

Capstone Project - 1

Hotel Booking Analysis

Team members

Arunkumar Gond
Anuj kumar Gaur
Adithya Karumuri
Nishant Verma

MAJOR POINTS TO BE DISCUSSED

- DATA SUMMARY
- UNDERSTANDING DATASET
- CLEANING DATASET
- PREPARING DATASET
- EXPLORATORY DATA ANALYSIS(EDA)
- CONCLUSION



DATA SUMMARY

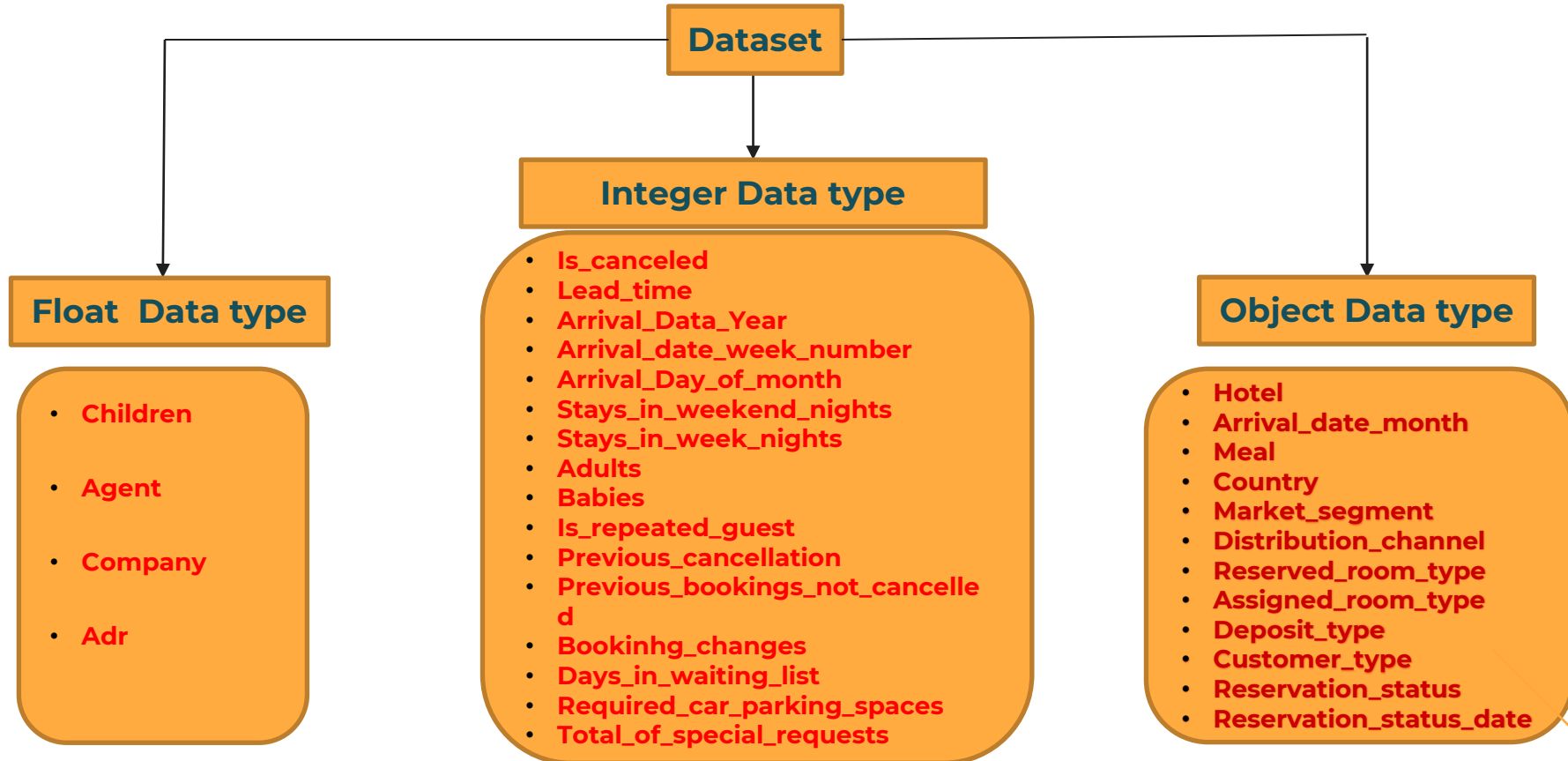
DATA SET PROVIDED IS THE COLLECTION OF DIFFERENT COLUMN AS BELOW-

- hotel
- is_canceled
- lead_time
- arrival_date_year
- arrival_date_month
- arrival_date_week_number
- arrival_date_day_of_month
- stays_in_weekend_nights
- stays_in_week_nights
- adults
- children
- Babies
- Meal
- Country
- market_segment
- distribution_channel

DATA SUMMARY(cont..)

- is_repeated_guest
- previous_cancellations
- previous_bookings_not_canceled
- reserved_room_type
- assigned_room_type
- booking_changes
- deposit_type
- agent
- company
- days_in_waiting_list
- customer_type
- adr
- required_car_parking_spaces
- total_of_special_requests
- reservation_status
- reservation_status_date

DATA SUMMARY(cont..)



UNDERSTANDING DATASET

1. Going through the Dataset, It is confirmed that all data is Structured Dataset with 32 columns and 119390 rows.
2. Dataset available are in 3 types , Integer , Float and Object .
3. Data provide the information of bookings of customers and there brief Details.
4. Data Provided ranges from 2015 – 2017 with 2 Hotels as main Subject line.
5. While going through the Data, We found that some of the Columns Values were found NAN , so we surely need it to replace with Values (0) that can be used for further process of Data, which will be processed in step Cleaning.
6. Other than the NAN values , we also found around 31994 Duplicate Values were found ,which can be removed in Cleaning Process.



CLEANING AND PREPARING DATASET

Why cleaning ?

The data set which we receive consist consist lot of Bad Data , which would result in errors while processing ,
Hence cleaning is required for the Dataset provided to us.

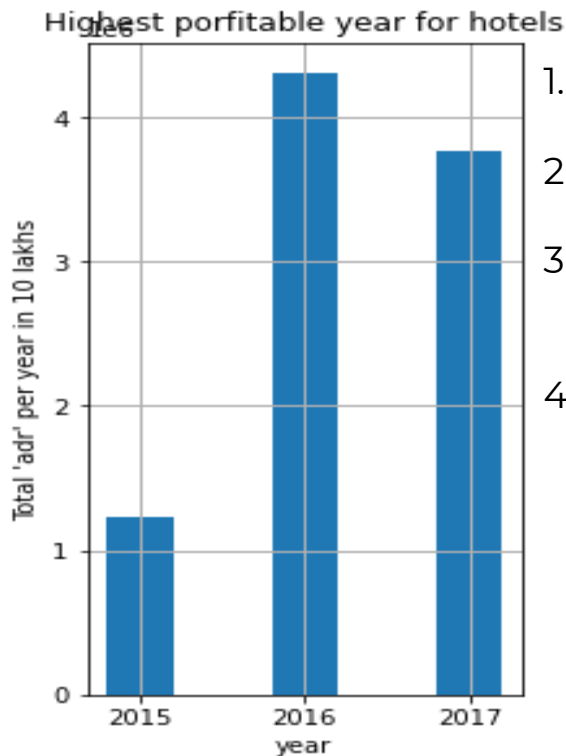
What we did?

1. First we Deleted Duplicate Values from the Dataset provided to us.
2. We Replaced all the NAN values by 0 for data type Int and float, for object Datatype we replaced it by others.
3. We changed the Datatype of the columns with suitable Datatype according to the values present inside.
4. We added Extra columns which would be essential for further process.



EXPLORATORY DATA ANALYSIS (EDA)

In which year hotels has the maximum profit ?



