

# Capstone Project

## Hotel Booking Analysis

By – Shikhar Satpute

# Point for Discussion

- Introduction
- Data Summary
- Overview of Hotel type
- Yearly , Monthly and Date
- Reserved, Assigned (rooms)
- Booking changes
- No of adults, children and babies
- Types of Guest
- Meals booked
- Market segment
- Distribution channel
- Repeated guest
- Days in waiting
- Busiest time
- Reservation status
- Cancelation
- Deposit type
- Special Request
- Week-end and week days
- Average Price
- Country – wise
- Whole data comparison

# Data Summary

- hotel – 1 : resort hotel ; h2: city hotel
- is\_canceled – 0 : not - cancelled ; 1 : cancelled
- lead\_time – the time taken between when a customer makes a reservation and their actual arrival.
- arrival\_date\_year – year of arrival ( 2015 -17 )
- arrival\_date\_month – month of arrival ( jan – dec )
- Arrival\_date\_week\_number – week of arrival ( 1 – 53 )
- Arrival\_date\_day\_of\_month – day of the arrival
- stays\_in\_weekend\_nights - No of weekend nights (Sat/Sun) the guest stayed or booked to stay at the hotel
- stays\_in\_week\_nights - no of week nights (Mon – Fri) the guest stayed or booked to stay at the hotel
- Meal – BB : Bed & Breakfast ; HB – Half board (breakfast and one other meal – usually dinner) ;  
FB – Full board (breakfast, lunch and dinner) ; Undefined/SC – no meal package.

- **Market\_segment** – (a group of people who share one or more common characteristics, lumped together for marketing purposes) TA: Travel agent ; TO : Tour operators
- **Distribution\_channel** – (A distribution channel is a chain of businesses or intermediaries through which a good or service passes until it reaches the final buyer or the end consumer) TA: Travel agent ; TO : Tour operators
- **Is\_repeated\_guest** - (value indicating if the booking name was from repeated guest) 1: Yes ; 0 : No
- **Previous\_cancellations** – No. of previous bookings that were cancelled by the customer prior to the current booking
- **previous\_bookings\_not\_canceled** – No. of previous bookings not cancelled by the customer prior to the current booking
- **reserved\_room\_type** - Code of room type reserved. Code is presented instead of designation for anonymity reasons.
- **assigned\_room\_type** - Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons.
- **booking\_changes** – No of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

- **deposit\_type** - No Deposit – no deposit was made; Non Refund – a deposit was made in the value of the total stay cost; Refundable – a deposit was made with a value under the total cost of stay.
- **Agent** - ID of the travel agency that made the booking
- **Company** - ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons
- **day\_in\_waiting\_list** – No of days the booking was in the waiting list before it was confirmed to the customer.
- **customer\_type** - Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking.
- **adr** (average daily rate) -measures the average rental revenue earned for an occupied room per day.
- **required\_car\_parking\_spaces** – No of car parking spaces required by the customer
- **total\_of\_special\_requests** - Number of special requests made by the customer (e.g. twin bed or high floor)
- **reservation\_status** - Canceled – booking was canceled by the customer; Check-Out – customer has checked in but already departed; No-Show – customer did not check-in and did inform the hotel of the reason why
- **reservation\_status\_date** - Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel

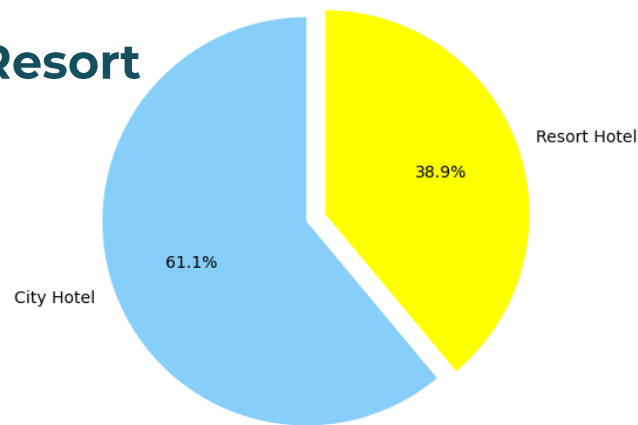
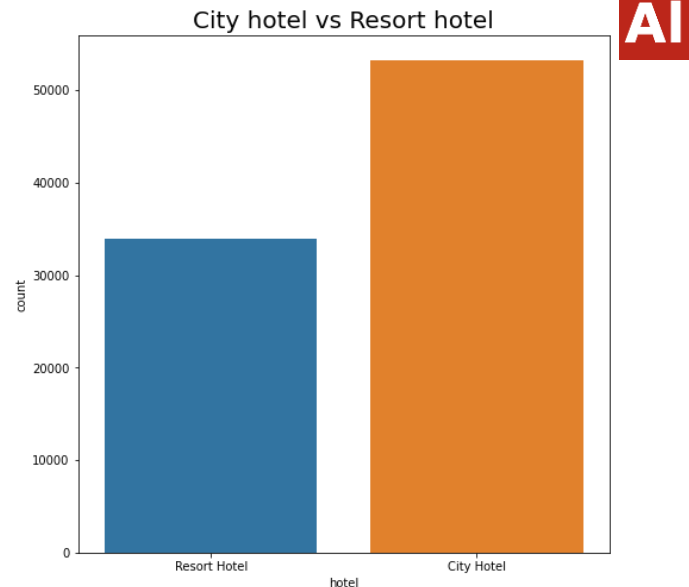
# Objective

**We will analyse hotel booking dataset and find out the following things :-**

- **The behaviour of the customer.**
- **Which season is best for Hotel Industry.**
- **Which customer is attract for maximum profit.**

