Capstone Project-1

Hotel Booking Analysis

Team Members :
Sahil Jagtap
Shweta Jagtap
Swapnil Wankhede

Outline of presentation

- **Introduction**
- **Problem Statement**
- **Data Summary**
- **Data Exploration and Data Cleaning**
- **Analysis on problems**
- **Challenges**
- **Conclusion**

Introduction

- Hotel Booking system is the online reservation system that allows all hotel reservations made through hotel websites and various apps.
- Hotel Booking is now widely used all over the world because of the online availability of various application from which we can book the hotel at any time and at anywhere.

- We are analysing this data because it is helpful in making the hotels plan better so that hotel can perform various things to boost their business.

- By analysing, it helps the hoteliers to understand the preferences and purposes of a customer.

Problem Statement

Objective:

- when is the best time of year to book a hotel room?

- The optimal length of stay in order to get the best daily rate.

- Which type of distribution channel are mostly used for which hotel type?

- Booking trend of hotels.

- Comparison of hotels' adr month wise.

- Many more things.

Data Summary

This dataset contains the information for a city hotel and a resort hotel, and includes several variables such as

Numerical Variable
- **Stays_in_week_nights, Stays_in_weekend_nights, arrival_date year, arrival_date_week, adults, children, babies, adr etc.**

Categorical Variable
- **hotel, Country, meal, arrival_date_month, distribution channel, deposit_type, reservation_status etc**

Binary Variable
- **Is_cancelled, is_repeated_guest**

Explore and analyse the data to discover important factors that govern the bookings.

Data Exploration and Data Cleaning

- **Loading dataset and exploration of data.**

- **Removing duplicates rows**

- **Handling Null values.**

- **Converting columns to appropriate data types**

- **Creating new colums.**

- The given dataset has nearly 119390 observations with 32 variables which are mix of float type (4), integer type (16) and object type (12) including some missing values and duplicates that affect the dataset.

- The duplicate values are 31994 and non duplicate values are   87396.

- Only company, agent ,children and country have the null values that we have to replace with the appropriate values.

- Converting datatype of columns 'children', 'company' and 'agent' from float to int.

- After cleaning and exploring data the updated rows are 87228 and columns are 33 .

Analysis on problems

AI

Percentages of bookings have been cancelled and not cancelled month wise
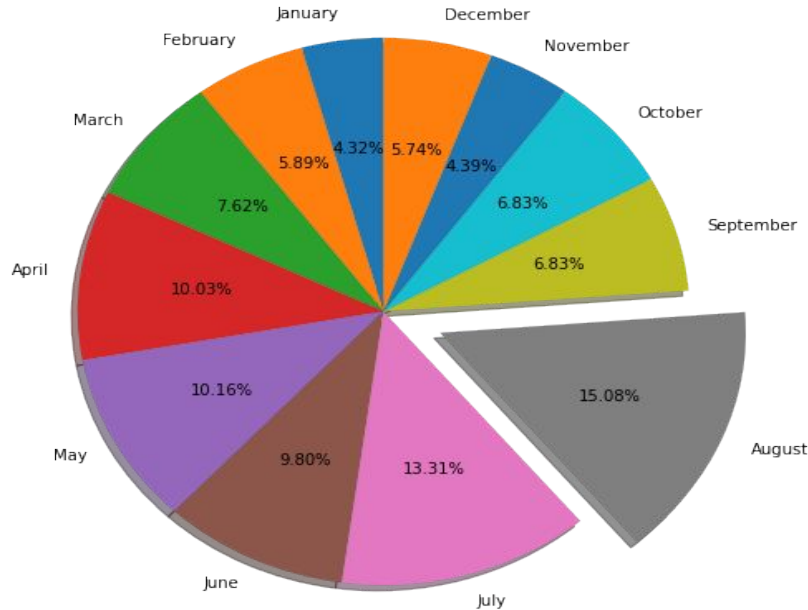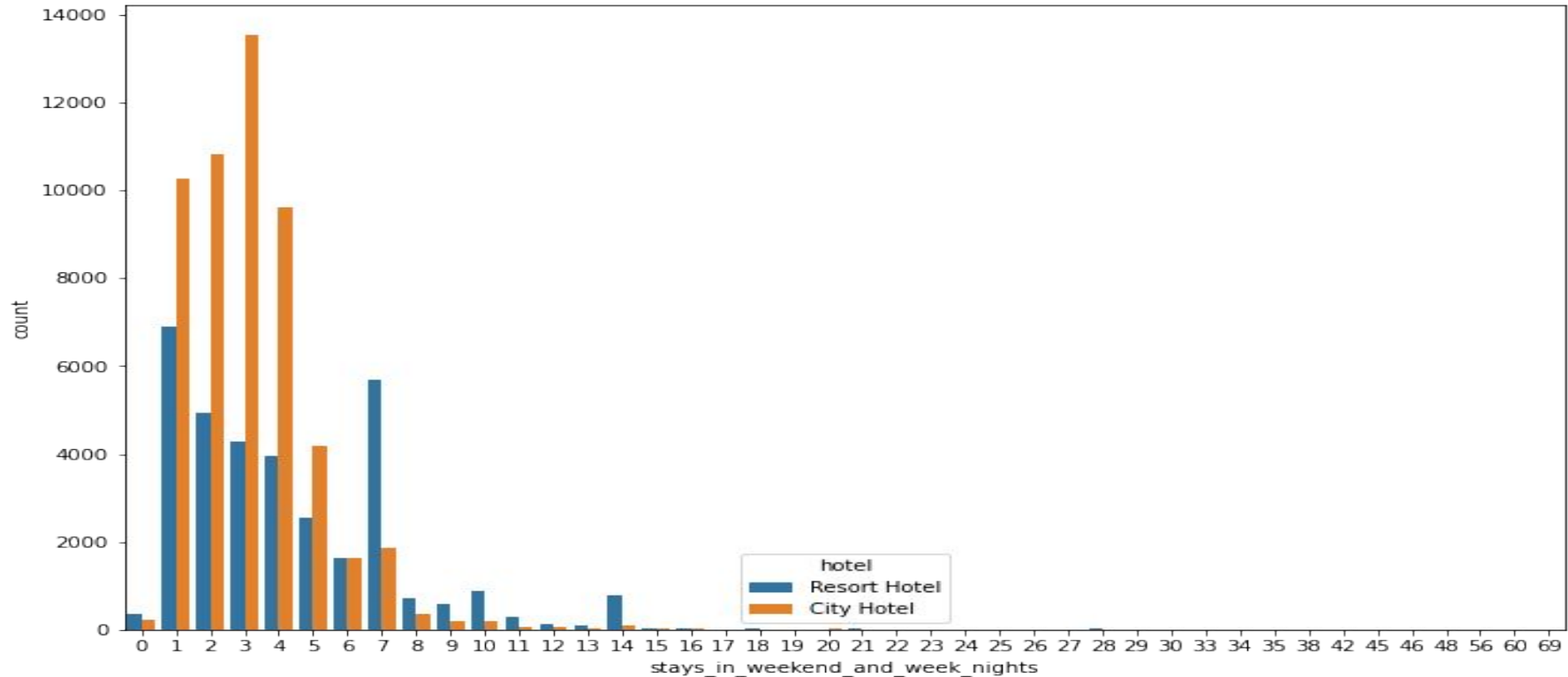

