

Capstone Project
Hotel Booking Analysis
By
Syed Sharin
Fathima.K
Arman Alam

Points for Discussion

- Data Summary
- The price for different hotel types each month
- Waiting time for Hotel booking with respect to market segment
- Type of Hotel preferred by the customers
- Busiest month
- Bookings made by each hotel every month
- Most preferred business segment
- Home country of Guests
- Bookings per market segment
- Adults and kids guests arrived per month in each hotels
- Conclusion

Exploratory Data Analysis

Data science is often thought to consist of advanced statistical and machine Learning techniques. However, another key component to any data science endeavor is often undervalued or forgotten: exploratory data analysis (EDA). It is a classical and under-utilized approach that helps you quickly build a relationship with the new data.

We will discuss exploratory data analysis and data visualization of the hotel booking data set. The goal of this project is to analyse which type of customer prefer which type of hotel, or which type of hotel is mostly preferred by customer, during which month the hotels make maximum business, which are the least occupied months of the hotel, which is the most preferred business segment, average daily rate for hotels

Data Summary

For this EDA project, we are using the "Hotel booking demand" database.

This data set contains the reservation details of two hotels namely the city hotel and the resort hotel, and includes details such as when bookings are made, duration of stay, number of adults, children, and the number of available parking spaces, among other things.

The following steps are performed on the Data set:

Step 1. Overview the whole data

Step 2. Identification of variables and data types.

Look at the number of rows and columns of the dataset and Look at each column information that is data type, null rows.

Step 3. Looking at missing data

Step 4. Outlier Treatment

Step 5. Corelation analysis

Step 6. Splitting the dataset into numerical and categorical(String)

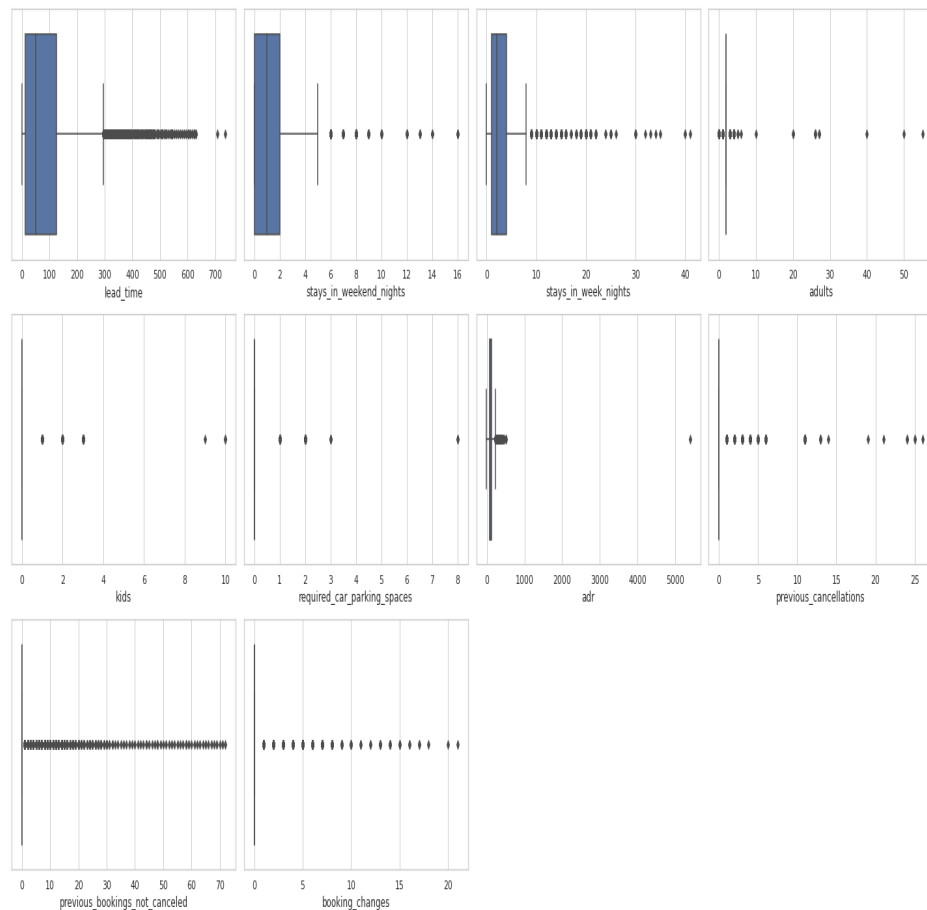
Step 7. Look at data distribution for numerical columns

Step 8: Look at Data Characteristics of text variables

Step 9: Conclusion

Exploratory Data Analysis (EDA)

- **Data Preparation and Cleaning:** Importing the libraries needed in this project for data manipulation , cleaning and visualization.
- Import pandas as pd, import numpy as np, import seaborn as sns, from matplotlib import pyplot as plt.
- Next will download the dataset to use it as pandas dataframe.
- Will check how many rows are there in dataset , `len(df.index)` and how many cells are missing, `df.isnull().sum()`
- we found four columns with missing values
- But Column agents and column companies has many missing values making the columns irrelevant for our analysis , hence removing them.



