EDA CAPSTONE PROJECT

HOTEL BOOKING ANALYSIS

BY:- AJAHAR DAROGA

PROBLEM STATEMENT:

- **Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!**

- **This data set contains booking information for a city hotel and a resort hotel, and includes 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. All personally identifying information has been removed from the data.**

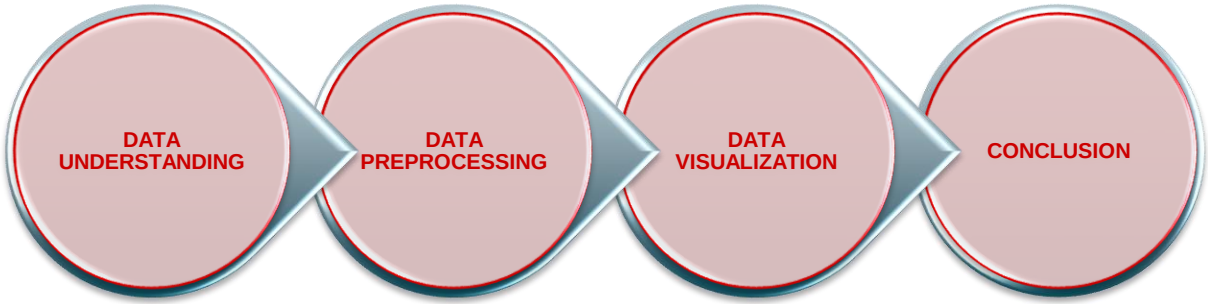
- **Explore and analyze the data to discover important factors that govern the bookings**

PROJECT AIM:

- Aim of this project to find out insight from given dataset to figure out below questions like:

  - The best time of year to book a hotel room.
  - The optimal length of stay in order to get the best daily rate.
  - Predict whether or not a hotel was likely to receive a disproportionately high number of special requests.

METHODOLOGY:

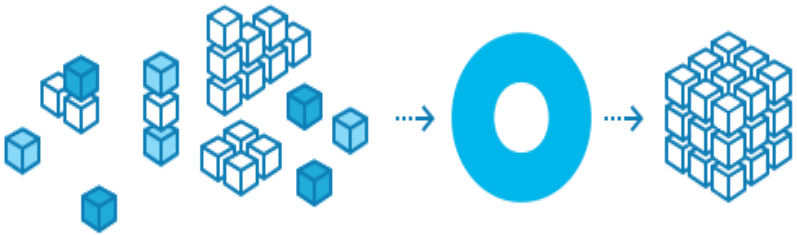


DATA UNDERSTANDING:

Sr. No	Features	Sr. No	Features
1	hotel	17	is_repeated_guest
2	is_canceled	18	previous_cancellations
3	lead_time	19	previous_bookings_not_
4	arrival_date_year	20	reserved_room_type
5	arrival_date_month	21	assigned_room_type
6	arrival_date_week_number	22	booking_changes
7	arrival_date_day_of_month	23	deposit_type
8	stays_in_weekend_nights	24	agent
9	stays_in_week_nights	25	company
10	adults	26	days_in_waiting_list
11	children	27	customer_type
12	babies	28	adr
13	meal	29	required_car_parking_s
14	country	30	total_of_special_reques
15	market_segment	31	reservation_status
16	distribution_channel	32	reservation_status_date

- **The given dataset has 119390 records and 32 features.**
- **Dataset have mixture of data type.**
- **Features like country, agent, company, children contain missing values.**

DATA PREPROCESSING:



- Dropping the feature which have more than 15% of data is missing (i.e. company).
- Missing value in dataset is impute with mean, mode, median. (i.e. NAN values).
- Converting the required categorical feature using dummy.
- Removing the outliers (adr)

DATA VISAULIZATION:

- **Visualization the dataset using**

1. Univariate Analysis

i. Histogram

ii. Box plot

iii. Count plot

2. Multivariate Analysis

a) Categorical variable

i. Bar plot

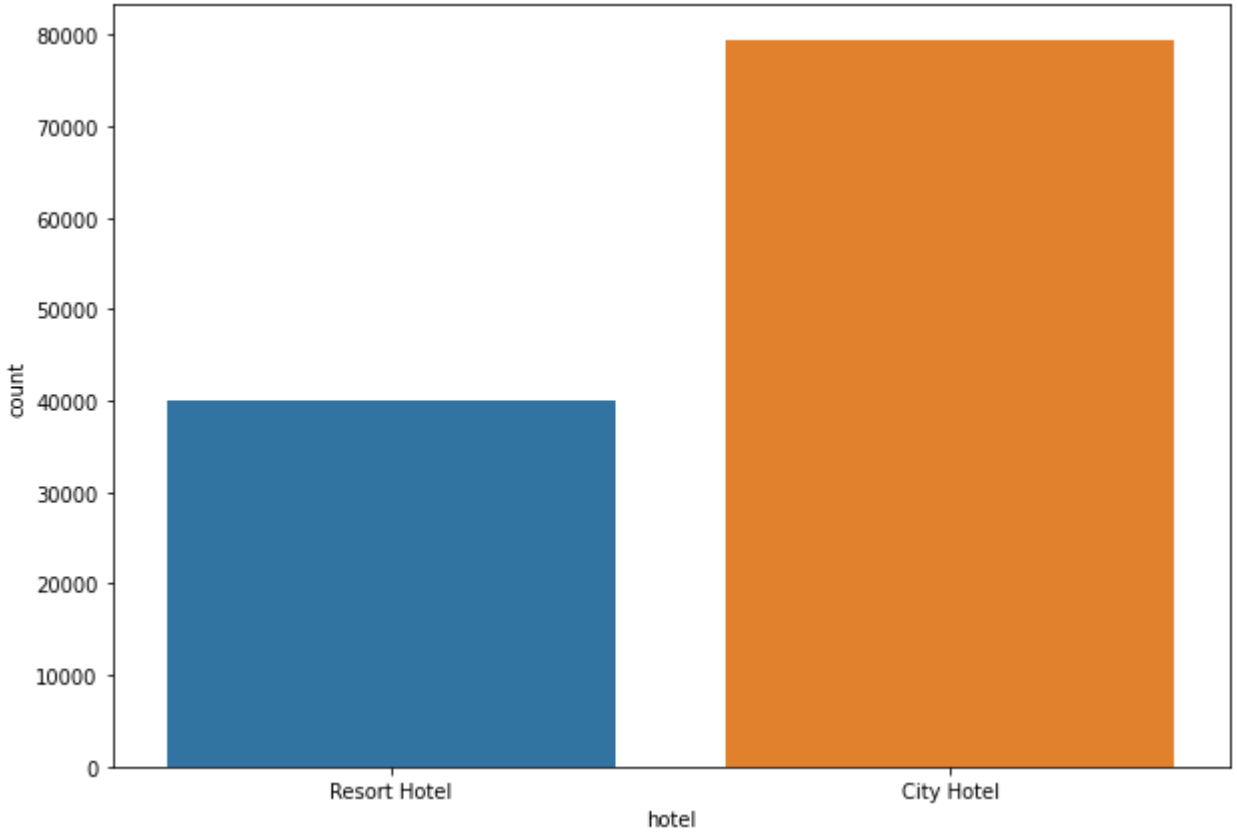
b) Continuous variable

i. Scatter plot

ii. Line plot

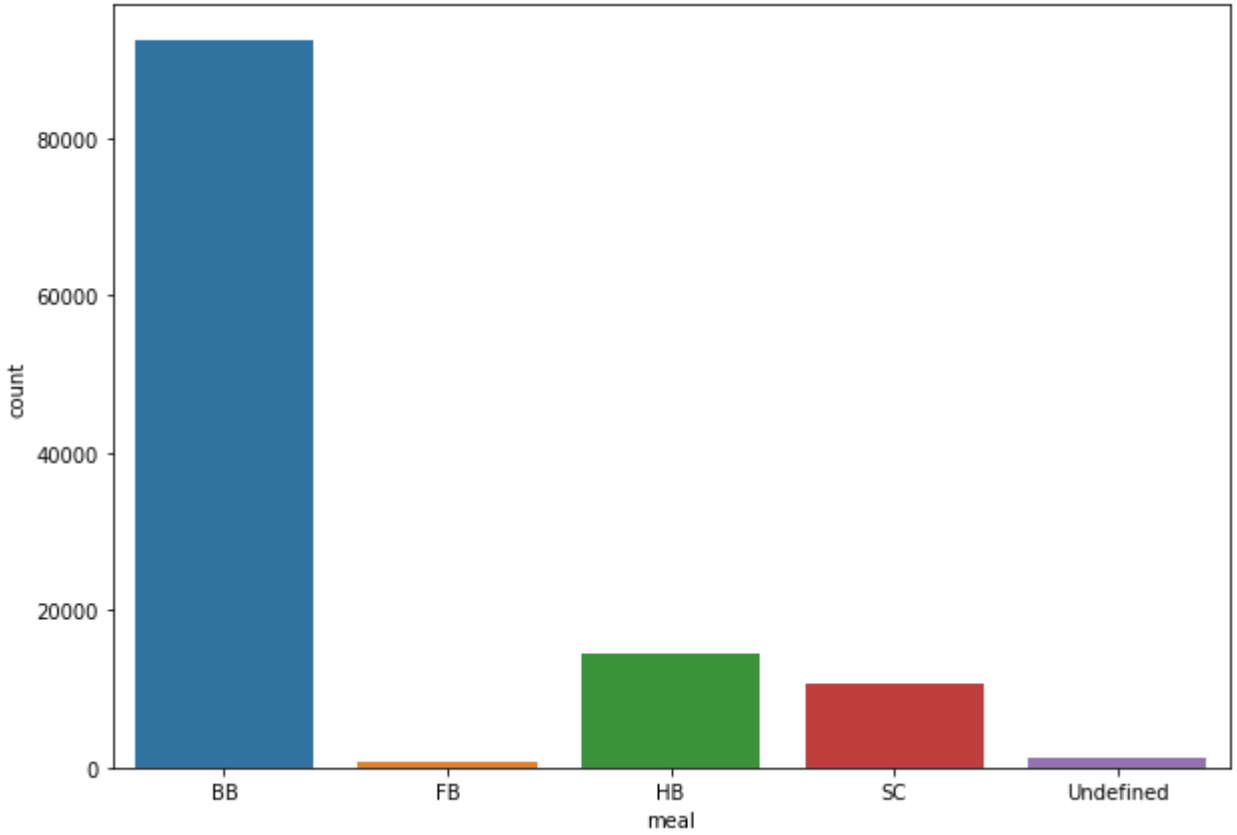
HOTEL

- Hotel Type
- Resort Hotel customer: 40060
- City Hotel customer : 79330



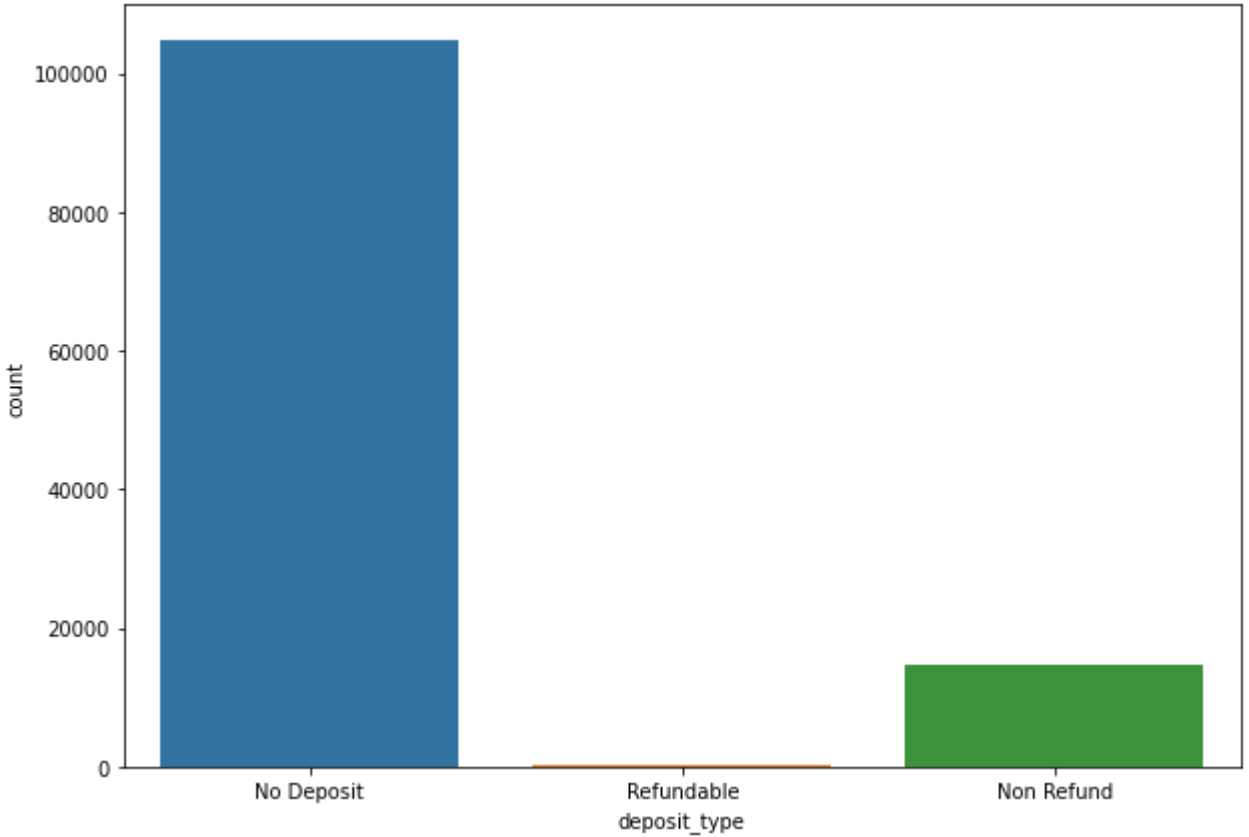
MEAL

- Based on meal type customer booking the hotel.
- Meal type BB like more number of customer compare to other meal type.



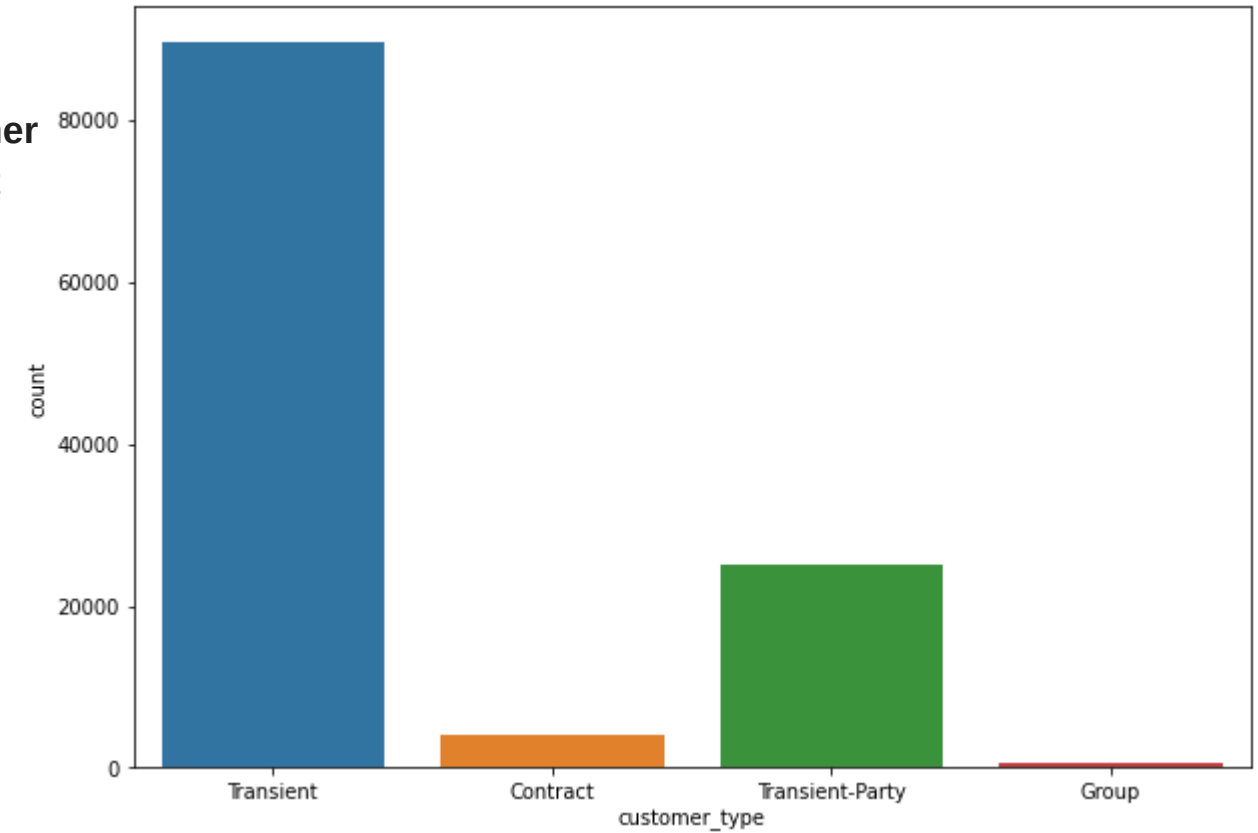
DEPOSIT TYPE

- More number of customer prefer the no deposit during the booking the hotel.



