# Capstone Project

## ‘Hotel Bookings Analysis’

### (Team Mrityunjay)

### Cohort Seattle


**By-**

**Luv Mutreja**

**Jivan Jadhao**

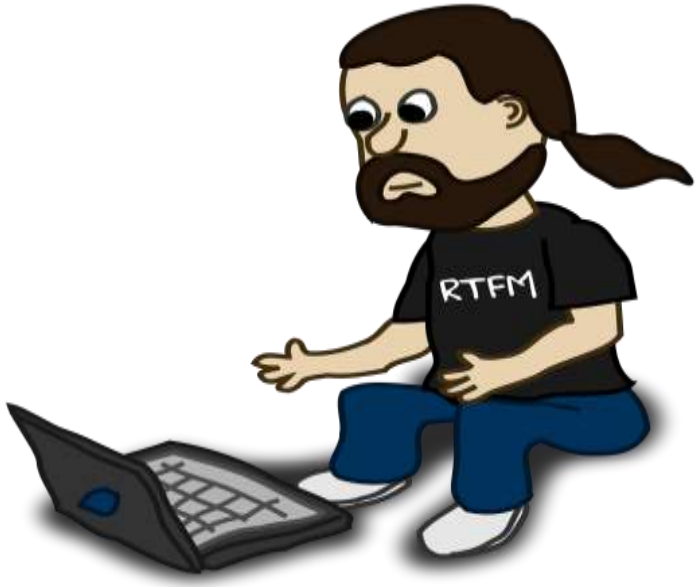
**Hitesh Kumar Khurana**

# TABLE OF CONTENT:

- Data Collection And Understanding
- Data Cleaning and Data Handling
- Exploratory Data Analysis
- Conclusion

# Data Collection and Understanding:

- We've collected our Data in the form of an Excel Sheet.
- After collecting the Data, it's very prerequisite to observe the Dataset thoroughly.
- Subsequently, we observed that our Dataset has a total of 119390 rows and 32 columns.



# The dataset we have has following columns:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Hotel | arrival_date_month | stays_in_week_nights | is_repeated_guest | country | assigned_room_type | company | required_car_parking_spaces |
| is_canceled | arrival_date_week_number | adults | previous_cancellations | market_segment | booking_changes | days_in_waiting_list | total_of_special_requests |
| lead_time | arrival_date_day_of_month | children | previous_bookings_not_canceled | distribution_channel | deposit_type | customer_type | reservation_status |
| arrival_date_year | stays_in_weekend_nights | babies | meal | reserved_room_type | agent | adr | reservation_status_date |

# Data Cleaning and Data Handling:

- After Observing Dataset, we found that there are 4 columns containing NaN or Null Values *i.e.* Company, Agent, Country, and Children.
- Since we have limited numerical value data, we didn't drop any of the aforementioned columns. Instead, we assign '0' to the Null value in column Children and 'unknown' to column Country.
- Column 'is_canceled' has only 2 values, Value 0, which means the booking is not cancelled and Value 1, which means the booking is cancelled. As we can see that 44224 bookings are canceled, so we will not be using the 44224 rows for our analysis. Only not canceled rows will be used for analysis.

```
[ ] #finding out first which columns have null values
    #using 'sum' function will show us how many nulls are found in each column of the dataset
    hoteldf1.isnull().sum()
```

```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        4
babies                          0
meal                            0
country                       488
market_segment                  0
distribution_channel            0
is_repeated_guest               0
```

```
[ ] #replacing the null values in children column with 0
    hoteldf1.fillna({'children':0},inplace=True)

[ ] #replacing null countries with 'unknown'
    hoteldf1.fillna({'country':'unknown'},inplace=True)
```

```
#checking again for null values
hoteldf1.isnull().sum()
```

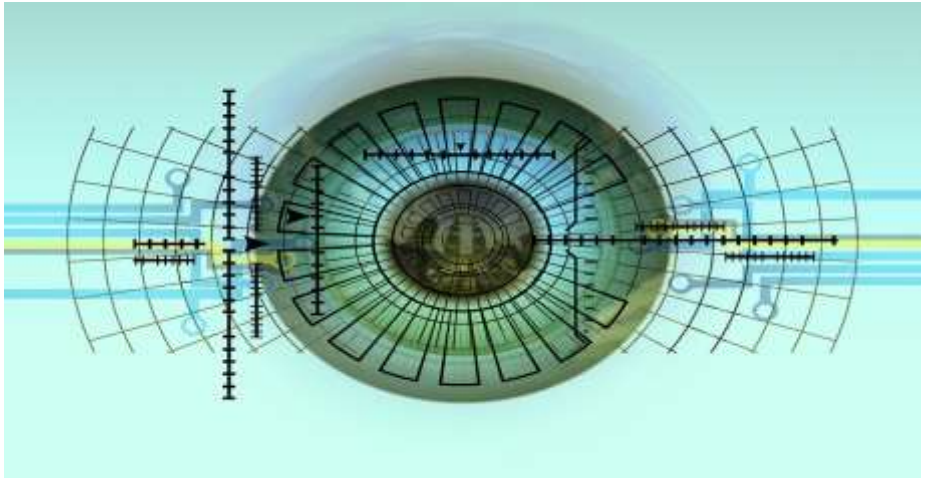
```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        0
babies                          0
meal                            0
country                         0
market_segment                  0
distribution_channel            0
is_repeated_guest               0
previous_cancellations          0
previous_bookings_not_canceled  0
reserved_room_type              0
assigned_room_type              0
booking_changes                 0
deposit_type                    0
agent                       16340
company                    112593
days_in_waiting_list            0
customer_type                   0
adr                             0
required_car_parking_spaces     0
total_of_special_requests       0
reservation_status              0
```

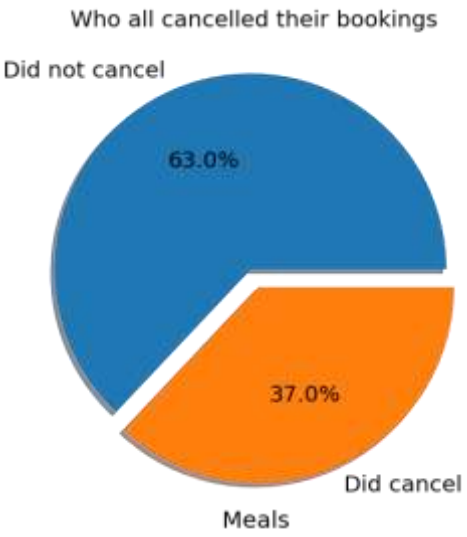

```
[ ] #viewing the values in adults columns and its values
    hoteldf1['adults'].value_counts()
```

```
2      89680
1      23027
3       6202
0        403
4         62
26         5
27         2
20         2
5          2
40         1
50         1
55         1
6          1
10         1
Name: adults, dtype: int64
```

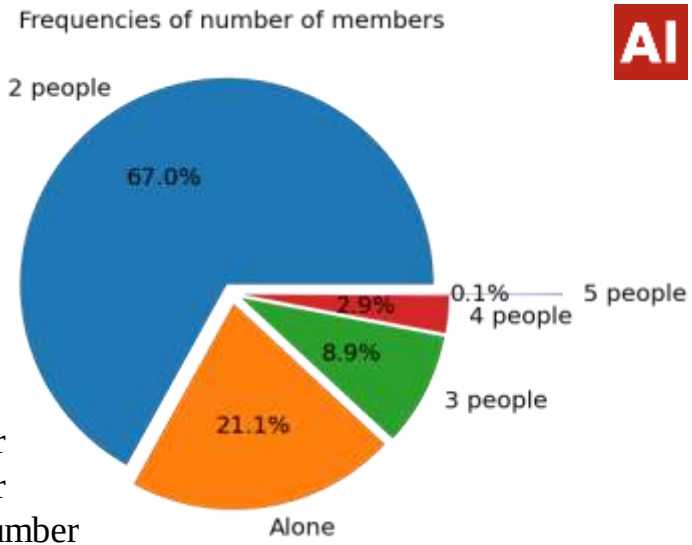
```
[ ] #creating a new column to store total members
    hoteldf1['total_members']=hoteldf1['adults'] + hoteldf1['children']
```