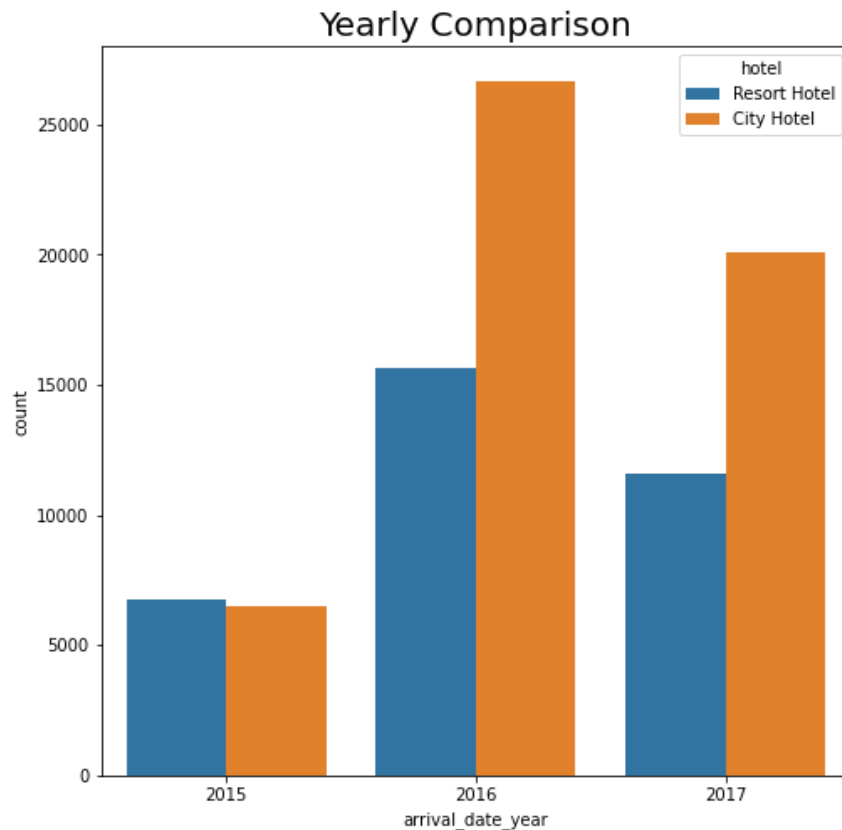
# City Hotel vs Resort Hotel

- As we can see from the graph that City Hotel has more booking than the Resort Hotel.
- People tend to book City Hotel more as Resort Hotel tend to be more expensive than.



# Yearly Comparison

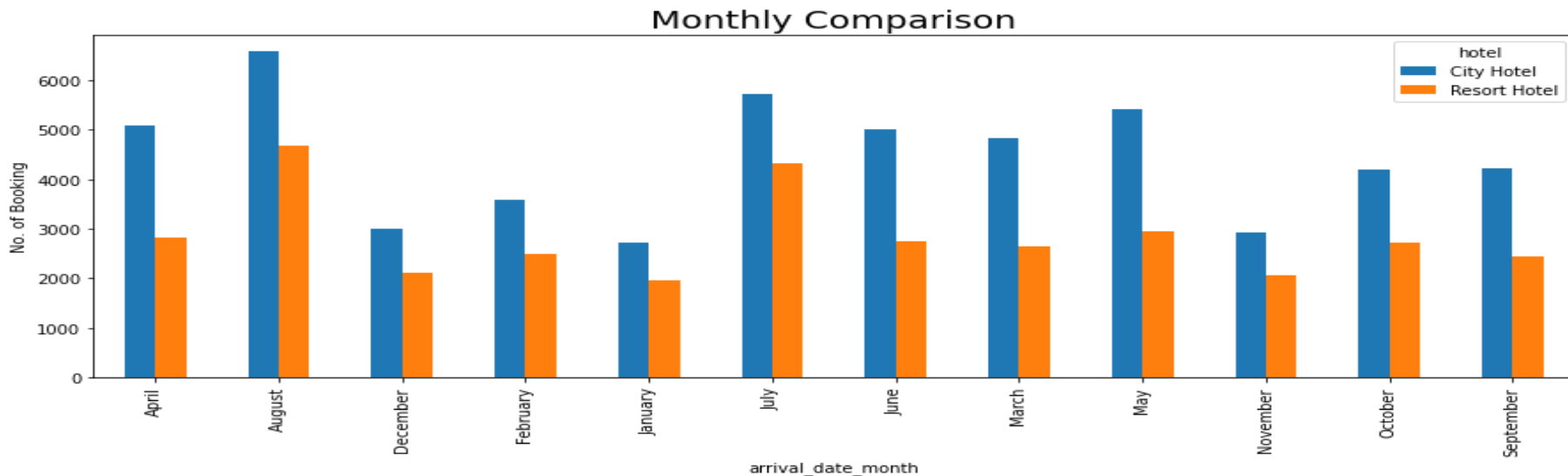
- In 2015, Hotel industry seems all time low.
- In 2016, Hotel industry seems blooming as City and Resort Hotel had a lot of booking.
- In 2017, booking count dropped.





