**Hotel Booking Analysis**

**Aman Sharma**

**Jitesh Kumar Srivastava**

**Pavani Billa**

**Data science trainees,**

**Cohort Milan**

**AlmaBetter, Bangalore**

**Abstract:**

We have been given a data set of hotel booking in which data set contains booking information for a city hotel and a resort hotel, and include 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.

Our job is to do Exploratory Data Analysis (EDA) and came up with excited analysis of the data set. For this, we have a team which comprises of three member.

**1. Focused Question**

As in team we have decided that each member can frame three or four questions so that we can have more than ten question in total. The questions are

1) Hotel wise confirmation and cancelation of the bookings

2) Given dataset consists booking status of three years viz. 2015, 2016 and 2017. We were curious to know what are the trend in booking status of both the hotel are in these years.

3) We wanted to know that how many number of guest arrived per month in these three years.

4) From which country maximum tourist came?

5) We were inquisitive about the number of repeated guest coming to the both hotels.

6) We have gone through the dataset and found that some guest come to hotel by their own vehicles so we wanted to know about the parking area requirement in all the three years.

7) How many bookings are done by different market segment?

8) After going through the dataset we found that there are 53 total weeks, so we wanted to know how many people have stayed in these hotel per week for a given period.

9) How different channel are contributing to the bookings?

10) What are the average daily rate of both the over a given period of months?

11)How many days in advance the two hotels are booked?

The main idea behind these questions were to think and learn to solve business related problem for various stakeholder in the future to come.

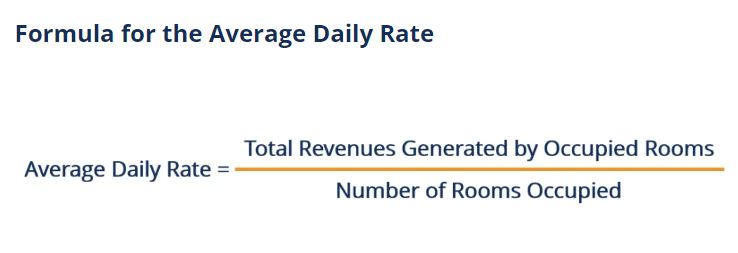
**2. Introduction**

After deciding the above question we engage ourselves to how to answer these question and how we can extract an insightful solution of our problem and it should be similar to real time project that we will going to solve in times to come. All the definition and graph used as example in this document are taken from internet and are taken from various sites through internet.

## **3. ADR.**

There is a kpi used in the dataset ‘**adr**’. So we go through different sites and found out that it stands for **average daily rate**, and it’s used to measure the average revenue that a hotel receives for each occupied guest room per day.

In mathematical term**;**



For e.g.

Suppose there is a Hotel A, there are total 50 rooms available for booking. On a particular day,

Numbers of rooms booked = 20

Revenue generated = ₹24000

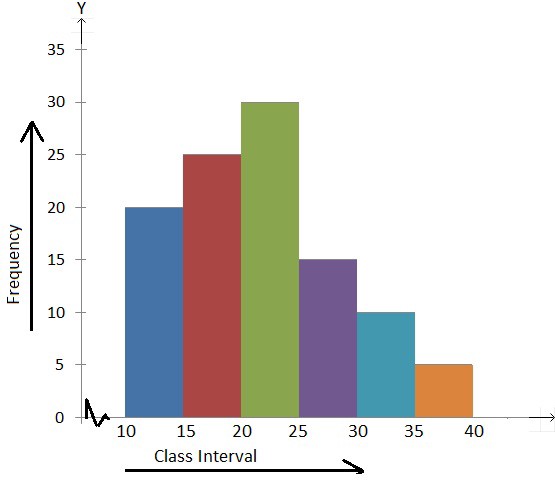
Adr = ₹24000/20

Thus, Adr = ₹1200 on that particular day. We can say that higher the ‘adr’, larger the revenue thus larger the profit of that particular hotel.

