

Capstone Project
‘Hotel Bookings Analysis’

(Team Mrityunjay)

-Presented 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 has following columns:

- hotel,
 - is_canceled,
 - lead_time,
- arrival_date_year,
arrival_date_month,

arrival_date_week_number,
arrival_date_day_of_month,
,stays_in_weekend_nights,
stays_in_week_nights,
adults,
- is_repeated_guest,
previous_cancellations,
previous_bookings_not_cancelled,
reserved_room_type,
children,

babies,
meal,
country,
market_segment,
distribution_channel,
assigned_room_type,
- booking_changes,
deposit_type,
agent,
company,
days_in_waiting_list
customer_type,

adr,
required_car_parking_spaces,
total_of_special_requests,
reservation_status,
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.
- From the Dataset itself, we created 2 more columns:
 1. Total_Stay = Sum of Columns Weekend night stay and Weeknight stay.
 2. Total_Members = Sum of Columns Adults, Children.

```
| | #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
| | hotel$fi.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
hotelfdf1.fillna({'children':0},inplace=True)

[ ] #replacing null countries with 'unknown'
hotelfdf1.fillna({'country':'unknown'},inplace=True)
```

```

booking again for null values
hotelIdIsNull().name()

hotel                0
is_cancelled         0
load_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
meals                0
country              0
market_segment       0
distribution_channel  0
is_repeated_guest    0
previous_cancellations  0
previous_bookings_not_cancelled  0
cancellation_type     0
assigned_room_type    0
booking_changes       0
deposit_type          0
agent                100000
company              -1123400
days_in_waiting_list  0
customer_type         0
ad                    0
required_car_parking_spaces  0
total_of_special_requests  0
recreation_status     0

```

```

1 # Cancelling has been handled, so we can now do our cancelled
2 bookings to cancelled [table_name]
3
4 # ...
5 # ...
6 done: if cancelled, status: failed
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

```
[1]: #viewing the values in adults column and its values
hotel_df1['adults'].value_counts()

2      89680
1      23027
3       6302
0        403
4         62
20         5
27         3
20         2
5          2
40         1
50         1
55         1
6          1
10         1
Name: adults, dtype: int64
```

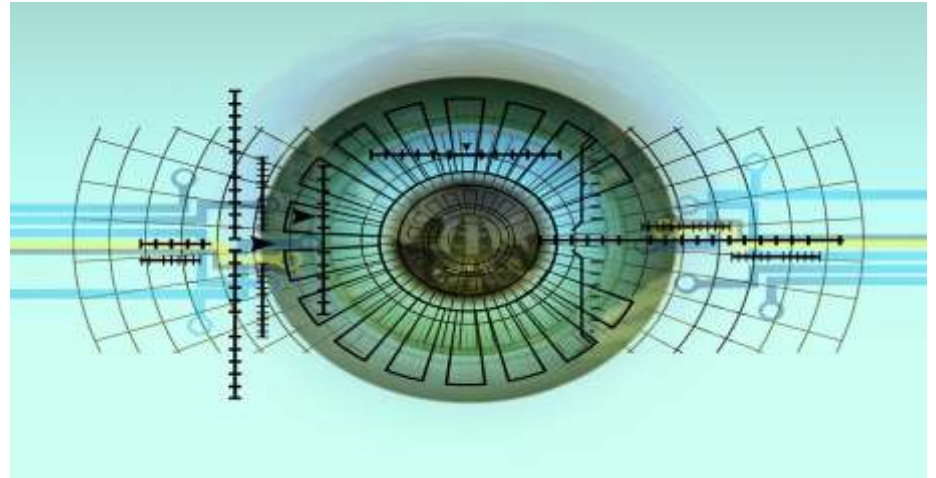
```

1 |> library("dplyr")
2 |> library("tidyr")
3 |> library("ggplot2")
4 |> library("rstatix")
5 |> library("readr")
6 |> library("readxl")
7 |> library("readr")
8 |> library("readr")
9 |> library("readr")
10 |> library("readr")
11 |> library("readr")
12 |> library("readr")
13 |> library("readr")
14 |> library("readr")
15 |> library("readr")
16 |> library("readr")
17 |> library("readr")
18 |> library("readr")
19 |> library("readr")
20 |> library("readr")
21 |> library("readr")