Cancelled bookings pie chart
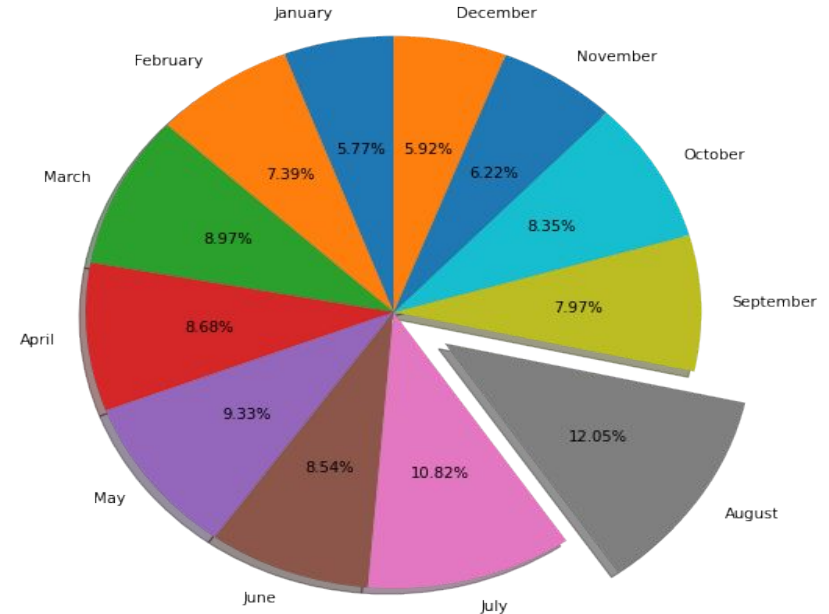
Not cancelled bookings pie chart

Most hotel rooms are booked in august month which is the best time of year to book a hotel room.

Which type of hotel are mostly booked in weekend and week nights?

City hotels are mostly booked in weekend and week nights.

Which type of hotel got maximum number of cancellation of bookings?

City hotel has maximum number of cancellation of bookings as compared to resort hotel.

cancelation of bookings in hotels

Which type of distribution channel are mostly used for which hotel type?

