**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  | | --- | --- | --- | | Names | Email ID | Contribution | | Ankush Kumar  (Team Leader) | dsankushkumar@gmail.com | Code: Different Variate Analysis and correlation.  Conclusion and Summary  Power Point Presentation Preparation | | Manoj K B | kbmanoj0@gmail.com | Code: Data Insights  Power Point Presentation Preparation | | Nayan Kumar Jha | Nayan8625@gmail.com | Code: Data Information  Power Point Presentation Preparation | |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - **https://github.com/nayankr8625** |
| **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 project consists of hotel booking datasets of two different types of hotels those are 1. City hotels and 2. Resort hotel from year 2015 till 2017, Including details such as bookings, cancellations, and guest information. The project's main goal is to comprehend and visualize data from the hotel and customer perspectives to get the proper insights from the dataset and to make proper data driven decisions.**  **Initially our dataset contains 119390 rows and 32 columns. Our first goal is to analyze the datasets and figure out the numerical and categorical features/columns. And also, to identify null values and duplicated value. We have also checked the statistical description of numerical columns and summary of categorical column to understand the data better.**  **There were several null values rows and many duplicate rows in our datasets so**  **our first goal was to cleanse the data for doing the further analysis. After that we have done data wrangling and added some new columns in our dataset for doing further analysis.**  **In the next part after fixing our data and making it ready for analysis, we have done some visualizations as well as analysis and gain some data-driven insights by using certain feature present in our datasets and also by using some general analysis technique to solve the data related question by following certain analysis method such as Univariate, Bivariate and Multivariate Analysis. I am going to list all of the analysis that we have done:**  **UNIVARIATE ANALYSIS: -**   1. **Observation from hotel type that which type of hotel has how many bookings** 2. **The count of bookings is canceled** 3. **Observation on Market Segment wise bookings** 4. **Observation on Distribution Channel wise bookings** 5. **Is Hotel booking is canceled and is it done by the repeated guest or not**   **BIVARIATE ANALYSIS: -**   1. **Market segment and days on waiting list for each of them** 2. **Comparing Distribution Channels and days on waiting list for each of them** 3. **Per month arrivals hotel bookings** 4. **Market Segment wise bookings for each Hotel Type** 5. **Distribution Channel wise bookings for each Hotel Type** 6. **Months of the year with the lead time** 7. **Arrival of customers per day of the months to the hotels** 8. **Demand of parking space by customers in each type of hotels**   **MULTIVARIATE ANALYSIS: -**   1. **We have done Multivariate analysis to check hotel bookings occurred according to the average daily rates per month.**   **After using these three analyses, we have also used correlation matrix to check the correlation in the whole dataset. (By generating heatmap)**  **Afterward we had also done several hypothesized questionings and did some analysis to figure out the data driven answers. Those questions are mentioned below**   1. **Do customers have any preference with or without children.** 2. **How daily average rate is impacting the reserved room type in hotels.** 3. **Does total stay per month impacted by customers with and without children.** 4. **Do reservation status impacted by type of customers.** 5. **Do customers cancel their booking if they are allotted with different room type.**   **We have asked these questions to our data and figure out relevant data driven answers to gain some proper insight.**  **Then after doing this all-exploratory data analysis, we have come to data driven conclusions and all the conclusions are mentioned below.**   * **The vast majority of reservations are for hotels in cities. Resort hotels have fewer cancellations than city hotels. The aviation industry has the shortest wait time.** * **The months of August, July, and May saw the most hotel bookings in the city. The lowest cancellation rate occurs when a consumer is repeated. The lead time for July is quite long, whereas the lead time for January and February is extremely short. In July, August, and September, the average daily rate for Resort Hotels is higher.** * **In June, July, and May, the average daily rate for City Hotels is higher. Customers travelling with children have little preference for the type of hotel they stay in.** * **In both cases, the number of stays with and without children is rather equal. Customers who stayed for 1-7 days with or without children had the same count.** * **If we take the average stay, three to four days will be spent with youngsters, while the rest of the time will be spent with adults solely.** * **Online TA customer group has the highest number of hotel reservations in the city and at resort hotels. As a result, the impact of room type allocation on clients is minimal. The majority of appointments are made by consumers of the Online TA group who book directly through the website.** |
|  |