**Hotel bookings Analysis project**

**Avinash Gangarde,**

**Data science trainee,**

**Almabetter, Bangalore**

**Abstract:**

Hotel industry is crucial part of the service industry as well as tourism industry. Managing the available resources would be challenge for them .So we want insure that they could use their resources in better way as well as avoid the waste of resources .This dataset could help them to understand the bookings and their different aspects of it as well as choices of customers.

Our experiment with this data is to analyze the data and give valuable insights for the hotels industry. We have organized the data in best possible way to get helpful insights from it.

***Keywords: EDA, hotel booking, data visualization, data cleaning***

**1. Problem Statement**

Data provided is about the hotels, we have do analysis of data and bring out helpful insights from that. We want to analyze the data with different aspect of hotel industry. We can help them to take better decisions and optimize the use of resources and save lot resources and money. They should able to understand the consumer and whole aspect of hotel performance.

### Our goal is to help people who are in hotels industry. We want help them to take better decisions .So they could manage resources better and run their campaign efficiently.

Our main objective is to get valuable insights in graphical format and plot as many as comparing plots from the given data.

* hotel: Name of hotel ( City or Resort)
* is\_canceled: Whether the booking is canceled or not (0 for no canceled and 1 for canceled)
* lead\_time: time (in days) between booking transaction and actual arrival.
* arrival\_date\_year: Year of arrival
* arrival\_date\_month: month of arrival
* arrival\_date\_week\_number: week number of arrival date.
* arrival\_date\_day\_of\_month: Day of month of arrival date
* stays\_in\_weekend\_nights: No. of weekend nights spent in a hotel
* stays\_in\_week\_nights: No. of weeknights spent in a hotel
* adults: No. of adults in single booking record.
* children: No. of children in single booking record.
* babies: No. of babies in single booking record.
* meal: Type of meal chosen
* country: Country of origin of customers (as mentioned by them)
* market\_segment: What segment via booking was made and for what purpose.
* distribution\_channel: Via which medium booking was made.
* is\_repeated\_guest: Whether the customer has made any booking before(0 for No and 1 for
  + - Yes)
* previous\_cancellations: No. of previous canceled bookings.
* previous\_bookings\_not\_canceled: No. of previous non-canceled bookings.
* reserved\_room\_type: Room type reserved by a customer.
* assigned\_room\_type: Room type assigned to the customer.
* booking\_changes: No. of booking changes done by customers
* deposit\_type: Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund)
* agent: Id of agent for booking
* company: Id of the company making a booking
* days\_in\_waiting\_list: No. of days on waiting list.
* customer\_type: Type of customer(Transient, Group, etc.)
* adr: Average Daily rate.
* required\_car\_parking\_spaces: No. of car parking asked in booking
* total\_of\_special\_requests: total no. of special request.
* reservation\_status: Whether a customer has checked out or canceled,or not showed
* reservation\_status\_date: Date of making reservation status.

**2. Introduction:**

We are given with the hotels booking dataset. We want perform EDA on this dataset and find out the general trends related to the hotel industry and analyze them to bring out valuable insights from it. We have almost 119398 entries of data related to city hotels and resort hotels .We kept our analysis mostly around the hotels and comparing different things around it.

Our goal is to get as many as possible insights from the dataset and help the hotel industry to take better decisions and manage their resources best possible way.

**3. Steps involved:**

* **Removing the duplicate**

After loading the data we removed the columns that are not needed or does not contain any important information .

* **Null values Treatment**

We have null values present in three columns and we deal with them best possible way.

We have children columns where we replace null values with zero because zero is the minimum value present in the columns.

Second columns is country where we replaced it with anonymous because we don’t know the information about the country.

* **Encoding of categorical columns**

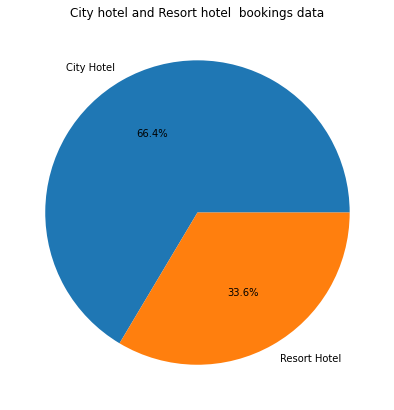
We have seen one columns in the data where the information present in it is encoded in the format of 1 for canceled booking and 0 for the booking accepted .We have replaced the 0 and 1 with the words of booked and canceled.

