Capstone Project - 1

Hotel Booking Analysis

Saurav kumar

Index

➢ Defining problem statement
➢ Data cleaning and feature engineering
➢ Hotel wise analysis
➢ Cancelation wise analysis
➢ Market segment wise analysis
➢ Average daily rate(ADR) based analysis
➢ Challenges
➢ Conclusion

Problem Statement

First of all before going through any code or analysis we must know what is the reason for doing this analysis.

Data provided to us have many attributes which refer to a particular booking, our task is to remove noise from that data, and then find relation among the different attributes, to visualize the behavior of attribute or relation of two or more attributes using eda. To gain understanding from the data, we will use Python to undertake exploratory data analysis.

Our task involves to analyze the data and determine the best time of year to book a hotel room , the optimal length of stay in order to get the best daily rate? To predict when a hotel was likely to receive a disproportionately high number of special requests? And many other important visualizations.

Data Summary

- Data of Hotel booking status from july 2015 to August 2017
- Some important Features in hotel booking dataset
  - **Hotel:** Two types of hotels are given in the dataset that are city hotels and resort hotels.
  - **IS_CANCELED:** Shows whether a booking made in advance is canceled(1) or not(0).
  - **LEAD TIME:** Contains the number of days between reservation date of booking and arrival date.
  - **Arrival_date_year:** Year of arrival date.
  - **Arrival_date_month:** Month of arrival date.
  - **Stay_in_weekend_nights:** Numbers of nights stay in weekend
  - **Stay_in_week_nights:** Number of nights stay in week
  - **Adults:** Number of adults
  - **Children:** Number of children
  - **Meal:** Type of meal booked.
  - **Country:** Country of origin.

Data Summary

- Some important Features in hotel booking dataset
  - **market_segment:** Market segment designation, way through which marketing of hotel is done
  - **Distribution_channel:** Booking distribution channel
  - **Is_repeated_guest:** Is a repeated guest (1) or not (0)
  - **Adr:** Average daily rate as defined by dividing the total revenue by total room sold.
  - **Required_car_parking_space:** Number of car parking space required by customer.

# Data Cleaning and Feature engineering

For data cleaning:

- We removed the duplicate rows.

- **Removing null values**
  - We remove the company column from dataset as it has more than 80% of entries unavailable(null).
  - We replace the NaN value from agent and children column with 0.
  - We replace NaN value in country column with other.

- **Feature engineering**
  - We add two columns :
    - Total_guest from columns children, adults
    - Day_stay from columns weekend night and week night

- **Remove outliers**
  - We remove outliers from adr and lead time column using Quantile

Removing Outliers and boxplot

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. They may reduce the accuracy of our prediction by changing the mean value to a great extent so outlier needs to be removed necessarily.

As adr and lead_time are continuous columns so we removed outliers from these columns. We used upper and lower quartiles for finding out and removing the outliers.

To visualize whether the outliers are removed or not we used a boxplot visualization. Boxplot Visualization for adr is show here.
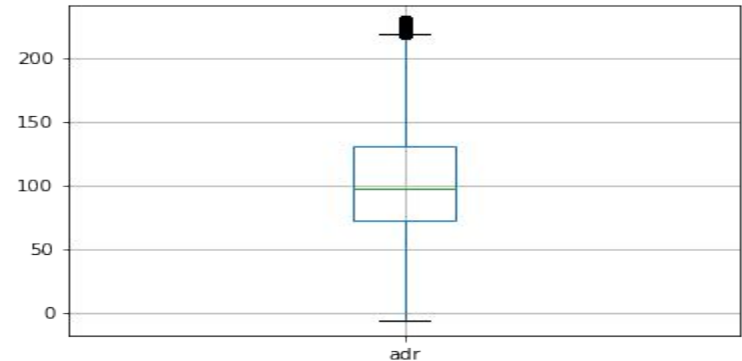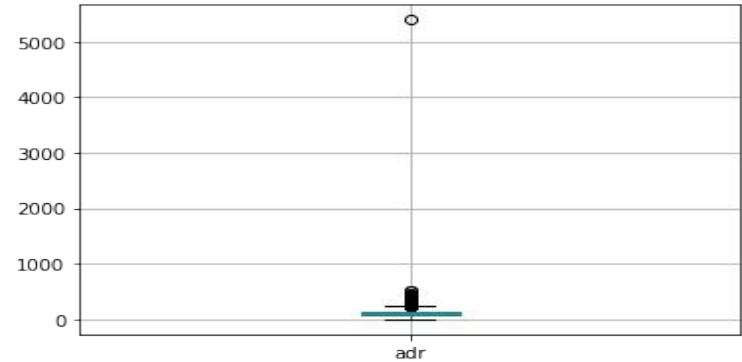
Hotel wise analysis

- City hotel was more booked from the month of march to june and maximum booked in the month of May.

- Resort hotel was more booked in the month April, may, october and maximum booked in April
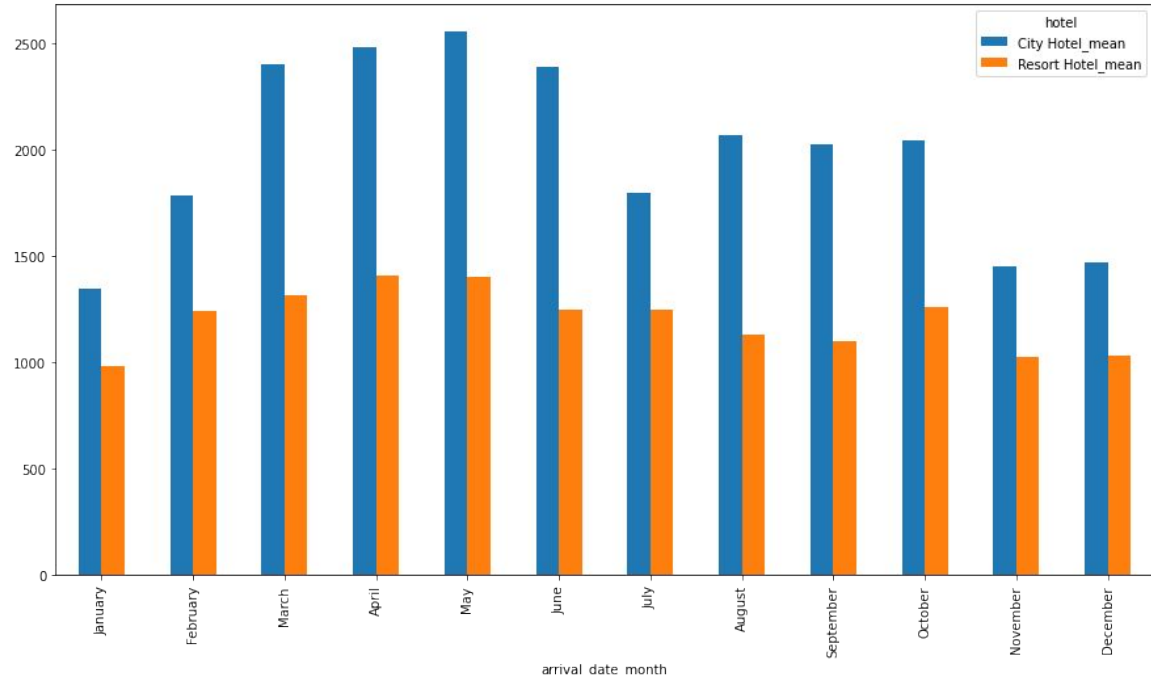
Hotel_booking_country_wise

- Protugal has maximum hotel booking followed by Great Britian and France.