22 |> library("readr")
23 |> library("readr")
24 |> library("readr")
25 |> library("readr")
26 |> library("readr")
27 |> library("readr")
28 |> library("readr")
29 |> library("readr")
30 |> library("readr")
31 |> library("readr")
32 |> library("readr")
33 |> library("readr")
34 |> library("readr")
35 |> library("readr")
36 |> library("readr")
37 |> library("readr")
38 |> library("readr")
39 |> library("readr")
40 |> library("readr")
41 |> library("readr")
42 |> library("readr")
43 |> library("readr")
44 |> library("readr")
45 |> library("readr")
46 |> library("readr")
47 |> library("readr")
48 |> library("readr")
49 |> library("readr")
50 |> library("readr")
51 |> library("readr")
52 |> library("readr")
53 |> library("readr")
54 |> library("readr")
55 |> library("readr")
56 |> library("readr")
57 |> library("readr")
58 |> library("readr")
59 |> library("readr")
60 |> library("readr")
61 |> library("readr")
62 |> library("readr")
63 |> library("readr")
64 |> library("readr")
65 |> library("readr")
66 |> library("readr")
67 |> library("readr")
68 |> library("readr")
69 |> library("readr")
70 |> library("readr")
71 |> library("readr")
72 |> library("readr")
73 |> library("readr")
74 |> library("readr")
75 |> library("readr")
76 |> library("readr")
77 |> library("readr")
78 |> library("readr")
79 |> library("readr")
80 |> library("readr")
81 |> library("readr")
82 |> library("readr")
83 |> library("readr")
84 |> library("readr")
85 |> library("readr")
86 |> library("readr")
87 |> library("readr")
88 |> library("readr")
89 |> library("readr")
90 |> library("readr")
91 |> library("readr")
92 |> library("readr")
93 |> library("readr")
94 |> library("readr")
95 |> library("readr")
96 |> library("readr")
97 |> library("readr")
98 |> library("readr")
99 |> library("readr")
100 |> library("readr")

```

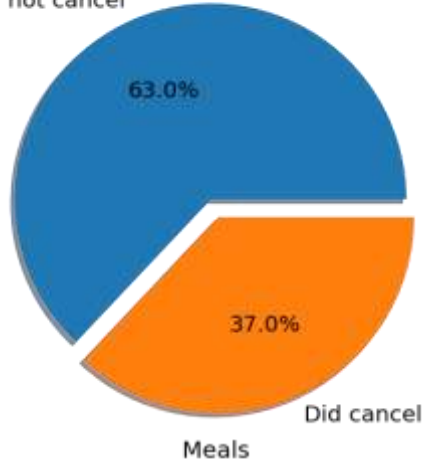
```
[ ] #creating a new column to store total members
    hotelddf1['total_members']=hotelddf1['adults'] + hotelddf1['children']
```

❖ Exploratory Data Analysis:



Who all cancelled their bookings

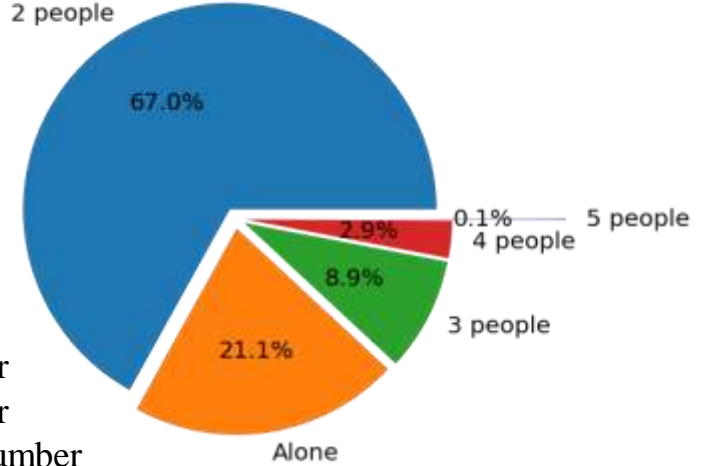
Did not cancel



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

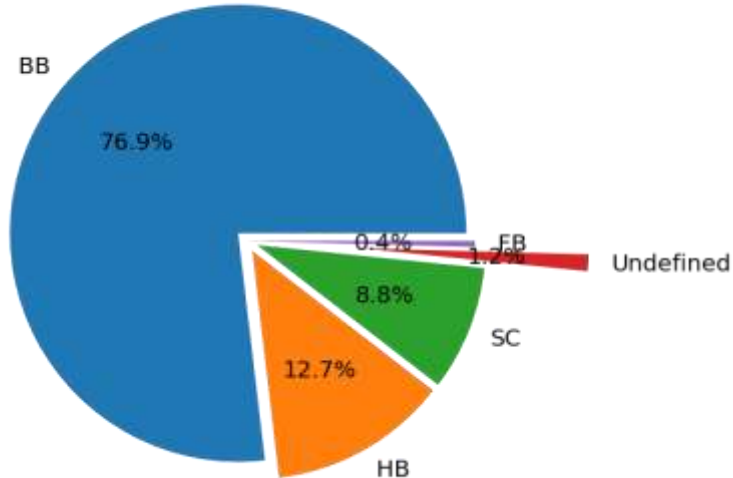
Frequencies of number of members

2 people



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

BB



76.9% people opted for bed and breakfast
12.7% people opted for 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 full board i.e., Breakfast, Lunch and Dinner

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

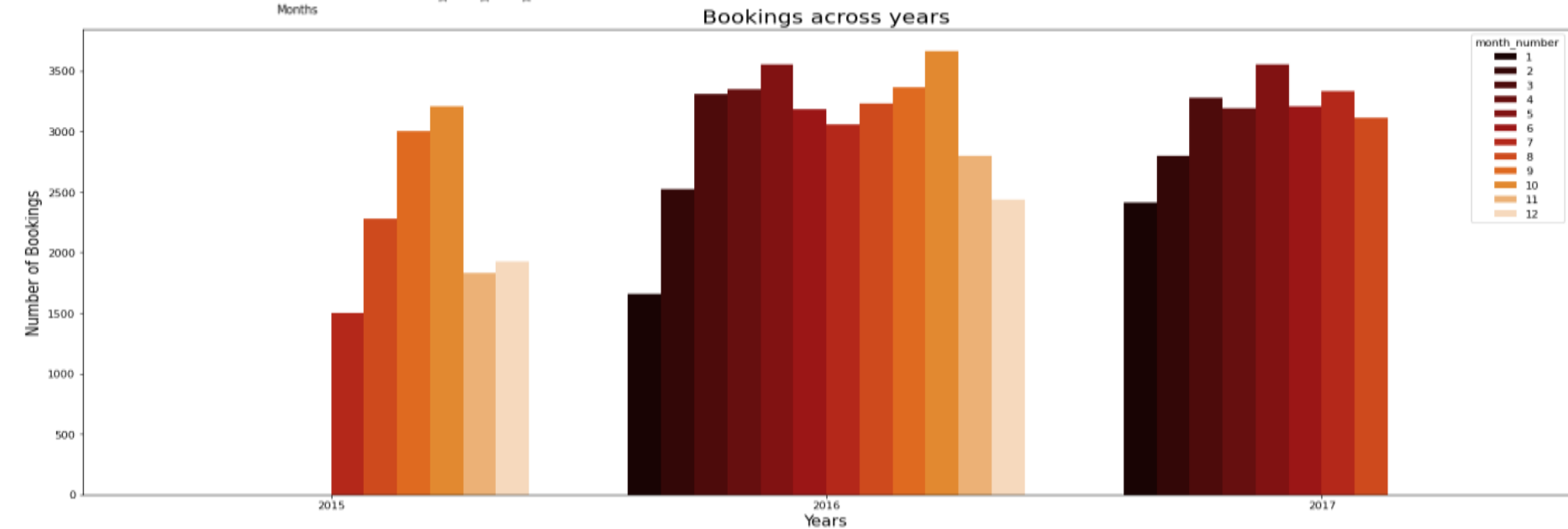
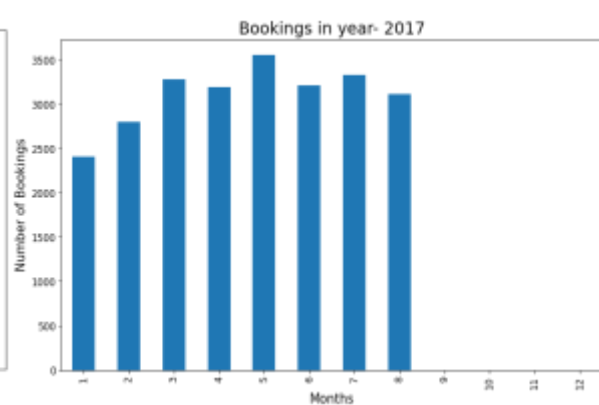
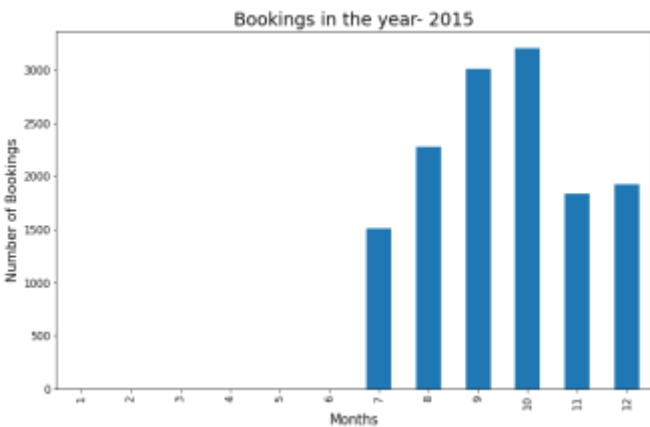
This question has two aspects-

#From customer's point of view-

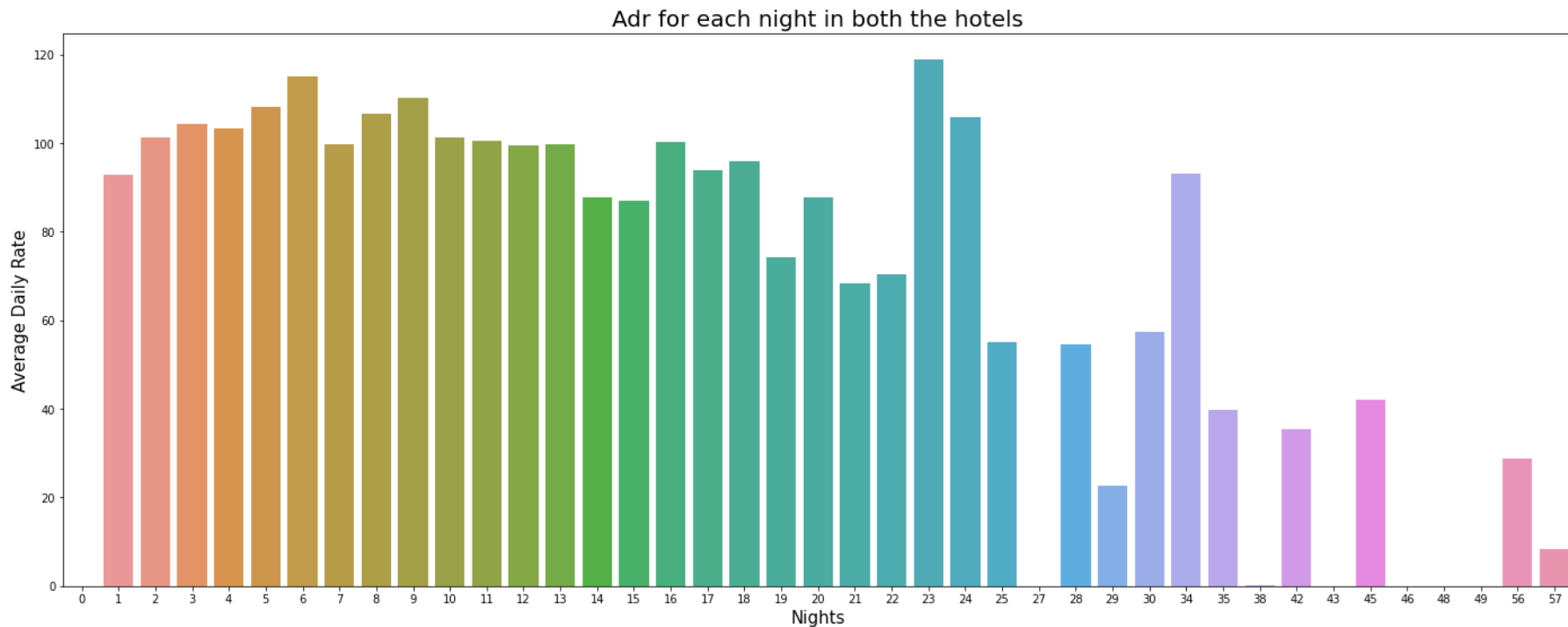
From customers point of view the best time to book a hotel room is when it is the cheapest i.e. during off seasons, when not too many people book hotel rooms.

