

# **HOTEL BOOKING ANALYSIS**

BY - ANKITA SINGH,  
PRASHANT GOUR  
AKHAND PRATAP SINGH  
KHUSHBOO KANT  
MANISH KUMAR SHARMA

## **EDA CAPSTON PROJECT**

ALMABETTER, BANGALORE

### **ABSTRACT:**

- 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. We will perform exploratory data analysis with python to get insight from the data
- This project on medium explains the entire process. In this data analysis, we are using different data visualization techniques which help us to determine trends followed in bookings.

### **PROBLEM STATEMENT:**

Explore and analyze the data to discover key understandings. Data analysis on millions of listings provided, These millions of listings generate a lot of data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around **119390** observations in it with **32** columns and it is a mix of categorical and numeric values.

## INTRODUCTION:

We live in an era where data is produced and circulated in an enormous amount. Those data can be collected and allow us to infer meaningful results and make well-informed decisions. However, as the number of data increases, we need to visualize the data to help us in conducting data analysis. By using visualization tools, we can deliver a message to our customers and inform them about our findings. We will explore and visualize the dataset using basic exploratory data analysis techniques. We will find out the distribution of every listing based on their distribution channels, including their price range, room type, listing name, and other related factors. This San-Francisco-based startup offers you someone's home as a place to stay instead of a hotel. Our goal here is to explore the data and find useful insights from the data and find out different relations between the columns.

Basic descriptions about datasheet containing following features.

- Hotel
- is\_canceled
- lead\_time
- arrival\_date\_year
- arrival\_date\_month
- arrival\_date\_week\_number
- arrival\_date\_day\_of\_month
- stays\_in\_weekend\_nights
- stays\_in\_week\_nights
- adults
- children
- babies
- meal
- country
- market\_segment
- distribution\_channel

- is\_repeated\_guest
- previous\_cancellations
- previous\_bookings\_not\_canceled
- reserved\_room\_type
- assigned\_room\_type
- booking\_changes
- deposit\_type
- agent
- company
- days\_in\_waiting\_list
- customer\_type
- adr
- required\_car\_parking\_spaces
- total\_of\_special\_requests
- reservation\_status
- reservation\_status\_date

## **STEP INVOLVED:**

### ☐ **1. Acquire and loading data**

For this project, we are using Google colab a web IDE with a python programming language to write our script. IDE or Integrated Development Environment is a software application used for software development. To get the data, we are using data that is publicly shared on the internet under the Creative Commons License. Before we are able to load the data into our IDE, first we need to import various external libraries/modules that are needed for visualization and analysis.

### ☐ **2.Cleaning datasheet**

The next step is cleaning up the data, oftentimes the data we load have various faults, such as missing value, incomplete data, etc. By cleaning up, the data quality will have better quality to be used for further analysis.

- Checking column with missing values
- Removing redundant variables
- Replacing all the missing values

#### ☐ **Exploring and visualizing data:**

- In statistics, exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods.
- Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images.
- A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

## **DATA ANALYSIS:**

After loading the dataset, we performed this method by comparing all the columns. This process helped us to figure out patterns, spot anomalies, detect outliers, and find fascinating relations among the variables. By conducting EDA, we can turn an almost usable or unusable dataset into a usable dataset. The major components of exploratory data analysis are Cleaning datasets, exploring and visualizing data.

### **Cleaning dataset**

The next step is cleaning up the data. Often the data we load has various faults, such as typos, missing values, incomplete data. By cleaning up, the data quality will be better to use for further analysis.

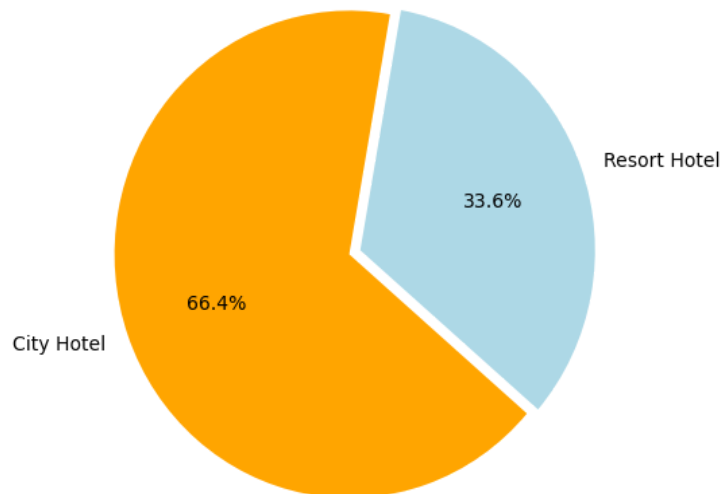
#### **Null & unfitting values Treatment**

Our dataset contains many null values and a tiny number of rows, so we tried to preserve as many rows as possible by replacing null values with suitable values.