- In Protugal and Britain resort hotel is more booked than city hotel, whereas in France city hotel is more booked than resort hotel.
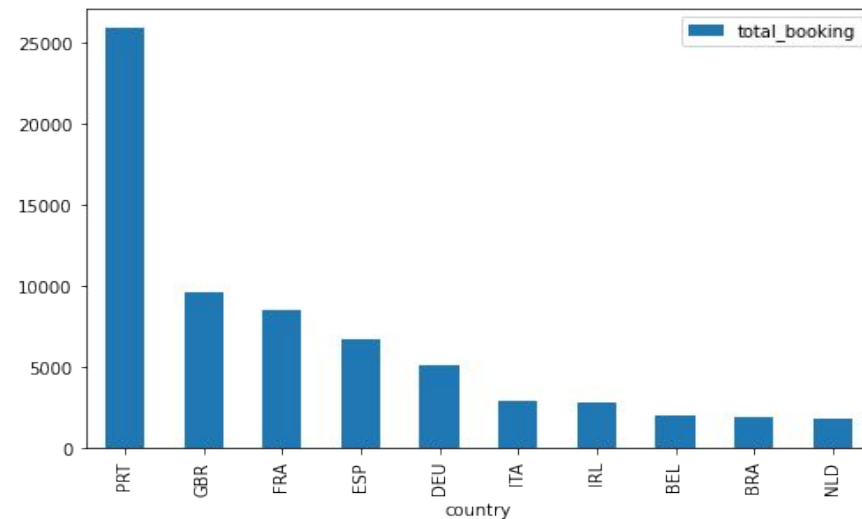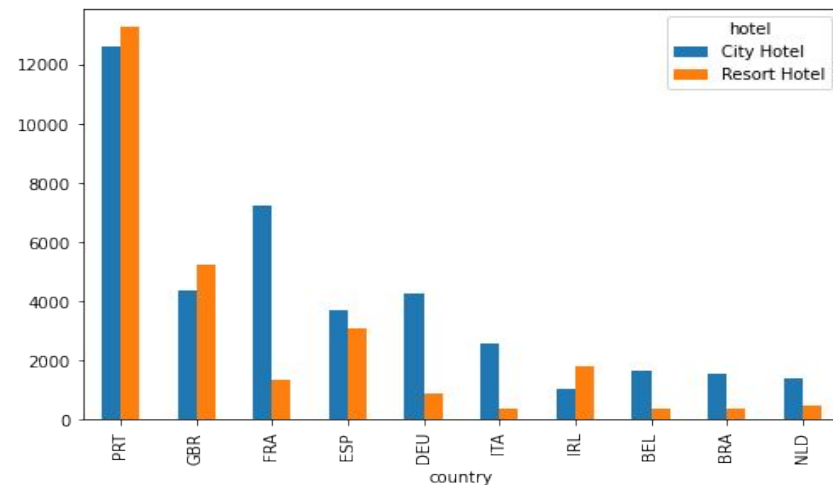
Most meal type prefered by guest

- Bed and Breakfast(BB) is the most common meal prefer by the guest. In BB meal, breakfast is included in the room rent

- FB meal is only chooses in resort hotels not in city hotel

Resort_hotel_meal
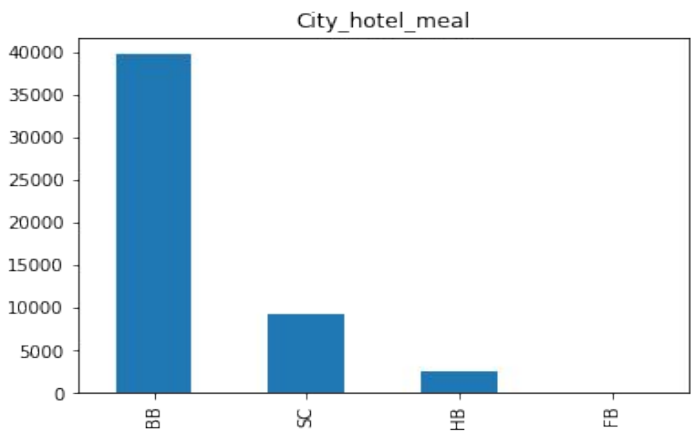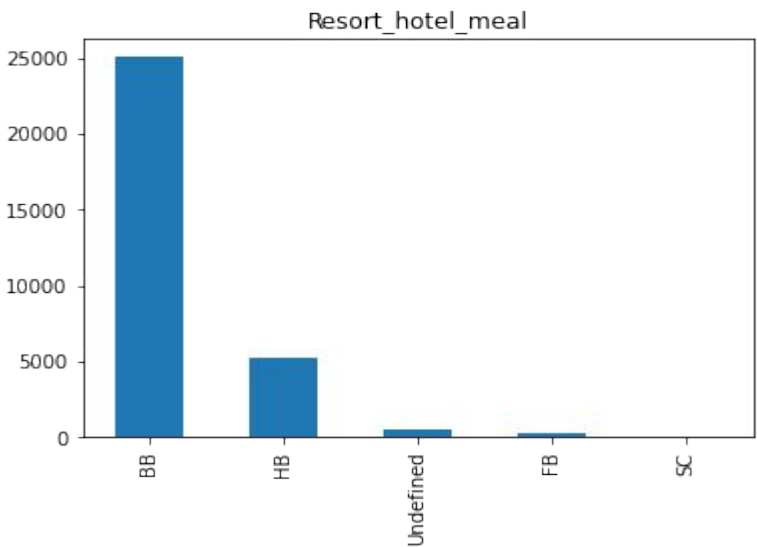
City_hotel_meal

Comparison of parking space between Resort Hotel and City Hotel

- Mostly City hotels does not required any parking space, only 2.5% of guest in city hotel needs 1 parking space.

- For resort hotel these percentage increases as about 13.5% of guest requried 1 or more parking space.

