Hotel Booking Analysis

**Rahul Raj**

**rahulraj.crp@gmail.com**

**Alma Better Capstone Project**

**Git-Hub link:** “https://github.com/RAHULRAJ666DX/Capston-Hotel-Booking-Analysis-”

# Abstract:

Data has become so important it’s sometimes referred to as digital gold. Unlike gold, however, data isn’t rare or scarce. In fact, it’s so abundant we sometimes don’t know what to do with it. To be useful, data needs to be organized, filtered, and cleansed. Thorough data analysis can provide hoteliers with vital business intelligence to help make better decisions for their property.

Personal customer data in [guest profiles](https://www.cloudbeds.com/templates/guest-persona-templates/) and payment transactions also need to be stored and used securely, not only because guests expect it but also because many jurisdictions are governed by strict data privacy and security laws. A data breach can cause huge headaches for hotels and erode trust with travelers.

So my project helps by analyzing all characteristical data and help the business owner how to cope up with business requirements and how to better the service to increase the value of the business and efficiency

# Problem Statement:

We have been put into the shoes of a marketing team for a new hotel startup that wants to know about the hotel industry. We have been provided a dataset of the hotel industry of the region in which our startup aims to begin its initial operations. Our primary goal is to understand the market of this particular region and try to use our data analysis techniques to draw out key features of the market. Our secondary goal is to draw out actionable insights from our analysis and give conclusions about key aspects of the market such as cancellation rate, distribution

channels, and alike.

# Data Summary:

Based upon the initial assessment we found that the data was pretty much clean except for some missing values in a few columns. Upon using the info() method, we draw out the following key insights about the data:-

1. The dataset has a shape of (119390, 32) which means that it contains approximately 1.2 lakh rows and 32 columns.
2. Our Dataset has 4 columns with float64 dtype, 16 columns with int64 dtype, and 12 columns with object dtype.
3. In our Dataset, we observed null values in the following columns:
   * 4 null values in the children column
   * 488 null values in the country column
   * 16,340 null values in the agent column
   * 112,593 null values in the company column

We have the following column names provided to us in the dataset,

* 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

# Steps involved in the Data Analysis: -

1. **Understanding the need for EDA to solve the problem:**

* This is the most important step of the analysis because we cannot solve a problem if we don’t know what is the need for the solution and in this step we study what is the need of doing a EDA analysis for Hotel Booking analysis
* To get ideas many creative thinking techniques such as brainstorming is done.

**2. Studying the Data and brainstorming how to work on it:**

* In this step the dataset provided by Almabetter is studied throughly and we make some assumptions based several selection criterias and what to be done on the data
* various required questions are written down from a customer and business perspective and look for answers to those questions

1.Which Hotel Type is more used

2.Which Country the Guests are coming from to the Hotels.

3.How many cancellations in both types of hotel.

4.To create a plot to see how many cancellations with respect to City Hotel and Resort Hotel