#From hotel's point of view-

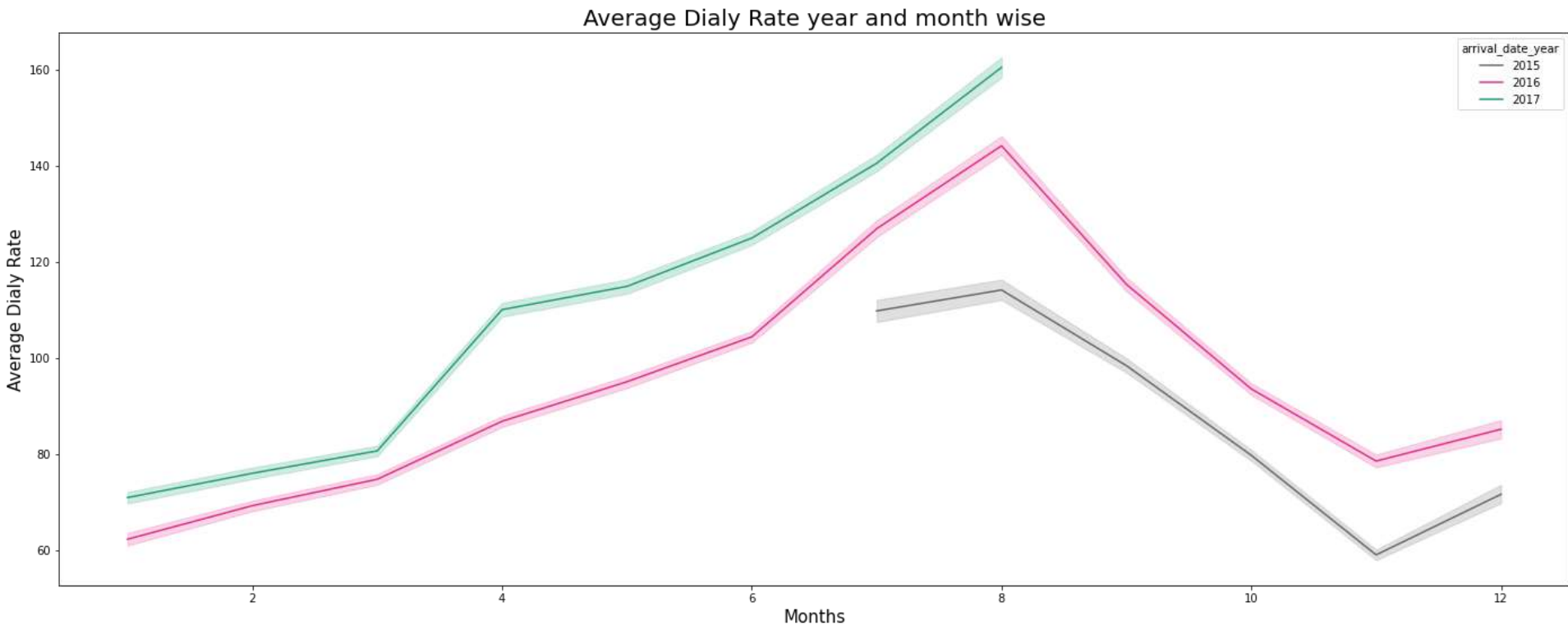
From hotel's point of view the best time to book the hotel is during onn seasons i.e when too many people are booking rooms and the demand for rooms is high, at that time the hotel can charge high prices from customers, because the supply is same but the demand is relatively high.



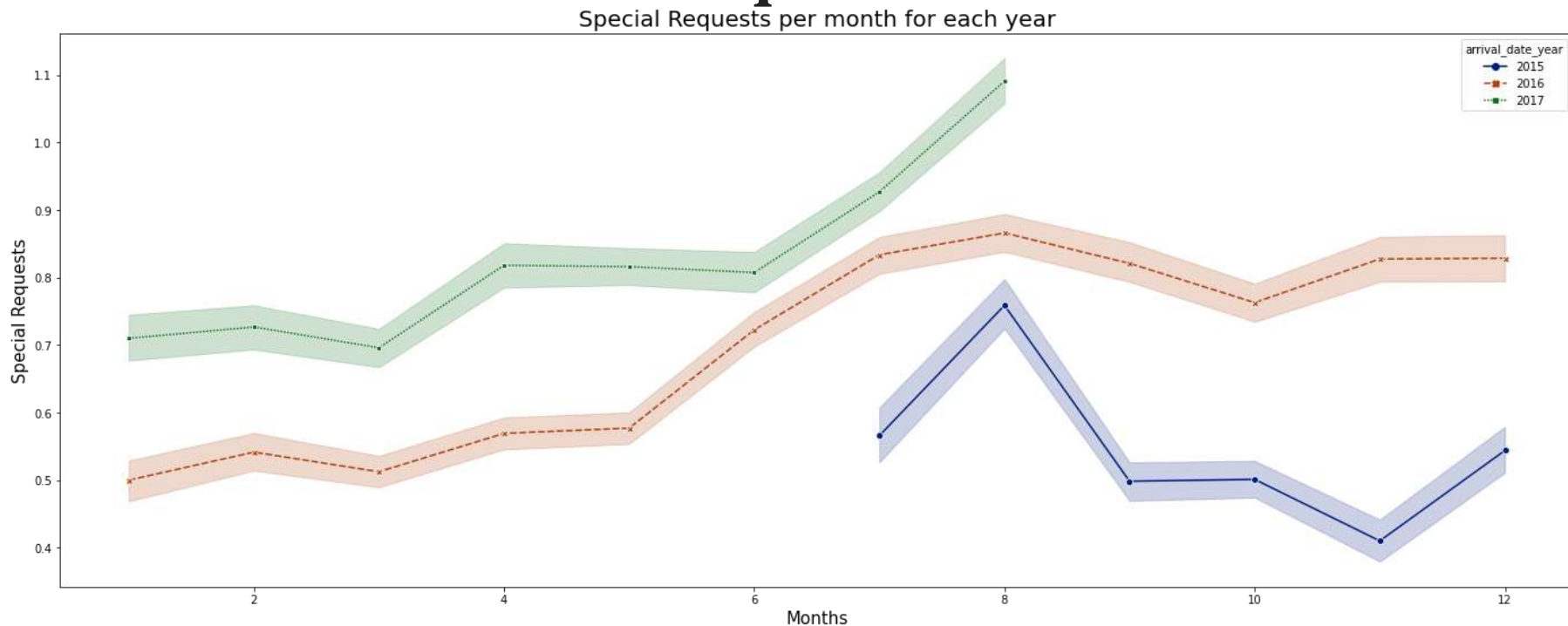
What is the optimal length of stay in order to get the best daily rate?