- Car parking requirements are more for resort hotels, so they must work on their parking facilities
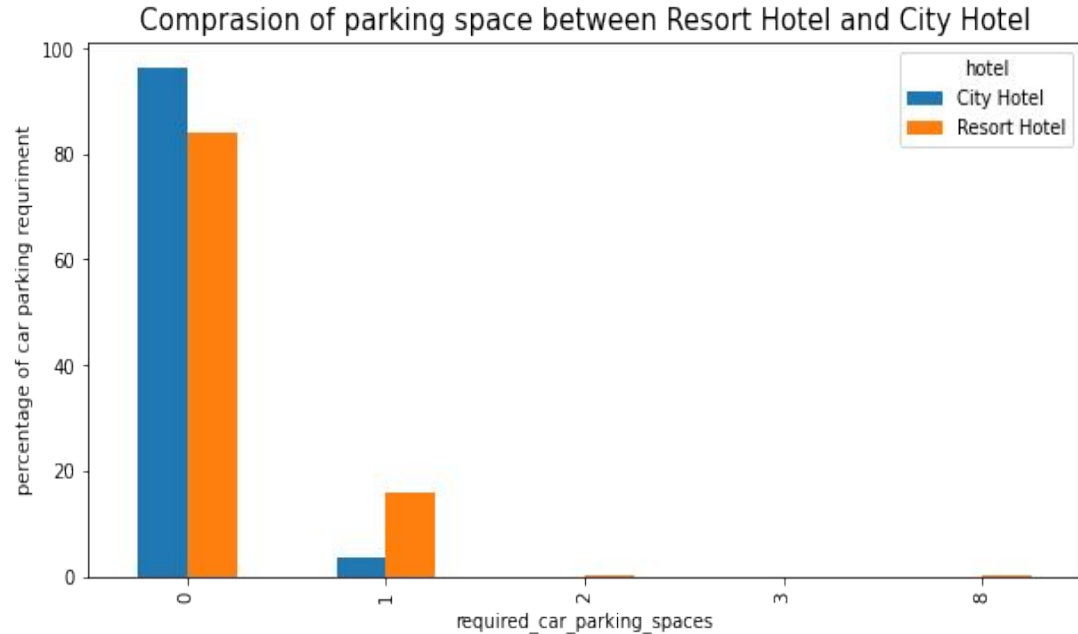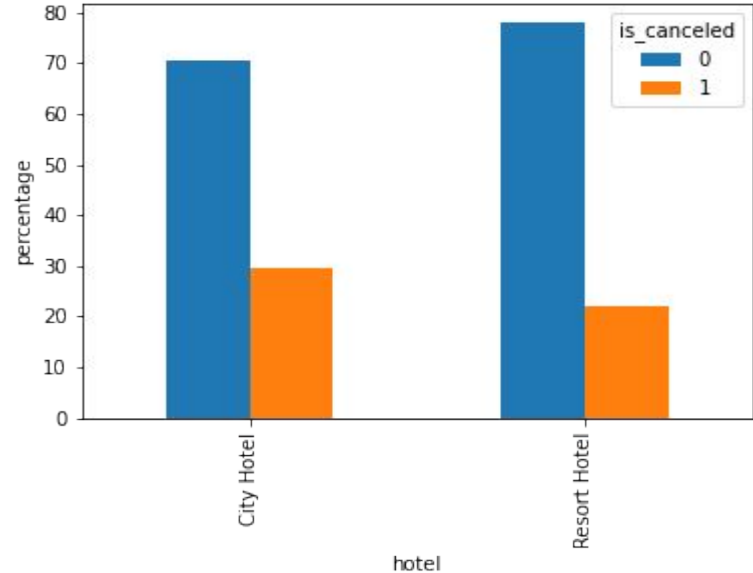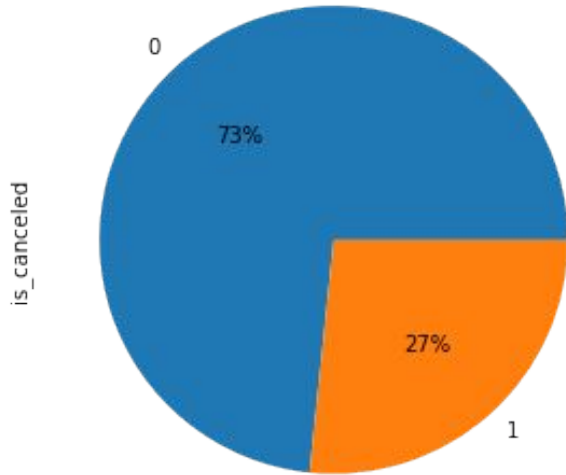
Number of day stay vs hotels
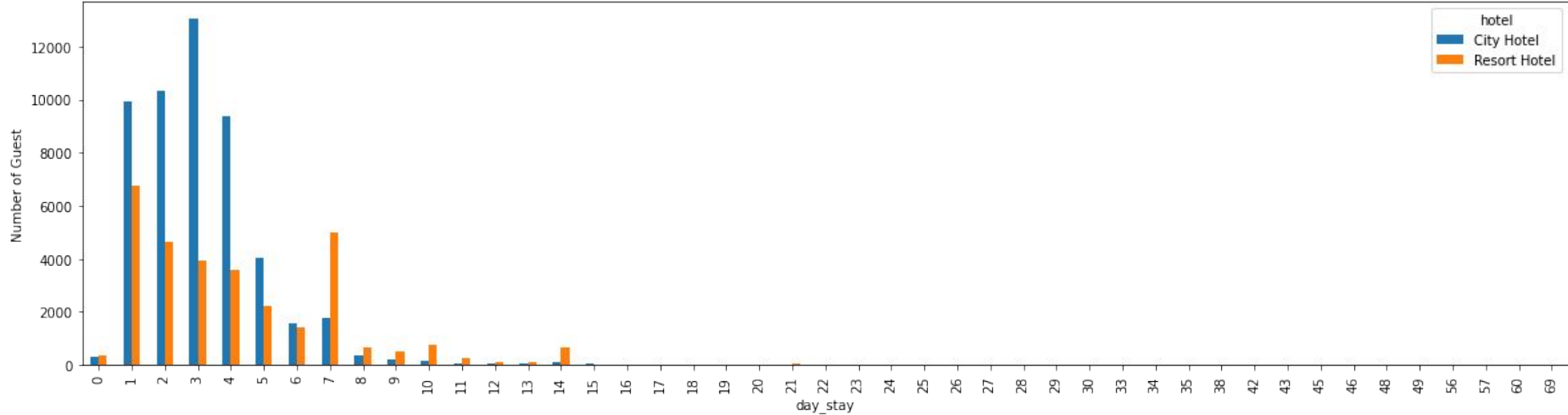
- Tourist generally booked the resort hotel for the long term stay and city hotel for the short term stay

Cancelation wise analysis

- About 27% reservation in hotels is cancel after booking

- The probability of cancelation of booking in both resort and city hotel is approx same. 29% of total booking in city hotel is cancelled whereas for resort hotel cancelation rate is 22%

Market Segment v/s cancelled percentage

- canceled percentage is high for undefined type of market segment that is bookings done through these sources have 100% chances of cancelation
- cancelation percentage is also high for grouped market segment, the bookings through them are more likely to be canceled.

