Hotel Booking Analysis

Data Science Trainees
Alma better, Bangalore
Vikas kumar

Abstract :

My project is on EDA(Exploratory Data Analysis) done on Hotel Bookings. We are given a data set of hotel bookings in which all the details of hotel bookings are provided such as types of hotel, bookings, cancellations, and bookings done by users according to country. We are doing Exploratory Data Analysis using python. In python, we use Pandas, Numpy, Seaborn, and matplotlib.

Keywords- Python for data science

1. Problem statement:

1. Finding the correlation between the numerical data.
2. Preferred hotel types.
3. Which month has the most peak booking?
4. Optimal length of stay in the hotel.
5. Which year has the most peak booking?
6. Which country has the most number of bookings done by the customer?
7. Preferred stay in each hotel based on the total number of bookings.

2. Introduction:

In this project, we are given a data set based on hotel bookings. In the given data set there are 32 columns and 1,19,391 rows are there.

The first column is about the hotel types used for bookings similarly each column comprises specific data such as cancellations, lead time, arrival date- years months, weeks, and days. Also total number of stays, number of adults, children, and babies, and the country whose hotel books more frequently.

Types of hotels used for holidays:

We have two types of hotels used for holidays.

- Resort Hotels.

- City Hotels.

Correlation between numerical data of columns:

- In the plotted heatmap adr(average daily rate) is correlated with no_of_people and it indicates that more revenue is generated when the number of people increases.
- Then, lead time and the number of stays are correlated which indicates that people take longer stay in hotels when booking is done more before they arrive to stay.

Steps involved in this project:

Step 1:

In the first step, we wrote the python programming to find some of the results. We had different types of datasets in which there were different columns. So we extracted the null and unique values. Then we wrote the programming to draw the graph.

Step 2:

In this second step, we designed our PowerPoint slides using google ppt. In this step, we discussed the result that we observed in python programming.

Step-3

In the last step, we prepared the presentation video for the project. In this part of the step, we discussed different results through the videos and audio with PowerPoint slides.

Conclusions:

1) Here adr(average daily rate) is correlated with no_of_people. It simply indicates that more revenue is generated when the number of people increases.
2) Then, lead time and the number of stays are correlated. It indicates that people take longer stay in hotels when booking is done more before than they arrive to stay. 3) The percentage of city hotels is 66.45% while the percentage of resort hotels is 33.55%.
4) August is the month or we can say that the time of the year in which hotel booking is at its peak. And January is the month in which hotel bookings are the least.
5) According to the given data set, the optimal length or favorable time of stay is less than 1 day for weekends and 2 days for weekdays.
6) The year 2016 is the highest booking year with a total booking of 56,707 then booking decreased to 40,687.
7) Most of the guests came from European countries, with most no. of the guests coming from Portugal.
8) Most people stay 1 to 4 days. Generally, people go to City Hotels for a short period of time but for longer periods of stay, people prefer Resort Hotels.