

# **EDA ON HOTEL BOOKING ANALYSIS**

**SURAJ KAD,  
TANVIR PATEL,  
SHUBHAM SAWANT,  
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. All personally identifying information has from the data.

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

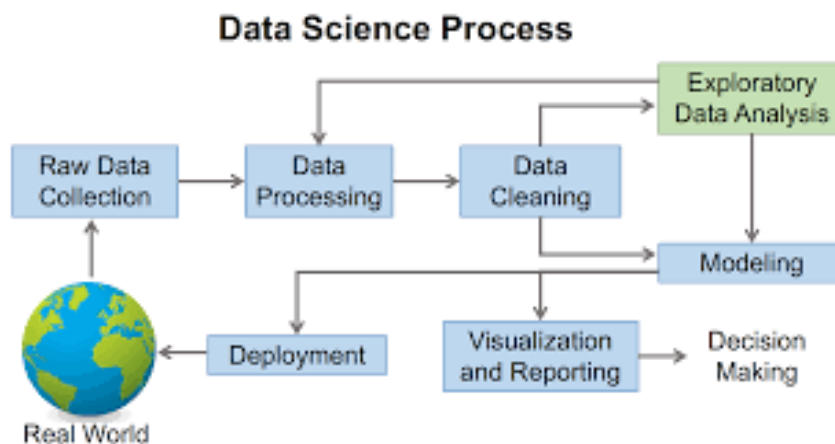
## **Introduction:**

When we hear about hotel booking we definitely hear about Trivago. Thanks to the large scale advertisement, I can't seem to get the ad out of my head. Needless to say, Trivago was the main platform I used for my hotel booking. The main few things I will usually consider include prices per night, distance of hotel from attractions and restaurants, availability of free breakfasts, scenery in hotel room, cleanliness of hotel room and of course, availability of free Wi-Fi. In this dataset, we are able to know different types of bookings (i.e. type of hotel, duration of stay, types of visitors, types of booking, etc.).

## **Objective :**

Create a in-depth analysis to figure out the standard patterns of booking

Generate a report for Michael Scott, the new marketing manager of Trivago so he can derive a strategy for the marketing team. As a newly hired data analyst, you have been assigned this task as majority of your colleagues (except for Kevin Malone) were quarantined after you spread the coronavirus to them. You need to win them back! Kevin Malone has completed his analysis from year 2018 to 2019 and is currently in a meeting with Michael in his office. You will work with data from year 2015 to 2017.



## **EDA (Exploratory Data Analysis) :**

After loading the dataset we performed this method by comparing our target variable that is Surge Pricing Type with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

## **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.

## **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\_number**
  - 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  

$$= \frac{\text{SumOfAllLodgingTransactionTotalNumberOfStayingNight}}{\text{SumOfAllLodgingTransactionTotalNumberOfStayingNight}}$$
- **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 go down slightly from June to September, which is also when the prices are highest. Thus, these months should be avoided for bookings.

Thirdly, Broadly, April 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, cancellations are high when done through agents compared to direct bookings. Hotels need to do marketing and give special incentives for direct bookings as these may establish personal one to one relationship promoting customer loyalty.

## **References-**

1. GeeksforGeeks
2. Analytics Vidhya
3. [www.kaggle.com](https://www.kaggle.com)