# Exploratory Data Analysis:

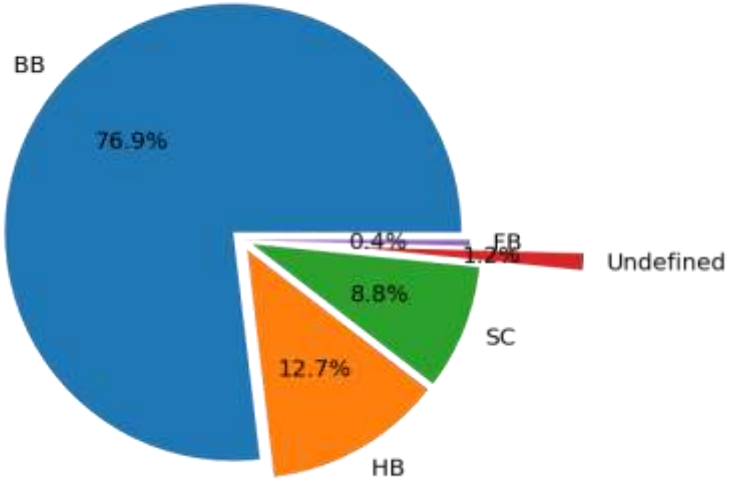





63% had 0 in is canceled column
Rest
37% had 1 in the same column




21% came alone
67% came 2 in number
8.9% came 3 in number
2.9% came 4 in number
And 0.1% came 5 in number




76.9% people opted for (BB)Bed and Breakfast
12.7% people opted for (HB)Half Board i.e., Breakfast and Dinner
8.8% people opted for Self catering
For 1.2% people meal was undefined
0.4% people opted for (FB)Full Board i.e., Breakfast, Lunch and Dinner

# What is the best time of year to book a hotel room ?

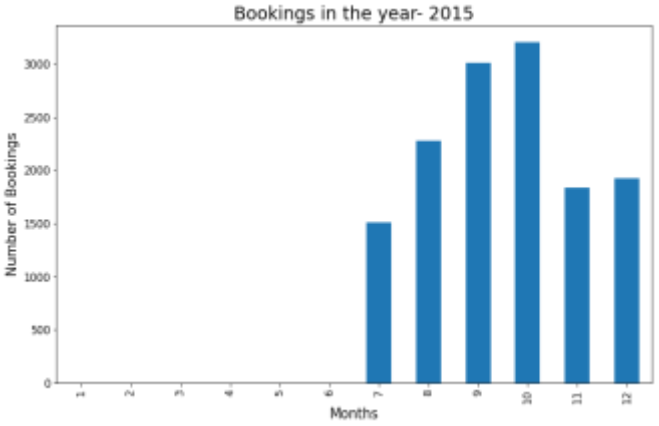
This question has two aspects-

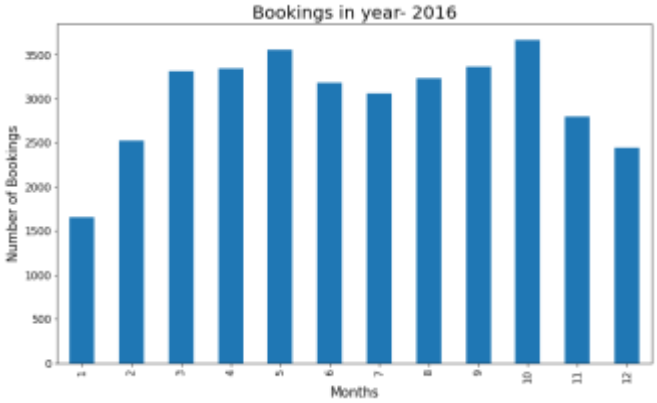
#From customer's point of view-

From customers point of view the best time to book a hotel room is when it is the cheapest i.e. during off seasons, when not too many people book hotel rooms.

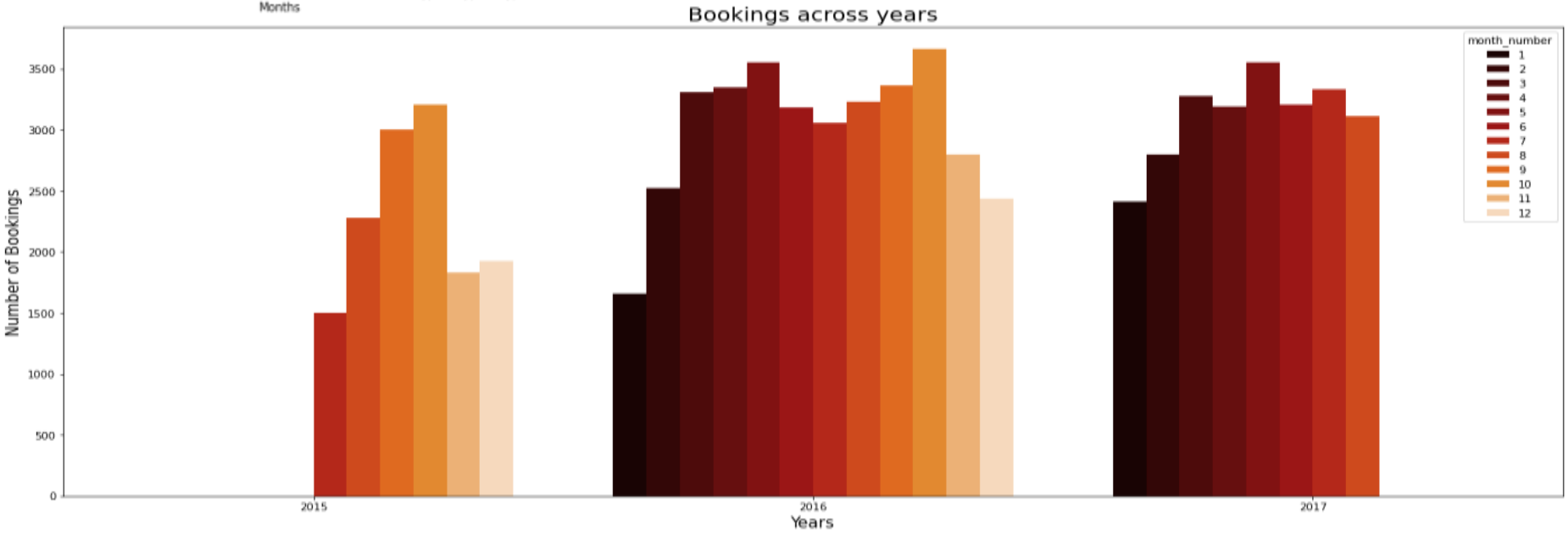
#From hotel's point of view-

From hotel's point of view the best time to book the hotel is during onn seasons i.e when too many people are booking rooms and the demand for rooms is high, at that time the hotel can charge high prices from customers, because the supply is same but the demand is relatively high.

