**Capstone Project Submission**

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| **Name- Deepanshu Kishore**  **Email –** [**deepanshu.kishore123@gmail.com**](mailto:deepanshu.kishore123@gmail.com)  **Contribution:**   * Upload dataset to Google colab and explain dataset to team members. * Analyze null values and filter them. * Data cleaning. * Correction of data types * Data wrangling * Data Visualizations * PowerPoint presentation * Project summary   **Name- Abhishekh**  **Email –** **abhisheksingh20yadav2002@gmail.com**  **Contribution:**   * Data cleaning. * Correction of data types * Technical Write up * PowerPoint presentation * Project summary   **Sayali Mankar**  **Email – sayalimankar877@gmail.com**  **Contribution:**   * Data wrangling * Data Visualizations * Technical Write up * PowerPoint presentation * Project summary   **Gourav Dhussa**  **Email – gouravdusss143@gmail.com**  **Contribution:**   * Upload dataset to Google colab and explain dataset to team members. * Analyze null values and filter them. * Data cleaning. * Correction of data types * Data wrangling |
| **Please paste the GitHub Repo link.** |
| Github Link:- https://github.com/deepanshu2415/Capstone-1\_EDA |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| This data set contains booking information for hotels and includes information such as when the booking was made, length of stay, the number of adults, children, or babies, and the number of available parking spaces, among other things.  when the best time of year to book a hotel room is? length of stay in order to get the best daily rate? This hotel booking dataset can help us explore those questions.  First, we explore the dataset. Then look for any duplicate or null values present in dataset then after locating some we cleaned the data. We replace all the null values with zero. We also checked the total number of rows and columns present. Then we checked the statistical parameter by describing the method.  Next, we extract favourable columns which can be helpful in data visualization then we visualize the data with graph representation. First, we check which hotel type have maximum number of cancellation percentage and represent the value by the graph. Then compare the booking with hotels and year. We also represent the demand for various hotels.  Then we analyze around distribution channel that which channel is contributing more  As per the data representation we see that TA/TO has made remarkable contribution in distribution sector  Then after that we find out which agent have done the most number of booking for this we just remove all the entries which have been cancelled so that we can have accurate data that how much bookings have been done by each agent and after representing it on graph we came to know that agent with id number 9 have done most number of booking.  After that we move forward to look that is there any correlation between any of the columns or not and which type of room is mostly preferred we can conclude that number of stays and adr is slightly correlated and room type –A is widely preferred |