Average Daily Rate year and month wise

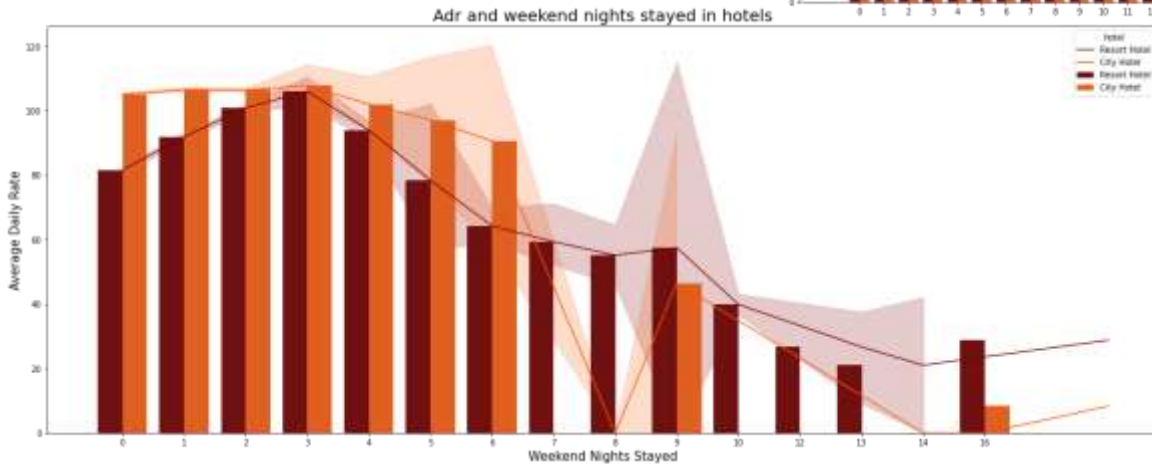
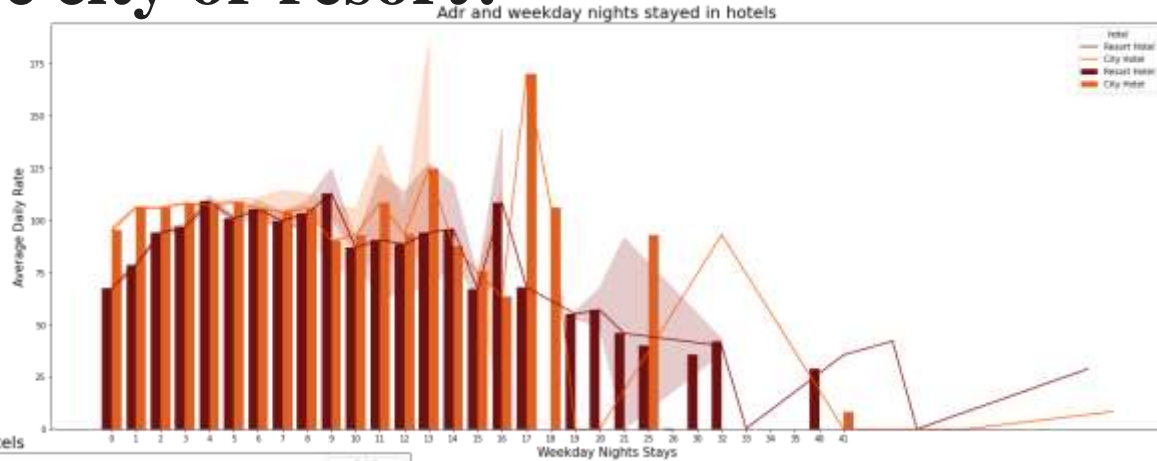


- **How to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?**



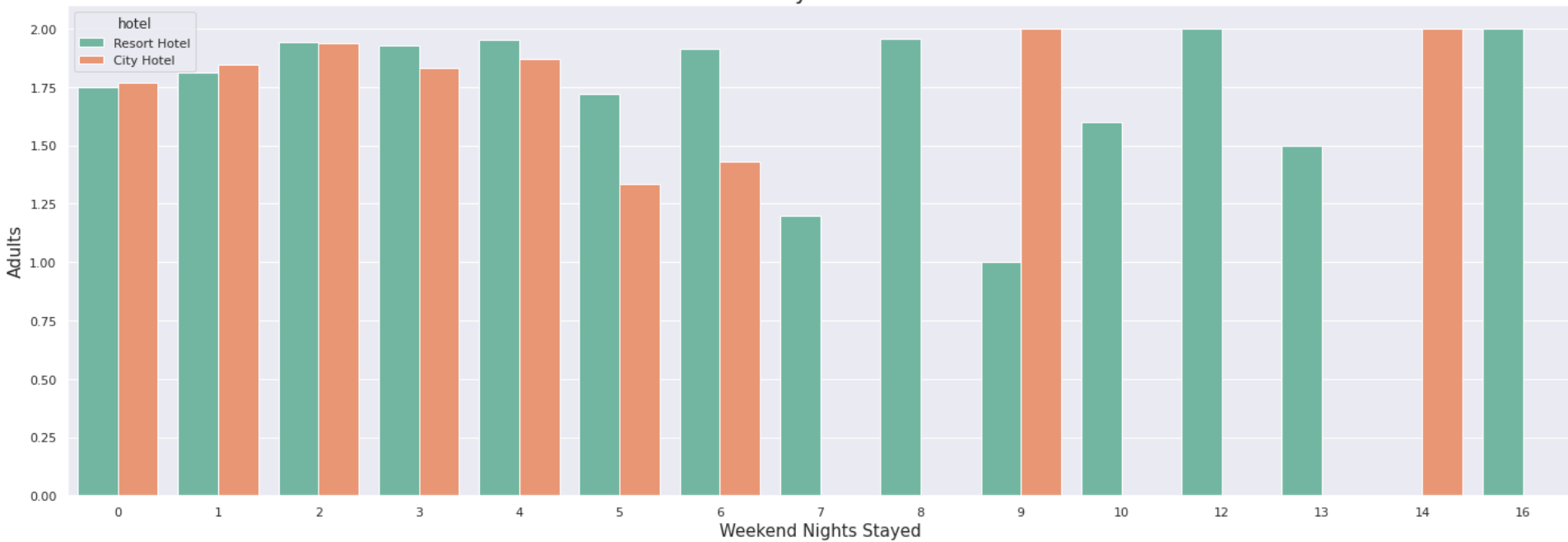
Which hotel to choose city or resort?

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

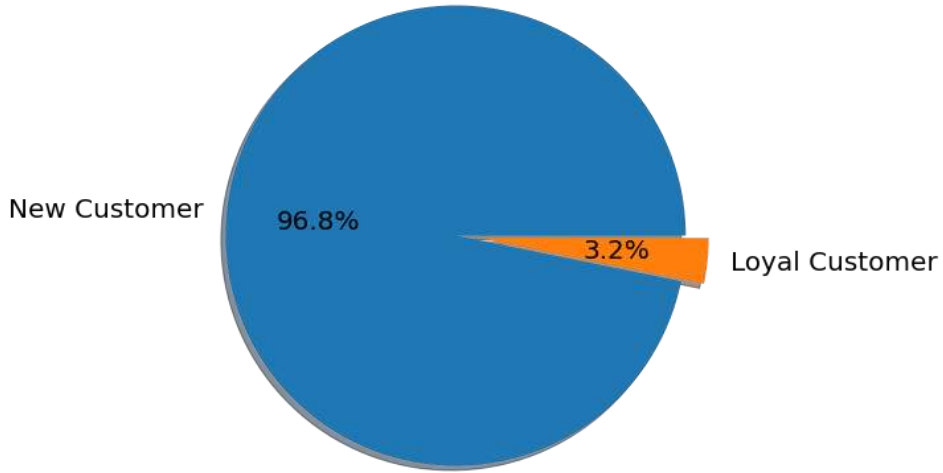


City Hotel's per day price for weekend is more than that of Resort hotel, but there is not much difference. Which was not the case for weekday's per day price.