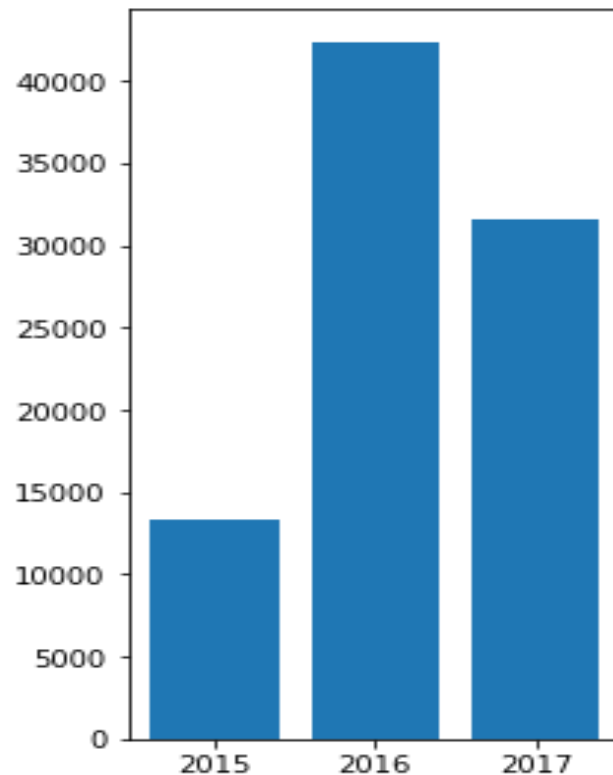
1. If we consider the graph , we can clearly see that the maximum ADR was found in the year 2016 .
2. In 2015 , ADR was found to be 1226930.09 , which was comparatively less than 2016 and 2017 , ADR sum.
3. In 2016 , there was sudden increase in the ADR sum value , which gone upto 4303105.84 , has noted as the maximum profitable years for the hotels.
4. In 2017 , we can see that it decreases to 3761531.89.

We can Conclude that, Maximum profitable year was found on 2016 and minimum on 2015.

EXPLORATORY DATA ANALYSIS (EDA)

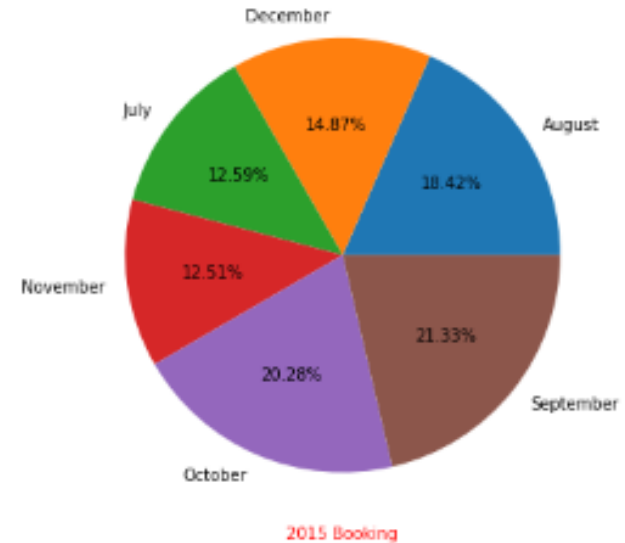
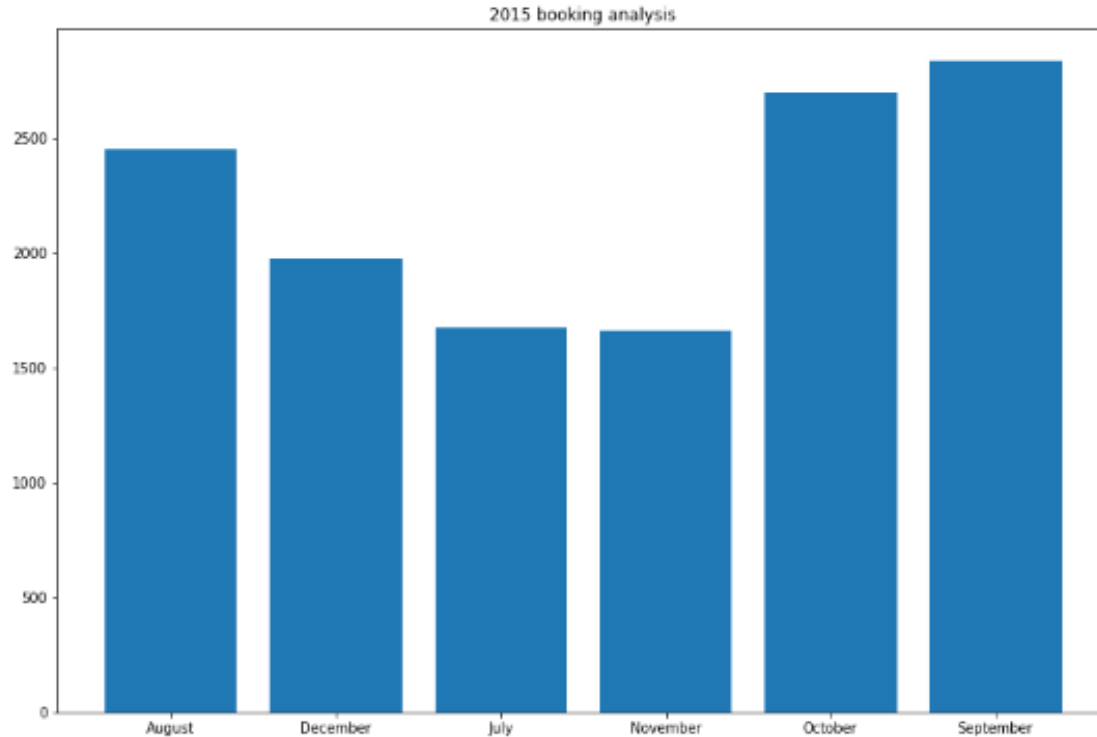
In which year hotels has the maximum booking?

Here, we can find the year in which maximum booking was made and Also, the minimum booking made.



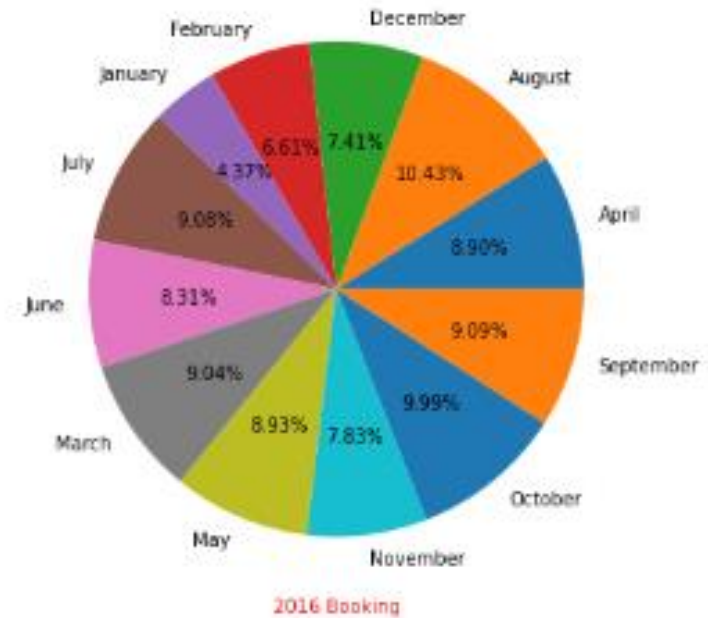
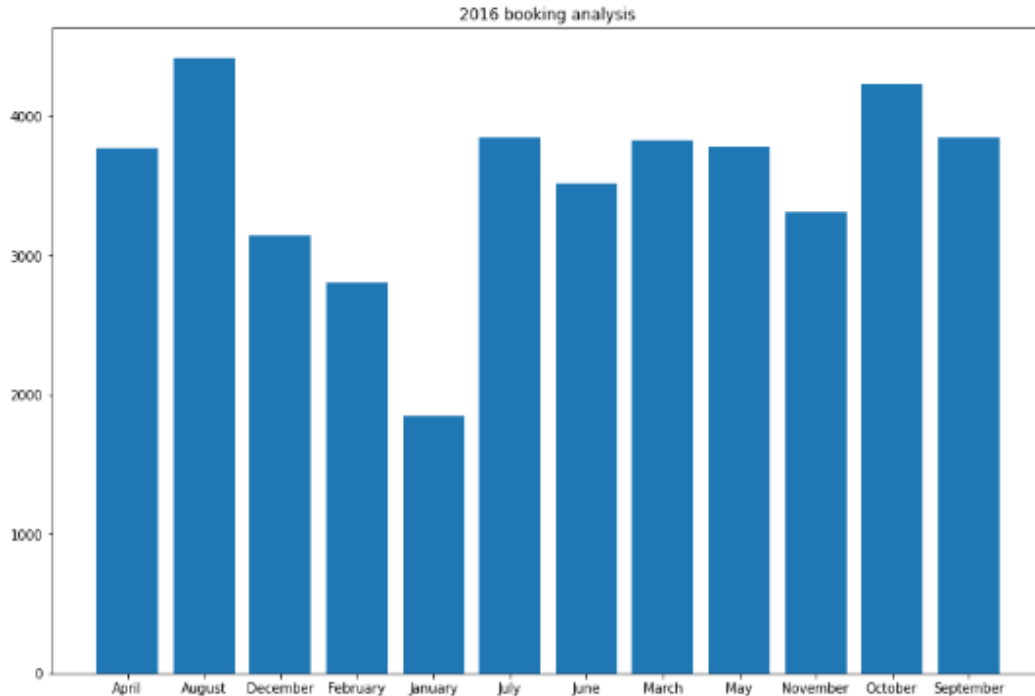
EXPLORATORY DATA ANALYSIS (EDA)

In which month of the year 2015, hotels has the maximum booking?



