Capstone Project Submission

Team Member's Name, Email, and Contribution:

Name: SHAHFAISSAL I DHARWAD
email id: shahfaissal21@gmail.com
Contribution:

- Analysis of hosts areas & data visualization
- Map wise distribution of properties
- Scatter plot of Properties density
- Room Type with price
- Minimum nights distribution

Minimum number of nights distribution
Name: Ajinkya Satish Jumde
email id: jumde.ajinkya@gmail.com
Contribution:

- Data Wrangling
- Handling Missing values
- Creating a word cloud
- Room Type vs neighborhood group Analysis
- Room Type with availability analysis

Name: Shreyash Movale
email id: shreyash9m@gmail.com
Contribution:

- Data Wrangling
- Handling Missing values
- Creating a word cloud
- Room Type vs neighborhood group Analysis ·
- Room Type with availability analysis

Name: Neha Gupta
email id: nehagupta5994@gmail.com
Contribution:

- Data Cleaning
- Handling Missing values
- Neighborhood Group vs Count
- Room Type distribution
- Density and distribution of prices of the neighborhood group

Name: Eshaan Sosa
email id: rgip4321@gmail.com
Contribution:

- Data Cleaning
- Handling Missing values
- Top busy Hosts with the most listings
- Review Analysis with areas
- Reviews dependency on price

Please paste the GitHub Repo link.

GitHub Link: - https://github.com/shahfaissal/capstone-project---1

Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Airbnb Data Analysis:

Business Problem: The objective of this capstone is to do EDA on the given dataset and find out the insights from it.

1. PROBLEM STATEMENT:

Problem statement is the most important part of any analysis as we need to understand the needs of the client / business owner of what job they need to analyze and work according to their needs. The main objective was to analyze the data provided to provide the company with information that would help analyze its business and make important decisions to improve its business**.**

2. Know your Data:

Before performing any operations, the main step is to understand what our data consists of, things like what is type of the data, how many observations are present in the data, what is categorical data and numerical data has to be known first before getting into the analysis Dealing with a large data set is a time-consuming part. In this database, we are provided with 16 features and data types of about 49k. To understand the database we are working on, initially, we find key columns and data using ".head ()" and "info()". In order to reduce work and effort we have to disseminate data and analyze content first.

3.Handling missing values:

Any dataframe with missing values makes the analysis complicated so it is necessary Uninstalling unnecessary component while managing shortages and cleaning data. During the analysis, we found that there are four columns with empty values. To minimize errors, we have replaced the columns with empty values by 0 using the "fillna ()" function**.**

4.Univariate and Multivariate analysis:

a) To find the most searched words, we first deleted all the default words using the nltk dictionary and then used word cloud in an attempt to display the most searched words in the image view. We then analyzed strangers based on their number of properties and based on location and Airbnb listing. We then analyzed the types of rooms based on location, and the price of the room using the price division as well.

b) Based on price differentiation, we have tried to find the feelings of people in all the different neighborhood groups, where we have found that most people prefer affordable and expensive rooms in Manhattan and Brooklyn and, by booking in Queens and the Bronx., relationships are postponed.

c) We used longitude and latitude to indicate the spatial distribution of structures across all areas
d) We have tried to find a link between the various variables on which we build a relationship strategy. All analyzes were consistent with image representation.
e) We tried to set up a site to distribute to different neighborhood groups according to price. These distributions and inflation were in line with forecasts. Manhattan has proved to be extremely expensive compared to other clubs. In addition, it appears to be the place with the highest booking value

5.Conclusion:

- The scatter plot of neighbourhood groups with property density shows that Manhattan and Brooklyn are the most favourable area for investors as well as customers. So, more property owners can be considered for business in this area.
- Analysis on room type showed that more people are interested in renting a private room or flat/apt than shared room making or business concern towards the private room or flat/apt category
- Price predictions show that Manhattan is the most expensive area followed by Brooklyn while the Bronx is the cheapest although price category analysis showed that most visitors come from the affordable category so concentrating our business towards affordable properties.
- Analysis for most bookings again showed that Manhattan and Brooklyn are best for investment.
- Violin type graphical analysis showed that the Bronx and Staten Island show that property rates are a little lower there as compared to others so it will be better to invest there for future business

  With the above insights the company will be able to determine in which areas they are leading or lagging, where they need improvisation and what necessary changes they need for better functionality of their business. Even guests can benefit from this analysis.