**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| 1. Ayush Goyal [erayushgoyal96@gamil.com](mailto:erayushgoyal96@gamil.com)  2. M Sameer Ahamed [sameerm8095@gmail.com](mailto:sameerm8095@gmail.com)  3. Nitesh bhowmick [nitesh.gnit@gmail.com](mailto:nitesh.gnit@gmail.com) |
| **Please paste the GitHub Repo link.** |
| https://github.com/Ayugoyal/hotel-booking-analysis.git |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| **Hotel booking analysis is done by 3 group members M Sameer Ahamed, Ayush Goyal, Nitesh bhowmick. In this project we got hotel booking analysis as csv file.**  **As we downloaded the data as csv file from alma better capstone project dashboard we encoded the file in colab notebook through mounting the drive we have no idea about the project because this was our first project .all the member from the group participated throughout the project with great efforts.**  **The data was huge and some of the column were not needed so we drop it from the data.**  **The cleaning data was done and created the new cleaned data frame consists of the column were compared to gain the knowledge for the analysis. Worked individually gaining same insights doing some EDA .**  **The first difficulty we faced was the column name and the missing data in the dataset ,As the first step we performed data cleaning of raw database, that is renaming the columns by using dictionary format, and we dropped the agent and the company column because it consists more missing data, we dropped the entire row which consists the missing data, by getting information of data frame we plotted for each data to understand and visualize thoroughly.**  **From cleaned data frame the length of stay, booking percentage by market segments, yearly booking, monthly booking, hotel prices variation by months and Country of guests visiting hotel analyzed through the cleaned data frame.**  **We identified that the guests most of the guests preferred to stay not maximum of 3 nights and some guests prefer to stay up to 5 nights, Most of the booking were done by online TA market segment,**  **The booking across years is higher for city hotel than the resort hotel and more booking are done in the year 2016 compared to 2015 and 2017, the booking across the months are higher in monsoon and summer seasons compared to other seasons, we observe that resort hotel prices are higher than city hotel, city hotel prices varies less but expensive in spring and autumn, most of the guests are from Portugal and other European countries and as the lead time increases the risk cancellation of booking also increases.**  **Contributors Roles:**   1. **Ayush Goyal:**   1. Data Wrangling:  1. Dropping of Agent and Company Columns  2. Visualizing hotel wise yearly booking  3. Visualizing hotel wise monthly booking  4. Visualizing Country wise guests  **2. M Sameer Ahamed:**  1. Data Wrangling:  1. Dropping rows which contains missing data  2. Adding of guests columns  2. Visualizing Correlation between the numerical data  3. Visualizing of hotel prices with month  4. Visualizing of relation with lead time and cancellation    **3. Nitesh Bhowmick:**  1. Data Wrangling:  1. Adding of Week stays columns  2. Renaming of columns  2. Visualizing bookings by market segment  3. Visualizing total length of stays |