

# Capstone Project

AI

## EDA -Hotel Booking Analysis



- Poonam Shevkar  
Sanjay Verma  
(AlmaBetter Trainee)

# Introduction

Hotel Booking Analysis dataset contains booking information for a city hotel and a resort hotel in different Countries, and includes information such as when the booking was made, length of stay, chosen meals, price etc. There are 119390 observations and 32 features in dataset.

In this project, we will try to predict the possibility of a booking for a hotel based on different factors and also try to predict if they need special requests based on different features.

From it, we can understand the customer's behaviour and it might help us make better decisions.

# Scope of the Project-

To perform meaningful analysis as -

- ❖ Data Prepping/ Wrangling.
- ❖ Best time for Hotel Bookings.
- ❖ Price For Hotel Bookings.
- ❖ Special Requests from Customers.
- ❖ Country wise Hotel Bookings.
- ❖ Meal Preference by Customer.
- ❖ Hotel Bookings/Cancellations.
- ❖ Correlation and Scatter Plot.



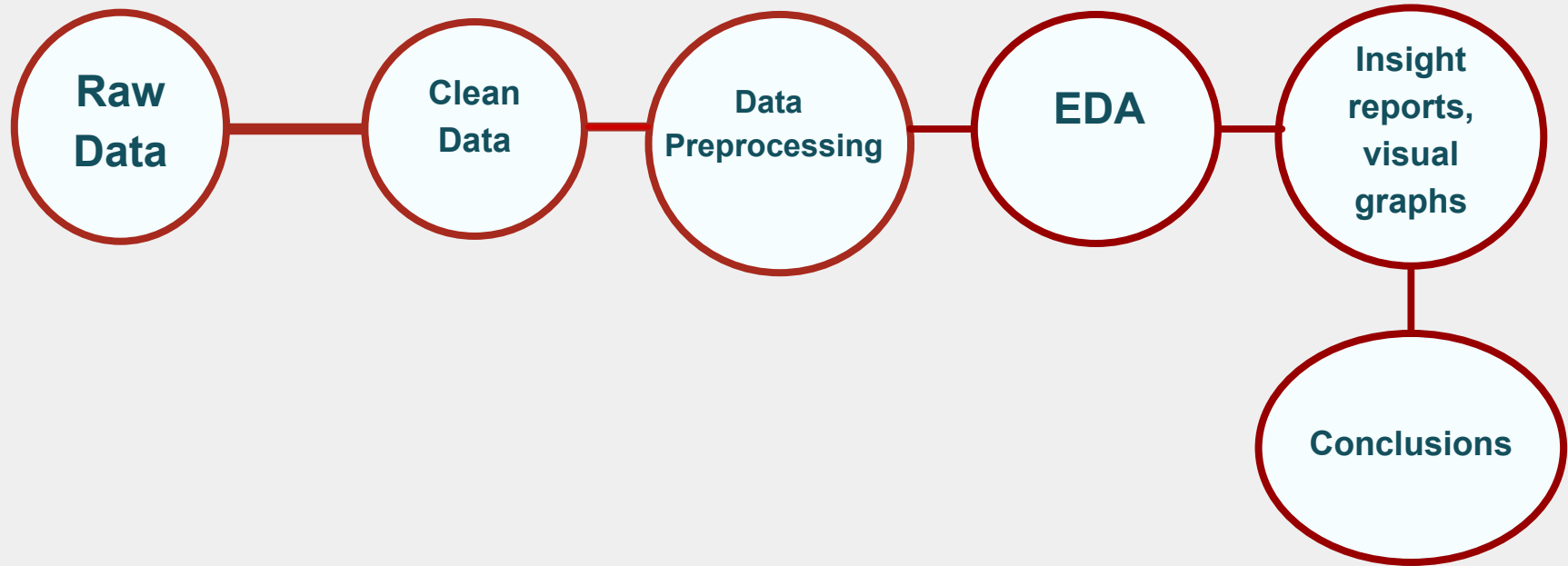
# Problem Statement-

- Have you ever wondered when the best time of year to book a hotel room is?
- 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!**

# Process Flow-



# Data Preprocessing -



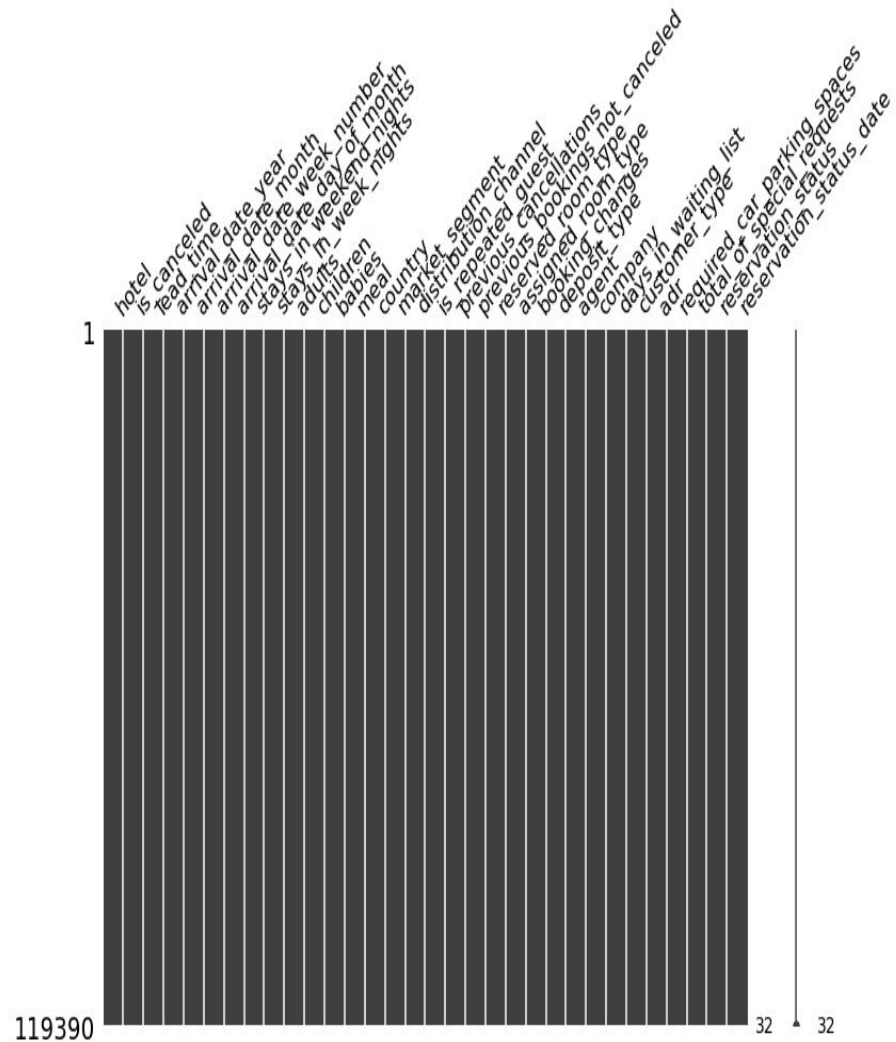
## Clean Data:-

If Null values are present in the dataset then chances of errors are maximum.