5.Which Type of Rooms is getting Booked

6.How the Room is Booked

7.Which distribution channel contributes most to ADR

8.Deposit type for the Booking

9.Meal statistics from the hotels

10.How many Guests are repeated guests

11.Customer Types

12.No of travellers in Various Months

13.Stay During Weekends

14.Stay During Week days

15.Quantity of people who stayed in Which Rooms

16.Optimal Stay for Guests

17.The Correlation Heatmap of the Dataset

**3. Cleaning and preparing the Data:**

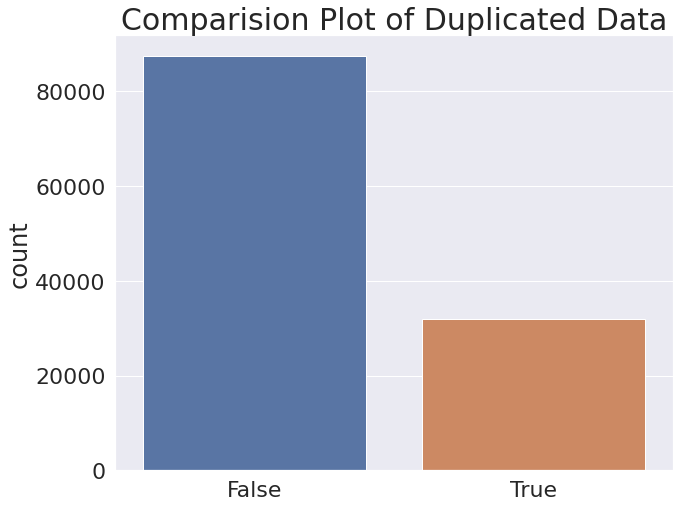
* For working on the data the data is to be cleaned and unwanted null values has to be removed or replaced with meaningful data
* There of 4 columns present in the dataset with null values we have removed two columns with the most null values and for the other columns we replaced the Nan values with mode and new data element

So we have 31994 duplicate rows in our data. So we will drop the duplicate rows from our dataset

dropping the duplicate rows df= df.drop\_duplicates()

df.shape #dataset reduced (87396, 32)

Dropping duplicates since we cannot derive meaningfull insights.

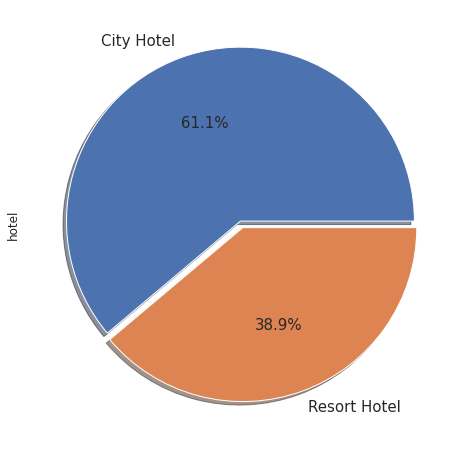


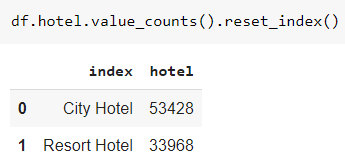
**EDA analysis:**

By EDA we mean exploratory data analysis. In this, we looked at the dataframe and decided our target variables (Important Columns) based upon which we were going to conduct further analysis. We started

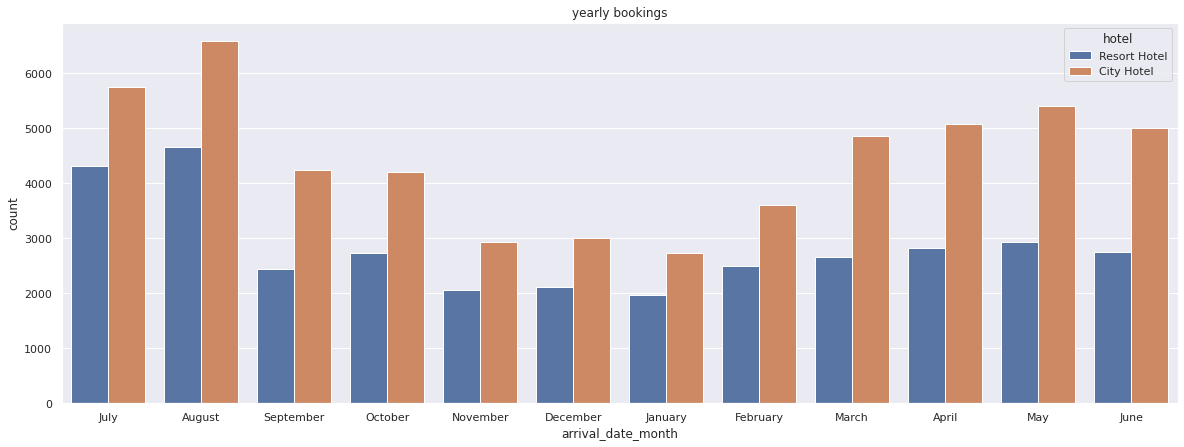
comparing our target variable with other independent variables (leftover columns) to get insights into the relationship between them. This also gave us a better picture of how the different variables affect the target variable.