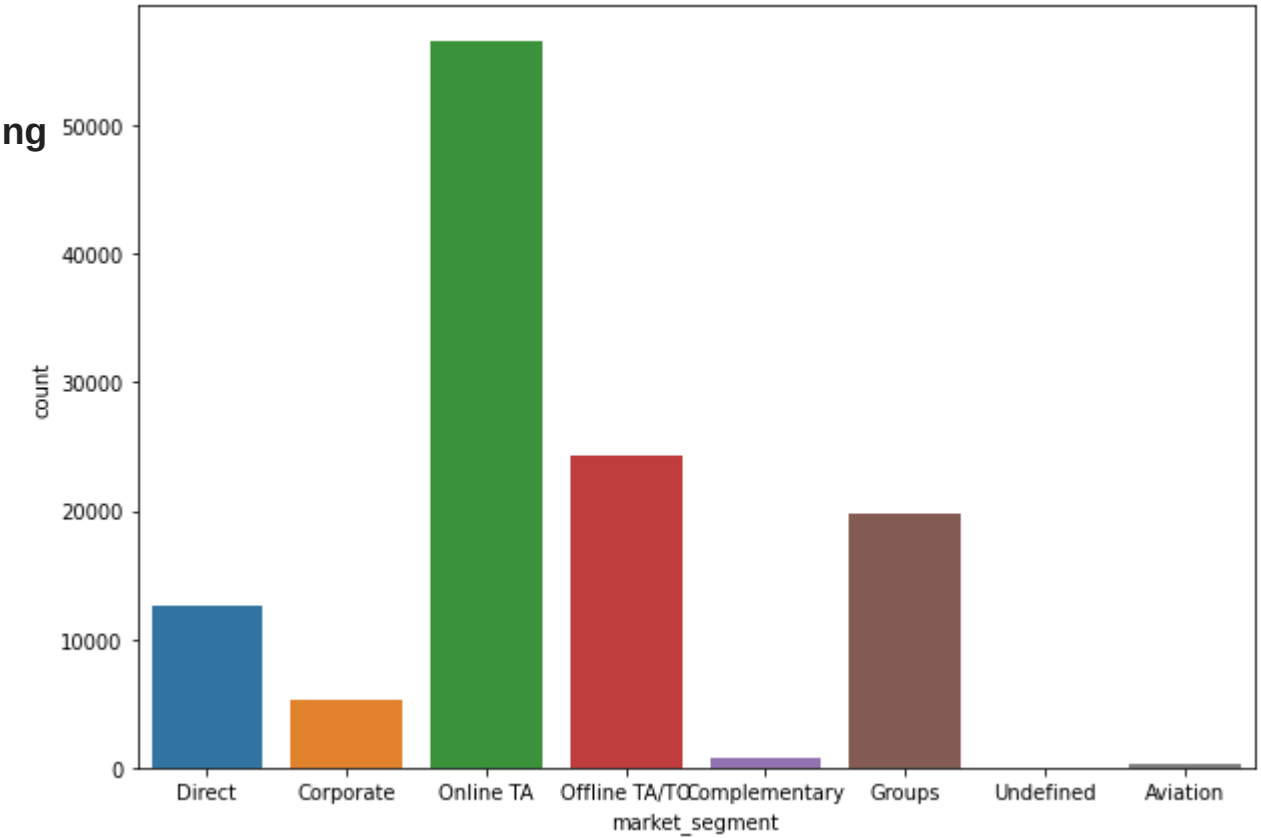
CUSTOMER TYPE

- More number of customer belong to the Transient type.
- Very less in the group.



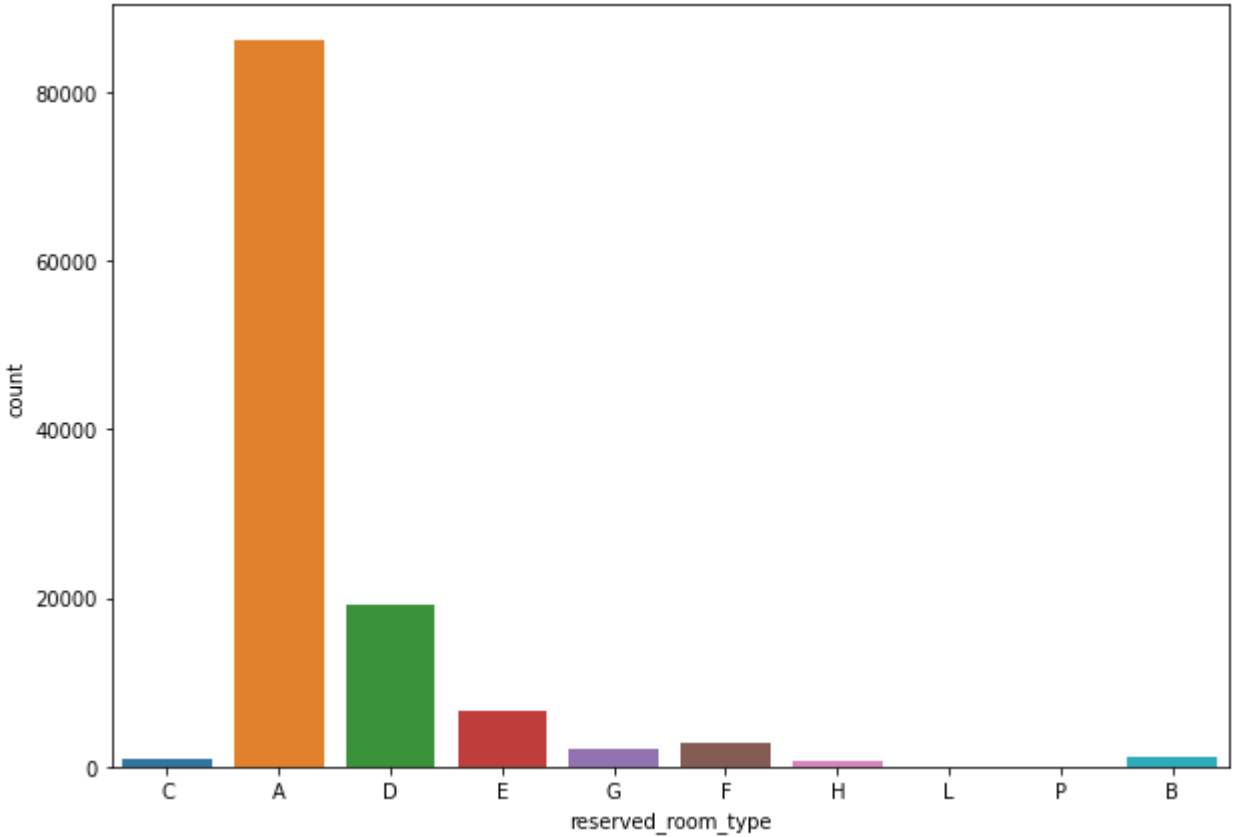
MARKET SEGMENT

- **More customer are booking hotel by using online platform**.



