**Hotel Booking Analysis**

**Rohit Sakharkar, Nikita Negi,**

**Mithlish Singh, Ashish Yadav, Shubham Sanklecha**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

In this project “Hotel Booking Analysis” from given dataset we found out that to know more details about each of these hotels before the actual booking. Like which customer more preferred hotel, repeated guest, what is the monthly booking of different hotels and many more.

Through this analysis we can understand the customer’s behavior and it might help us make better decisions further.

**1.Problem Statement**

The main aim of this Hotel Room Booking Analysis Project is to give simple application which provides all facilities like room booking, room class type, hall booking, hotel booking, etc… These are the below methods used in this project:-

Find null values

Find outliers

Data wrangling

The main objective is to build a predictive model, which could help them in predicting the hotel booking. This analysis would help the customer to find the right hotel quickly and efficiently.

**2. Introduction**

### The given data contains two datasets with hotel booking information. One of the hotels (H1) is a resort hotel and the other is a city hotel (H2). Both datasets share the same structure, with 31 variables describing the 40,060 observations of H1 and 79,330 observations of H2. Each observation represents a hotel booking. Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017.

The dataset contains booking information such as the number of adults, children, and/or babies, and the number of available parking spaces, arrival date month, arrival date day of month, arrival date year, is repeated guest, previous cancellations, previous bookings not cancelled, reserved room type, booking changes, among other things. We have done exploratory data analysis with python to get insight from the data.

### Our goal here is to build a predictive model, which could help customers in predicting which hotel would be best for booking surge pricing type proactively.

**3. Steps involved**

* **Data Preparation and Cleaning**

Data cleaning in this part we remove incorrect, corrupted, incorrectly formatted, duplicate, and incomplete data within a given dataset.

* **Null values Treatment**

The data which we use it consist of large numbers of null values. We handled by deleting the rows and columns having null values and the columns which having more then half of the null values that entire column we drop and rows as well. For in order to get accuracy.

* **Handling Outliers**

An outlier is any data point that falls above the 3rd quartile and below the first quartile is called as outliers. So we use boxplot to show the outliers. Then we remove those outliers by adding same conditional to get and update those values Now our data is clean.

* **Exploratory Data Analysis (EDA)**

The main purpose of EDA is to help look at data before making any assumptions. It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, find interesting relations among the variables. In this part, we looked at the data frame and decided our target variables (Important Columns) based upon which we were going to conduct further analysis from that it easy to understand on the given database that how many guest where preferred the resort hotel or city hotel, yearly booking of different hotel, the booking and cancelled status of different hotel, relation between price and month, average daily rate and many more.

* **Confirmed booking**

we have found out in this part that how many confirmed bookings are there according to the months.

The confirmed bookings goes from their lower value (4115) in January to their highest value (8624) in august.

* **Cancellation Booking**

In this part, we have found out the total number of cancellation of bookings**.**

During the year, total booking cancelled is 44199.

And, in percentage 37.13%

Bar chart of cancelled booking.

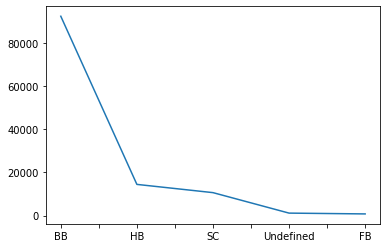


* **Meal**

In hotel meal is very important factor. If meal and services of hotel is good then the percentages of repeated guests alternately increase. there are many types of meal are available in hotels according to children, adults and for aged persons. different types of meal plans are as below.

1. Undefined/SC — no meal package;
2. BB — Bed & Breakfast;
3. HB — Half board (breakfast and one other meal -usually dinner);
4. FB — Full board (breakfast, lunch and dinner);

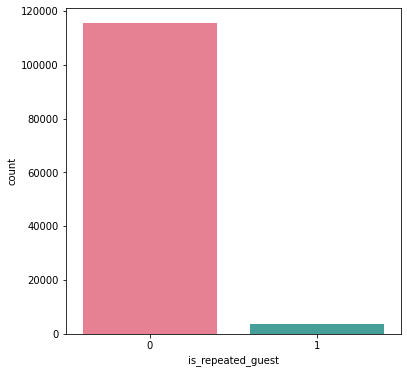
So on the basis of our project most guests prefer B&B type meal



* **Repeated Guests**

If the overall services of hotel are good then the most customer prefer to stay in that same hotel again and again. All services like free wi-fi, wake up calls, pool, meeting room, restaurant, gym etc.