**1.Which Hotel Type is more used**

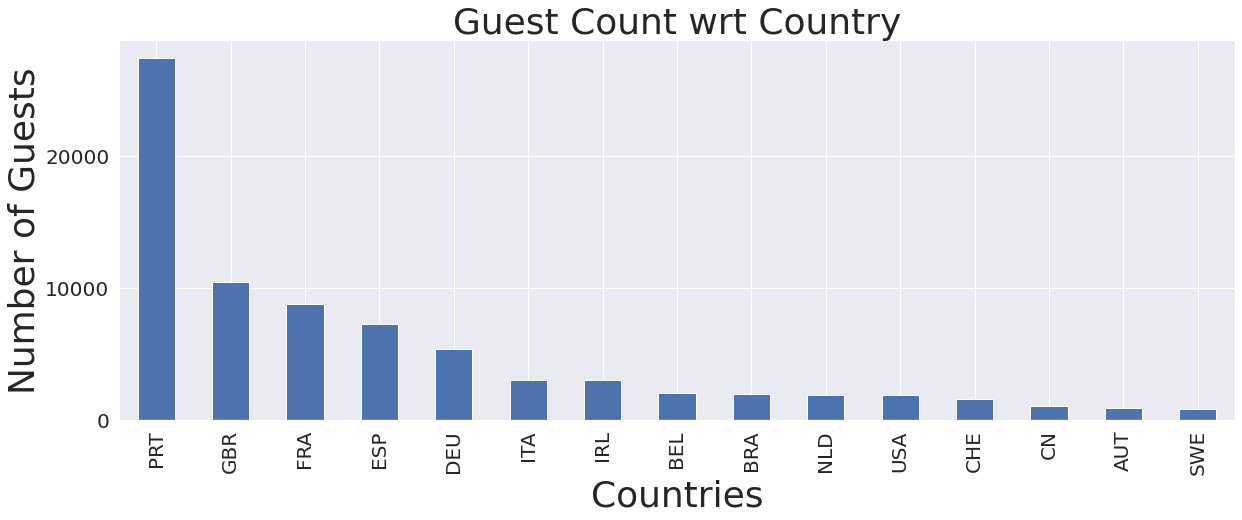




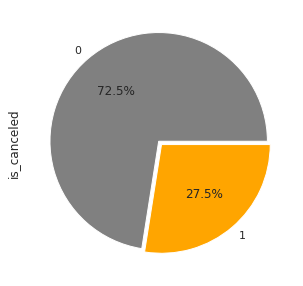
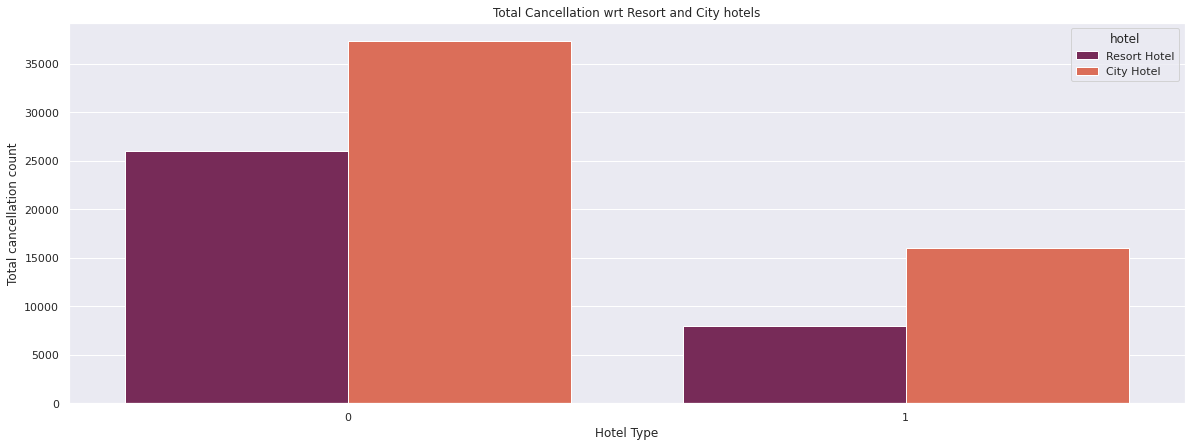
**2.Arrival Date Month Comparing how many Bookings were made for Hotels to Compare**