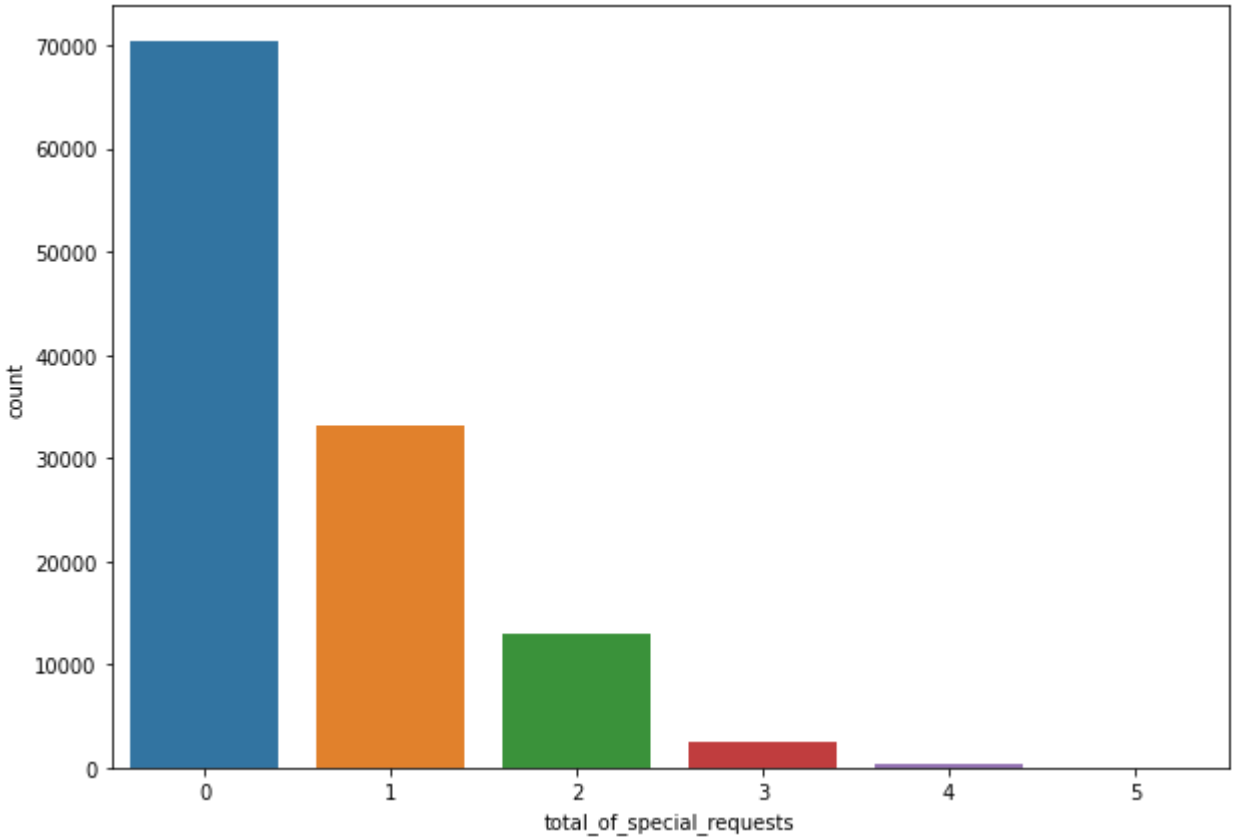
RESERVED ROOM TYPE

- **More customer are booking hotel A category room type**.



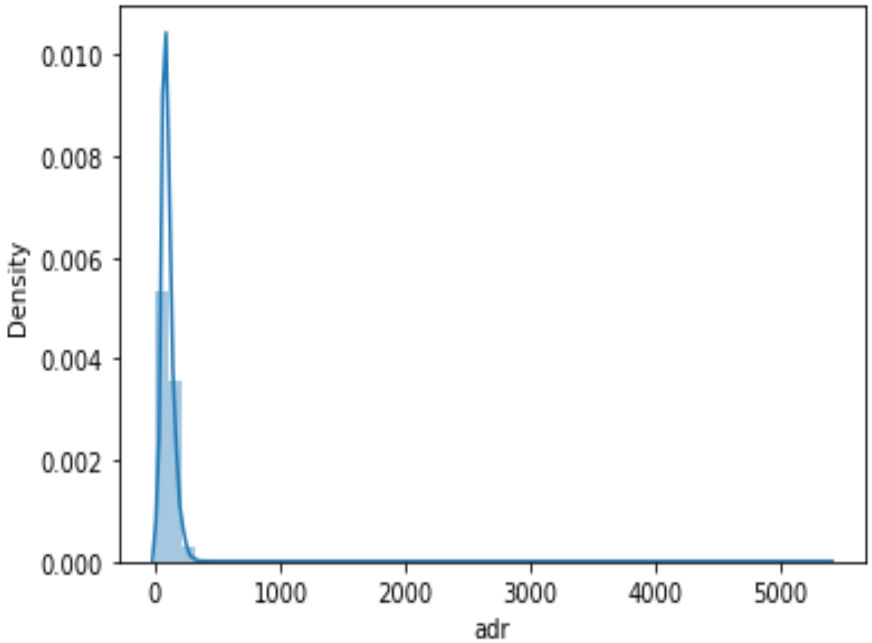
SPECIAL REQUEST

- **Number of booking without special request.**



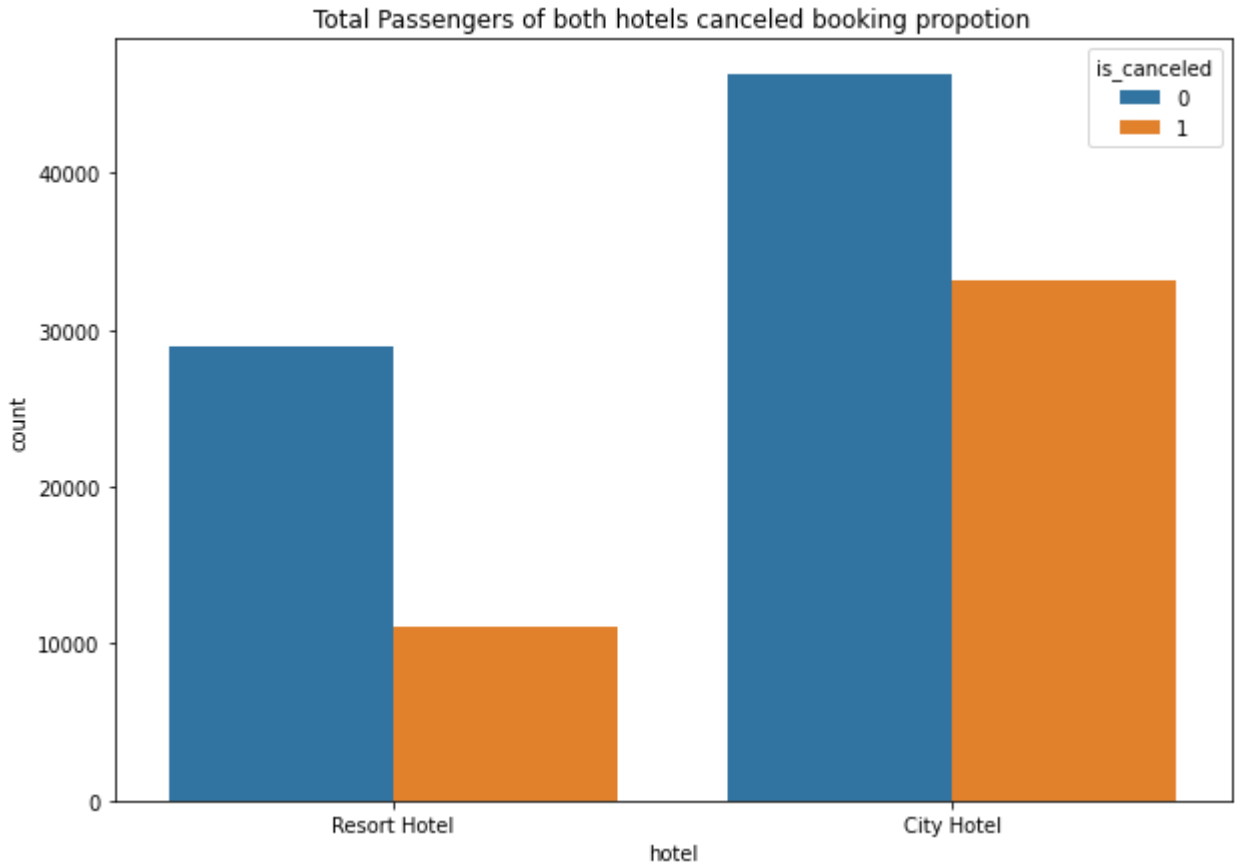
adr

- The average adr is 101 and its varies in 50 to 150.
- Contains the outlier.



