# Capstone Project – 1

## Hotel Booking Analysis


**Team Members**

**Prashant Bhatt**

**Manish Kumar**

# Introduction

Businesses around the world are waking up to the importance of data analytics. Getting the right insights gives you an advantage in a competitive business environment.

Why…..?

- Introduce better hotel pricing strategies
- Managing Hotel Booking Channels
- Demand Forecasting
- Enhancing productivity
- Marketing strategy

# Data Summary

**hotel:** There are only two types of hotels Resort hotel and City hotel

**market_segment:** We have eight unique market segments from where customers are coming i.e. 'Direct', 'Corporate', 'Online TA', 'Offline TA/TO', 'Complementary', 'Groups', 'Undefined', 'Aviation'

**meal:** These are the four types of meal in the given data :

- SC: self-catering (no meals are included).
- BB: bed and breakfast.
- HB: half board.
- FB: full board.

**country:** We have data of 177 countries coded in short form.

**Lead_time:** At a hotel, the time taken between when a customer makes a reservation and their actual arrival is called the Lead Time.

# Data Summary(continued)

**is_canceled:** It contains only two values 1 and 0.

- 1: booking has cancelled
- 0: booking is currently active

**Stays_in_weekend_nights:** We are getting two unique values for column 'stays_in_weekend_nights' i.e. 1 and 2 , that means someone booked for one weekend night(saturday or sunday) and some booked for two weekend nights(saturday and sunday).

**Stays_in_week_nights:** We are getting five unique values for column 'stays_in_weekend_nights' i.e. 1 to 5 , that means someone booked for at least one week night to maximum five week nights.

**Adr:** Average daily rate for individual order.

**required_car_parking_spaces:** this column contains number of car parking spaces required as per customer demand.

# NaN value Handling

- Total no. of rows / data = 119390
- Here we can observe that four columns have some null values:
  - Children 119390 - 119386 = 4 null values.
  - Country: 119390 - 118902 = 488 null values.
  - Agent: 119390 - 103050 = 16340 null values.
  - Company : 119390 - 6797 = 112593 null values.
- Here we can observe that in 'company' column 94.3% values are null.

```
sum([True for idx,row in hotel_df.iterrows() if any(row.isnull())])
```