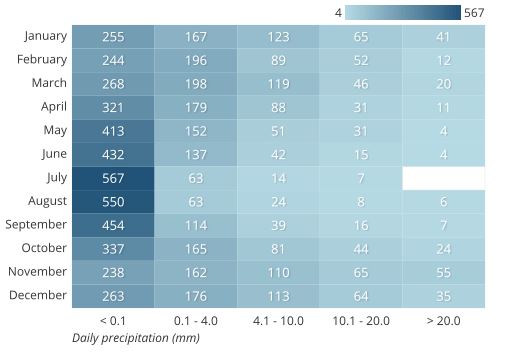
## **4. Types of graphs:**

In this EDA, We have used different types of graph for better visualization of our analysis and these are

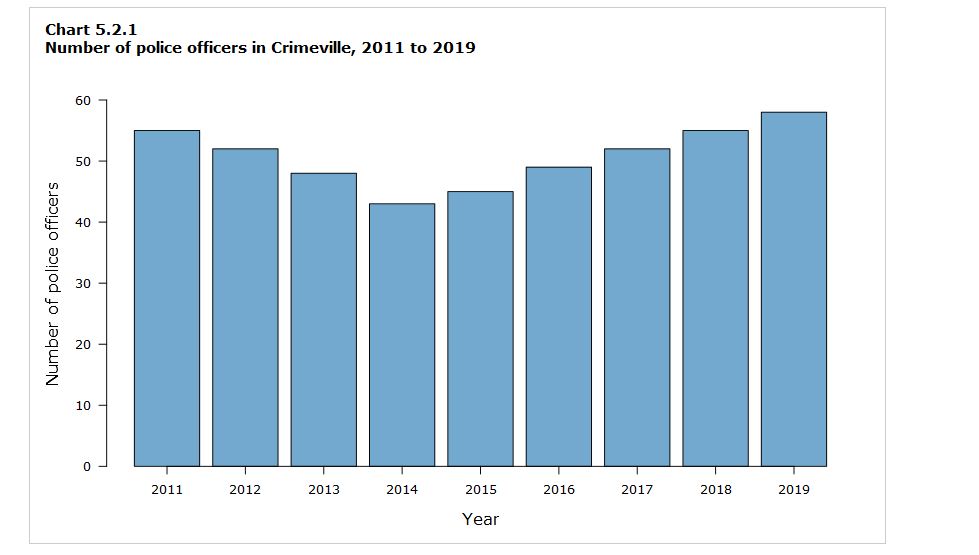
1. Histogram - It is used to summarize discrete or continuous data that are measured on interval scale.



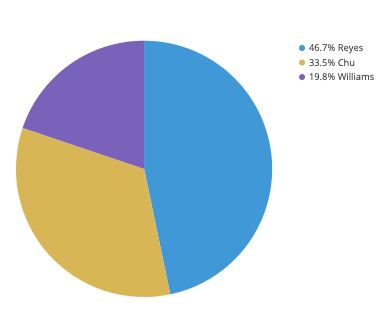
1. Heat map - A heatmap (aka heat map) depicts values for a main variable of interest across two axis variables as a grid of colored squares. The axis variables are divided into ranges like a bar chart or histogram, and each cell’s color indicates the value of the main variable in the corresponding cell range.



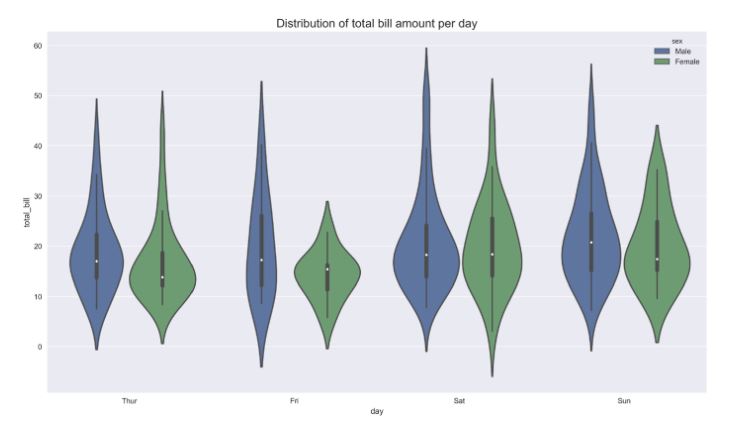
1. Bar Graph - The bar graph is used to compare the items between different groups over time.



1. Pie Chart - A pie chart shows how a total amount is divided between levels of a categorical variable as a circle divided into radial slices. Each categorical value corresponds with a single slice of the circle, and the size of each slice (both in area and arc length) indicates what proportion of the whole each category level takes.