Adults and their stay in hotels on weekends



How many guests are repeated guests

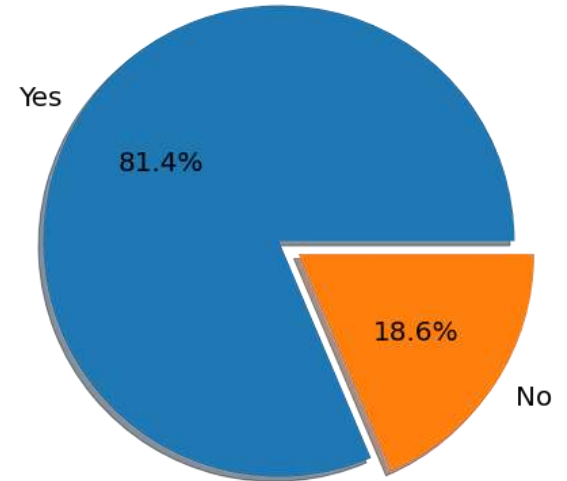


3.2% customers are repeated guests

Rest

96.8% customers are new

Where people assigned the room they wanted

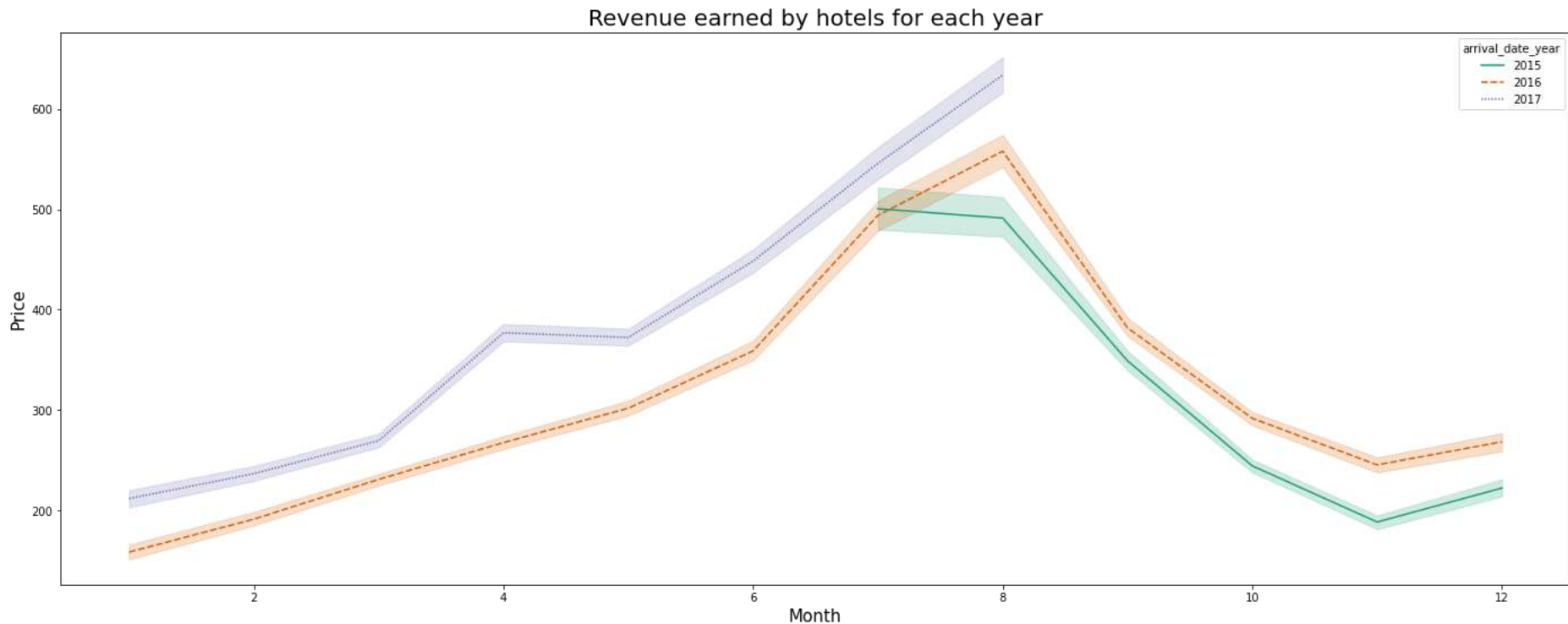


81.4% customers were allotted the room they reserved

Rest

18.6% did not receive the room they wanted

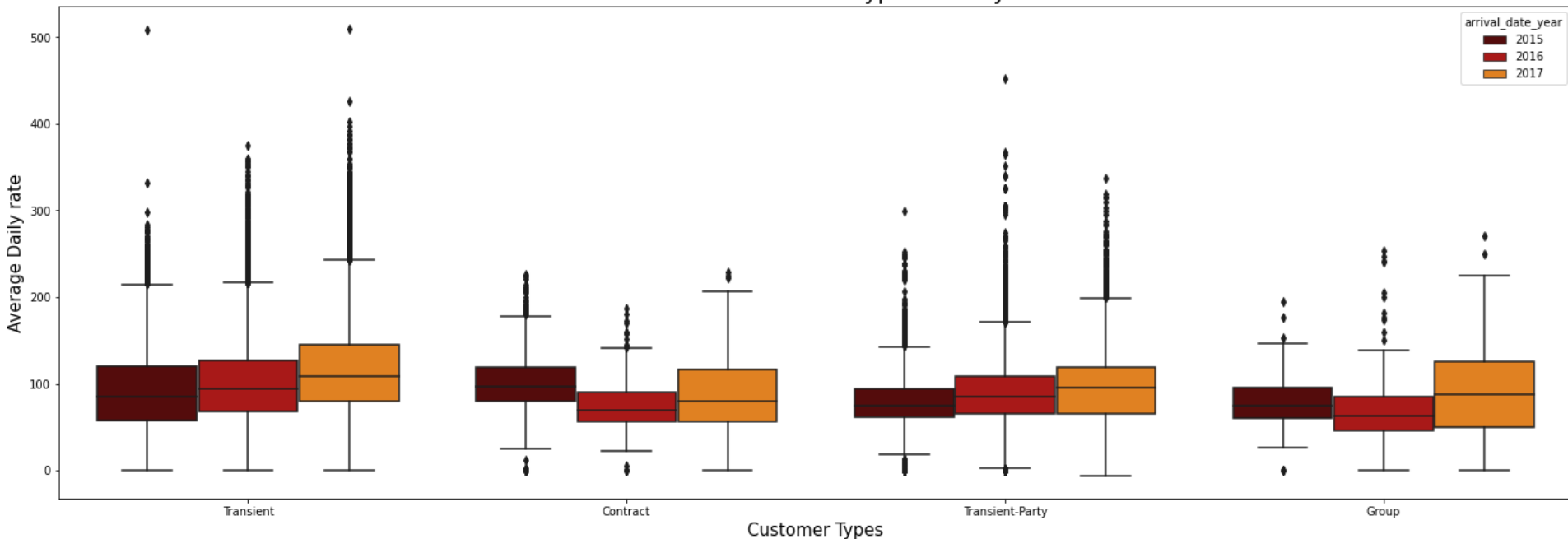
Revenue earned by hotels for each year



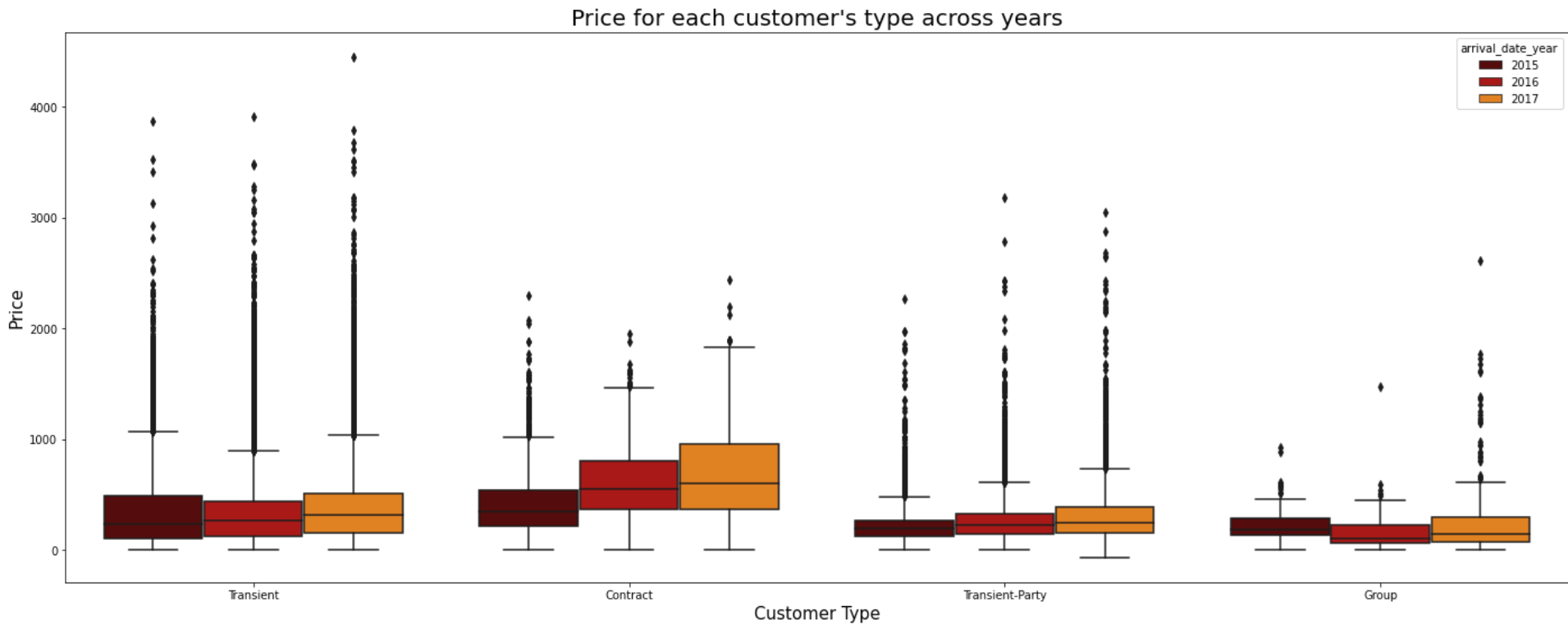
From which customer's type did hotels earned the most revenue?

