**Capstone Project Submission**

**Instructions:**

1. Please fill in all the required information.
2. Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  | | --- | --- | --- | | NAME | EMAIL | Contribution | | TEJAS PATIL | Tejas9patil@gmail.com | DATA ANALYSIS AND VISUALIZATION | | RIYA GOYAL | goyalriyamarch@gmail.com | DATA ANALYSIS AND PRESENTATION | | KAILASH SING | Kailash.rid.444@gmail.com | DATA PARSING AND DOCUMENTATION. | | SAMRUDDHI PARIHAR | Samruddhi.parihar9049@gmail.com | ANALYSIS AND PRESENTATION | | RUPESH BHERANE | rupeshbherane@gmaill.com | ANALYSIS | |
| **Please paste the GitHub Repo link.** |
| Github Link:- |
|  |
| **SUMMARY:**  **The data set that we are using for our project is an information of hotel industry comprise of two sorts of hotels. One is Resort Hotel and other one is City Hotel. Total entries in the data set are approximately 119390. Based on this data, we created some problems and tried to get a solution for these problems by analyzing the information from the data set. There are 16 convenient problems and their answers. We tried to make the code understandable to all and the conclusion we get are valuable to clients as well as business perspectives.**    **Problems:**  **1)Problem:** Best time to Book a room ?    **Approach :**We first filter the data based on total monthly bookings done and then plot thebar graph    **Conclusion :** In the year of 2015,september got highest traffic and in year 2016 April was most booked month and in the year 2017 month of May received high traffic    **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**    **2) Problem:** Which Market segment is most profitable (ADR)?    **Approach:** We group the column of hotel and market segment on the basis of market segment count and then we get a new Data Frame |

|  |
| --- |
| **Conclusion:**Total ADR generated from online travel agent is the highest both in the city hotel and the resort hotel.     1. **problem:** Calculating Which Hotel is likely to get Special Request     **Approach**: First we take data from two columns. We findtotal special request on hotel basis then we calculate total bookings done and simply use the probability formula for calculating series.      **Conclusion:** From the graph red bar is indicating probability in percentage of getting special request to city hotel and blue is indicating probability in percentage of getting special request to Resort hotel.  Resort Hotel is more Likely to get special request than City Hotel.     1. **Problem:** Which Nationalities Customer visit most?     **Approach:** We made two groups of hotel and country based on number of country,sorted tge values and then we take last five values of data to get top 5 most visited nationalities  **Conclusion:** As you can see from above graphs and tables portugal cutsomers are dominating in both types of hotel ,so we can conclude hiring staff which knows portuguese better would always benefits APA group new hotel , and using this data they can also run profitable ad campaigns.     1. **Problem:** Total Population Proportion.   **Approach:** We filter number of Adults, children and babies based on their age and find their proportion from the total population.    **Conclusion:** We find that total adults in data are 221636 total children in data are 12403.0 total babies in data are 949.     1. **Problem:** How many days does customers prefer to stay the most in hotel?   **Approach:** From the data we find how many nights does customers book their hotel and then find how many days most of the customers prefer to stay in hotel.  **Conclusion:** Three days stays preferred by most hotel.     1. **Problem:**  What is the percentage of cancellation in city hotel and resort hotel.   **Approach:** From the dataset we find the booking and cancellation rate in both resorts and city hotel. Then we find which hotel type has higher cancellation rate.  **Conclusion:** Observed that 37% customer were cancelling booking of city or resort hotels**.**     1. **Problem:** Time Series Analysis on most lucrative market segment Online TA   **Approach**: From the data we find how many bookings were done every month by both hotel types individually. Then we find for which month most of the bookings are done**.**  **Conclusion:** We see from the graph that there is seasonal variation in hotel bookings both hotel is following same trend but with different numbers ,August is most visited month and january is least visited in both types of hotel |

|  |
| --- |
| 1. **Problem:** How Many customers require car parking spaces   **Approach**: From the data we will find how many customers would like to have parking facility in hotel.  **Conclusion:** From graph and pie chart we can observe 93% customers require car parking space.     1. **Problem:** What type of meal customers prefer to have in hotel   **Approach**: From the data we will find which is the most preferred meal by the customer booking the hotel.  **Conclusion:** Observed here that Highly preferable meal type is BB(bed and breakfast) which is required 77.3% and second most is HB(half board) which is required 12.1%     1. **Problem:** Predicting whether or not a hotel was likely to receive a disproportionately high number of special requests?   **Approach**: From the data we will find how many special request are made by the customers.    **Conclusion:** Both city hotel and resort hotel receive most number of requests in the month of August     1. **Problem:** How Many Guests Are Repeated   **Approach**: This data will be very helpful in retargeting customer base for the marketing campaigns and building a strong customer base.  **Conclusion:** Out of total 87230 customers, 83866 are repeated and 3364 are not repeated.     1. **Problem:** Hotel Demand by Market Segment   **Approach**: This data will help hotel businesses to set their market price based on the increasing or decreasing demand  **Conclusion:** Both City Hotel and Resort hotel is in demand among customers belonging to market segment of online TA.     1. **Problem:** Number of arrivals every day of the month and year   **Approach**: Analysis of this data will help hotel businesses to understand how the market is growing every year.  **Conclusion:** From the graph we can say in year 2015 number of arrivals were same but in 2016 city hotel has more arrivals than the resort hotel and same trend is followed in the year 2017 as well.     1. **Problem:** Total Bookings Canceled in every months of the year   **Approach**: This data can help hotel management to understand why the cancellations occurred in the particular month and hence they can improve their service for the future  **Conclusion:** From the graph month of August faced most cancellation of bookings and month in of January customers canceled bookings least. |
| **16) Problem**: Co-relation of the column  **Approach**: Co-relation is measure of strength of relation between two variables like, if there is positive co-relation between two variable then the are strongly co-related for example if demand is increasing supply is increasing ,and negative co-relation specify just vice versa  **Conclusion:** In the above co-relation matrix we have co-relation between different columns of Data Frame and we can compare the two columns based on their annoted values take "is\_cancled" and "lead\_time' column for example you can there is positive co\_relation between these two columns but not very large                                        **THANK YOU** |