### **Non-Graphical and Graphical Univariate Analysis**

We analyze just one variable to describe the data and find patterns that exist within it. For which we used different graphs like a bar, histogram, box, etc. By which we came to know about some outliers.

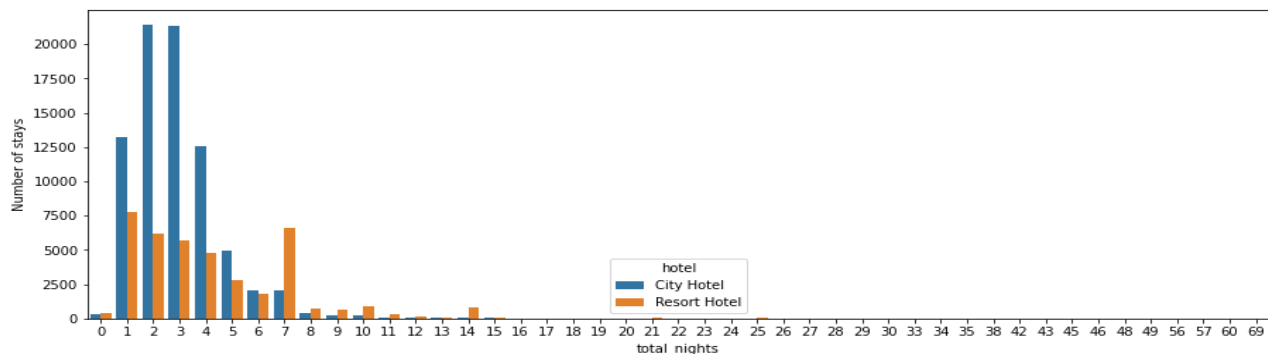
## Percentage of booking in each hotels



It would be easy to say that the majority of the customers belong to the city hotels(66.4%) while resort hotels having very less percentage(33.6%).

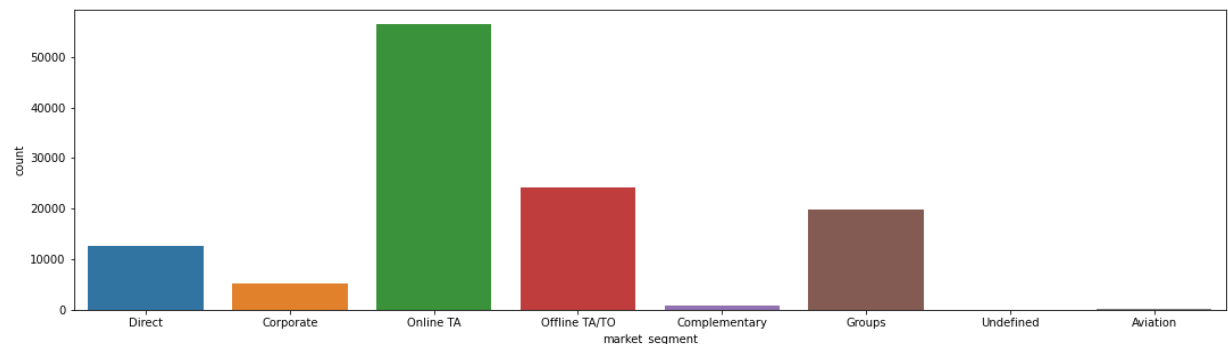
## What is most preferred stay weekend night length in each hotel?

City hotel is the most preferred hotel for weekend night rather than Resort hotel and customer stays number of time in City hotel.



DISTRIBUTION CHANNEL WISE ANALYSIS:

Types of distribution channels used by people to book hotels

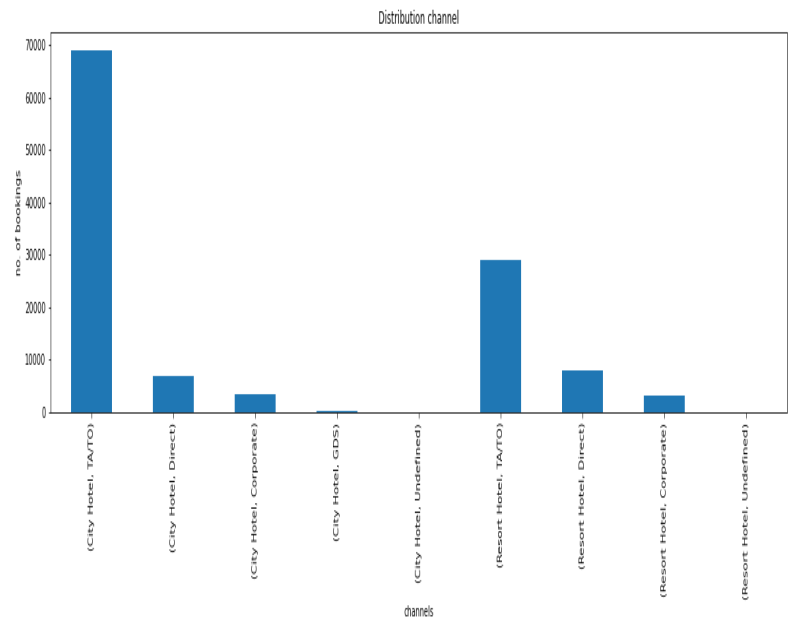


Online TA(travel agent) is the most common channel for booking hotels compare than other hotels. Here we can see that most number of guest are making reservation through TA/TO channels which is travel agent tour operator.