**3.Which Country the Guests are coming from**



**4.How many cancellations in both types of hotel**



**5.Which Type of Rooms is getting Booked**

There are a total of 10 Room Types

'C', 'A', 'D', 'E', 'G', 'F', 'H', 'L', 'P', 'B'

1.Most of the Guests Book room type 'A' which is 64.70% with 56552 bookings

2.Second with type 'D' with 17398 bookings and type E with 6049 bookings

3.Room type 'L' and 'P' is booked very few times with only 6 booking

**6.Which market segment the guest was from**

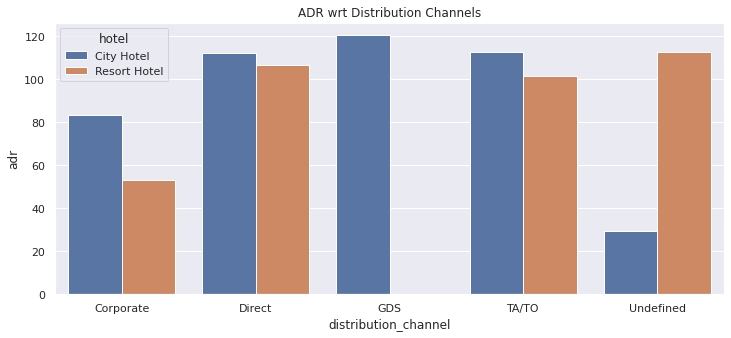
**isualization of Insights:**

There are a total of 8 different Marketting segments

['Direct', 'Corporate', 'Online TA', 'Offline TA/TO', 'Complementary', 'Groups', 'Undefined', 'Aviation']

1.Most of the Guests are from 'Online TA' which is 59.06% with 51618 bookings

2.Followed by 'Offline TA' and 'direct' are almost same at 15.89 and 13.5%

**7.Which distribution channel contributes most to ADR**

1.Most of the ADR is contributed by GDS and TA and Direct

2.Most of the Guests are Booking from 'GDS' and in most cases the City hotels gets more bookings

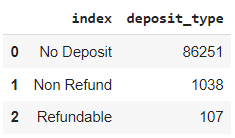
3.Undefined- Bookings are undefined so most probable bookings are on arrival.

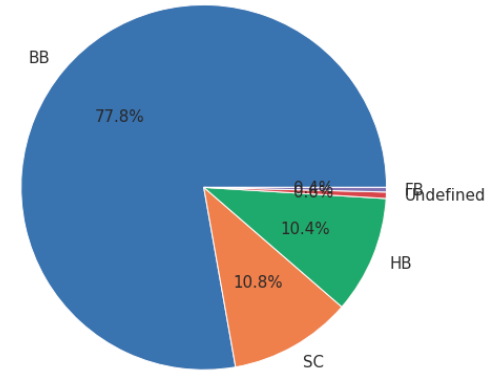
**8.Deposit type for the Booking**

There are 3 different Deposit types

'No Deposit', 'Refundable', 'Non Refund'

1.98.68% of the people don't make a Deposit with 86251 records

2.1.1%% of the people make non refundable deposits with 1038

**9.Meal statistics from the hotels**

Out of the meals

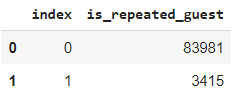
BB (Bed & Breakfast) is the most ordered meal which is around 77.31%,

HB(Half Board) is 10.84%

SC(no meal package) is 10.39%

Undefined and FB (Full Board) with 0.5%

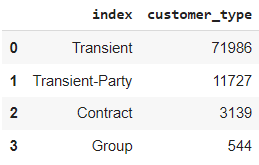
**10.How many Guests are repeated guests**