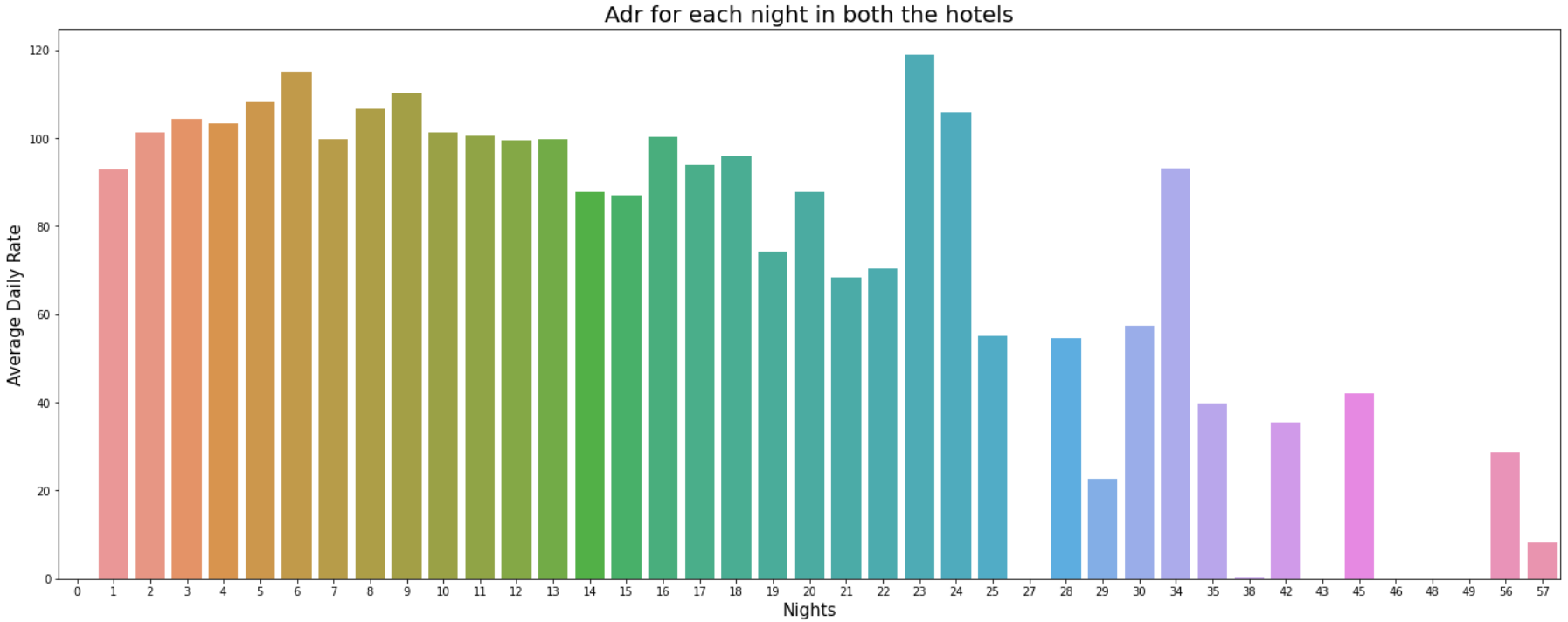

Bookings in the year- 2015
Number of Bookings
Months


Bookings in year- 2016
Number of Bookings
Months

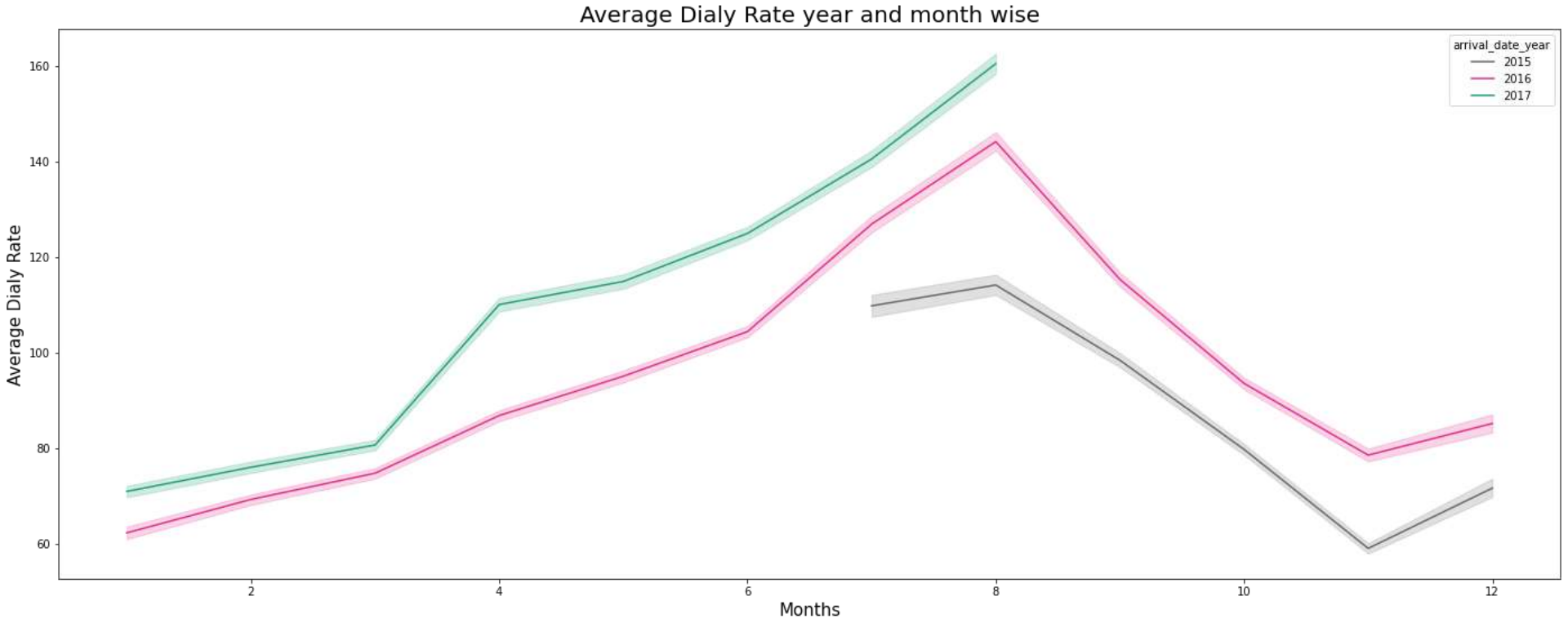

Bookings in year- 2017
Number of Bookings
Months


Bookings across years
Number of Bookings
Years
2015
2016
2017
month_number
1
2
3
4
5
6
7
8
9
10
11
12

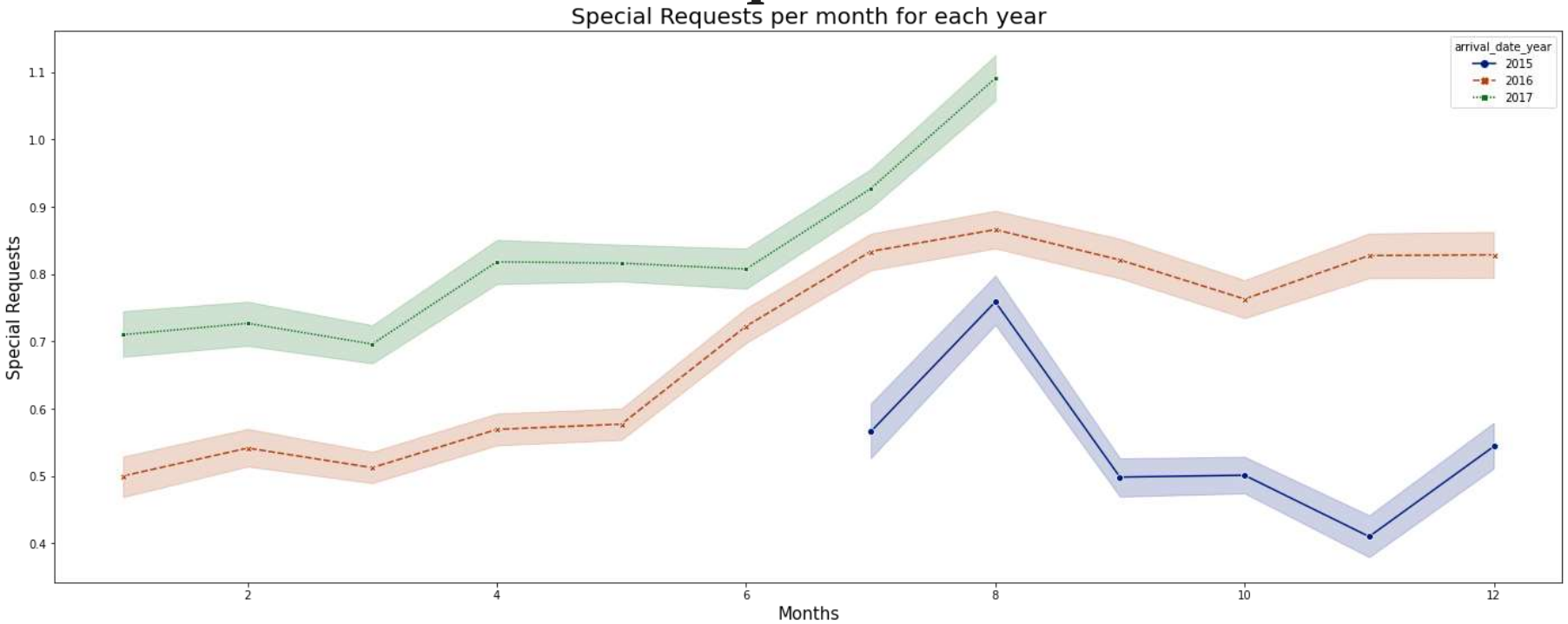
# What is the optimal length of stay in order to get the best daily rate?