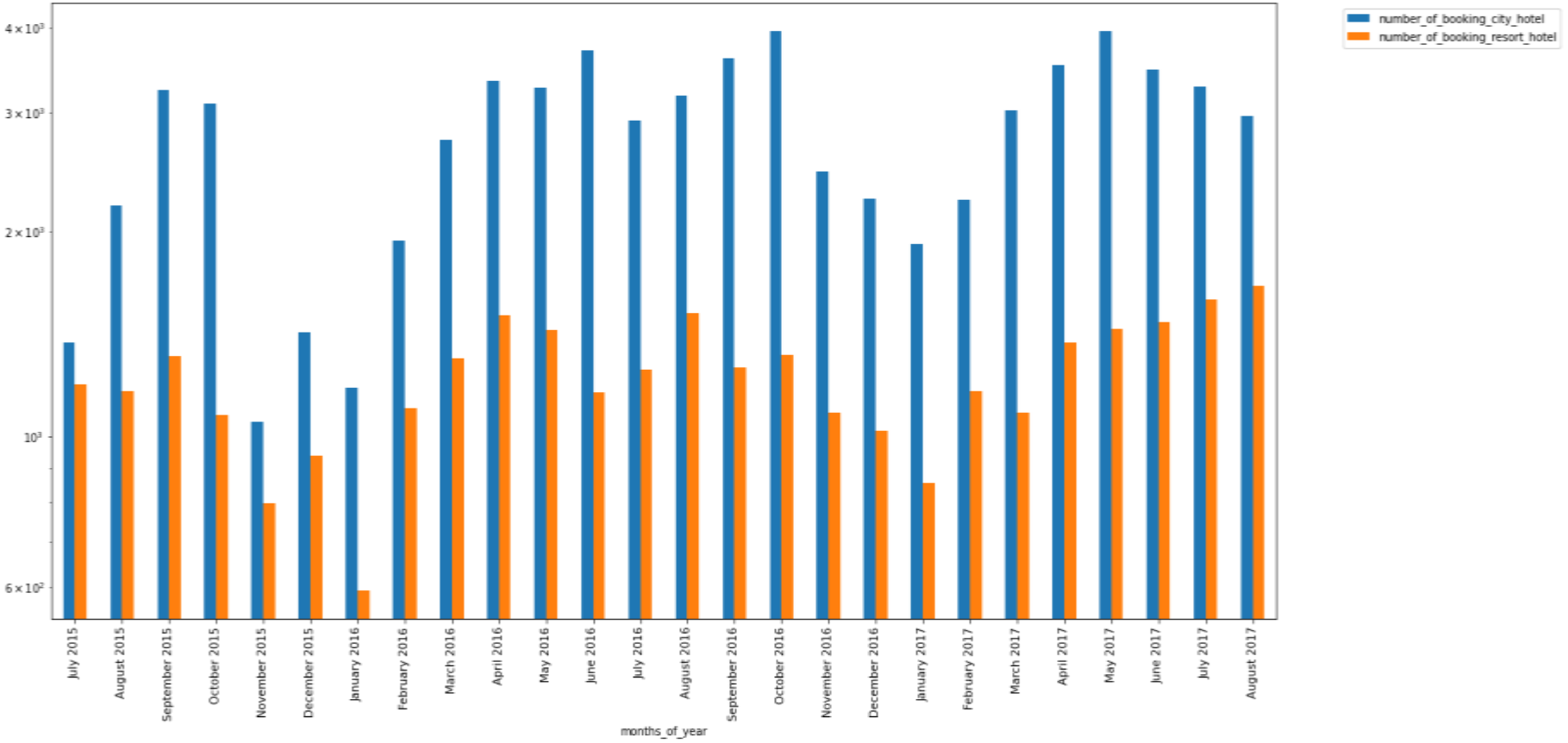
```
16496
```

Total rows which has null values = 16496

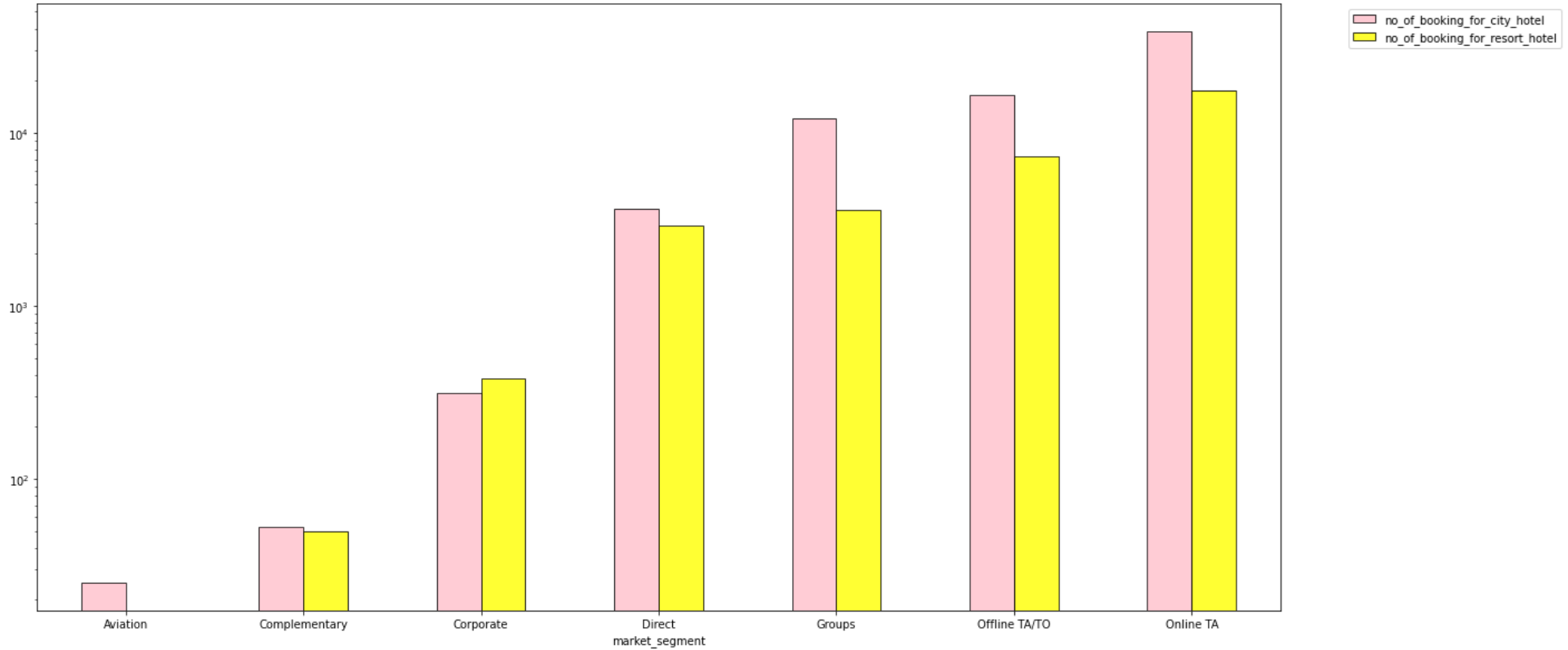
Total percentage of rows which has null values = 13.6 %

| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| 0 | hotel | 119390 non-null | object |
| 1 | is_canceled | 119390 non-null | int64 |
| 2 | lead_time | 119390 non-null | int64 |
| 3 | arrival_date_year | 119390 non-null | int64 |
| 4 | arrival_date_month | 119390 non-null | object |
| 5 | arrival_date_week_number | 119390 non-null | int64 |
| 6 | arrival_date_day_of_month | 119390 non-null | int64 |
| 7 | stays_in_weekend_nights | 119390 non-null | int64 |
| 8 | stays_in_week_nights | 119390 non-null | int64 |
| 9 | adults | 119390 non-null | int64 |
| 10 | children | 119386 non-null | float64 |
| 11 | babies | 119390 non-null | int64 |
| 12 | meal | 119390 non-null | object |
| 13 | country | 118902 non-null | object |
| 14 | market_segment | 119390 non-null | object |
| 15 | distribution_channel | 119390 non-null | object |
| 16 | is_repeated_guest | 119390 non-null | int64 |
| 17 | previous_cancellations | 119390 non-null | int64 |
| 18 | previous_bookings_not_canceled | 119390 non-null | int64 |
| 19 | reserved_room_type | 119390 non-null | object |
| 20 | assigned_room_type | 119390 non-null | object |
| 21 | booking_changes | 119390 non-null | int64 |
| 22 | deposit_type | 119390 non-null | object |
| 23 | agent | 103050 non-null | float64 |
| 24 | company | 6797 non-null | float64 |
| 25 | days_in_waiting_list | 119390 non-null | int64 |
| 26 | customer_type | 119390 non-null | object |
| 27 | adr | 119390 non-null | float64 |
| 28 | required_car_parking_spaces | 119390 non-null | int64 |
| 29 | total_of_special_requests | 119390 non-null | int64 |
| 30 | reservation_status | 119390 non-null | object |
| 31 | reservation_status_date | 119390 non-null | object |

