## Abstract:

Hotel Booking Analysis

Shabana Ali,

Data science trainees,

AlmaBetter, Bangalore

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, or babies,

and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

## Introduction

The Hotel booking data set includes

Numerical, Categorical and Binary data. The data set has columns like the hotel type, is\_canceled, arrival\_date\_year,

arrival\_date\_month,

stays\_in\_weekend\_nights,

stays\_in\_week\_nights,country, market\_segment, distribution\_channel, etc. which helped us draw major insights from the data set. Our aim here is to understand the important factors that governs the hotel bookings.

**Problem Statement**

Hotel Booking is ruled by way of many elements consisting of the time of the year, wide variety of guests, distribution channel, hotel type, etc.

The predominant thing of this undertaking is to operate Exploratory information evaluation and draw insights to recognize all the vital elements that govern the Hotel bookings.

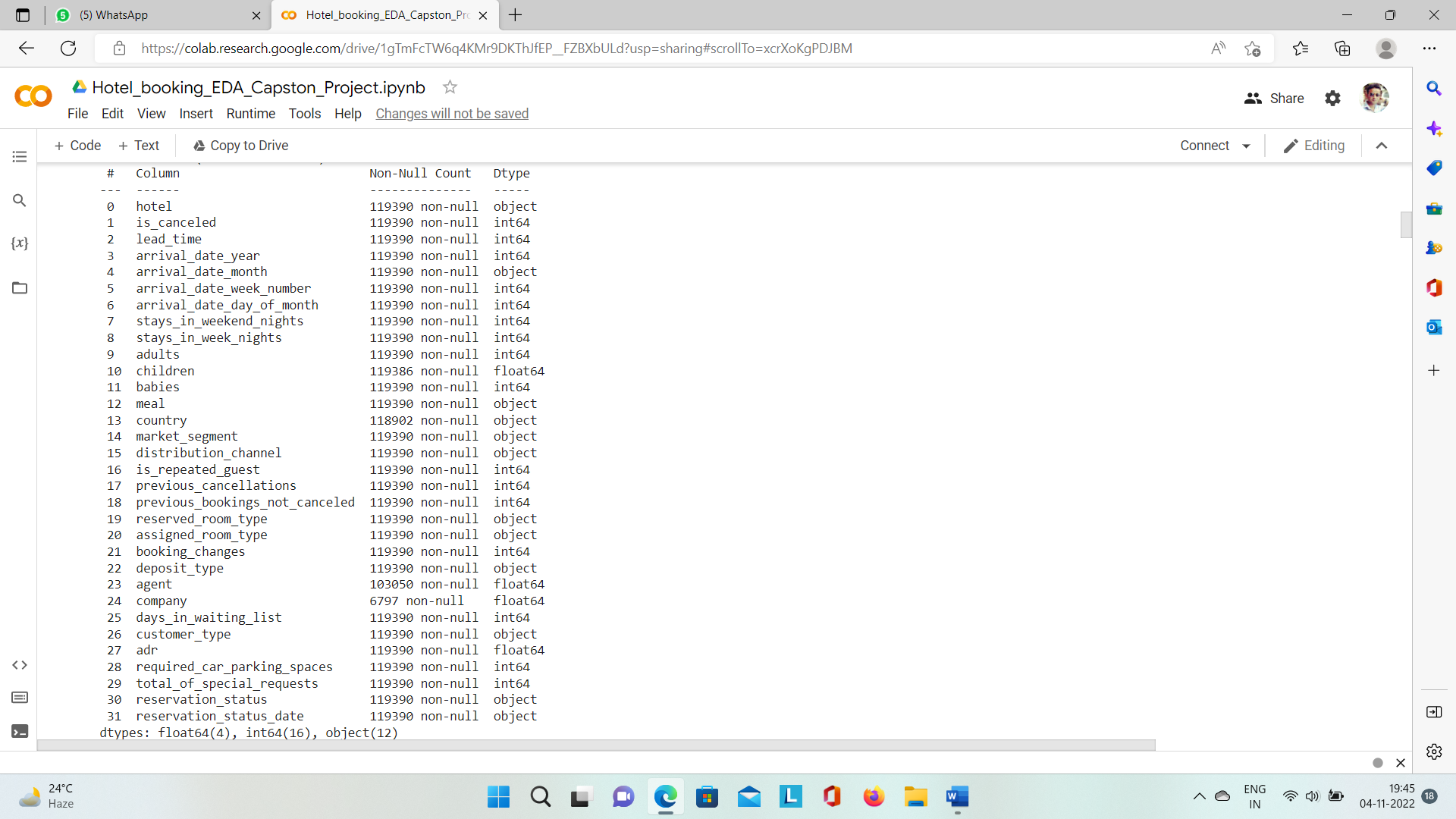
**Understanding the data**

The given data set has 3 types of data:

* + Numerical data
  + Categorical data
  + Binary Data

## Data Summary

The provided data set has following different columns of variables necessary for hotel bookings:



* hotel: The category of hotels, which are two city hotel and resort hotel.
* is\_cancelled : The value of column

shows if the booking was canceled or not.

Values[0,1], where 0 indicates not

canceled and 1 indicates the cancellation.

* lead\_time : The time between reservation and actual arrival.
* stayed\_in\_weekend\_nights: The number of weekend nights stay per reservation.
* stayed\_in\_weekday\_nights: The number of weekday night stays per reservation.
* meal: Meal preferences per reservation, which are
* Country: The origin country of guest.
* market\_segment: This column shows how reservation was made and what is the purpose of reservation.
* distribution\_channel: The medium through booking was made.[Direct,Corporate,TA/TO,undefined,GDS.]
* Is\_repeated\_guest: Shows if the guest is who has arrived earlier or not.Values[0,1]-->0 indicates no and 1 indicated yes person is repeated guest.
* days\_in\_waiting\_list: Number of days between actual booking and transact.
* customer\_type: Type of customers (Transient, group, etc.)
* Undefined/SC – no meal package and dinner)

**distribution\_channel:** The medium of booking was [Corporate, Direct,

GDS, TA/TO, undefined]

**Data Summary**

* + - 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)
    - **TA** –Travel agency
    - **TO** –Tour operator
    - **GDS** – Global Distribution System.

## Types of Hotels

* + Resort Hotel
  + City Hotels

## How hotel booking works?

* + Hotel reservations are made by guests before they arrive at the hotel. Bookings are received from the following distribution channels:
    - Direct
    - Corporate
    - TA/TO
    - GDS
* Depending on availability and demand, the hotel assigns a room. If the requested room is unavailable a different room is assigned.
* The Hotel receives no deposit, or refundable deposit or non-refundable deposit against the bookings.

# Steps involved:

**Importing important libraries -** During this step, our main goal was to import all the necessary libraries to help us explore the problem statement and perform EDA to draw conclusions from the data.

**Understanding the data set** Next, we worked on checking the data set. We need to know how many rows and columns are available and what columns could be important in solving the problem statement.

**Null values Treatment**

Our data set contains a large number of null values which might tend to disturb our insights. As a result, we replaced them with '0' for numerical data and 'undefined' for categorical data.

**Exploratory Data Analysis**

After treating the null values, started with the EDA. I performed EDA.

**Exploratory Data Analysis:**

The following analysis was used to solve the problem statement during EDA

**Uni-variate Analysis:**

We answered the following questions when performing univariate analysis on a hotel booking data set:

* Which distribution channel gave most of the bookings?
* Most customers are from Portugal.
* Type A room is in most demand.
  + Which room type is in most demand?
* From which country most of the customers are coming?
* What is the most popular meal among customers?