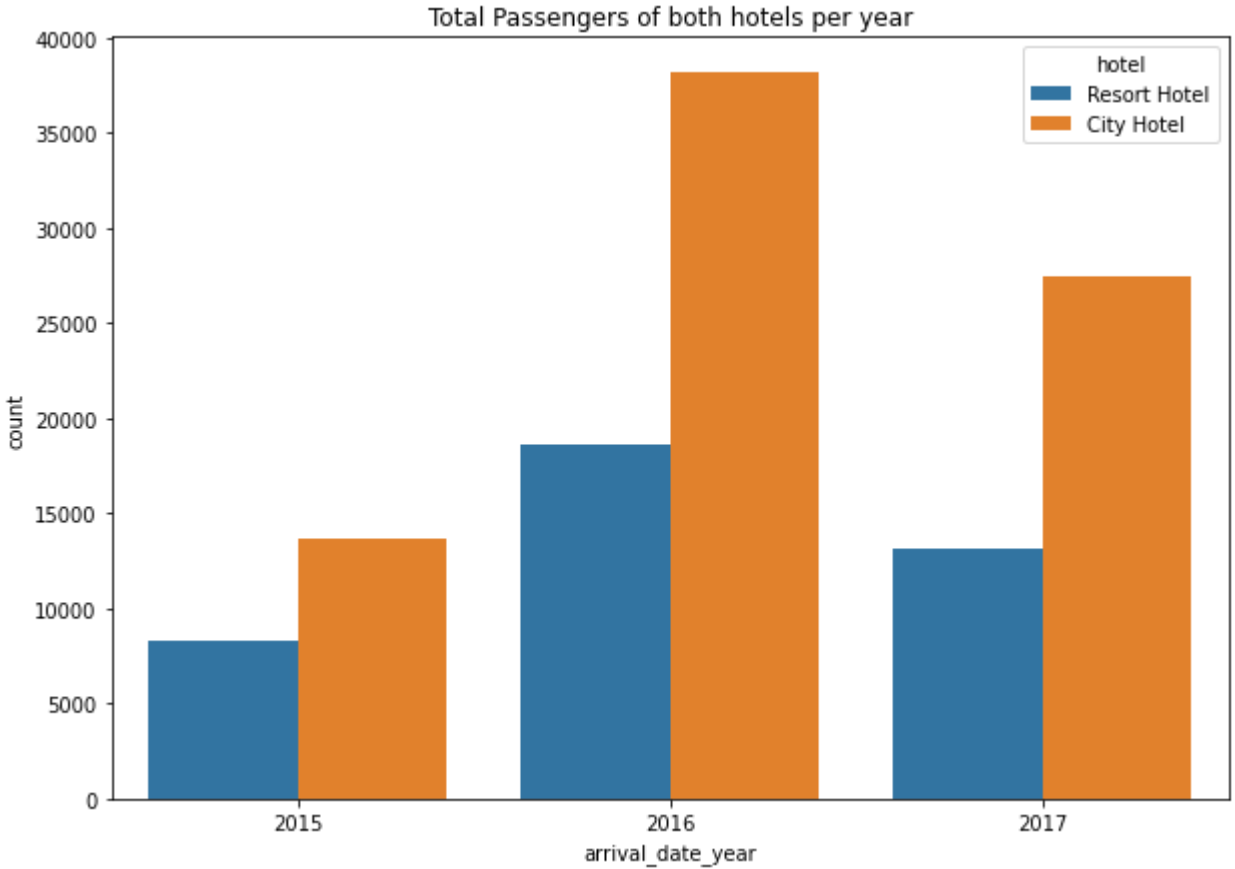
BOOKING CANCEL BASED ON TYPE OF HOTEL

- More number of booking cancel in City hotel type.



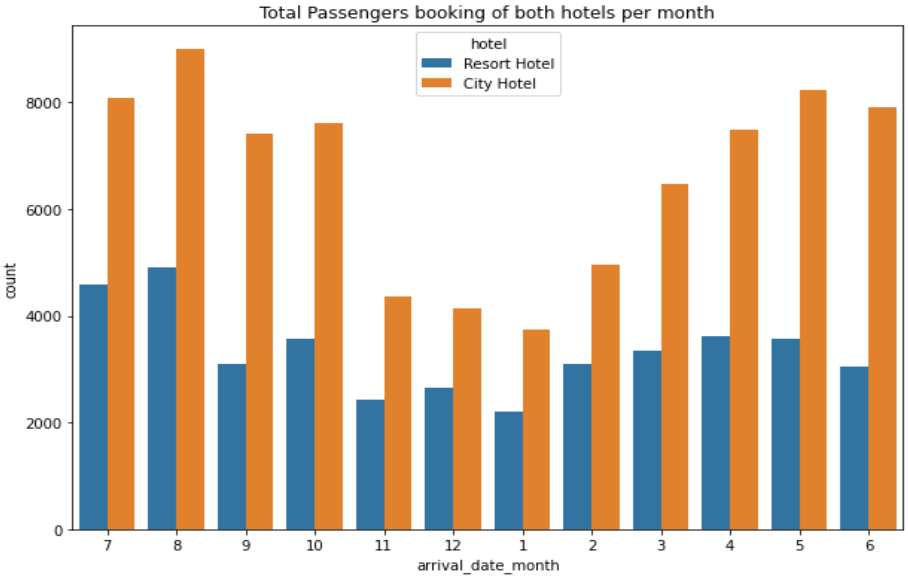
HOTEL BOOKING BASED ON YEAR

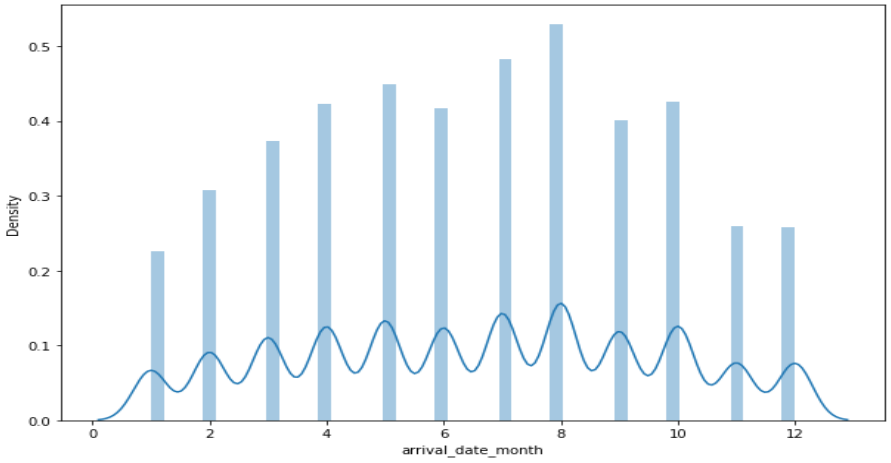
- **Proportional of booking based on year wise.**



HOTEL BOOKING BASED ON MONTH

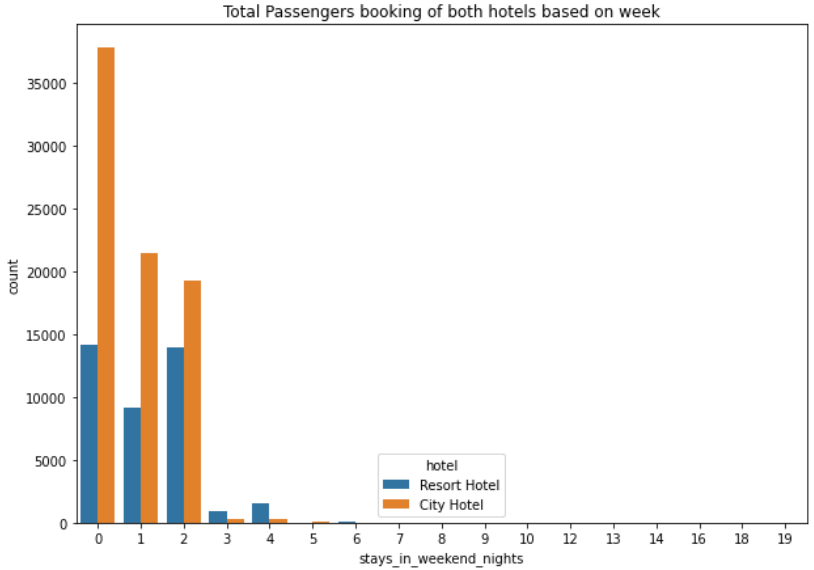
- We can decide after seeing this data less number of booking in Nov, Dec, Jan month.
- So customer get more discount this period.