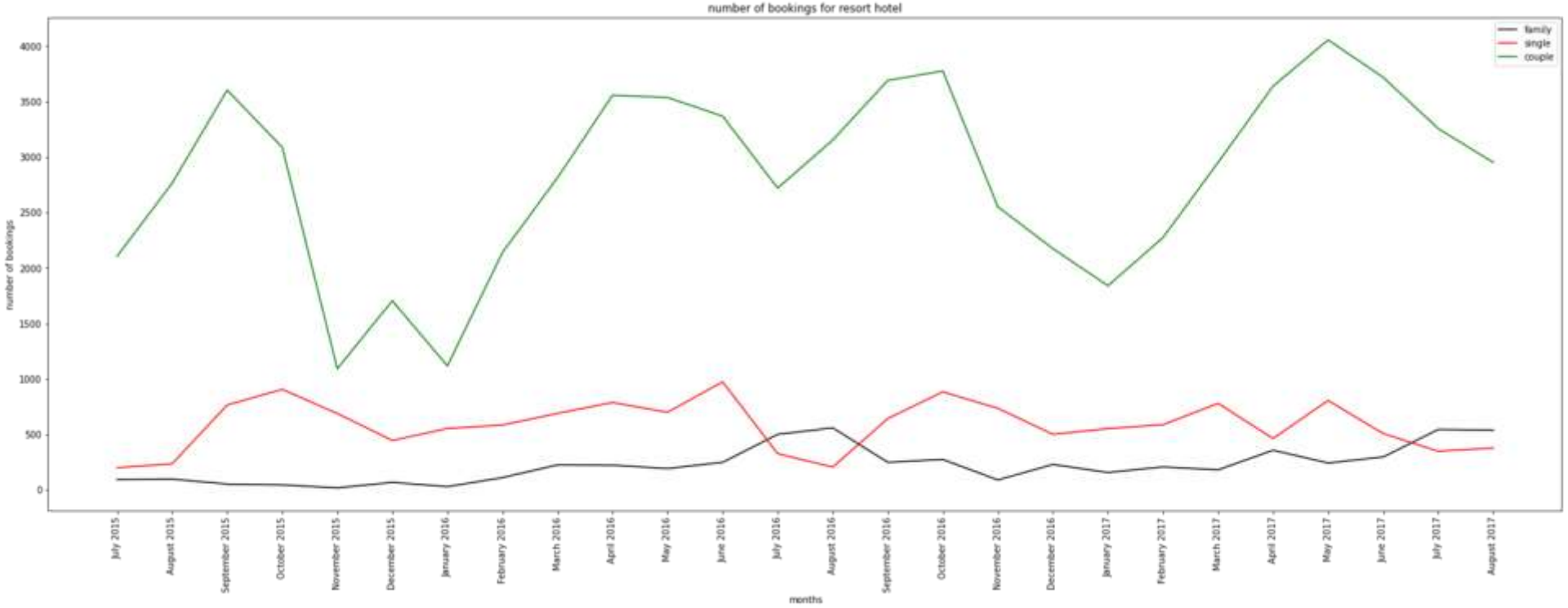
# Total number of bookings



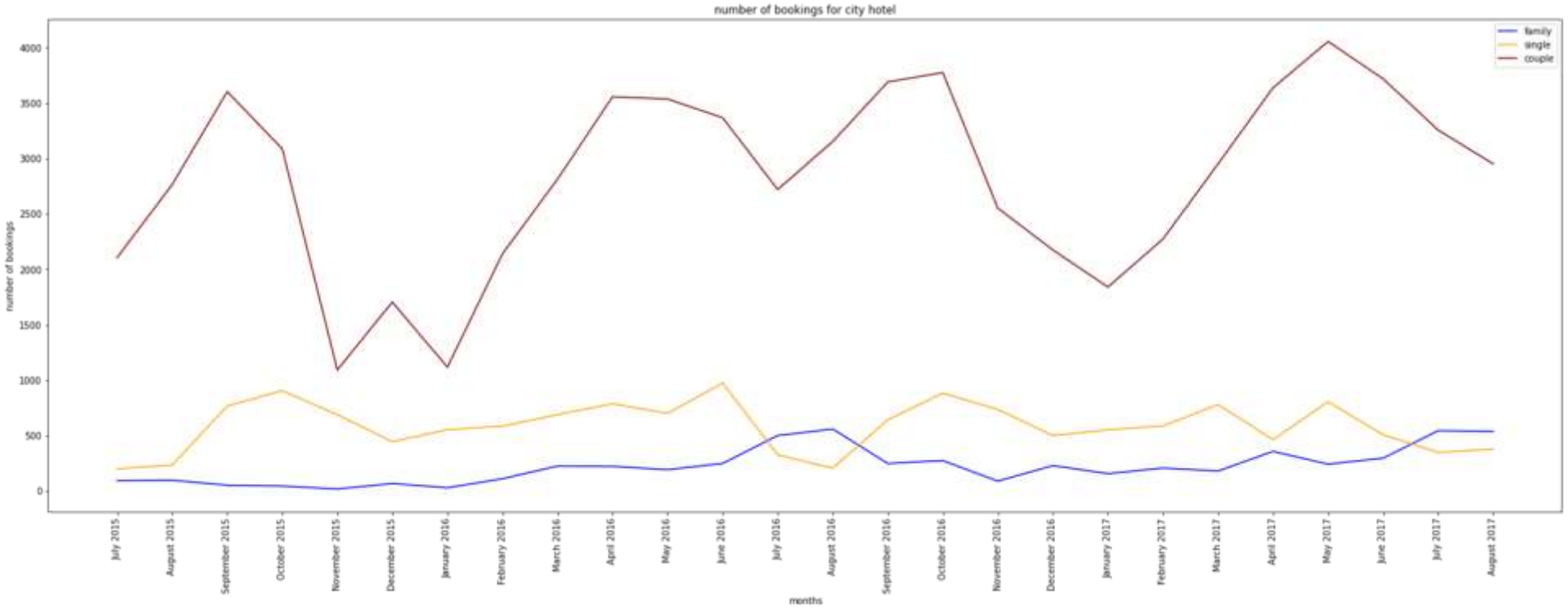
# Bookings from different market segments



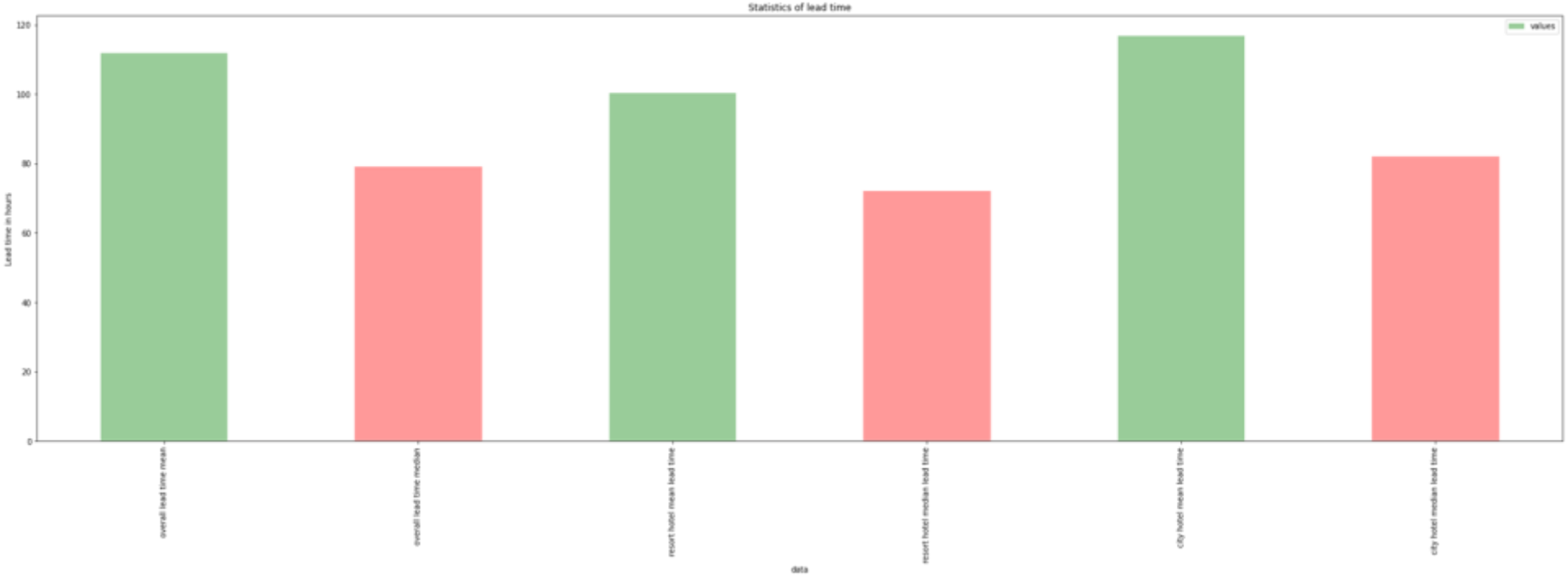
# Average number of booking of couples, families and single for resort hotel