so hotels can have extra bookings in the case rooms are booked by one of the two segments

Market Segment v/s cancelled percentage
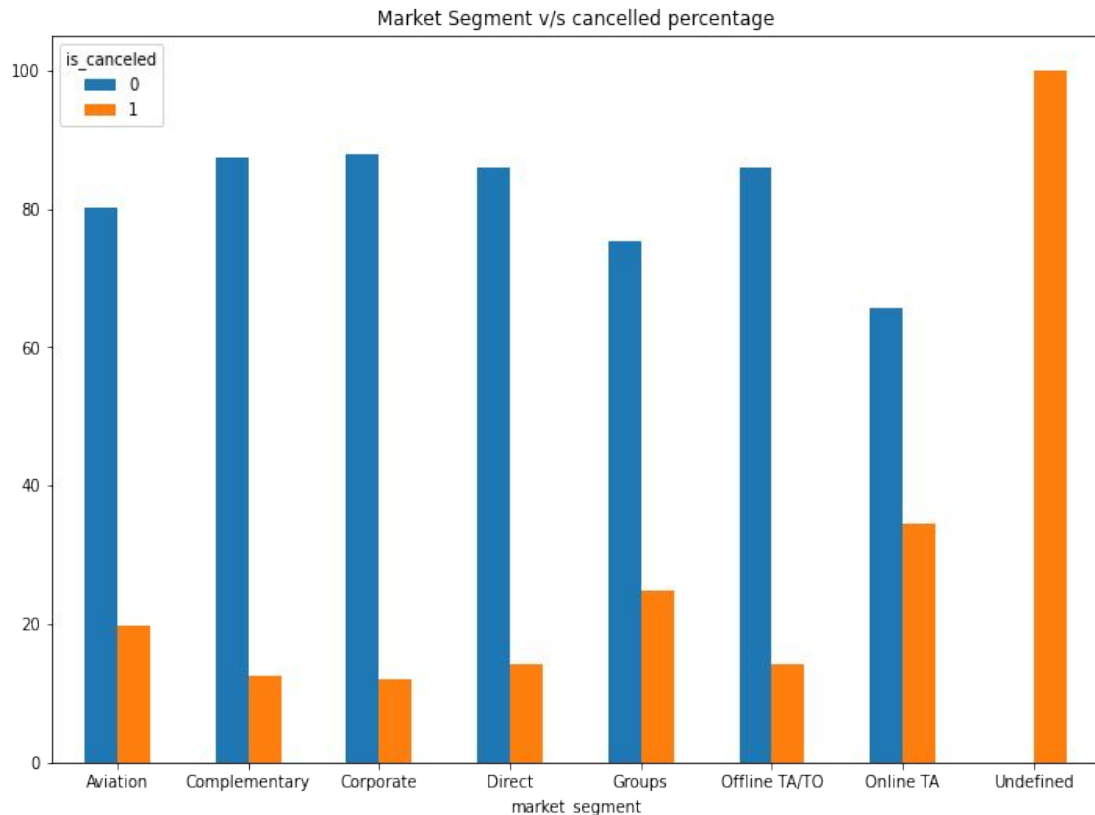
Cancelation and lead time Dependence

- Lead time is a very important feature for hotels it tells the number of days between the time a guest books their room and the time they are scheduled to arrive at the hotel

- Lead time of 0-10 days is most common. We notice if the lead time exceeds 75 days than there is a higher probability that the booking gets canceled


Cancelation And Lead Time Dependence
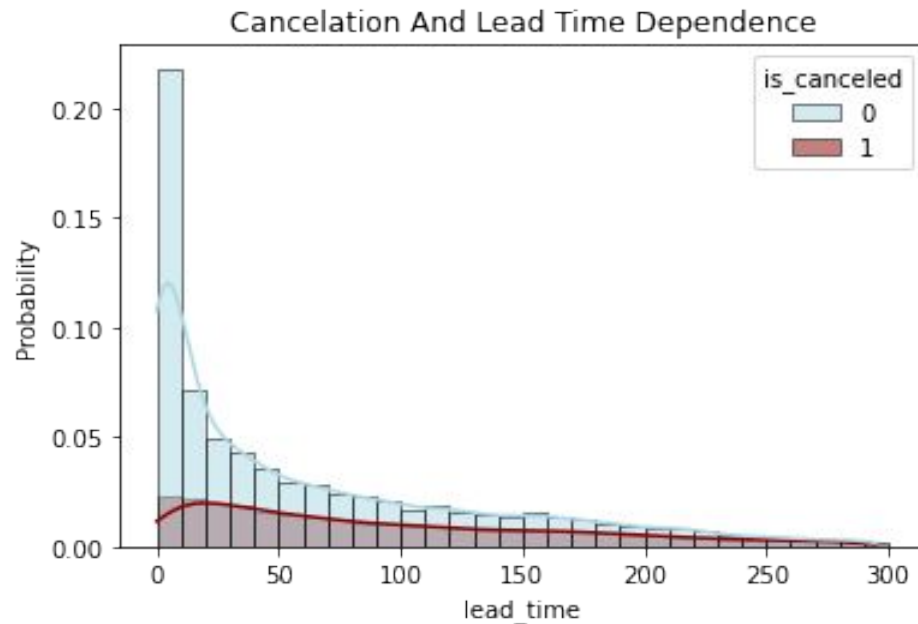
Lead time v/s adr

- More the average daily rate and more the lead time dominance the high probability of cancelation in hotel industries.
- Also if the lead time is greater than 250 days, cancelation rate is high whereas for higher adr cancelation depend on lead time, i.e if the lead time is less than 10 days and adr is greater than 150 then the cancelation rate is very low.
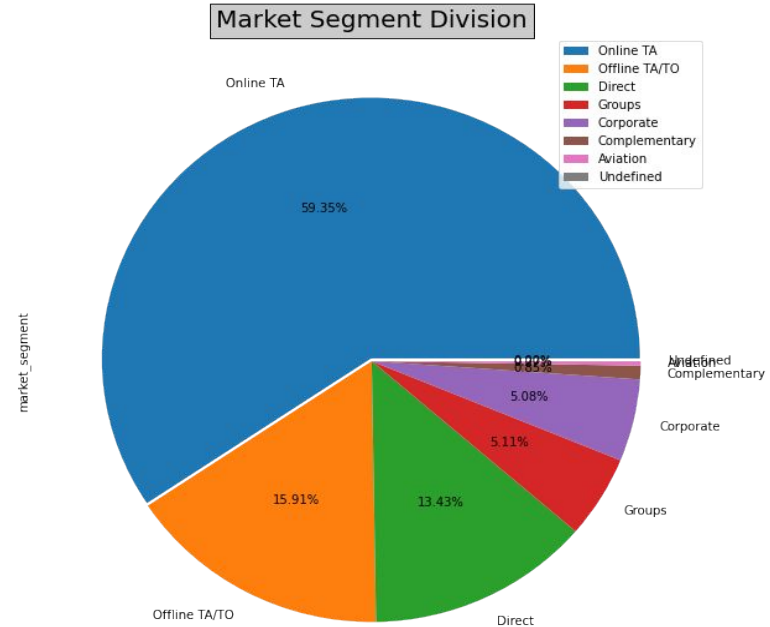- so basically we conclude that cancelation is not affected much by adr

