**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Name:** Manisha Dhanuka  **Email Id:** [manishadhanuka5135@gmail.com](mailto:manishadhanuka5135@gmail.com)  **Contribution:**   1. Data Cleaning    * Handling Null values    * Duplicates Consideration and the assumptions 2. Data Exploration- Univariate Analysis:    * Qualitative measures of Data i.e., Descriptive Statistics    * Correlation between the features    * Booking Distribution per hotel    * Exploring the other options like yearly and monthly booking trends, meals demand, tourist’s native country, contribution of various market segments and distribution channels to the bookings etc. 3. Analyzing the factors related to cancellation:  * Are number of days in waiting list is linked to cancellation or not? * If most of the cancellations were from repeated guests or the new guests? * The number of tourists who have previously cancelled the bookings too * Tourist Cancellation with respect to deposit type  1. Average Daily Rate Analysis:  * Do bookings on weekends vs on week days affect the revenue * How ADR is varying over the years. * ADR with respect to various room types * Contribution of various market segments to ADR in these years * Contribution of various distribution channels to ADR in these 2 years |
| **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)** |
| The hotel industry is one of the most important components of the wider service industry, catering for customers who require overnight accommodation. It is closely associated with the travel industry and the hospitality industry. In this project, I explored data related to bookings of two hotels: Resort Hotel and City Hotel.  First exploration was to find out if there are any null or duplicates  in the dataset. There were many, handled them by replacing it with 0 in case of number of children and ‘No intermediaries’ in case of agents and company assuming these to be self-booked. Also, go along with the null values of country since filling /imputing it with any random value won’t have been right choice and would have affect the data accuracy. For the duplicates, I decided not to drop them as the features do not have any unique id indicating these are same or different observations, and by any chance if the duplicates were because of mass-bookings then dropping them would have affect the EDA.  Later, I did the Analysis and the main conclusions are:   1. City hotel accounts for nearly 66% of the bookings. 2. Bookings in mid-year is comparatively more in all the three years. 3. Nearly 42% of City Hotel and 22% of Resort hotel bookings were cancelled. 4. People from Portuguese visited the hotel most. But the cancellations are maximum from the too. The potential countries are Portugal, Great Britain, France, Spain and Germany. 5. ‘A’ type room is really in demand whereas the other rooms like L, K, P are rarely demanded maybe because of the price. Hotel officials can consider converting those less demanded (if number of unoccupied rooms are considerably high) into the more demanded one. 6. Travel Agents and Tour Operators are actually contributing a lot to the hotel revenue but this creates a discretion of them. They could move to other hotels if they are given more benefits. So, to avoid any kind of human- discretion, focus should also be given towards Direct bookings and that can be done through media Advertisements. Leveraging Online TA would be a great benefit using market campaigns or discounts. 7. Also, contribution of Global Distribution Systems to the ADR is maximum |