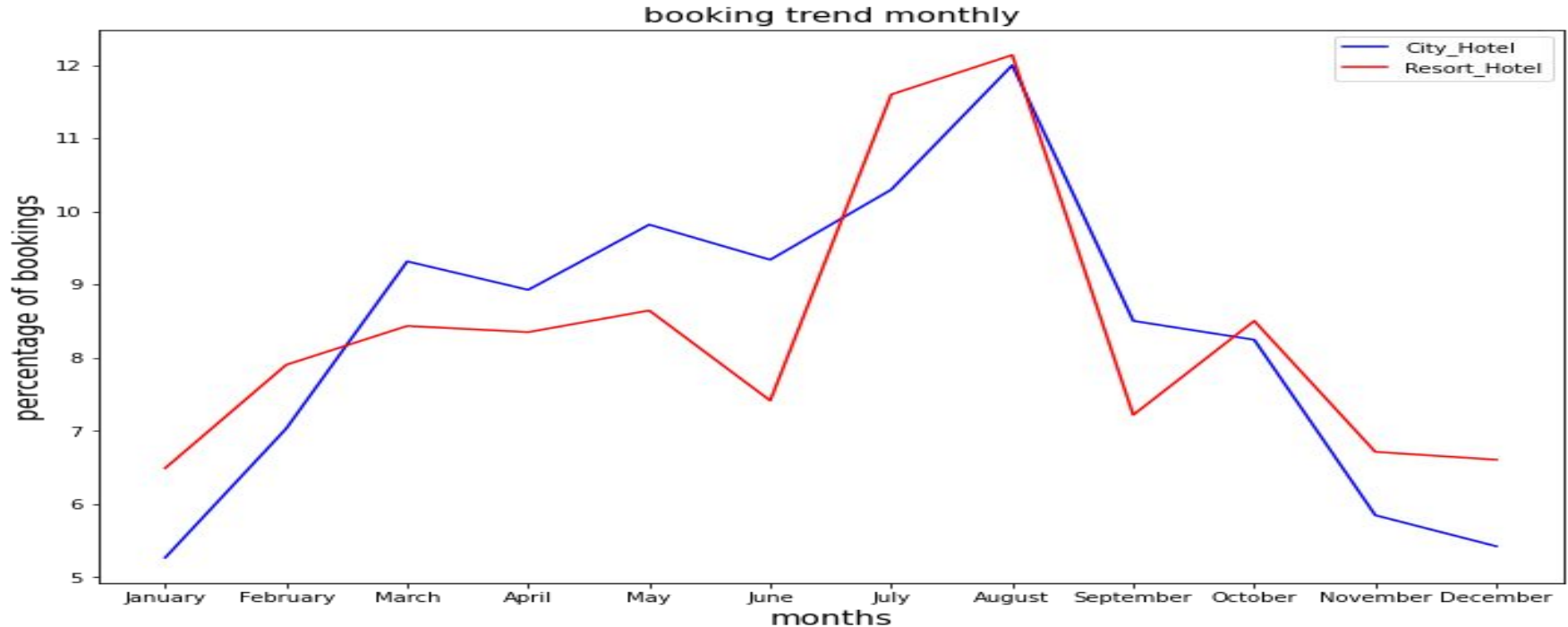Most of the rooms of city hotels are sold by TA/TO distribution channel as compared to resort hotel.
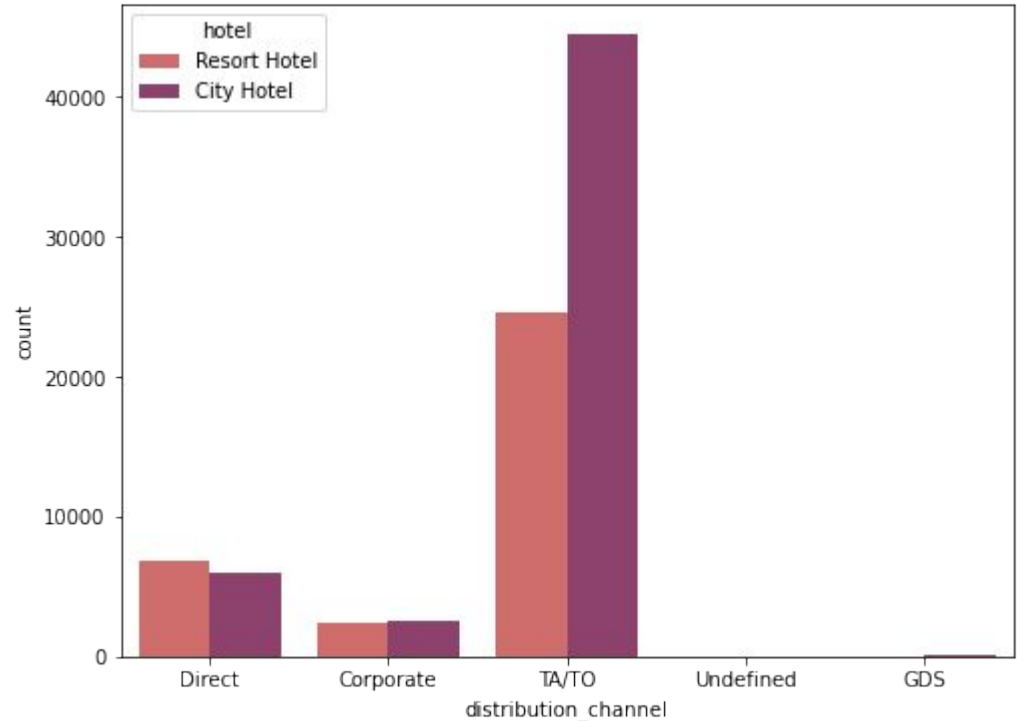
Which type of hotels have maximum repeated guests?

Resort hotel type has maximum repeated guest as compared to city hotel.