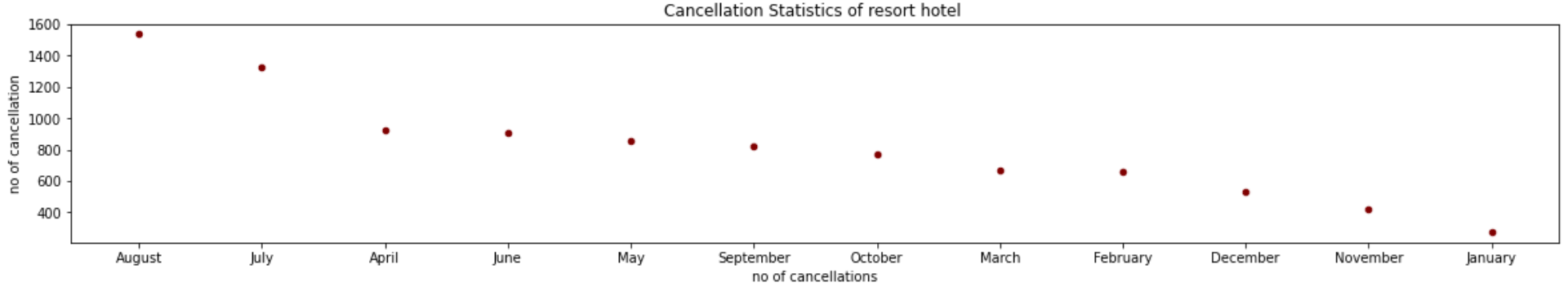
# Average number of booking of couples, families and single for city hotel

