Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| Airbnb, Inc. is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities.Airbnb stands for “Air Bed and Breakfast” Airbnb does not own any of the listed properties; instead, it profits by receiving [commission](https://en.wikipedia.org/wiki/Commission_(remuneration)) from each booking. This EDA project includes "Airbnb booking Analysis" database.In this work, we present analytical results obtained by data mining on Airbnb booking dataset.  As the first step, perform data wrangling over the raw data, and handle the outliers over the price column using mean. After doing the data cleaning we are left with 45923 rows and 16 columns.  Our exploratory data analysis is sliced into 3 main parts: they are 1)Busiest host and reviews per month. 2)Average price of properties. 3)Availability and number of minimum nights.  **Busiest host:**  we would consider the host to be busiest, if the host frequently receives bookings. Since we have not been provided with any booking data. We can consider the number of reviews received by the host as the minimum number of bookings the host could get.host with maximum number of reviews is not the busiest, but the host with the highest rate of reviews (i.e max number of reviews per month) is the busiest. Sonder has the highest review per month, highest listing count, and most of his properties are in Manhattan which is the most densely populated among the 5 boroughs of New York City. Hence Sonder is the busiest host.  **Average price of properties:**    Entire home/apt is costly than other room types and shared rooms are cheaper. Properties in Manhattan are costly as compared to other neighborhood groups. The average price of properties in Queen and Staten Island are almost the same. Properties in the Bronx and Queens are cheaper.The Entire home/apt in the Bronx, shared room in Queen and Private room in the Bronx are cheaper.Most of the properties listed on Airbnb cost between 100 to 200 dollars per day.    **Availability and minimum nights:**  One of the factors on which the availability of property depends is booking.Availability of properties for booking in Manhattan and in the north part of Brooklyn is less as compared to other  neighborhood group, This shows that the properties in this region are often booked, hence People prefer renting an apartment in Manhattan and in the northern part of Brooklyn. And they mostly prefer private rooms and the entire home/apt over the shared room.We see that most of the hosts on Airbnb have set the minimum\_nights stay as 1, 2 and 3 nights  **Team Member’s Name, Email and Contribution:**  **1) Abdul Rehman**  [iamabd53@gmail.com](mailto:iamabd53@gmail.com)  **Data cleaning:**   1. Fill the NaN values in host\_name by 'no name'.  Replace NaN values in the 'name' column by 'no description'. **EDA and Visualization:**   1. Properties Count On Airbnb. 2. The average price of properties on Airbnb.   Detailed plotting   1. Properties located on the map and the percentage of each property type in different neighborhood groups listed on Airbnb. 2. Price of properties in different neighborhood groups shown on a map. 3. Availability Property for booking in different neighborhood group shown on a map   **2) Amani Chivilkar**  [amanichivilkar@gmail.com](mailto:amanichivilkar@gmail.com)  **Data cleaning:**  1) Replace NaN values in 'reviews\_per\_month' column by 0’.  2) If the price is 0 , replace it by the average price of all similar room\_type with similar neighborhood groups.    **EDA and Visualization:**   1. Properties Count On Airbnb 2. The average price of properties on Airbnb 3. Rate of reviews per month, received by each host for all its properties 4. Top 5 hosts with highest listing counts 5. Overall Correlation with Price   **3) Shivam**  [shiv01212@gmail.com](mailto:shiv01212@gmail.com)  **Data cleaning:** Check how many NaN values are present in each columns  1. Drop unnecessary column like "id"   **EDA and Visualization:**   1. Properties Count On Airbnb 2. The average price of properties on Airbnb 3. Top Hosts by their Number of Listings 4. Price at which most of the properties are available 5. Most Popular Descriptive words used by the host to describe their properties   **4) Vinay Kumar**  [vinaykumar.nigudgi@gmail.com](mailto:vinaykumar.nigudgi@gmail.com)  **Data cleaning:**  1) Find the outliers of the price column.  2) Convert the time string to datetime.  3) Verify cleaned Data .  **EDA and Visualization:**   1. Count of unique hosts on Airbnb. 2. Availability percentage of properties in different neighborhood groups. 3. Average minimum\_night of different room types. 4. mimimum\_nights stay set by most of the hosts . 5. Top 5 hosts with minimum possible earnings. |
| --- |