Capstone Project
EDA- Hotel Booking Analysis

Presented by….
Rajat Mishra

Introduction

- The hotel sector is very unpredictable, and bookings depend on a wide range of variables, including hotel type, seasonality, day of the week, meal type, hotel type and many more.

- To assist the guest for booking the hotel in optimized value with safe stay. It is now even more crucial to analyze the patterns present in the historical data. I

Continued…

With the use of bar chart and line chart, we can use the patterns to predict the future reservations.

Step Involved

➤ **Step-1**

➤ Loading the Dataset in Notebook

➤ Checking the first view and first information of data

➤ Checking the Null values and cleaning the same

➤ Understanding the variables

➤ Analyzing Through Data Manipulation & Visualization Tools

➤ Deriving Conclusions

Some keywords

- **hotel** – Two hotels are given: Resort Hotel, City Hotel
- **is_canceled-** 1: Canceled, 0: Not canceled
- **lead_time** - gap between booking and arrival
- **arrival_date_year** - arrival year
- **arrival_date_month** - arrival month
- **arrival_date_week_number** - arrival week
- **arrival_date_day_of_month-** arrival date
- **stays_in_weekend_nights** - count of nights the guests booked the hotel during Sat-Sun
- **stays_in_week_nights** - count of nights the guests booked the hotel during Mon-Fri
- **adults** - count of adults
- **meal**- meal type (no meal package; BB; HB; FB)
- **country**- country of guests

Continued….

- **previous_cancellations-** count of previous bookings that were cancelled by the customer before final booking

- **previous_bookings_not_canceled-** count of no canceled bookings

- **days_in_waiting_list-** count of days the booking was in the waiting list before it was confirmed

- **customer_type-** Transient, Contract, Group, Transient-party

- **adr** - average daily rate for the booking

- **total_stay_nights** - duration of stay including weekend nights and week nights stay

- **price**- total price spent by a guest entity

Step 1

➢ **Loading the dataset in Notebook:**

● I have initially uploaded the given data in google drive and given the path of Hotel Booking analysis into the colab using python function and code.

➢ **First view of Data:**

➢ After loading the data set, I have checked the information using Python function. Data contains the shape of 119390 rows and 32 columns. Also, I have checked and lookout the variables

function and code.

Step 2

➤ **Cleaning the Data**:

● Data cleaning is the process to remove the unwanted values, undesired features, etc. It is very crucial part of dataset.

● In the data it is found that column 'company' have huge amount of null data which might get wrong result while doing the analysis hence we removed the particular column .

● And for remaining column we replaced the Null value to 0.

Step 3

➤ **Data visualization:**
- is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools ..

- We have done the analysis on the basis of pie chart, Bar chart, Line chart and correlation.
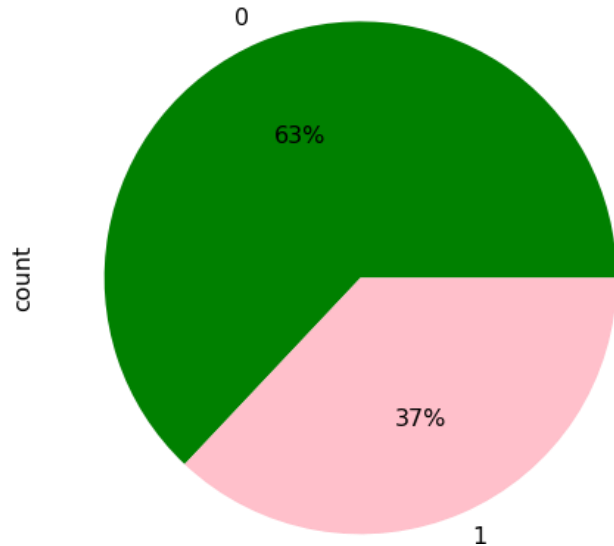
- Let's us see now the analysis!!!