To overcome this, we have to remove/replace the Null values in data set for features name as- Children, Country, Agent and Company. Also, Company has huge Null values in data set.

## Libraries used-

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Plotly
- Missingno



# Exploratory Data Analysis - Visualization and Descriptive Statistics

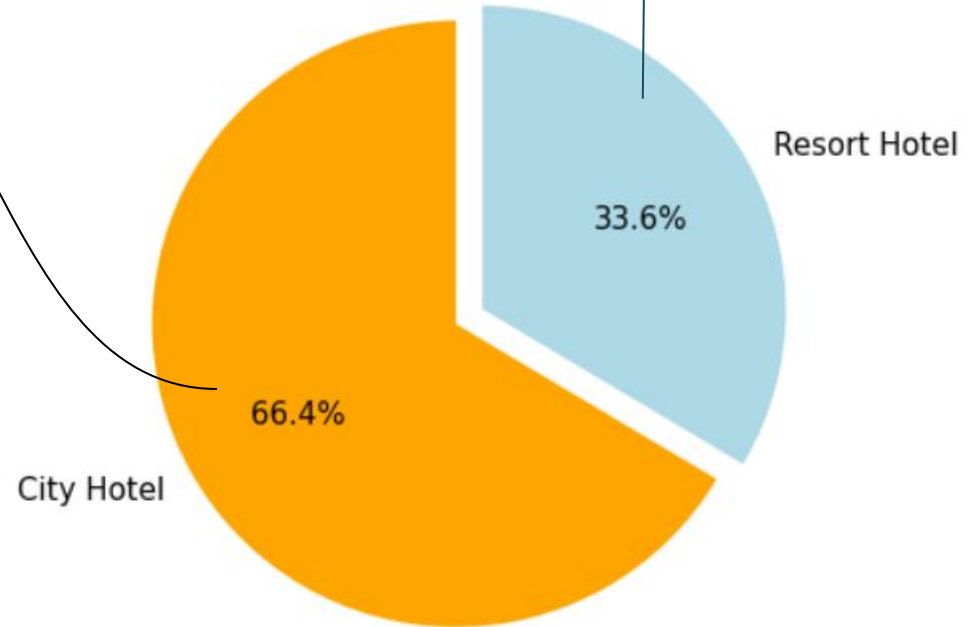




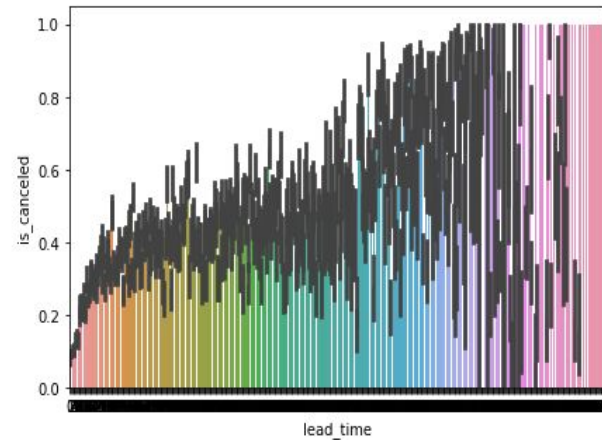
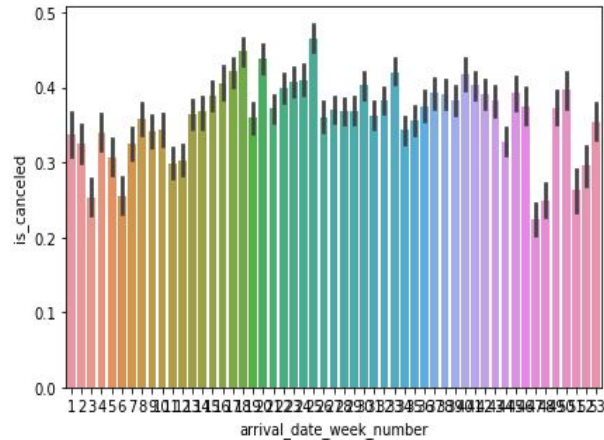
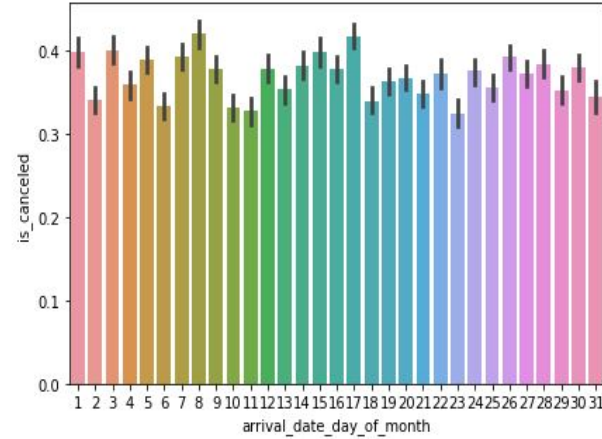
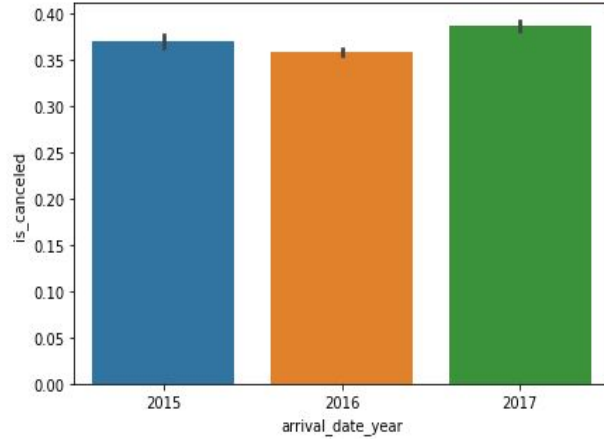
# Percentage of Hotel bookings-

Cheap than resort hotels, hence more bookings.

Expensive and preferred on weekends.