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

Adr for each customer's type across years

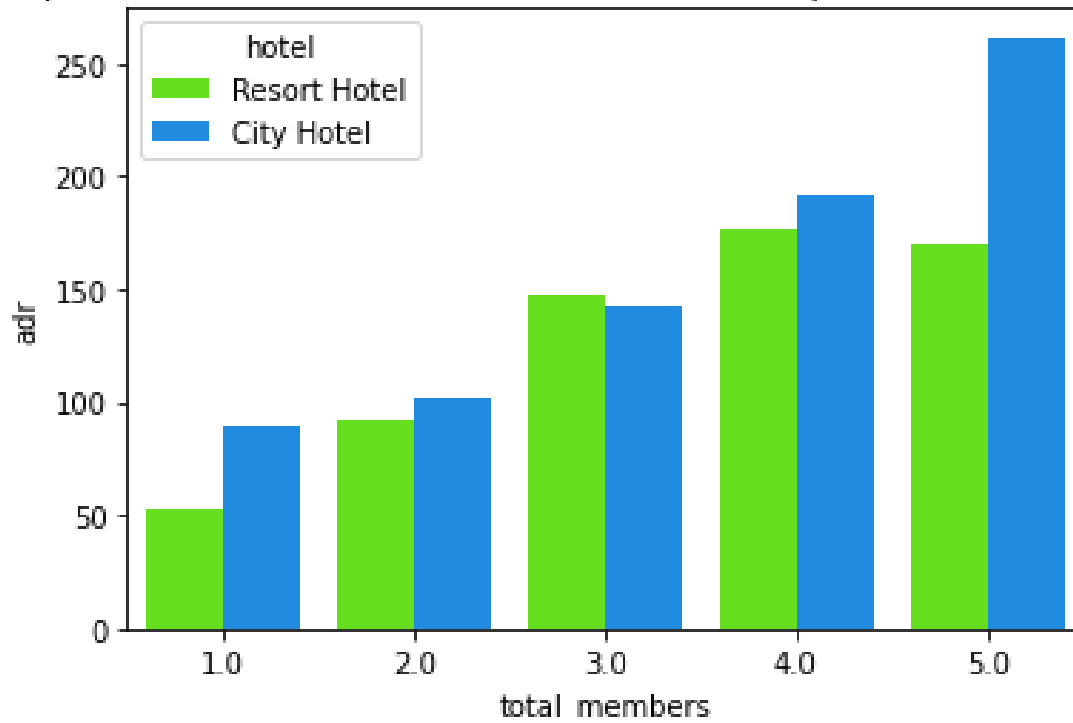


- ii) Price and Customer type

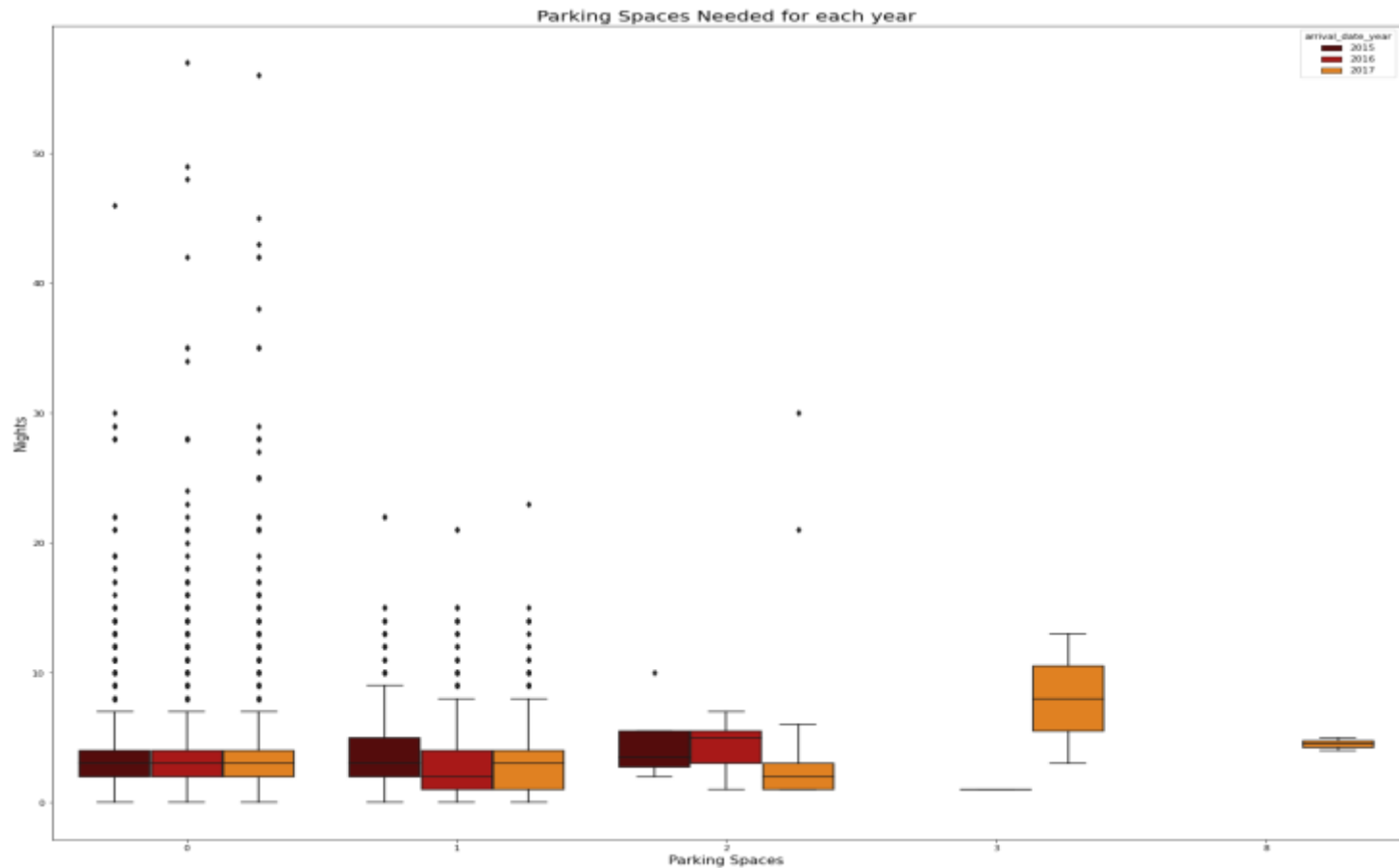


Which Group of Family Members Generates a Large Amount of Revenue for Hotels?