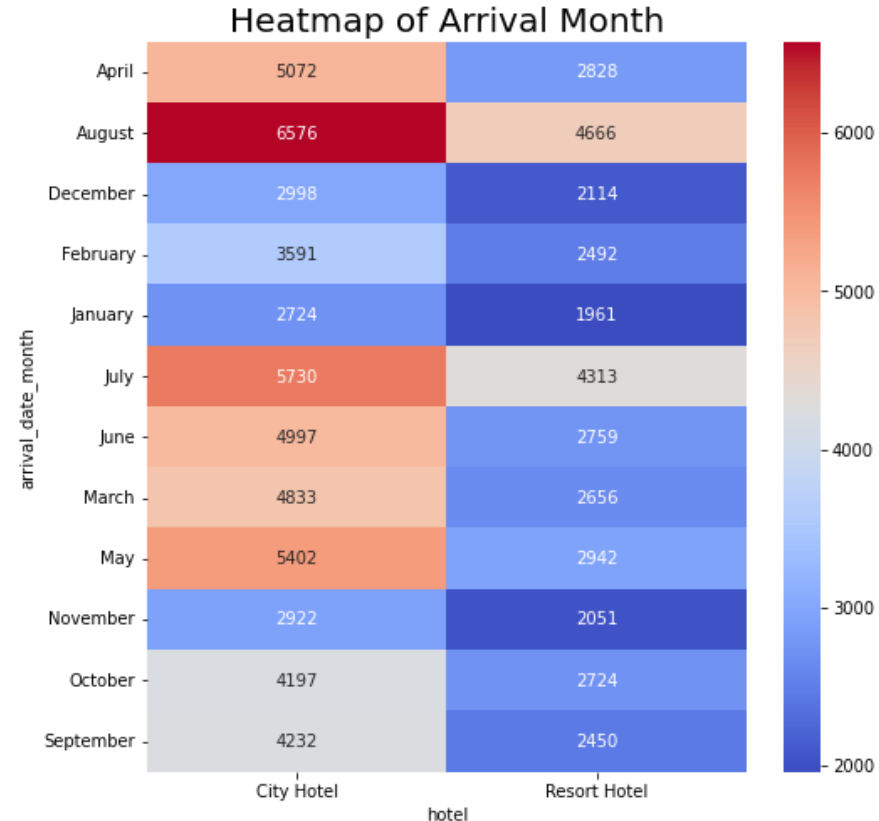
# Monthly Comparison

We can see an increasing trend in booking around the middle of the year, with August being the highest. Summer ends around August, followed straight by autumn. It seems that summer period is a peak period for hotel booking.



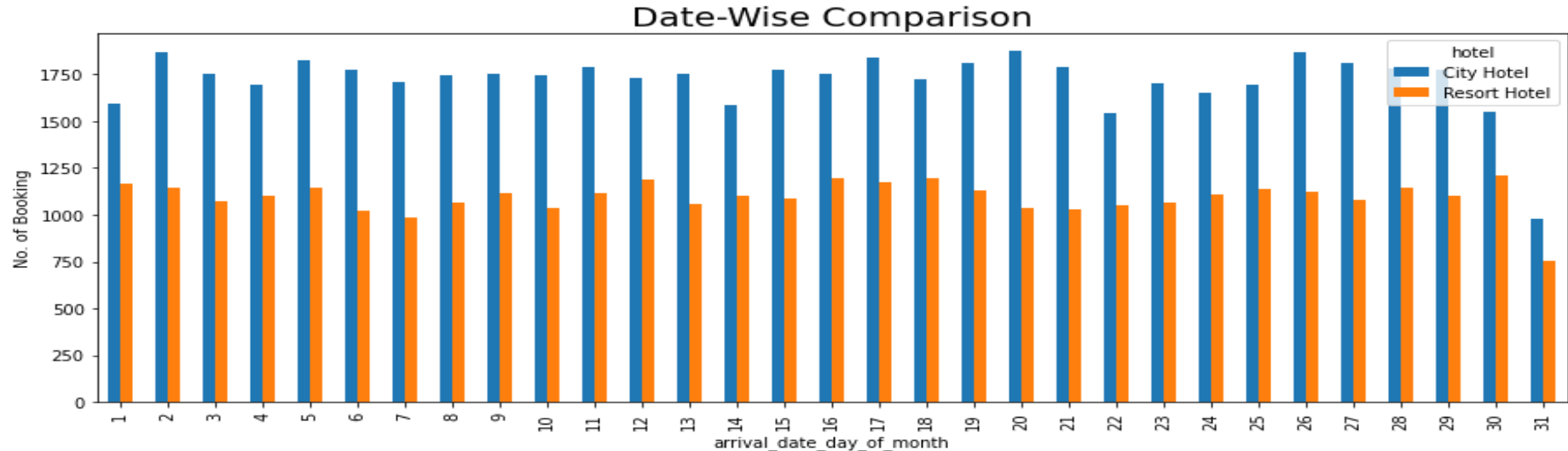
# Arrival Month

- **City Hotel :-**  
**Peak Time**  
**May – August**  
**Off – Peak Time**  
**November – January**
  - **Resort Hotel :-**  
**Peak Time**  
**July – August**  
**Off – Peak Time**  
**November – January**
- Peak Time – Summer Season**



