**Capstone Project**

**EDA OF Hotel Booking**

## **Project Overview**

I perform exploratory data analysis with **Hotel Booking Dataset**.

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.

First I explore the data, cleaned and preprocessed the data and then I performed the exploratory data analysis to extract information by data wrangling over the raw data and plotted the graph to visualize the data by using some important libraries such as Matplotlib, Seaborn, Pandas, and Numpy.

Let’s understand the terms.

* **Hotel**: Name of hotel ( City or Resort)
* i**s\_canceled**: Whether the booking is canceled or not (0: Non-cancelled, 1: Canceled)
* **lead\_time**: Time (in days) between booking transaction and actual arrival.
* **arrival\_date\_year**: Year of arrival
* **arrival\_date\_month:** Month of arrival
* **arrival\_date\_week\_number:** Week number of arrival date.
* **arrival\_date\_day\_of\_month:** Day of month of arrival date
* **stays\_in\_weekend\_nights:** No. of weekend nights spent in a hotel (Monday to Friday)
* **stays\_in\_week\_nights:** No. of weeknights spent in a hotel (Saturday and Sunday)
* **adults:** No. of adults in single booking record.
* **children:** No. of children in single booking record.
* **babies:** No. of babies in single booking record.
* **meal:** Type of meal chosen
* **country:** Country of origin of customers (as mentioned by them)
* **market\_segment:** Which segment via booking was made and for what purpose.
* **distribution\_channel:** Via which medium booking was made.
* **is\_repeated\_guest:** Whether the customer has made any booking before

(0: No, 1: Yes)

* **previous\_cancellations:** No. of previous canceled bookings.
* **previous\_bookings\_not\_canceled:** No. of previous non-canceled bookings.
* **reserved\_room\_type:** Room type reserved by a customer.
* **assigned\_room\_type:** Room type assigned to the customer.
* **booking\_changes:** No. of booking changes done by customers
* **deposit\_type:** Type of deposit at the time of booking

(No deposit/ Refundable/ No refund)

* **agent:** agent Id for booking
* **company:** Id of the company who is booking
* **days\_in\_waiting\_list**: No. of days on waiting list.
* **customer\_type:** Type of customer (Transient, Group, etc.)
* **adr:** Average Daily rate.
* **required\_car\_parking\_spaces:** No. of car parking asked in booking
* **total\_of\_special\_requests:** Total no. of special request.
* **reservation\_status:** Whether a customer has checked out or canceled,or not showed
* **reservation\_status\_date:** Date of making reservation status.

We analyze the data by considering the below **problem statement.**

## When is the best time of year to book a hotel room?

## What is 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?

And correlation of the data.

**Conclusion**

Considering all of these we **concluded that**

City hotels having maximum booking, which are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.

Considering the years, Bookings for 2016 were highest with 48%. Bookings were increasing on a yearly basis. For 2016 increment was too high so there was a steep fall in bookings for 2017.

The month of August saw the highest number of hotel bookings for both types of hotels. The lowest hotel bookings were in January. Daily rates are also high in the month of August where low in January so January is the best time to book a hotel with cheaper rate & confirmed booking.

Guests preferred no deposit type bookings because of the flexibility to cancel or change bookings without losing any money. About 87.63% bookings are without any deposit. 12.24% bookings deposit are nonrefundable.

Out of total booking 37% booking were canceled. With high no. of booking, cancellation is also higher in city hotels and out of that canceled booking, hotels with no deposit have higher cancellation rate.

Repeated guests are not more likely to cancel their bookings.

Considering the rates City hotel has the highest ADR. That means city hotels are generating more revenues than the resort hotels.

Most people prefer a 2 night stay. Optimal length of stay is 7 days. As an increase in length of stay ADR decreases. To get the best adr you need to stay more than a month.

For long stay guests prefer resort hotels.

More involvement in bookings with Online TA and Offline TA/TO that means maximum hotels are booked with agents only.

Special requests are more for an adult. As an increase in special requests and increasing in people ADR will be more.

Most of the guests are coming from Portugal (PRT), Great Britain (GRB), France (FRA), Spain (ESP) and Germany (DEU).

Finally checkout the reservation status, 36% of bookings got canceled.1% of bookings never arrived at hotel without canceling bookings.

**Github Link:** <https://github.com/YashwantRaul/EDA-of-Hotel-booking>

**Drive Link:**

https://drive.google.com/drive/folders/1pr2LXyWN0t\_gmaE\_vue4TSaFqM5QDHHo?usp=sharing