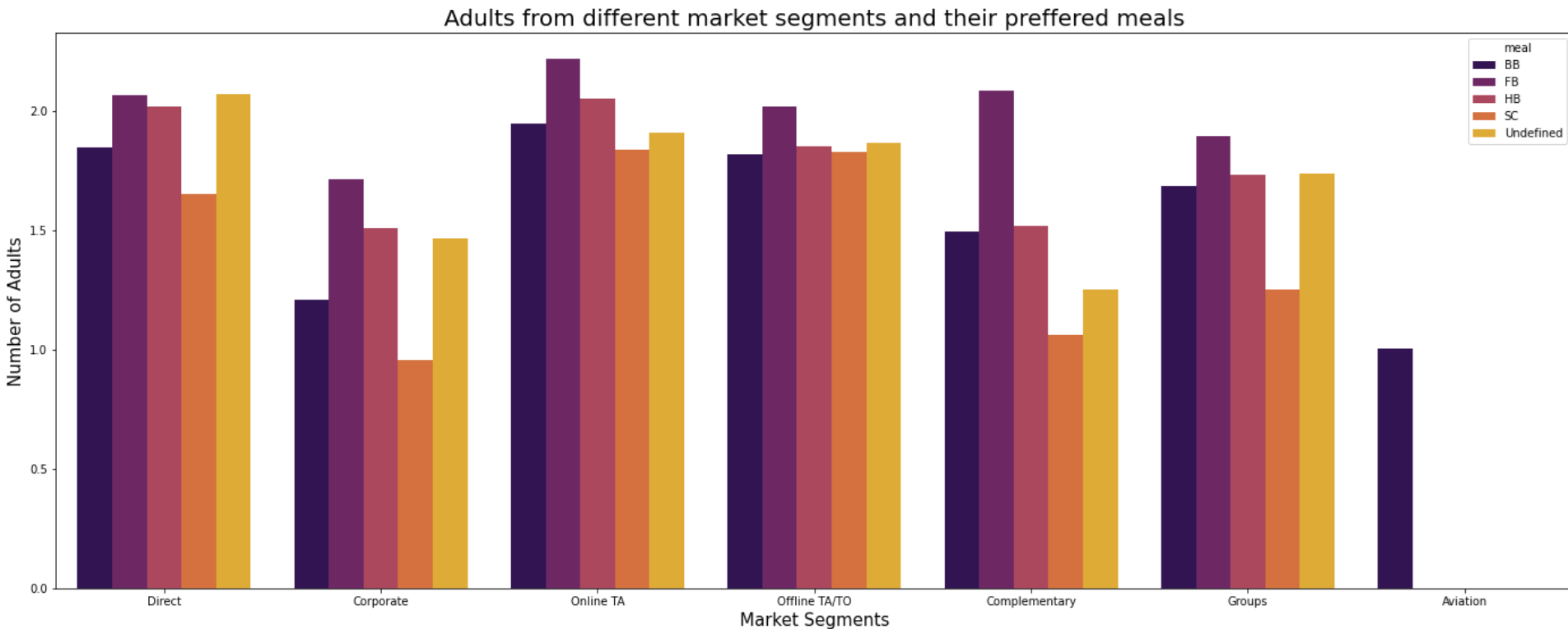
Per Capita Revenue Generated from Different Groups of Total Family Members



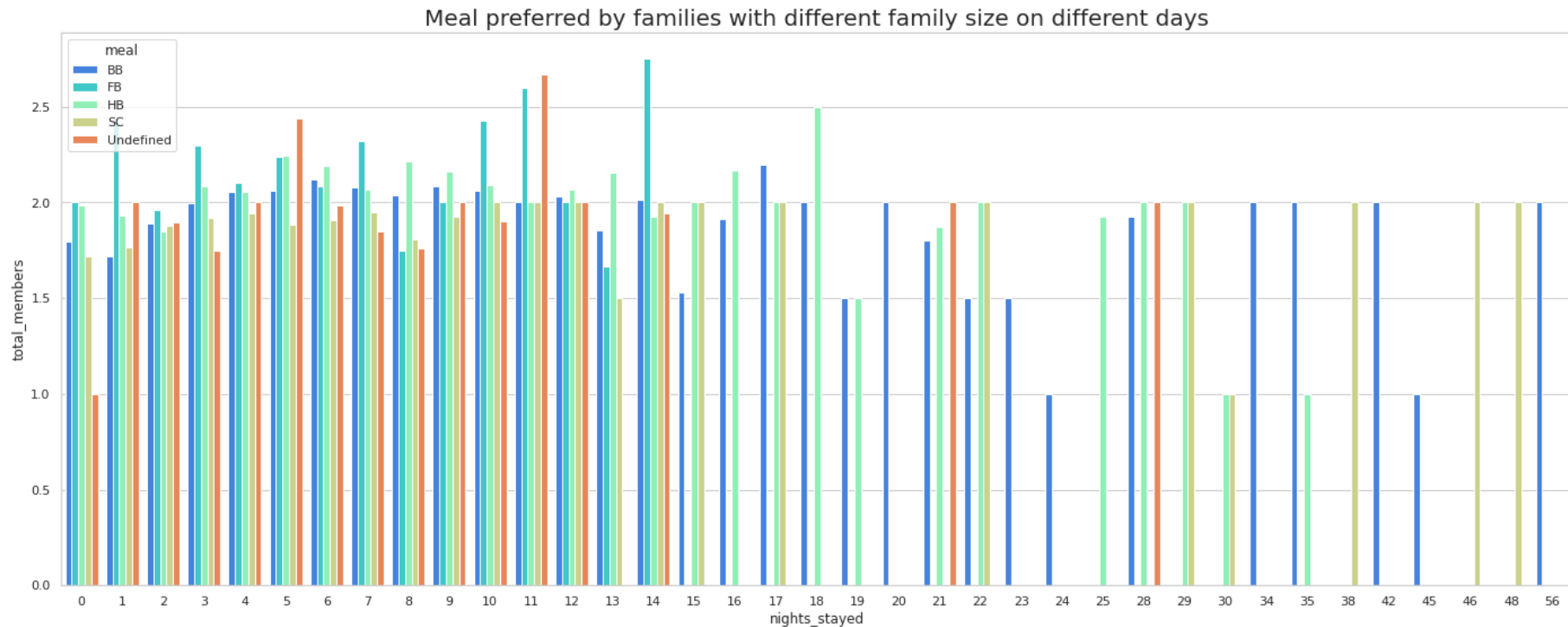
Car parking Spaces required in hotels



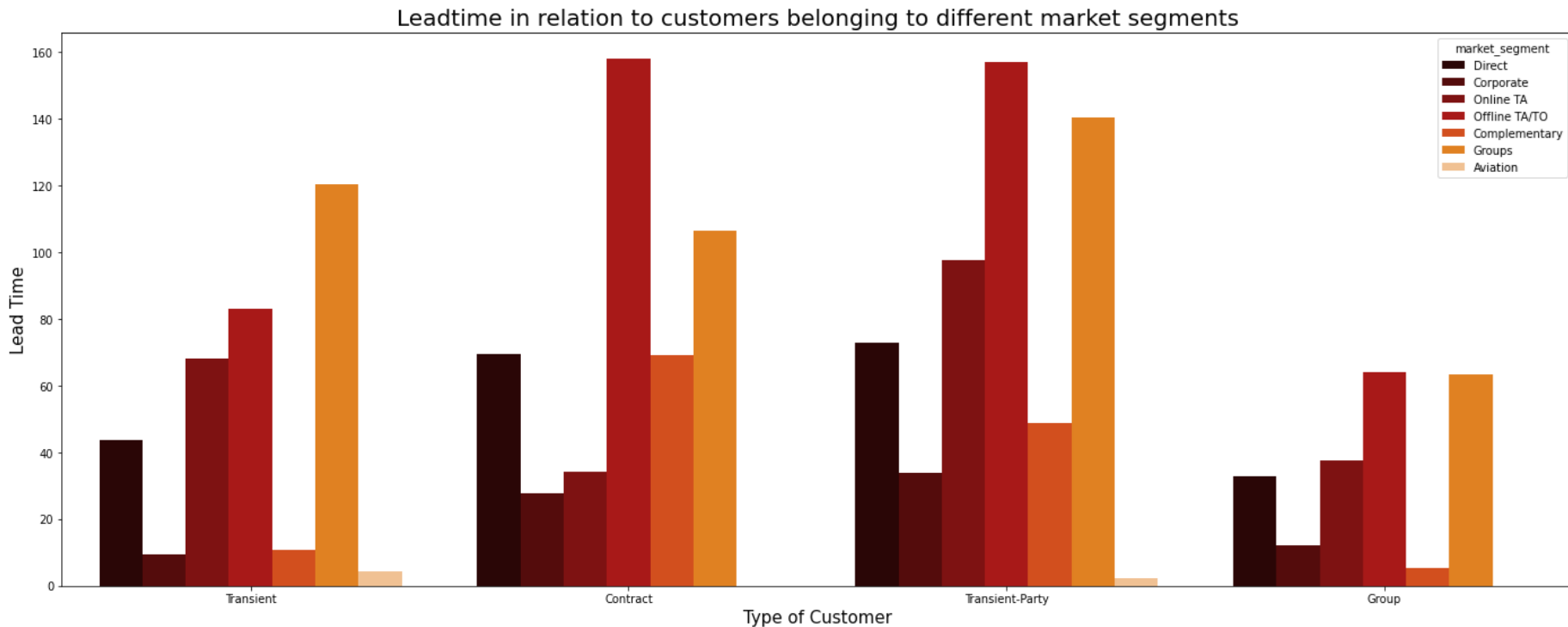
Which meal is mostly preferred by people belonging to different market segments?