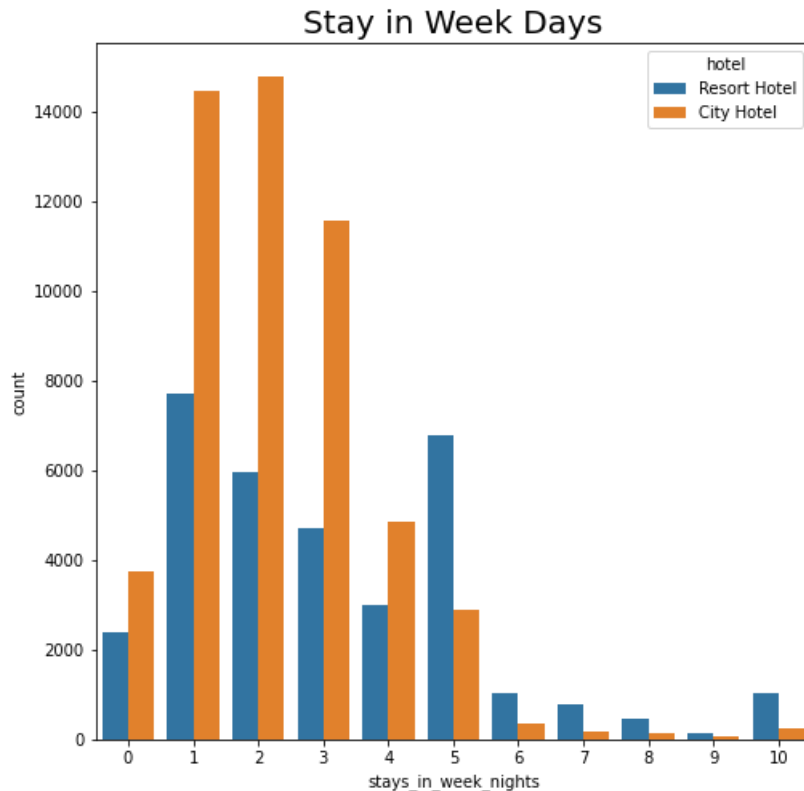
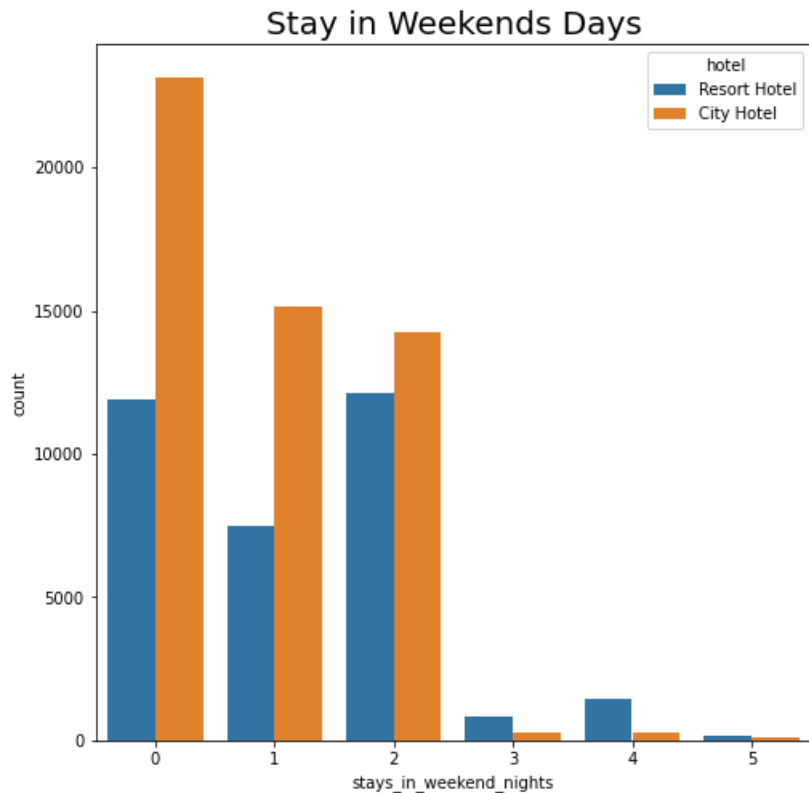
# Date – Wise Comparison

Month end seems to have low arrivals compared to rest of the dates



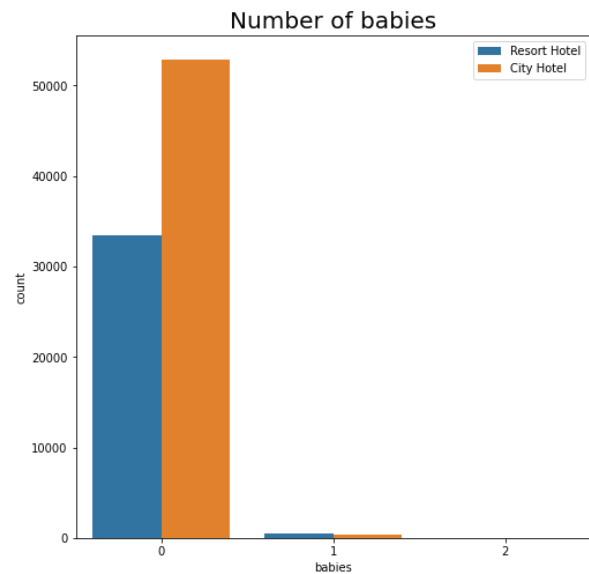
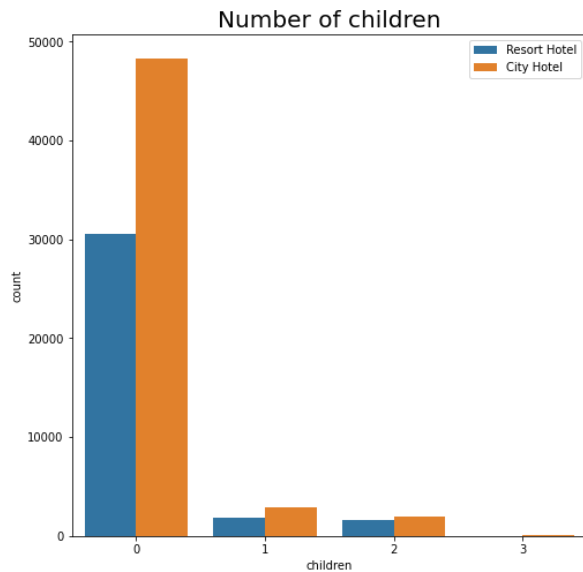
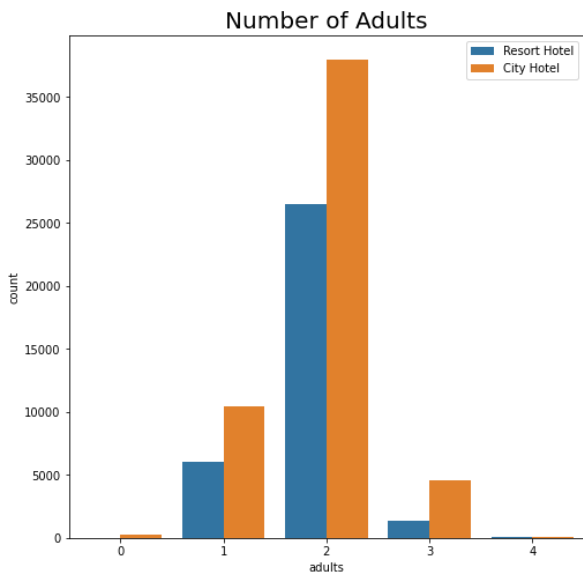
# Stay in Week – Days vs Weekends

It seems that majority of the stays are over the weekday's night.



