**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)   **Contribution:**  **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/priyalaxmii/Hotel-Booking-Analysis>  Github Profile Link:- <https://github.com/priyalaxmii> |
| **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:**  The hotel industry has a very big role in the world of business and is a very big field that depends upon different factors such as types of hotels, booking cancellations, type of booking data, date, year, month, type of meal, etc. By analyzing past datasets, 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 closer to the interests, wishes, and needs of customers and provide them with 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 year, arrival data per month, arrival data per day, types of guests (children, adults, and babies), etc. We have a total of 119390 rows and 32 columns.  Firstly, we understand the meaning of each column, as it is very important to understand the data first to work effectively on the problems using the provided dataset.  Secondly, we have done filtering and manipulation in which we did renaming of columns, found missing values in which we found only 4 columns containing null values (i.e., company, country, agent, and children), made changes to null values (we changed null values with zero), found duplicate values (total 31980 null values and 87230 are not null values), and dropped duplicate values, after which we have totaled 87230 rows and 34 columns.  Through data analysis and visualization, we have performed various problems and reached their conclusions by plotting a bar graph, a pie chart, a count plot, a graphical representation (booking data by the country of origin), a heat map ( correlation between variables), and a, line plot.  **Conclusions:**   * 'PRT' (Portugal) has the highest number of visitors (27355), while 'GBR' (Great Britain) has the second highest number of visitors (10424), and 'FRA'(France) has the third highest number of visitors (8823). * August is the highest reservation month (11242), while the lowest reservation month is January (4685). * BB is the most popular meal ordered by the visitor (67907). * City hotels have the highest number of reservations from visitors as compared to resort hotels. * 72.4% of the bookings were not canceled, and 27.5% have been canceled. * The city hotel with the highest cancellation rate also had the highest bookings. * The bare minimum number of guests was repeated. * Most of the time, the visitors arrived in the years 2016 and 2017 at the city hotel. * The majority of visitors like city hotels. * The prices for resort hotels were higher and fluctuated more than those of city hotels. * City hotel: the ADR was more costly during the months of July, August, May, and June; resort hotel: the ADR was more costly during the months of July, August, May, and June. * Most of the visitors are coming from City hotel and Resort hotel through Online TA. * Highest average daily rate month in 2016 is August. |