**Capstone Project Submission**

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Ajay Singh, Email Id: [ajaysingh.4198@gmail.com](mailto:ajaysingh.4198@gmail.com)  Ajeet Kumar, Email Id: [a.k.r979374@gmail.com](mailto:a.k.r979374@gmail.com)  Laxmi Priya, Email Id: [priyalakshmi5876@gmail.com](mailto:priyalakshmi5876@gmail.com)  Shruti Sharma, Email Id: [shruti21sharma11@gmail.com](mailto:shruti21sharma11@gmail.com)  **Contruibution:**  **Ajay Singh:**   * Understand the dataset * Remove null values in the dataset and find the unique values in each column. * Unnecessary columns are dropped in the dataset and rename the column and check the outliers in the dataset and drop all the outliers in the data set. * Visualizations for the following point in the below: * Check the correlation given the columns in the dataset * Which is the most visited country by the visitors. * In which month the highest reservations are done. * Which is the most popular meal ordered by the visitors? * In which type of hotel the reservation is highest between City hotel and Resort hotel. * To check the average cancellation booking of the year of 2015-17 * What is the cancellation rate between a city hotel and resort hotel? * To check if the guest are repeated or not * In 2015 to 17 when the hotel is mostly booked by the visitors * Calculate the ADR(average daily rate ) per person * Find the ADR according to Months.   **Ajeet Kumar:**   * Exploring the data * Null values treatment * Visualizations for the following: * How many people Reserve Hotel in 2017? * How many adults, children and babies are in hotel * Cancellation in each year * Repeated guest * Booking in each year * Canceled vs not canceled booking * Online vs offline mode * Highest booking in 2015 to 2017   **Laxmi Priya:**   * Exploring the data * Null values treatment * Visualizations for the following: * Year wise Hotel bookings * Most busy month * Online – offline booking * Top 5 country most visitors come   **Shruti Sharma:**   * Exploring the data * Null values treatment * Visualizations for the following: * Which hotel is most preferred by people, Resort hotel or city hotel? * In which month, people most preferred to come in hotel * Which year had the highest booking? * Form which country most guests are coming * In which hotel people are more preferring to come in with their children and babies * What is the percentage of required car parking spaces also compare hotel wise |
| **Please paste the GitHub Repo link.** |
| Github Repositories Link:- <https://github.com/ShrutiSharma111/Hotel-booking-Analysis>  Github Profile Link:- <https://github.com/ShrutiSharma111> |
| **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:**  Hotel Industry has a very big role in the world of business and is a very big field depends upon the different factors of such as types of hotels, booking cancellation, it’s type of booking data , date ,year , month ,types of meal etc. By analyzing past dataset it is easier to find various flaws and by correcting the flaws we can improve the required business strategy, to be more effective than the others and most importantly to get more close to the interest, wishes and the need of customers and provide them satisfying service.  We are working on a hotel dataset that contains booking information for City and Resort hotels with their corresponding variables such as canceled bookings, arrival data per annum, arrival data per month, arrival data per day , types of guests( children , adults , babies) and company etc. We have total 119390 rows and 32 columns.  Firstly, we understand the meaning of each columns as it is very important to understand the data first to work effectively on the problems using provided dataset.  Secondly, we have done Filtering Data and Manipulation in which we did renaming of data set columns, finding missing values in which we found only 4 columns containing null values (.e., company, country, agent and children), made changes in null values ; we have changed null values with zero, finding duplicate values (total 31980 null values and 87230 is not null values) and dropping duplicate values after which we have total 87230 rows and 33 columns.  Lastly, Data Analysis and Visualization . we have performed various problems and their conclusions by plotting bar graph, pie chart, count plot, graphical representation( booking data by the country of origin), heat map( finding correlation between variables), line plot.  **Conclusions:**   * According to our visualization PRT (Portugal) has the highest number of visitors(27355) and GBR(Great Britain) and FRA(France) have the second and third highest number of visitors respectively. * City Hotel has the highest cancellation(61.50%) as well as highest reservation(66.4%). While Resort Hotel has lowest reservation (38.50) as well as lowest reservation (33.6%). * Year 2016 has the highest bookings in hotels and lowest bookings in year 2015. * August month has highest hotel bookings. * Most people preferred online booking mode as they find it easy. * In 2016 cancelation is maximum in the month of February (11.8%) and is minimum in the month of July (4.1%). * Average daily price for per person in Resort Hotel is high in the month of August and minimum in the month of January. For City Hotel it is almost same throughout the year. * According to our analysis only 3.2% guests do repeat in the same hotel whereas rest 96.8% guests don’t repeat. * In City Hotel ADR(Average Daily Rate) is more in the month of May, June, July and August and is minimum in January. In Resort Hotel it is more in the month of June, July and August and is minimum in November. * Most guests preferred BB meal (79.6%). Bed and Breakfast is a small lodging establishment that offers overnight accommodation and breakfast. |
|  |