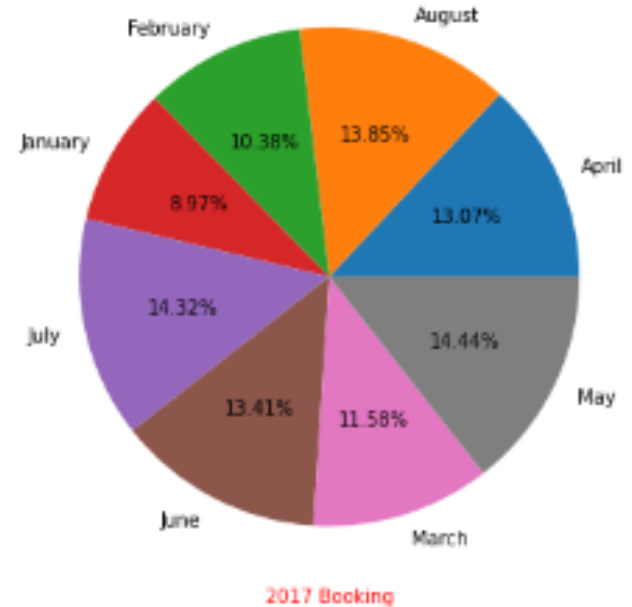
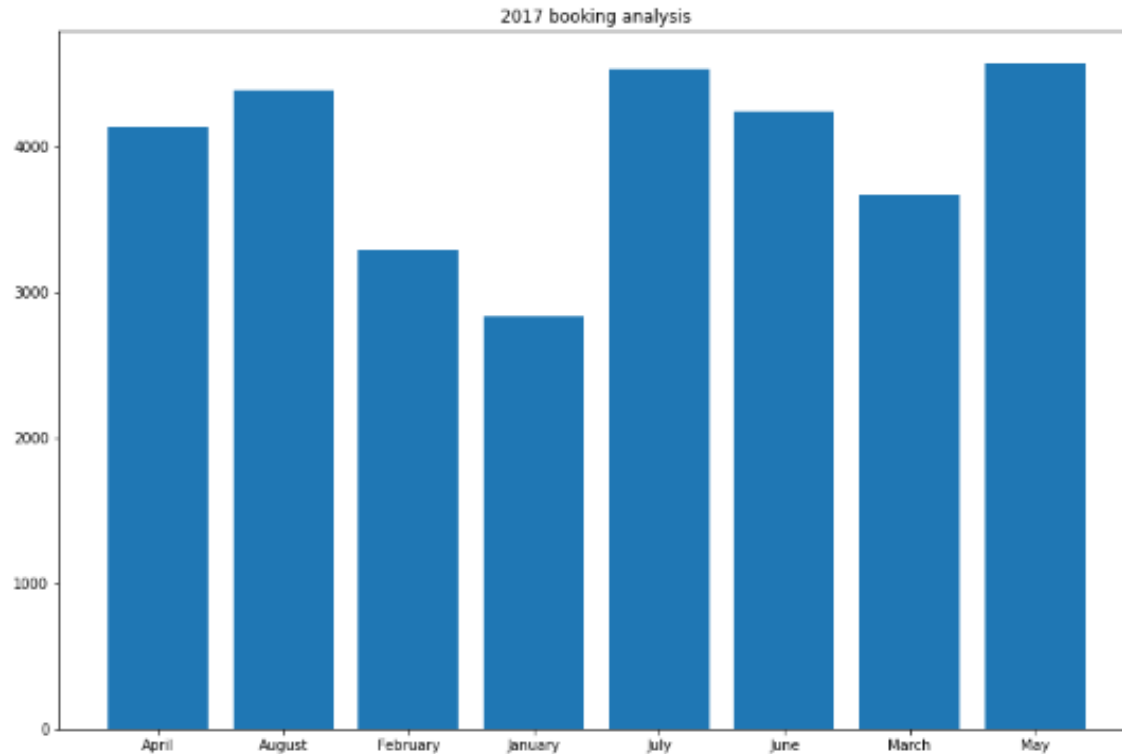
EXPLORATORY DATA ANALYSIS (EDA)

In which month of the year 2016, hotels has the maximum booking?

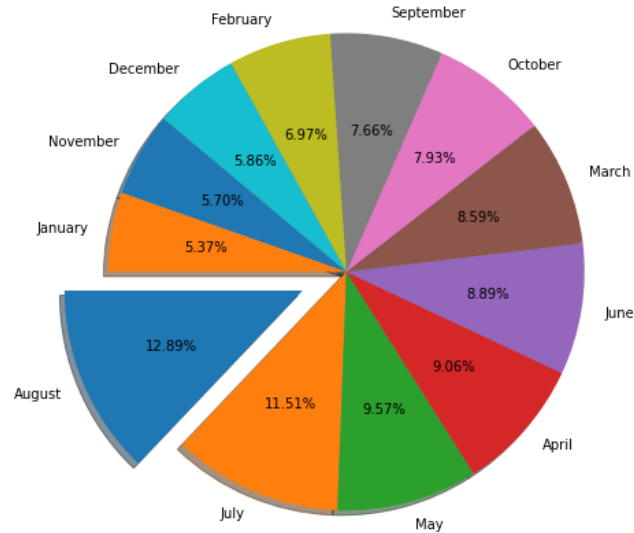


EXPLORATORY DATA ANALYSIS (EDA)

In which month of the year 2017, hotels has the maximum booking?

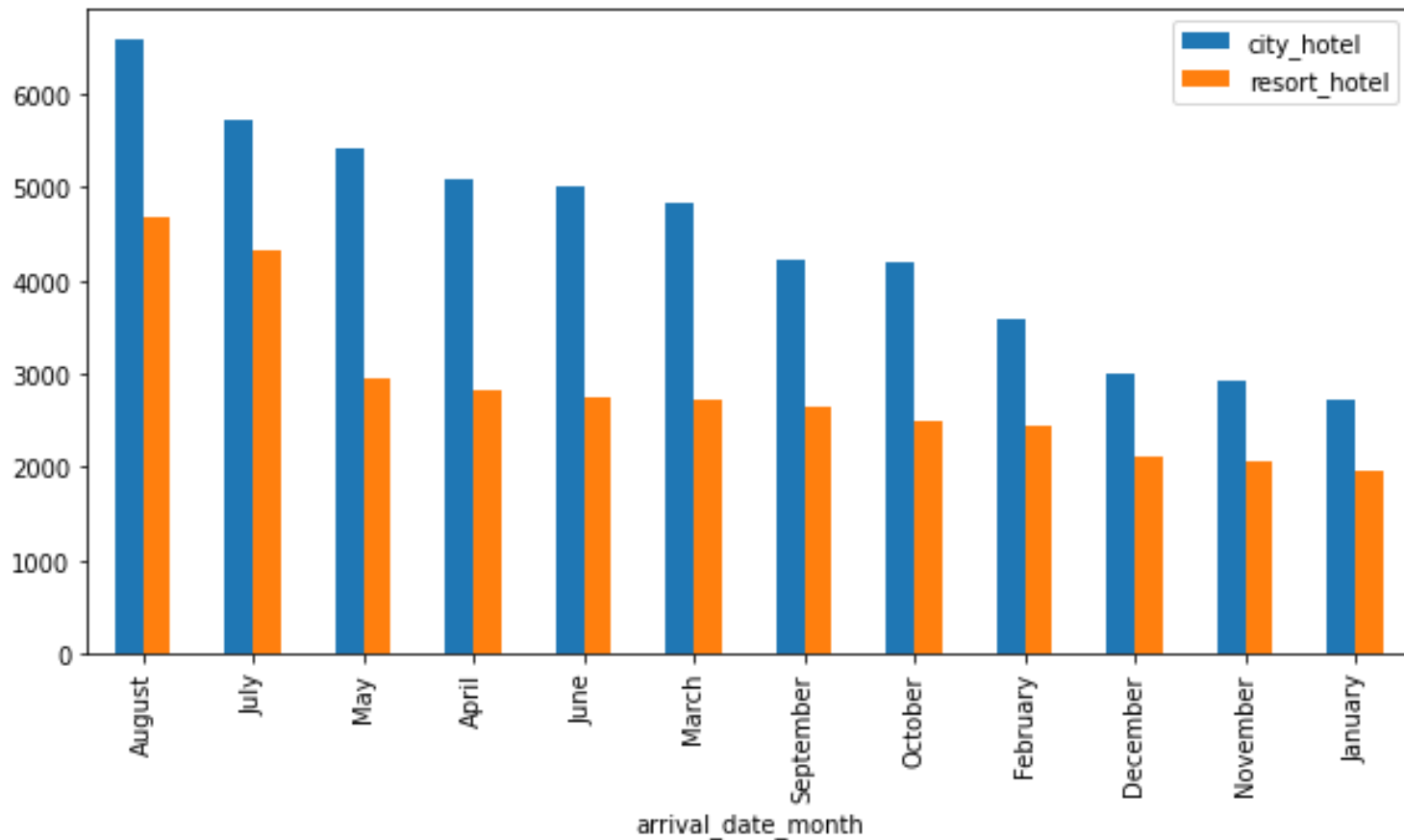


EXPLORATORY DATA ANALYSIS (EDA)



