**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email, and Contribution:** |
| --- |
| Name: Mohan Prakash Vishe  Email: [mohanvishe1999@gmail.com](mailto:mohanvishe1999@gmail.com)  Contribution:  1. Mohan Vishe([mohanvishe1999@gmail.com](mailto:mohanvishe1999@gmail.com)):  1. Introduction  2. Feature information  3. Data Wrangling:  l. Data Cleaning and Data Manipulation  2. Univariate analysis  4. Bookings Cancellation analysis  5. Stay Analysis  6. Meal booking analysis  7. Correlation between essential features  8. Relation between ADR and the Number of People  9. Country-wise bookings  10. Hotel Comparison by bookings  11. Conclusion  2. Gaurav Yadav([gauravyadav7939@gmail.com](mailto:gauravyadav7939@gmail.com)):  1. Data Wrangling:  1. Data Cleaning and Data Manipulation  2. Univariate analysis  2. Hotel booking analysis  3. Time-wise analysis  4. Cheapest bookings  5. ADR price analysis  6. Assigned and Reserved room type  7. Correlation heatmap between features  8. Market and distribution channel analysis  3. Rahul Ray(ryrahul124@gmail.com):  1. Data Wrangling:  1. Data Cleaning and Data Manipulation  2. Univariate analysis  2. Yearly analysis  3. Monthly analysis  4. Country-wise analysis  5. Meal analysis  6. Assigned and Reserved room type  7. Reservation Status  8. Repeated Guests  4. Shambhuraj Desai(shambhurajdesai97@gmail.com):  1. Data Wrangling:  1. Data Cleaning and Data Manipulation  2. Univariate analysis  2. Required Parking spaces  3. Meal analysis  4. Reserved and Assigned Room Type  5. Total special requests  6. Bookings Cancelled or not  7. Stays on weeknights |
| **Please paste the GitHub Repo link.**  **GitHub link-**  [**https://github.com/MohanVishe/Data-Analysis---Hotel-Booking-**](https://github.com/MohanVishe/Data-Analysis---Hotel-Booking-) |
|  |
| **SUMMARY**  Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions! Whenever we think about hotel booking the words that come to mind  are money required and budget, location, services, cleanliness, and facilities. Feels like this is more important. Some things are adjustable according to price and some features can be rewarded such as free wifi and breakfast.  The “Hotel Booking Dataset “ gives us information about whether the booking was cancelled or not. There are two types of hotels in the dataset called city hotels and resort hotels which contain booking date and time, the number of adults, children, and babies, duration of stay, reservation type, etc. All personally identifying information has been removed from the data.  The dataset is in table formation which contains rows which are also called datapoints there are 119390 rows and columns are also called  features. There are 32 columns. To extract data, clean, and visualize we used the libraries such as Pandas, Numpy, seaborn, and matplotlib.pyplot. As this dataset has many rows and columns it is hard to find information about each feature. Some features have numerical and some have categorical values. So it is hard to find the importance of various features. The data type “object” is hard to interpret some have a large category of values. There are various features with missing or null values. So we have to remove some features. So we first remove less important features such as “company” and “agent” with high null values.  There are some features such as country and children which are important and contain fewer null values. So we remove rows only. As feature children should be integer values so we convert the column into a data type integer.  For EDA I started to find the answer to some questions with the help of visualization of the data set. For visualization, we used the library such as matplotlib.pyplot and seaborn. Sometimes I have to add, subtract and merge columns to get a new column for visualization such as “People” and “stay”. As the data contain booking canceled information also we have to remove that data and create a new dataset for some visualization. There is also the presence of outliers so we have to ignore them.  There is various data analysis task that we have done on this dataset they are booking cancellation analysis, stay analysis, meal booking analysis, the correlation between essential features, Relation between ADR and number of people, country-wise booking, Hotel comparison by bookings, etc. From visualization, we get to some conclusion as follow:  **Conclusion**  1. There is less number of feature with missing values and outliers so dealing with them is easy.  2. After the peak in 2016, there is a fall in bookings in the year 2017. After 2015 bookings increased in 2016 then decreased in 2017.  3. City hotels have a higher number of bookings so we have to focus on them and May to Aug period is targeted as there is the peak of the summer period.  4. 1/3 of bookings were canceled Which is worse.  5. As the booking rate for country PRT, GBR, and FRA is high so we have to focus on advertisement in that country.  6. As the average daily rate increases with an increase in the number of people so we have to encourage a large group of people/families to book by giving offers and discounts.  7. Hotels should focus on the months of summer as there is more rate in that time.  8. We can say that more than 93% of customers do not prefer the parking, so we can commit that most of the customers reach the hotel by transport.  9. The hotel should be near stops of public transport    10. We should give less space for parking  11. From the above, we see that the majority of the distribution channels and market segments involve travel agencies (online or offline). We have to focus on targeting these travel agencies' websites and work with them since the majority of the visitors tend to reach out to them..  12. For both city and resort hotels, November to January has a cheaper average monthly rate. |