Analysis for types of Hotel
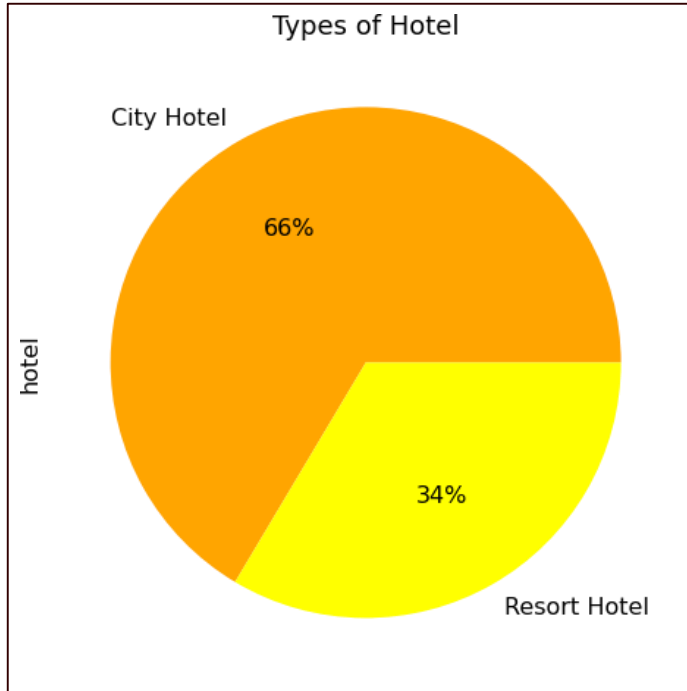
For below analysis, I have considered the whole data of booking, and it is found that city hotel has 66% preference and Resort hotel have 34% of preference. City hotel have more preferred by customers rather than the resort hotel.
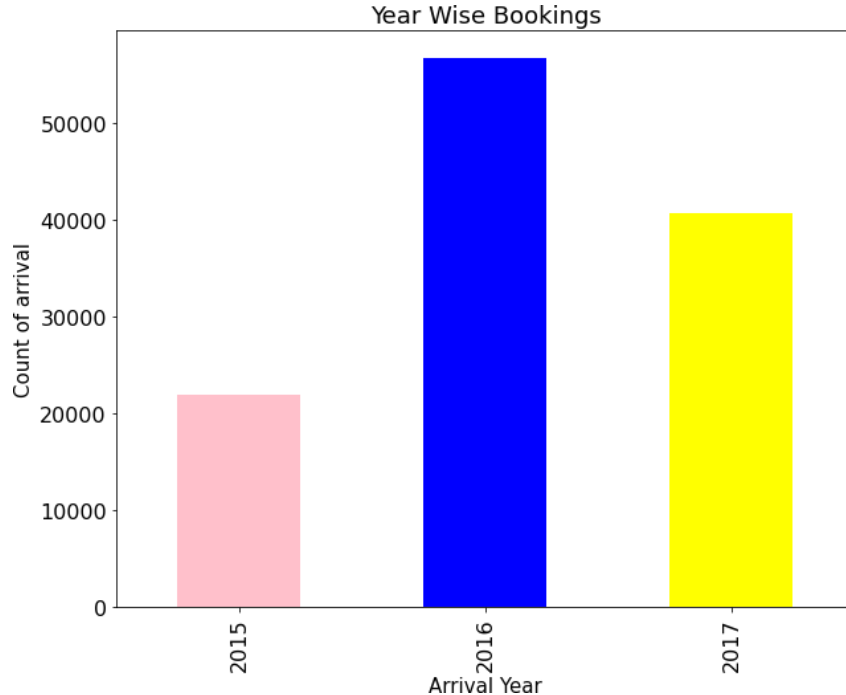
Hotel Cancellation Plot