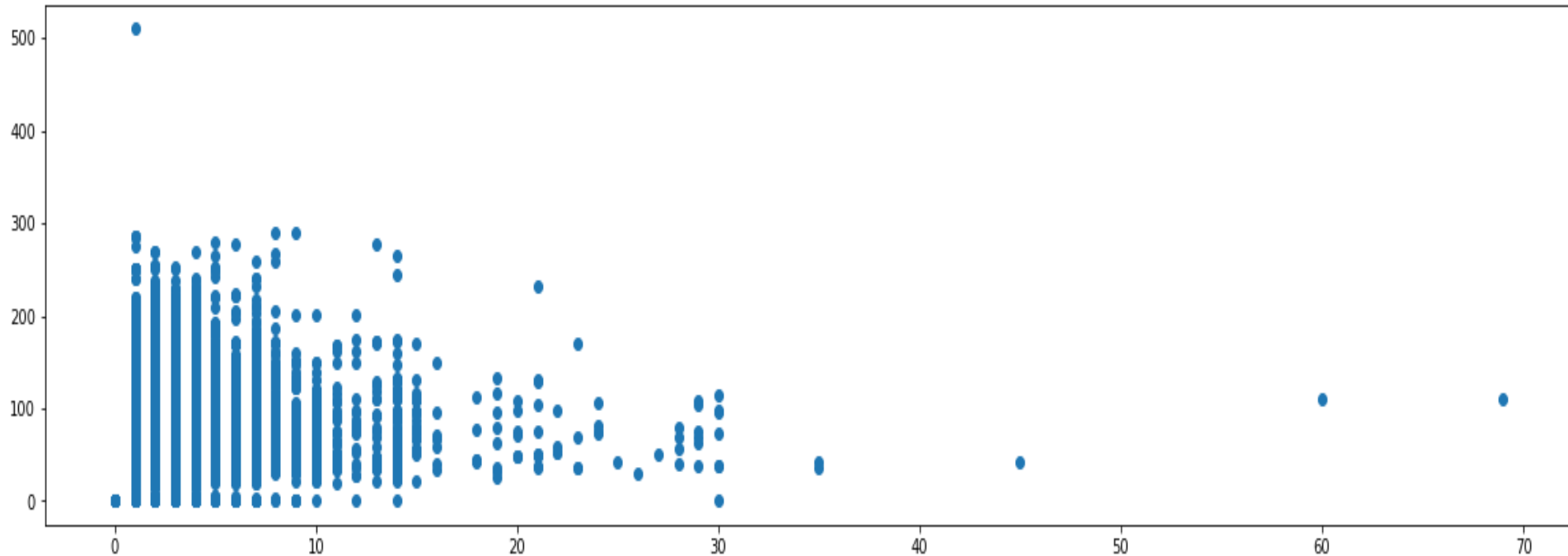
According to the chart we can find that , maximum no. of Bookings in Average of 3 years is found on August Month.

EXPLORATORY DATA ANALYSIS (EDA)



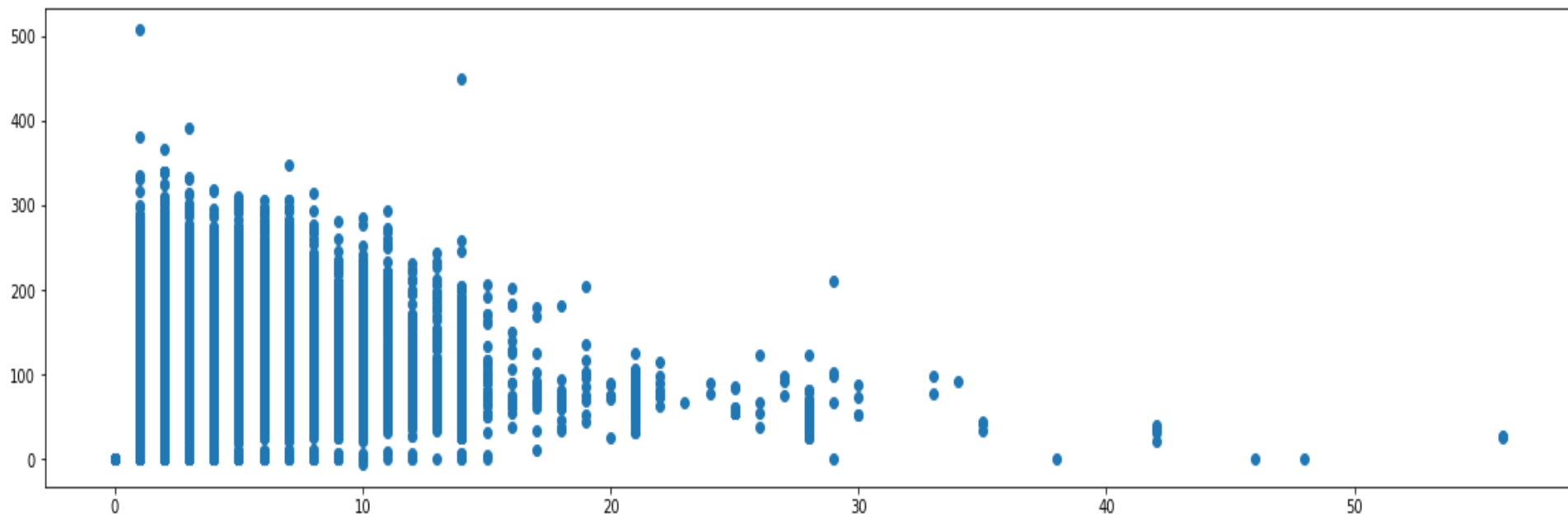
EXPLORATORY DATA ANALYSIS (EDA)

Optimal length of stay in order to get the best daily rate for 1 Person?



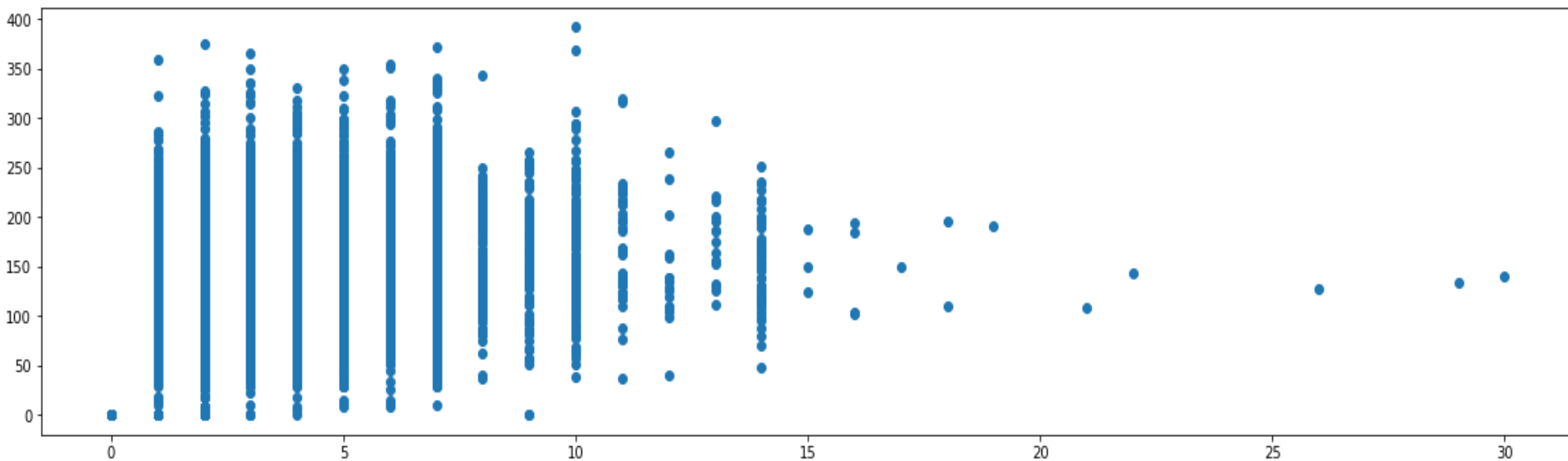
EXPLORATORY DATA ANALYSIS (EDA)

Optimal length of stay in order to get the best daily rate for 2 people?



