Capstone Project Submission

Done By : PRAVEEN SIVAKUMAR (Cohort Everest)

Team Member's Name, Email and Contribution:

1. Praveen.Sivakumar (spr87654@gmail.com)
    1. Data Understanding:
        1. Columns which should be dropped and Used
        2. Providing step by step approach

    2. Median price for zero price
    3. Classifying the Features of Dataset
    4.Top 10 host IDs that owns the maximum number of Airbnbs in the New York City
    5.Price distribution of the Airbnbs across the New York City
    6.Price distribution of Airbnbs on the basis of neighbourhood groups
    7.Number of reviews received by the Airbnbs on the basis of neighbourhood groups and room types
    8.Distribution of number of reviews received by the Airbnbs on the basis of price and room type

2. Raj Sonar(rsonar821@gmail.com):
    1.Data Preprocessing:
        1.Approaching null values
        2.Quantile method for price
    2. Analyzing trends of the Features
    3.Exploiting Duplicate values
    4.Distribution of the Airbnbs in the neighbourhood groups
    5.Distribution of Airbnbs on the basis of room types
    6.Final Code Editing
    7.Power Point Presentation
    8.Data Visualization Techniques

3. Rahul Jha(rahuljha0904@gmail.com):
    1.Data Visualization:
        1.Providing Which graph is suitable
        2.Correlation of Dataset Features
    2.Exploiting Over Priced Airbnbs
    3.Colab Notebook Presentation
    4.Correlation between all the variables
    5.Average availability of Airbnbs on the basis of neighbourhood groups and room types
    6.Average number of nights spent in the Airbnbs on the basis of neighbourhood groups and room types
    7.Top 20 neighbourhoods on the basis of mean price
    8.Power Point Presentation Visualization

GITHUB LINK -https://github.com/prav87654/Airbnb-EDA
GOOGLE DRIVE LINK -https://drive.google.com/drive/folders/1UMya-T37E-x9JAIhOcMoTuW3aywNjeYD?usp=sharing

SUMMARY

The Airbnb services provide people an easy,relatively stress-free way to earn some income for the hosts from their property as well as for the customers such as home benefits,personalized services,authenticity and social connection around the world.Thus it has curated a large dataset having both dependent and Independent variables.

In this EDA Project,we were provided with only one dataset which consists of approximately 49000 rows and 16 columns. We given alias for the dataset as df.

For the first step,we performed data wrangling over the raw data and concluded what are the features or columns should be used or dropped. Since the columns last_review and reviews_per_month has most of the null values,we have to either drop or replace the values. We faced some challenges along the way as some price values given zero in which we replaced with median price on the basis of room_type and neighbourhood_group.

We divided the analysis part into two variates such as univariate and multvariate analysis for clear evaluation of conclusion and also for visualization.

Some of the univariate analysis that has been done for evaluating are top 10 host IDs that owns the maximum number of airbnbs,Distribution of the airbnbs in the neighbourhood groups,Distribution of Airbnbs on the basis of room_types.

Some of the multivariate analysis that has been done for evaluating are Comparison of neighbourhood_group in terms of price,Average availability of Airbnbs on the basis of neighbourhood groups and room types,Correlation of every features and so on.

The analysis are evaluated using various type of graphs and plots such as scatter plot,bar graph,violin plot etc.

The conclusion that we have derived can be used to implement business decision,understanding customer and provider behaviors,guiding marketing initiatives and implementing innovations and so on.