## Visualization of Hotel bookings in different periods

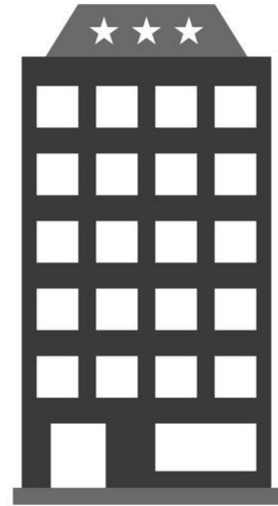


↑ Cancellation due to Lead time

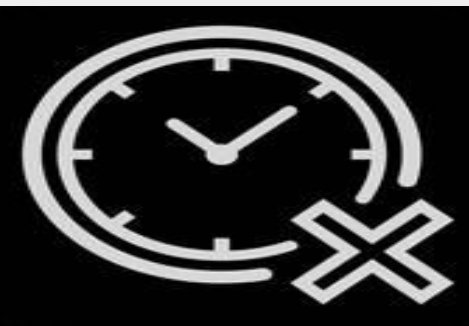
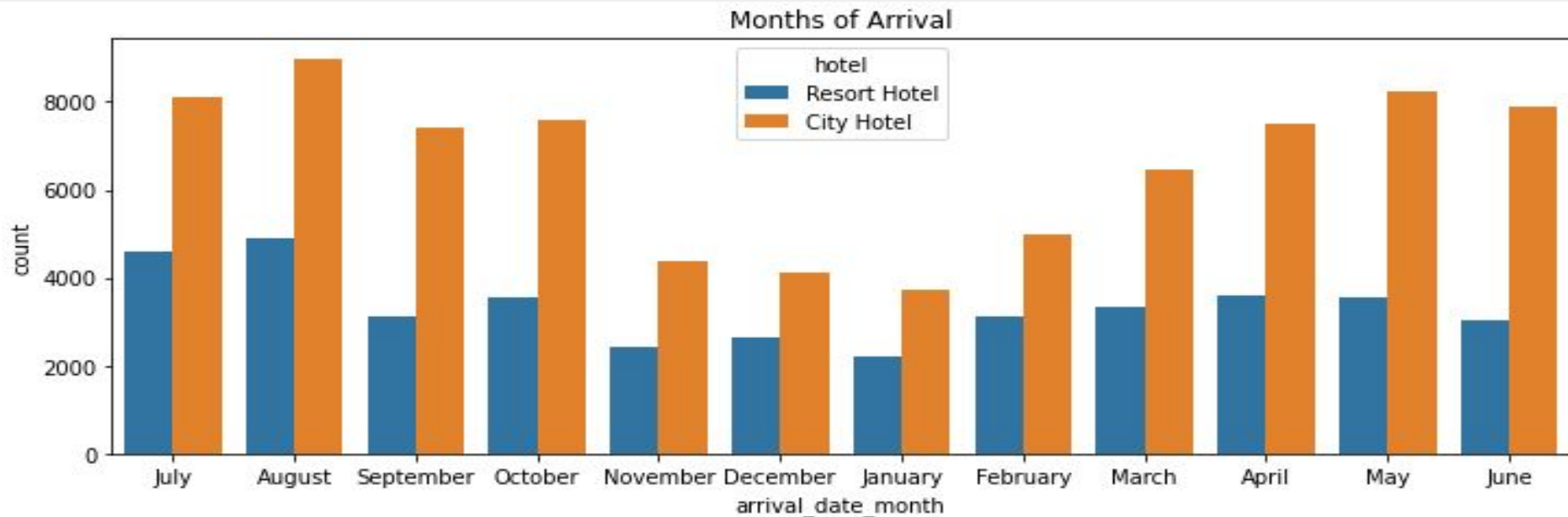
↑ Booking in 2017

Months and weeks bookings.