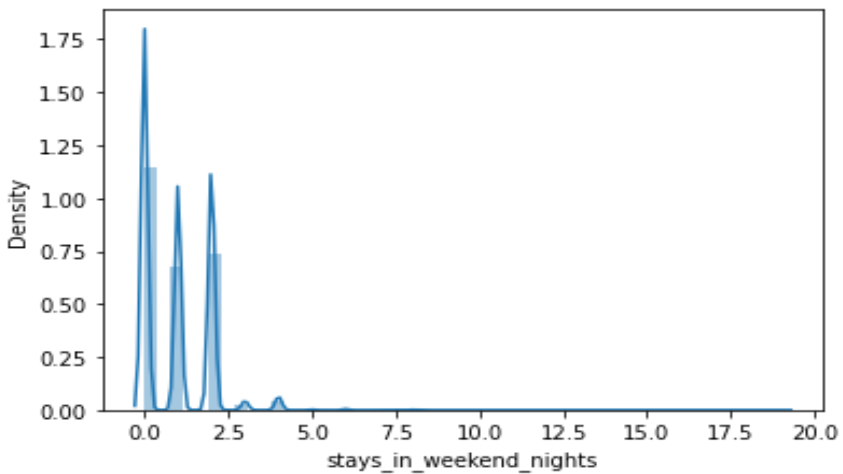




LENGTH OF STAY IN HOTEL ON WEEKEND

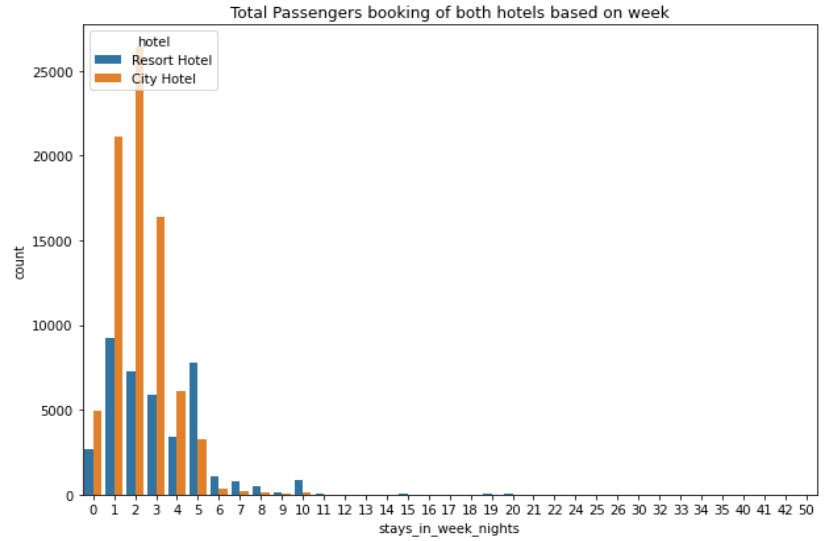
- Optimum length for stay in hotel is more than 2 days to get best daily rate.
- Applicable for both type of hotel.

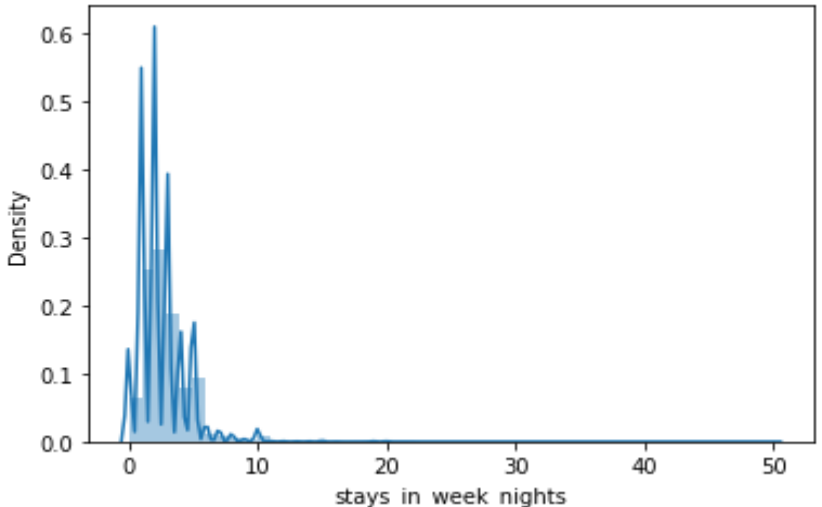




LENGTH OF STAY IN HOTEL ON WEEK

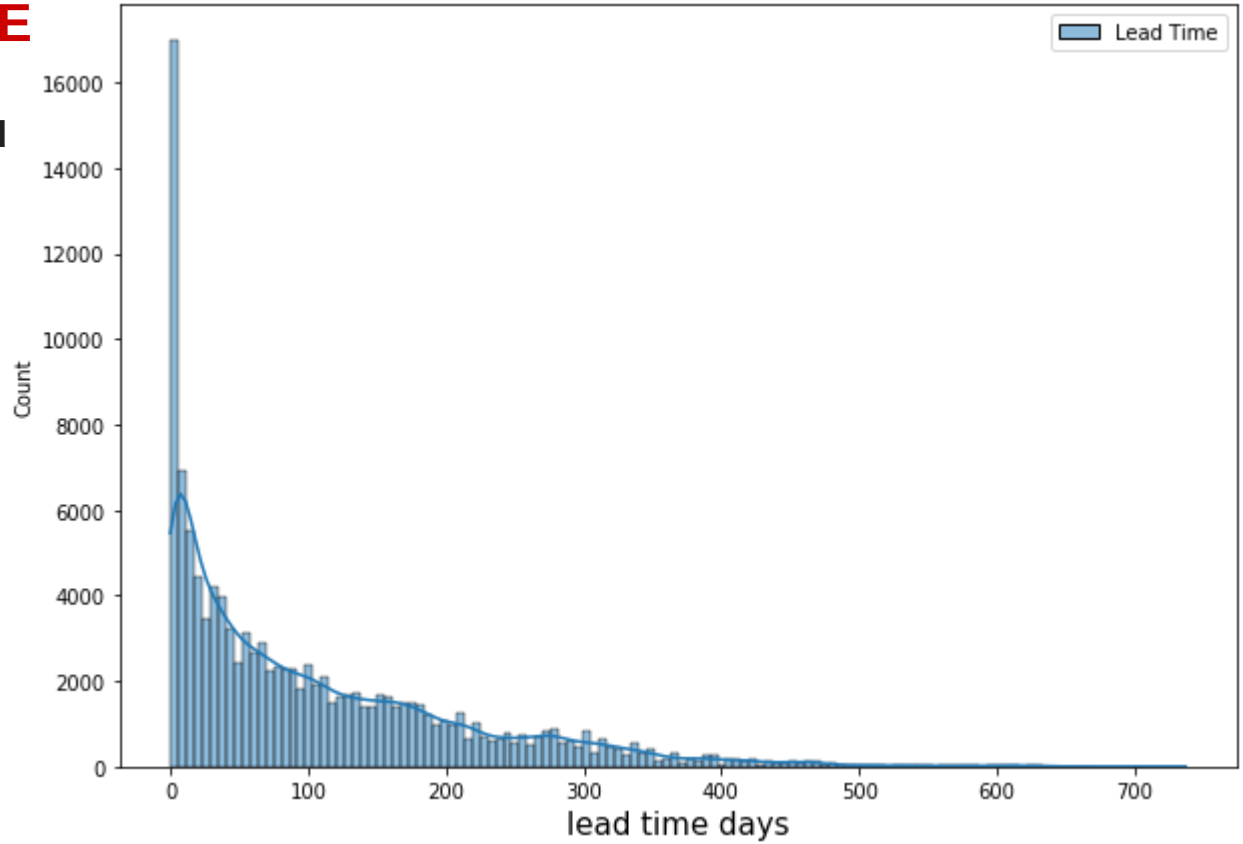
- Optimum length for stay in hotel is more than 6 days to get best daily rate.
- Customer stay more days in Resort hotel compare to City hotel.