Cancelation Plot for Hotel Booking Customers

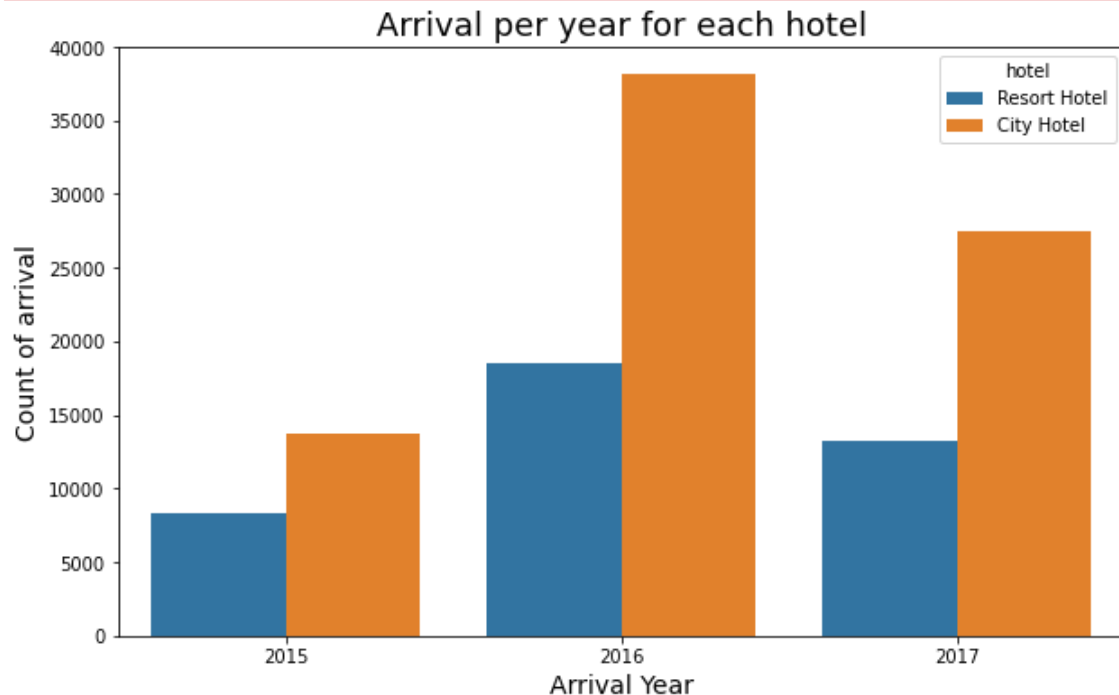It is found that over 37% of booking were cancelled by the customers. Which is slightly high

Arrival data using Bar Plot


Year Wise Bookings

Analyze the data year wise

We can observe that number of arrivals seems to be high in year 2016 while the bookings seem to be less in 2015 and 2017.

Arrival data for Type of Hotel

Arrival per year for each hotel

We can observe that number of arrivals seems to be high in year 2016 for city hotel compared to Resort Hotel, while the bookings seem to be less in 2015 and 2017 for both resort hotel and city hotel.

