Capstone Project Submission

HOTEL BOOKINGS

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

IRSHAD, GMAIL: irrshadkhan@gmail.com
ROLES: 1. Data Wrangling 2.Count number of rows and columns in the dataset 3.Find the missing values in the dataset 4.Drop low percentage of missing values 5.Find and remove outliers 6.Count number of person book a room 7.Merge male , female and children come in the room 8.Arrival date 9.Cancelled booking 10. Repeated guest 11.Check arrival date by month 12.Count arrival date month by count plot 13.Determine market segment for the hotel booking 14.Count year of arrival 15.Meal package for the guest in the hotel 16.Costumer relation with the hotel booking 17.Reserved type room 18.count number of booking changes by the costumer 19.Car parking facility for the customer 20.Type of hotel means 3 star , 5 star ect 21.Relation between price and month for the hotel booking
Please paste the GitHub Repo link.
Github Link:- https://github.com/irshad9873/Hotel-Bookings
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

This data set contains booking 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, among other things. All personally identifying information has from the data.

The main aim of this chapter is to build model which predicts whether guest cancels reservation and use some explanation methods to analyze the reasons of customer behavior

This dataset contains booking information for two types of hotels - city and resort - 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, among other things.

The data is imported into R and subjected to preliminary data analysis to understand the structure and variable types. Summary statistics are computed to reveal distributions of each variable, missing values and outliers. The outliers and missing values are treated as required.

Univariate and bivariate plots help determine any hidden patterns or relationships within the variables. We will also derive features from existing variables that shed light on customer behavior using feature engineering.

Based on the inferences from the aforesaid exploratory analysis, we will attempt to implement a regression model to attempt to estimate the length of stay and/or booking cancellations. We will also see the feasibility of using clustering techniques to identify customer groups that follow certain booking patterns.

The objective of this study is to help the end-user understand any underlying pattern in hotel bookings considered for the study and provide the proposed models to estimate length of booking or to predict whether a particular booking will be cancelled.