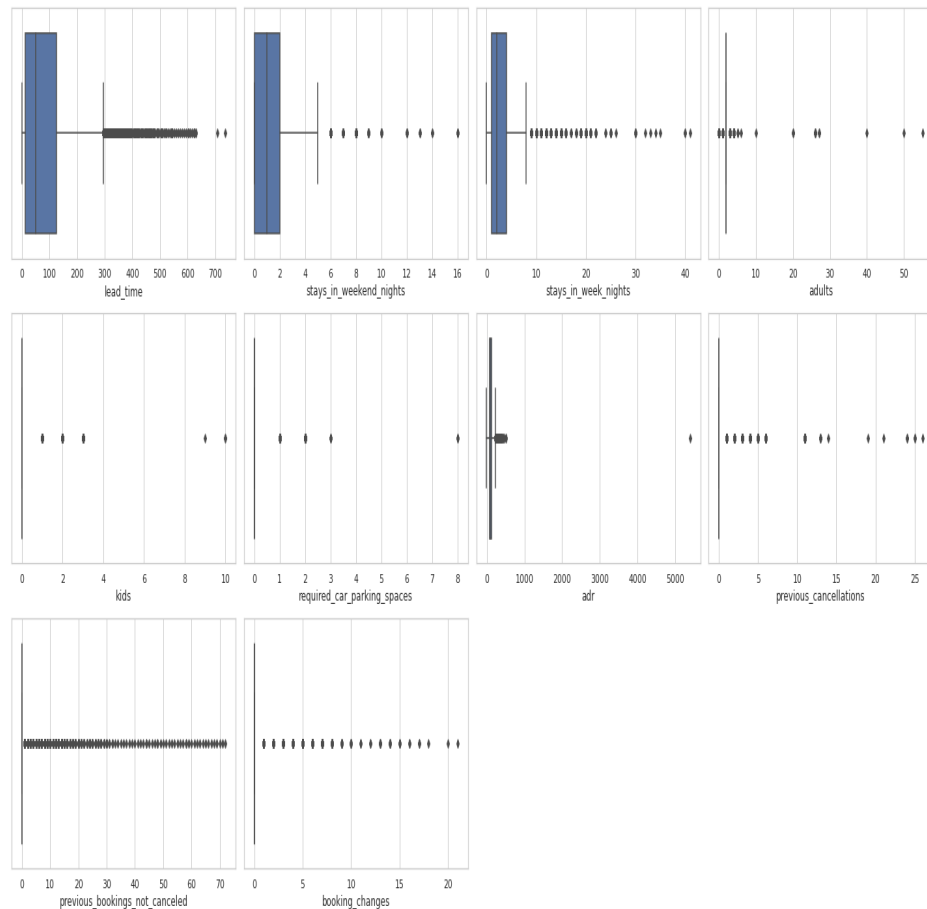
Exploratory Data Analysis (EDA)

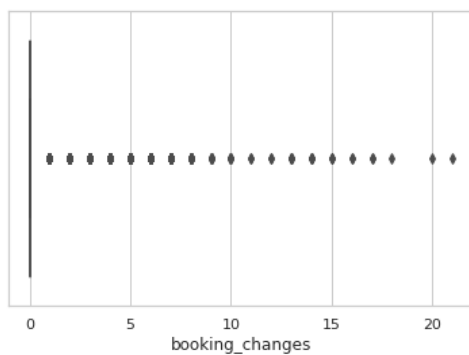
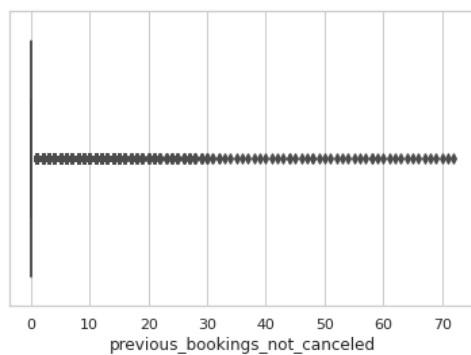
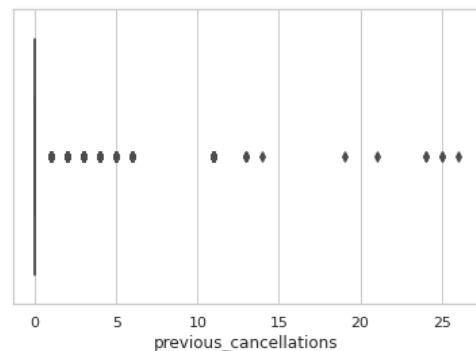
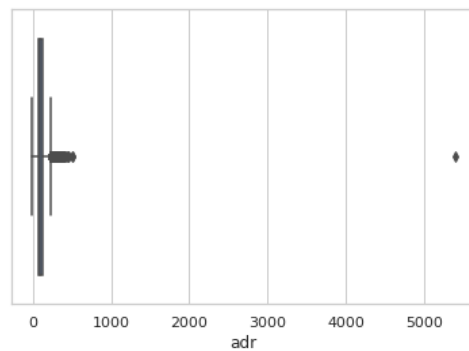
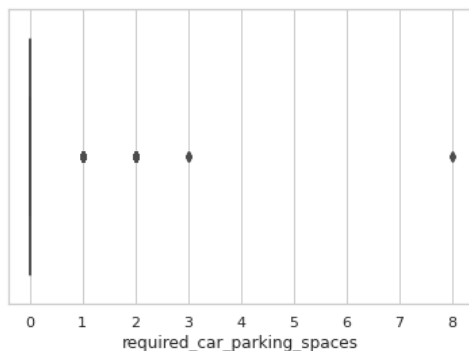
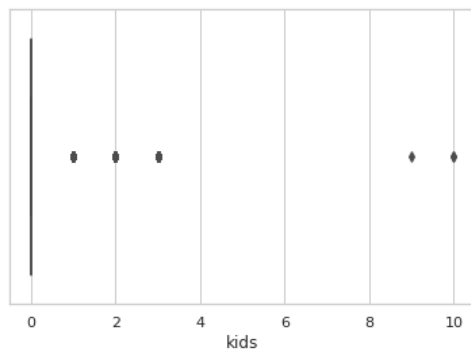
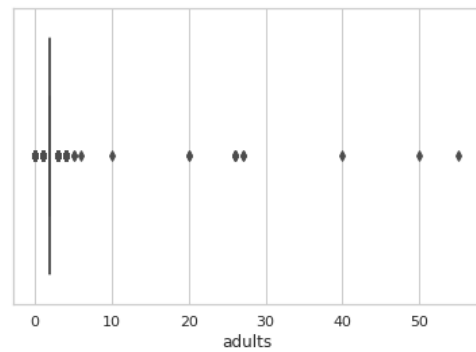
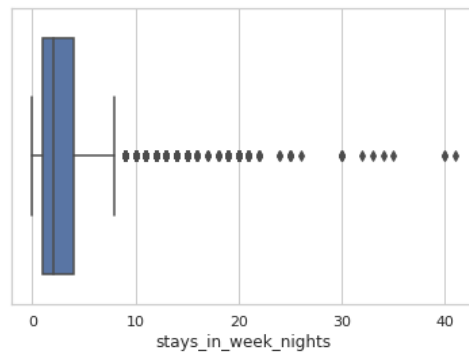
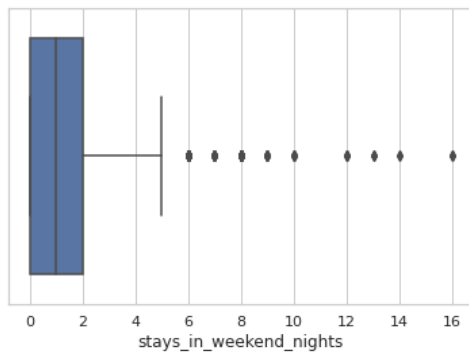
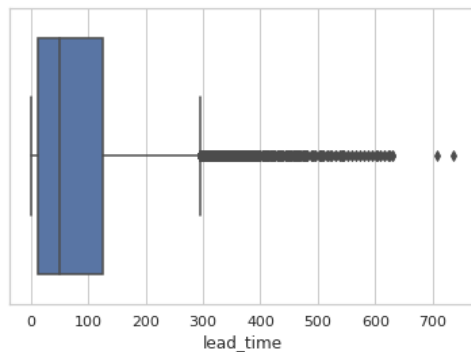
- Finding all Null and Duplicate values & dropping them, `df.dropna(inplace=True)`
After doing the data cleaning we are left with 86913 rows and 29 columns.