Customer Arrival Data Per Year for Each Hotel


Arrival per month for each hotel
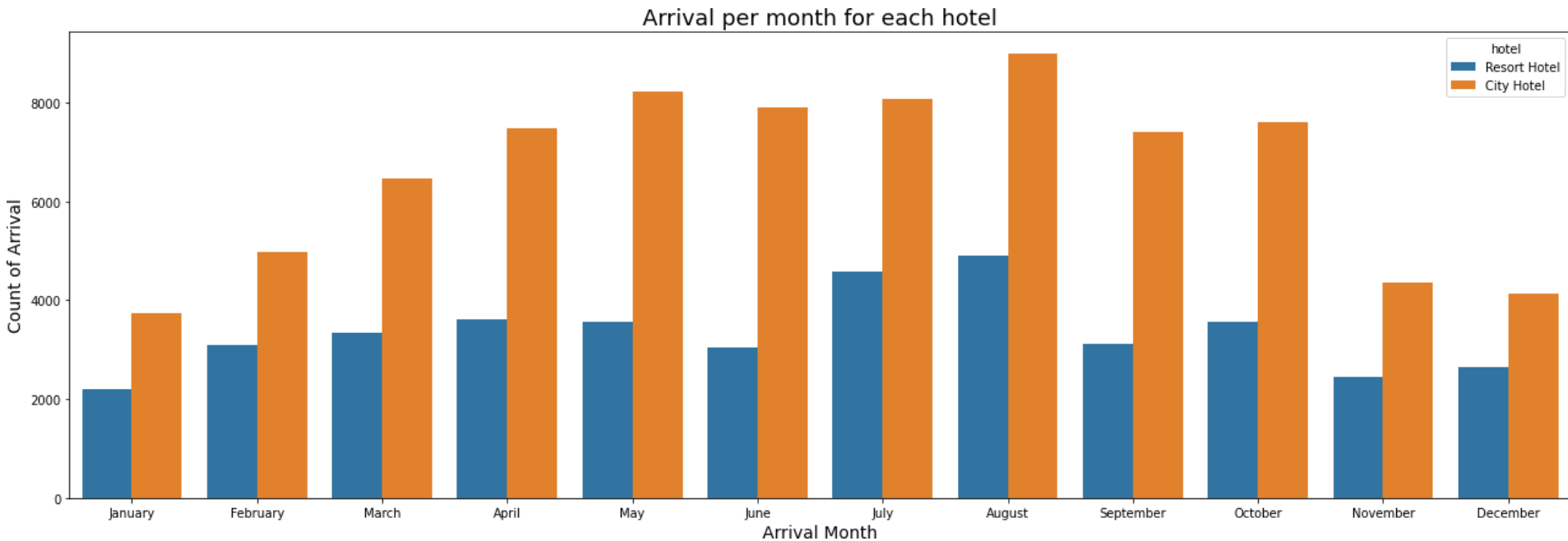
Continued…