As per our project there are 3755 no. of repeated guests that means There were good share of guests who visited and stayed in the hotel more than once, which shows that the overall service of the hotel was good.



* **Deposit Type**

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.

There were 876277 No deposit,122364 Non refund and 1359 Refundable.

* **Average Daily Rate(ADR)**

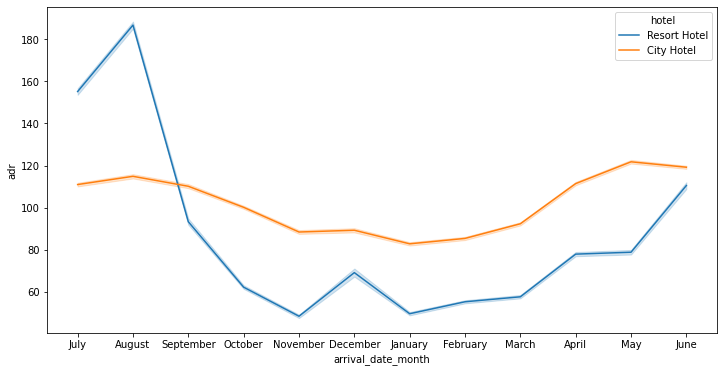
The average daily rate (ADR) measures the average rental revenue earned for an occupied room per day. The operating performance of a hotel or other lodging business can be determined by using the ADR. Multiplying the ADR by the occupancy rate equals the revenue per available room.

In these we used line plot graph

to check the average daily rate of

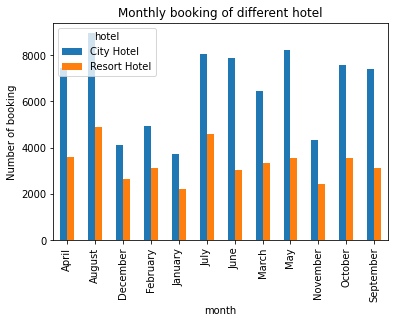
different hotel.

Where we conclude that- resort hotels, the average daily rate is more expensive.



* **Monthly Booking Of Different Hotel**

Our dataset contains a large number of hotel booking. So, In this step we are checking monthly wise booking of different hotel. When the data from april to september was extracted through the graph, it found that booking done in each month of city hotel is higher than resort hotel.

****

* **Yearly Booking Of Different Hotel**

In this section we have found the

number of booking held in the three

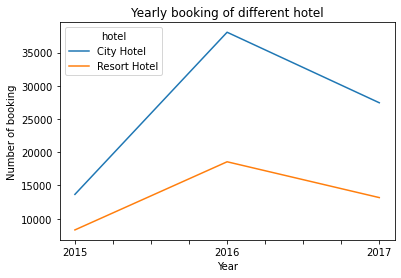
different years i.e. 2015, 2016, 2017

of City Hotel and Resort Hotel. At

last concluded that bookings done in

City Hotels are larger than bookings

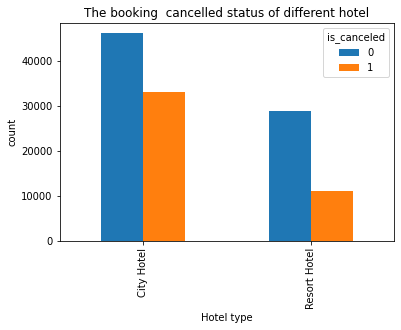
done in Resort hotels.



* **The Booking And Cancelled Status Of Different Hotel**

In this Part we try to find the Booking and Cancelled Status of City Hotel and Resort Hotel and described ‘0’ as Booking status and ‘1’ as Cancelled status and finally comes to a conclusion that Booking cancelled by customers in City Hotel

is larger than Resort Hotel.



**4. Conclusion:**

That's it! We reached the end of our exercise.

In this project first loaded the data and then cleaned the data, null values treatment, encoding of categorical columns, EDA, feature selection and then model building.

Using the results from the analysis, we can determine which hotel would be best for booking city hotel or resort hotel and can understand customer behavior and business can make key decisions regarding the customer experience they desire to deliver.

**References-**

1. Kaggle
2. GeeksforGeeks
3. Analytics Vidhya
4. GitHub