# Average Daily Rate year and month wise

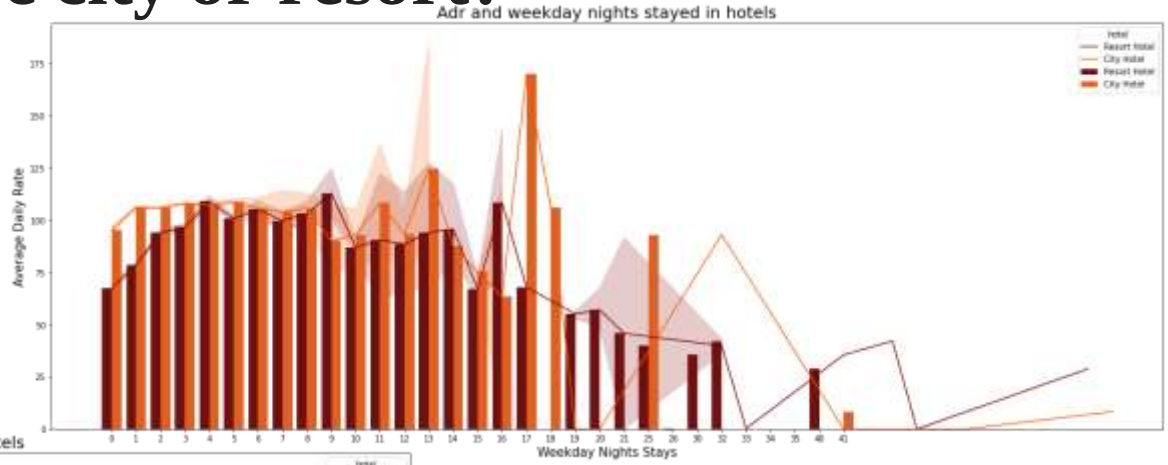


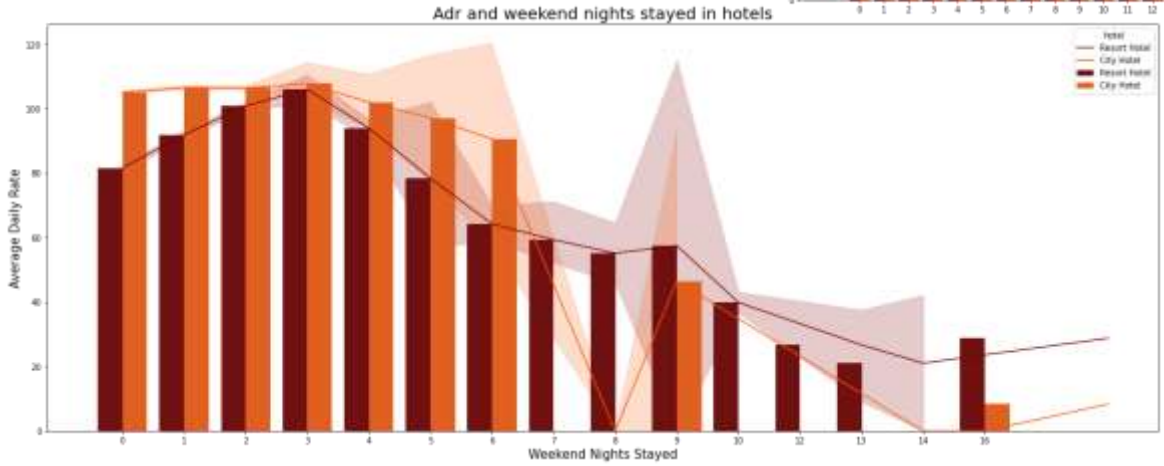
# How to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?



# Which hotel to choose city or resort?

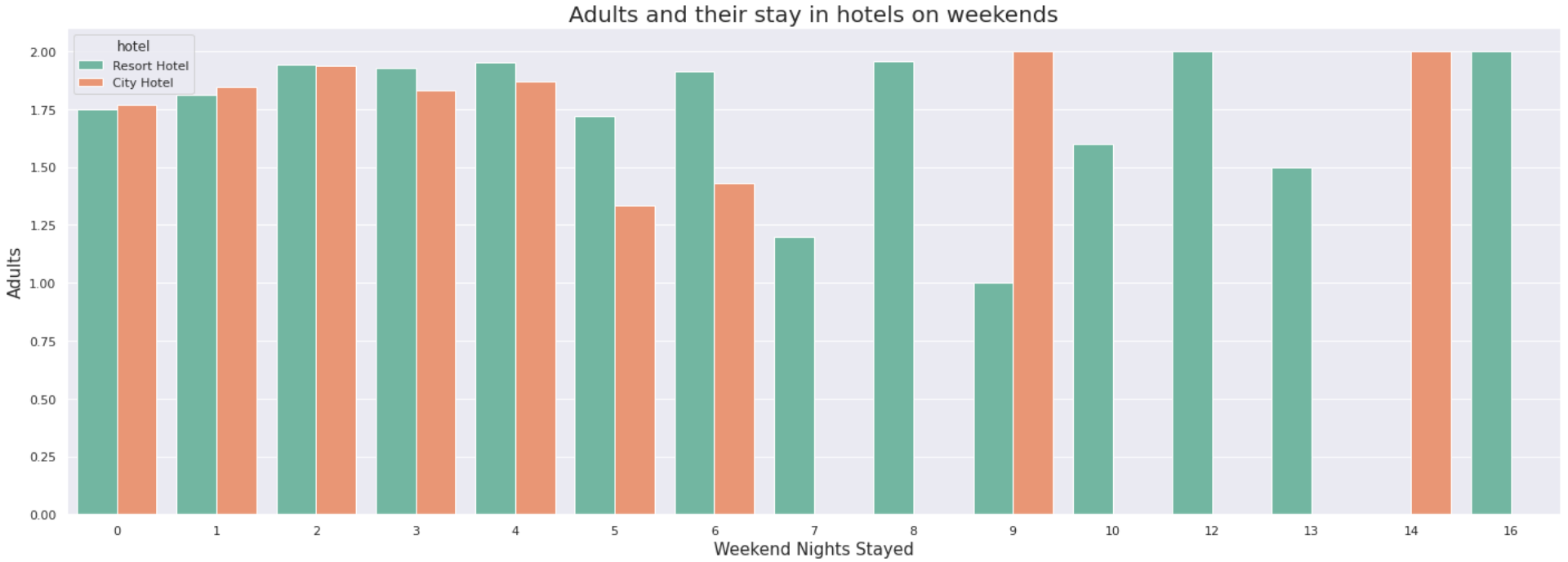
City Hotel's per day price fore weekdays is quite high as compared to Resort hotel



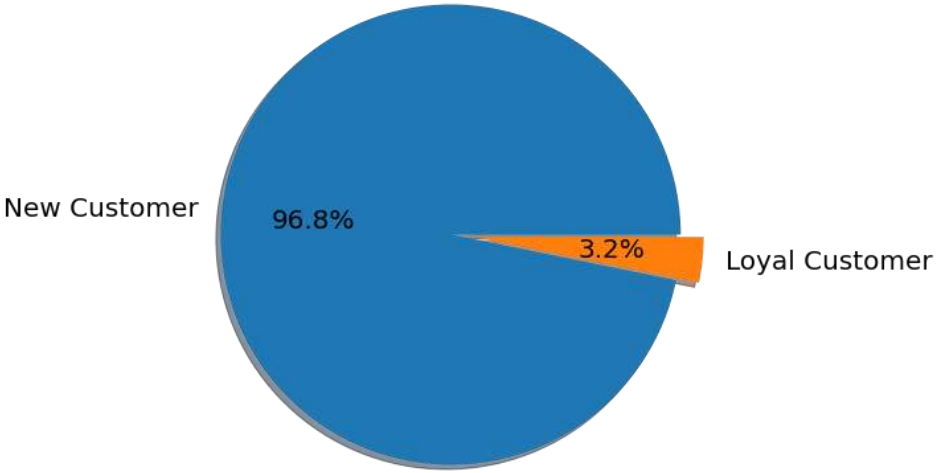




City Hotel's per day price for weekend is more than that of Resort hotel, but there is not much difference. Which was not the case for weekday's per day price.