- We can observe that number of arrivals seems to be high in month of May, June, July and August for city hotel. But number of arrivals for Resort Hotel is seems high only for July and August. For initial month number of arrival is less compared to mid-months
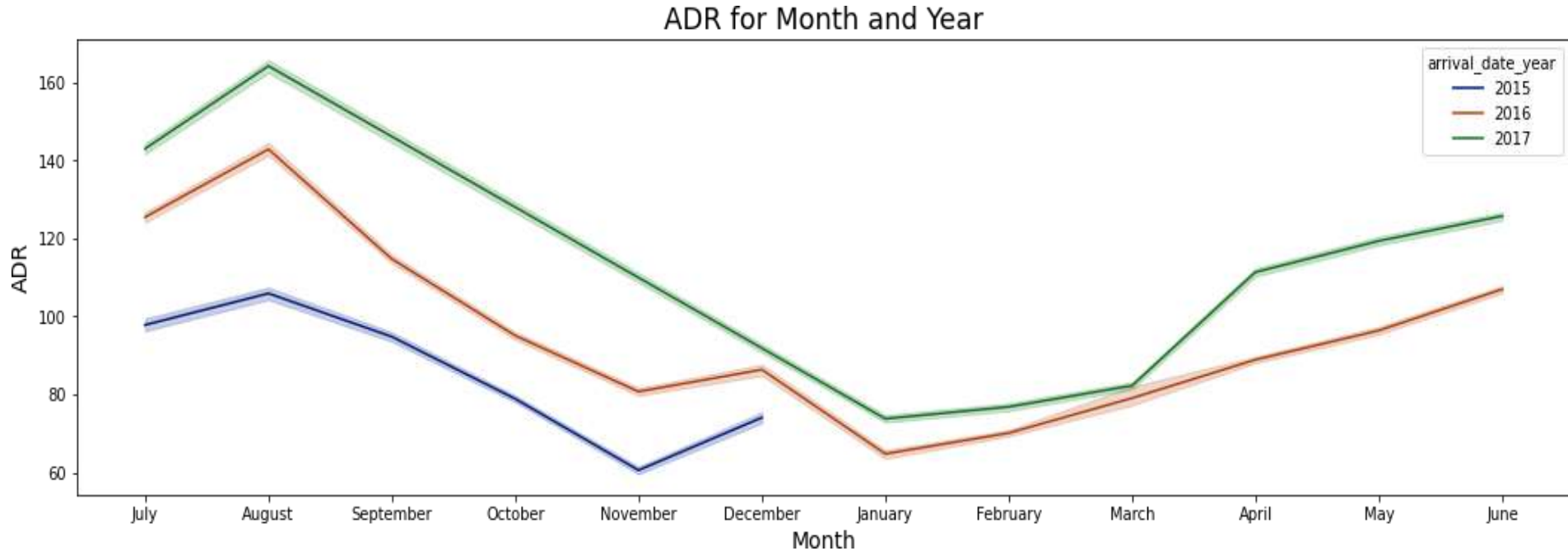
Number of Bookings vs Country
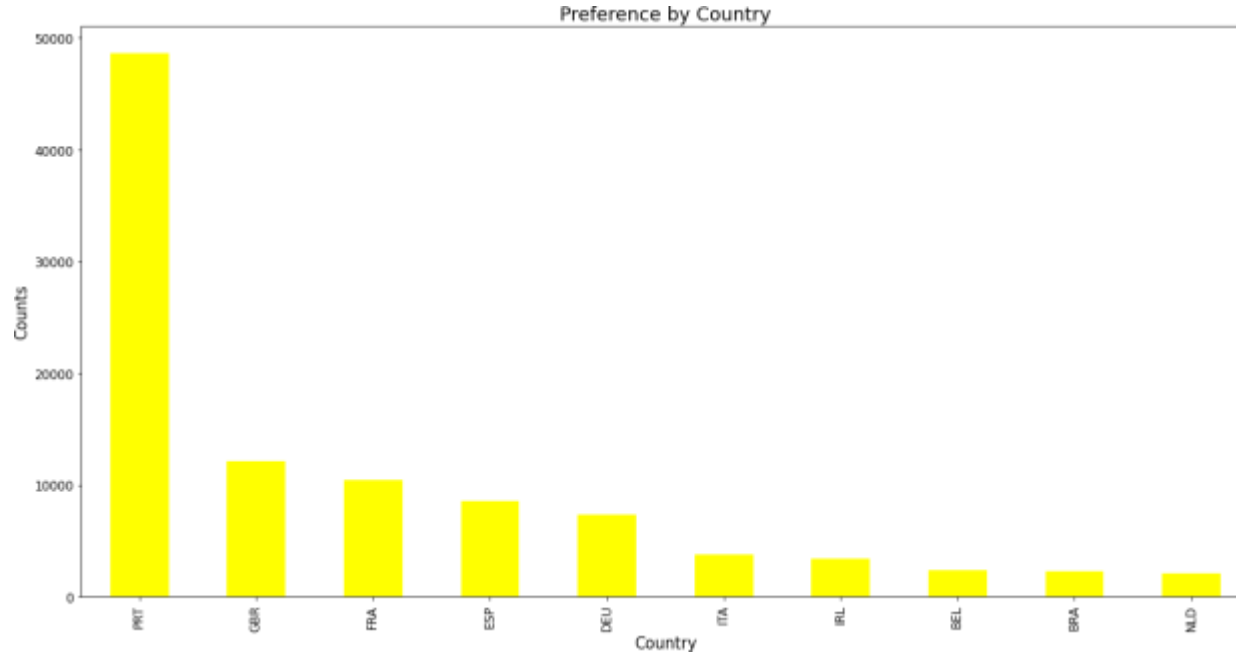
Most píefeííed countíy is PRΓ followed by GBR