Distribution channel brings better revenue genrating deal for hotels

Resort hotel has more revenue generating deals by direct and TA/TO channel. Resort hotel need to increase outreach on GDS to increase revenue.

- GDS channel brings higher revnue generating deals for city hotel.City hotel can work to increase outreach on GDS channel to get more higher rev



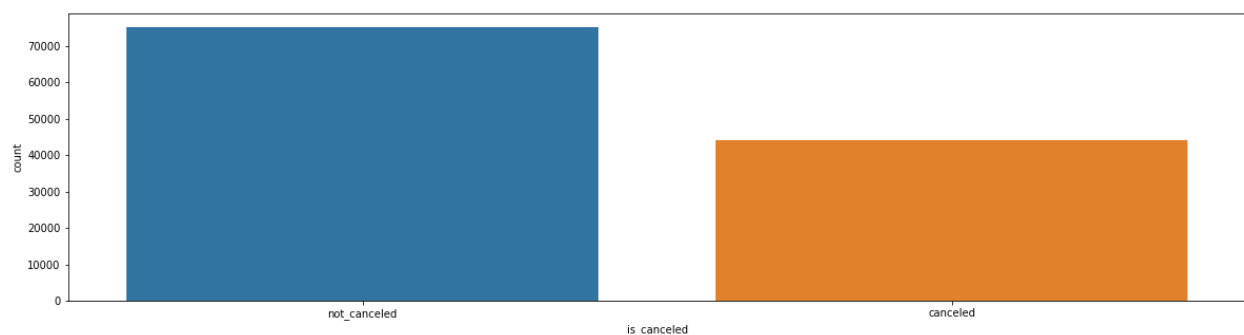
Resort hotel has more

revenue generating deals by direct and TA/TO channel. Resort hotel need to increase outreach on GDS to increase revenue.

- GDS channel brings higher revenue generating deals for city hotel. City hotel can work to increase outreach on GDS channel to get more higher revenue generating deal.

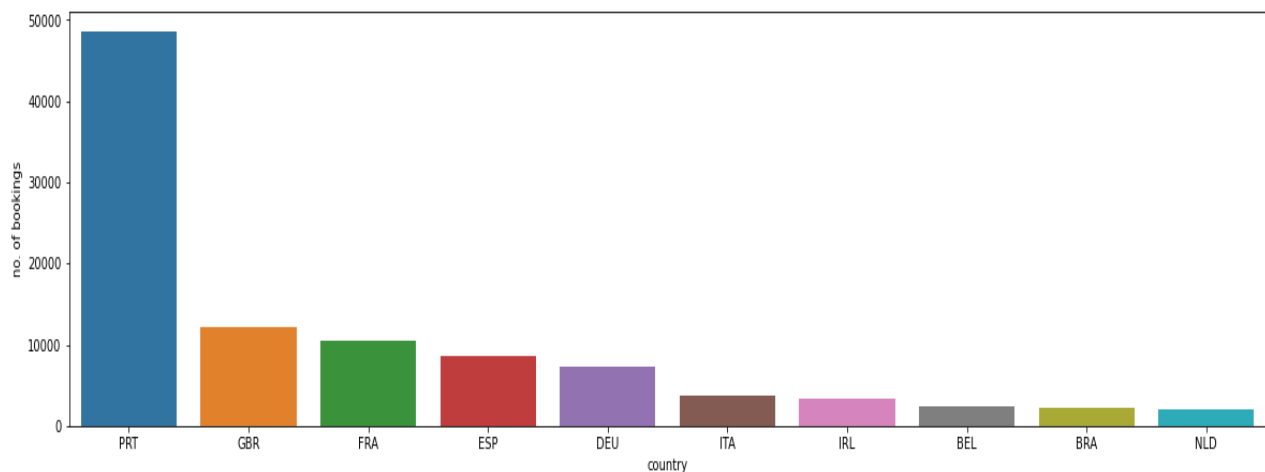
### ☐ The hotel having high cancellation rate

**Blue** stands for City hotel. **Orange** stands for Resort hotel. so far, we can conclude here that City hotel has higher booking cancellations rate than Resort hotels.