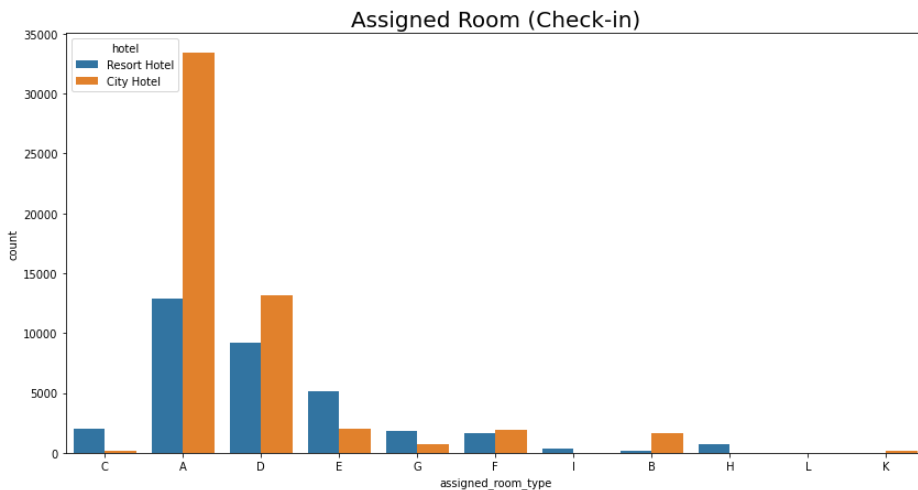
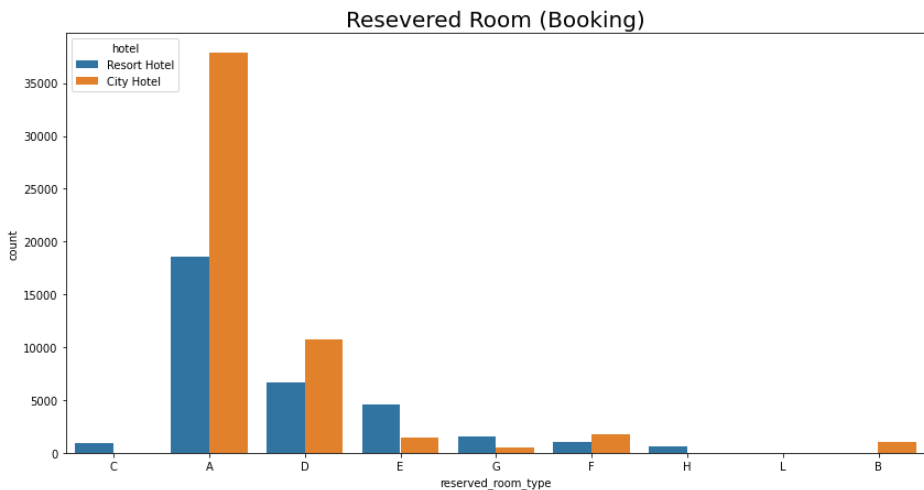
# Type Of Visitors

It seems that majority of the visitors travel in pair.



# Reserved Room & Assigned Room

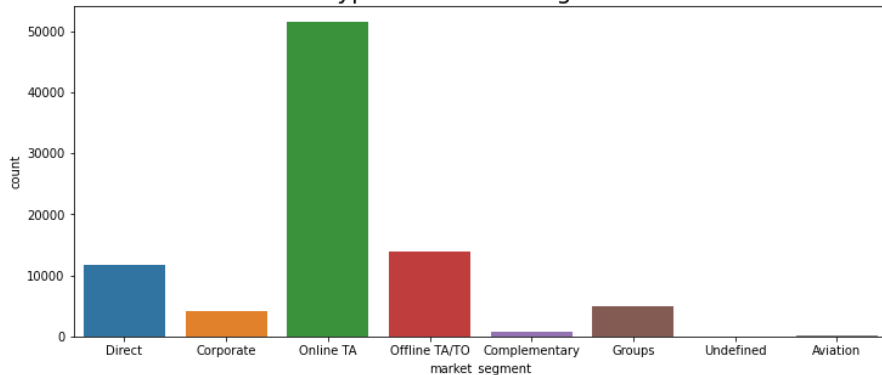
Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request.



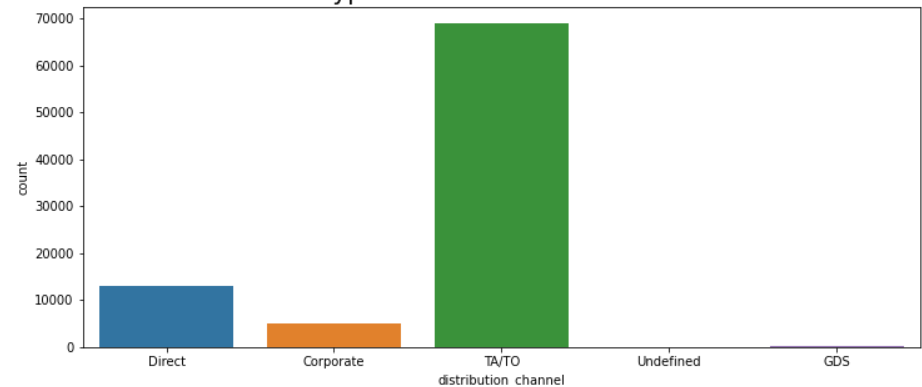
# Market Segments And Distribution Channel

- Majority of the distribution channels and market segments involve travel agencies (online or offline).
- We can target our marketing area to be on these travel agencies website and work with them since majority of the visitors tend to reach out to them.

