Capstone Project Summary

Team Member's Name, Email and Contribution:

1. Tito Varghese
   Email - tito.varghese1992@gmail.com

   - Removing the Outliers from the Dataframe
   - Factors Governing the Hotel Bookings
   - To predict whether or not a hotel was likely to receive a disproportionately high number of special requests?
   - Factors Governing the Cancellation
   - Find the best customer type with maximum Average Daily Rate in Different Year

2. Anmol Raj
   Email – actuallyanmol@gmail.com

   - Removing the Duplicate rows from the Dataframe
   - Hotel with highest Revenue
   - Countries from which most guests are coming
   - Which Hotel has highest waiting time
   - Which distribution channel brings highest revenue generating deals

3. Lakshmi Keerthana
   Email - keerthana826@gmail.com

   - Handling the missing values in the Dataframe
   - The most used Distribution Channel
   - The most preferred meal type by the customer
   - Market segment used by the guest in different hotels
   - Heat Correlation Map

4. Kamatam Harshith
   Email - 1harshasai@gmail.com

   - Converting Columns to appropriate data type and adding important columns
   - Which hotel has a higher chance that its customer will return for another stay
   - Which hotel has higher booking cancellation rate
   - When is the best time in a year to book a hotel room
   - Type of customer booking the most

Please paste the GitHub Repo link.

GitHub Profile Link: - https://github.com/7692TITO

GitHub Repository Link: - https://github.com/7692TITO/EDA-CAPSTONE-PROJECT

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Hotel Booking Analysis

In this Capstone Project on Exploratory data analysis, the prime aim has been to understand all the key concepts taught in python modules and to use the tools like python coding, pandas libraries and data visualization libraries in order to extract, analyze and to get insight into the given hotel bookings csv dataset from the timeline 2015-2017.

Additionally, we have analysed the data to understand when the best time of year to book a hotel room is? 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?
Factors governing cancellation of bookings?

Our approach to solving these problems were as follows.

- Loading the data from the drive into dataframe.
- Cleaning the dataframe
- Data Analysis using pandas library
- Data Wrangling and Visualization

The dataset consists of 119390 rows and 32 columns.
Initially, we mounted our drive into Collab notebook and using pandas library we successfully read the given csv file. The next step was data cleaning and removing all the duplicate rows and outliers from the data. we have also handled the missing values, converted the columns to appropriate datatype and add some useful columns from the data.

Thereafter, we have used various data wrangling techniques and different data visualization plots to get insight into our dataset and tried to come up with useful conclusions for various exploratory data analysis problems.

Some of our important conclusions and results from our exploratory data analysis were as follows

- Nearly 60% of guests prefer City Hotel and only 40% of guests prefer Resort Hotel while booking.
- Most of the customers come in the month of August and we can see the least customers in December
- Most bookings are done by couples. Hence providing good deals for couples can increase the number of bookings
- The reason behind hotel booking cancellation is due to no availability of car parking space.
- The number of days stay preferred by the most number of customers is 1 day with nearly 25% secondly, the preferred no. of days stay is 3 days.
- If the number of adult guests is 3 and if the number of kids is 1 or 3, then we can expect the highest number of special requests.