ADR on the basis of Hotel Types

ADR on the basis of Hotel Type
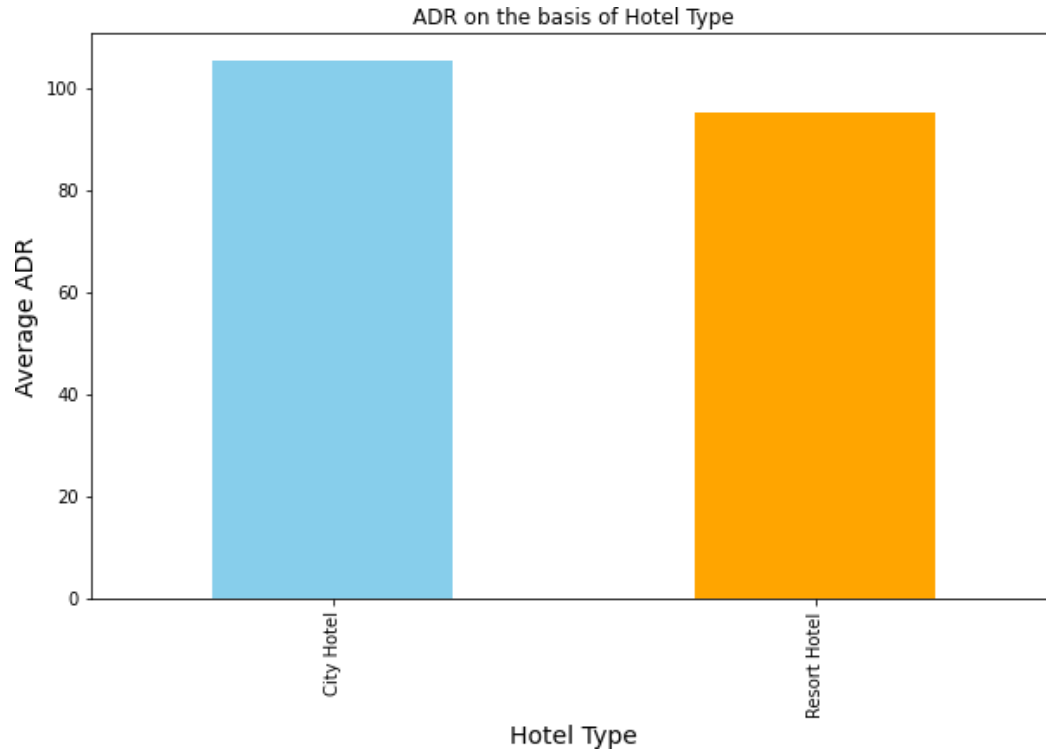
ADR for city hotel is slightly more than the Resort hotel, not much difference found in this.

# Average Daily Rate vs Hotel Types per Month and Year

ADR for Month and Year

Continued….

- The line plot is used to plot the average daily rate (ADR) per month and year. The year is shown in different colors,

- Blue indicates year 2015

- Red indicates year 2016

- Green indicates year 2017.

- Above graph clearly shown that hotel business is growing every year.