Now taking a closer look `df.describe()`

We can see some outliers.

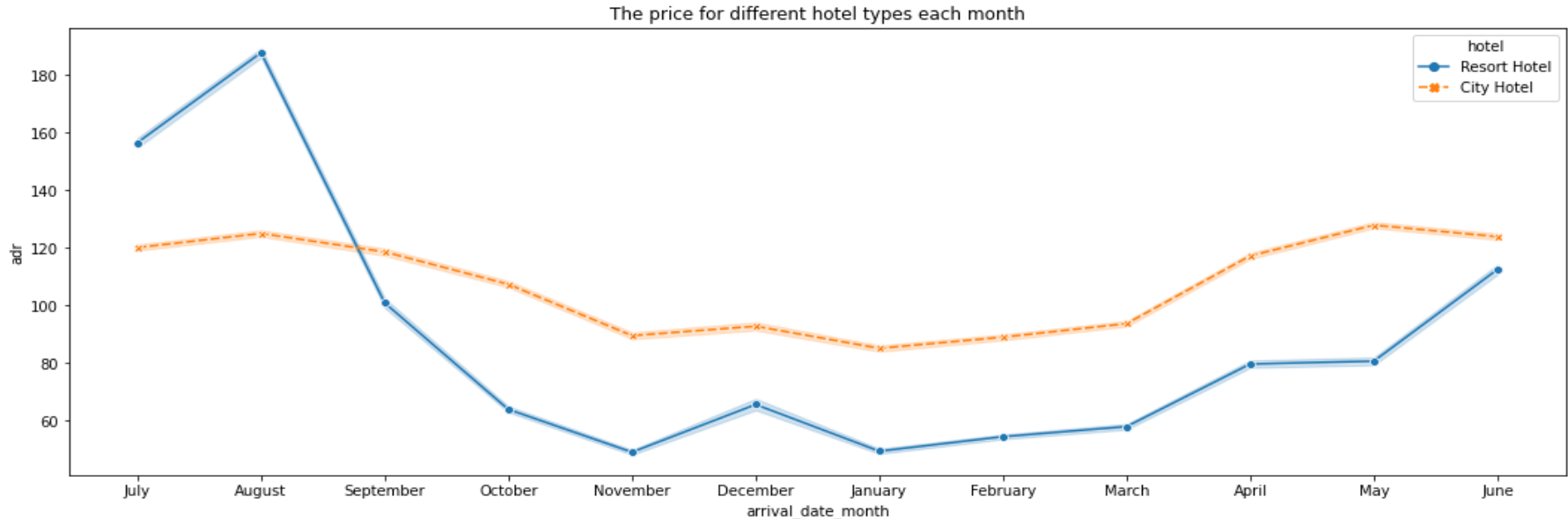
- Handling Outliers:** By plotting boxplot, we see some outliers. We have used median for removing outliers. Will check outliers in numerical column with Seaborn boxplot. change the value which is greater than 500 to median of lead time column and so on





