EDA ON

HOTEL BOOKING ANALYSIS

Contributor Role:-

SHIVAM MISTRY (mistryshivam619@gmail.com)

DATA SCIENCE TRAINEES,

TEAM ALMABETTER

Abstract:

The 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.

We will be conducting Exploratory data Analysis on the dataset and understand the dataset by Data Visualization.

Our experiment can help understand what could be the reason for the classification of such labels by feature selection, data analysis.

Problem Statement:

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions! 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 been removed from the data. Explore and analyze the data to discover important factors that govern the bookings

Introduction:

Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business. We will be using the data available to analyses the factors affecting the hotel bookings. These factors can be used for reporting the trends. Our goal is to explore and analyze the dataset to discover the important factors that govern Hotel bookings.

Objective:

- Create an in-depth analysis to figure out the standard patterns of booking.
- Explore the dataset to find the best time of year to book a hotel, optimal length of stay in order get the best daily rate, rate of cancellation and many more of such question.
- Visualise the data using Graph, Bars and Charts for the better understanding.
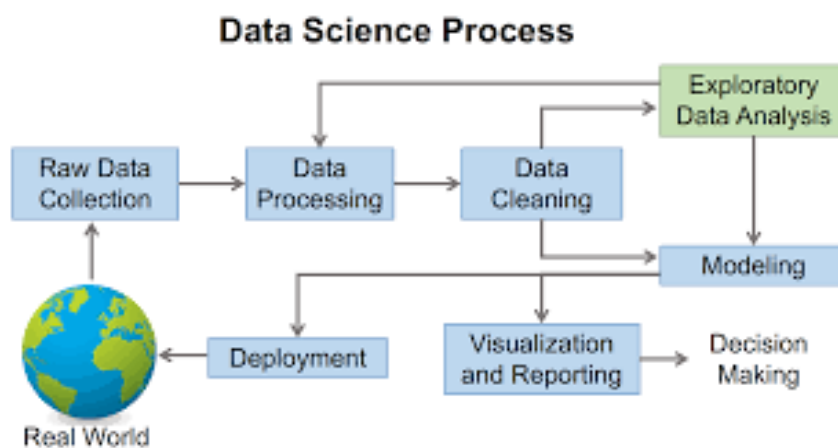
EDA (Exploratory Data Analysis) :

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

Exploratory Data Analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods.

A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling and thereby contrasts traditional hypothesis testing.

Null values Treatment :

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

Data Visualization :

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In this dataset it will be most helpful in distinguishing the difference between hotels, customers and see the data graphically.

Column information:

- **Hotel**
    - H1: Resort hotel
    - H2: City hotel
- **is_cancelled**
    - 1: Cancelled
    - 0: Not cancelled
- **lead_time**
    - No of days that elapsed between entering date of booking into property management system and arrival date
- **arrival_date_year**
    - Year of arrival date (2015-2017)
- **arrival_date_month**
    - Month of arrival date (Jan - Dec)
- **arrival_date_week_numberr**
    - Week number of year for arrival date (1-53)

- **arrival_date_day_of_month**
  - Day of arrival date
- **stays_in_weekend_nights**
  - No of weekend nights (Sat/Sun) the guest stayed or booked to stay at the hotel
- **stays_in_week_nights**
  - No of week nights (Mon - Fri) the guest stayed or booked to stay at the hotel
- **Adults**
- **Children**
- **Babies**
- **meal**
  - Type of meal booked. Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)
  - **country**
- **market_segment** (a group of people who share one or more common characteristics, lumped together for marketing purposes)
  - TA: Travel agents
  - TO: Tour operators
- **distribution_channel** (A distribution channel is a chain of businesses or intermediaries through which a good or service passes until it reaches the final buyer or the end consumer)
  - TA: Travel agents
  - TO: Tour operators
- **is_repeat_guest** (value indicating if the booking name was from repeated guest)
  - 1: Yes
  - 0: No

- **previous_cancellations**

- ▪ Number of previous bookings that were cancelled by the customer prior to the current booking

- **previous_bookings_not_canceled**

  - ▪ Number of previous bookings not cancelled by the customer prior to the current booking

- **reserved_room_type**

  - ▪ Code of room type reserved. Code is presented instead of designation for anonymity reasons.

- **assigned_room_type**

  - ▪ Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons.

- **booking_changes**

  - ▪ Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

- **deposit_type**

- Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.

- **agent**

  - ▪ ID of the travel agency that made the booking

- **company**

  - ▪ ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons

- **day_in_waiting_list**

  - Number of days the booking was in the waiting list before it was confirmed to the customer

- **customer_type**

  - Contract - when the booking has an allotment or other type of contract associated to it;

  - Group – when the booking is associated to a group;

  - Transient – when the booking is not part of a group or contract, and is not associated to other transient booking;

  - Transient-party – when the booking is transient, but is associated to at least other transient booking

  - **adr (average daily rate)**

  - average   daily   rate = SumOfAllLodgingTransactionTotalNumberOfStayingNightSumOfAllLodgingTransactionTotalNumber Of Staying Night

- **required_car_parking_spaces**

  - Number of car parking spaces required by the customer

- **total_of_special_requests**

  - Number of special requests made by the customer (e.g., twin bed or high floor)

- **reservation_status**

  - Cancelled – booking was cancelled by the customer;

  - Check-Out – customer has checked in but already departed;

  - No-Show – customer did not check-in and did inform the hotel of the reason why

- **reservation_status_date**

  - Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when was the booking cancelled or when did the customer checked-out of the hotel.

Summary:

- Majority of the hotels booked are city hotel. Definitely need to spend the most targeting fund on those hotels.

- We also realise that the high rate of cancellations can be high due to no deposit policies.

- We should also target months between May to Aug. Those are peak months due to the summer period.

- Majority of the guests are from Western Europe. We should spend a significant amount of our budget on those area.

- Given that we do not have repeated guests, we should target our advertisement on guests to increase returning guests.

Inferences and Conclusion

Firstly, higher lead time has higher chance of cancellation. Also, history of previous cancellations increases chances of cancellation.

Secondly, the City hotel has more guests during spring and autumn, when the prices are also highest, in July and August there are less visitors, although prices are lower. Thus, customers can get good deal on bookings in July and August in city hotel.

Guest numbers for the Resort hotel goes down from June to September, which is also when the prices are highest. Thus, these months should be avoided for bookings.

Thirdly, May to August is the peak season of bookings. Both hotels have the fewest guests during the winter.

Fourthly, no deposit cancellations are high compared to other categories but these should not be discouraged per se as bookings in this category are also very high compared to non-refundable type bookings.

Fifthly, Hotels need to do marketing and give special incentives for direct bookings as these may establish personal one to one relationship promoting customer loyalty.