Price paid for stay vs Month for Both Hotel

Month-wise Price Paid for stay by Guests
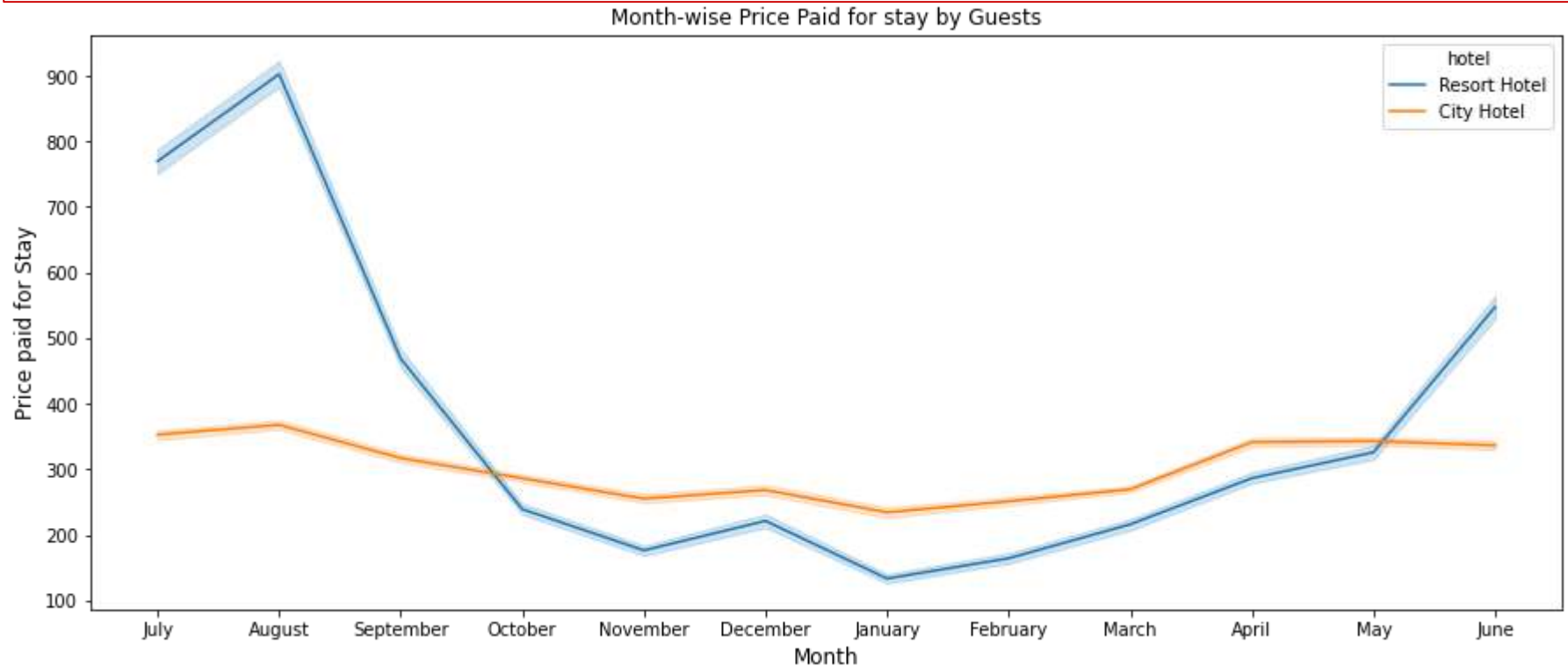
Continued….

- As per above graph it is clearly seen that price paid by city hotel for month of July, August and September was less than the Resort hotels. But for rest of the month price of City hotels were consistently higher than the Resort hotels.