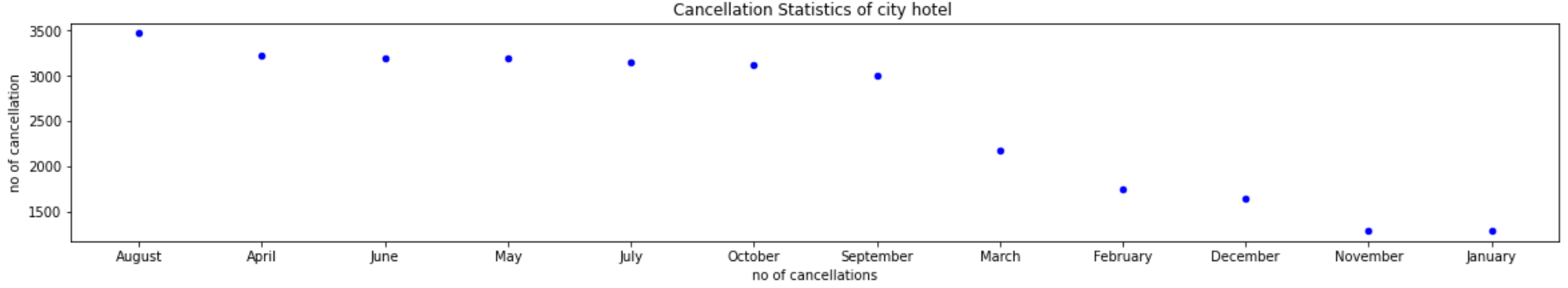


# Statistics of lead time

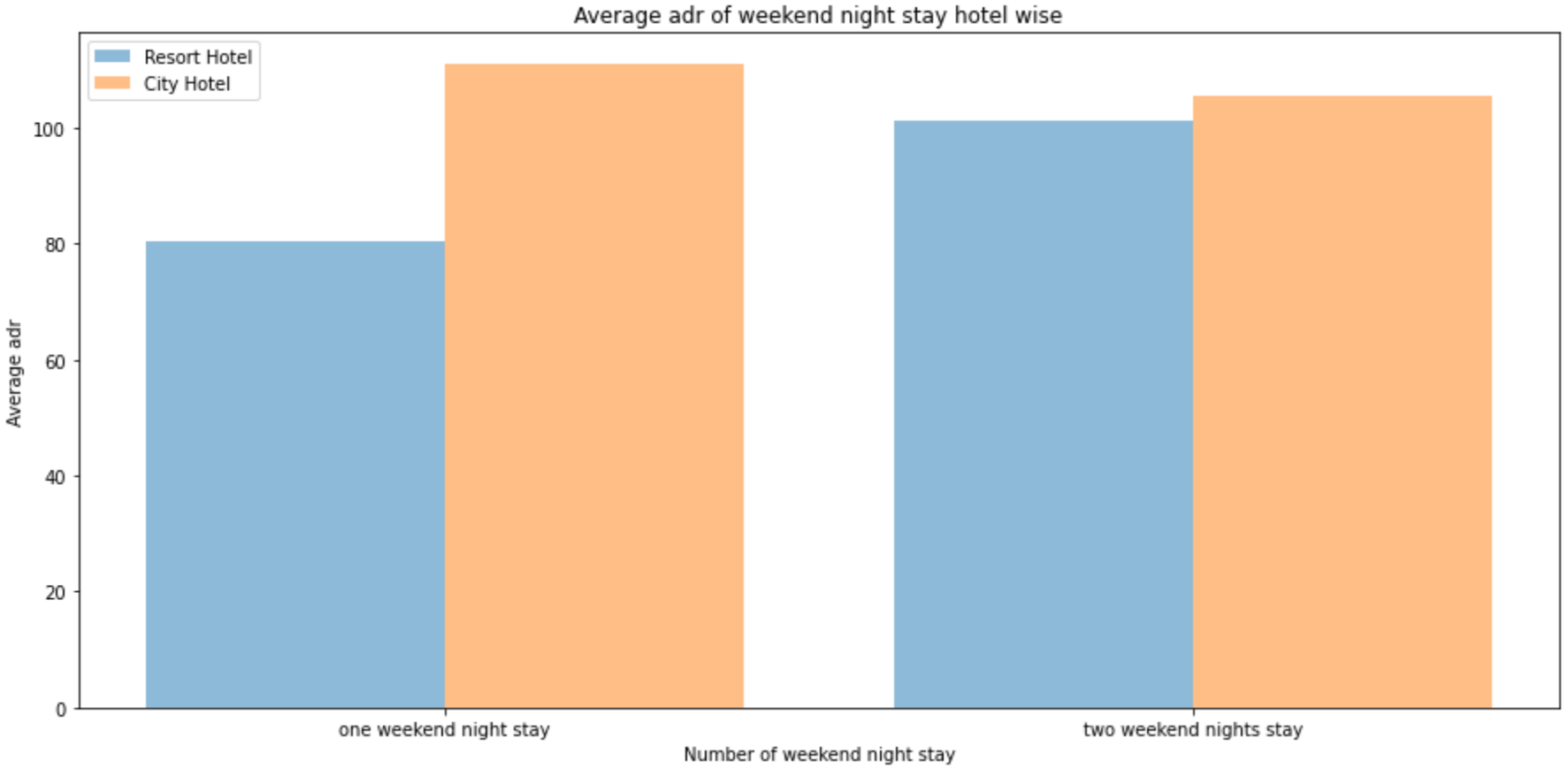


# Cancellation statistics

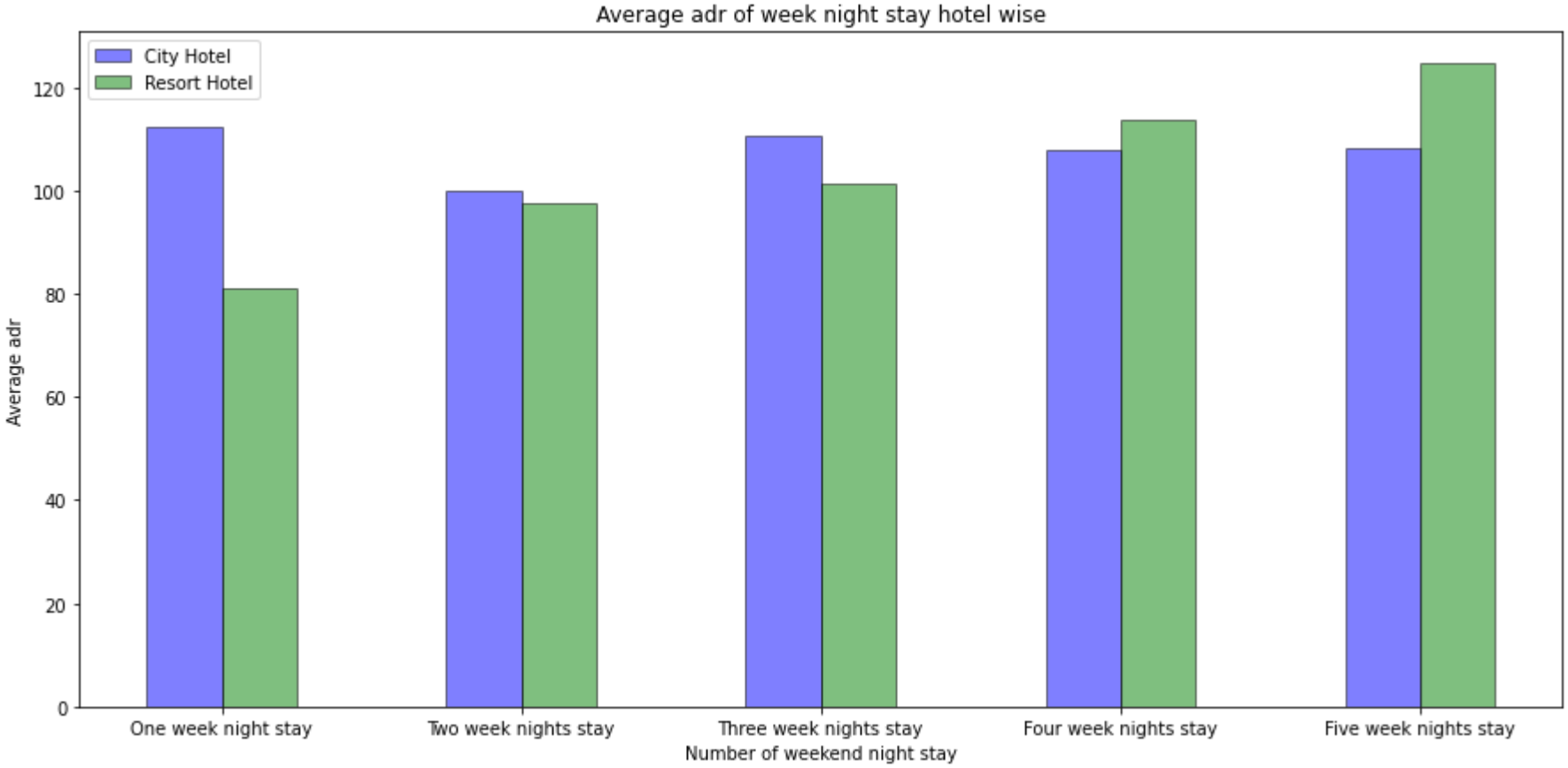