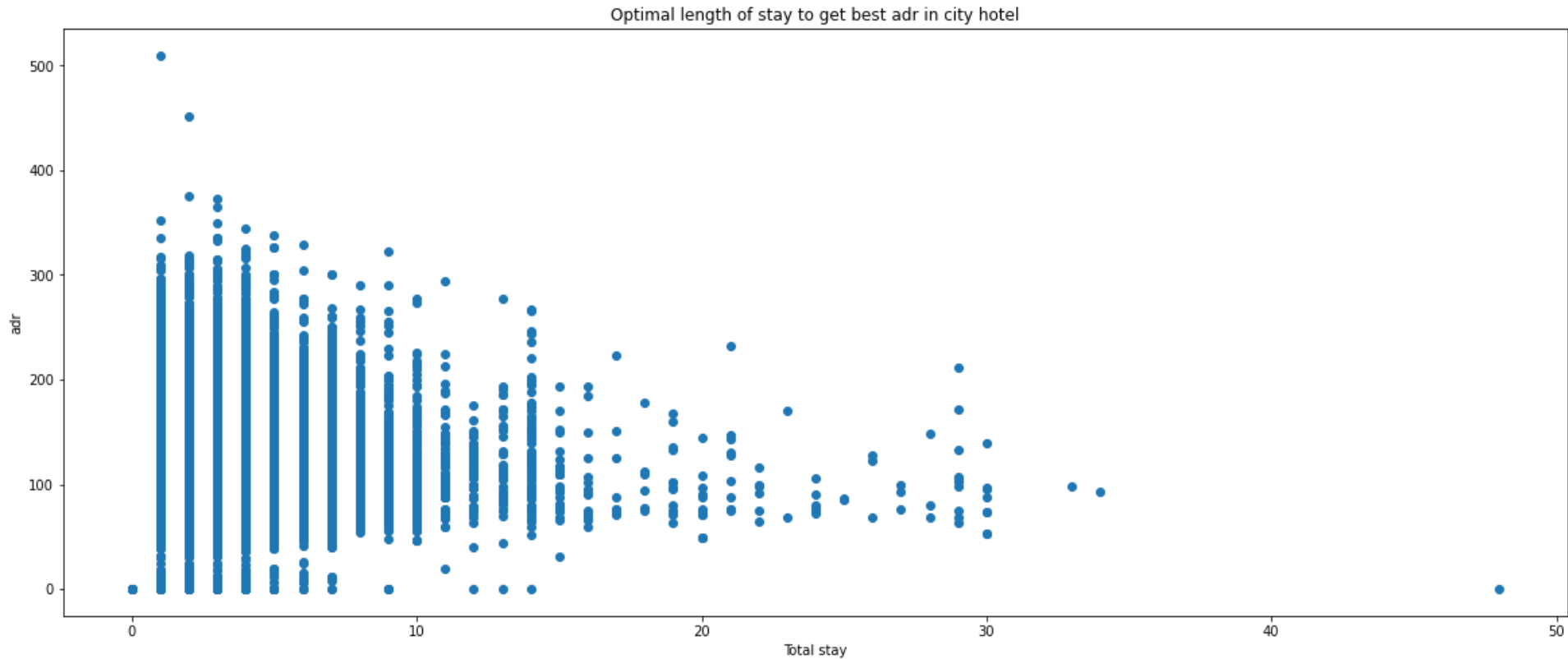
EXPLORATORY DATA ANALYSIS (EDA)

Optimal length of stay in order to get the best daily rate for 3 people?



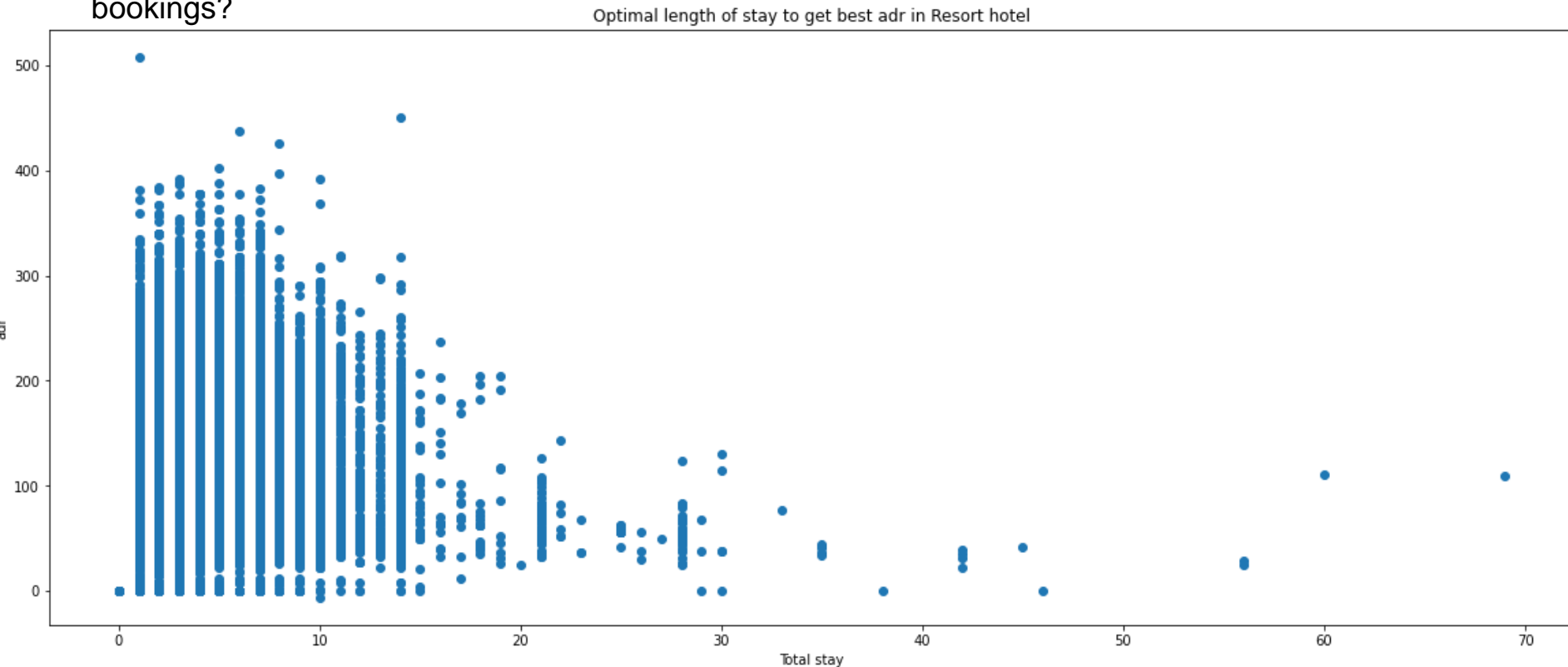
EXPLORATORY DATA ANALYSIS (EDA)

Optimal length of stay in order to get the best daily rate considering City Hotel for overall bookings?



EXPLORATORY DATA ANALYSIS (EDA)

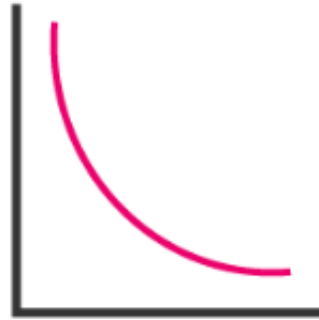
Optimal length of stay in order to get the best daily rate considering Resort Hotel for overall bookings?



EXPLORATORY DATA ANALYSIS (EDA)

Optimal length of stay in order to get the best daily rate ?