# Best Time for Hotel Booking



# Best time of year to book a hotel room -



*January*

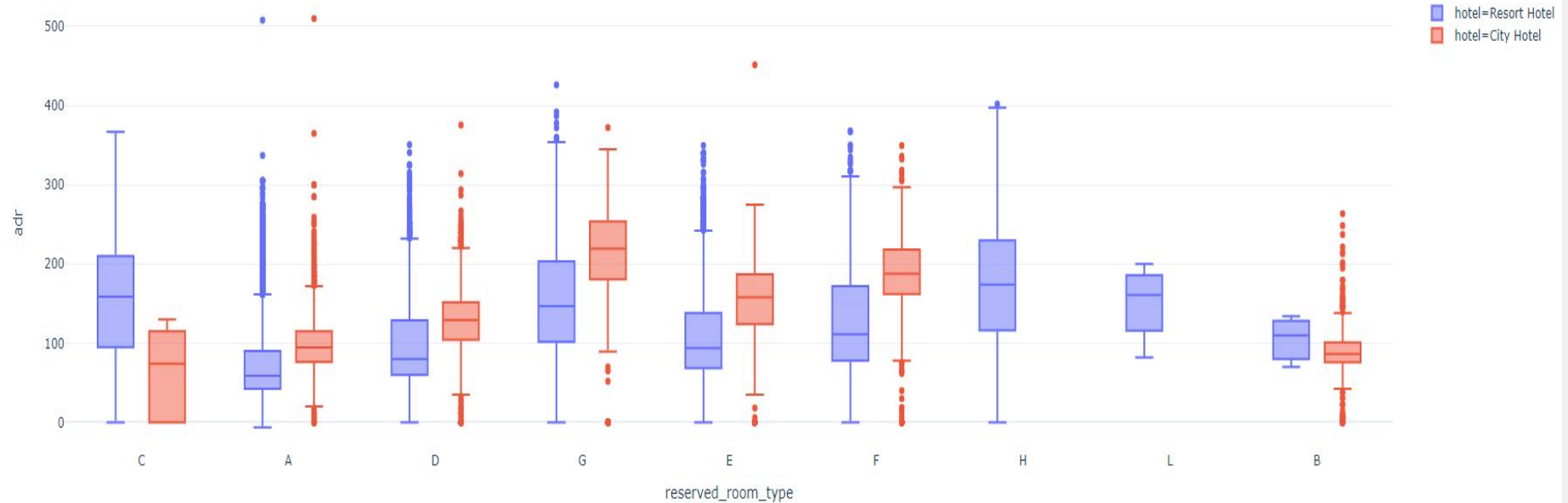
*August*



# Rates of Hotel & Stay at Hotel

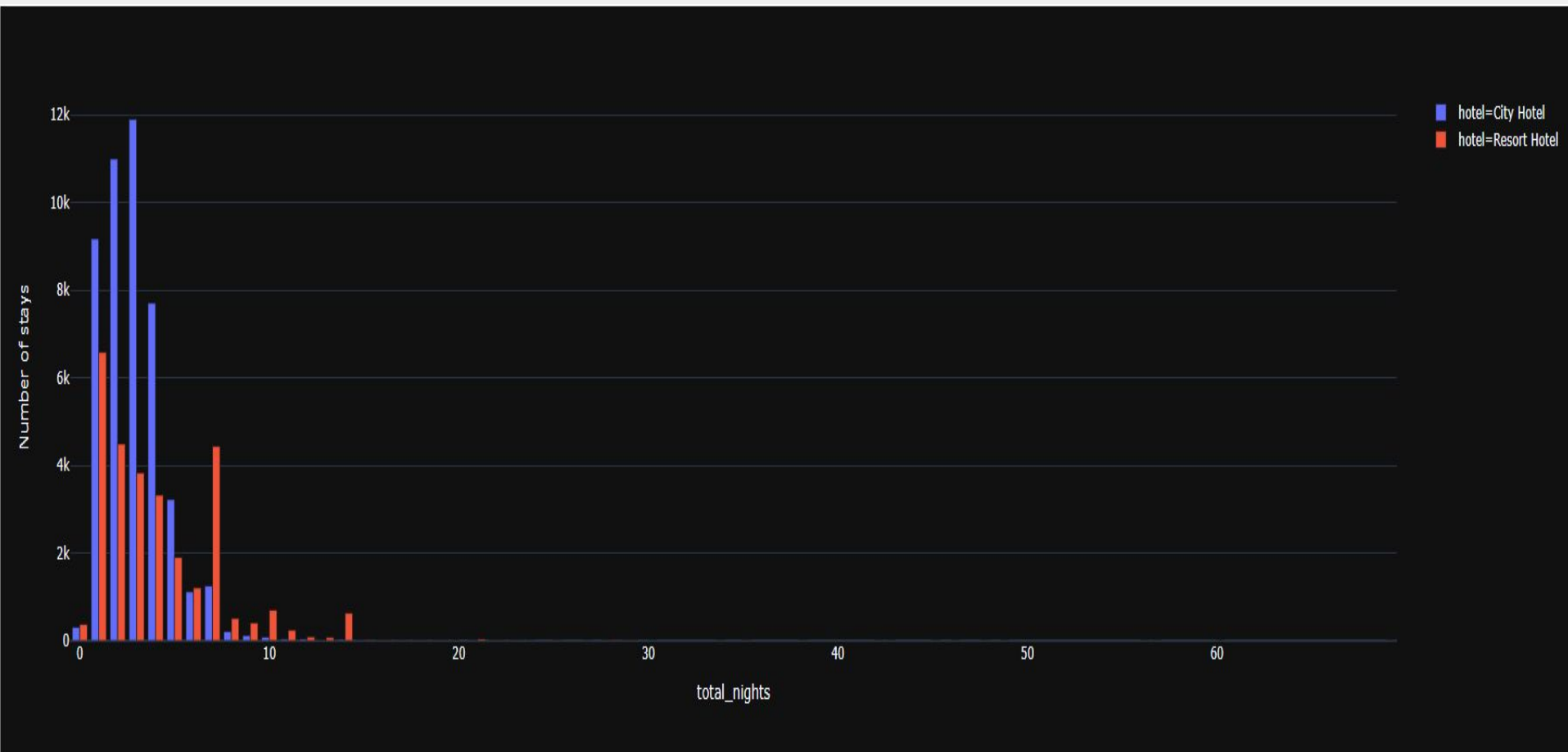


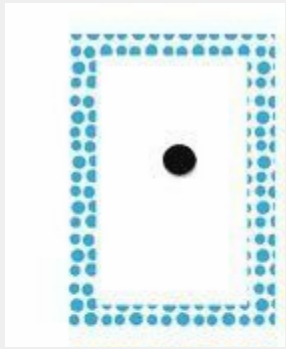
# Rates of Hotel-



- **Guests pay for a room per night-**  
*its type and the standard deviation*

# Lets observe the stays of people at the hotels

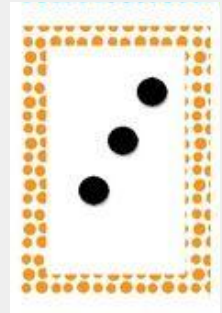




City Hotels



Resort Hotels



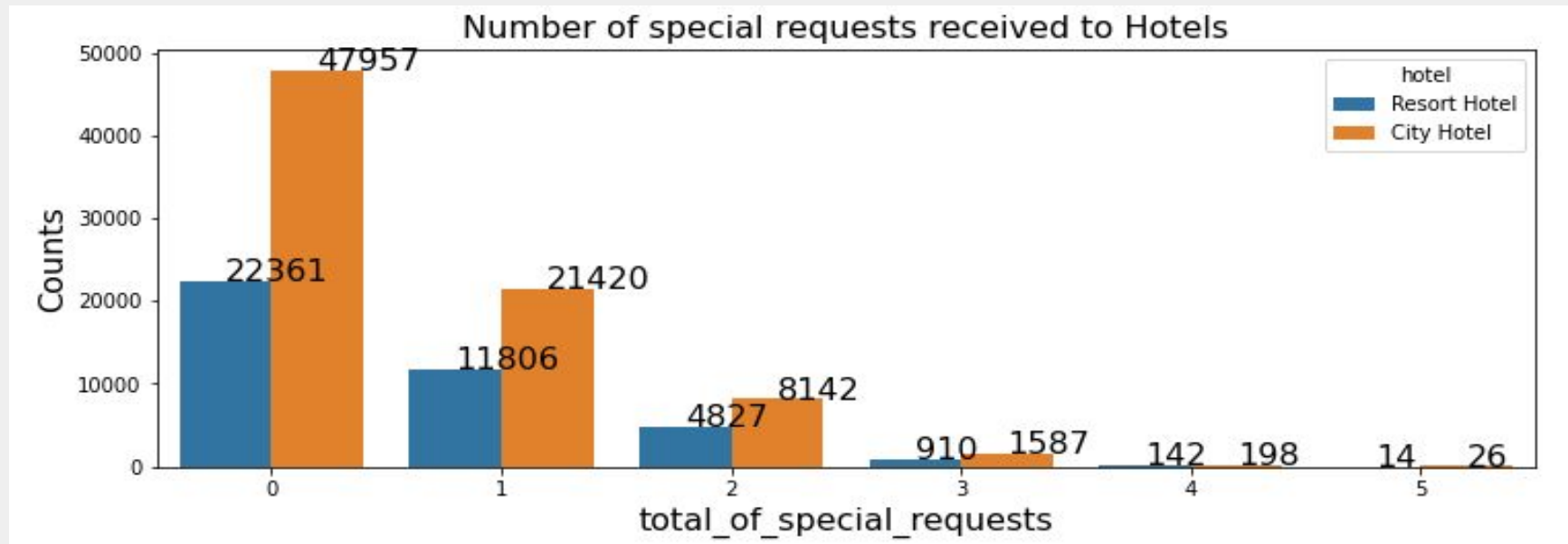


# Special Requests



Customer Requests

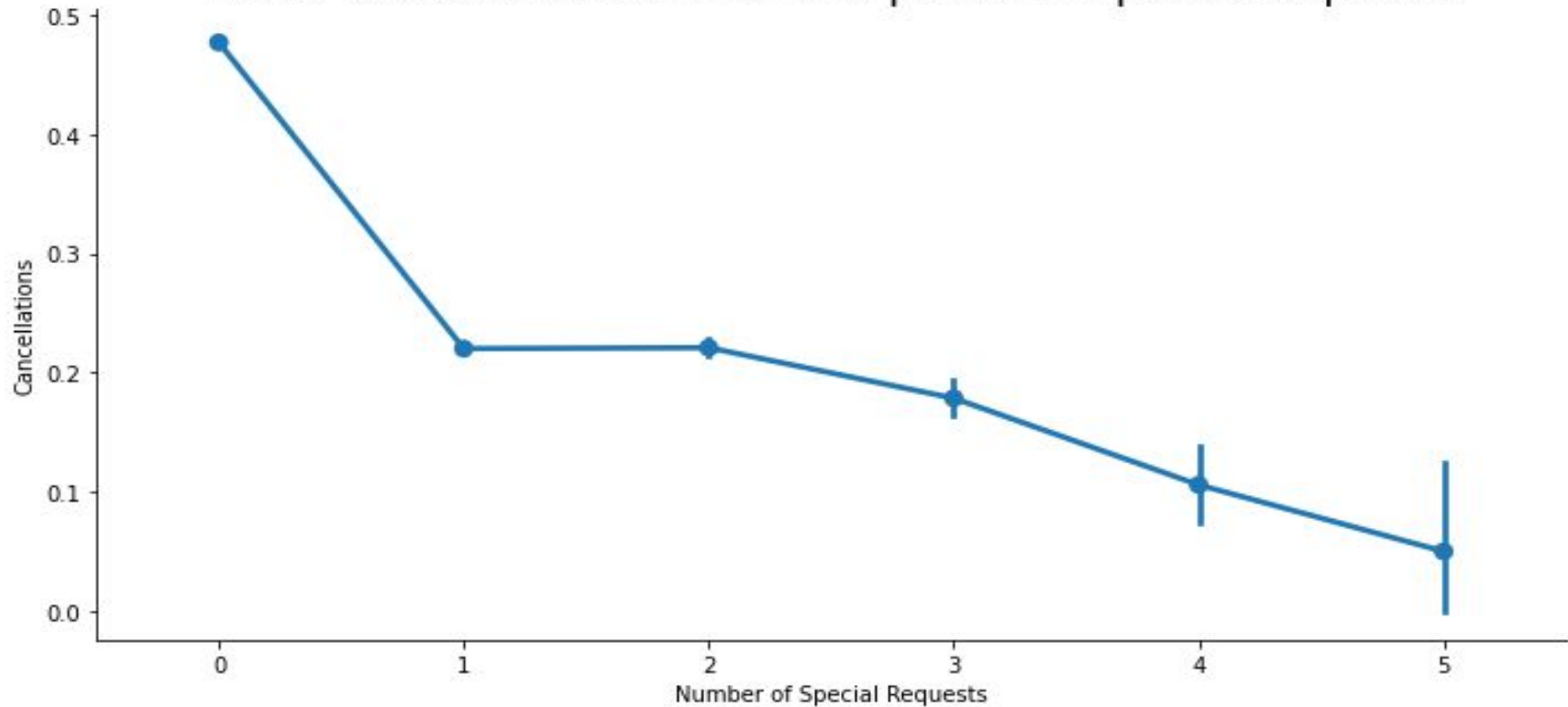