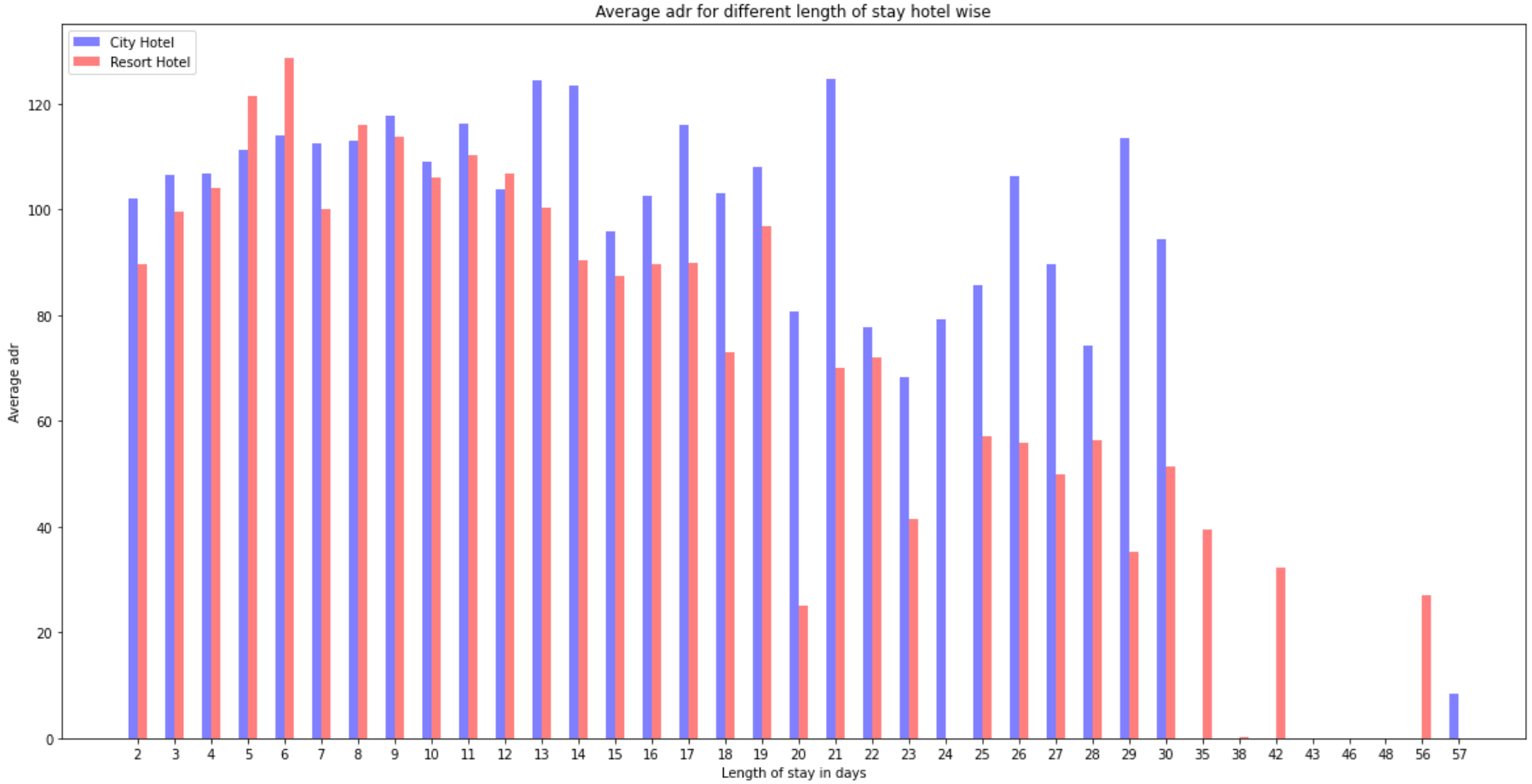
# Adr for weekend nights



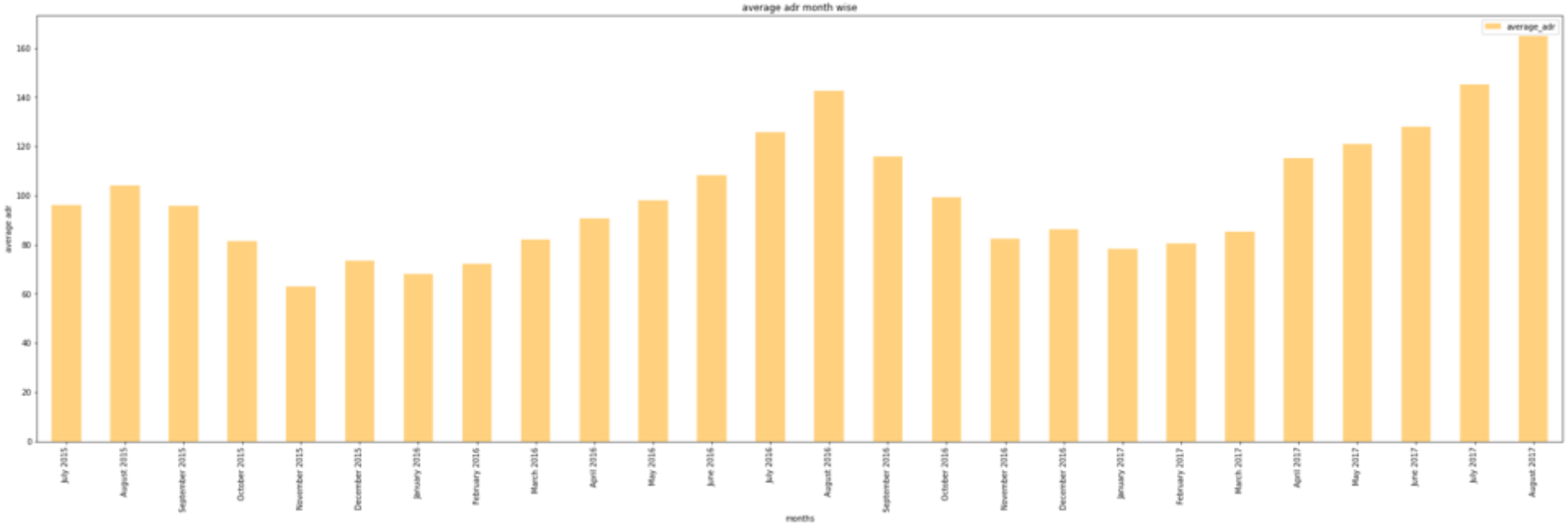
# Adr for weekday nights



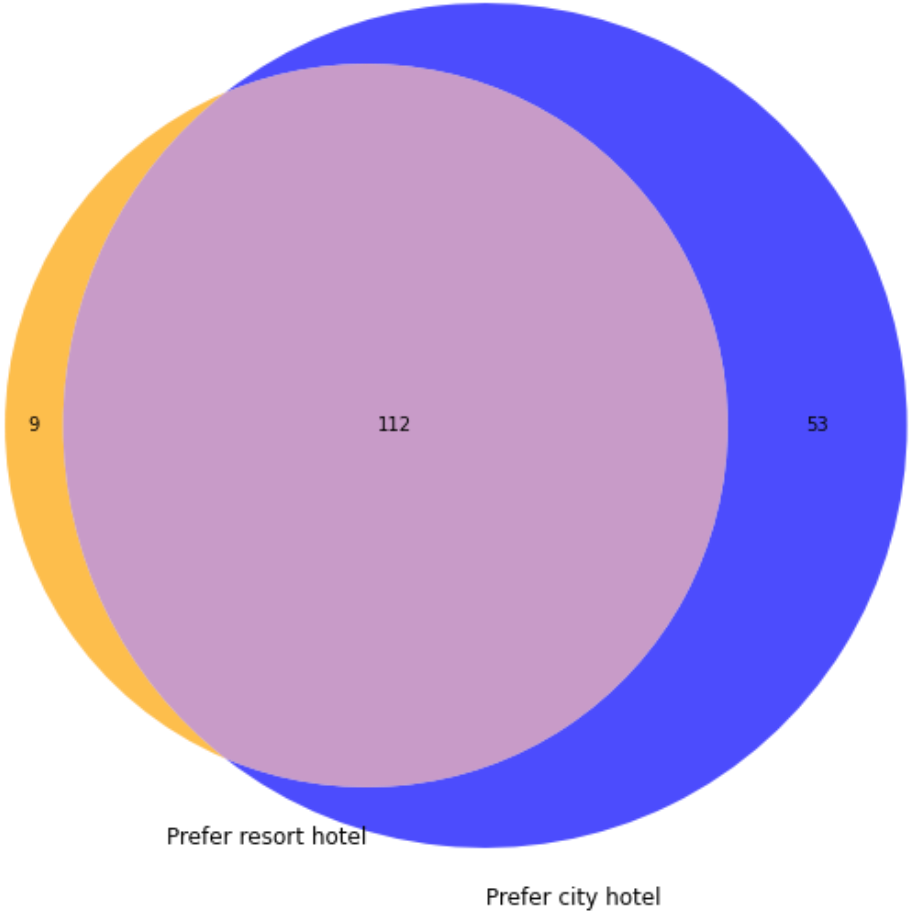