# Which is the favourite place to stay in weekend nights for adults?

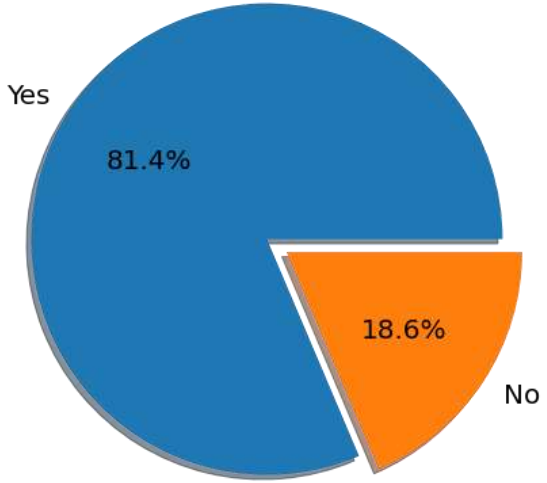


How many guests are repeated guests



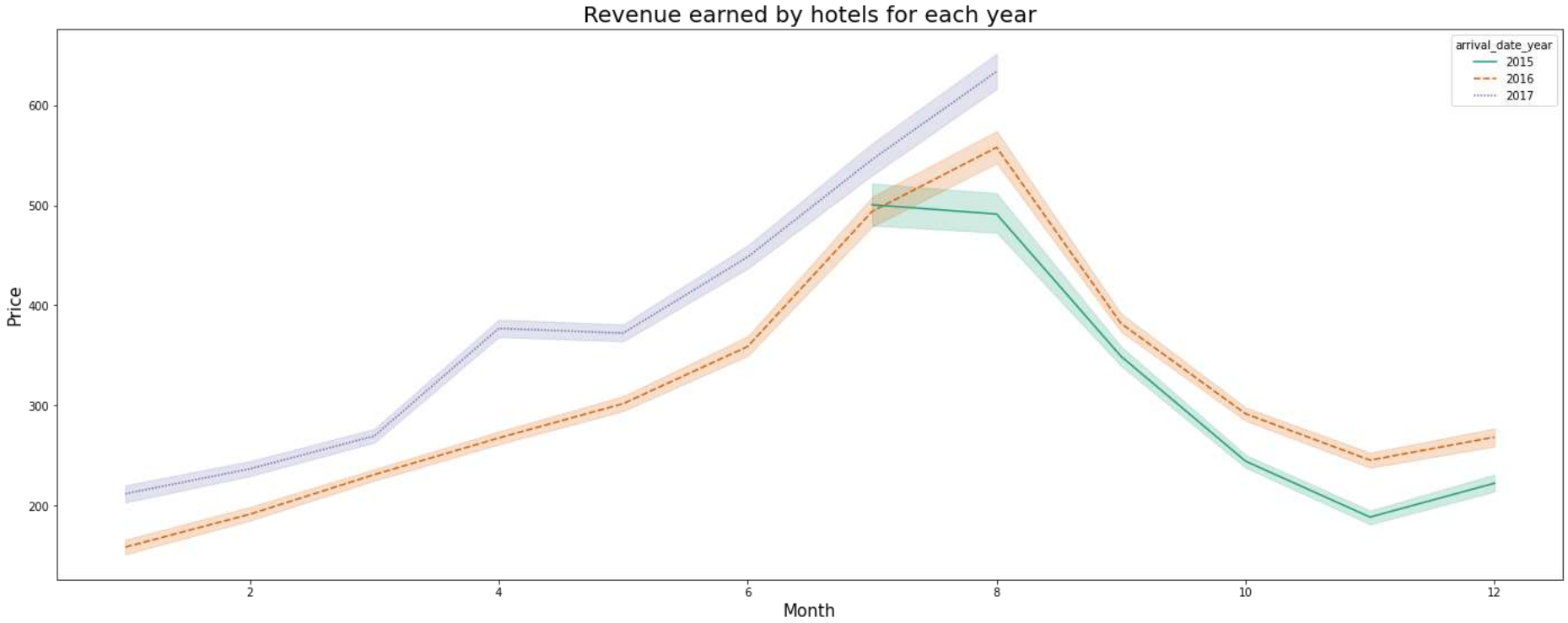

3.2% customers are repeated guests
Rest
96.8% customers are new

Where people assigned the room they wanted



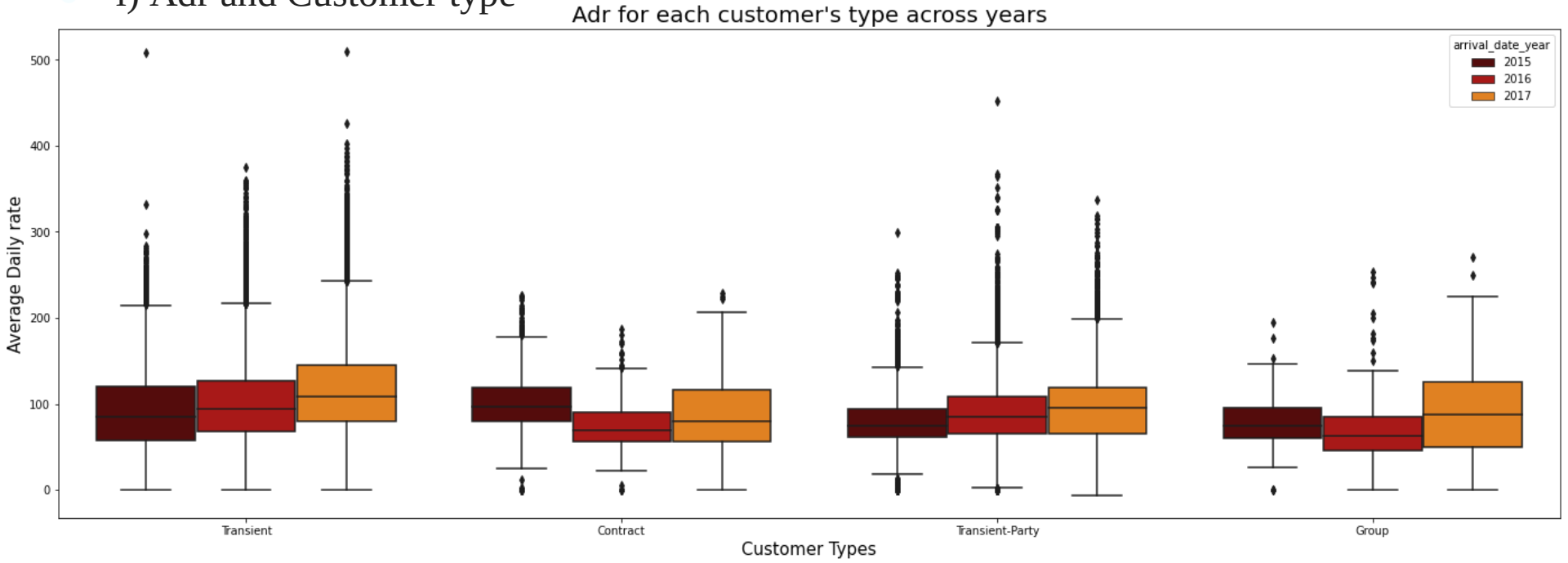

81.4% customers were allotted the room they reserved
Rest
18.6% did not receive the room they wanted