Only 3.19% of the guests are repeated guest

City Hotels has 2032 repeated guests

Resort Hotels has 1778 repeated guests

**11.How many Guests are repeated guests**

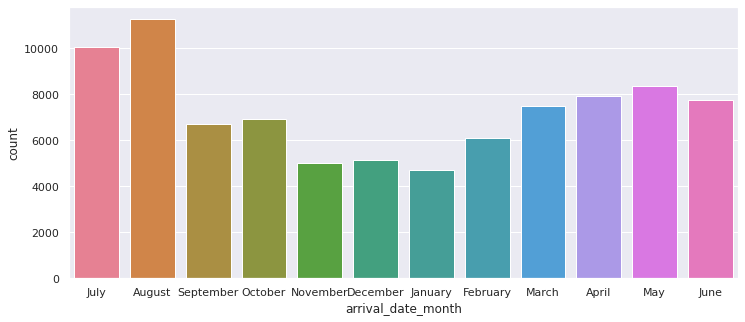
Observation

75.05% of the guests are Transient guests

Followed with 21.04% of Transient-Party and

3.41% of guests are from Contract

**12.No of travellers in Various Months**

August month had most travellers with 12.88% which is 13877 guests July with 11.5% which is 10057 guests,May and October with 8355 and 7908 guests

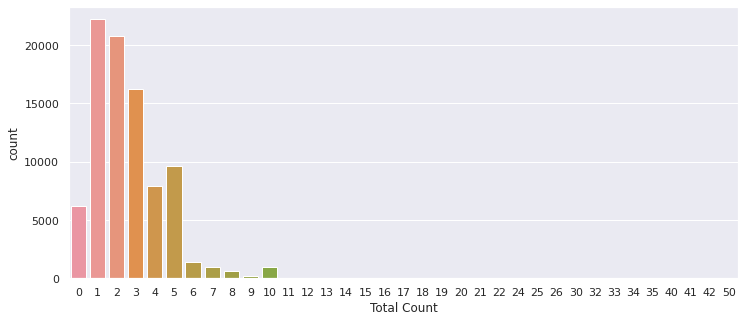
**13.Stay During Weekends and Weekdays**