# Adr vs. length of stay



# Variation of adr with time

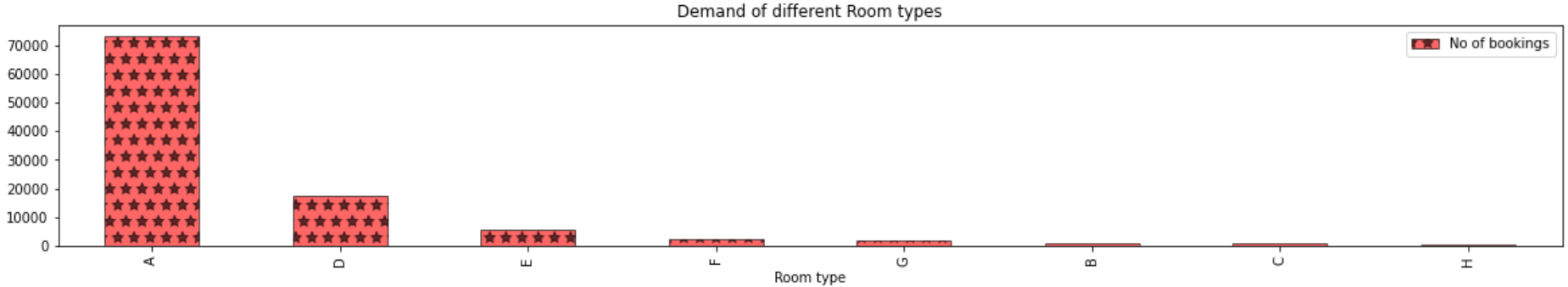


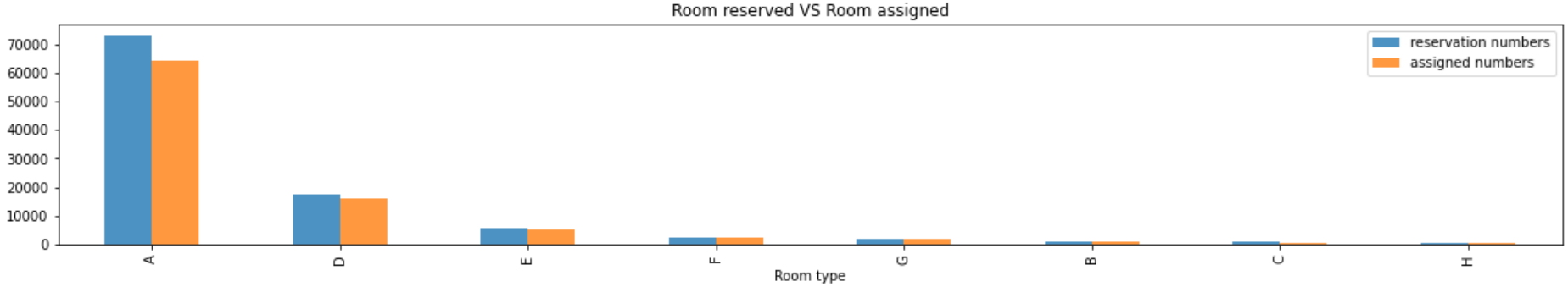
# Choice of hotel from people of different countries