# Revenue earned by hotels for each year

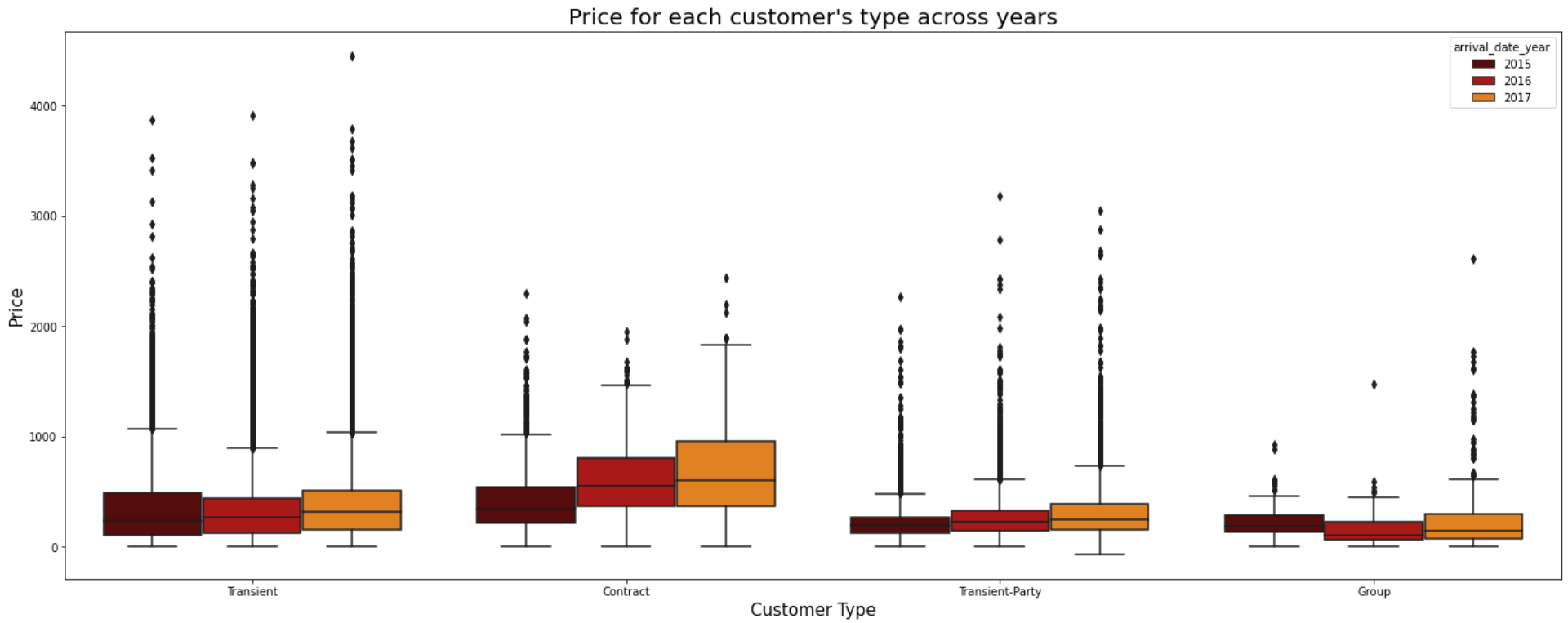


# From which customer's type did hotels earned the most revenue?

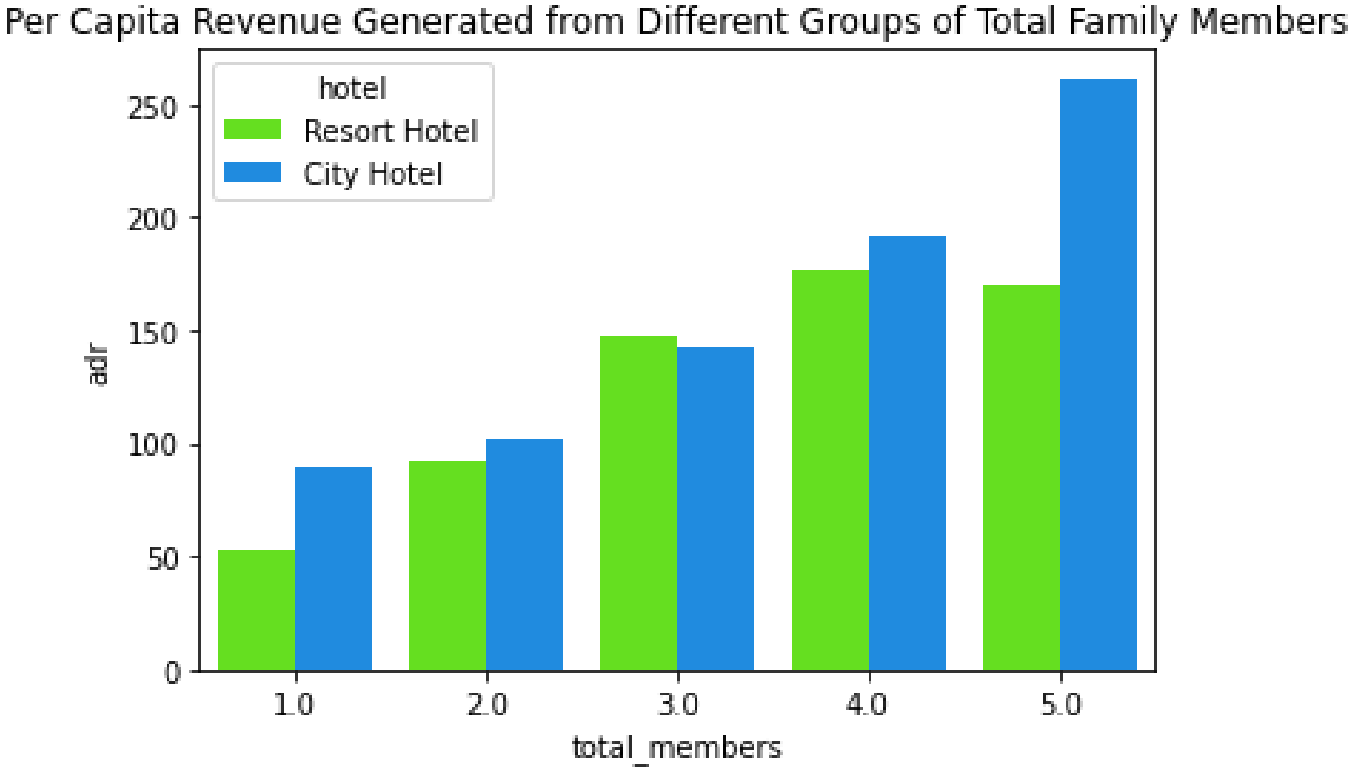
- For this we need to compare 2 things-
- i) Adr and Customer type



