**Hotel Booking Analysis**

**Rohit Sakharkar, Nikita Negi,**

**Mithlesh 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**

Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business. We can use the patterns to predict the future bookings using time series or decision trees.

We will be using the data available to analyze the factors affecting the hotel bookings. These factors can be used for reporting the trends and predict the future bookings.

We will be analysing some key metrics for hotel bookings like:

1. Cancelled booking
2. Confirmed booking
3. Arrival date by month
4. Year of arrival
5. Preferred meal type
6. Customer type
7. Booking change when cancelation
8. Repeated guest

* **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 (4068) in January to their highest value (8618) 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 37.07%

Bar chart of cancelled booking.

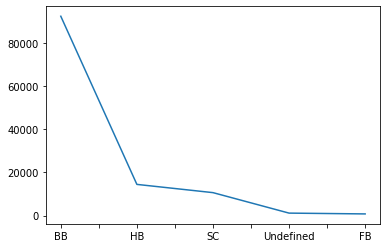


* **Meal**

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);

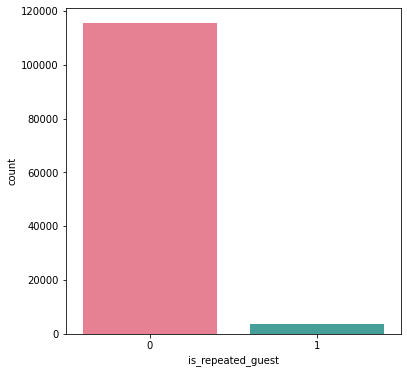
So on the basis of our project most guests prefer B&B type meal



* **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 31499 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)**

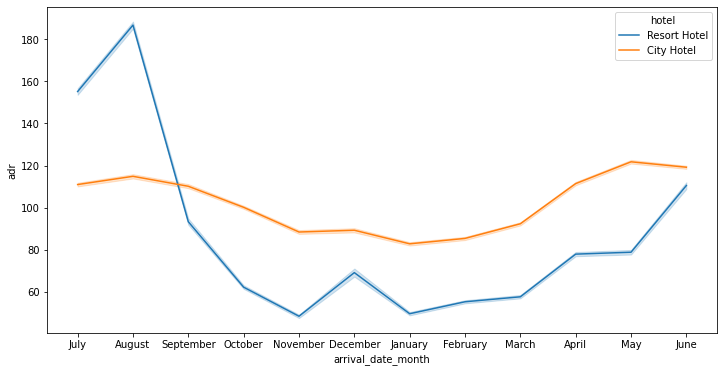
The average daily rate (ADR) measures the average rental revenue earned for an occupied room per day. The operating performance of a hotel or other lodging business can be determined by using the ADR. Multiplying the ADR by the occupancy rate equals the revenue per available room.

In these we used line plot graph

to check the average daily rate of

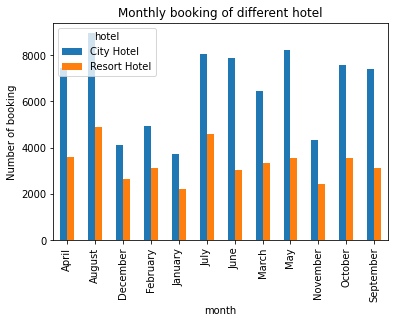
different hotel.

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

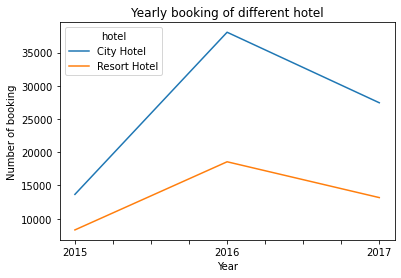
different years i.e. 2015, 2016, 2017

of City Hotel and Resort Hotel. At

last concluded that bookings done in

City Hotels are larger than bookings

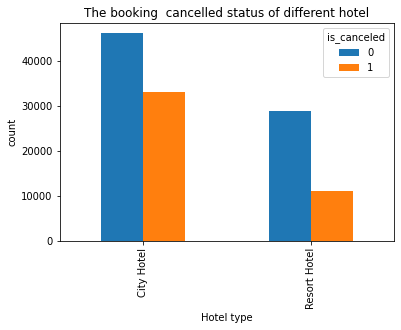
done in Resort hotels.



* **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.

We used the dataset that contains data about hotel bookings.

We cleaned and preprocessed the data and then we performed the exploratory data analysis to extract information from the data to answer the following questions:-

* Maximum number of booking has been observed in Month of August and lowest in January.
* Total number of booking was cancelled 75011
* Customer most preferred hotel is “city hotel.
* Highest number of booking have seen for both hotels is 2016.
* Cancellation is more in city hotels as compare to resort hotel.
* August is the best month for the hotel business in all years(2015-2017) because average “ADR” is the highest of August for all years.
* Almost 77.3% of people preferred Bed and Breakfast at the hotel where they stay.

**References-**

1. Kaggle
2. Github
3. GeeksforGeeks