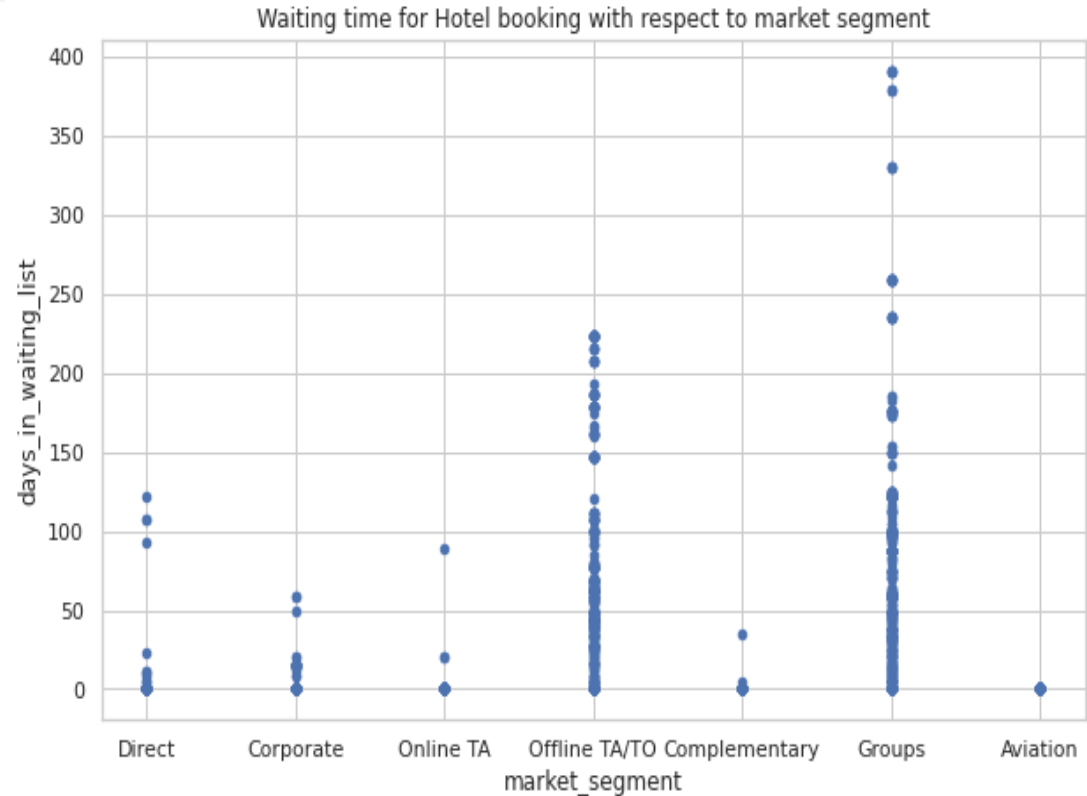
Which months have highest and lowest booking rates?

From graph we can conclude , that both city and resort hotels , from the month of November to January have lowest Average Daily Rate (adr) and month of August has the highest Average Daily Rate (adr).we can also say For both city and resort hotels, Nov to Jan have cheaper average daily rates



Which segment of the market usually has the Maximum number of days in the waiting list?

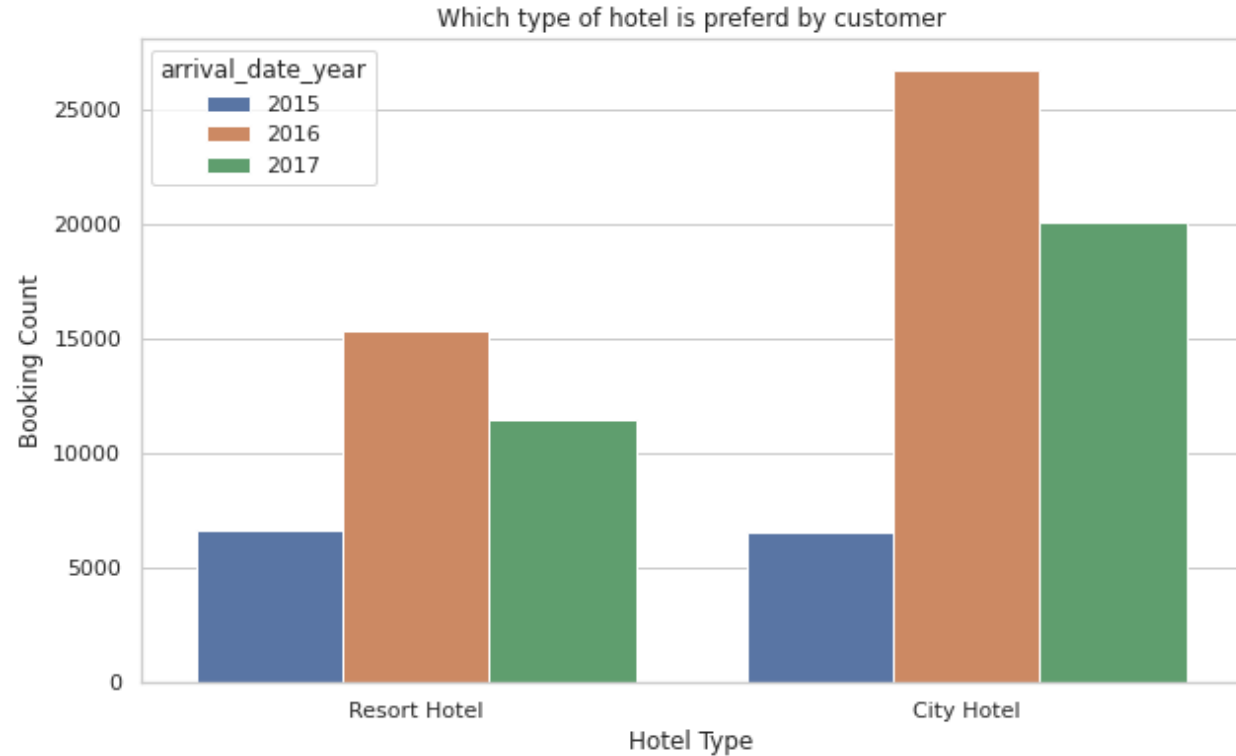
From the graph we can conclude that through the 'Groups' market segment customers have to wait more, this means that the waiting time is more before the confirmation of booking through groups market segment. Groups waiting time is 391 days for a data of 3 years which is maximum.