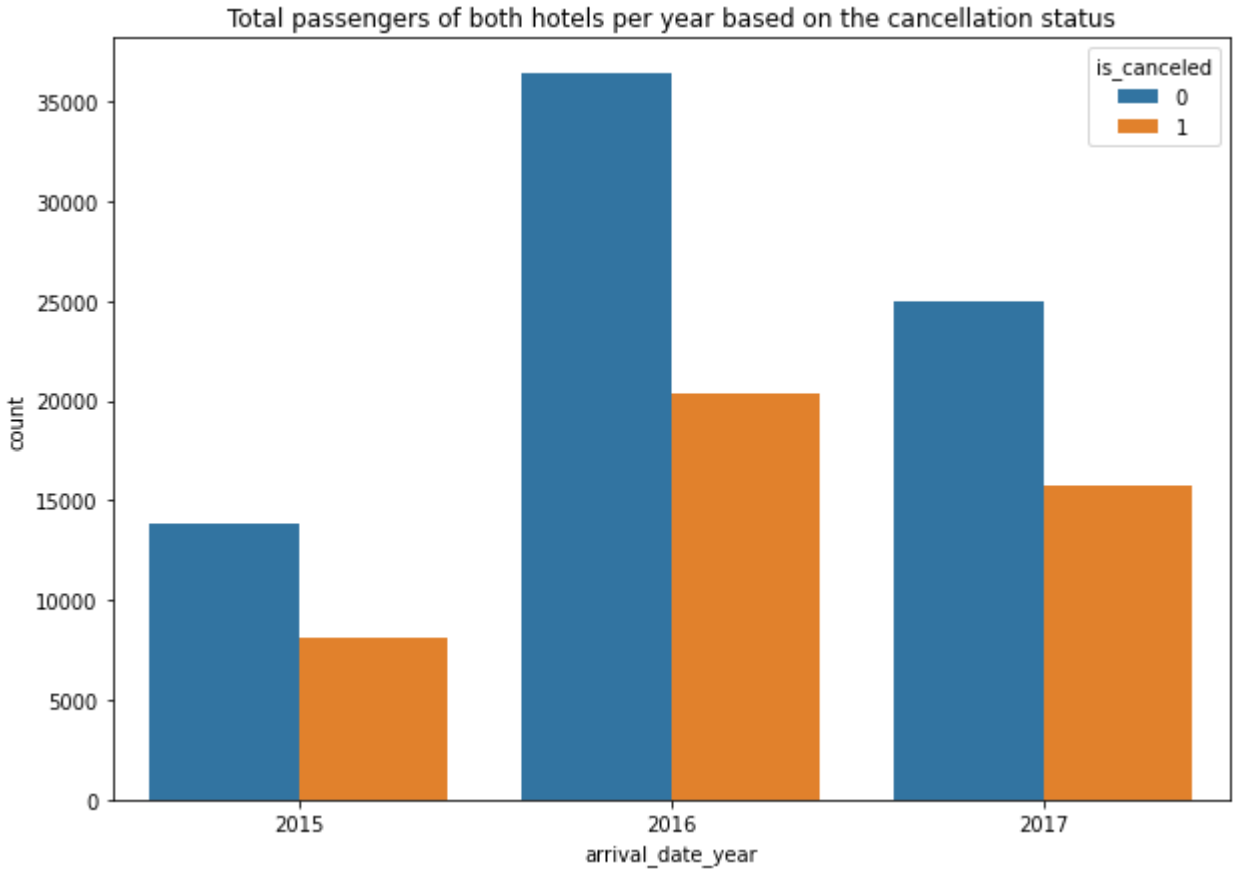
CANCELATION OF HOTEL BASED ON LEAT TIME

- **We can easily understand from graph lead time increase the count of bookings reduces.**



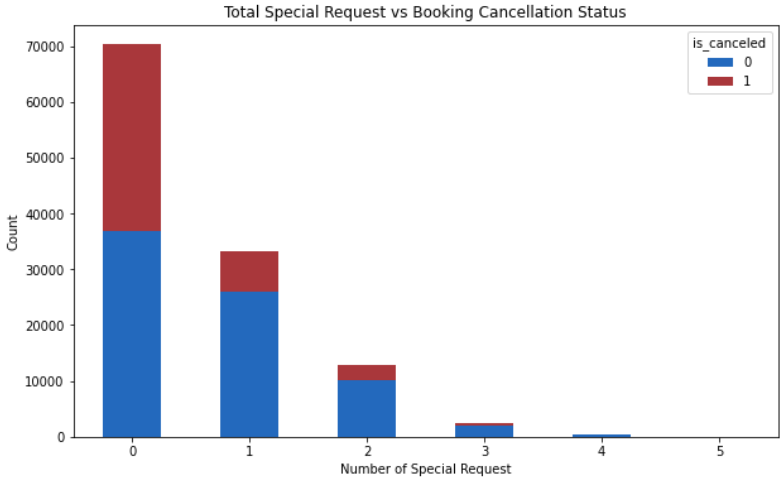
CANCELATION OF HOTEL BASED ON YEAR

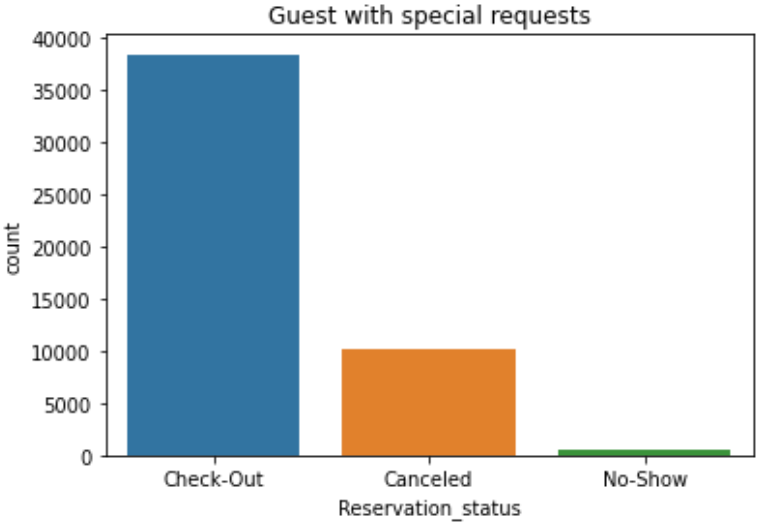
- More than 50% of hotel booking is cancel every year.

