HOTEL BOOKING ANALYSIS

Naramalla shirisha Yembaiah

DATA SCIENCE TRAINEES
EDUREKA-KPHB(Hyderabad)

ABSTRACT: This data article describes two datasets with hotel demand data. One of the hotels is a city hotel and the other is resort hotel. Both datasets share the same structure with 32 variables describing 40,060 observations of resort hotel and 79,330 observations of city hotel. Each observation represents a hotel booking. Both datasets comprehend bookings due to arrival between 1st July 2015 and 31st of August 2017 including bookings that were cancelled. Since this is hotel real data, all data elements pertaining hotel or consumer identification were deleted. Due to the scarcity of real business data for scientific and educational purposes, these datasets can have an important role for research and education in revenue management, machine learning, data mining as well as in other fields.

This project main objective is to explore and analyse the data to discover important factors that govern the hotel bookings.

1.Problem statement

Hotel bookings data base provides information on two hotels business during three years 2015 to 2017. This data base gives the opportunity to understand the pattern and trends of bookings made during the three years of duration. By ED Analysis of this data set, we will infer and find out the scope for improvement, in order to conclude the opportunities for business improvement (to increase the hotel bookings).

2.Introduction

Hotel bookings are done in two hotels - one in city hotel and other in resort hotel. Bookings are governed by various factors such as location of the hotel, rate per day, type of booking (no deposit, refundable and no refund etc) ,
EDA - Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns , trends and distribution of data to infer and draw the conclusions.

3.Steps involved

Importing Libraries:

1.Numpy:- it is a python library used for working with arrays(mathematical operations).

2.Pandas;- is a high level data manipulation tool.
3.Matplot :- is an amazing visualization library in python for 2D plots of arrays.
4.Seaborn:- is a library for making statistical graphs in python.

Importing Dataset:- it has been imported from Google drive and read this dataset using read_csv.

Null values Treatment: As our dataset contains null values, we removed them at the beginning of our project to improve the data accuracy
.
Duplicate values:- Length of duplicate values found to be 31994, which were removed(dropped).There are 87396 rows and 32 columns after removing null values

Observations:

1.City Hotel is found to be a "most preferred" hotel than "Resort Hotel".

2.Country PRT (Portugal) found to be contributing to more business with 27449 guest
entries in the data set.

3.Online TA market segment is top one among other eight segments with 56477 guest
entries.

4.TA/TO distribution channel is top one with 82% business - leading other channels.

5.NO DEPOSIT category found to be attractive as 87.6% entries are (guests) enjoyed
the benefit out of it.

6.TRANSIENT customer type is occupying 75% of business.

7.77% of "BB" Meal type liked by guests.

8.Year 2016 has done more business than 2015 and 2017

9.Stays in "Week nights" are more than "Week end nights"

10. Week number 28 and 31 (pertains to July) contributed to more business irrespective of year.

Conclusion:

With this data analysis, we conclude that guests from Europe continent are more during July (Summer holidays).