Correlation Map
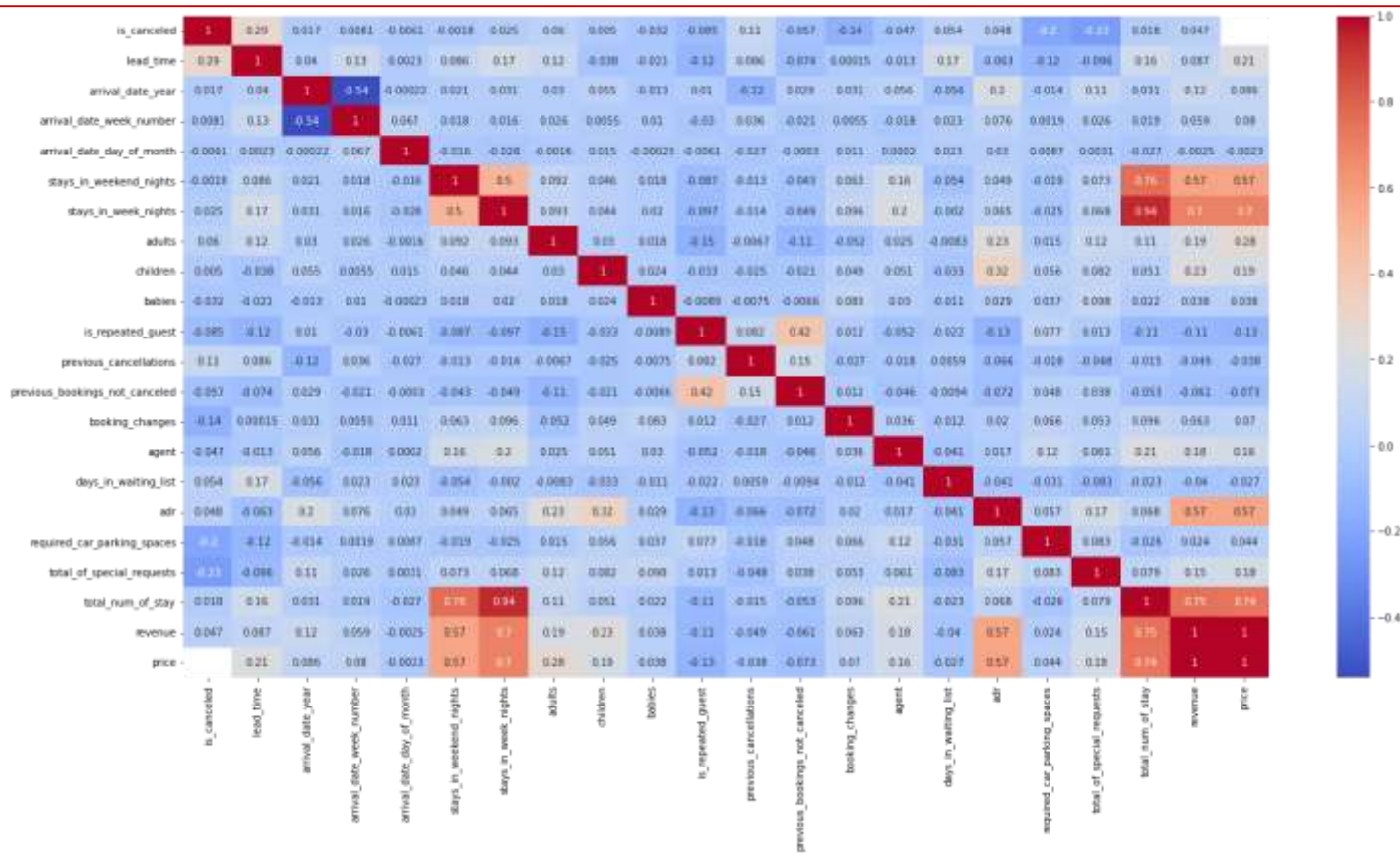
Continued….

- In the above correlation map we can clearly observe that, stay in week nights was more than the weekend nights. Also focus on revenue, stay in week nights, total number of stay we can say that revenue was almost same for both the parameters.

Conclusion

➢ **City hotel have more preferred by customers rather than the resort hotel.**

➢ **In overall data and it is found that over 37% of booking were cancelled by the customers.**

➢ **Number of cancellations of City Hotel seems very higher than the Resort Hotel.**

➢ **Number of arrivals seems to be high in year 2016 while the bookings seem to be less in 2015 and 2017.**

➢ **Number of arrivals seems to be high in month of May, June, July and August for City hotel. But number of arrivals for Resort Hotel is seems high only for July and August.**

➢ **PRT have more preference followed by GBR.**

➢ **As per line graph of ADR per month for three years, clearly seen that hotel business is growing every year.**

➢ **As per line graph Price paid for stay for each month it is clearly seen that price paid by City hotel for month of July, August and September was less than the Resort hotels.**

➢ **In the correlation map we can clearly observe that, stay in week nights was more than the weekend nights.**

Thank You !!!
Have a Great Day