Which Type of Hotel is prefer by the customers ?

From the graph we can conclude that more number of customers preferred city hotel with 61.45% .

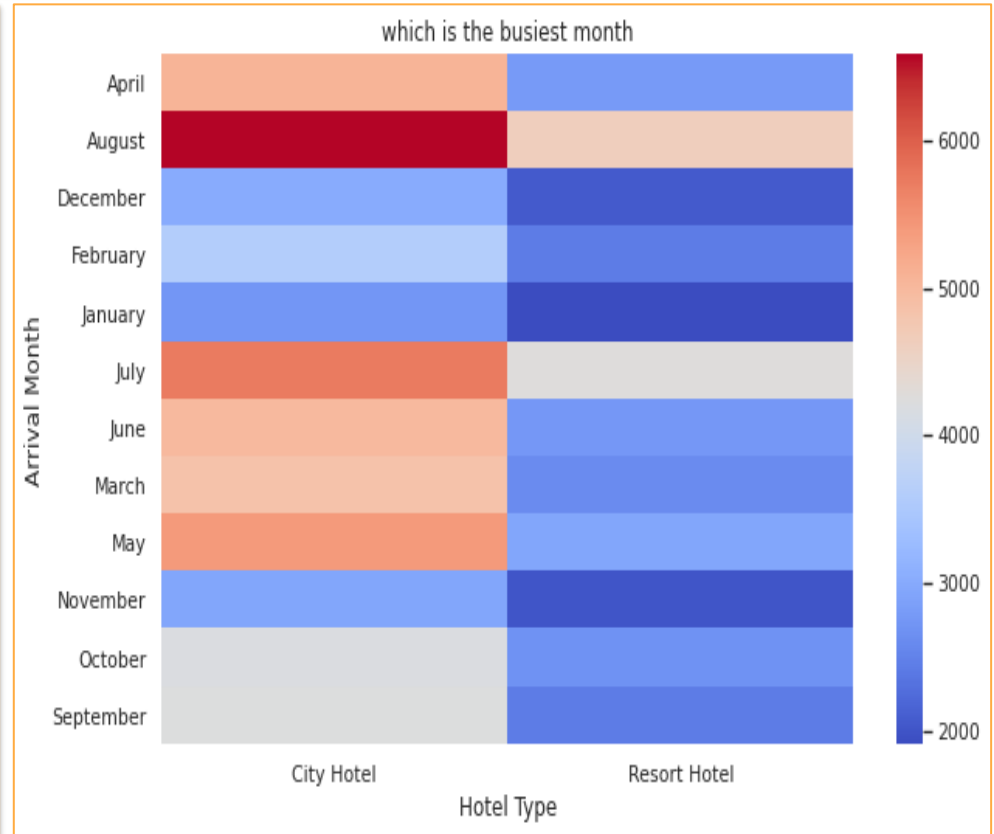
And in 2016 Both Hotels got maximum business



Which is the busiest month of the hotel ?

