**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. **Jatin (guptajatin281@gmail.com)**    * **Analyzed all the columns.**    * **Acquire and loading data.**    * **Understanding the variables.**    * **Cleaning Data.**    * **Exploring and visualizing.**    * **Plotting graphs like Count Plot, Bar Plot etc.**    * **Analizing relationship between variables.**    * **Checking for NaN values and handling NaN.**    * **Filled NaN values for columns Company and Agent with ‘0’. While for Country I filled by ‘X’.**    * **Analyzed the relationship between repetition of guest for both the hotels.**    * **Analyzed from where maximum number of guests are coming.**    * **Analyzed average waiting period for both of the hotels.** 2. **Abhilasha.M (**[**abhilasha.m03@gmail.com**](mailto:abhilasha.m03@gmail.com)**)**  * **Plotted correlation map using klib** * **Analyzed the repeated customers from various countries for both the hotels.** * **Analyzed number of bookings for both of the hotels.** * **Analyzed the busiest month for both the hotels.** * **Found out the average rent for all the months for both of the hotels and also the total revenue for both of the hotels.** * **Analyzed the waiting time and average rent for different types of customers for both of the hotels.** * **Found out the agent who has done maximum number of bookings for both of the hotels.** * **Analyzed the various ticket booking mode through which hotel booking was made.** * **Analyzed the most popular type of rooms booked and their respective rents.** * **Plotted various graphs like bar plot, line plot, scattered plot and pie chart.**   **3.) Krishan Bafna (**[**cakrishanbafna@gmail.com**](mailto:cakrishanbafna@gmail.com)**)**     * + **Analyzed Columns name and relation between the columns.**   + **Understanding the variables.**   + **Loading Data.**   + **Cleaning Data.**   + **Understanding the correlation between the graphs.**   + **Analyzed which column is not in need and can be dropped.**   + **Analyzed maximum number of guest from top 10 countries.**   + **Calculated the percentage of Repeated number of customers.**   + **Plotted the graph for hotels for longest waiting period of time.** |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/Link/to/Repo> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **About the Dataset:**  This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things from 2015 to 2017. Main aim of the project is to understand and visualize the Data from guest and hotel point of view.  All personally identifying information has been removed from the data.  The data is originally from the article Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019.  **The Approach we used in this project:**   1. **Loading the data:** In this section we just loaded the data in the colab notebook and read csv file. 2. **Data Cleaning and Processing**: In this section we have tried to remove null values and replaced all the null values by some integer values and with some appropriate values. 3. **Analysis and Visualisation**: In this section we have tried to explore all variables which can play an important role for the analysis. In the next part we have to explore the effect of one over the others. 4. **Future scope for further analysis**: We observed that the August month has the busiest month over all the years. It means the maximum business happened in the August month. City hotel is busier that Resort hotel which means the City type hotel is more demanding that is why City hotel has the longer waiting time as compare to Resort hotel. In August, July and may month most number of guests are coming and most common room booked by the guests are Room Type ‘A’. So that Hotels can arrange their facilities accordingly. There are various columns which can play an important role in the analysis such as, is\_cancelled, days\_in\_waiting\_list, avg\_waiting\_time, arrival\_date\_month etc can play an important role for analysis.   **Challenges faced**:   * I faced challenges in plotting different types of graphs. * In getting total number of bookings for each type hotel. * Some of the methods are not applied in once so I have to try a few times. * Faced problem while doing Data Wrangling. |
| **Insights we get from the Data Set:**   1. Most of the guests come from Portugal, Great Britain, France and Spain. 2. Almost 41% of City Hotel and 21% of Resort Hotel were cancelled and maximum cancellation was happened in City Hotel. 3. August month has highest number of bookings over all the years. 4. Repeated customer is approx 3.2% for both of the hotels. 5. City Hotel significantly longer waiting time, so we can say that city hotel is busier than the Resort hotel. 6. There are 46228 customers who booked City Hotel and not cancelled by the guest and 28938 customers are booked Resort Hotel and not cancelled by the customer. 7. Maximum waiting day period for Resort Hotel is 185 days and for City Hotel is 379 days. 8. August, July and May are the most busiest month over all the years. 9. Average rent Resort Hotel is maximum in the month of August and minimum in the month of January. 10. For Contract, Transient, and Group the waiting period is ‘0’ and for Transient-Party is 3.12 which means the hotel is booked immediately for Contract, Group and Transient in Resort Hotel. 11. For Contract, Transient and Group the waiting period is ‘0’ and for 7.88 days so, the hotel is booked immediately for Contract, Transient, and Group in Resort Hotel. 12. The maximum three types of room booked the most, type ‘A’ booked the most and type ‘E’ stands in third place. 13. The types of room and rent of the room are co-related. It is observed that the most common room booked are types of ‘A’, ‘D’, and ‘E’ have minimal and moderator room rents whereas rarely booked hotels have the high room rents. So, if the room rent is minimal or moderator then we can expect maximum number of bookings. |