**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Raneev K** – [raneevk36@gmail.com](mailto:raneevk36@gmail.com)   **Data wrangling, data cleaning, and data visualisation**   * Area vs Host\_listings analysis * Host\_name vs Number of reviews analysis. * Room Type vs price analysis * Neighbourhood\_group vs availability analysis * Price vs Number of reviews analysis * Reviews vs neighbourhood area analysis  1. **Radha -** [radhagarima@gmail.com](mailto:radhagarima@gmail.com)   **Data wrangling and data visualization**   * Room type vs number of reviews * Neighbourhood vs number of reviews * Neighbourbood\_group vs minimum number of nights |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/Raneevk/Airbnb-Booking-Analysis |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Abstract**  Airbnb is an online marketplace connecting travelers with local hosts. On one side, the platform enables people to list their available space and earn extra income in the form of rent. On the other, Airbnb enables travelers to book unique homestays from local hosts, saving them money and giving them a chance to interact with locals. Catering to the on-demand travel industry, Airbnb is present in over 190 countries across the world.  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**  We have a dataset of Airbnb having almost 16 columns and 48k entries,we are going to analyse this dataset to find solution for some business problems,they are,   * What can we learn about different hosts and areas? * What can we learn from predictions? (ex: locations, prices, reviews, etc) * Which hosts are the busiest and why? * Is there any noticeable difference of traffic among different areas and what could be the reason for it?   **Approach**  **Exploring Dataset**  This process involves Understanding given dataset like shape of the dataset, different column values, check whether any null values present etc.  **Data Cleaning**  In this process we explore the number of rows & columns, ranges of values etc. Then Handle missing, incorrect and invalid data and Perform any additional steps (parsing dates, creating additional columns, merging multiple dataset etc.)  **Exploratory Analysis and Visualization**  Exploring the data by analyzing its statistics and visualizing the values of features and correlations between different features. Explaining the process and the results. Now that we are ready for an exploration of our data, we can start working from left to right. The reason some may prefer to do this is due to its set approach - some datasets have a big number of attributes, plus this way we will remember to explore each column individually to make sure we learn as much as we can about our dataset.  **Conclusions:**   1. From the top 10 observations accordings to highest calculated\_host\_listings\_count, we can found that 7 results are from “Manhattan” area,2 from “Brooklyn” and 1 from “Queens”.So it is clear that airbnb is a popular businesss model in “Manhattan”.The host who has most host\_listings are “sonder(NYC)”. 2. From top 10 list of busiest hosts we can see that 8 results are from private room type. 3. The name of the busiest host is “Dona” from “queens” area having 629 reviews. 4. Even though shared rooms are much cheaper as compared to private rooms most reviews are for private rooms which means that most number of people prefer private rooms over other room types.probably couples or small group of peoples who need more privacy and prefer budget friendly stay will go for private rooms,so we can say that most of the airbnb customers are from this category. 5. 'Manhattan' is the area having highest median price,followed by “Brooklyn”.So we can say that most costly hosts are situated in "manhattan". 6. Room availability is high in "Staten Island" ,the average value shows that rooms are available for 246 days in each year. 7. Room availability is very low in "Brooklyn", as per available data rooms are available only for 93 days in each year. 8. From the previous observations we found that maximum host listings are in "manhattan area" but room availability is very low as compared to other areas that means number of people visiting "manhattan" is more than available rooms. Which creates high demands for rooms and thats the reason for high price. 9. From the scatterplot it is clear that most number of people prefer budget friendly rooms. 10. by analysing available datas of "number\_of\_reviews","minimum nights" we found "manhattan" and "brooklyn" are the high traffic areas. |