SUMMARY

In this project, we have tried to predict the possibility of a hotel booking based on different factors. It was important to understand the guest preferences (locations, special requests, and room types), purchase behavior (market segment, length of stay, time of year). it might help us make better decisions.

The process of our analysis has the following step: understanding the Datasets, Data preparation, and cleaning, analyzing the data, and conclusion. We started with importing some of the useful python libraries like pandas, matplotlib, and seaborn needed for the project. After the basic operations of importing the dataset, seeing at the overview of the dataset we started with cleaning the dataset. We mostly focused on the null values, duplicates, and dropping unnecessary columns in cleaning the dataset.

After that, we started Exploratory Data Analysis to analyze the data with the help of inbuilt seaborn and matplotlib libraries, plot various graphs including histogram, bar graph, scatterplot, boxplot, heatmap, line graph, etc. and got some of the following conclusions.

- The data set contains two hotels one was a city hotel and the second was a resort hotel, which has 1,19,390 rows and 32 columns. The city hotel was more preferred than the resort hotel. Months between May and August have maximum bookings and the year was 2016.
- Most of the market segment was online TA or offline TA/TO, even the distribution channel was occupied by Online TA/TO
- Most customers prefer Bed and Breakfast package.
- The majority of the people were assigned the room which they reserved.
- The Repeated guests were very few and Maximum cancellations were from the new guests.
- We conclude that Portugal (PRT) was the country with the highest bookings after that we have Great Britain "GBR”.

- Group segment has maximum waiting time whereas aviation segment with the minimum waiting time.

Contributor Roles:

1. Tanjul Gohar
    - Data Cleaning
    - Handling outliers with mean
    - Univariates analysis
        a. Preferred Hotel
        b. Arrival date month
        c. Arrival date year
        d. Preferred Meal
        e. Assigned room type
        f. Reserved room type
        g. Market segment
        h. Distribution Channel
        i. Booking changes
        j. Customer types
        k. Repeated guest
    - Bivariate analysis
        a. Percentage of canceled bookings
        b. Cancellation by repeated guest
        c. Special request for cancellation
        d. Arrival month for each hotel type
        e. Bookings for top 10 countries
        f. ADR in each month
        g. Market segment with a waiting list
    - Correlation map for whole data

GitHub Link: **https://github.com/Tanjul5/Hotel_Booking_Analysis.git**

Drive Link:
https://drive.google.com/drive/folders/1gOa2Z0s379vBUDdngqKoBCtDZGhvm3r7?usp=sharing