Meal preferred by families with different family size on different days during their stay



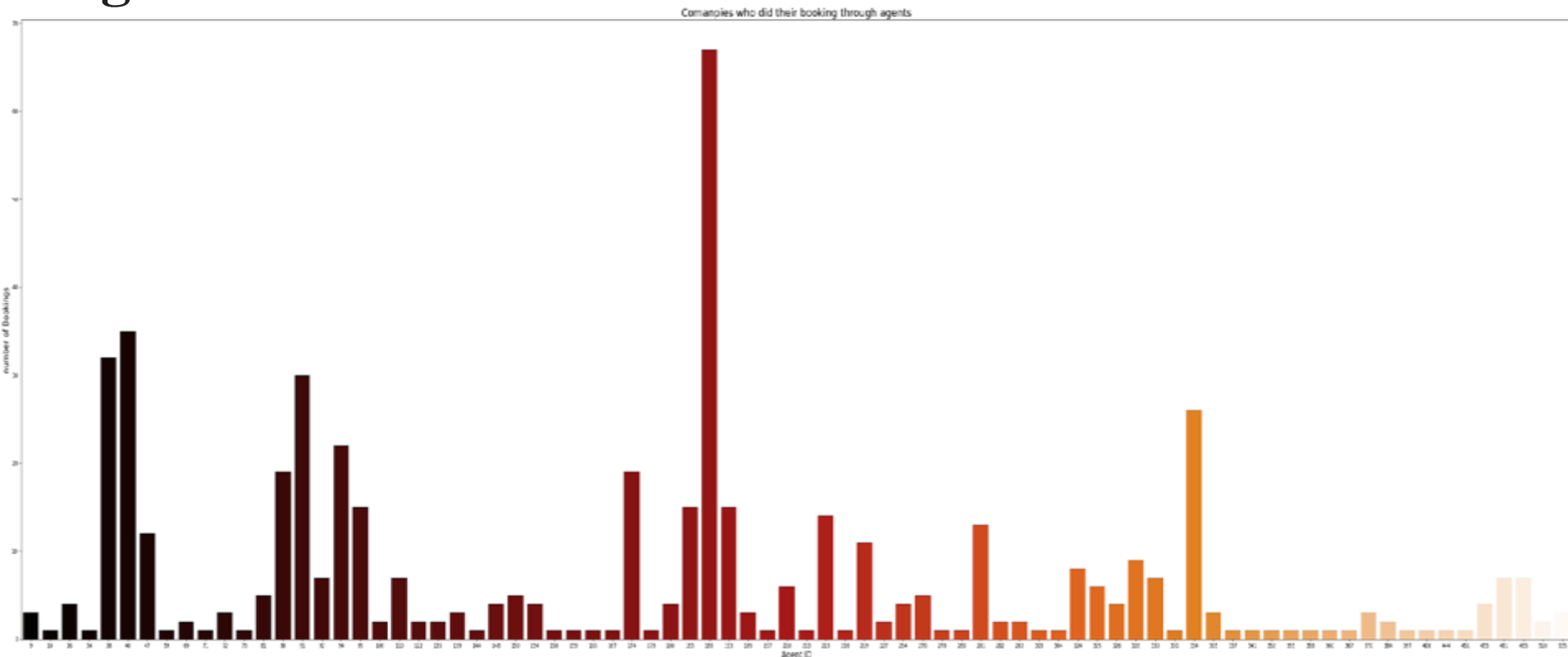
Lead time for each type of customer and the market segment they belong to



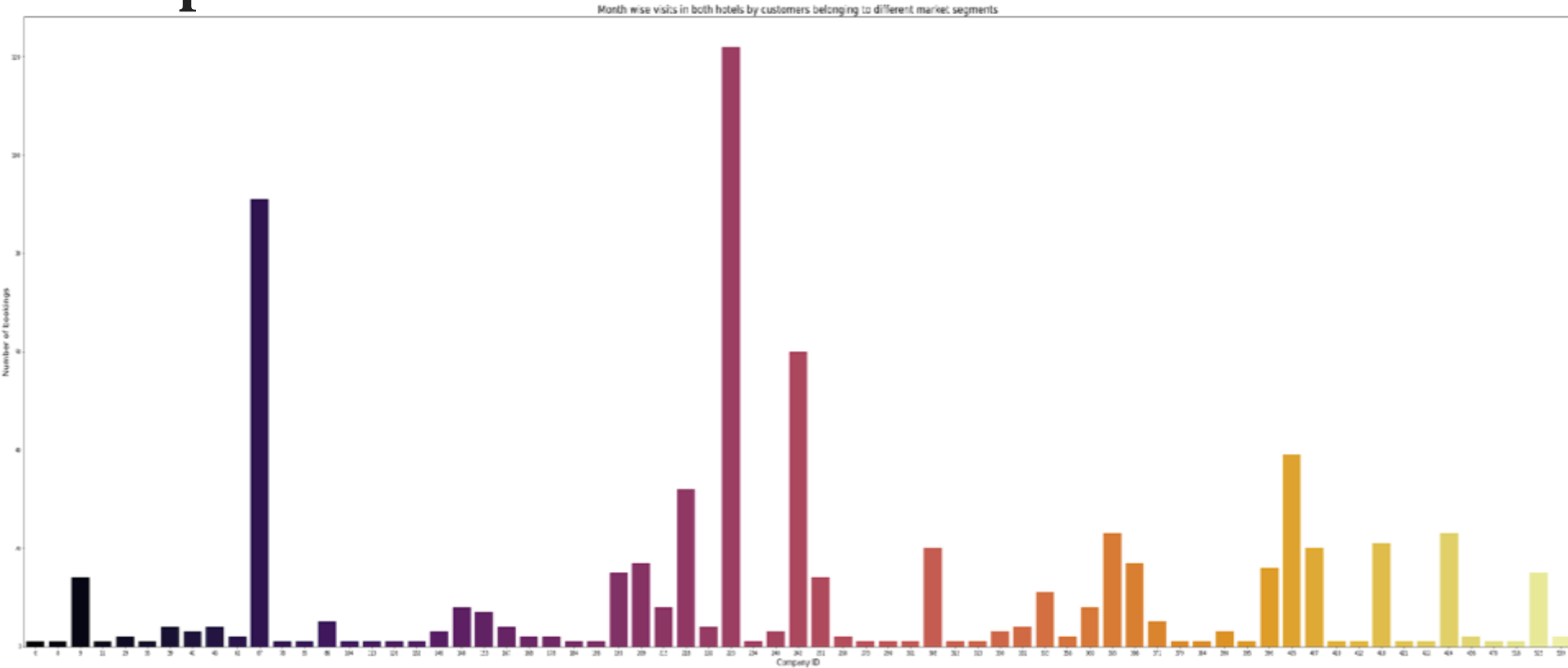
Agents with highest bookings



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.