Repeated guest hotel type wise
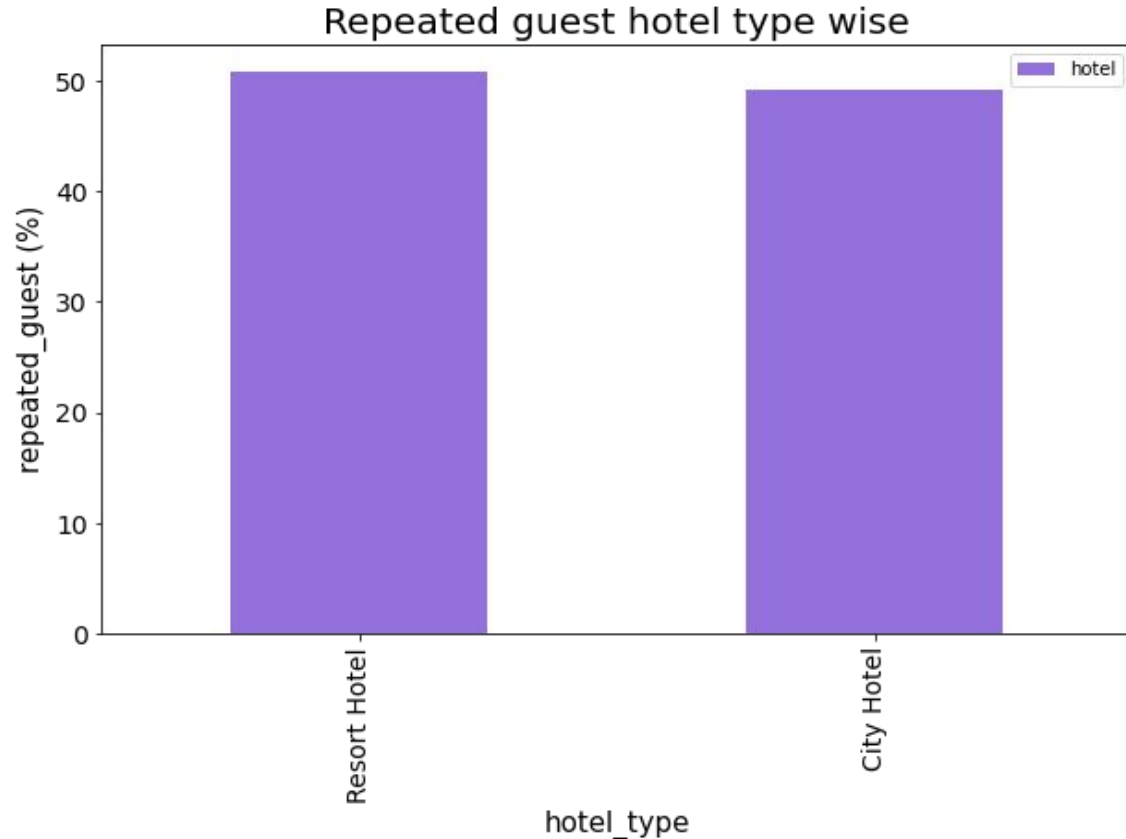
# Bookings trend of hotels (city hotel and resort hotel) month wise.

Graph of trend of city and resort hotels is kind of similar with some difference.For the both hotels, more booking is in august month followed by july month and in october month, percentage of booking for both hotels is same.

# Comparison of Hotel's ADR (Average Daily Rate) week wise.
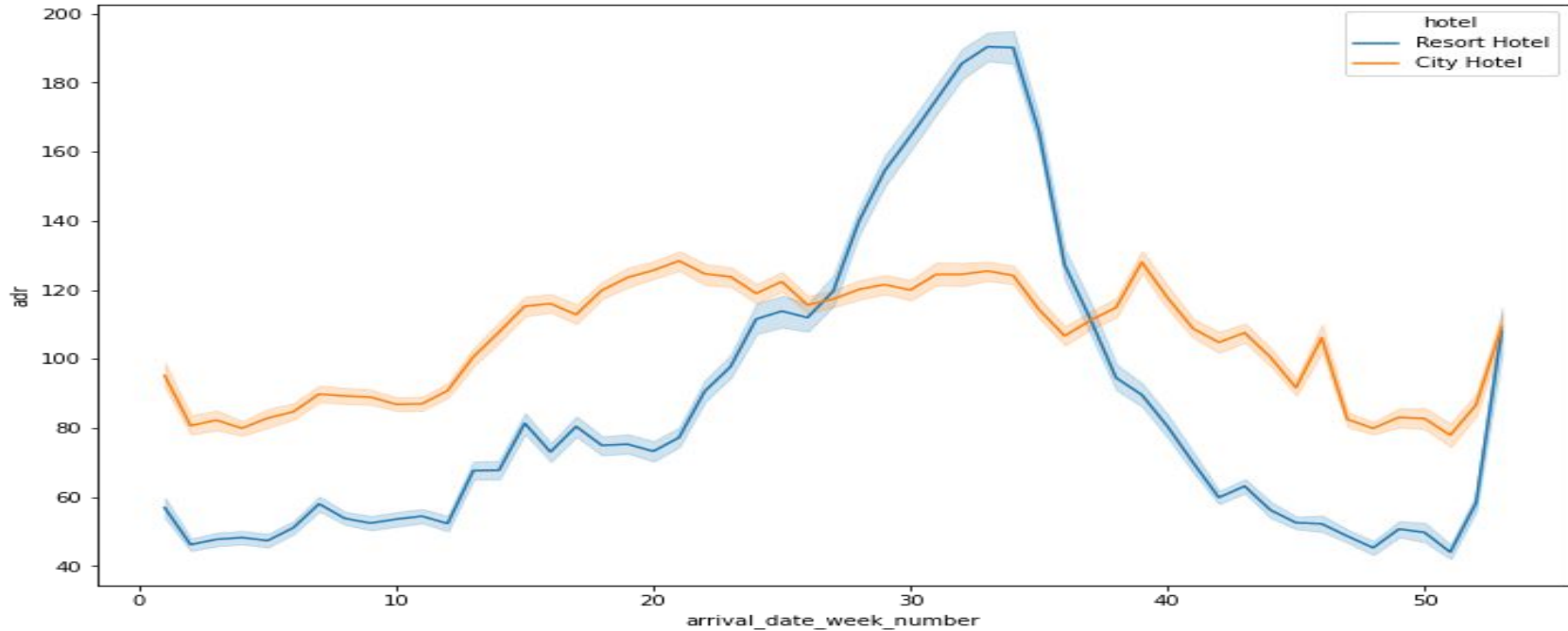
Difference of Adr in both hotel types is high at the starting week of year i.e january month. At the end of week, ADR of both hotel types is same. Avarage Daily Rate (ADR) for the week 28 to 35 (approx) i.e the month of july and August are more for the resort Hotel than city hotel.