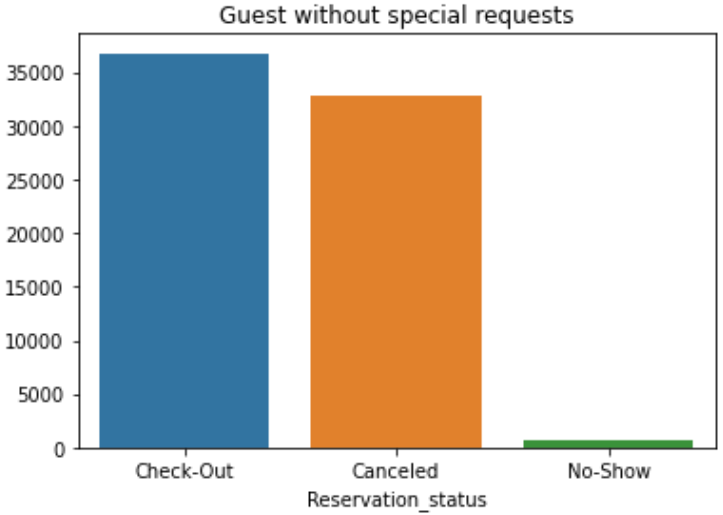


HOTEL BOOKING BASED ON SPECIAL REQUEST

- **Sending Special request to customer get increased the hotel booking.**

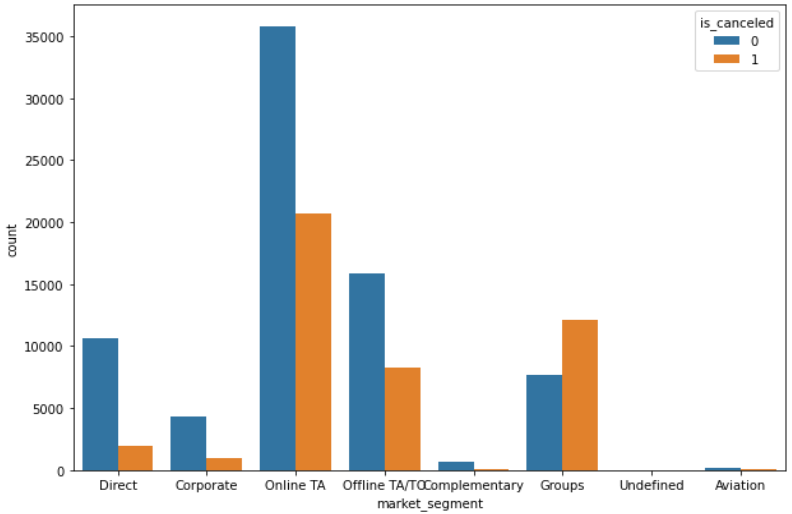


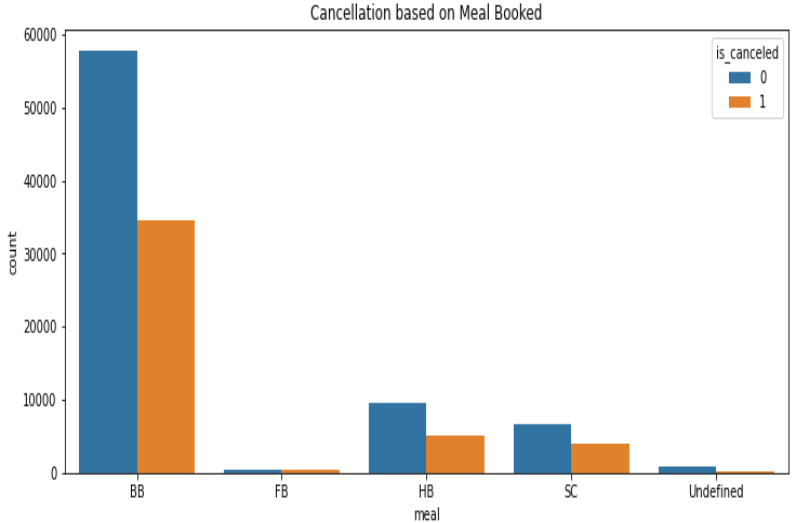




HOTEL BOOKING CANCELATION BASED ON MARKET SEGMENT & MEAL

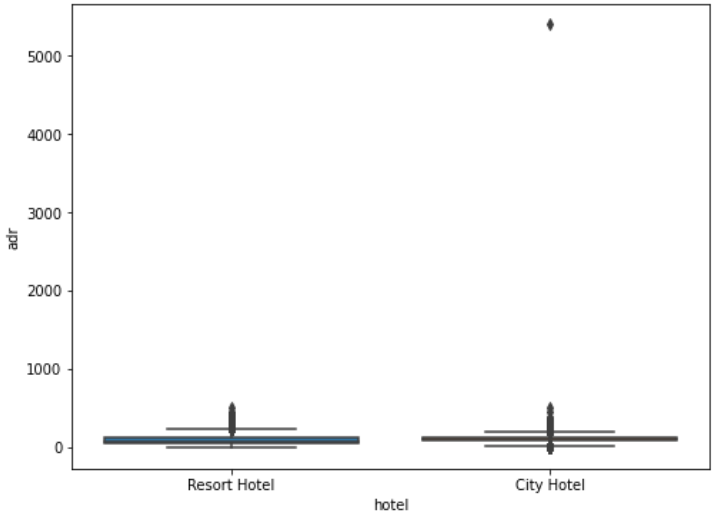
- More number of booking cancel in Group market segment.

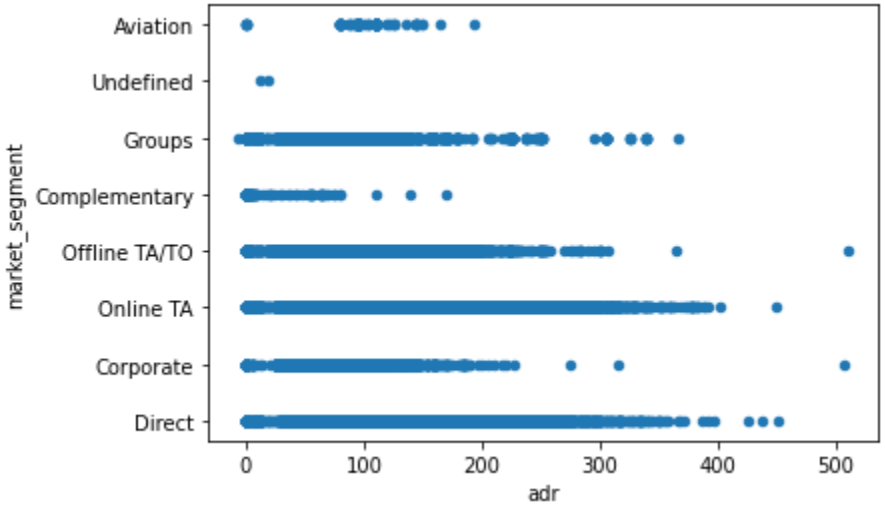


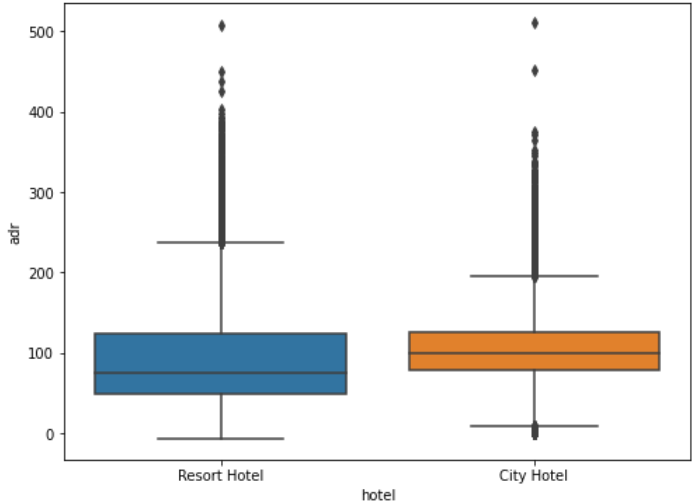


DISTRIBUTION OF ADR BASED ON HOTEL TYPE

- Almost same average rate is applied on both hotel type.
- Slightly higher side on Resort Hotel.

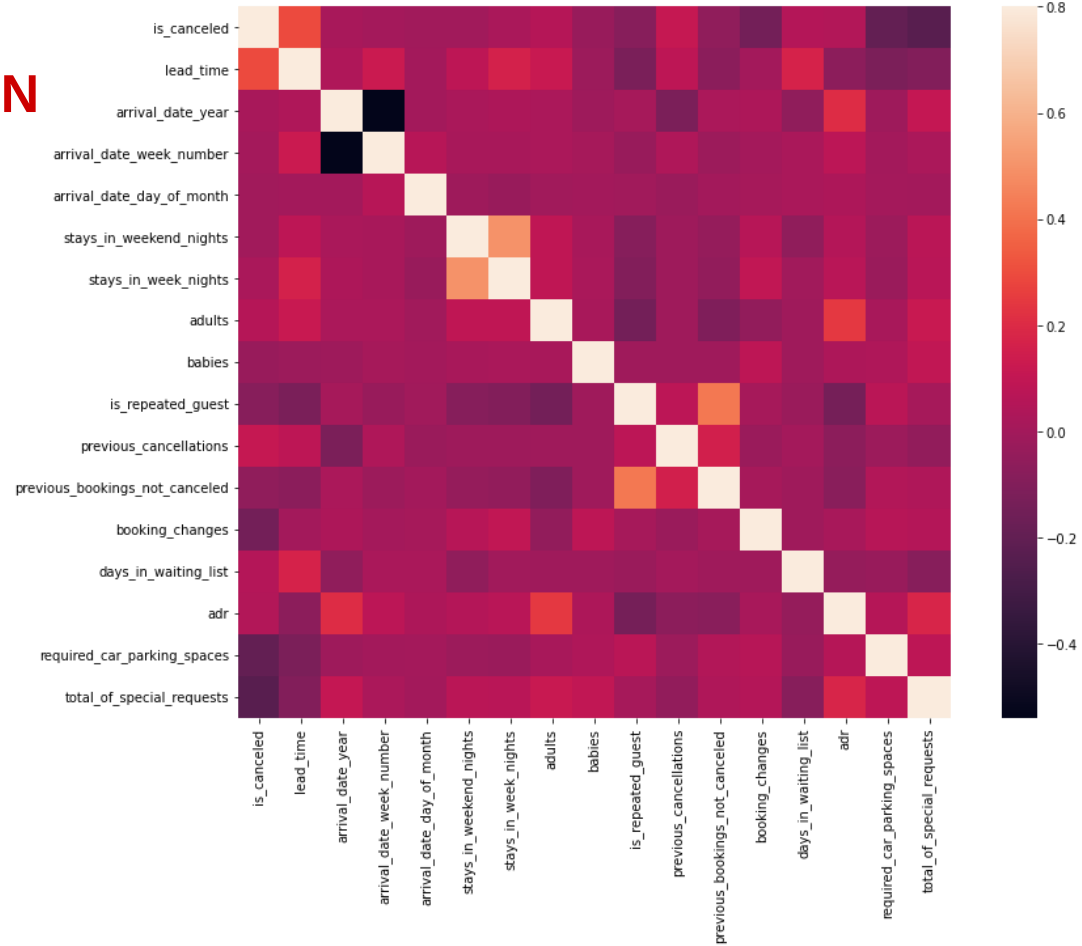






CORRELATION BETWEEN THE FEATURES

- Good relation between two feature using heatmap are below
  1. stay in week & Lead time
  2. adult & adr



PAIRE PLOT

- **Better understanding relation between the features.**

CONCLUSION

FEATURE SPECIFIC

- **Hotel Type :-** City Hotel
- **Market Segment :-** Online
- **Deposit Type :-** No deposit
- **Meal Type :-** BB
- **Customer Type :-** Transient
- **Resaved Room Type :-** A Type
- **Special Requests :-** 1
- **Lead Time :-** as less as possible
- **Customer :-** Adults

COMMENTING ON OVERALL DATASET

- **Booking month to get best daily rate on Nov, Dec, Jan.**
- **Optimum length for better daily rate on weekend more than 2 days.**
- **Optimum length for better daily rate on week more than 6 days.**
- **Customer stay more days in Resort hotel compare to City hotel.**
- **More number of booking cancel in City hotel.**
- **More than 50% of hotel booking cancel in every year.**
- **Sending special request increases chance of hotel booking.**


Thank you!