**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. **Vaibhavkumar Gupta** ([vaibhavguptavkg@gmail.com](mailto:vaibhavguptavkg@gmail.com))  * Analyzed availability\_365 column * Filled price column price=0 with respective median price * Plotted Categorical Plot using Klib library * Analyzed host\_name and host\_id * Analyzed the relation of reviews\_per\_month with each neighbourhood * Analyzed listings according to neighbourhood\_group * Analyzed distribution of price column across each room\_type * Found out average price in each neighbourhood\_group for each room\_type for getting good number of reviews * Figured out a metric for knowing busiest hosts  1. **Bhavik Ashokkumar Verma** ([vermabhavik585@gmail.com](mailto:vermabhavik585@gmail.com))  * Analyzed price column * Replaced NA values in reviews\_per\_month by 0 * Plotted Pearson’s correlation matrix using Klib library * Analyzed host\_name and host\_id * Analyzed the relation of total\_number\_of\_reviews with each neighbourhood * Plotted Map using Folium and added the markers by using list of latitudes and longitudes. * Analyzed room\_type distribution across each neighbourhood\_group * Found out average price in each neighbourhood\_group for each room\_type * Helped to understand correlation with various columns to figure out good metric for knowing busiest hosts  1. **Dilkhush Sharma** ([kumardilkhush.rds@gmail.com](mailto:kumardilkhush.rds@gmail.com))  * Analyzed neighbourhoods * Analyzed room\_type distribution across each neighbourhood * Analyzed price trend across neighbourhood * Analyzed room\_type distribution on Airbnb * Plotted multiple scatter plots and bar plots * Analyzed hosts having highest number\_of\_reviews * Helped in knowing busiest hosts  1. **Priyanka Pal** ([palpriyanka00029@gmail.com](mailto:palpriyanka00029@gmail.com))  * Analyzed neighbourhoods * Analyzed room\_type distribution across each neighbourhood\_group * Analyzed hosts having highest reviews\_per\_month * Analyzed price trend across neighbourhood * Analyzed room\_type distribution on Airbnb * Helped in knowing busiest hosts  1. Shayan Somanna ([shayan.somzz@gmail.com](mailto:shayan.somzz@gmail.com))  * Analyzed neighbourhood\_group * Analyzed room\_type distribution across each neighbourhood\_group * Analyzed price trend across neighbourhood\_group * Analyzed room\_type distribution on Airbnb * Analyzed hosts having highest reviews\_per\_month * Helped in knowing busiest hosts |
| **Please paste the GitHub Repo and Google Drive Folder link.** |
| Github Link: - <https://github.com/bhavikk001/airbnb-eda>  Google Drive Folder: - <https://drive.google.com/drive/folders/1tqhcXYB23-WaMJiR69ahkQnfijktuxUf?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)** |
| Airbnb is an internet marketplace that connects folks that need to farm out their homes with people who are trying for accommodations in specific locales.  Airbnb revolutionized the hospitality industry. Before 2008, travelers would likely have booked a hotel or hostel for their trip to another city. Today, many of these people are turning to Airbnb.  Today, Airbnb is a unique service that the entire globe uses and recognizes. For the business, data analysis on the millions of listings offered through Airbnb is essential. These millions of listings generate a tone of data, which can be analyzed and used for a variety of purposes, including security, business decisions, understanding customer and provider (host) behavior and performance on the platform, directing marketing initiatives, putting into practice cutting-edge extra services, and much more.  **Understanding the variables**   * **id**: Unique listing ID * **name**: Name of the listing * **host\_id**: Unique host ID * **host\_name**: Name of the host * **neighbourhood\_group**: Location * **neighbourhood**: Area * **latitude**: Latitude coordinates * **longitude**: Longitude coordinates * **room\_type**: Listing space type * **price**: price in dollars * **minimum\_nights**: Amount of nights minimum * **number\_of\_reviews**: Number of reviews * **last\_review**: Latest review * **reviews\_per\_month**: Number of reviews per month   **The approach we have used in this project is defined in the given format**:  **1) Loading our data:**  In this section we just loaded our dataset in colab notebook and read the csv file.  **2) Data Cleaning and Processing:** In this section we have tried to remove the null values and for some of the columns we have replaced the null values with the appropriate values with reasonable assumptions.  **3) Analysis and Visualization:** In this section we have tried to explore all variables which can play an important role for the analysis. In the next parts we have tried to explore the effect of one over the other. In the next part we tried to answers our hypothetical questions.  **4) Future scope of Further Analysis:** There are many apartments having availability as 0 and date of last\_review is very old, which can mean that they must have stopped their business, we can find the relation with neighbourhood with these apartments if we could dig much, various micro trends could be unearthed, which we are not able to cover during this short duration efficiently. There are various columns which can play an important role in further analysis such as number of reviews and reviews per month finding its relation with other factors or other grouped factors can play an important role.   1. **Conclusions**  * The host Sonder (NYC) has the most listings on Airbnb in NYC. * The area around Williamsburg has the most listings. * The most expensive listings in NYC are in the Upper West Side, Astoria, and Greenpoint neighbourhood. * The Bedford-Stuyvesant neighbourhood have the most reviews overall and monthly, respectively. * Manhattan and Brooklyn neighbourhood group have the most listings. The quantity of listings in the Staten Island and Bronx neighbourhood group is relatively very low. * Private Room or Entire Home/Apartment rentals make up the majority of Airbnb listings in New York City. People who want to stay in an entire home or apartment are likely to stay longer, whereas those who prefer to stay in a private room are probably going to stay for less time than those who prefer to stay in an entire home or apartment. * The price field has several rows with values of 0, suggesting an error that Airbnb should fix.   Keeping the listing's price high and having no availability doesn't help the host because the customer is willing to pay the amount, but what is the benefit if there are no rooms available even after that?   * The host, Maya, has received the most reviews overall. * The average cost of every room type in Manhattan is higher than the cost of every room\_type in the other neighbourhood group. The average cost of every room type in the Bronx neighbourhood group is lower than the cost of every other neighbourhood\_group. * To learn more about the busiest hosts, refer to the answer to Question 12 in the Colab notebook. |