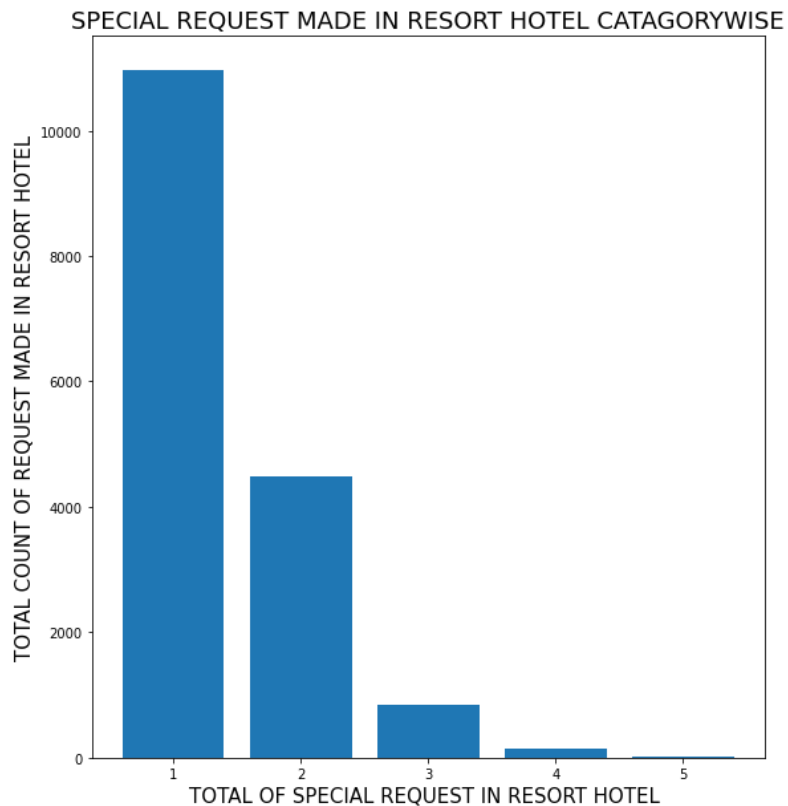
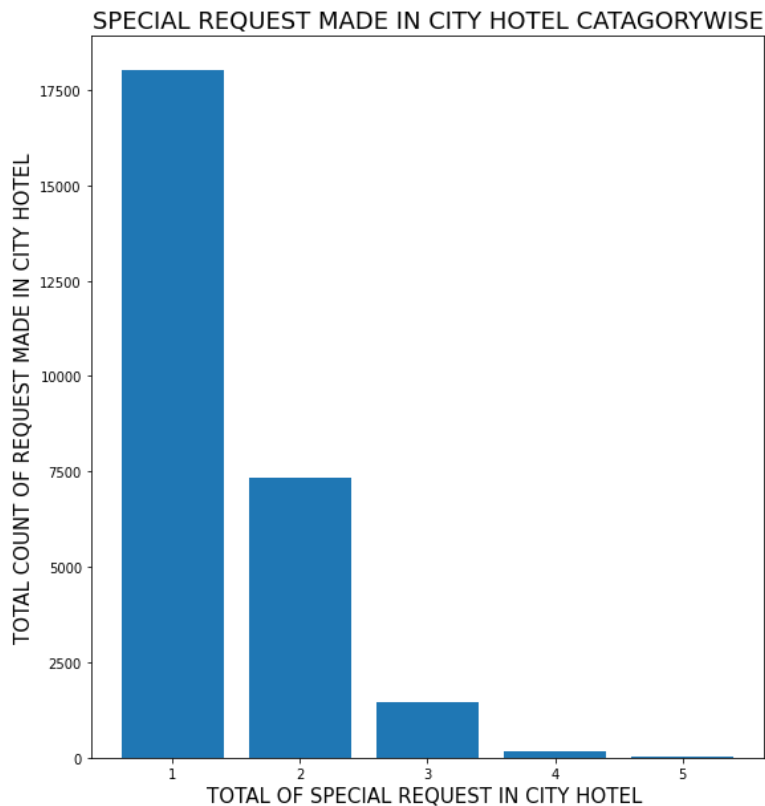
Considering Total no. of people as 1 , 2 and 3 , as shown in previous scatter plots , we can conclude if the optimal length of stay increase , the ADR value get Decrease.



Inverse proportion

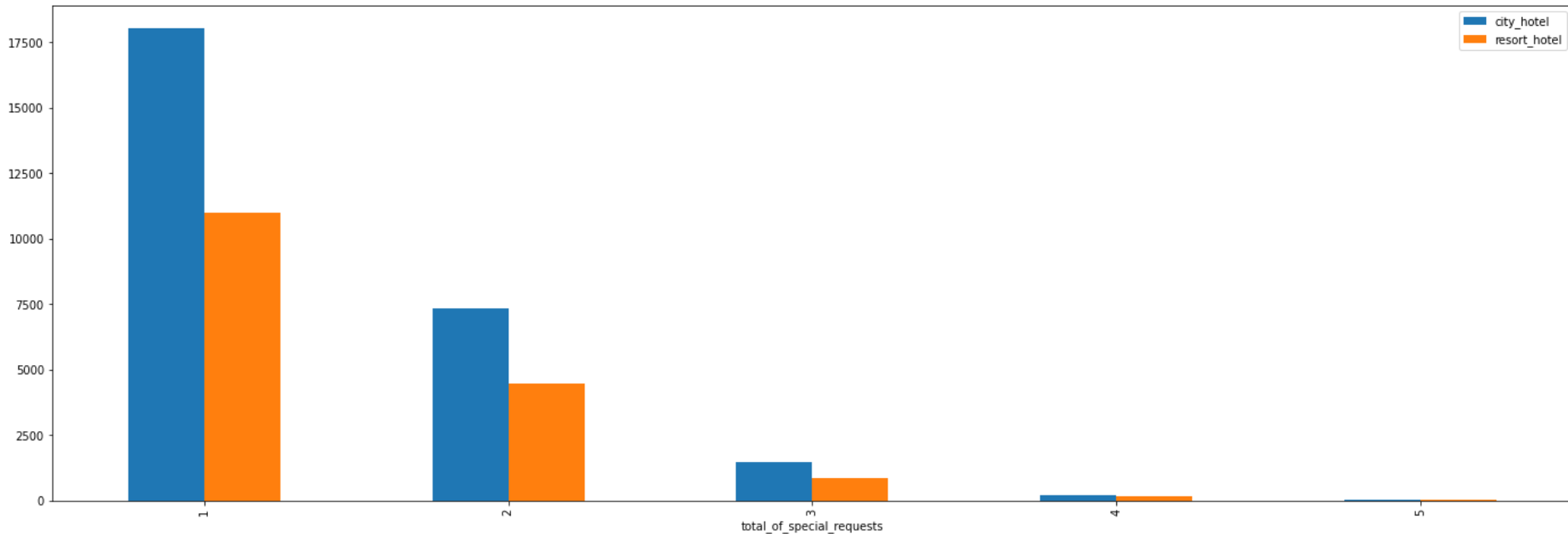
EXPLORATORY DATA ANALYSIS (EDA)

Now lets analyse , which Hotel has Received the max special request ?



EXPLORATORY DATA ANALYSIS (EDA)

