**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **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, is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Based in San Francisco, California, the platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking.  Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world. Today, Airbnb became one-of-a-kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.  Firstly, we have imported the raw dataset and started to format in a desired state to draw any useful information out of it. We have selected the features that are required, eliminated the columns which are having a greater number of null values and replaced the null values in columns which are low in number with a sense so as there is not a significant loss of information. Then we used ‘describe’ method to get the idea of the numerical columns present in the dataset and their value range. We have also identified categorical columns and the total categories that are present in each.  Then we started analyzing our dataset. Firstly, we started with univariate analysis like hosts with most listings, number of hotels in each neighborhood groups, top reviews of neighborhood groups and top 10 most expensive hosts with the help of bar-plot and pie-plot. We have also  Then, we have filtered our dataset on the basis of 10th and 90th quantile value to eliminate the outliers that are present to get the unbiased conclusion w.r.t. price. Then we moved forward by doing bivariate analysis using features like ‘neighbourhood\_groups’, ‘price’, ‘number\_of\_reviews’, ‘reviews\_per\_month’ and ‘room\_type’. We have also determined the top ten busiest host according to the maximum number of reviews and determined the area which has maximum traffic.  We are able to draw out various inferences from our analysis and able to figure out that customers mostly prefer popular places with a lot of tourist attractions and corporate sectors, for short stay private rooms are preferred. Shared rooms are least in number and are least preferred by the guests. |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Gaurang Bagga**  * Feature selection * Null values handling * Outliers’ detection * Analysis of data through different techniques * Visualizing the data  1. **Sushant Tripathi**  * Feature selection * Null values handling * Outliers’ detection * Analysis of data through different techniques * Visualizing the data  1. **Prerna Dave**  * Feature selection * Null values handling * Outliers’ detection * Analysis of data through different techniques * Visualizing the data |
| **Please paste the GitHub Repo link.** |
| GitHub Link:- https://github.com/sushant-tripathi/Airbnb-Dada-Analysis |
| **Please paste the drive link to your deliverables folder. Ensure that this folder consists of the project Colab notebook, project presentation and video.** |
| Colab Link:- https://drive.google.com/drive/folders/1dARdbs\_lh767ELA3YU22Q4yj0T-A0GVg?usp=sharing |