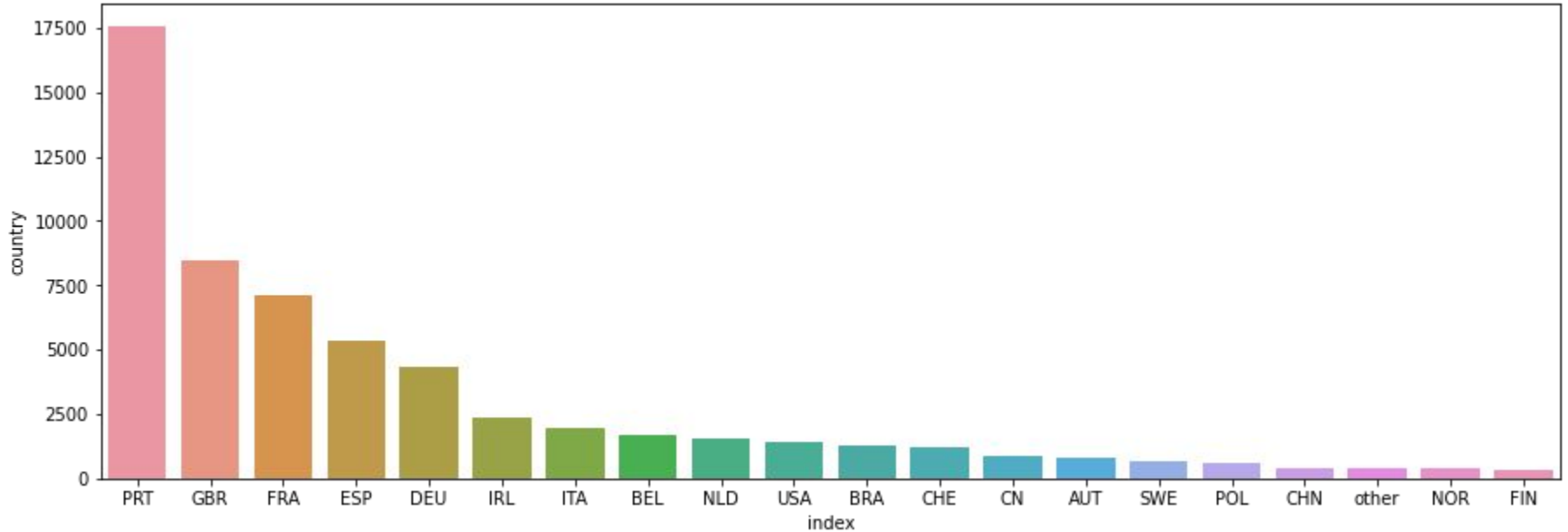# In which year maximum number of reservation status is checked out?
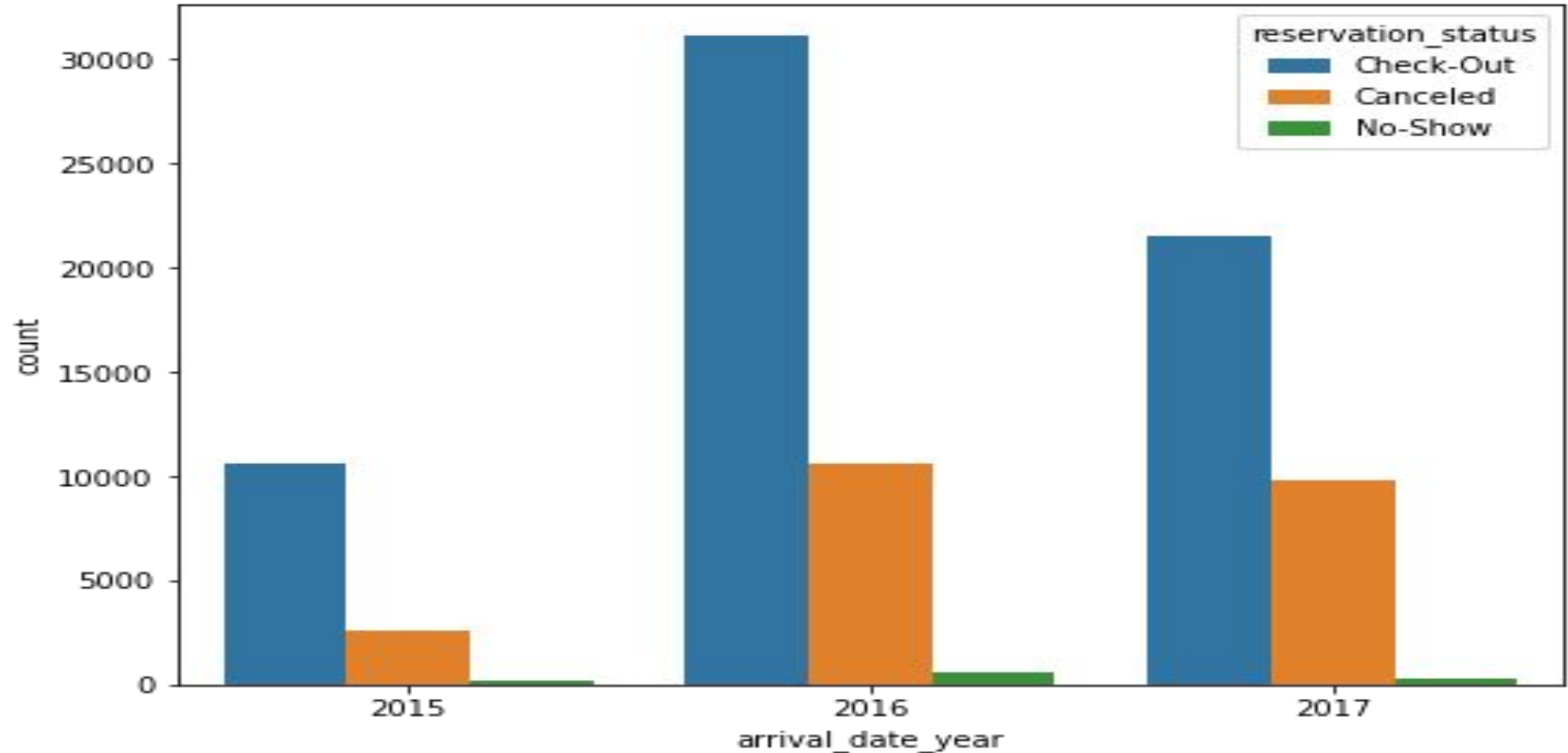
In 2016, maximun number of reservation_status is checked_out.