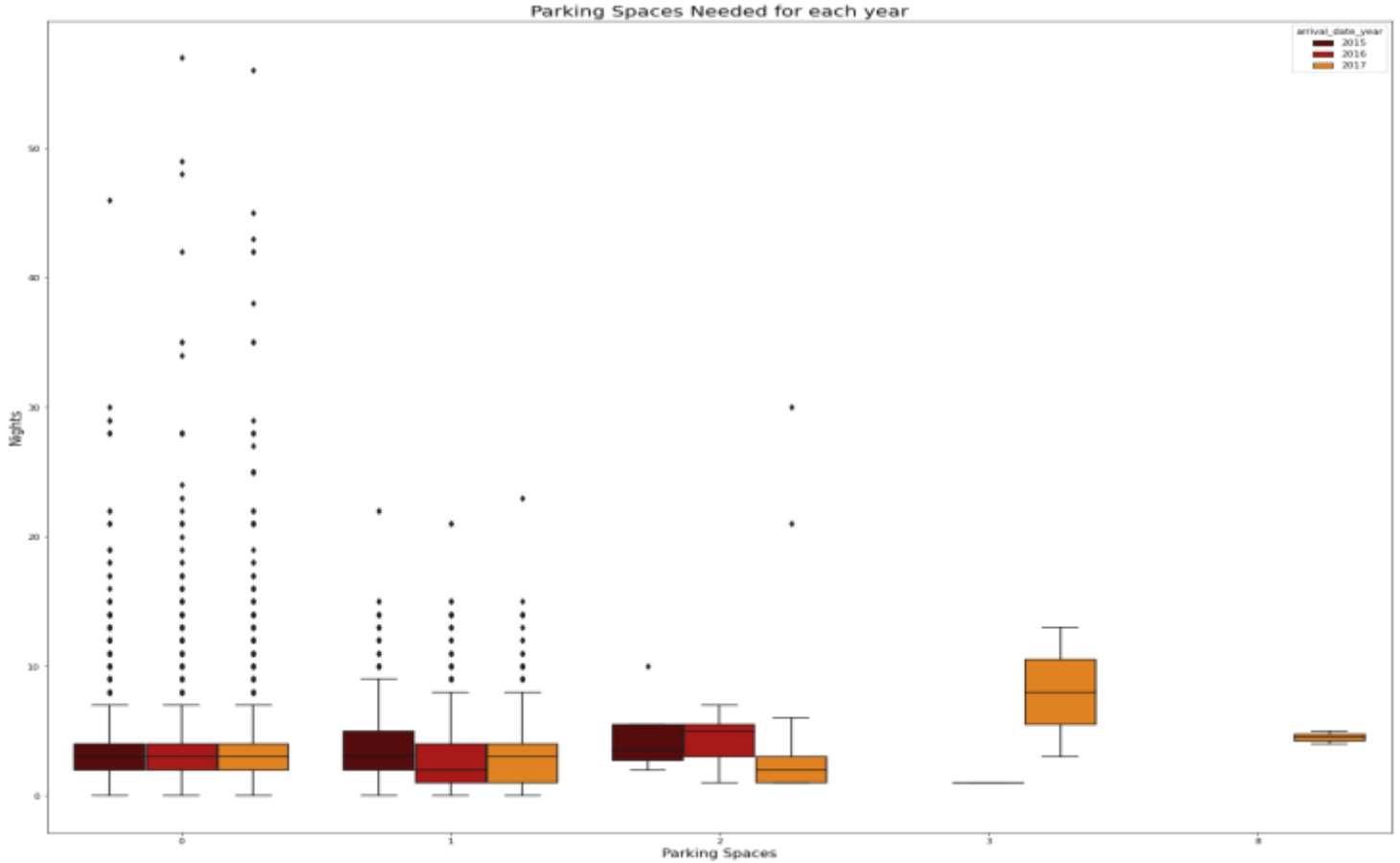
# ii) Price and Customer type



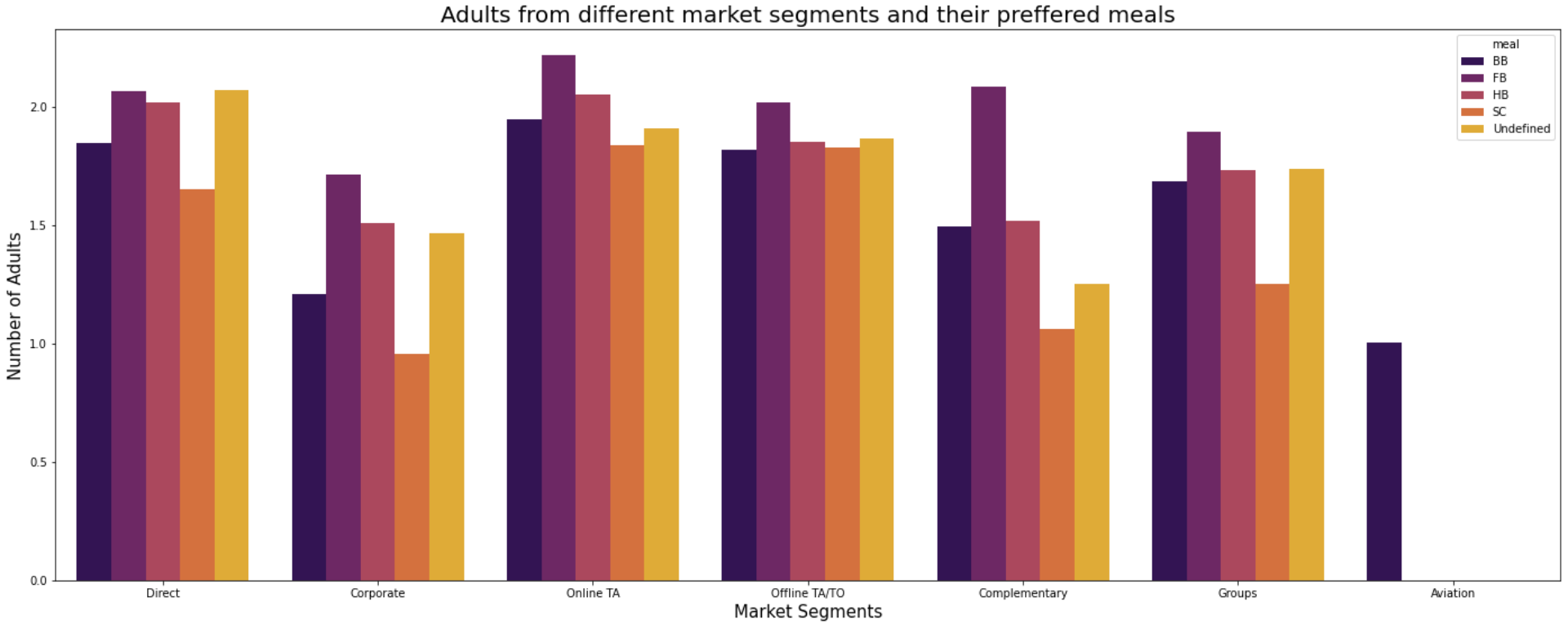
# Which Group of Family Members Generates a Large Amount of Revenue for Hotels?



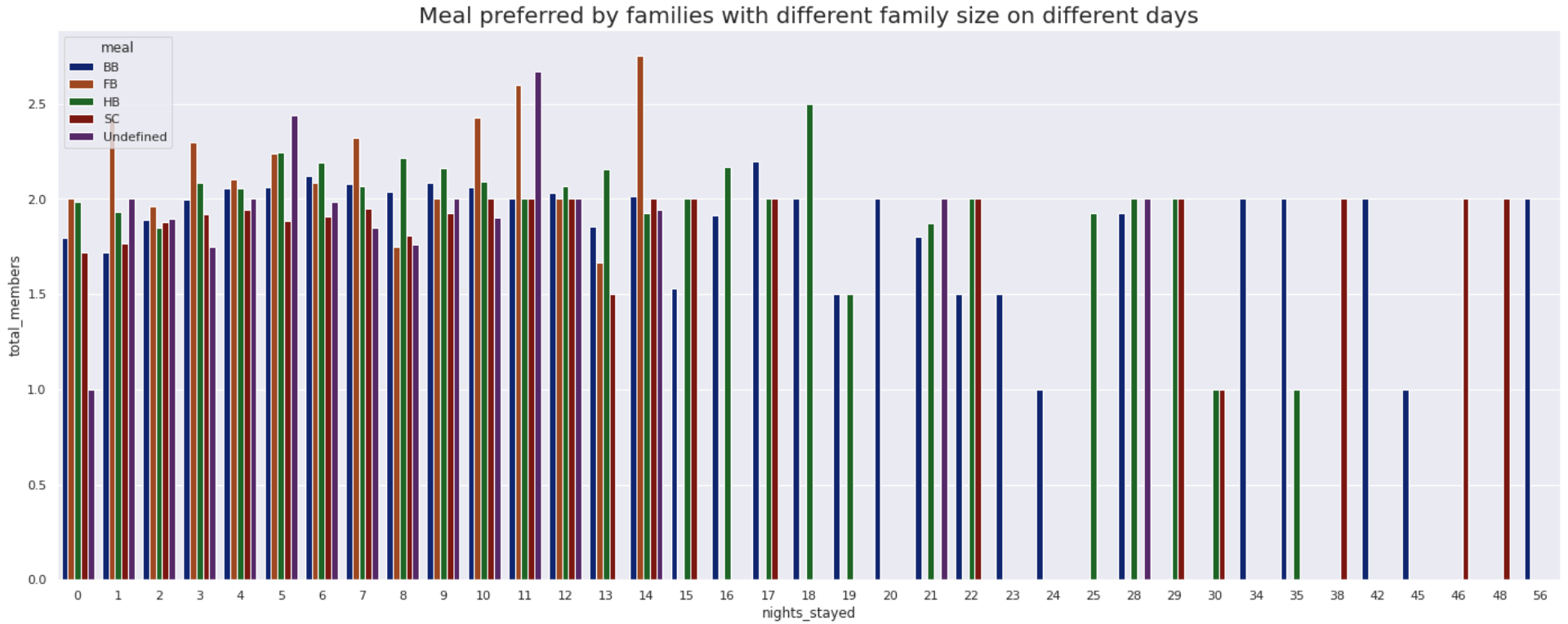
# Car parking Spaces required in hotels



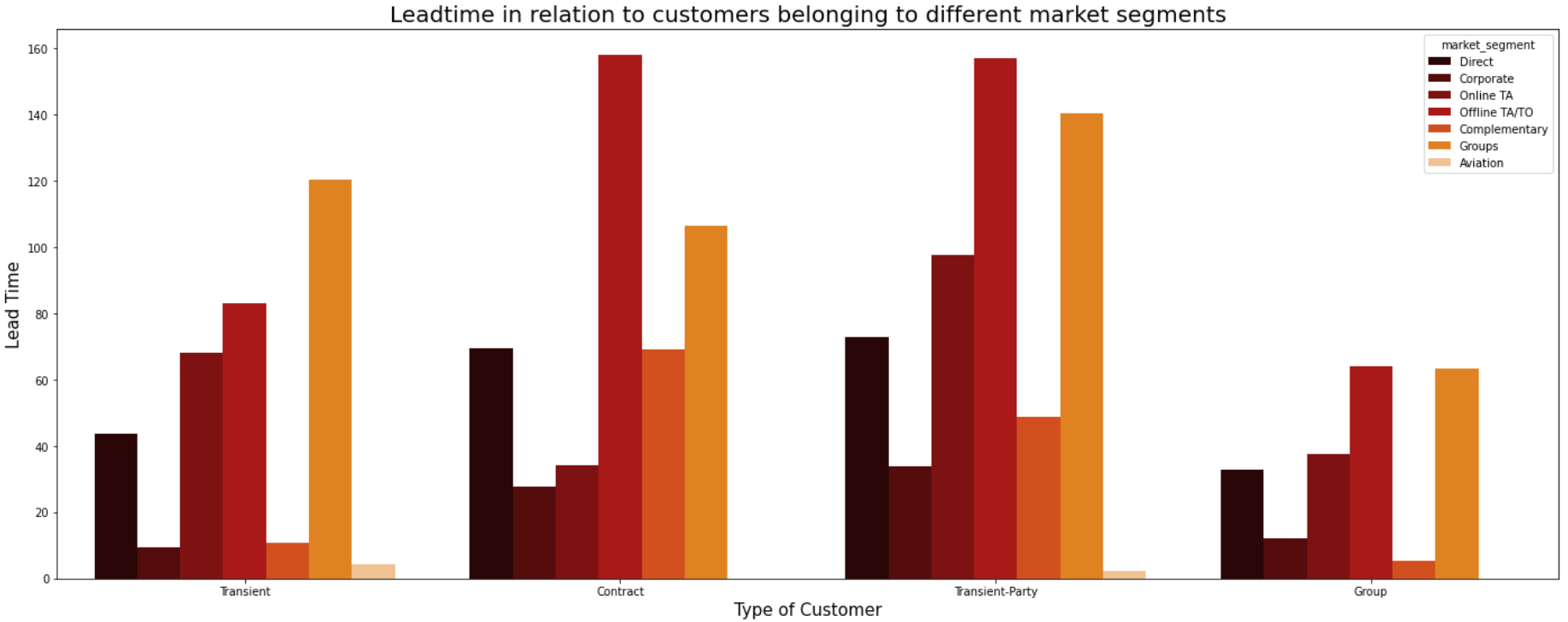
# Which meal is mostly preferred by people belonging to different market segments?