Types of Market Segment

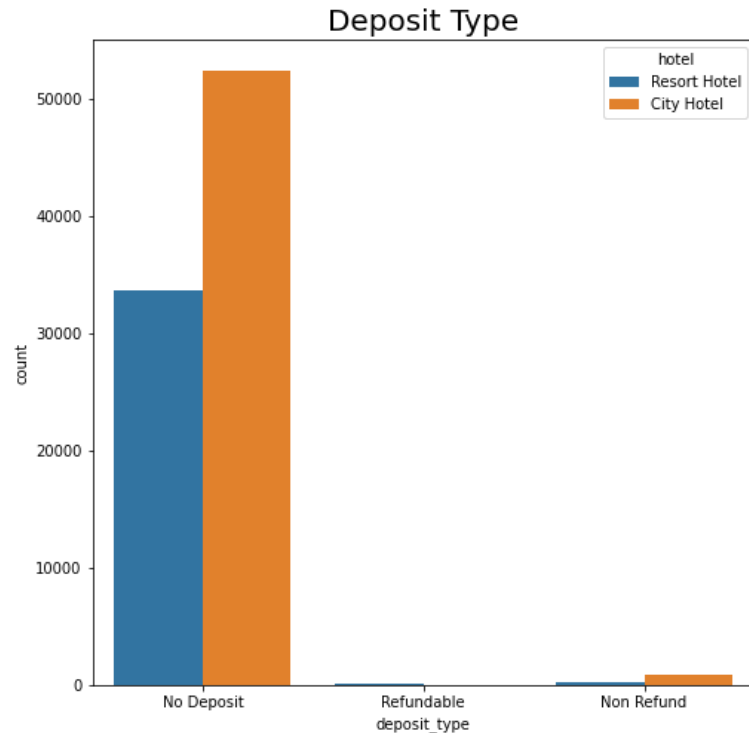


Types of distribution channel



# Deposit Types

- Majority of the booking does not require deposit regard-less of the Hotel type.
- We also realize that the high rate of cancellations can be due high no deposit policies.



**No Deposit** — no deposit was made;

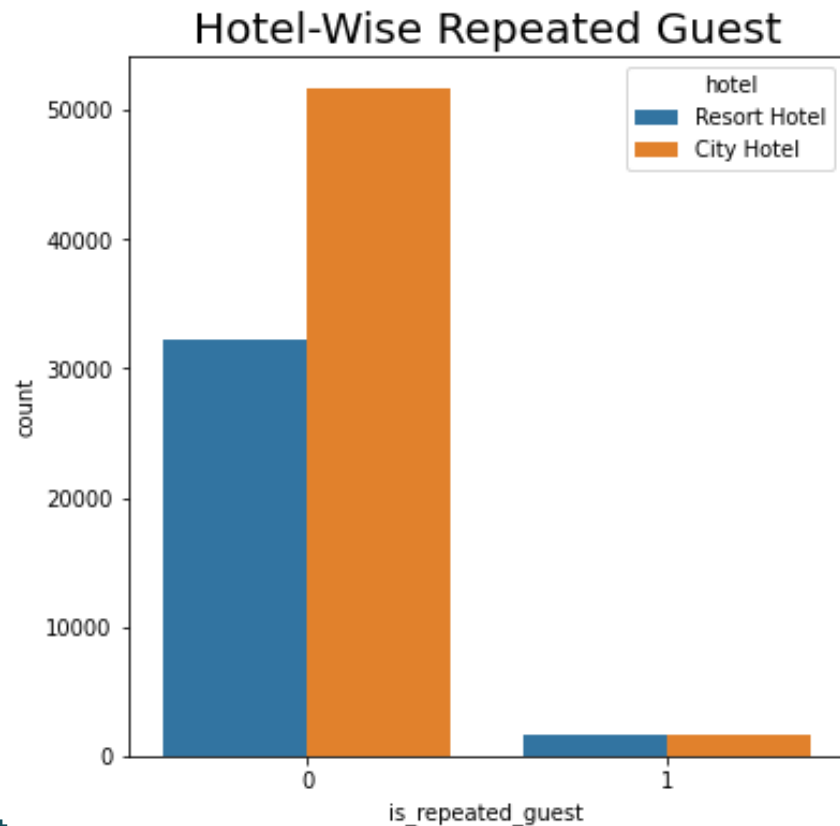
**Non Refund** — a deposit was made in the value of the total stay cost;

**Refundable** — a deposit was made with a value under the total cost of stay.



# Repeated Guest

- Low number of repeated guests.
- A need to target repeated guests since they have booked before.