Which is the most preferable meal Types?
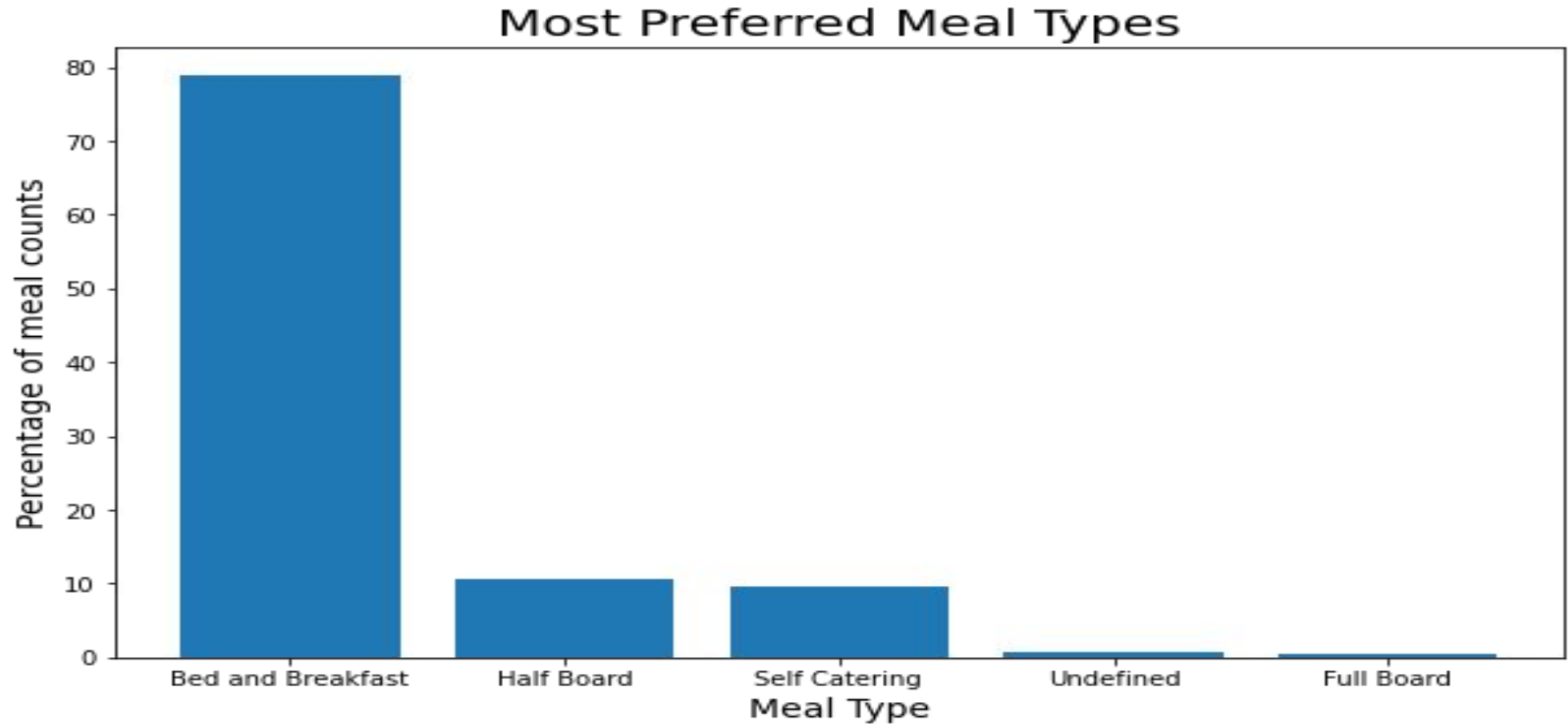
Bed and Breakfast is the most preferable meal types.