# Special Requests received from guests to Hotels-



*special request*



# Hotel Cancellation due to overpass the special requests



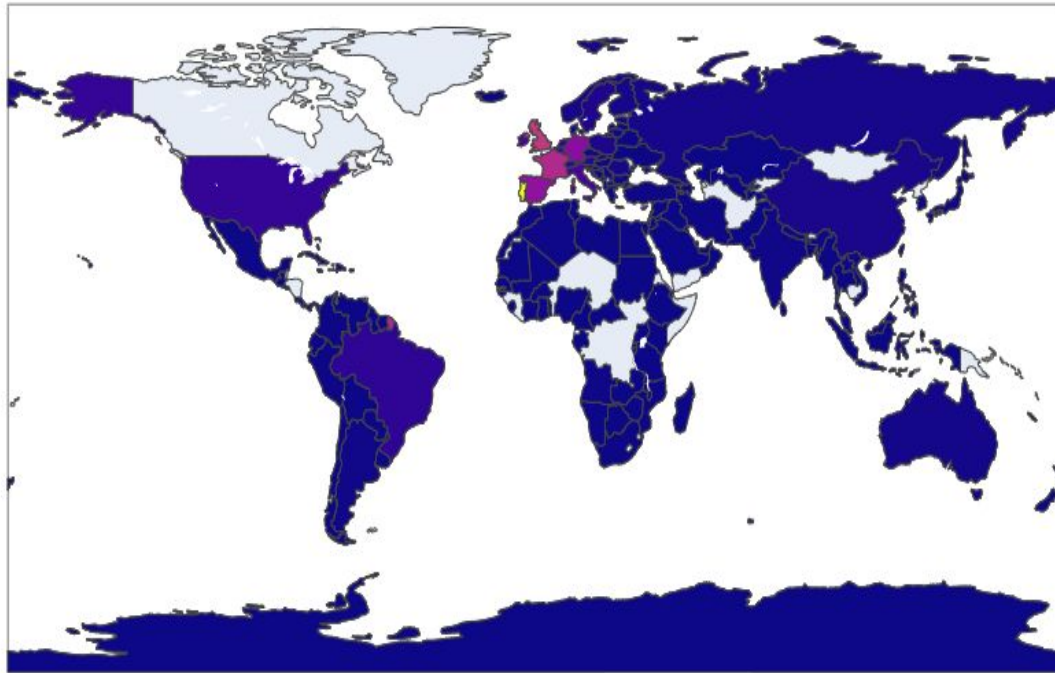
***High negative total of special requests with cancellation. My hypothesis is that when a customer has more special requests and they are met, the customer is more likely to stay with the hotel***

# Country Bookings





# Visualization of countries with the no. of guests-



No of guests

20k

15k

10k

5k

0

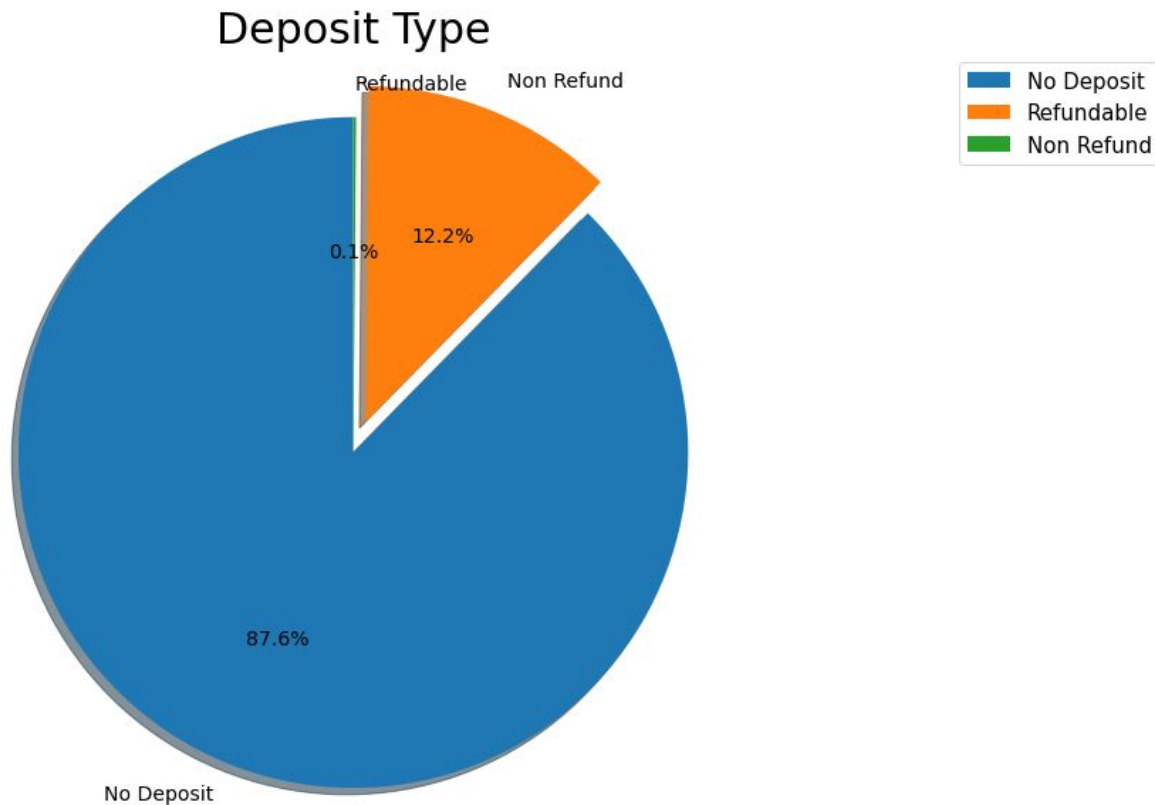
*People from all over the world are staying in City and Resort hotel. Most guests are from Portugal and other countries in Europe.*



