

# Capstone Project-1

## Exploratory Data Analysis

### Team Members

1. Arvind Kale
2. Priyanka Jain
3. Mohd Sakib Quraishi
4. Nitish kumar

# Hotel Booking EDA



# Information Given

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.

We have to analyse the data but before that we should do the data cleaning ,pre-processing and outlier treatment to get the accurate and meaningful output.

You will discover in this presentation everything step by step as we have been thoroughly shown the charts and graph.

# PYTHON LIBRARIES USED

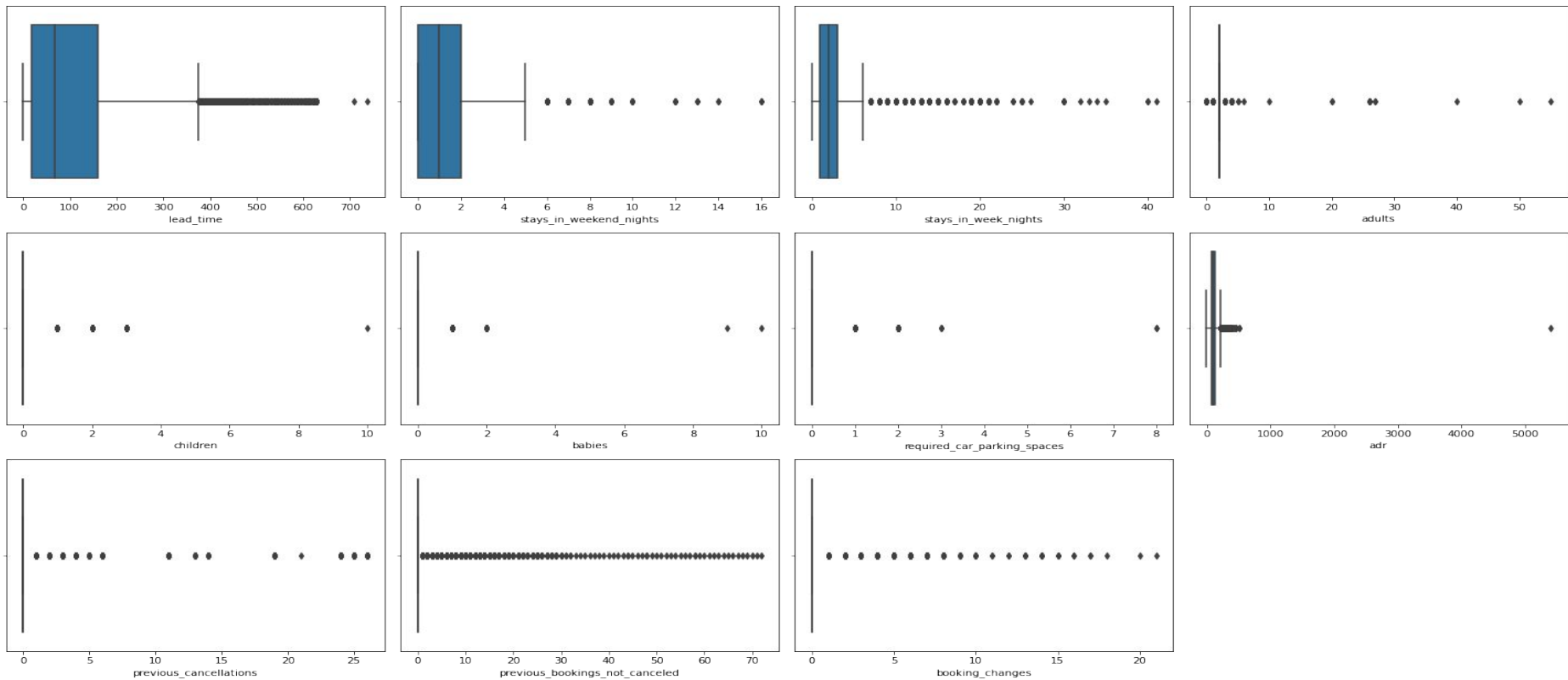
PANDAS

SEABORN

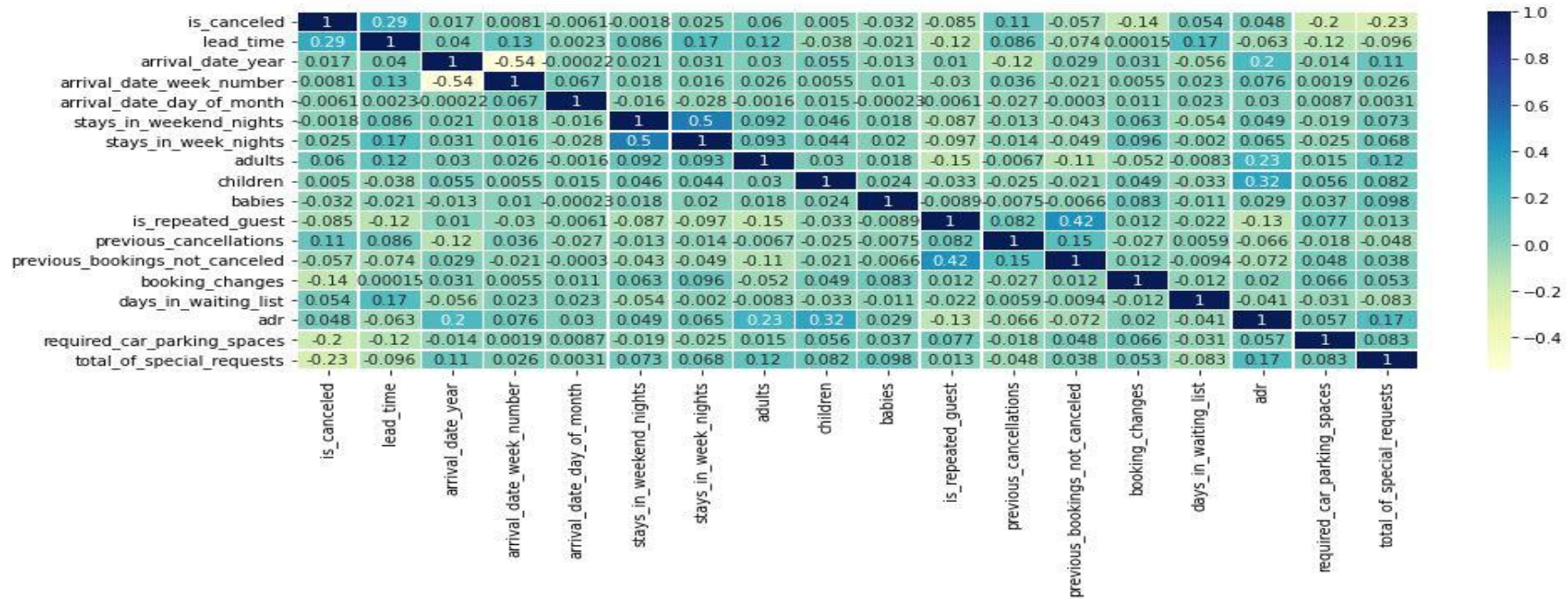
MATPLOTLIB



# Outliers

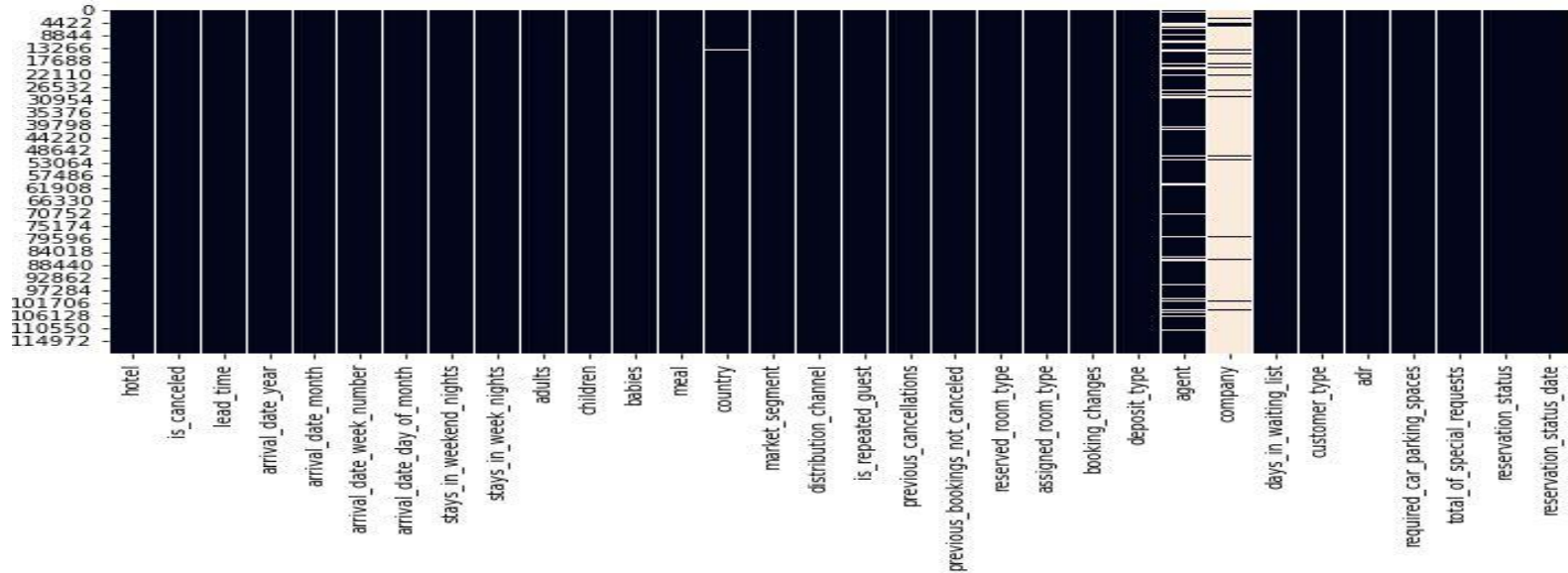


# CORRELATION HEATMAP



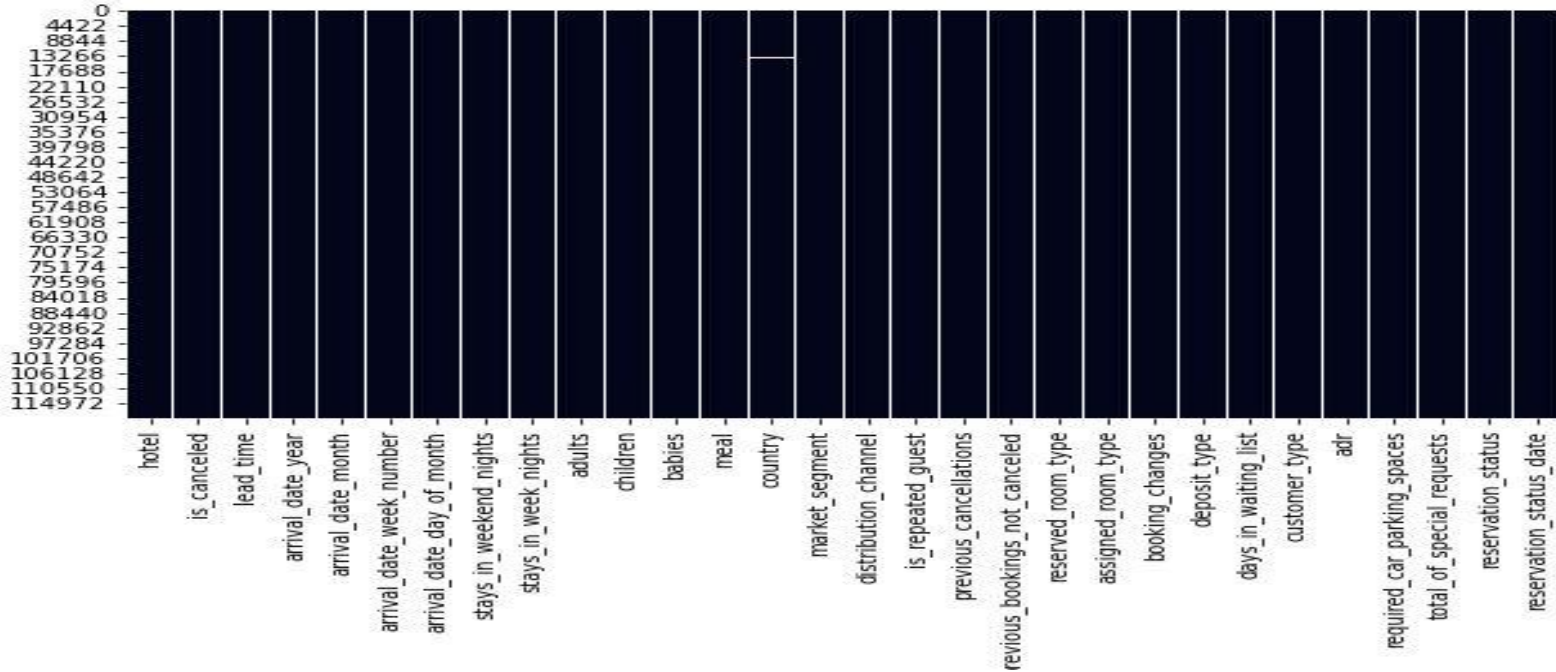
- Apart from diagonal elements there seems no strong correlation between the variables
- We can conclude that variables are independent of each other