Maximum Bookings From Top Four Countries

PRT (Portugal), GBR (United Kingdom), FRA (France) and ESP (Spain) are the top four countries where the bookings are maximum

The Best Optimal Length To Get The Best Daily Rate

- The best optimal length of stay is 1 to 10 week and weekends nights in order to get the best daily rate.

- As length of total_stay increases the adr decreases

- This means for longer stay, the better deal for customer may be.
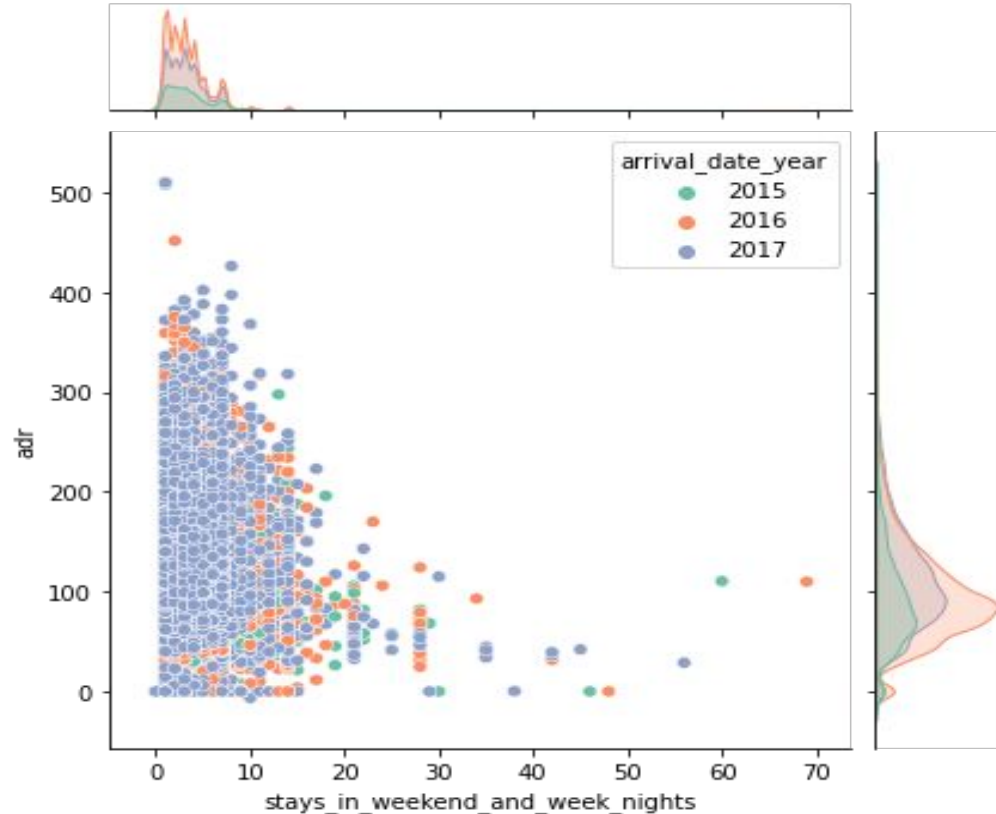
Corelation between features using Heatmap

- Average daily rate(ADR) is positively correlated to children, adults, and babies with 33%, 25% and 2% respectively.It means that maximum and minimum revenue is generated by occupied rooms with children and babies respectively.

