**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| **Gaurav Jindal:**   * Uploading the dataset to google colab. * Analyzing the data, objective. * Removing null values and redundant columns. * Data wrangling. * Data visualization * Concluding from the various observations. * Project Presentation. * Technical Documentation. * Project Summary * Presentation Video |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/igauravjindal/Capstone-Project-1>  Google Drive Link - [Click\_Here](https://drive.google.com/drive/folders/1IMsXiNUlEMCeyEMQ5E4cBcB1ors8d5yi?usp=sharing) |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| 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.  This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.  **Problem Statement:**  In this project we are analyzing Airbnb’s New York City(NYC) data of  2019.  Our main objective is to explore and analyze the data to discover key  understanding about listing of properties on the platform. We will  perform basic Exploratory Data Analysis(EDA). We will be find out key  metrics that influence every Airbnb listing based on their location,  different hosts and areas, prices, reviews, room type, listing  name, traffic and other related factors. |
| At the first step, imported important libraries and from the drive Airbnb NYC 2019.csv were loaded and the airbnb\_df Data Frame was created.  Then Data cleaning was performed as there were various columns having null values and the ‘id’ column was removed from the dataframe as it was not needed for our purpose.  Then various problems were taken upon one by one and Exploratory Data Analysis was performed. We explored the data using various techniques and found some useful information from it from which we could have few good conclusions and we took help of visualizations to get more clarity about various aspects of the data and finally various conclusions were written upon.  **Conclusion:**   * Host Sonder(NYC) has the most listings and these listings are in the Manhattan area. * Manhattan has the most number of Airbnb properties available followed by Brooklyn. * As Queens has the most reviews so most visitors in Queens followed by Manhattan. * Brooklyn and Manhattan neighborhood groups have maximum price tags on Airbnb and Staten Island has low prices as compared to Brooklyn and Manhattan. * Manhattan has the most expensive Airbnb properties * Entire home/apt is the most expensive Airbnb Room Type * Dona is the busiest Host and this is in Queens, followed by Jj in Manhattan and so on. * Private Rooms has the most number of reviews and Queens has the most private rooms among all the neighborhood groups. * Manhattan has the most number of minimum night stays, so Manhattan is the most Traffic area. * 'Entire home/apt' room type is more demanding in traffic areas as compared to 'private room' and 'shared room'. And 'shared room' is the least demanding. |