# PREPARATION OF THE DATA



- Columns company and agent are having most of the null rows
- It will not add any significant value if more than 50 percent data is missing

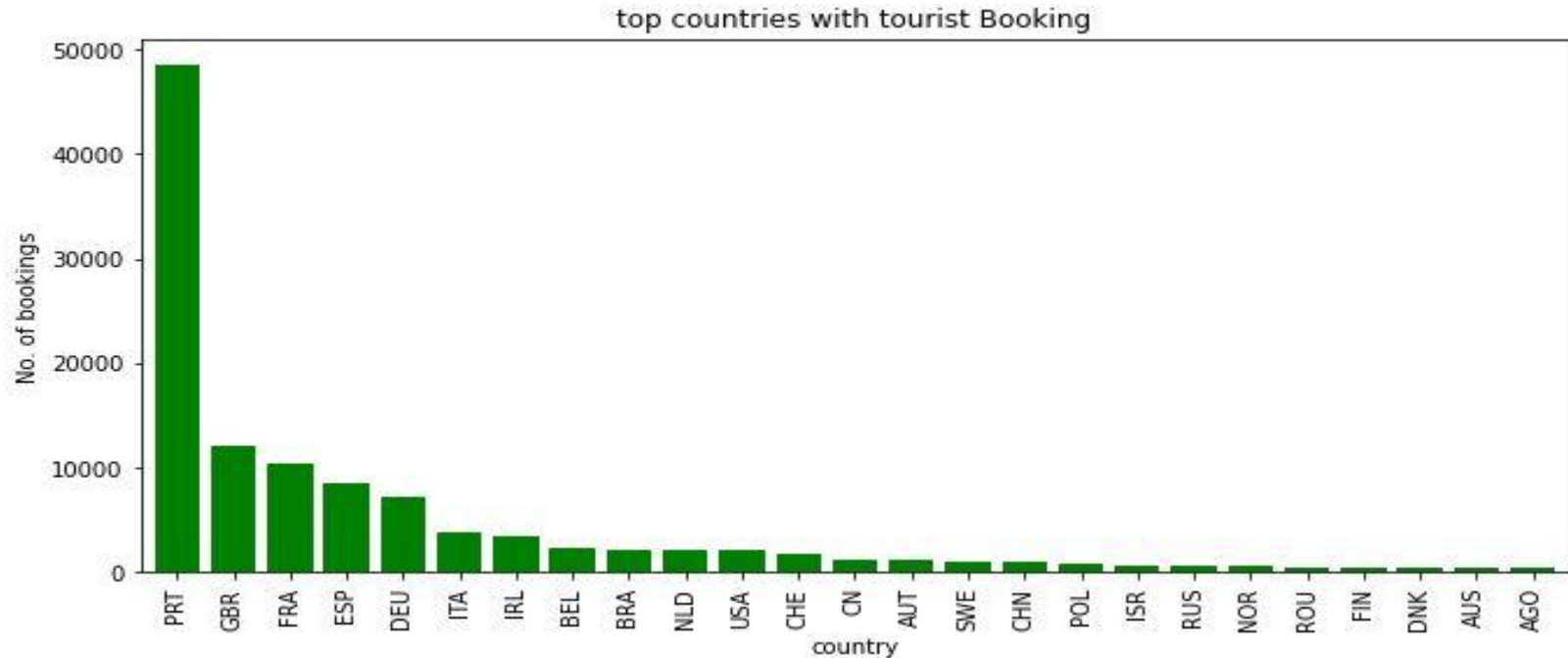
# REMOVAL OF NULL COLUMNS



We can see that now there is no null values in the dataset.



# COUNTRIES WITH MOST NUMBER OF BOOKING



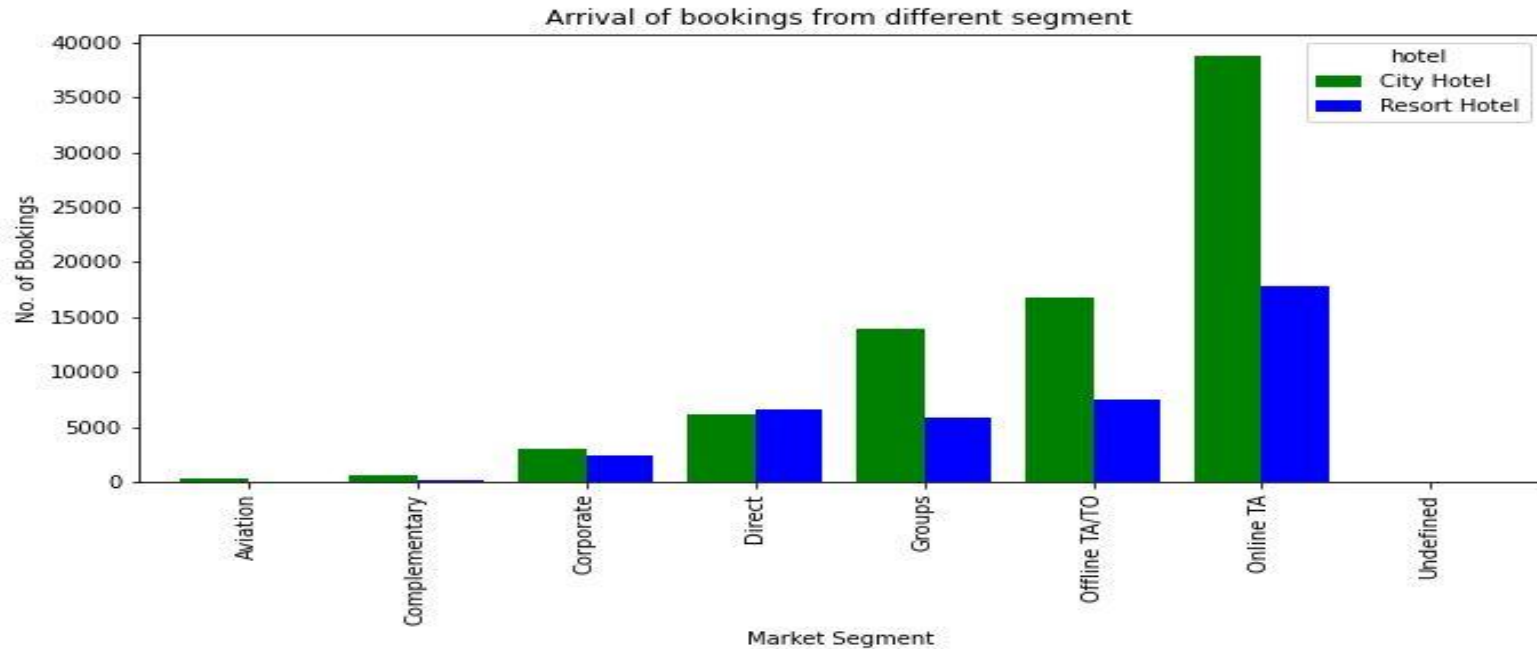
- **Portugal is the country from where almost 50k bookings were being done**
- **Portugal is followed by great Britain, France and Spain**

