Hotel Booking Analysis

Technical documents

KISHORE KUMAR,
DATA SCIENCE TRAINEE,
ALMABETTER, BANGALORE

Keywords: City hotel, Resort hotel, data mining, revenue management

***ABSTRACT*:**

This dataset demonstrates the hotel booking pattern of basically two types of hotels, 'City hotel' & 'Resort hotel'. Initially, the dataset had 32 variables describing 119390 observations. Each & every observation in the dataset represents a hotel booking. This dataset comprehends the booking data between 2015 & 2017. Since, this is a real hotel booking data, all the data elements pertaining customer as well as hotel identification were deleted. This dataset can definitely play an important role for research & analysis on revenue management, data mining as well as in other fields.

***BUSINESS PROBLEM*:**

'Hotel industry' is a very volatile industry & the bookings depend on a variety of factors. This makes analyzing the patterns available in the dataset which in turn would help the hotels plan better. Different factors affecting the booking pattern, is used to report the trends & predict the future bookings. So, the main objective is to perform exploratory data analysis with the past data on hotel bookings & extract some fruitful insights in order to help the hotels boost their business.

STEPS INVOLVED:

PART-(1) DATA COLLECTION & INSPECTION OF DATASET:

- **Data Understanding:** Simply inspect the data set as following steps like loading the dataset, checking shape of data set, information about data set etc.

PART-(2) DATA CLEANING & MANIPULATION:

- Simply here identify and dealing with null values of various features and manipulating various features.

- **Segregation of Continuous & Discrete variables:** Simply here identify and separating the continuous & Discrete variables(Nominal or Ordinal).

PART-(3) EXPLORATORAY DATA ANALYSIS:

- **Descriptive Analysis:** Here investigating the booking dataset and summarizing its main features which is a kind of descriptive analysis of dataset.

- however here plotting the different kind of graphs by using python tools/library like pandas, numpy, matplotlib, seaboarn to estimating the measure of central tendency, measure of variability, measure of spreads for all features.
- **Visualization of insights:** It's about presenting the insights derived from EDA via graphical plotting. where ever analyzing the all features as Univariate, Bivariate and Multivariate analysis.

PART-(4): CONCLUSION:

- It's about summing up the whole process and stating the final results.

CHALLENGES FACED:

- Data cleaning threw a major challenge as data needs to get cleaned for accurate results but in a way such that the data does not lose its meaning.

- Choosing between the type of visualization to be used and also making a choice between matplotlib & seaborn.

CONCLUSION:

- Right from loading the dataset, cleaning and treating the null values, analysis of the data on the basis of different variables to visualization of the trends & patterns of the data, gave us dynamic answers to some real business questions. This particular analysis has unraveled some key findings in terms of the hotel type, customer type, customer preference & time preference.

REFERENCES:

- The Python Standard Library.
- GeeksforGeeks.