# Deposit Type



# we will observe the Deposit type for hotels-

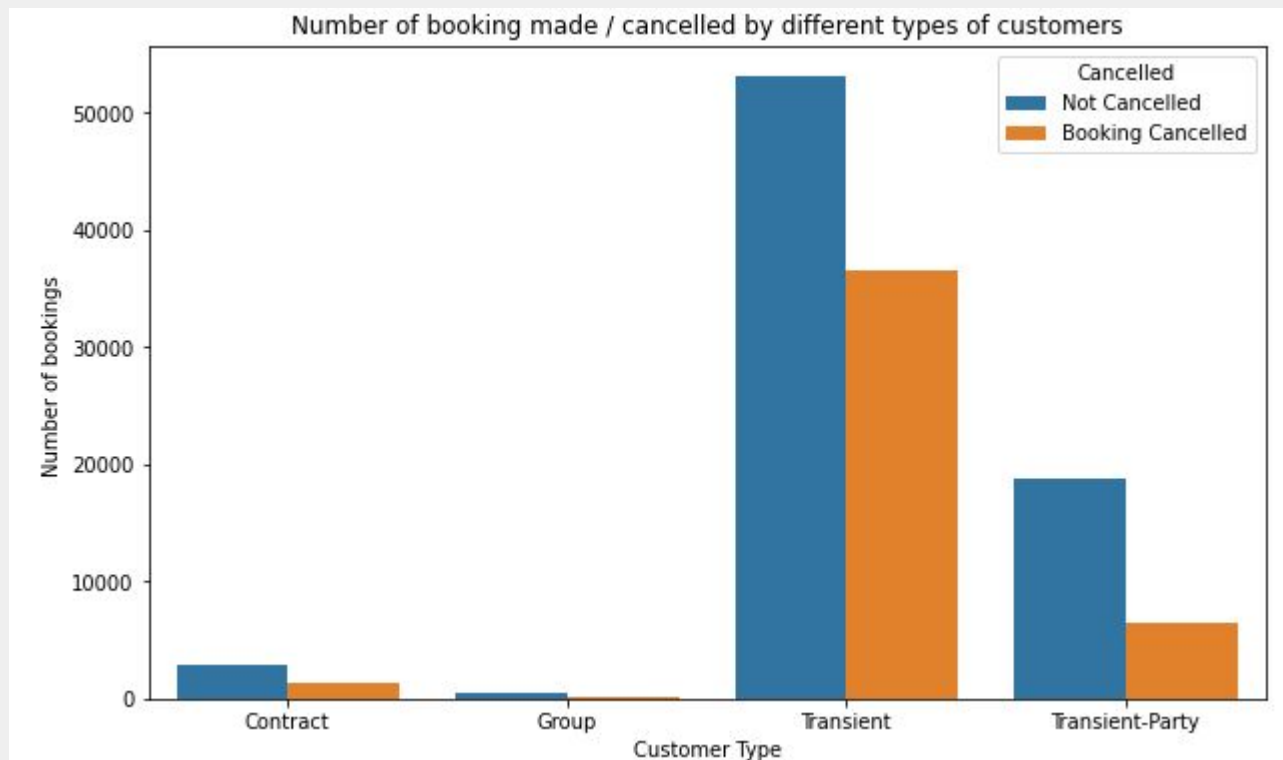






# Booking/Cancelled Bookings

# Bookings / Cancelled Bookings by different types of customers-



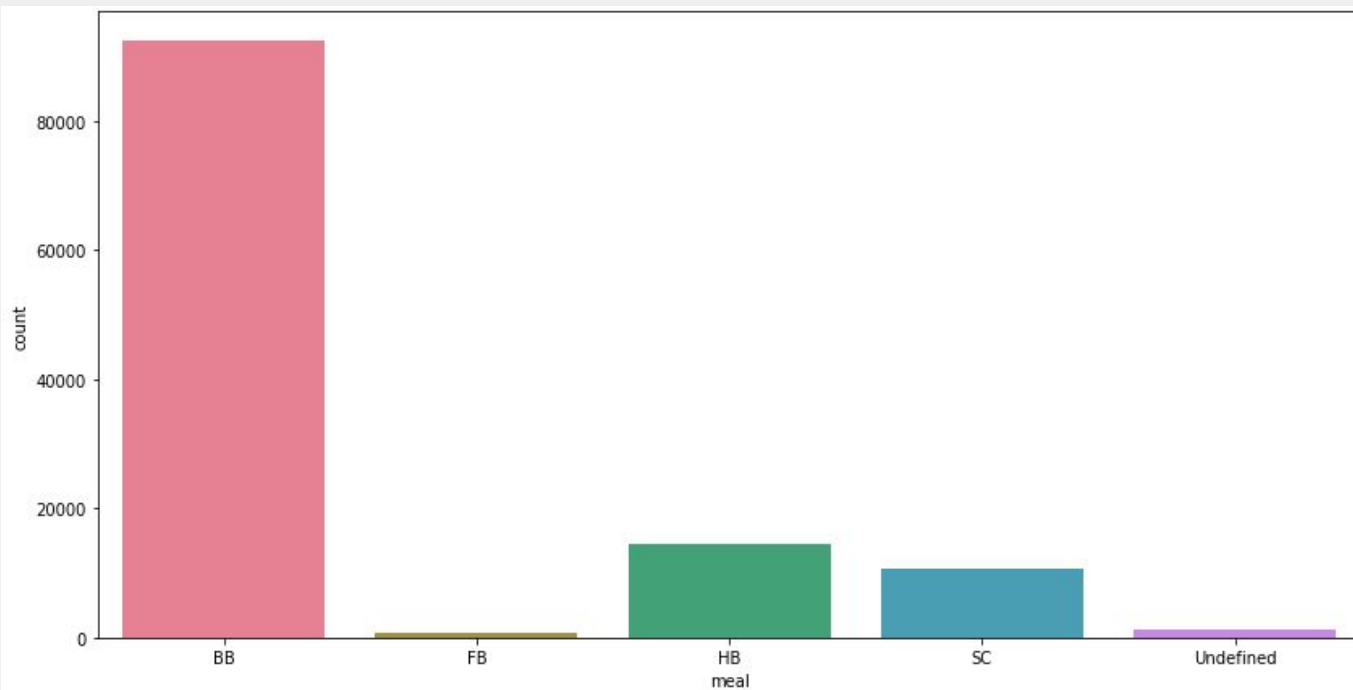
***Transient and Transient-party have more Bookings / Cancelled Bookings than Contract & Group.***