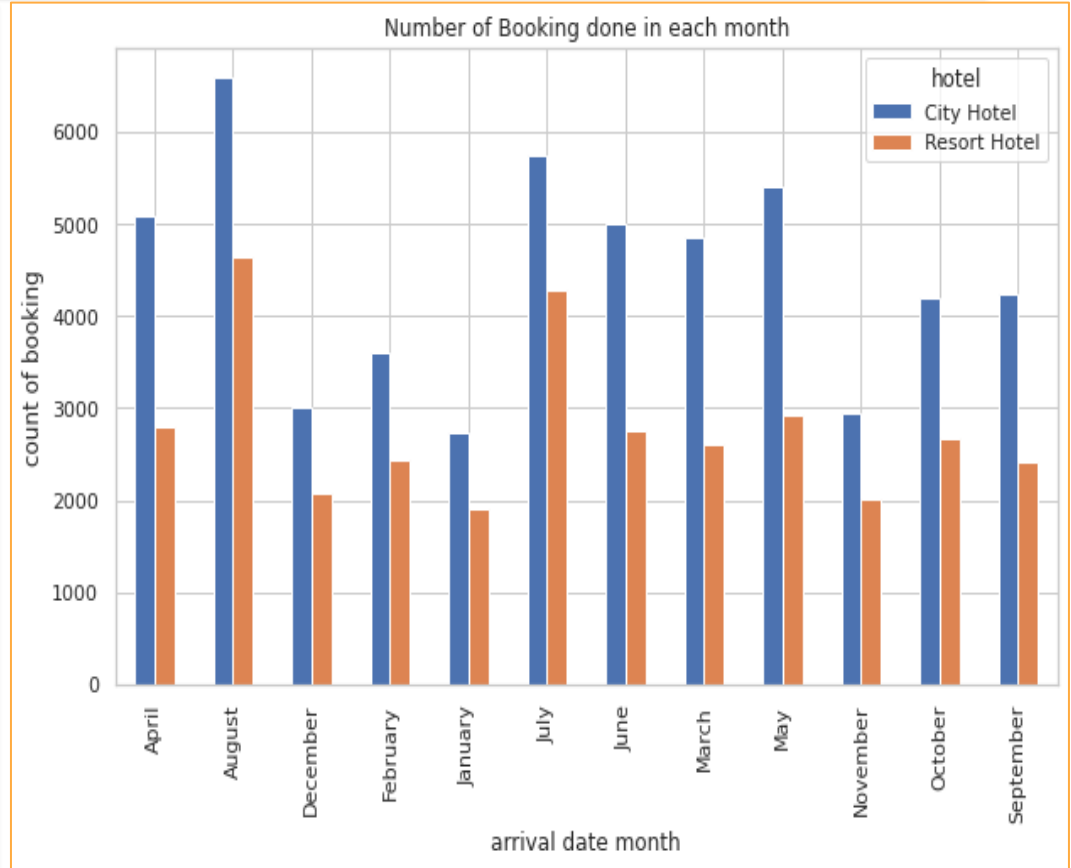
From graph we can conclude, that both city and resort hotels, from November to January have lowest bookings and August have highest bookings, which makes August month as the busiest month for both the Hotels.

August month City Hotel has 6586 booking and Resort Hotel has 4643 numbers of bookings.



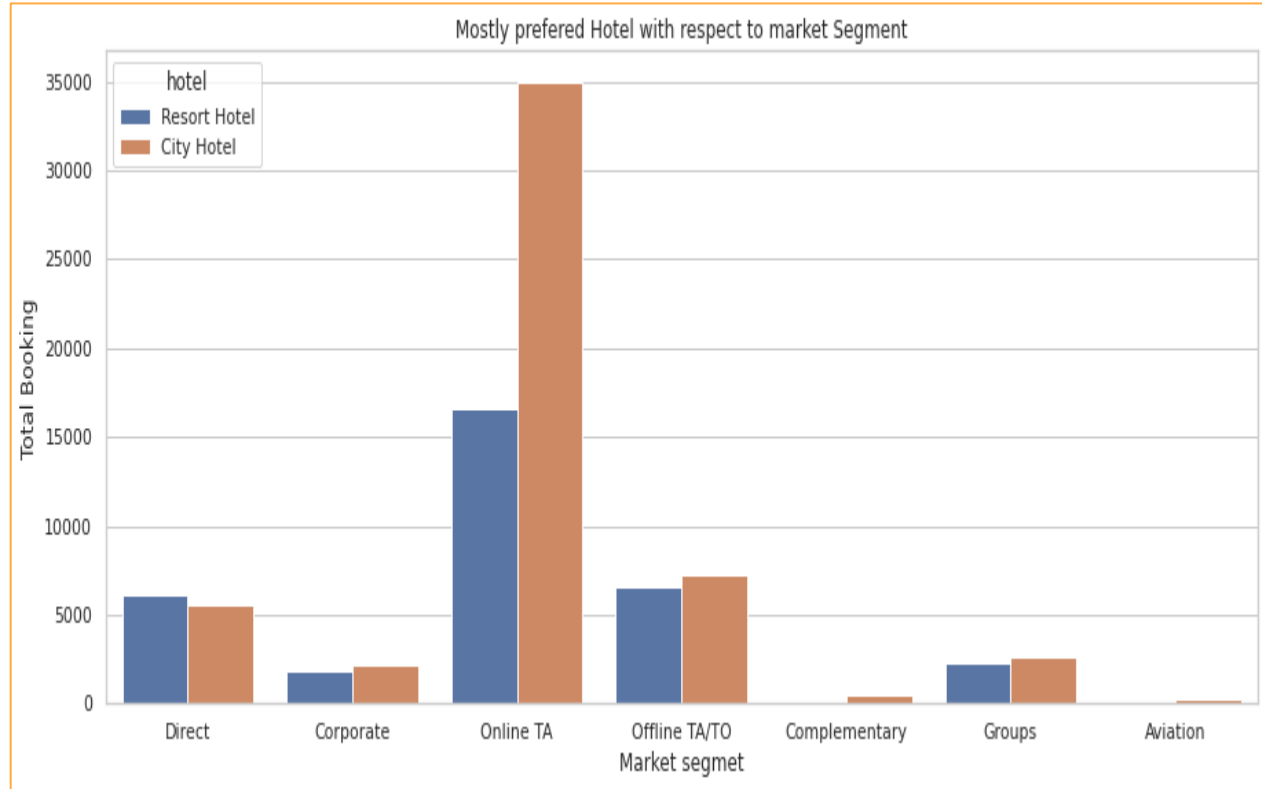
How many Bookings are done in each month ?