Lead Time V/S Adr
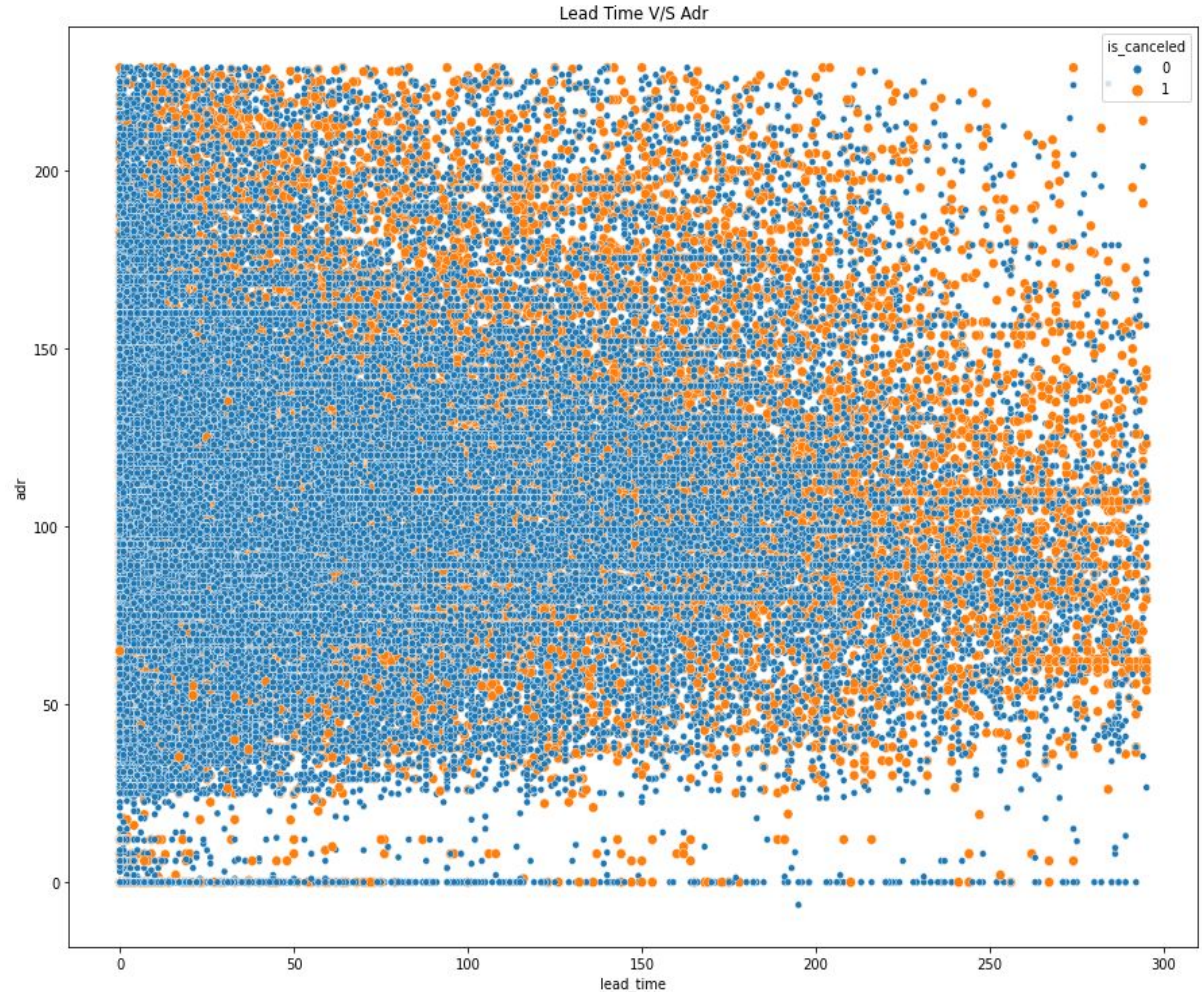
Percentage of cancelation when reserved room is change
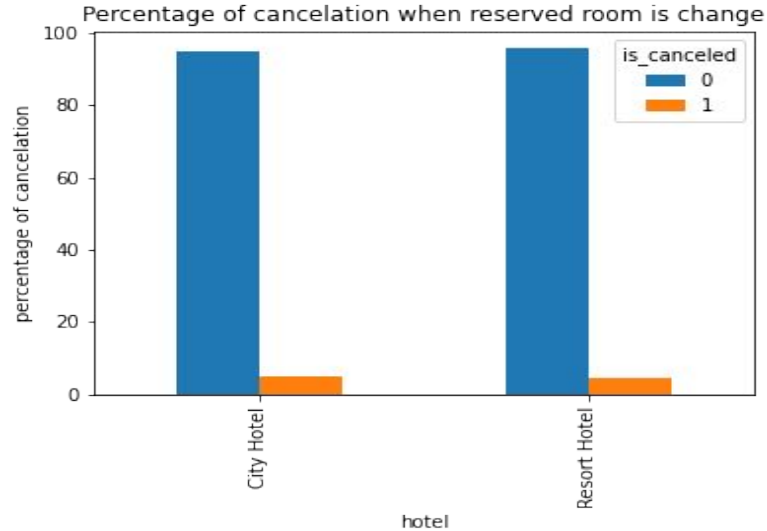
- Cancelation rate on changing reserved room is very low in both city and resort hotels. Its approx 5% in both types of hotels

Market segment wise analysis

- Market_segment is a very important feature for hotels, as it is the way how guests are coming to hotels for booking so hotels must take care of those marketing segments which are proving to be very beneficial for hotels, and invest more on important marketing segments.