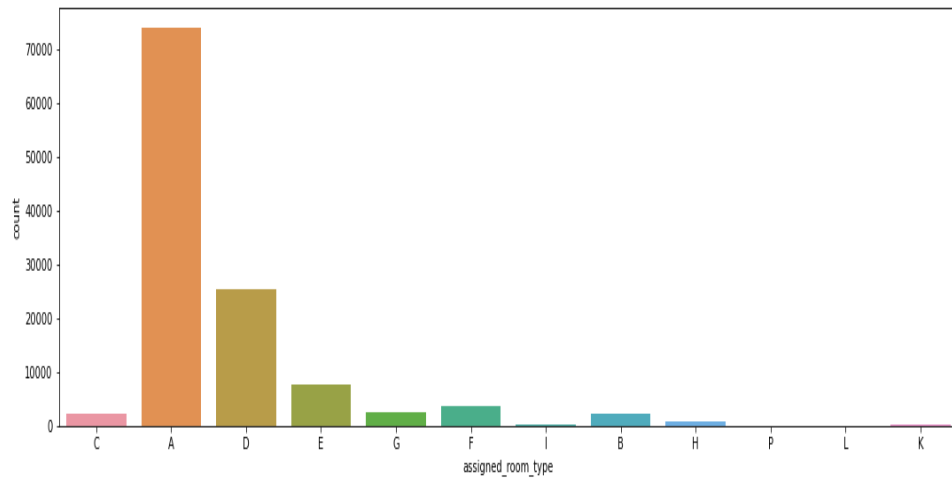
## UNIVARIATE ANALYSIS:

**From which country most of the guest are coming?**



We can see in above graph that PRT,GBR,FRA are the countries whose number of bookings are respectively more than other countries.

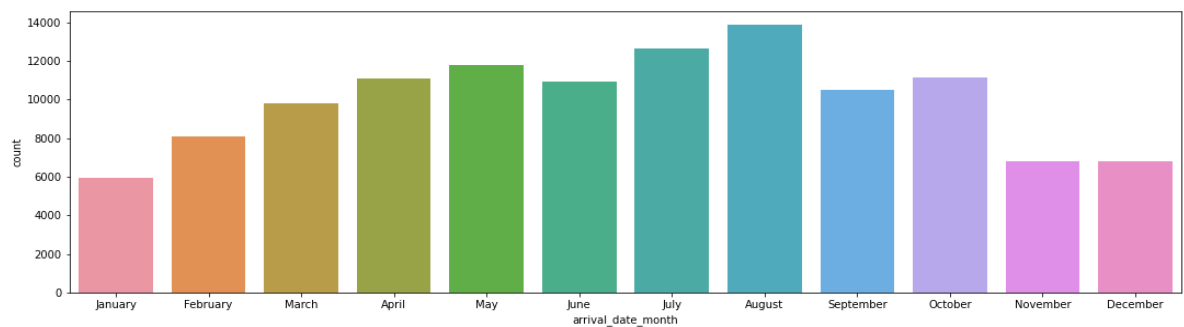
**Which room type is in most demand and which room type generates highest Average daily revenue?**



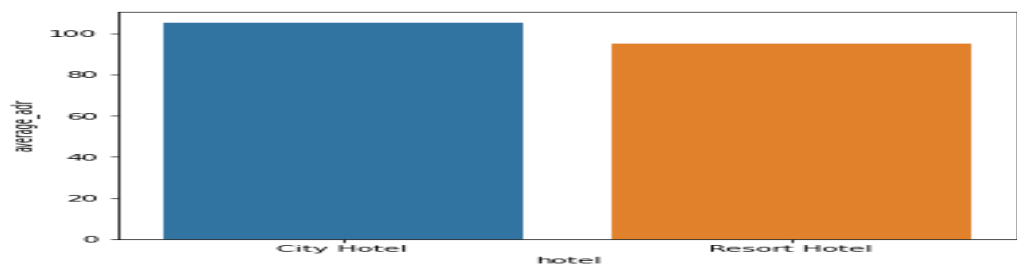
Room type A is in most demand

## TIME WISE ANALYSIS:

☐ The busy months of hotels



☐ The hotel charges high daily rate



City hotel charges more average daily rate because it is more in demand. And on the other hand Resort hotel charges less average daily rate as it is in less demand



## CONCLUSION:

- Around 66.4% bookings are for City hotel and 33.6% are for Resort hotel, therefore City hotel is busier than Resort hotel. Also the overall Average Daily Rate of City hotel is slightly higher than Resort hotel.
- Mostly guests stay for less than 23 days in City hotel and for longer stays Resort hotel is preferred.
- Guests use different channels for making booking out of which most preferred way is TA/TO.
- City hotel charges more average daily rate because it is more in demand. And on the other hand Resort hotel charges less average daily rate as it is in less demand.