# Meal Preference



# Meal preference by customers-

***The Bed & Breakfast : 77.26%***



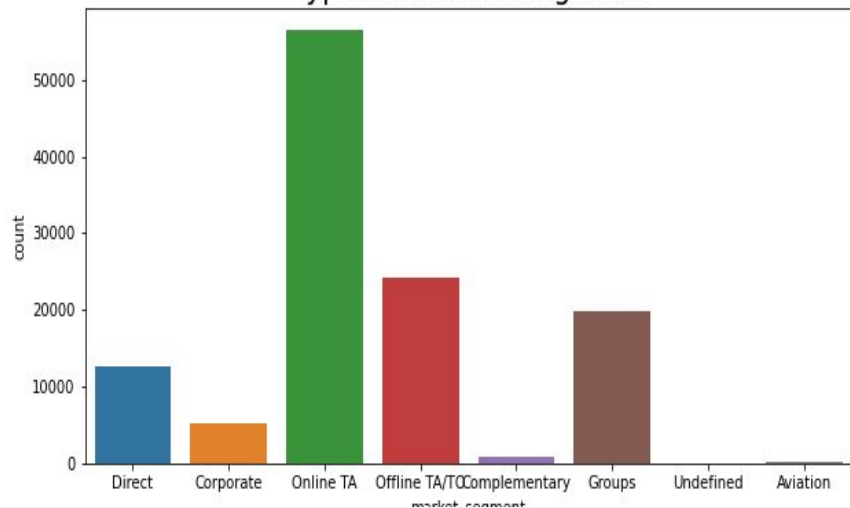
***Here, we observe that BB i.e Breakfast Bed Meals are preferred more by guests and undefined is less chosen by guests.***