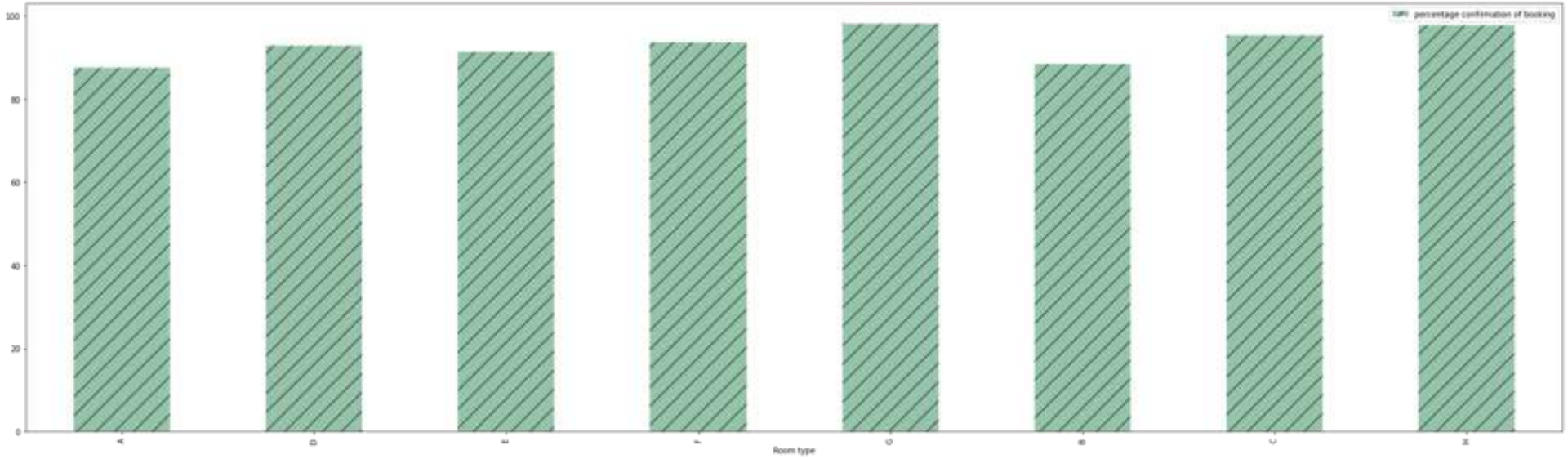
# Bookings from different countries

# Demand of different Room type

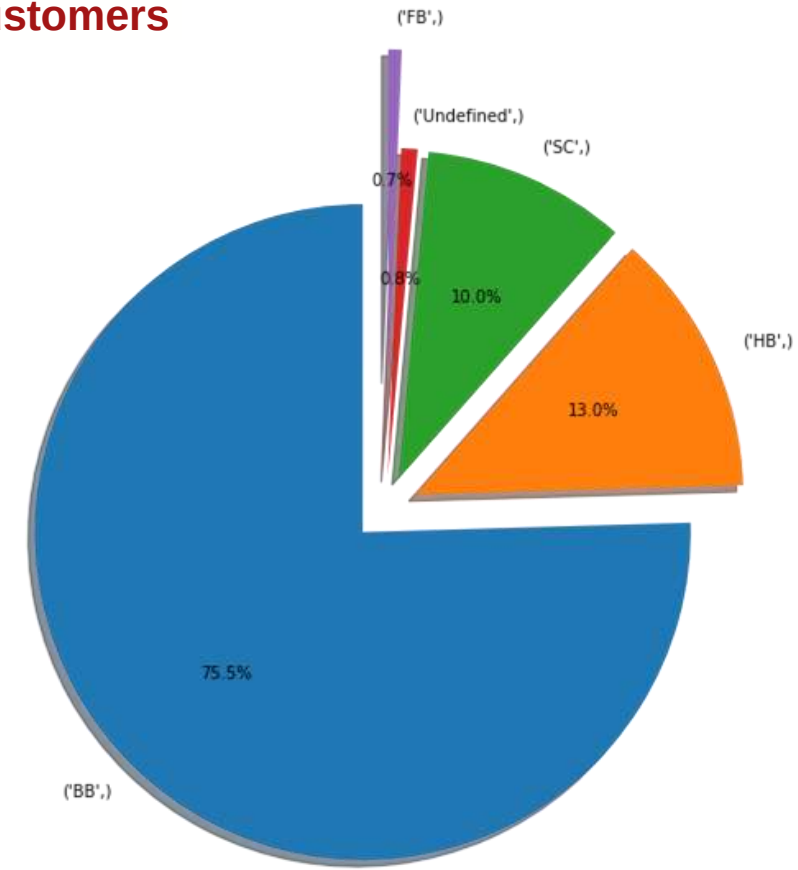


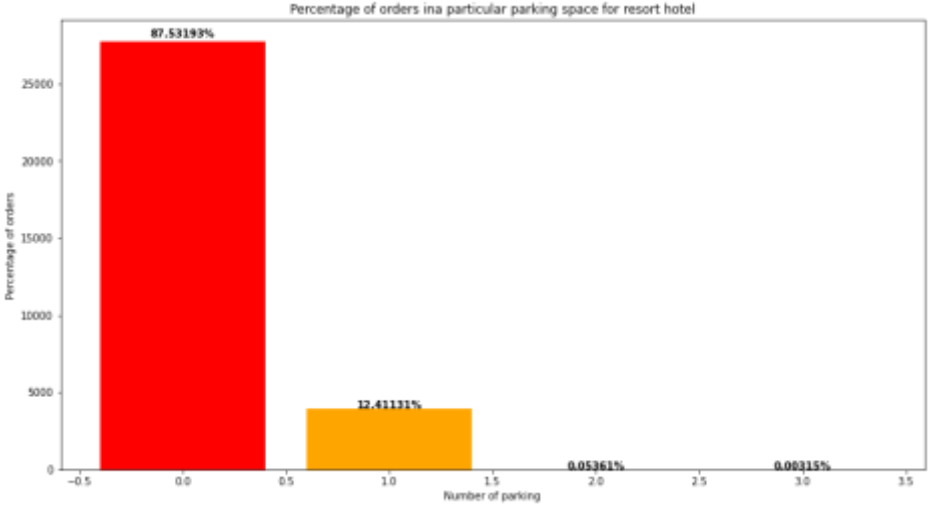
# Percentage of confirmed booking for various room types

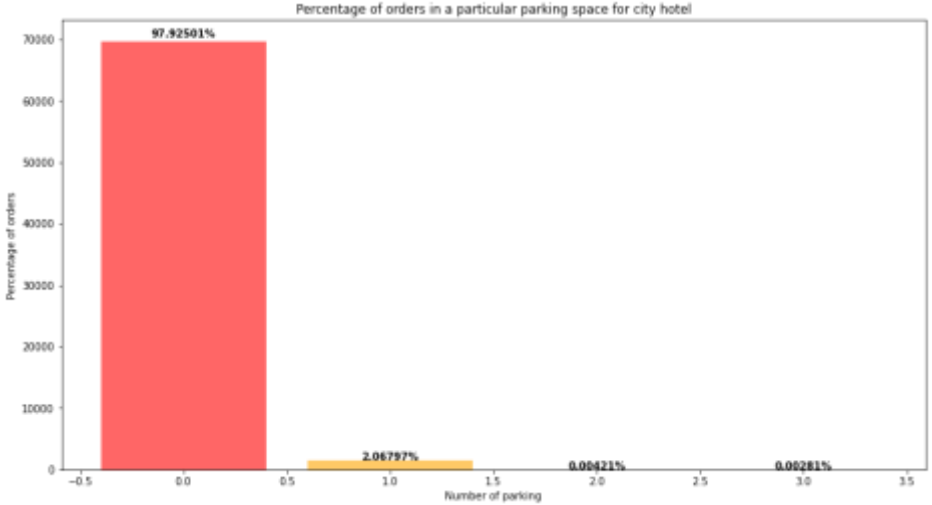


# Food preferences of customers



# Parking space demand analysis



# Challenges

- Data Cleaning
- More Data or Better Data
- Understanding the business problem
- Team coordination
- Lack of clarity in data

# Conclusion

- Comparing the magnitude of booking we see that city hotel has always scored higher booking counts.
- Number of booking start increasing from spring to summer season and then decrease till winter season.
- TA(travel agents) are the major contributors of booking
- The most number of cancelled booking is in the month of August.
- Adr starts decreasing from month of september and it gets its minimum value in between november and january, after that adr starts increasing and it gets its peak in august.
- BB is most preferred and FB is the least preferred meal.
- Demand of **type A** room is highest followed by **type D** room.

THANK YOU