, and much more.

**Steps involved: -**

* **Null values Treatment**

Data cleaning (preparation) is one of the most crucial steps before performing any operation on the data. The dataset I was working on contains some null values, which might affect our EDA or visualization. We have to handle null values by replacing all null values present in columns 'name' and 'hostname' with 'no name' and 'review per month' with its mean.

* **Exploratory Data Analysis**

Afterloading thedataset, I performed initial investigations to understand on data to discover patterns. This process helped us figure out various aspects and relationships among different variables. It gave us a better idea of which feature behaves in which manner.

* **Visualization**

Data visualization is **the graphical representation of information and data**. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

To perform data visualization in python, I used various python data visualization modules such as **Matplotlib, Seaborn, Plotly**.

**(Bar Graph, Pia chart, countplot, Box plot, scatter plot)**

**Library used:**

**Pandas, Numpy, Mateplotlib, Seaborn**

**Features: -**

1. **Price: -**

The minimum price was zero, so I assumed it was a closed business. Most of the rooms available at price range from 10 $ to 180 $ (+35,000 rooms). only 1044 rooms are available for more than 500 $.

**2.neighbourhood\_group: -**

Airbnb service is available in five different neighbourhood groups (boroughs)

1. Brooklyn
2. Manhattan
3. Queens
4. Staten Island
5. Bronx

* Most of the rooms are listed from Brooklyn and Manhattan locations.
* Manhattan has the most number of the entire apartment and shared accommodations.
* Brooklyn has the most private rooms.

**3. neighbourhood: -**

There are 221 neighbourhoods that had Airbnb listings.

**4. Room type: -**

There are three types of rooms on the Airbnb platform.

1. Entire home/apt
2. Private room
3. Shared room

* We can see that the Entire Home

/Apt has the highest preference of people, followed by the Private Room, and the least preferred is Shared Room.

* We can conclude that the Entire home/apt type has an average price for reservations around 211.79 $, which most expensive of all types of rooms.
* Private room type is the second most expensive.
* Shared room is the least expensive.

**5. availability 365: -**

Staten Island has the properties with the highest availability and Brooklyn has the properties with the least availability.

**Conclusion:**

* Simply by performing EDA on the dataset, we have identified various new insights into how Airbnb listings are distributed, how the price range varies according to location, and we know where the listings are located.
* There were three types of room onto which the majority of people preferred the Private room or the Entire home/apartment.$.
* I found out that Manhattan and Brooklyn are dominating (Prime locations) locations.
* The Manhattan accommodation price is higher and affordable for high−class people.
* Sonder (NYC) (Host\_ID = 219517861) is the busiest host on the Airbnb platform.
* Manhattan and Brooklyn had the most listings, covering more than 80% of the total listings.
* Manhattan and Brooklyn are the most traffic-generating locations.

**References:**

* Analyticsvidya.com
* GeeksforGeeks.org
* Towardsdatascience.com (Medium)
* Pandas Documentation