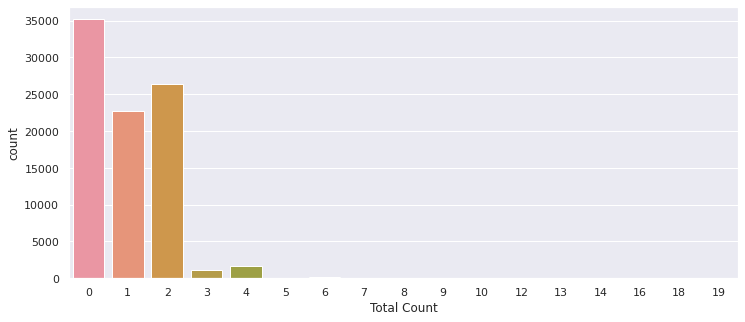
Weekends

Most of the time people stayed 2 to 1 days during Weekend

Weekdays

1.52.50% of the travellers were during Weekend

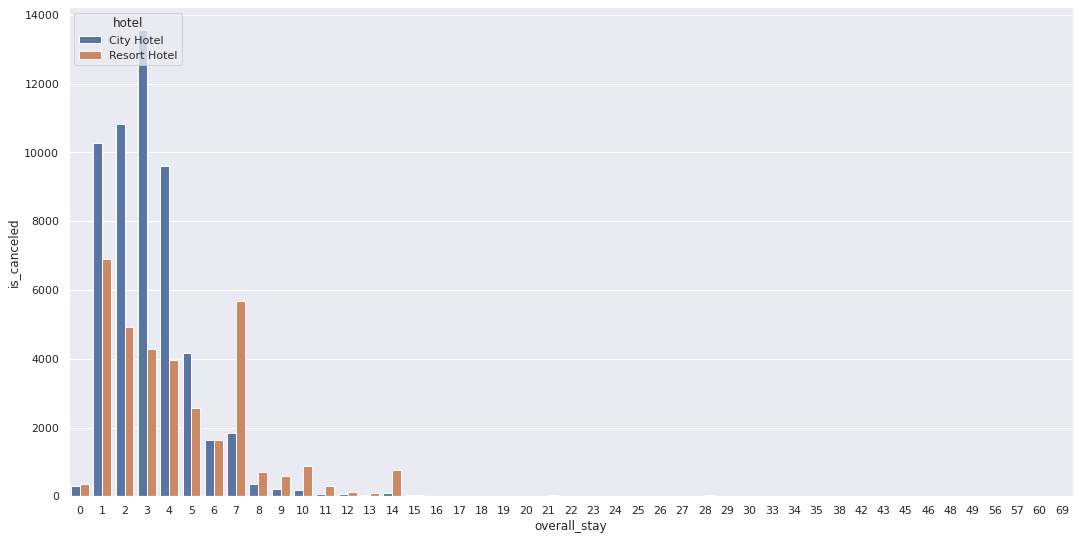
2.Most of the time people stayed 1 day and followed by 2 and 3 days during Week days

**14.Number of people per Room**

1. 73.79% of the time 2 guests stayed in the room

2. There were 16503 Single rooms bookings made which is 18.88% of total bookings

**15.Optimal Stay for Guests**

So most of the people stay 1 to 4 days and very rarely they stay above a week

**16.Optimal Stay for Guests**

Previous bookings not canceled" has 44% Correlation with "is repeated guest" which shows that those who has not cancelled their previous booking has more probability to be a Repeated Guest

adr is slightly correlated with total\_people, which makes sense as more guests means more service to deliver, therefore more adr.

**Conclusion**:

1. City hotels were more preffered by the guests because it is more easily accessible.

2. Portugal country has most guests with 31.41% followed by Britain=11.93 & France=10.11

3. 27.48% of total bookings were cancelled which is 24025 cancellations

4. City hotels have 16049 cancellations which is more then Resort hotels which had 7976

5. Type 'A' room was booked 64.70% with 56552 bookings followed by type 'D' with 17398 bookings and type E with 6049 bookingsRoom type 'L' and 'P' is booked very few times with only 6 booking

6. Most of the Guests are using 'Online TA' which is 59.06% with 51618 bookings Followed by 'Offline TA' and 'direct' are almost same at 15.89% and 13.5%

7. Most of the ADR is contributed by GDS and TA and Direct reservations

8. Set Non-refundable Rates, Collect deposits, and implement more rigid cancellation policies.

8.1 98.68% of the guests did not make a Deposit and very few made Non refundable deposits

9. 77.31% of the guests order Bed & Breakfastal and around 10% of the guests select Half Board and 10% don't opt for meal packages

10. Only 3.19% of the guests were repeated guests with only 3810 of total guests

10.1. Resort hotels had more repeated guests then City hotels

11. 75.05% of the guests are Transient guests and 21.04% are Transient-Party guests

12. August month had most travellers with 12.88% which is 13877 guests

12.1. July and May has the next most travellers with 10057 and 8355 guests

13. 50.26% of total guests stay in Hotels during Weekend

13.1. Most of the guests stay 2 days or 1 day during Weekend

14. 25.39% of the guests stay only 1 day during Weekdays and 23.73% guests stay for 2 days and 18% stay 3 days during Weekdays, Most of the guests stay 1 day Very negligable percentage of guests stay above 3 days

15. 19% of the bookings made were for Single Rooms

16. Guests who has not cancelled their previous booking has more probability to be a Repeated Guest

**Improvement Ideas**

1. Resort hotels can increase their business by improving transportation availability for guests

2. Improve marketting in countries which have very low guest counts

3. Mandate deposits to prevent cancellations

4. Improve and advertise the other types of rooms which are not getting booked

5. Try to convert Other rooms to Type A room which is preffered most

6. Give attractive discounts to increase repeated guests

7. Create attractions in other months to improve business

8. Give Discount offers if anyone is staying over 7 days to get a stable income