1. Violin Graph - A violin plot is a hybrid of a box plot and a kernel density plot, which shows peaks in the data. It is used to visualize the distribution of numerical data.



1. Line Chart - A line chart is a type of chart used to show information that changes over time. Line chart are created by plotting a series of several points and connecting them with a straight line. Line chart are used to track changes over short and long periods.



# **5. Preparations**

In the preparation of solution we face many challenges and subsequently we solved that by discussion, going through the lecture video and discovering it through internet.

Before starting the analysis we look for the NaN value and replace it with according to data. After doing this we go through the column name and data type of column. It gives us idea about the data we are going to operate. After doing this we sign up on GitHub individually. Everyone created their own repository and started our respective work. Working on GitHub gives us many joy and we enjoyed it thoroughly. Going through the python and its libraries like NumPy, Pandas, Matplotlib and Seaborn was not an easy task.

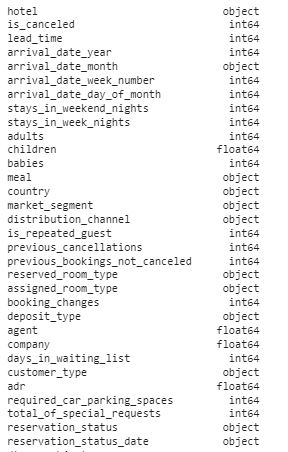


Fig 1 Column name and type of data.

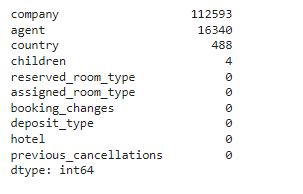


Fig.2 Sum of all Null values in a particular column.



Fig.3 Null values after removing of null values.

**6. Final Step:**

After preparation each team member solve their respective question and everyone came up with their respective interesting insights.

During this we gel good as a team and every member supported each other. After solving this we submitted our findings.

**7. Challenges:**

We encountered with many challenges as we were new to python and its libraries so it’s was a tough task and we overcame this by helping each other and most of the time surfing through internet.

There was many ambiguity regarding number of member in a team and the important various step, but thanks to Almabetter dashboard and community it solves all our queries.

**8. Conclusion:**

So, we comes to our final part with our findings which are listed below:

1. The City hotel is having greater number of bookings as compared to Resort hotel. But, the cancellation percentage is high of the City hotel.
2. In the year 2016 both the hotel saw a massive increase in their bookings and by far the year 2016 is the year of the highest bookings of both hotel. In each year that is 2015, 2016 and 2017 the city hotel is having the highest number of bookings.
3. Most number of the guests arrived in the month of August and least was in the month of January.
4. Most number of lead time was generated in the July month and least was in January month.
5. We can see that there is very low number of repeated guest in both the hotel. A lot of attention should be drawn to solve this problem because repeated guest know the services of hotel very well.
6. Most of the visitors do not carry children with them every year. Very less number of children has been carried by them around 1 and 2 only every year.
7. In the year 2015 from July the requirement of the car parking spaces was decreasing till September and then in the year 2016 in the month of August the requirement was on peak. Whereas in the year 2017 there was a steady requirement of parking spaces.
8. Online TA has the maximum booking followed by offline TA/TO in maximum booking and Aviation has the least booking.
9. From the week 28 to 31 has shown the highest days of stay whereas from the week 1 to 11 has shown a very steady trend in the number of stays and also the week 18 to 22 has shown the least number of stays by the visitors in aggregate of all 3 years 2015, 2016 and 2017.
10. TA/TO of the hotels have maximum contribution to the booking followed by Direct. Both the hotels need to expand their distribution channel which in results will expand their businesses.
11. Resort hotel the highest average daily rates is in the month of July and City hotel the highest average daily rates is in the month of May. For Resort hotel the least average daily rate is in month of November and for City hotel it is in month of January. We can also say that average price(adr) of city hotel is somewhere between 85 to 125 and for Resort hotel the average price(adr) is between 55 to 185 thus we can say that the average price of hotel Resort is volatile. The graph depicts that for the month of July to mid-August stay at Hotel Resort is costlier and from month of September to June stay at Hotel city is costlier.
12. Most of the booking of both the hotels are done within 100 days of check-in date to hotel and maximum booking are done on the same check-in day.