From the bar plot we can see the number of bookings made by city hotel are more than resort hotel and it is maximum during August month. Most number of hotel booking request came in the month of July and August followed by May and October. One reason for this may be the weather impact as these are the months of pleasant weather in Portugal.



Which is the most preferred business segment ?

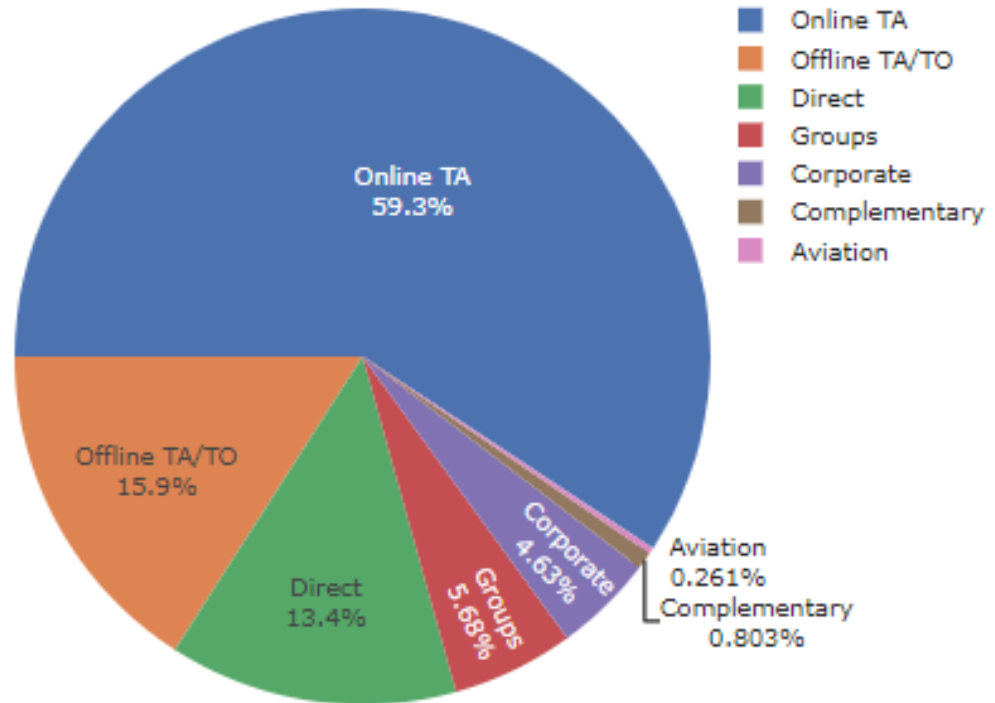
From the Bar graph we can see that 40.25 % city hotel booking are done by online TA and 19.04 % of Resort hotel are booked by Online TA.



Total Bookings by market segment

From the chart we found that maximum bookings are done by Online Travel Agents which is 59.3% and 15.9% bookings are made by offline Travel Agents and almost 13.4% bookings are Direct without any other agents.

Bookings per market segment

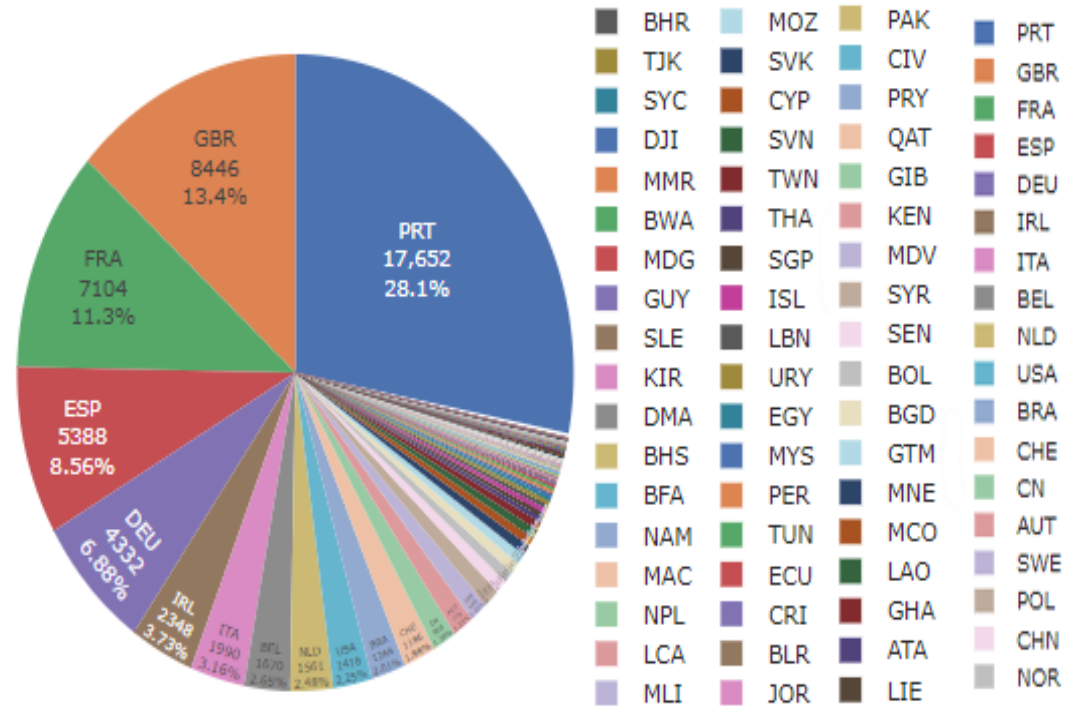


Which are the top five countries of origin of Guests ?

