**EXPLORATORY DATA ANALYSIS ON AIRBNB**

COHORT RIMO,ALMABETTER

BY:- NITIN KUMAR

Abstract :

Airbnb is a online market place that connects people who want to rent out their homes with people who are looking for accommodation in specific locales.

EDA of Airbnb helps us to investigate the data to understand the customers of Airbnb, spot anomalies, cleaning the data, & check the assumption with the help of the summary statics & graphical representation.

1. Problem Statement

Airbnb is providing a way for people to monetize their property while also providing a way for people who are traveling to have a different experience than a hotel. Additionally, hotels are also heavily taxed by the cities, and while cities are rushing to tax airbnb, it has been a way to avoid a charge that should mostly be considered unconstitutional.

1. Introduction

Airbnb dataset has information about hosts, costumer’s reviews, availability of rooms, neighbourhood groups, neighbourhoods, prices of room types and many more. So, with Exploratory Data Analysis we are analysing the patterns, changes, fluctuations in the prices and availability, the behaviour of hosts, expenditure and priority and relevancy of the people, and many things can be taken out as a conclusion from this dataset which can be used for the future development of the company.

Our goal here is to provide future-oriented conclusions for Airbnb. So that they can take decisions based on that conclusions and earn fruitful results.

1. Challanges Faced

* Understanding the dataset specially it’s columns
* Understanding the business model of Airbnb
* Removing Null values, changing datatype, fixing invalid data
* Designing the presentation to visualize the summarization such that the data is easily accessible to the reader

1. Neighbourhood Groups in Dataset

The sub-division of New York City is known as “Neighbourhood Groups”. There are 5 cateogeries in the Neighbourhood group :-

* Manhattan
* Brooklyn
* Queens
* Bronx
* Staten Island

1. Types of Rooms in Dataset

There are 3 room types in the 5 neighbourhood group :-

* Entire Home/Apartment: High Price and High Demand
* Private Rooms: Mid Price and Mid Demand
* Shared Rooms: Low Price and Low Demand

1. Libraries Used in EDA

* Pandas: Pandas is a python library used for data manipulation & statistical analysis. It is a fast & easy way to use open-source library that enables several data manipulation tasks. These include merging, reshaping, wrangling, statistical analysis and much more.
* NumPy: NumPy can be used to perform a wide variety of mathematical operations on arrays. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.
* Matplotlib: Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making 2D plots from data in arrays. It provides an object-oriented API that helps in embedding plots in applications using Python GUI toolkits.
* Seaborn: Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with Pandas data structures. Seaborn helps you explore and understand your data. Its plotting functions operate on DataFrames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.
* Folium: Folium is a Python library used for visualizing geospatial data. It is easy to use and yet a powerful library. Folium is a Python wrapper for Leaflet.js which is a leading open-source JavaScript library for plotting interactive maps.

1. EDA Procedure

* Importing Libraries
* Analysing the Dataset
* Understanding the dataset
* Data Wrangling
* Removing null values
* Changing Dtype of last review
* Fixing invalid data
* Removing the duplicates
* Outliers handling
* Plotting graph using matplotlib & Seaborn
* Observing the graphs & plots
* Final Conclusion

1. Conclusion

* Manhattan gives lead to every neighbour in all the fields like highest number of visitors, highest number of apartments, highest price, etc.
* Staten Island has least number of apartments and gets the least number of visitors in the neighbourhood.
* Majority of the people prefer Entire home/apt as they do not want to compromise on their comfort and privacy.
* Entire home/apt and Private rooms were available from 2011 were as the Shared rooms was introduced by Airbnb in 2014
* Number of reviews started increasing from 2015 and receives the highest number of reviews in 2019 by Airbnb.