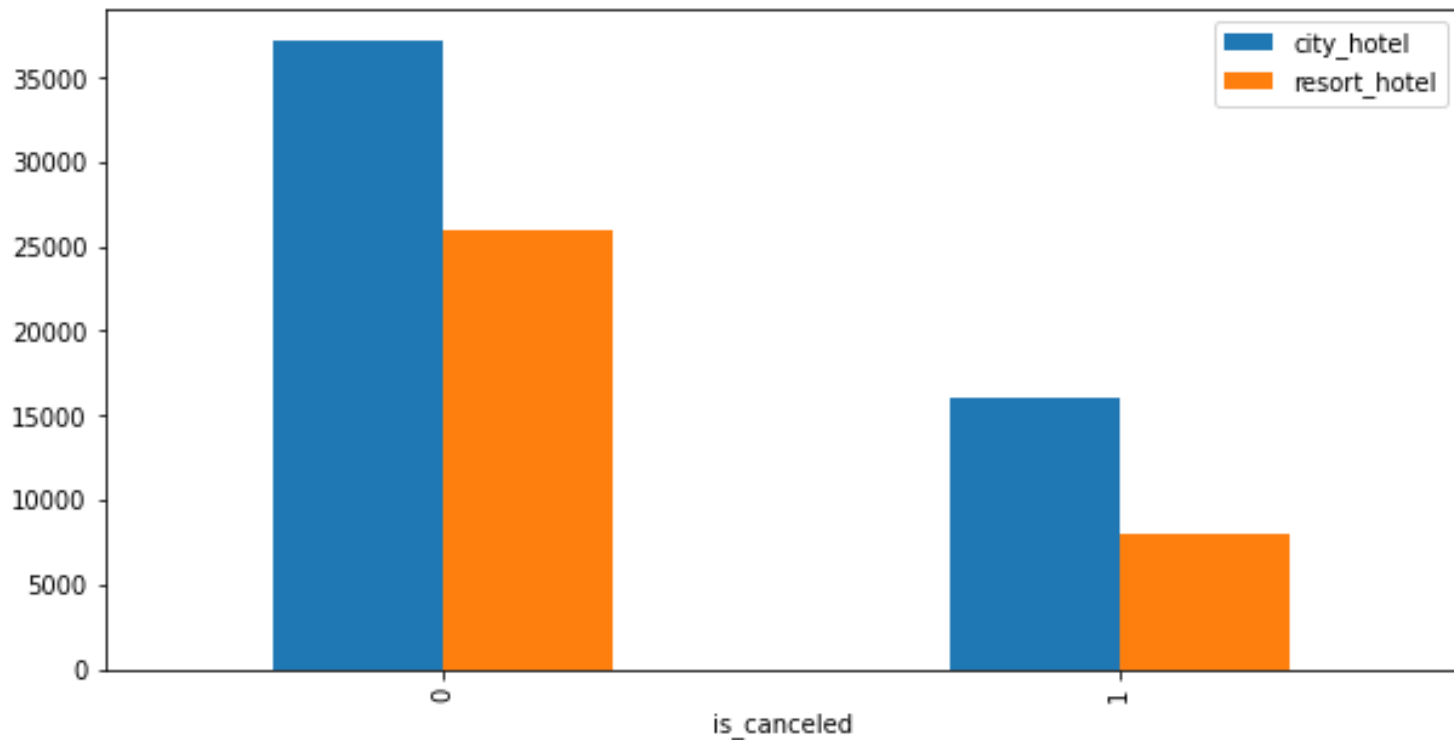
Now lets analyse , which Hotel has Received the max special request ?



Comparing the Hotels, it is clear that City Hotel is supposed to be received maximum special request as compare to the Resort Hotel.

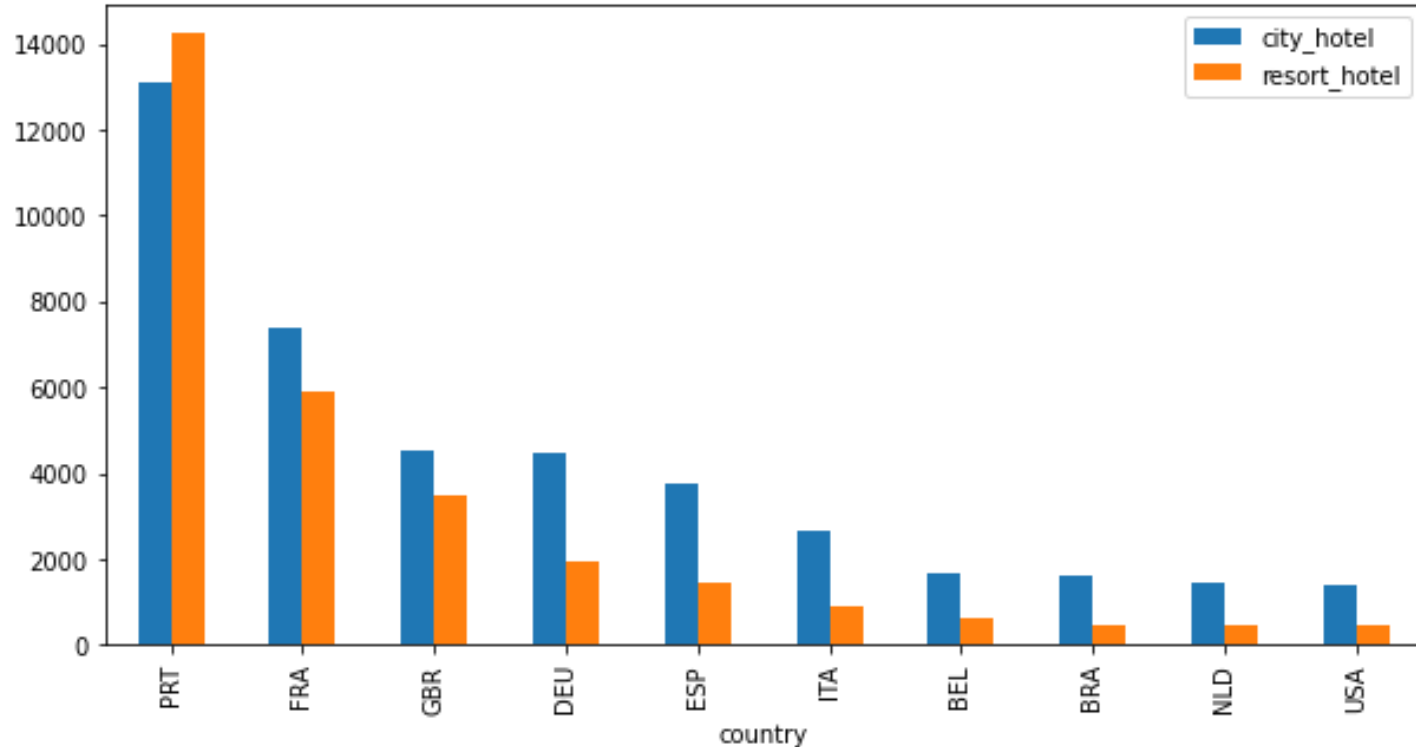
EXPLORATORY DATA ANALYSIS (EDA)

Which Hotels have the maximum Cancellation request ?



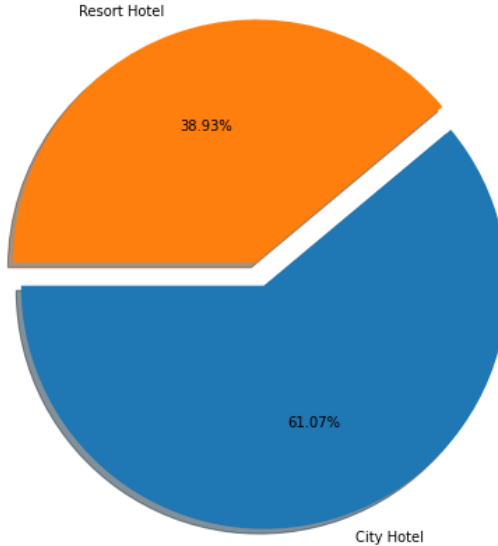
EXPLORATORY DATA ANALYSIS (EDA)

Which country residence made the maximum booking ?



EXPLORATORY DATA ANALYSIS (EDA)

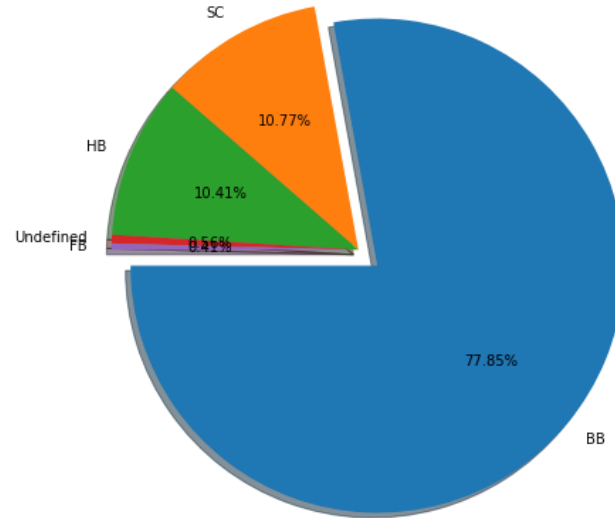
Which hotels are mostly preferred by the Guests ?



City Hotels are the most Preferred Hotels by the Guests.

EXPLORATORY DATA ANALYSIS (EDA)

Which meals are mostly preferred by the Guests ?