As we can see from the pie chart ,maximum numbers of guests are form Portugal with 17652 count, followed by Great Britain with 8446 count , and then France with 7104, and so on .

Around **28.1%** of all bookings were booked from **Portugal** followed by Great Britain(13.4%) & France(11.3%).

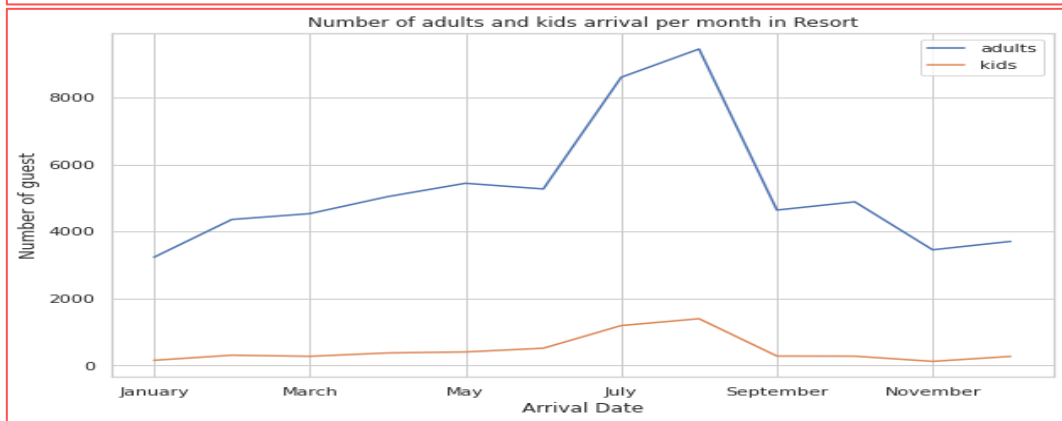
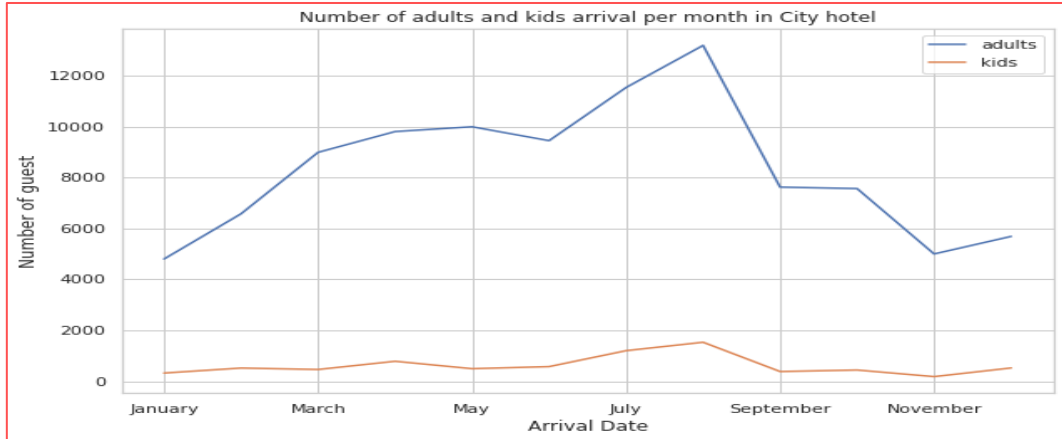
Home country of guests



How many adults and kids guests arrived per month in each hotels ?

The number of adults and kids in Resort Hotel in August month is 9463 and 1400 count, and for City Hotel its 13188 adults and 1532 kids.

It seems that the number of adults and children are higher in case of city hotel .



Conclusion

- ❑ From this EDA, we have observed that the top 5 most important features in the data set which will help the hotel in predicting the booking demand and in Revenue management are that most hotels are booked is the city hotel ,that is around 61.45% ,so You really need to spend more money on those hotels.Resort hotel tend to be on the costlier side and most people will go with city hotel.
- ❑ Most number of hotel booking request came in the month of July and August followed by May and October. One reason for this may be the weather impact as these are the months of pleasant weather in Portugal another reason can be vacation during summer season.
- ❑ Knowing that the hotels do not have repeated visitors, this can be taken as an opportunity to target on this area to attract repeated customers since they have already booked before , by offering them discount on booking, giving privilege services and taking feedback seriously for improvement.

Conclusion

- ❑ And most of the bookings are done through online TA (travel agencies), this can be taken as business opportunity by marketing and advertising on their website since majority of customers tend to reach out to them for booking, this can be because of the ease of booking from the website and to skip middleman. So we can say Most of the market segments and distribution channels involve travel agencies either online or offline.
- ❑ August is the busiest month and January month is the least occupied month. The marketing teams attention can be drawn to this, in order to come up with ideas to capitalize on this. Mostly the visitors are from Western Europe. Portugal being the highest followed by Great Britain and France. This gives another reason to spend a part of budget in that area to increase revenue.

Conclusion

- ❑ Analysis showed that the number of adults and children is higher in the case of the city hotel. This means that the city hotel is the better choice for large families.
- ❑ Majority of the reservation gets converted into success business transactions.

References

1. [Towardsdatascience.com](https://towardsdatascience.com)

2. [Kaggle.com](https://www.kaggle.com)

3. [Linkedin.com](https://www.linkedin.com)

4. [Analyticsvidya.com](https://www.analyticsvidya.com)

Thank You