**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| Name : Md Ismail Quraishi  Email : [mdismailquraishicse@gmail.com](mailto:mdismailquraishicse@gmail.com)  Contribution : Everything |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/mdismailquraishicse/Capstone-Project-1--AirBnb> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Problem Statement** : Airbnb is a room/apartment/lodge booking platform. The dataset has 49000 observations and 16 columns like location, price and details of the hosts and rooms or apartments. Given the dataset, we have to perform data analysis and find meaningful data which are useful to stakeholders.  **Approaches** :  First of all I performed data wrangling over the raw data. Then I defined a function to calculate the null value proportions in each column.  After that I performed column wise analysis, I used ‘name’ column to calculate total number of unique listings, ‘host\_id’ to calculate total number of hosts, neighborhood to calculate number of locations, neighbourhood\_groups to find groups, room\_type to find types of rooms, latitude and longitude to plot location map,  Calculated\_listings\_count to calculate maximum number of listings according to different criteria, price to calculate average and max price for different cases.  Graphs and charts are plotted for each important column to visualize and compare average price vs room type, sum of listings vs locations etc.  Using last\_review column I converted the column from object type to date time then I performed date wise analysis and plotted date wise graphs.  **Conclusions** :  There are 3% null values in name, 4% in host\_name, 20.56% in last\_reviews and 20.56% in reviews\_per\_month columns.   1. Price column was positively skewed but after the Log transformation it became almost normally distributed. 2. Number of unique hosts : 37457 3. Number of unique listings : 47906 4. Number of unique neighborhoods : 221 5. Number of unique neighbourhood\_group : 5 ['Brooklyn' 'Manhattan' 'Queens' 'Staten Island' 'Bronx'] 6. Number of unique room type : 3 ['Private room' 'Entire home/apt' 'Shared room']      1. Neighborhood groups with maximum number of listings : Manhattan (277073). 2. Top five unique hosts are : Sonder (NYC) with 327 hostings, Blueground with 232 hostings, Kara with 121 hostings, Kazuya with 103 hostings, Jeremy & Laura with 96 hostings. 3. Highest average price for Entire home/apt in Manhattan is 249.24 4. Highest average price for Private Room in Manhattan is 116.81 5. Highest average price for Shared Room in Manhattan is 88.98 6. Highest minimum nights spent for Entire Home/apt in Manhattan is 1250 7. Highest average price for year 2013 is 256.20 8. neighborhood with highest listing is : Williamsburg ( 3920) 9. sum of listings for all neighbourhood\_group : Bronx (2435), Brooklyn (45908), Manhattan (277073), Queens (23003), Staten Island (865) |