# ARRIVAL OF GUESTS PER MONTH



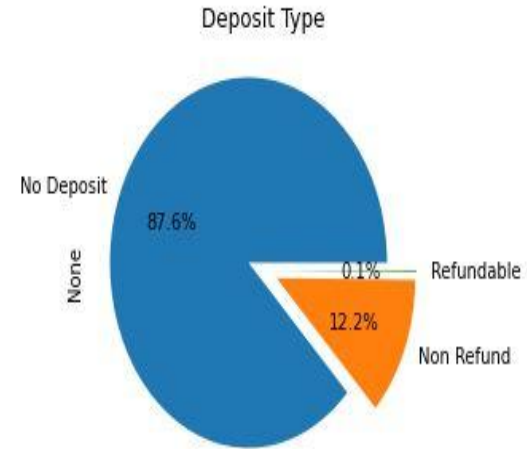
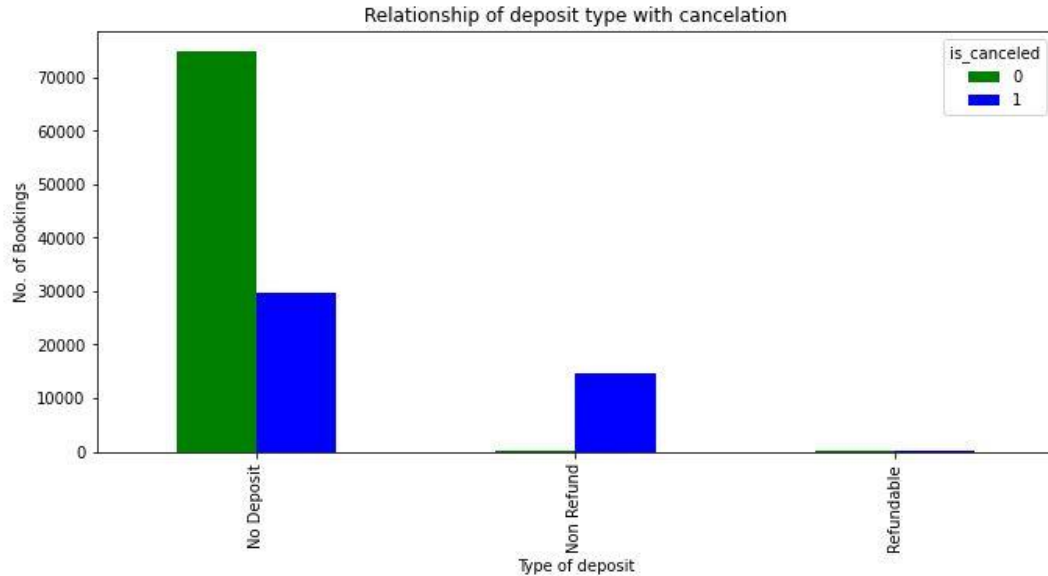
- Number of arrival rises steadily till august and sharply declines from November
- Highest arrival of are in the month of July and August
- November ,December and January saw the least number of guest

# ANALYSIS BASED ON MARKET SEGMENT



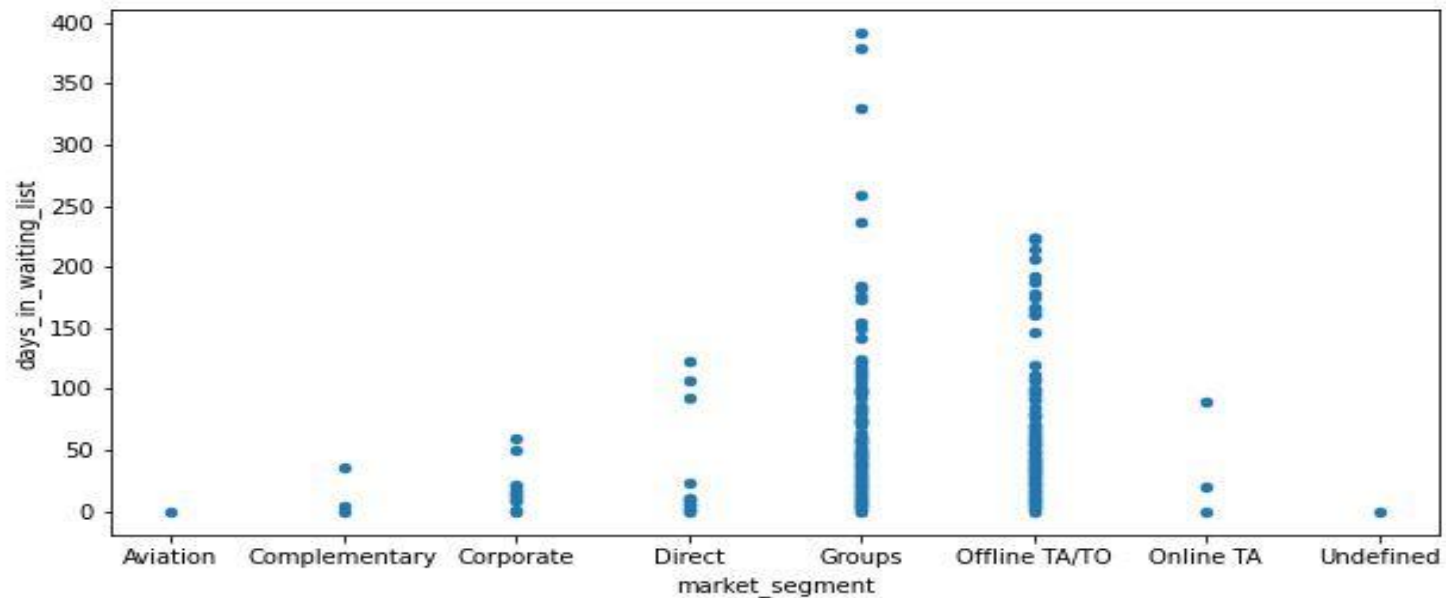
- **Aviation and complementary bookings are negligible compared to the online and offline TA**
- **Corporate ,direct and group bookings together add up to online TA**