- Market_segment of onlineTA is used for nearly 50% of bookings of hotels, so hotels must work on their online adds and must register their hotel in popular online Travel agency for increasing the number of bookings they are receiving

Market Segment Division

Legend:
- Online TA
- Offline TA/TO
- Direct
- Groups
- Corporate
- Complementary
- Aviation
- Undefined

Online TA 59.35%
Offline TA/TO 15.91%
Direct 13.43%
Groups 5.11%
Corporate 5.08%
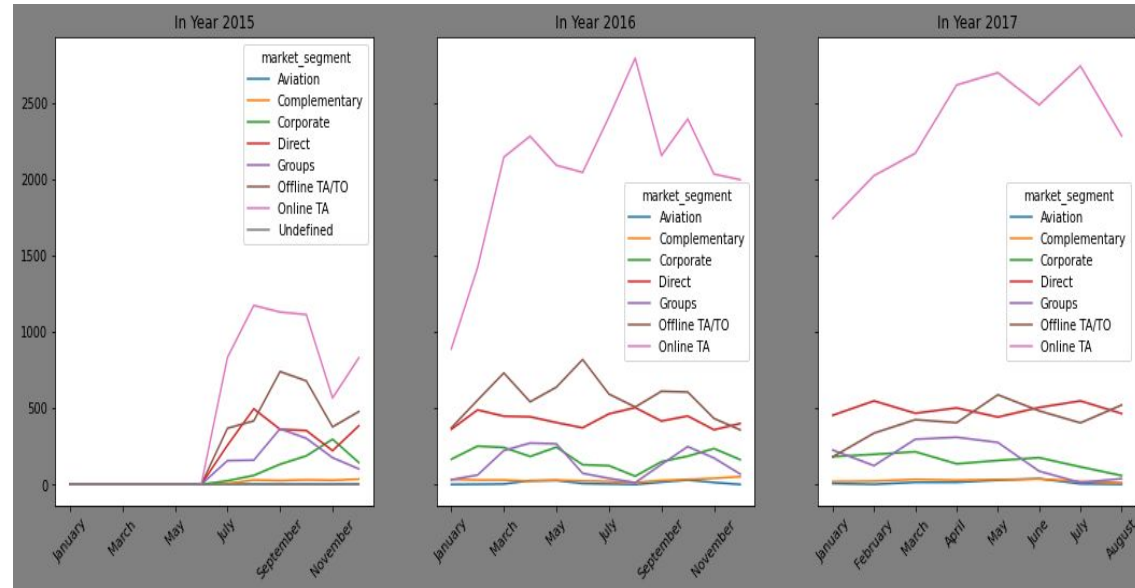Complementary 0.83%
Aviation/Undefined 0.02%

Hotels vs market_segment

- Resorts are having less number of bookings through online TA in comparison to city hotel, so resorts must invest on online TA to receive more number of bookings.

- City hotels must work on their structure and facilities to receive direct bookings as they are more in case of resort hotels
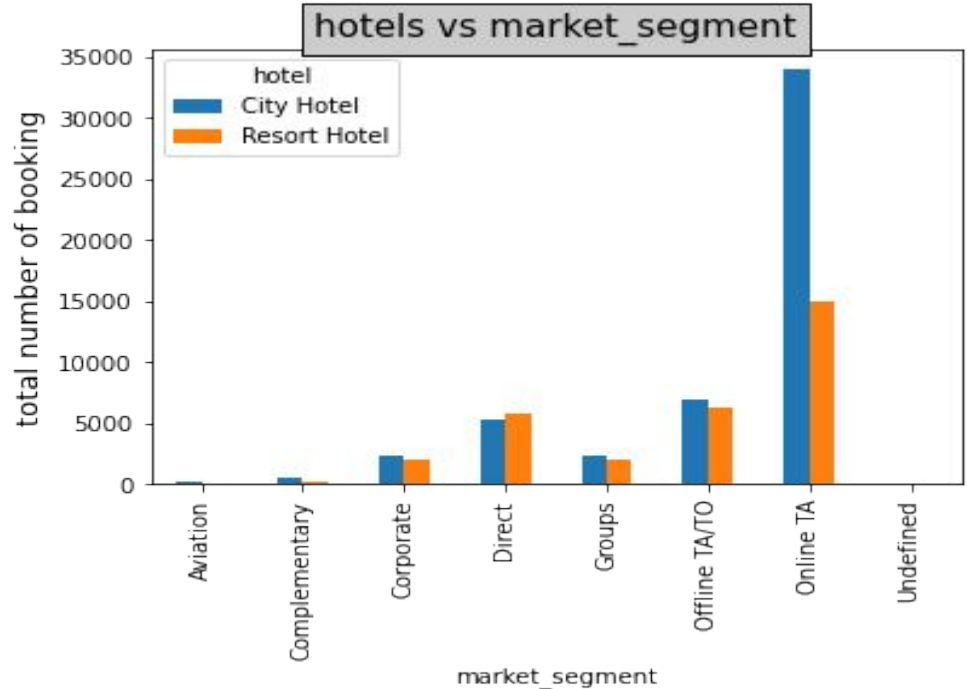
Year wise market segment

- The number of bookings through online TA is maximum since 2015, also we observe a rapid growth in number of bookings from 2015 to 2016 so hotels must invest properly in development of online sources, through adds or may create their own websites and must have contacts with various online Travel Agencies.

- The booking through Offline TA/TO is becoming less effective in preceding years so hotels have to pay less attention to offline sources

- Direct method of booking is nearly constant since last three years.

Retention rate of hotel industry based on market segment

- Retention rate refers to the percentage of customers who continue paying for a product over a given timeframe.

- Complementary market segment has maximum retention rate of 31% followed by corporate 27.81% and Aviation 27%. Whereas other market segment Direct, groups, online offline travel agencies and tour organiser have very low retention rate.
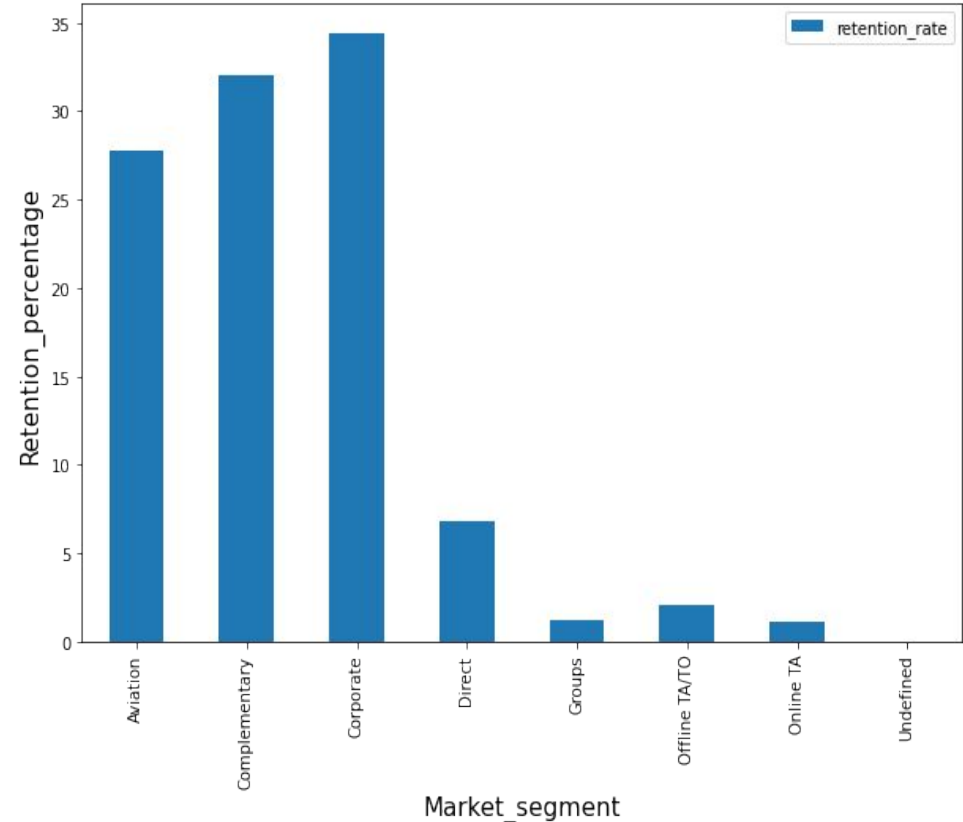
Average daily rate(ADR) based analysis

- ADR is average daily rate, is a statistical unit that often used in logging industry. ADR tells the ability of hotels to maximize the revenue by managing the inventory

- ADR = (Room Revenue/Room Sold).