# Market Segment & Distribution Channel

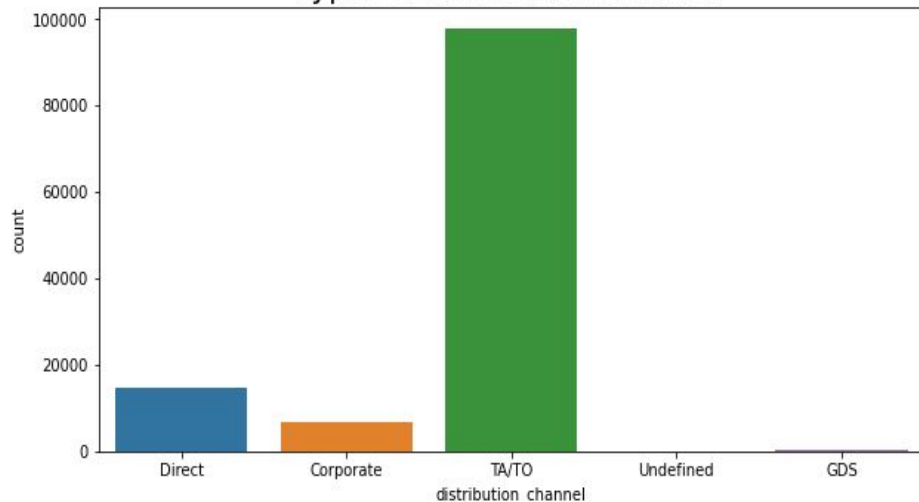


# Overview for Market Segments and Distribution Channel For Hotel Booking Publicity-

Types of market segment

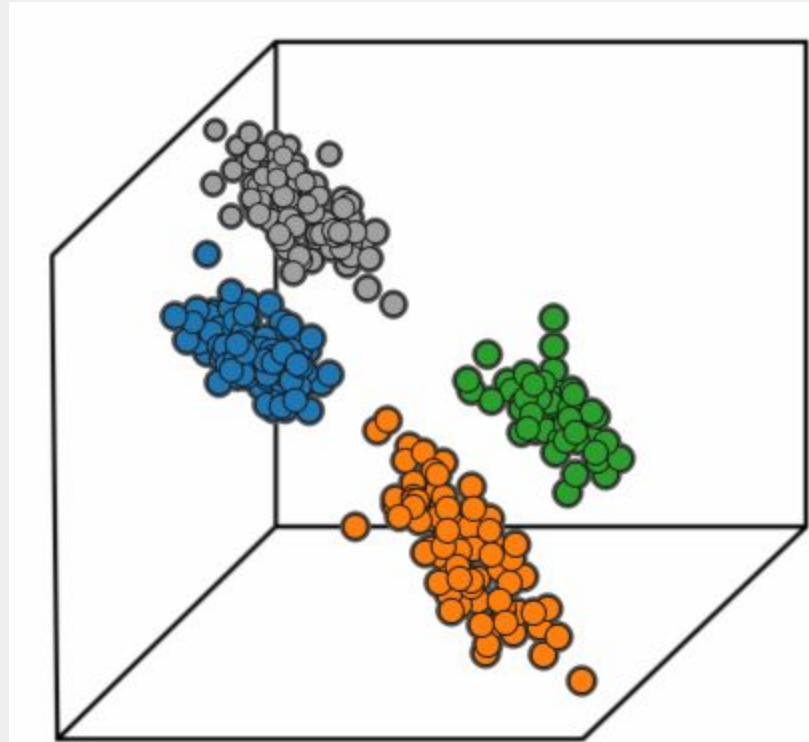


Types of distribution channel



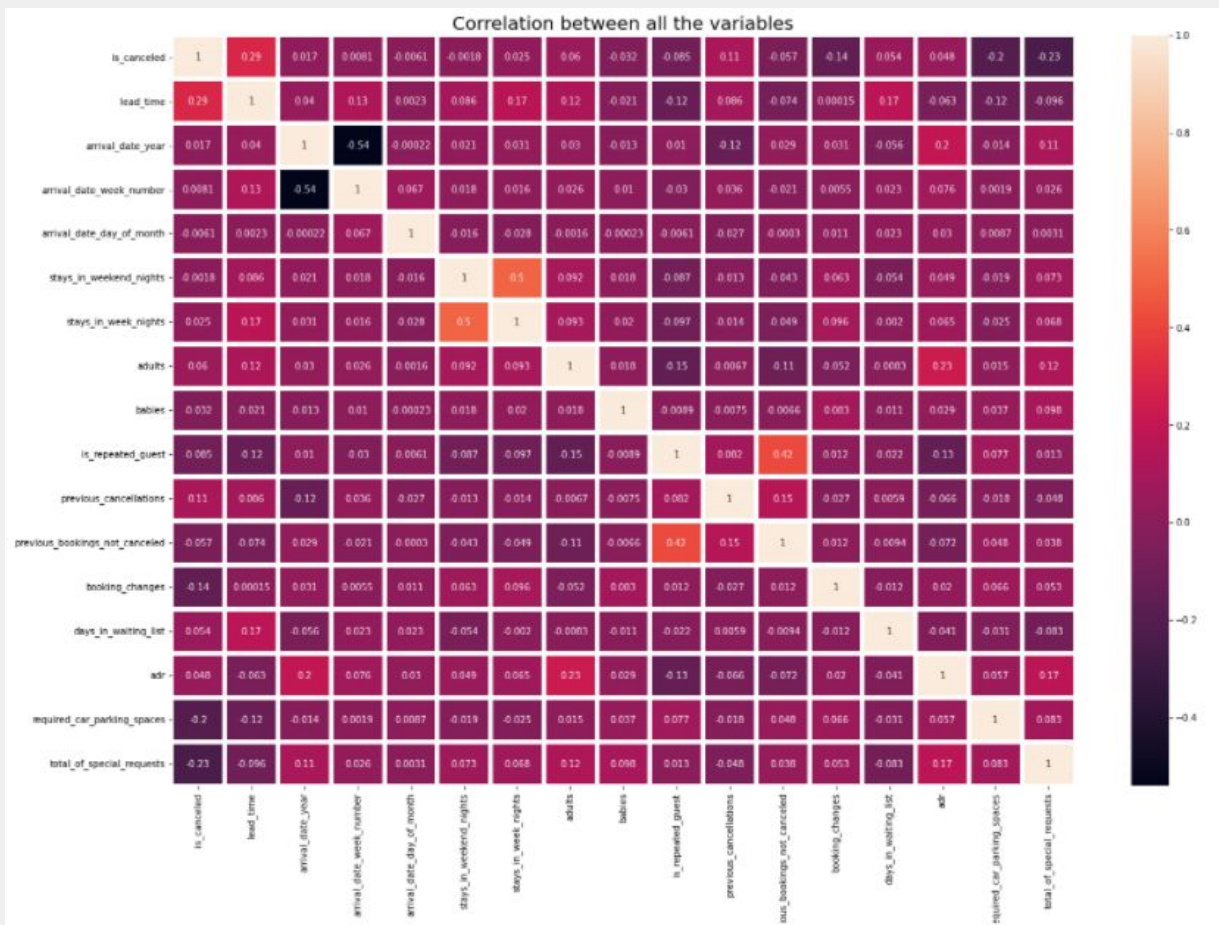
Majority of the distribution channels and market segments involve travel agencies (online or offline). We can target our marketing area to be on these travel agencies website and work with them since majority of the visitors tend to reach out to them.

# Correlation and scatter plot





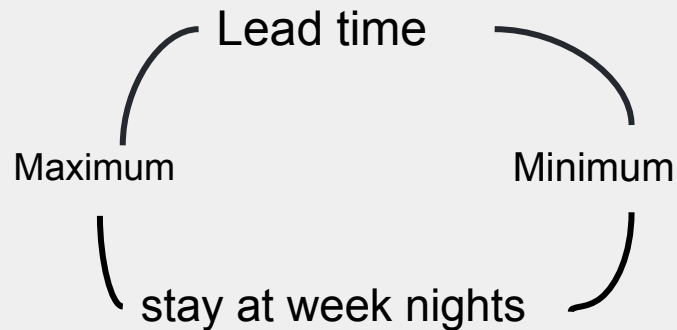
# Heat Map to visualize the correlated variables-



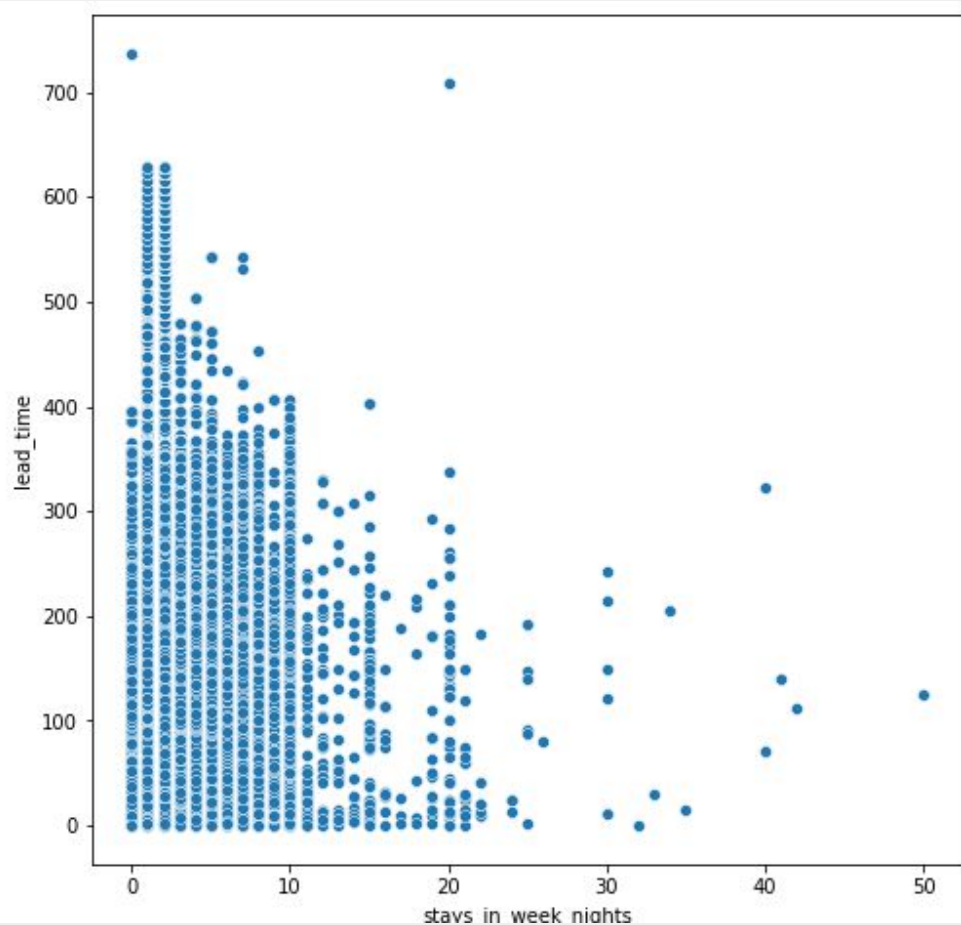
*The first 5 variables (lead\_time, total\_of\_special\_requests, required\_car\_parking\_spaces, booking\_changes, previous\_cancellations) can be used for building a model in future, since they are the ones with correlation > 0.10 for is\_canceled.*



# Demonstration of the relationship between the variables-



**Negative  
correlation**



# Conclusions-

1. More than 60% of the population booked the City hotel.
2. Most bookings were made from July to August. And the least bookings were made at the start and end of the year.
3. Average daily night per person price is 94.95 in resort Hotel whereas 105.30 in City Hotel.
4. Special Requests from guests increases cancellation rate decreases.
5. Portugal, UK, France, Spain and Germany are the top countries from most guests come, more than 80% come from these 5 countries.
6. Most people stay for 1 night at City Hotel and 3 night at Resort Hotel.
7. Customer preferred Bed & Breakfast Meal in maximum quantity.
8. Majority of bookings did not require a deposit, this tells the high cancellation rate.
9. Transient & Transient-party have more bookings including cancelled bookings than contract & group.
10. 5 Correlated variables are seen for is\_cancelled feature in dataset and negative pattern of Scatter Plot for lead time & stays in week nights.

Thank you!!