# DEPOSITS VS CANCELATION RATE



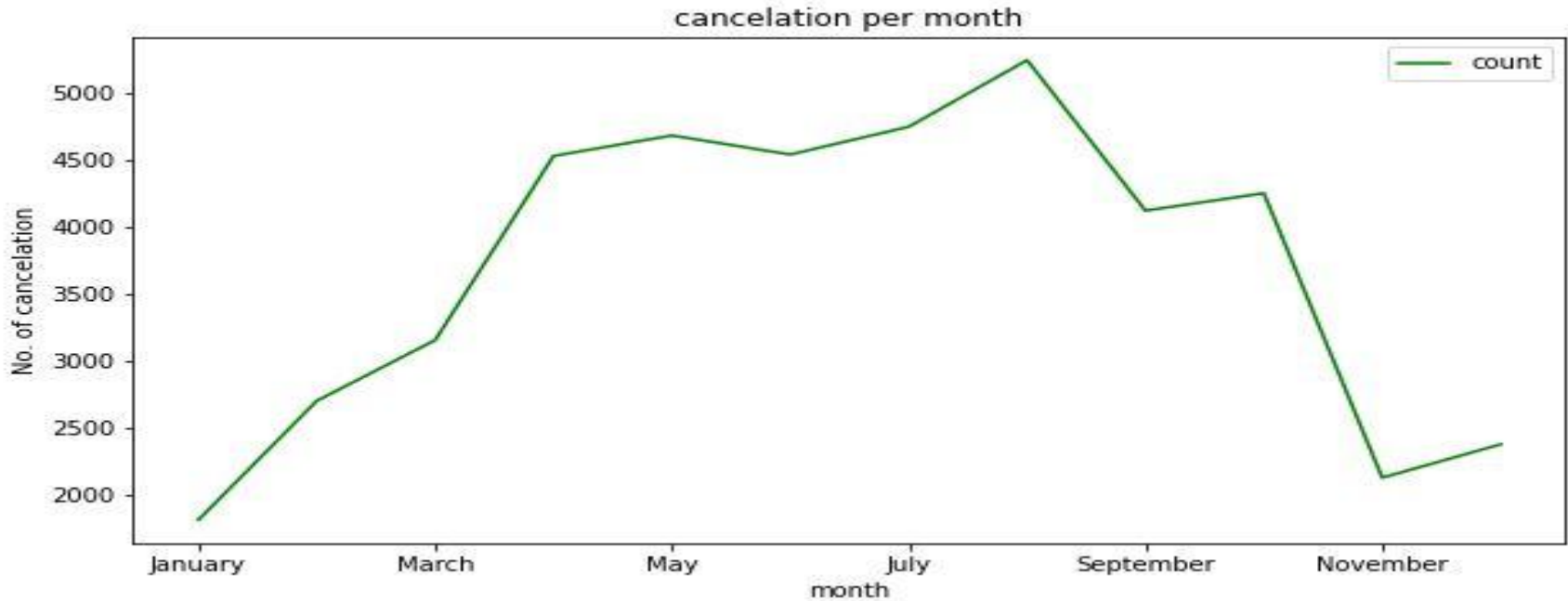
- **Refundable is least type of preferred deposit amongst all**
- **Most Guests with Non Refundable deposits have canceled**
- **More than 70 percent guest with no deposit have not canceled bookings**

# WAITING TIME FOR VARIOUS SEGMENT



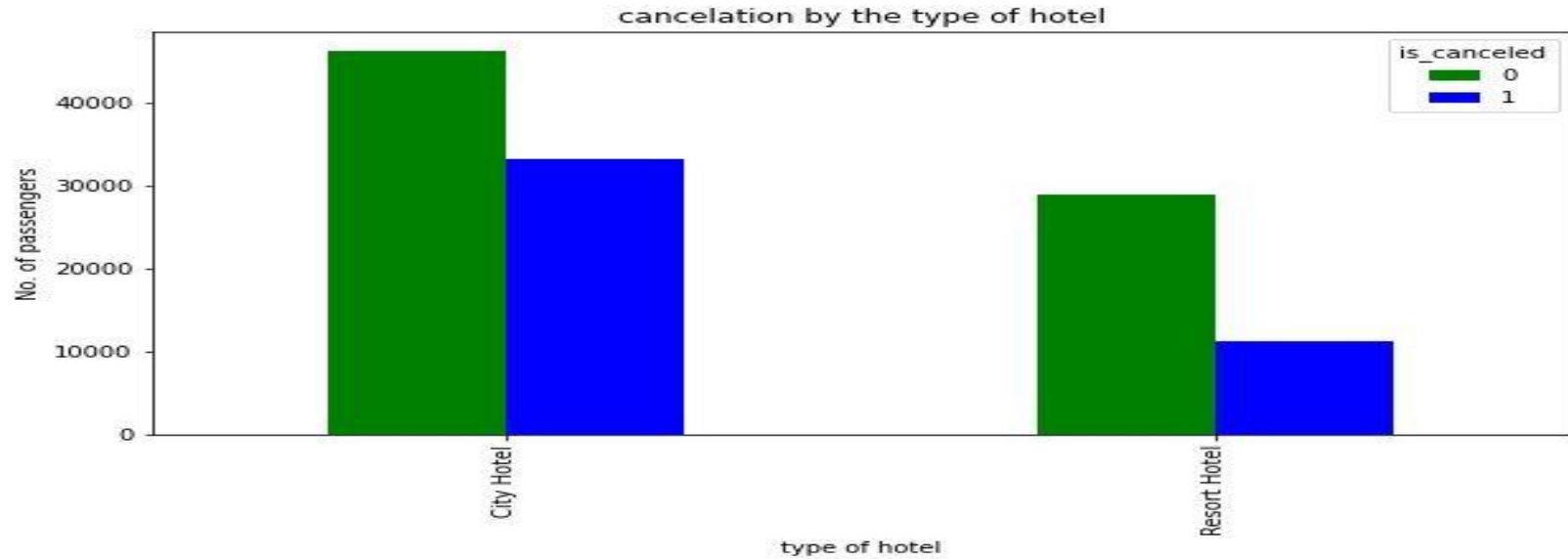
Aviation has least waiting time and Groups has highest waiting time compared to all other segment.