- It is observed that arrival_date_week_number and arrival_date_year are 51% negatively correlated.
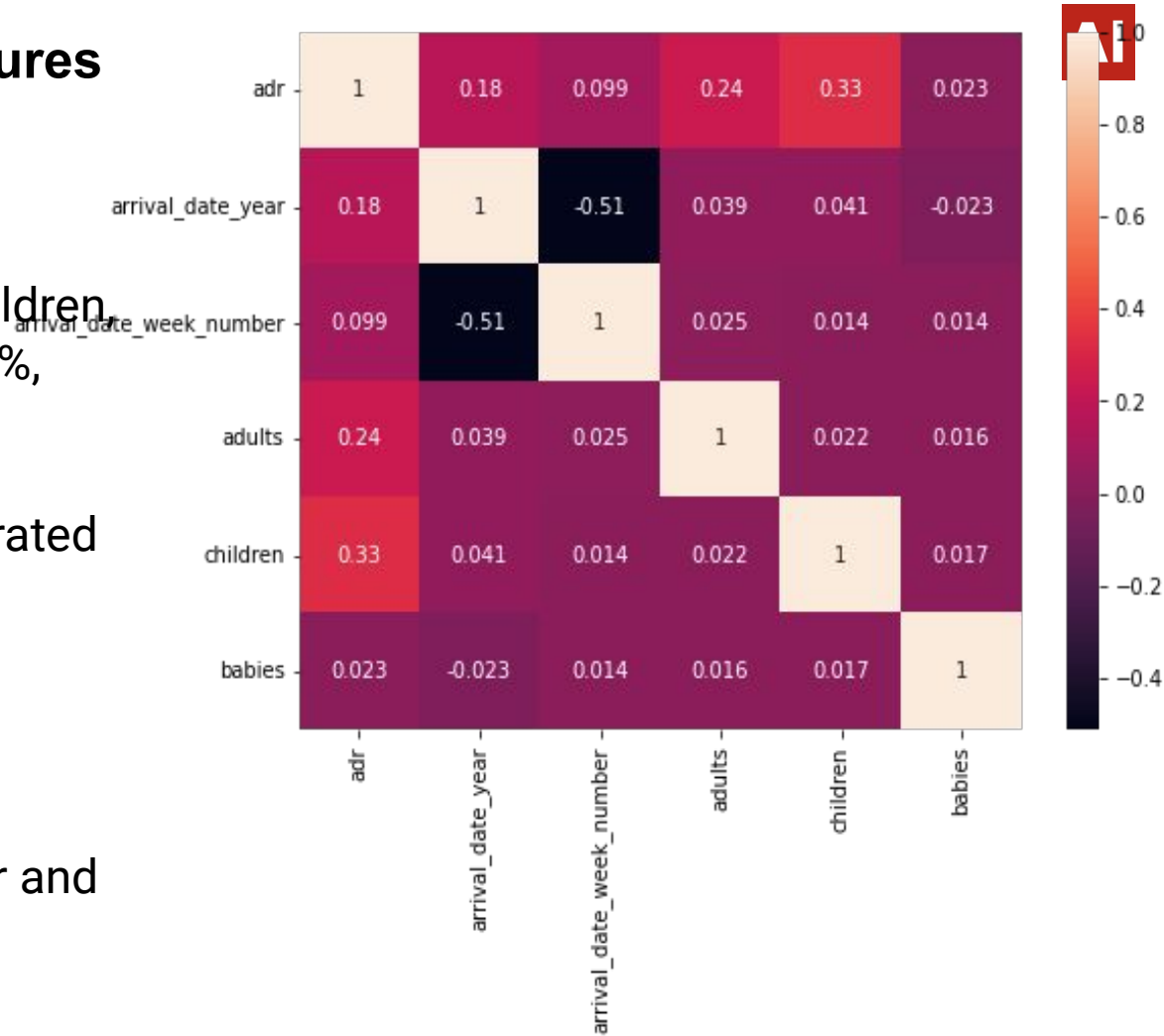
Challenges:

- Lots of Duplicate data, Null values.

- Face difficulty to use particular graph while visualizing the data.

Conclusion:

We learned that,

- Most hotel rooms are booked in august month which is the best time of year to book a hotel room.
- City hotels are mostly booked in weekend nights and week nights.
- City Hotel has maximum number of cancellation of bookings as compared to Resort Hotel.
- Most of the rooms of city hotels are sold by TA/TO distribution channel as compared to resort hotel.
- The city hotel got the higher number of special requests.
- Trend of booking for resort hotel and city hotel is kind of similar with a small difference. for the both hotels, more booking is in august month followed by july month. In october month, percentage of booking for both hotels is same.

- Difference of Adr in both hotel types is high at the starting week of year i.e january month. At the end of week, ADR of both hotel types is same. Average Daily Rate (ADR) for the week 28 to 35 (approx) i.e the months of july and August are more for the resort Hotel than city hotel.
- In 2016, maximum number of reseervation_status is checked_out.
- For every customer, BB (Bed and Breakfast) is most preferable meal type.
- PRT (Portugal), GBR (United Kingdom), FRA (France) and ESP (Spain) are the top four countries where the bookings are maximum.
- The best optimal length of stay is 1 to 10 week and weekends nights in order to get the best daily rate. As length of total_stay increases the adr decreases. This means for longer stay, the better deal for customer may be.
- Average daily rate(ADR) is positively correlated to children, adults, and babies with 33%, 25% and 2% respectively. It means that maximum and minimum revenue is generated by occupied rooms with children and babies respectively.

THANK YOU