- City Hotel generates more revenue than resort hotel

Comparison adr between Resort Hotel and City Hotel
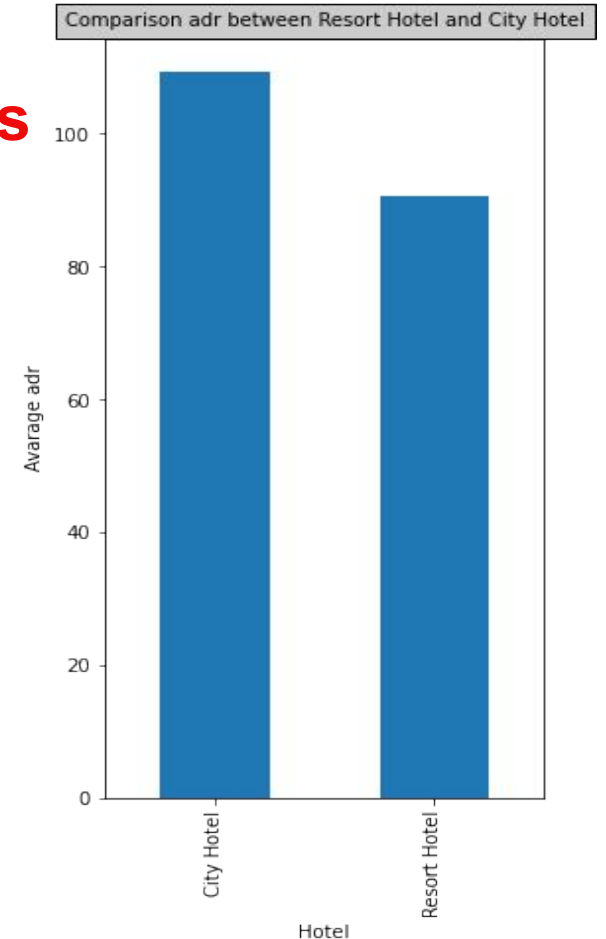
Comparison adr between Resort Hotel and City Hotel on the basis of month

- City hotel generally collect 20-30% more revenue from the resort hotel
- Resort hotel adr start increases in the month of june to september and maximum revenue collected in the month of august which is 14.70% greater than the city hotel.
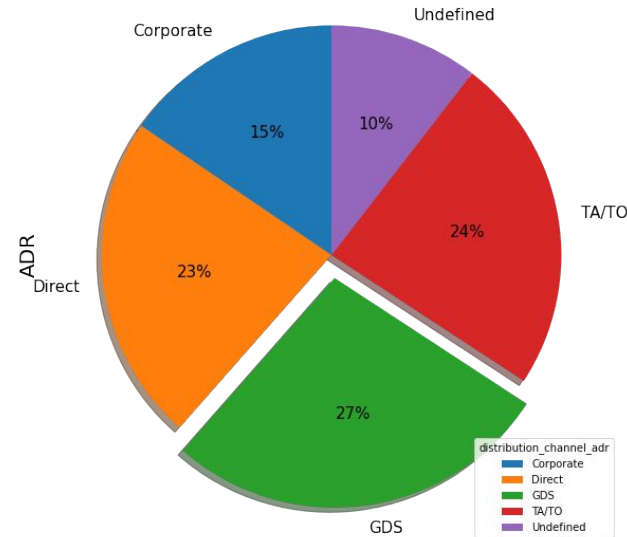

Comparison adr between Resort Hotel and City Hotel on the basis of month
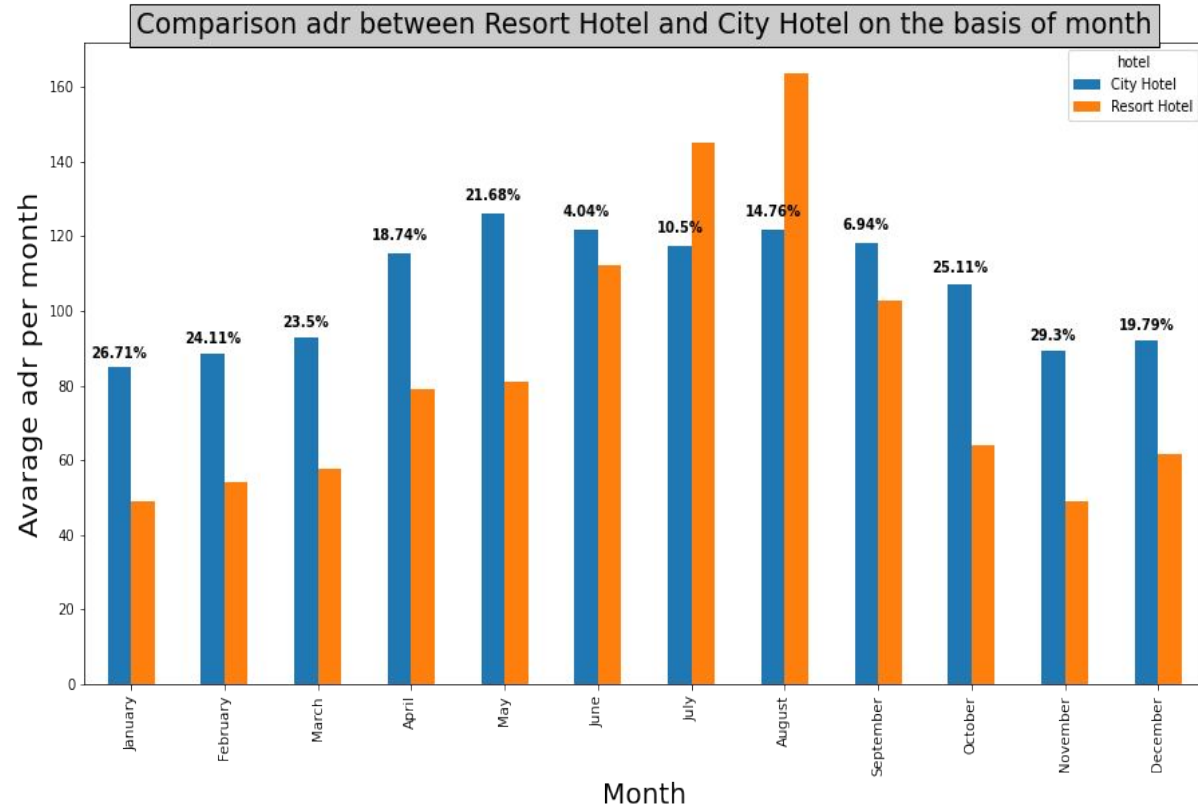
Analysis based on ADR