* **Data analysis**

We have done the analysis of data using seaborn library and matplotlib as well as the counterplot to plot the graphs and charts.

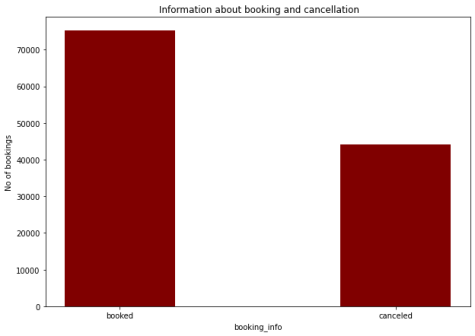
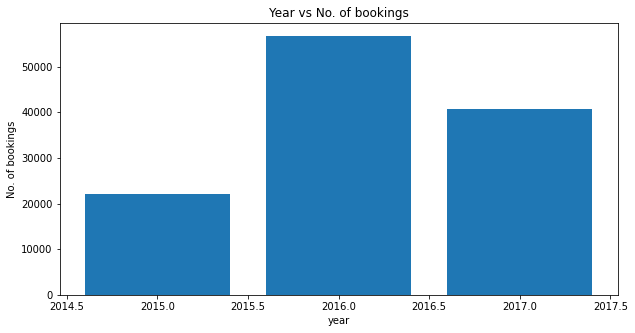
**4. Univarient Analysis:**

Here we are going to see the few important plots that would help us to understand about data analysis .

****

Fig(a)

From above pie chart we found that there are 66.4% hotels are city hotels and 33.6% hotels are Resort hotels.

** **

Fig(c)

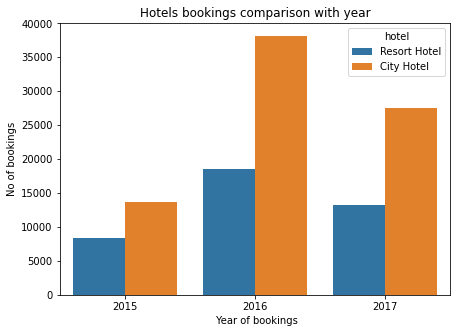
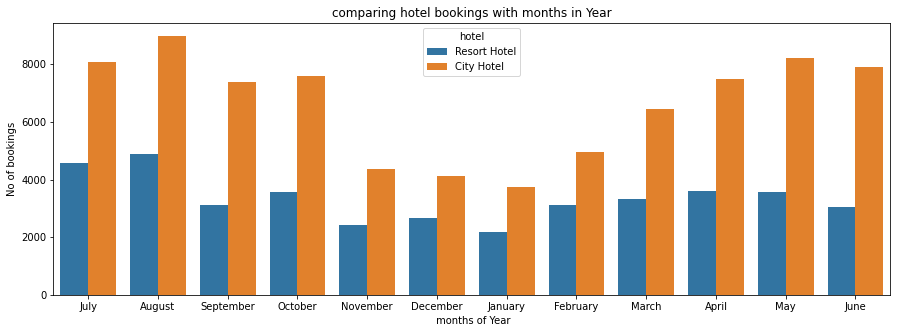
Fig(b)

We can see in fig (b) that the booking rate is higher than cancelation rate and the in fig(c) we can see that the no of bookings were high in 2016 followed by the 2017 and 2015

These are some important analysis chart that we have analyze and there are also some other analysis chart also present in EDA

