Capstone Project Submission

Instructions:

i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Team: Alma Titan Name: Arun Prasath R Email: arunraviprasath@gmail.com Contribution: 1. Prepared 14 different questions for analysis. 2. Understanding the data and describing the data variables. 3. Data cleaning by dropping the columns of high null values, filling the null values in some columns. 4. Removing the outliers. 5. Converted different object types into proper types for example values which should be in date format and integer format. 6. Completed the analysis for all 14 questions prepared. 7. Generated the plots such as bar graph, line graph, pie chart, box plots. 8. Contributed to combine all individual notebooks to make team Collab. 9. Contributed to make the PPT. 10. Write the technical report.
Please paste the GitHub Repo link.
Github Link:- https://github.com/ArunPrasath1205/Exploratory-Data-Analysis Drive link:- https://drive.google.com/drive/folders/1_YQ7OfzLvgO6_u-SOFrmdCYBFdPxAdZQ?usp=sharing
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)
Data set name - Hotel Booking Analysis Shape – 119390 rows × 32 columns Columns - 'hotel', 'is_canceled', 'lead_time', 'arrival_date_year', 'arrival_date_month', 'arrival_date_week_number', 'arrival_date_day_of_month', 'stays_in_weekend_nights', 'stays_in_week_nights', 'adults', 'children', 'babies', 'meal', 'country', 'market_segment', 'distribution_channel', 'is_repeated_guest', 'previous_cancellations', 'previous_bookings_not_canceled', 'reserved_room_type', 'assigned_room_type', 'booking_changes', 'deposit_type', 'agent', 'days_in_waiting_list', 'customer_type', 'adr', 'required_car_parking_spaces', 'total_of_special_requests', 'reservation_status', 'reservation_status_date' Problem Statement : - For this project we will be analyzing Hotel Booking data. This data set contains looking information for a city hotel and a resort hotel , and includes information such as when the booking was made, length of stay ,the number of adults ,children and/or babies, and the number of available parking spaces. - Hotel industry is a very volatile industry and the bookings depend on the above factors and many more. - The main objective behind this project is to explore and analyze data to discover important factors that govern the bookings and give insights to hotel management , which can perform various campaigns to boost the business and performance. Approaches : 1. Understand the data. Data understanding focuses on the comprehension of the information available in the project. In this step we basically check on the kind of variables provided with the dataset, dtype of the columns, shape of the data frame. 2. Basic cleaning. Our dataset contains numbers of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result. Pandas isnull() and notnull() methods are used to check and manage NULL values in a data frame

3. Selecting the variables to analyze the data:

We prepared some questions by selecting the different variables and relation between them. Then we generated the graphs such as bar chart, box plot, line chart for each question.

Conclusion :

Some conclusions drawn from the analysis are as follows.

- Customers preferred City Hotel more than Resort Hotel.
- The maximum length of stay is higher (than city hotel) in resort type as the resort is mostly used for vacation purposes.
- Median value of staying days of both the hotels are approximately equal.
- Bookings in the month of August are highest and January found lowest number of bookings.