**Capstone Project Submission**

| **Team Member’s Name, Email and Contribution:** |
| --- |
| Name : Ajinkya Satish Jumde  email id : [jumde.ajinkya@gmail.com](mailto:jumde.ajinkya@gmail.com)  Contribution : Multivariate analysis in airbnb capstone project analysis |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/ajinkyajumde/airbnb-cap |
| **Please write a short 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 : Objective of this capstone is to do EDA on the given dataset and find out the insights from it.**   * **Understanding the problem statement :**   **This is the most important part of any analysis as we must understand the needs of a customer/business owner for what task they require analysis on and work as per their requirements. The main motive was to analyze the given dataset in order to provide the company with the information that will help to analyze their business and take strategically important decisions to further grow their business to the new heights.**   * **Know your Data:**   **Dealing with a huge data set is a time consuming part. In this dataset we are provided with 16 features and around 49k data instances. To understand the dataset we are working on, initially we find the important columns and the data using “.head()” and “.info()” . To minimize the workload and efforts we must have to distribute data and analyze the content first.**   * **Handling missing values :**   **Removing the unnecessary part while handling missing values and data cleaning.During the analysis we found that there are four columns with null values. In order to minimize the errors, we replaced the numeric columns having null values with 0 using “fillna()” function.**   * **Univariate and Multivariate analysis:**  1. **In order to find out the most frequently searched words, we first removed all the stop-words using *nltk* dictionary and then using *wordcloud* we tried to display frequently searched words in an image view. Then we analyzed the hosts based on the number of properties owned by them and based on the area and Airbnb listings. Then we analyzed room types based on area, and the price of room using price categorization as well.** 2. **Based on the price categorisation, we tried to find out the sentiment of the people across the different neighbourhood groups, where we found that most of the people preferred affordable rooms over cheap and expensive in Manhattan and Brooklyn whereas for the bookings in Queens and Bronx, the relation was reversed.** 3. **We used longitude and latitude to demonstrate the geographical distribution of properties across the places.** 4. **We tried to get the correlation between different variables for which we plotted a correlation plot as well. All the analysis was accompanied by graphical representation.** 5. **We tried to plot a distribution plot for different neighborhood groups in terms of prices. This distribution and density of prices were in line with the predictions. Manhattan proved to be the costliest as compared to other groups.Moreover it proved to be the place with the highest number of bookings.**  * **Conclusion :**   **After the analysis we drew the conclusion that this analysis will help the company to decide in which areas they are leading or lagging, where they need improvisation and what necessary changes they need for better functionality of their business. As the analysis was done based on price category, different locations, hosts and room types, the company will be able to decide which type of room is prefered in what area and strategize accordingly for the betterment of their customer satisfaction and** |