# CANCELATION PER MONTH

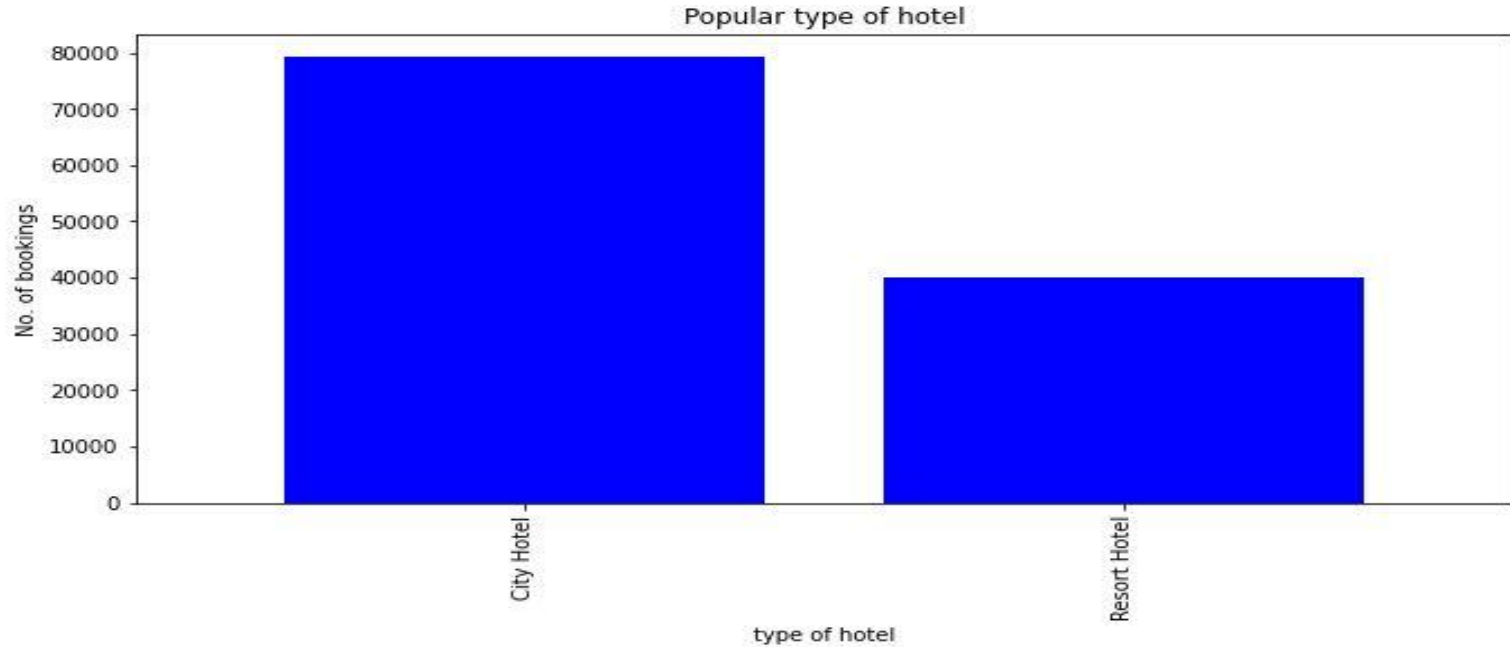


- **Month of august has the highest number of cancelation**
- **Although November and December has almost same bookings ,cancelation rate in December is more**
- **There is negligible change in cancelation between September and October**