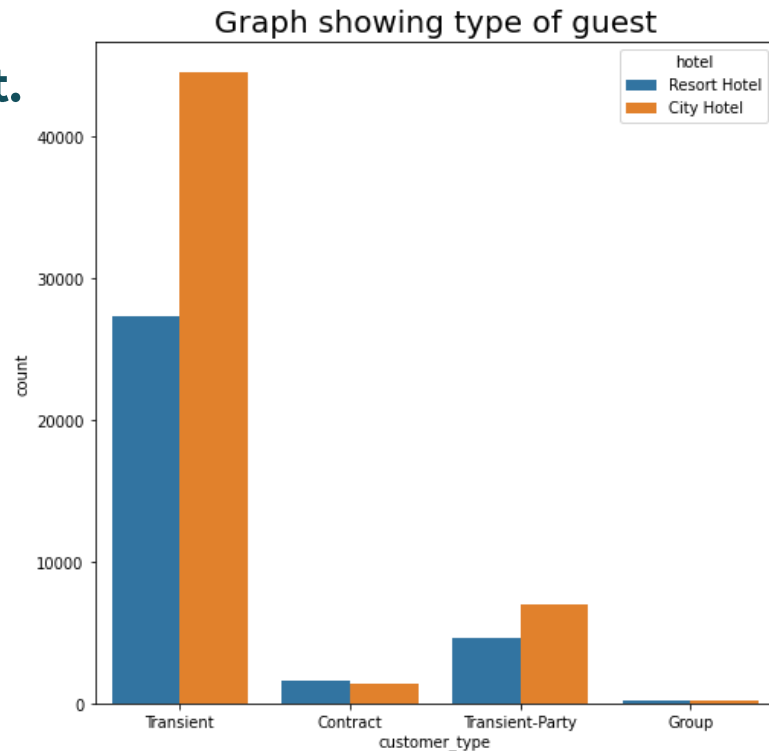


\* 0 – Not Repeated Guest

1 – Repeated Guest

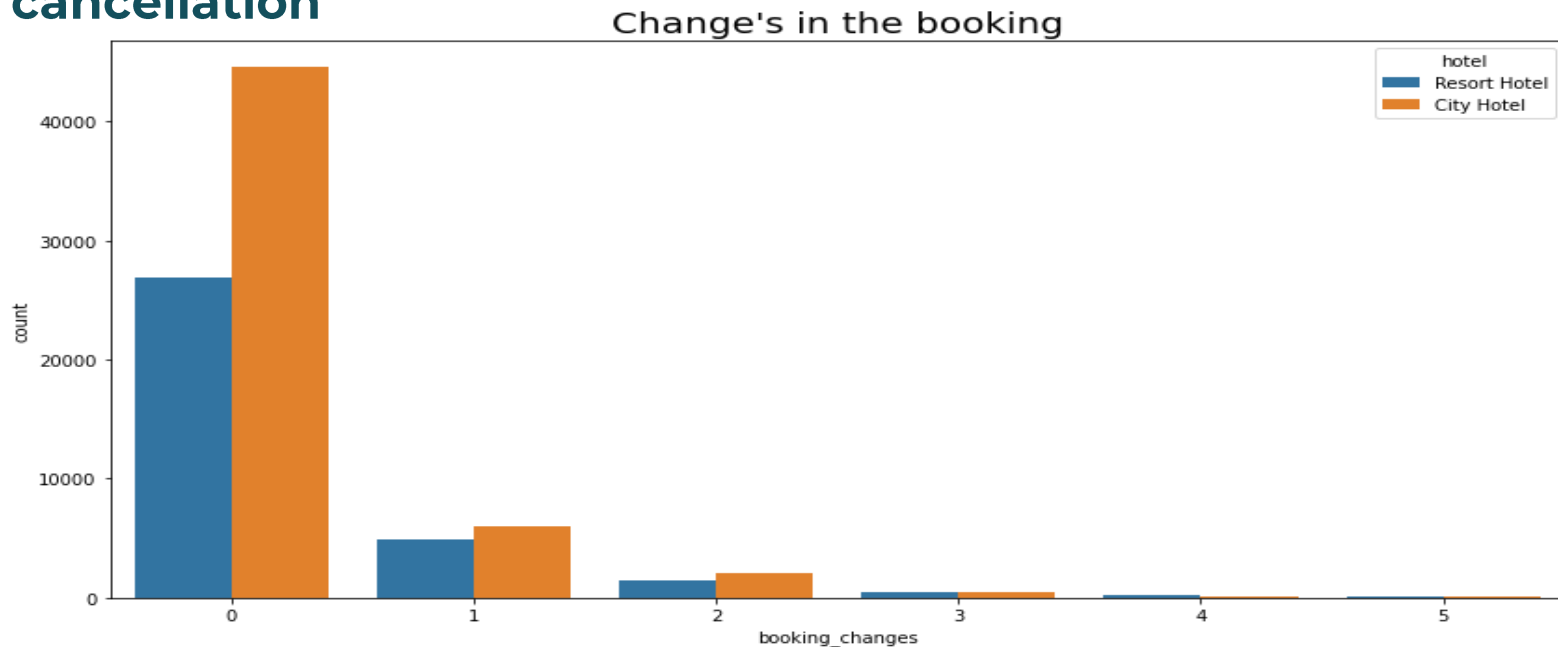
# Types of Guests

- **Majority of the bookings are Transient.** This means that the booking is not part of a group or contract.
- **With the ease of booking directly from the website, most people tend to skip the middleman to ensure quick response from their booking.**