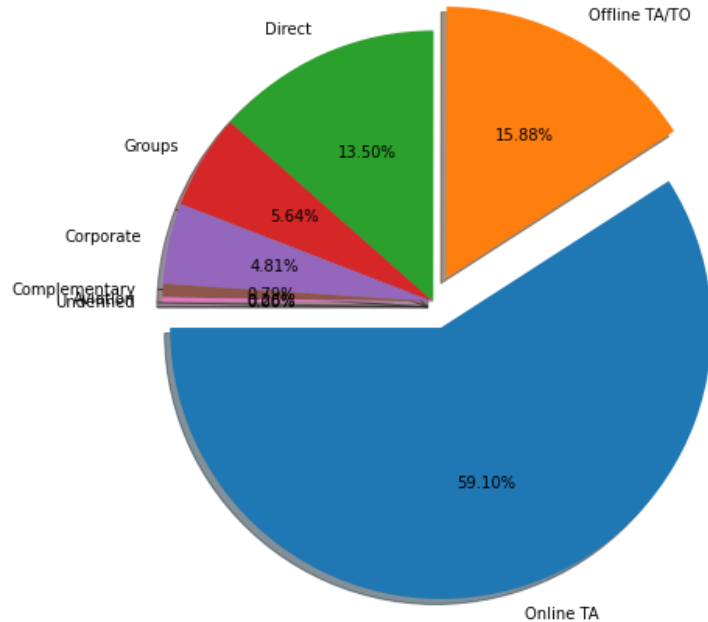


Breakfast are the most Preferred meals by the Guests.

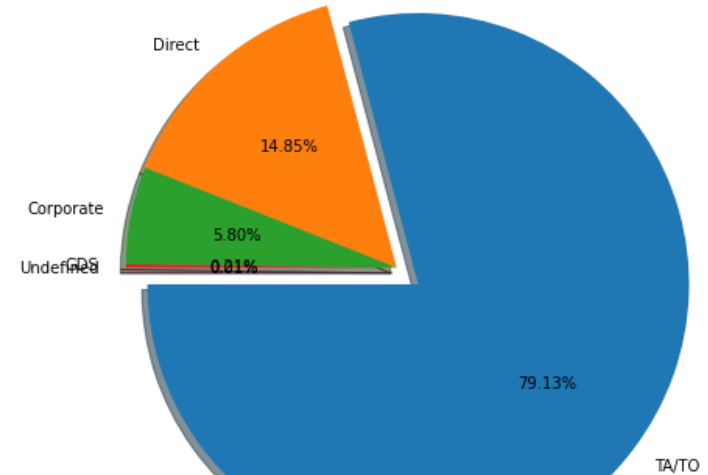
EXPLORATORY DATA ANALYSIS (EDA)

How do customer prefer to Book Hotels ?

Market segment



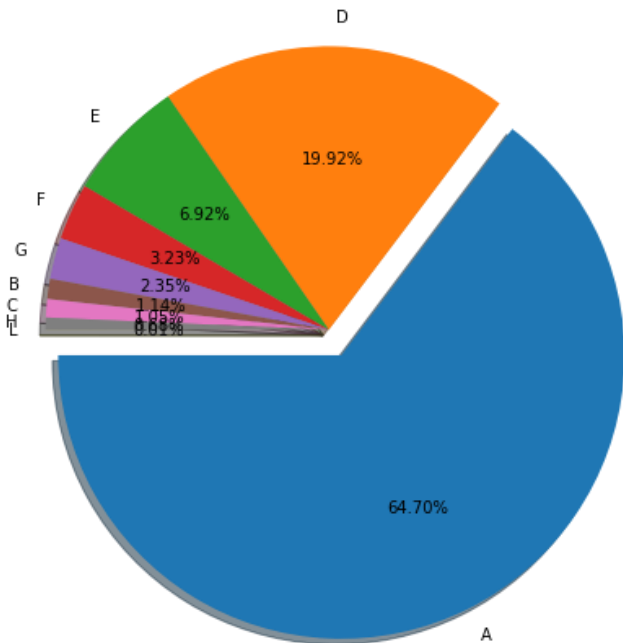
Distribution channel



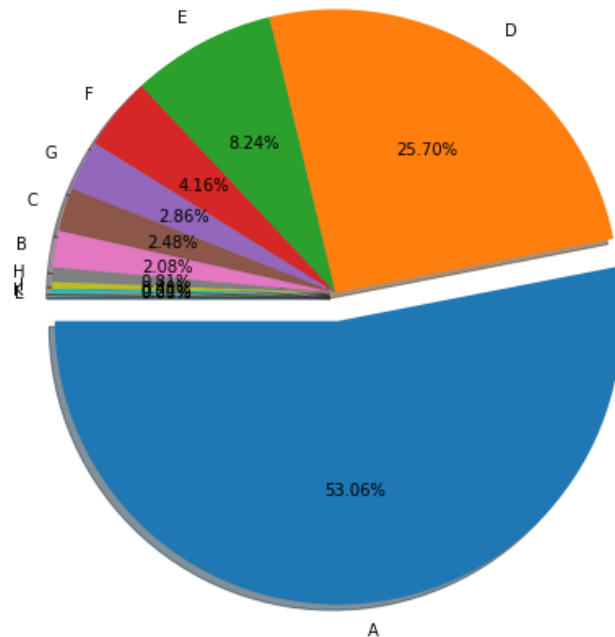
EXPLORATORY DATA ANALYSIS (EDA)

Which rooms are mostly referred by the guests?

Reserved Room Type

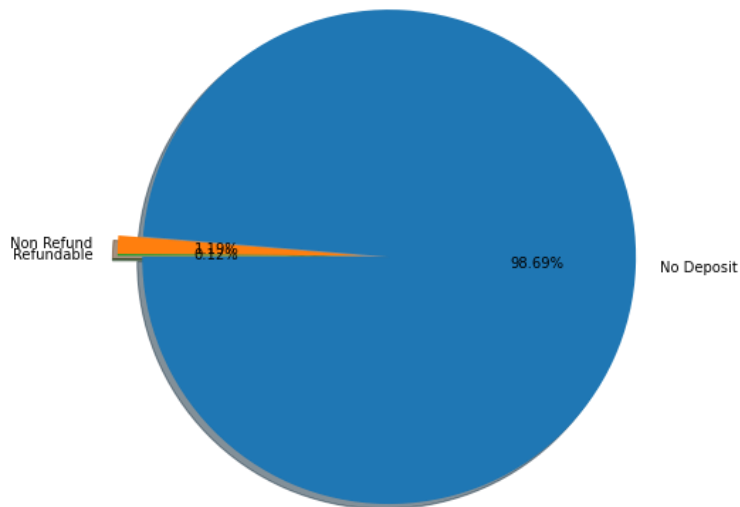


Assigned Room Type



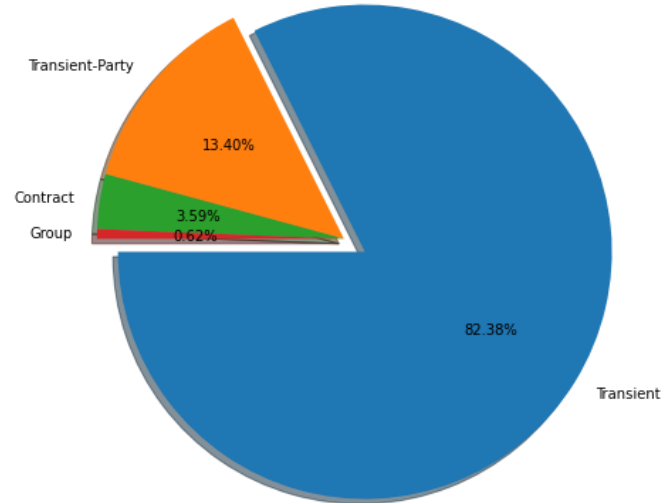
EXPLORATORY DATA ANALYSIS (EDA)

Which payment mode customer refers the most?



EXPLORATORY DATA ANALYSIS (EDA)

Which type Guests visit the hotel mostly?



CONCLUSION

1. 2016 Was the year of Maximum Bookings Made in the Hotels, As well as Highly Profitable year for the Hotels.
2. Maximum no. of Bookings in Average of 3 years is found on August Month.
3. City Hotels was Mostly preferred in comparison to resort Hotels.
4. we can conclude if the optimal length of stay increase , the ADR value get Decrease.
5. Comparing the Hotels, it is clear that City Hotel is supposed to be received maximum special request as compare to the Resort Hotel.
6. Maximum Cancellation request was made in City Hotels.
7. PRT country residence have the max Bookings in Hotels , In which they mostly prefer City Hotels for Booking in comparison to Resort Hotels.
8. 61.07 % People prefer the City Hotels and 38.93 % people prefer the resort Hotels.
9. Breakfast are the most Preferred meals by the Guests
10. Online & Offline TA/TO are mostly preferred by guests for booking.
11. Type A rooms are the most preferred rooms by the Guests.
12. No Deposit payment mode is mostly preferred by Guests.
13. Hotel bookings are mostly done by transient type Guests.



THANK YOU