**5. Bivarient Analysis:**

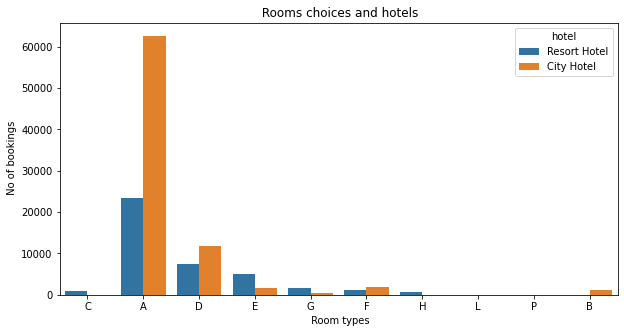
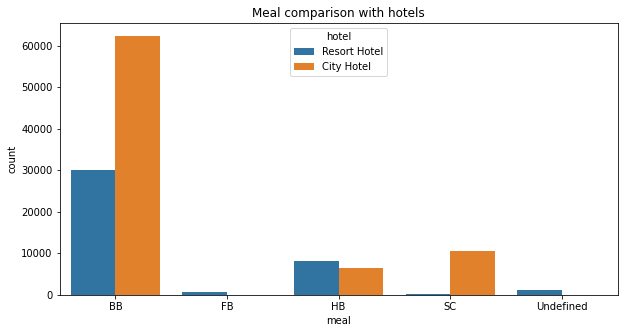
In bivarient analysis we have taken two variable and compared with each other with keeping the hotels are important varient to be analyze with different variable in terms of booking. Let’s see it.

Fig(e)

Fig(d)

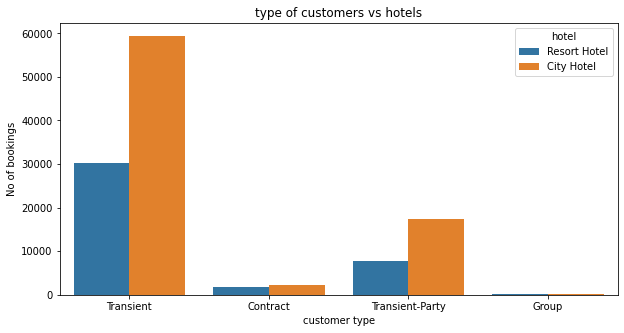
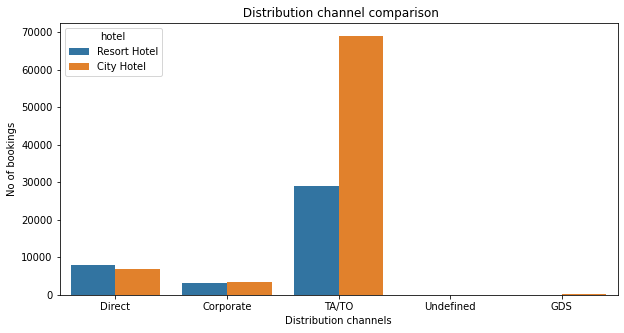
From fig (d) we can see that the city hotels have more bookings than the resort hotels and in fig (e) we can see that the August is the busiest month and December is the least busiest month amongst all.



Fig(g)

Fig(f)

We can see in fig (f) that the BB is topmost preferred meal and in figure (g) we have seen that the A room type is mostly chosen in both hotels most number of time .

****

Fig(i)

Fig(h)

We can see here in fig (h) that the TA/TO has the most number of bookings in city hotels is top distribution channel in terms of booking and direct distribution channel is having more bookings in Resort hotels than the city hotels . In fig (i) we can see that the city hotels and resort hotels have more Transient type of customer , followed by the transient party and contract .

**6. Conclusion:**

Here we are at the conclusion after performing EDA on this dataset. We want conclude the following things.

* 66.4 % hotels are City hotels and 33.6 % are Resort Hotels.
* 62% is booking rate and 38% are cancelation rate
* city hotels have more cancelation rate and vice versa.
* Lead time is more in city hotels than the Resort hotels.
* August is busiest month and December is least busiest months.
* BB ( bed and breakfast) mostly preferred meal type in city and Resort hotels.
* Portugal, Great Britain, France are top three countries with most number of bookings.
* TA/TO is top performing distribution channel among rest of all.
* Both hotels have mostly transient type of customers.
* Chances of getting same assigned room are more in Resort hotel than city hotel.
* Bookings are mostly done by couples followed by family and friends and thereafter singles.

**References-**

1. www.medium.com
2. analytics vidya
3. GeeksforGeeks
4. Web3school