# Meal preferred during stay by families with different family size

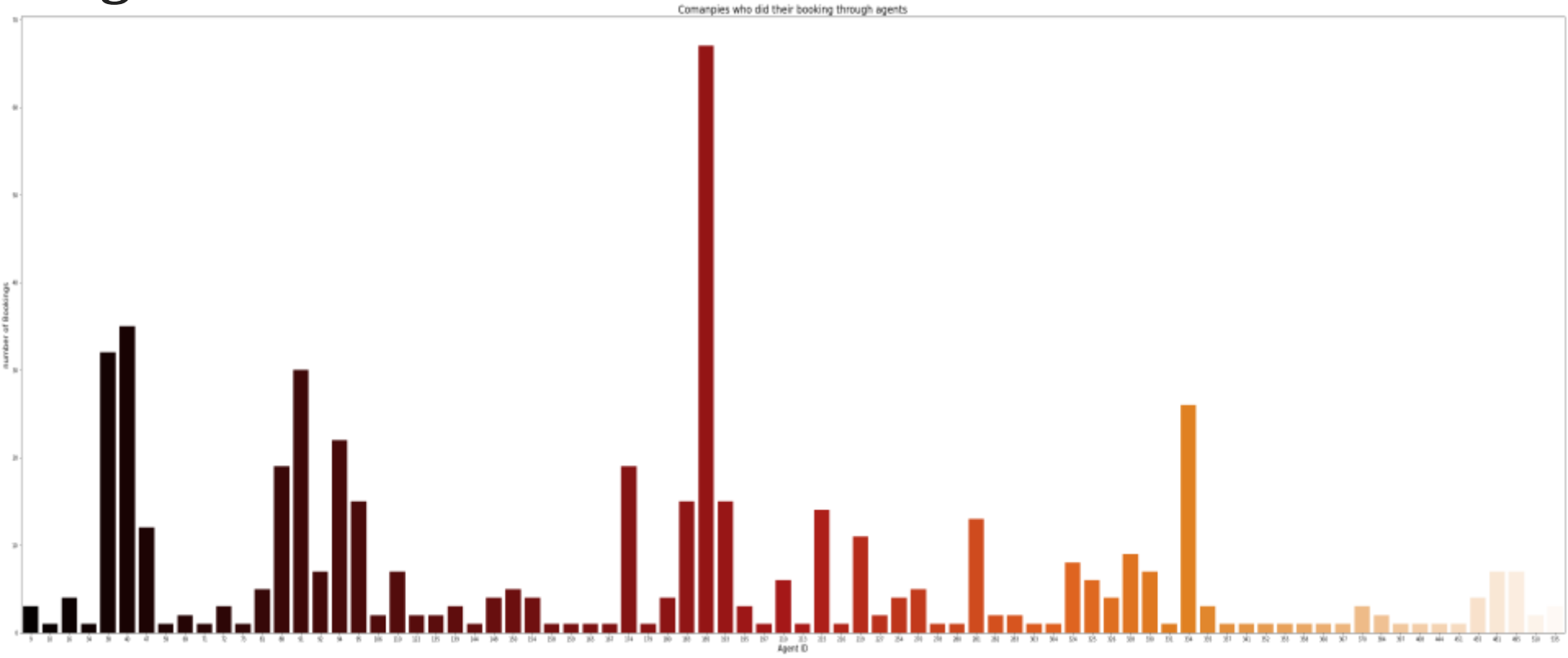


# Lead time for each type of customer and the market segment they belong to

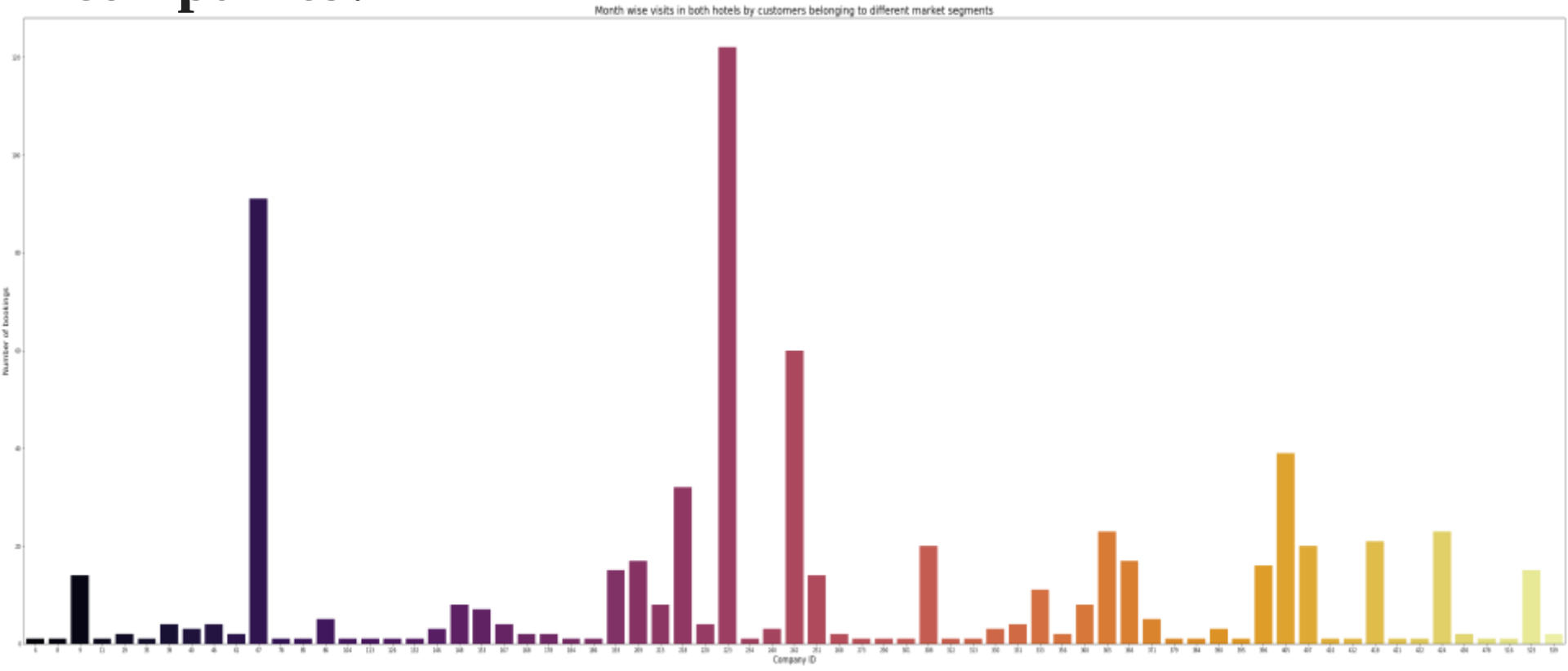


# Agents with highest bookings

# How many companies did their booking through agents?



# How many agents did their booking through companies?



# Conclusion

- We can conclude that, In the months of summer and winter, Hotels have more business than in the rest of the months of the year. Most of the bookings are made online.
- This means that for the rest of the month, hotels should give lucrative offers to customers through their online mode of interaction.
- In the years 2015 and 2017 booking traffic was not present for the complete year, despite this fact the revenue for hotels was maximum in the year 2017.
- Resort hotel is more affordable than city hotel.


Thank you