Market_Segment_ADR
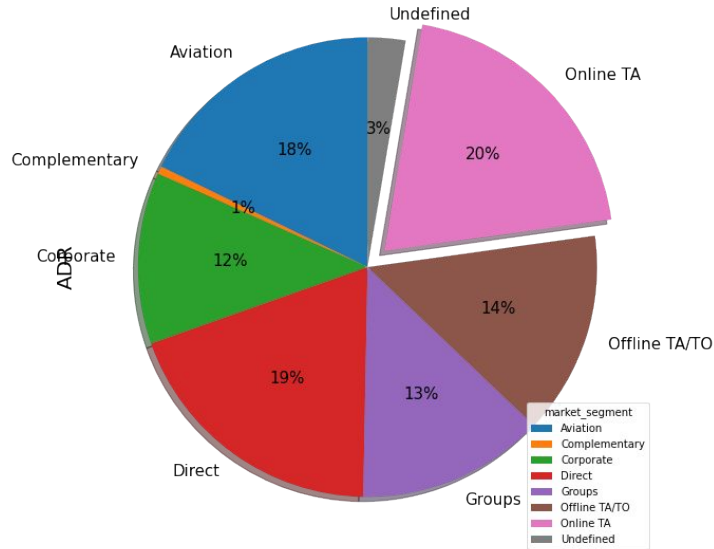
Distribution_Channel_ADR

- From global distribution system(GDS) channel hotels generate maximum profit, followed by travel agencies and tour organiser
- About 39% of revenue generated by the hotels from the online travel agency and direct market segment.

Guest_wise_adr

- Increase in number of guest leads to increase the ADR in hotels industry.
- Here 0 represent that the number of guest who booked a hotel room but not check in.

Guest_wise_ADR
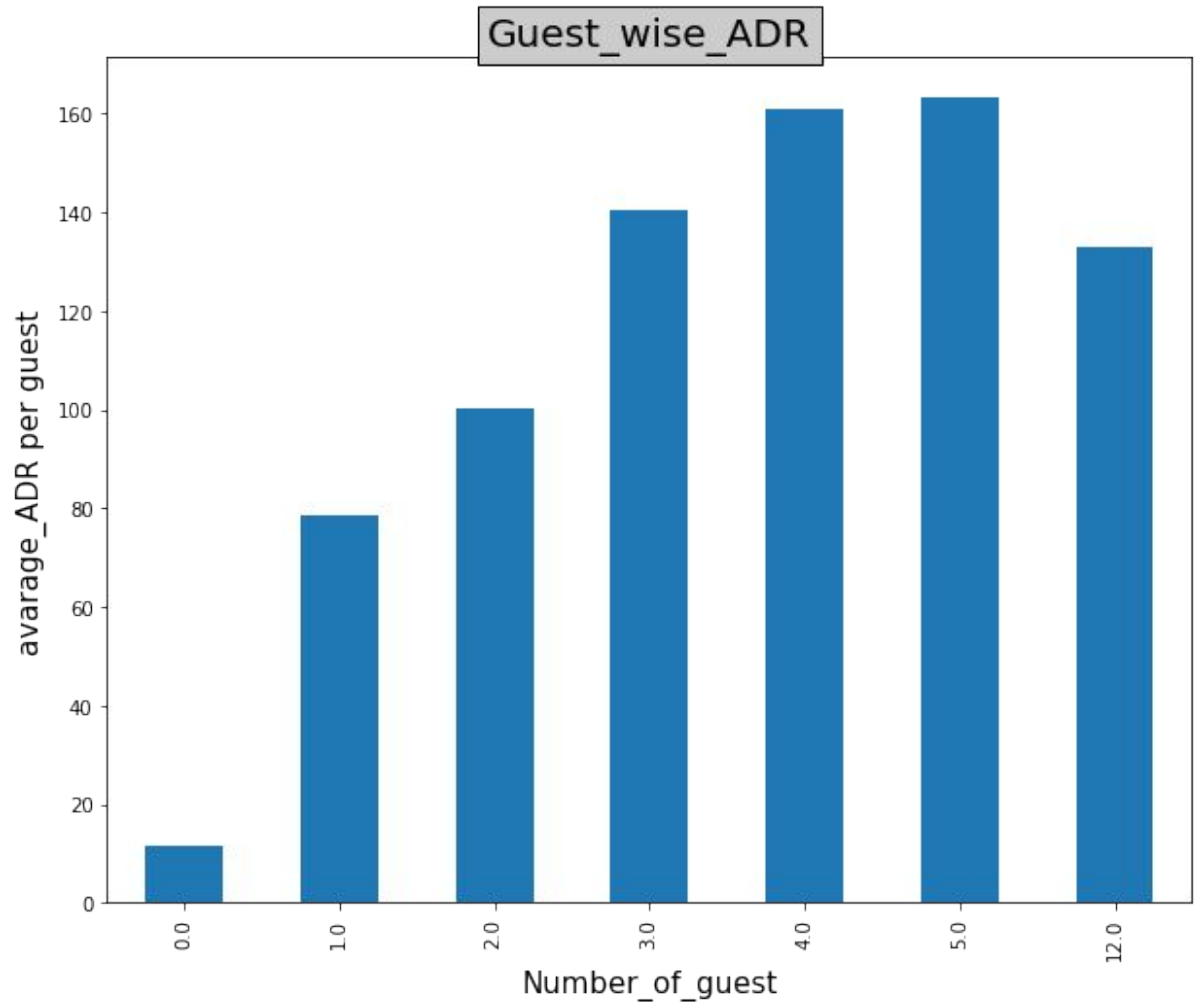
Challenges

- Knowledge of logging industries was not up to marks, due to which found difficulty in some of terminologies.
- Some plots didn't come up as planned.
- There was some missing data and outliers, which had to be interpreted well.

Conclusion

- Hotels are more booked in the month of April and May.

- Based on the above analysis cancelation is generally depends on lead time. Cancelation rate is very low when the reservation room is changed.

- This project analysis the important market segment for hotels and also visualised the retention rate of each market segment.

Conclusion

- This project compare the average daily rate with different market segment and different distribution channel.

- Generally parking space is required for resort hotel.

- Optimal stay in both hotels is less than 7 days. For stay more than a week people likes to stay in resort hotels

Thank you