**Capstone Project Submission**

|  |
| --- |
| **Contributor Roles** |
| * **Rachna Gupta** * Understanding Data * Understanding different column * Having overview of data * Availability analysis * Data Cleaning * Removing null values * Replacing zero value of price with mean value * Dropping unused rows. * Exploratory Data Analysis * Price analysis * Listing analysis * Inference * Summarization * **Aditya Bhople** * Understanding Data * Understanding different column * Having overview of data * Data Cleaning * Removing null values * Replacing zero value of price with mean value * Dropping unused rows. * Conclusions * **Rupali Dawkore** * Summarization * Conclusions * **Piyush Sujeeth** * Data Cleaning * Removing null values * Dropping unused rows. * **Subhi Yadav** * Data Cleaning * Removing null values * Dropping unused rows. * Data Analysis |
| **Please paste the GitHub Repo link.** |
| Github Link:- |
| **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 (ABNB) is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in specific locales.  In Airbnb people list their own housing for rent. Since 2008, it has grown in popularity and specially for those community which frequently use to travel. It is becoming a strong competitor to the hotel industry. It hasmillions of listing, which generate lots of data. We are analyzing these data for making business decision, for looking best room type etc.  As a first step we take the overview of data, where we specially made our focus on understanding what each column means.So that we can be clear from what perspective we have to analyze our data. After understanding different column,we marked few important columns. These columns are neighborhood group, room type, price, minimum nights, reviews per month. Then we did some basic visualization to see is there any correlation among columns.  Now we started cleaning our data. So we first identified the null values and we replaced these null value according to their data type. After dealing with null values we moved on to those columns which we don’t need. So we removed last review column. Then we replaced few data which don’t make sense with other values. Here we replaced zero price of property with mean price according to their room type and neighborhood type. Finally, our data is ready for EDA.   * Price Analysis: The first focus we put on is ‘Price’. Here first we looked in to average price of different room type across New York. From this we get to know that costly room type is ‘Entire home’. Then we looked at average price of room type according to different neighborhood group. From this analysis we made inference that if a salaried employee wants to increase his saving then he will prefer to work in Bronx. Then we did few more on analysis on finding cheapest neighborhood and the cheapest listing throughout New York. * Listing Analysis: Here we focused on different listing. In this we take a look at listing according to their neighborhood group. From the result we made inference that is someone want to do advertisement or marketing he should focus on Manhattan and Brooklyn. Then we deep dive in data and looked different listing according to neighborhood group. We get to know that in Manhattan, entire room type is highly listed. * Availability analysis: Here we focused on availability of different room according to their neighborhood group and then the average availability of different room type. We get to know that private room has highest availability and entire home has least availability. The inference which come out from this result if host is having entire room then he will be making good money. But from customer point of view private room is the best as half of the year it is available. * Profitability analysis: To confirm the above inference we created some new columns which talks about revenue generated by different host and their property. First we did analysis on revenue generation of different room type by their neighborhood group. Seeing the result, we were astonished that irrespective of any neighborhood group, entire home is making way ahead revenue then other room types. Then we deep dived and look in to which host is having the highest no. of property, which neighborhood group has highest review etc.   **Challenges faced**   * Reading the dataset and understanding of columns. * For answering some of the questions we had to understand the business model of airbnb that how they work. * Handling NaN values,null values and duplicates. * Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader.   **Scope of Improvement**   * As dataset has few qualifying attributes to value a property, more features can beadded like bedroom, bathroom, property age (it might be one of the most important one),applicable tax rate, distance to nearest airport, hospital or schools. * In presence of ratings, hosts can be classified and ranked, special discount or offer can be given to highest rated hosts following marketing strategy. * Time series analysis can be done to make prediction on occupancy rate based on tourist season.   This was all about our analysis that we did, and based on result and our inferences we are making following conclusion:   * Entire home/apt is highly expensive. * Manhattan living cost is highest, Bronx living cost is lowest. * Cheapest neighborhood is Bulls head. * Cheapest listing is Bronx apart. * Manhattan have highest no. of listing. * In Manhattan entire home is mostly preferred but in Brooklyn ratio between entire home and private room is 50:50. * Private room has highest availability; Entire home has least availability. * Revenue generated by Entire home is highest irrespective of neighborhood group. * Sonder have maximum property in New York. |