# CANCELATION BY TYPE OF HOTEL



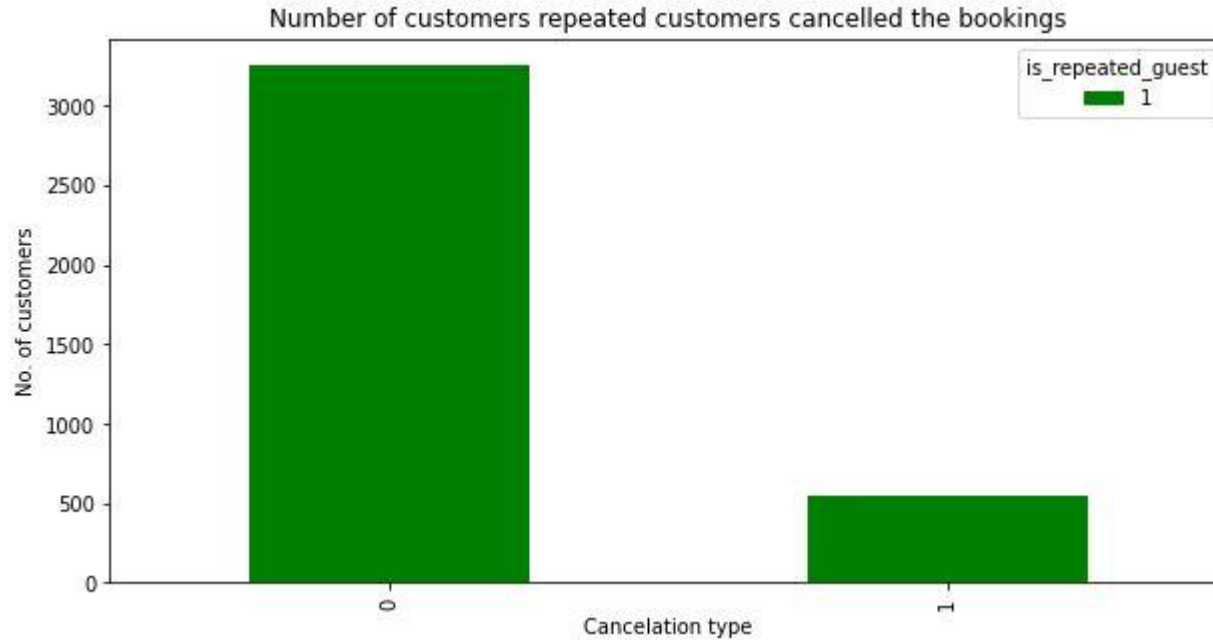
# TYPE OF HOTEL PREFERRED



popularity of the city hotel is 2 times more than the resort type of hotel

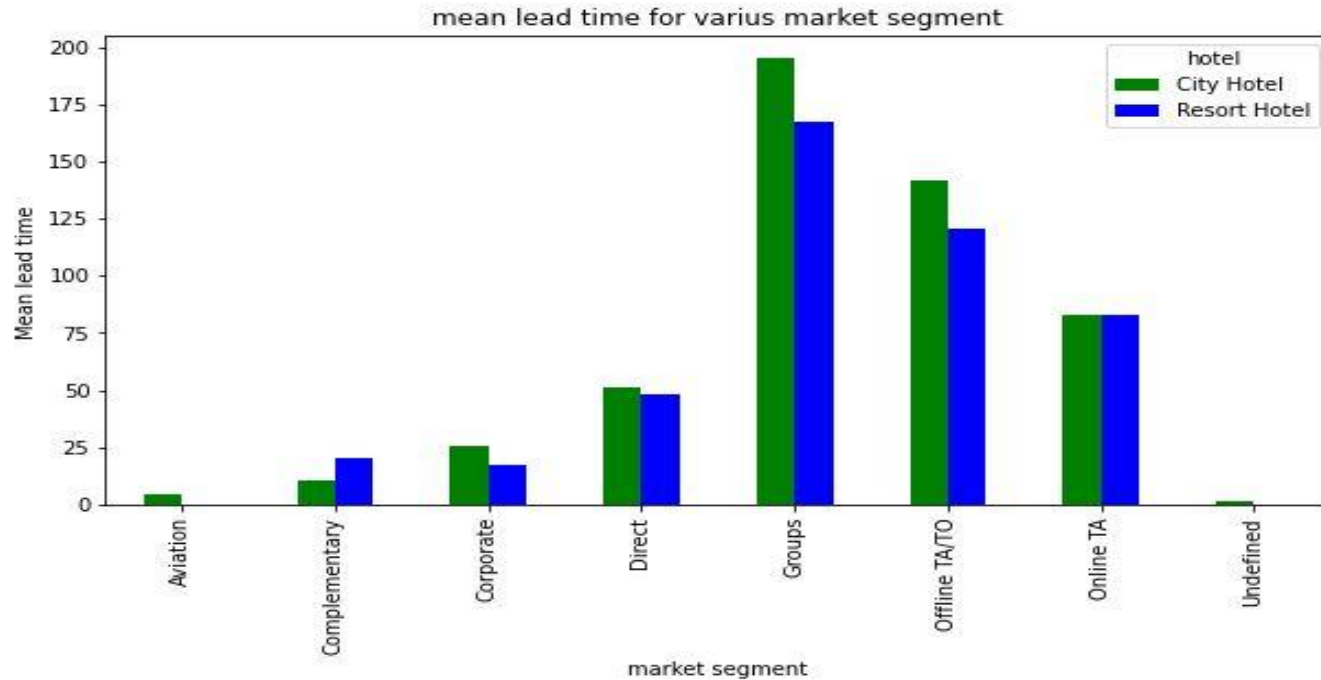


# CANCELATION BY THE REPEATED CUSTOMER

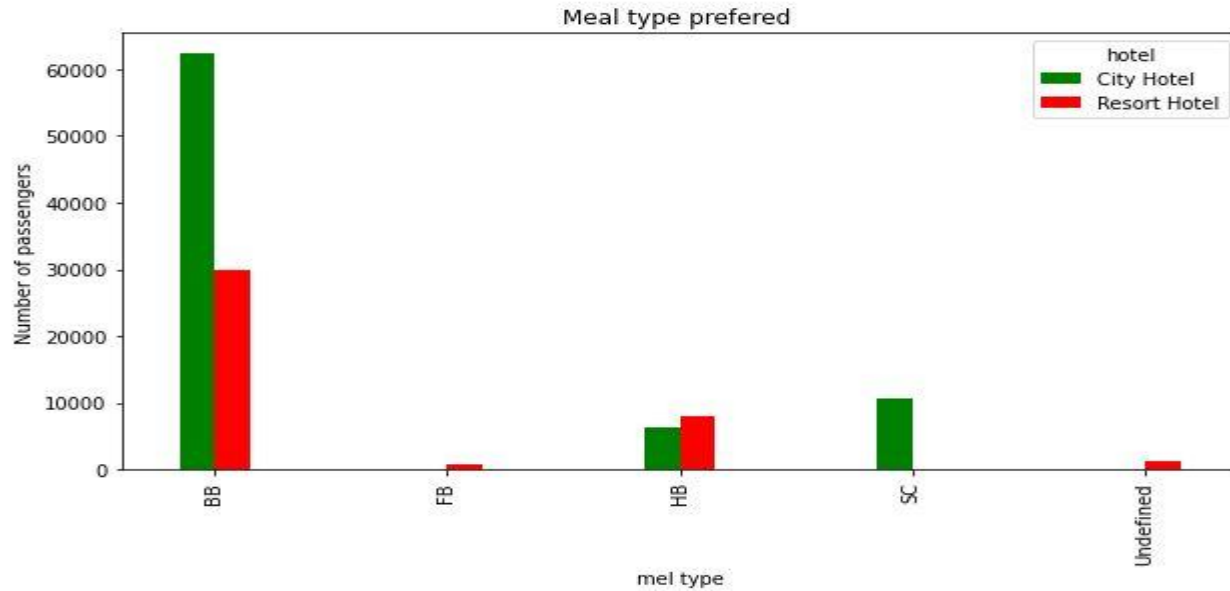


- Most of the Repeated customers did not place cancellation request

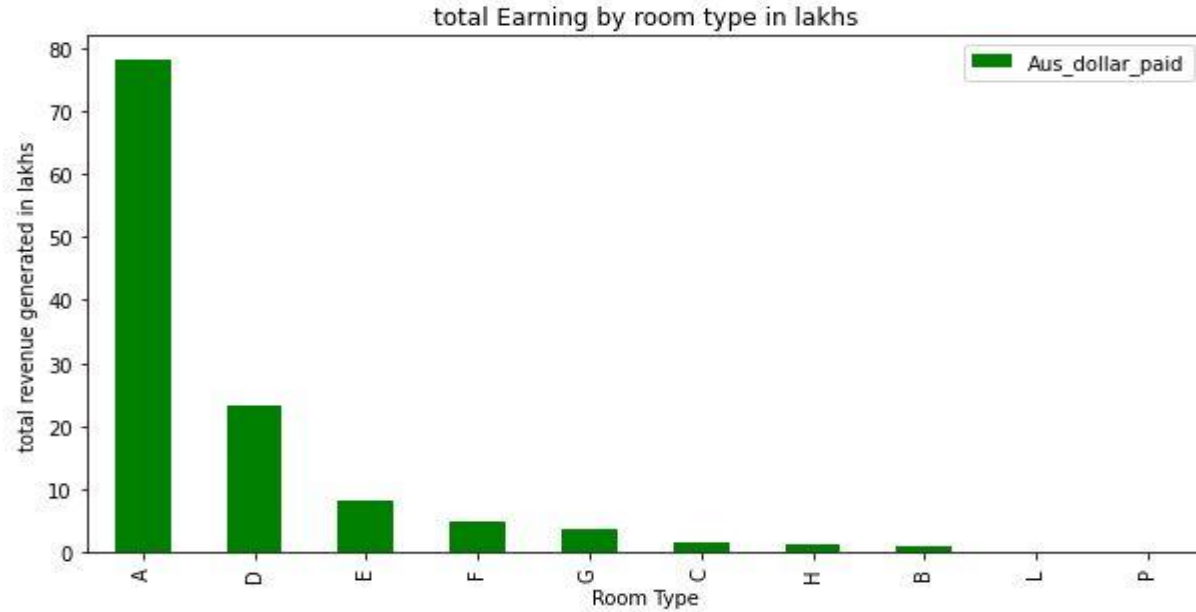
# LEAD TIME FOR VARIOUS SEGMENTS



# TYPE OF MEAL PREFERRED

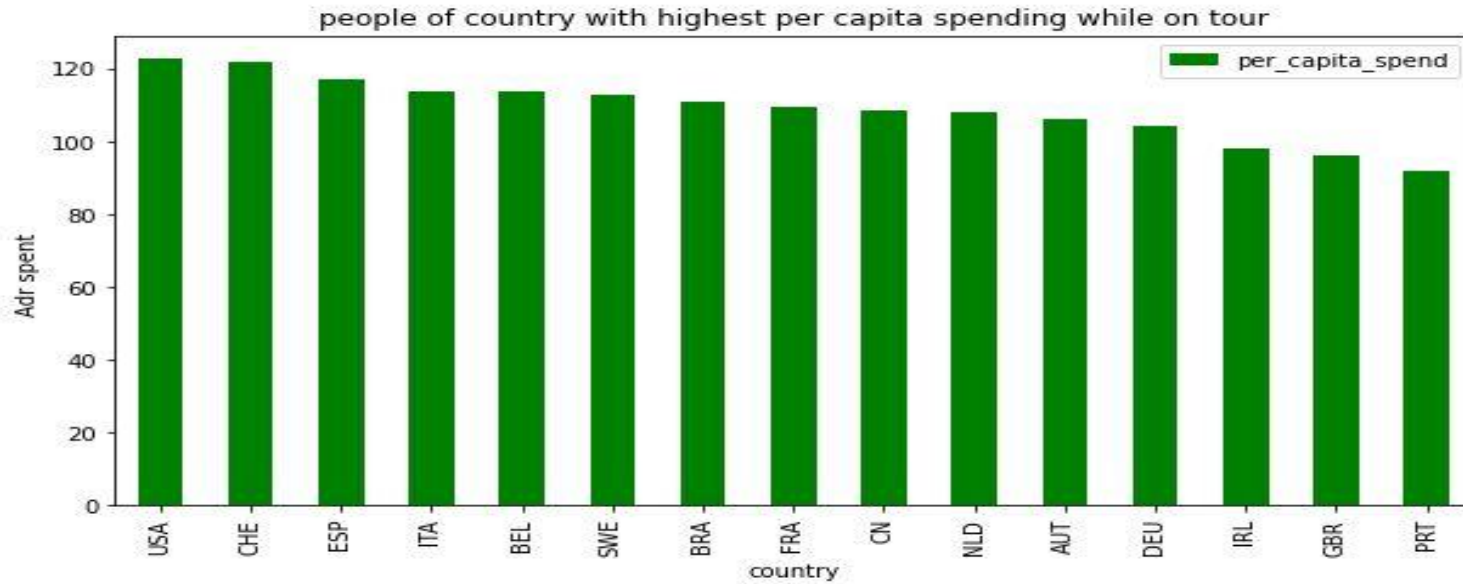


# TOTAL EARNING PER ROOM FOR GIVEN DATA



- Room type A is most preferred and while B,L and P are least preferred

# SPENDING PER BOOKING BY COUNTRY WISE



- **Guests from USA, Switzerland and France spend more on rooms**
- **While guest from Portugal and great Britain prefer economical rooms**

## Conclusion:-

1. We have seen that most of the guests were from western Europe countries
2. Portugal is the country from where almost 50k bookings were being done.
3. Mostly the city hotels are reserved.
4. Repeated customers are very less.
5. Mostly the rooms which are reserved converted to successful transaction.
6. Guests from USA, Switzerland and France spend more on rooms While guest from Portugal and great Britain prefer economical rooms.
7. Month of august has the highest number of cancelation.

**THANK YOU**