**Capstone Project Submission**

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Anjali Tidke ( [anjalitidke123@gmail.com](mailto:anjalitidke123@gmail.com) )  * 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. * Exploratory Data Analysis * Data inspection * Analyzed different hosts and areas * Predictions * Profitability analysis * Visualizations * Bar graphs and Pie charts * Pair plot * Heat map * Summarization * Conclusions  1. Shubham Dukare ([shubhamdukare98@gmail.com](mailto:shubhamdukare98@gmail.com))  * Understanding Data   Understanding different column  Having overview of data  Busiest hosts   * Exploratory Data Analysis * Graph of null values * Data Visualization * Visualizations * Pair plot & co relation matrix * Conclusions |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/shubham4955/Airbnb-Analysis  Drive Link:- https://drive.google.com/drive/u/0/folders/1o-At9sevNvazMQsp7L8Igt1gLVSlIkQt |
| **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 open online platform where people list their own housing for rent. Since 2008, it has grown in popularity and especially for those communities that frequently use travel. It is becoming a strong competitor to hotel industry. It has millions of listings, which generate lots of data. We are analyzing these data for making a 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.   1. What we can learn about different hosts and areas?   As we can see most number of listings are from Manhattan created Sonder (NYC), Blueground, kara, Kazuya.   1. What can we learn from predictions? (ex: locations, prices, reviews, etc.)          we can say that most people prefer to stay in place where price is less.   1. Which hosts are the busiest and why?   We can see that Dona host is busiest.   1. Is there any noticeable difference of traffic among different areas and what could be the reason for it?       People are preferring Entire home/apt or Private room which are present in Manhattan, Brooklyn, Queens and people are preferring listings which are less in price.  This was all about the analysis that we did, and based on result and our inferences we are making following conclusion:   1. Entire home/apt is highly expensive. 2. Manhattan living cost is highest, Bronx living cost is lowest. 3. Cheapest neighborhood is Bulls head. 4. Cheapest listing is Bronx apart. 5. Manhattan have highest no. of listing. 6. In Manhattan entire home is mostly preferred but in Brooklyn ratio between entire home and private room is 50:50. 7. Private room has highest availability; Entire home has least availability. 8. Revenue generated by Entire home is highest irrespective of neighborhood group. 9. Sonder have maximum property in New York. |