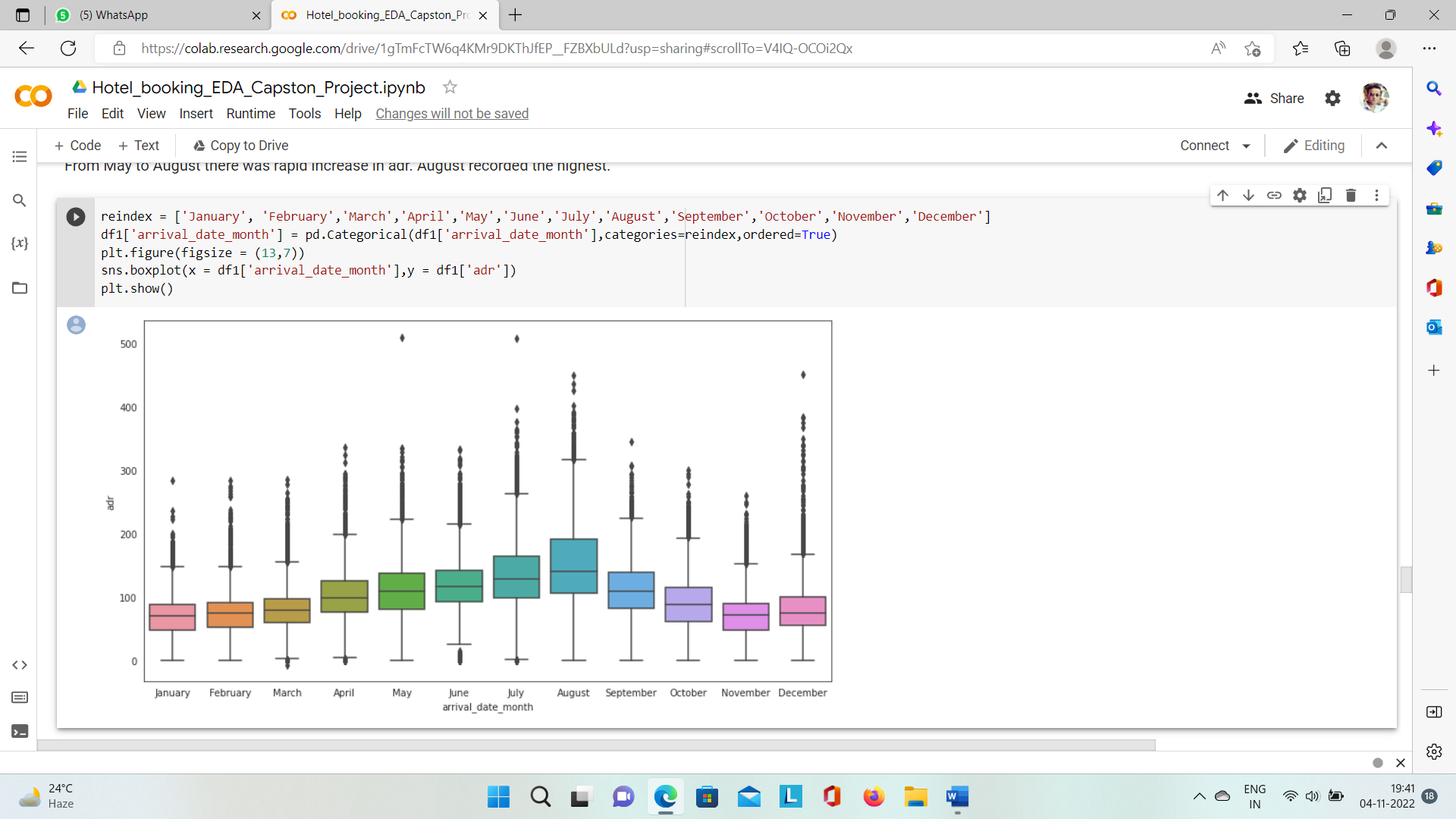
**Hotel wise Analysis:**

While doing hotel-wise analysis of the given hotel booking data set, we answered following questions:

* + Which hotel type was more engaging and in most demand?
  + Which hotel type receives more guests?
  + Which hotel type makes more revenue?
  + Which type of guests have the most check-ins?
  + What is the most preferred length of stay in each hotel?
  + Which hotel has higher and how much customer returning rate?

### Time wise Analysis:

* + We answered the following question based on time-wise analysis of given hotel booking data:
  + What are the busiest months for hotels?



### Some other questions:

* + Which hotel type received the most special requests?
  + What is the effect of deposit type on profit?
  + How many special requests were received?

# Observations:

As a result of the univariate analysis, hotel-wise analysis, and time-wise analysis, the following conclusions were reached:

# Conclusion:

That's all there is to it have reached the end of our exercise. The data has been loaded, null values have been treated categorical columns encoded, and major reasons that govern hotel bookings have been identified, along with steps to increase them.

* The majority of bookings were made through online travel agents.
* Most customers prefer the

Bed and breakfast meal type.

* City hotels receive a highest number of guests than the resort hotels. However, the cancellation rate is more for city hotels.
* City hotels have a higher number of repeat customers. However, the ratio of repeat customers is more for resort hotels as the city hotels receive a higher number of guests than the resort hotels.
* Deposits that are non-refundable end to result in successful transactions.
* May is the busier and most profitable month for the hotels in 2017.
* On an average, May-June is the busier and most profitable month for the hotels.
* City hotels have received the most special requests.
* 1 number of requests is the highest.
* 1 car parking was needed for most of the guests.

**References-**

* 1. Geeks for geeks
  2. X-mind