Author's Name: Sajal Sinha

Course Title: Hotel Booking Analysis (EDA)

19 August 2021

Keywords: Data Science, Data Analysis, EDA

Hotel Booking Analysis

ABSTRACT

We had Hotel Booking Dataset of City hotel as well as resort hotel which was to be analysed and proper insights should be taken out which can be useful to provider in future for making important decisions.

The Dataset which was provided include columns like cancelled Booking, arrival Day/Date/month/year of customer, market segment type, number of family members, parking space, type of meal, type of room etc. The Dataset is of size (119390,32)

PROBLEM STATEMENT

Hotel Bookings depends on various factors and if those aren't properly managed can lead to fall of hotel. Factors which affect bookings include food type, prices, month of year, country etc.

Our main objective is to perform Data Analysis of Hotel bookings and to give insights to hotel management which will boost their performance.

INTRODUCTION

1. What is EDA?

Exploratory data analysis (EDA) is employed by data scientists to research and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to control data sources to urge the

answers you would like , making it easier for data scientists to get patterns, spot anomalies, test a hypothesis, or check assumptions.

EDA is primarily wont to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a provides a far better understanding of knowledge set variables and therefore the relationships between them. It also can help determine if the statistical techniques you're considering for data analysis are appropriate. Originally developed by American mathematician John Tukey within the 1970s, EDA techniques still be a widely used method within the data discovery process today.

2. **Why is EDA important in Data Science?**

The main purpose of EDA is to assist check out data before making any assumptions. It can help identify obvious errors, also as better understand patterns within the info , detect outliers or anomalous events, find interesting relations among the variables.

Data scientists can use exploratory analysis to make sure the results they produce are valid and applicable to any desired business outcomes and goals. EDA also helps stakeholders by confirming they're asking the proper questions. EDA can help answer questions on standard deviations, categorical variables, and confidence intervals. Once EDA is complete and insights are drawn, its features can then be used for more sophisticated data analysis or modeling, including machine learning.

3**. Types Of EDA**

There are four primary sorts of EDA:

• Univariate non-graphical. this is often simplest sort of data analysis, where the info being analyzed consists of only one variable. Since it's one variable, it doesn't affect causes or relationships. the most purpose of univariate analysis is to explain the info and find patterns that exist within it.

• Univariate graphical. Non-graphical methods don't provide a full picture of the info . Graphical methods are therefore required. Common sorts of univariate graphics include:

• Stem-and-leaf plots, which show all data values and therefore the shape of the distribution.

• Histograms, a bar plot during which each bar represents the frequency (count) or proportion (count/total count) of cases for a variety of values.

• Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

• Multivariate nongraphical: Multivariate data arises from quite one variable. Multivariate non-graphical EDA techniques generally show the connection between two or more variables of the info through cross-tabulation or statistics.

• Multivariate graphical: Multivariate data uses graphics to display relationships between two or more sets of knowledge . the foremost used graphic may be a grouped bar plot or bar graph with each group representing one level of 1 of the variables and every bar within a gaggle representing the amount of the opposite variable.

Other common sorts of multivariate graphics include:

• Scatter plot, which is employed to plot data points on a horizontal and a vertical axis to point out what proportion one variable is suffering from another.

• Multivariate chart, which may be a graphical representation of the relationships between factors and a response.

• Run chart, which may be a line graph of knowledge plotted over time.

• Bubble chart, which may be a data visualisation that displays multiple circles (bubbles) during a two-dimensional plot.

• Heat map, which may be a graphical representation of knowledge where values are depicted by colour.

PROCEDURE FOR DATA ANALYSIS:

1. Importing packages
2. To check and treat Null Values
3. Outlier Detection
4. Data Manipulation
5. Data Visualisation
6. Insights

Types of plots used in analysis:

- Barplot
- Line Diagram
- Count Plot
- Histogram
- PieChart

CONCLUSIONS:

A. Portugal, Great Britain, and France are native place of most customers.

B. August is the MOST busiest month.

C. Most of our customers were brought in by Online Travel Agents.

D. Bed and Breakfast is most preferred meal package.

E. In 2016, we had most customers considering overall customers whereas city hotel had more bookings compared to resort hotels.

F. Most of the time, people come in pair.

G. Room Type A is MOST favourite room type among customers

MEASURES THAT CAN BE TAKEN:

- Very Less amount of customers revisit the hotel, so hotel staff should consider taking valuable feedback of their services.
- As Portugal, Britain, and France have most customers, marketing team can target those countries as well.
- As July to October has most number of visitors, hotel should focus on trained staff during this time.

REFERENCES: