**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| |  |  |  | | --- | --- | --- | | NAME | EMAIL | Contribution | | TEJAS PATIL | Tejas9patil@gmail.com | DATA ANALYSIS AND VISUALIZATION | | RIYA GOYAL |  |  | | KAILASH SING | Kailash.rid.444@gmail.com | DATA PARSING AND DOCUMENTATION. | | SAMRUDDHI PARIHAR | Samruddhi.parihar9049@gmail.com | Analysis and presrntaion | | Rupesh Bherane | rupeshbherane@gmail.com | Analysis | |
| **Please paste the GitHub Repo link.** |
| Github team repo:- [Team project repo link](https://github.com/samruddhi9049/Project1.git)  Github personal contribution in a team link: [Personal Contribution to team](https://github.com/samruddhi9049/own-contribution.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)** |
| **SUMMARY:**  **Data provided is a data of hotel industry consist of two types of hotel namely Resort Hotel and City Hotel total entries are 119390 ,Our (Data Digger Team) Analysis of data comes in 10 handy problems and their solutions precautions have been taken to make the code comprehensible to all and provided conclusions are useful to customers as well as business point of view.**  **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 :** As you can see from graph 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 DataFrame    **Conclusion:**Total ADR generated from online travel agent is the highest both in the city hotel and the resort hotel.  **3) 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.  **4) 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. **5) 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. **6) 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. **7) 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**.** **8) 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 |

9) problem : What is the percentage distribution of required\_car\_parking\_spaces?

Approach : we have made pie chart of number of customers requiring car parking space

Conclusion : 93% customers required car parking space

10)problem:which meal type is highly preferable

Approach : we have made a bar chart of different meal type

Conclusion: 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%

11)problem:hotel likely to get disproportionately high request

Approach : we have made bar graph showing high request

Conclusion: Highly disproprtionate special requesst is coming in month of june ,in which huge share of special reuest goes to city hotel and very low share to resort hotel

12)problem : is guest repeated or not

Appeoach: we have made bar graph showing how many guest are repeated

Conclusion: 83866 are repated guest and 3364 are not repated guest

13)problem :hotel demand in different segment of market

Aprroach:we have made a bar graph showing total demand variation in different market segmnet

Conclusion: Both City Hotel and Resort hotel is in huge demand among customers belonging to market segment of online TA

14)problem : Number of Arrival in a year

Approach : we have made a bar graph showing number of arrivals in a year and on daily basis also

Conclusion: from first graph we can say in year 2015 number of arrivals were same but in next year 2016 city hotel ovrtake the resort hotel and in year 2017 same trend continues

15)problem : Number of room booking cancelation on monthly basis

Approacch :we have made a stem graph showing canceled booking in a month

Conclusion : you can see from graph month of august faced most cancellation of bookings adn month in of januray customers canceled bookings least

16)problem :Corelation Between column

Approach:we made a heat map showing correlation between two columns