**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Purushotam [civilengineer.purushotam@gmail.com](mailto:civilengineer.purushotam@gmail.com)  Highly committed coordinated team work for successful completion of the project.  Rajat Sharma [rajatsharma6870@gmail.com](mailto:rajatsharma6870@gmail.com)  Contributions:  Highly committed coordinated team work for successful completion of the project. |
| **Please paste the GitHub Repo link.** |
| GitHub Link: - <https://github.com/Purhsotam/projects.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)**  **Problem Statement**  Question 1. Which customer type preferred the most resort hotel?  Question 2. Which type of hotel refunded the most?  Question 3. Which distribution channel is mostly used for booking City hotel?  Question 4. Which type of hotel is preferred by customer having children?  Question 5. Which type of hotel is most preferred for weekend night?  Question 6. Which year has highest booking and cancellation based on market segment (complementary)?  Question 7. preferred food by customer based on top 5 country?  Question 8. Most favorable month for booking :-  Question 9.Which are the top five agents for city hotel?  Question 10. Hotel data based on month where repeated guest is 1:  **Approach**  We have divided the data into three parts  Data collection: -  Dataset has been provided by the Alma Better https://grow.almabetter.com/data-science/projects/Hotel-Booking-Analysis  Data loading: -  Dataset was loaded on google drive was loaded on goggle colab ('content/Hotel Bookings raw.csv’)  Data Cleaning and manipulation: - For data cleaning NumPy, pandas, and matplotlib is used  Removing the duplicates: - Analysing the data we found out of 119390 rows 31994 rows are duplicates so we removed them.  Removing the company column with most nan values: - Analysing the company column we found it has only 5259 values are not nan remaining does not provide any significant data. So, we deleted the column.  Beginning with EDA: - |
| **Conclusion**   1. **Transient types of customers visited to Resort Hotel most.** 2. **City hotel has non refunded the most 845 compared to Resort Hotel having 193.** 3. **TA/TO is most preferred for City Hotel 44549.** 4. **City hotel is more preferred by customers having children with age group as per dataset.** 5. **City hotel is most preferred for weekend nights.** 6. **The out of 2015, 2016, 2017 2016 has the highest booking and cancellation based on complimentary** 7. **BB is most preferred by the Portugal** 8. **August month is most preferred for the booking** 9. **9.0 is among the top five agents for the City Hotel** 10. **So, March has the most Repeated guest 418**   **From the above conclusions we can say that City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.** |