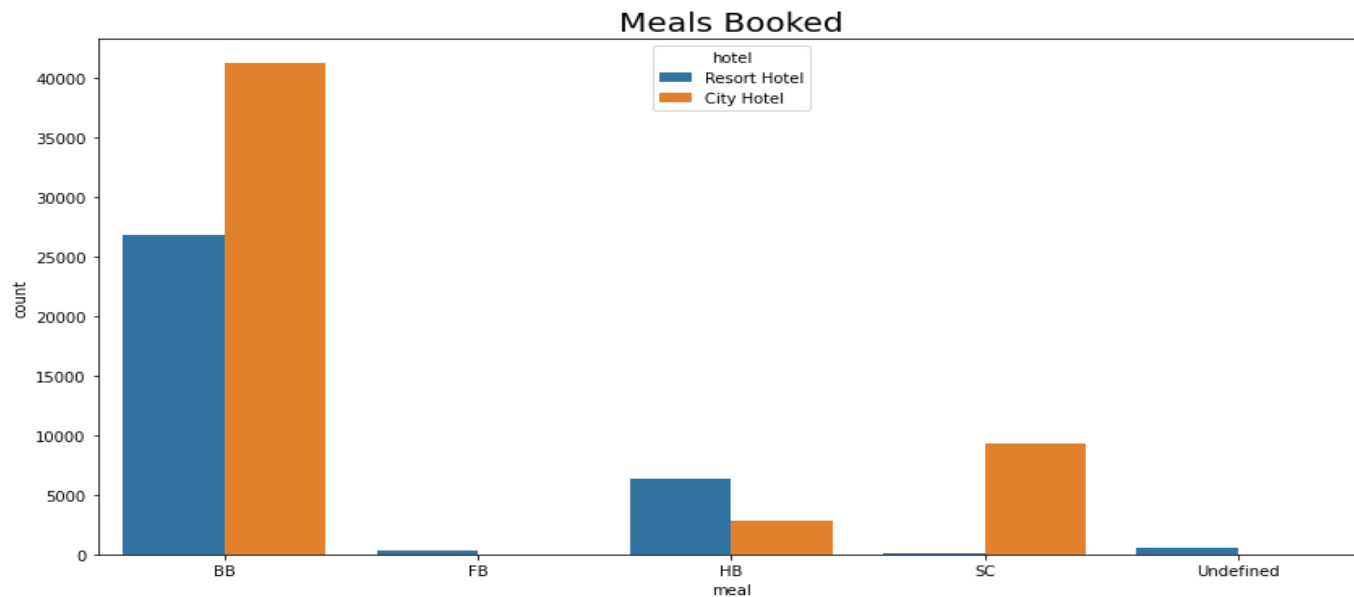
# Booking Change

Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation



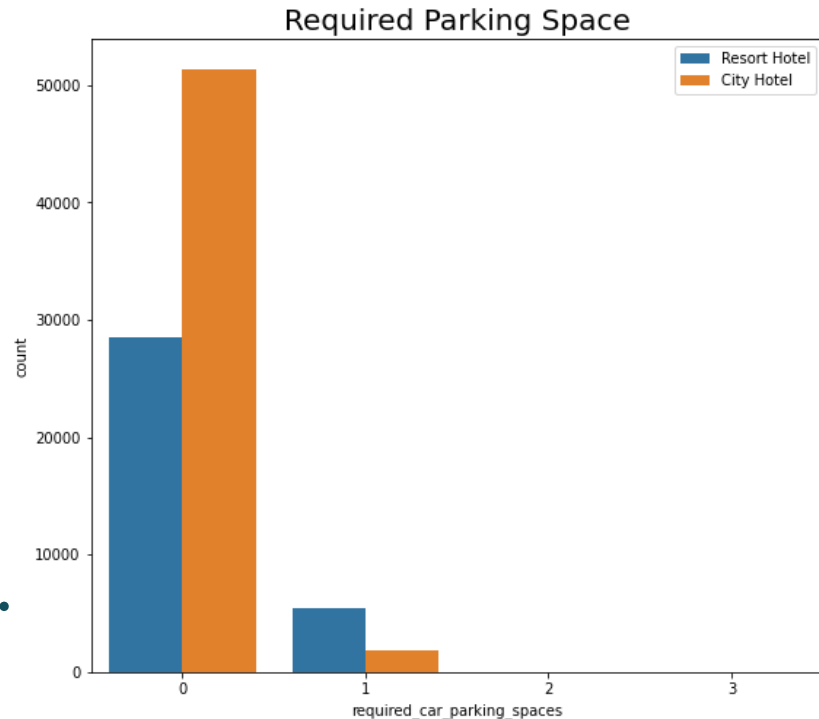
# Meals Booked

- Type of meal booked.
- Undefined/SC — no meal package;
- BB — Bed & Breakfast; ( Majority Booking )
- HB — Half board (breakfast and one other meal — usually dinner);
- FB — Full board (breakfast, lunch and dinner)



# Required Parking Space

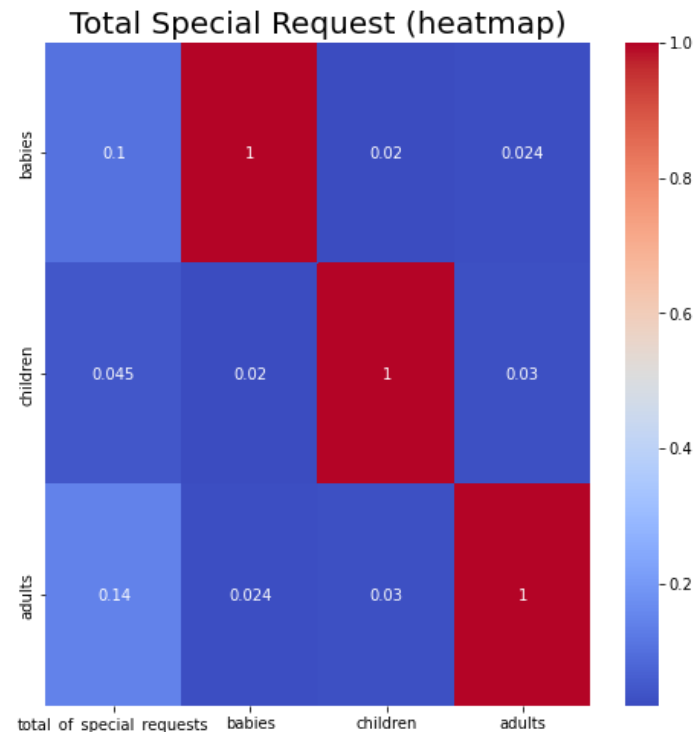
- Number of car parking spaces required by the customer.
- Most of the customer doesn't require parking space.
- The one's who require parking space are mainly from Resort Hotel.



# Total Special Request

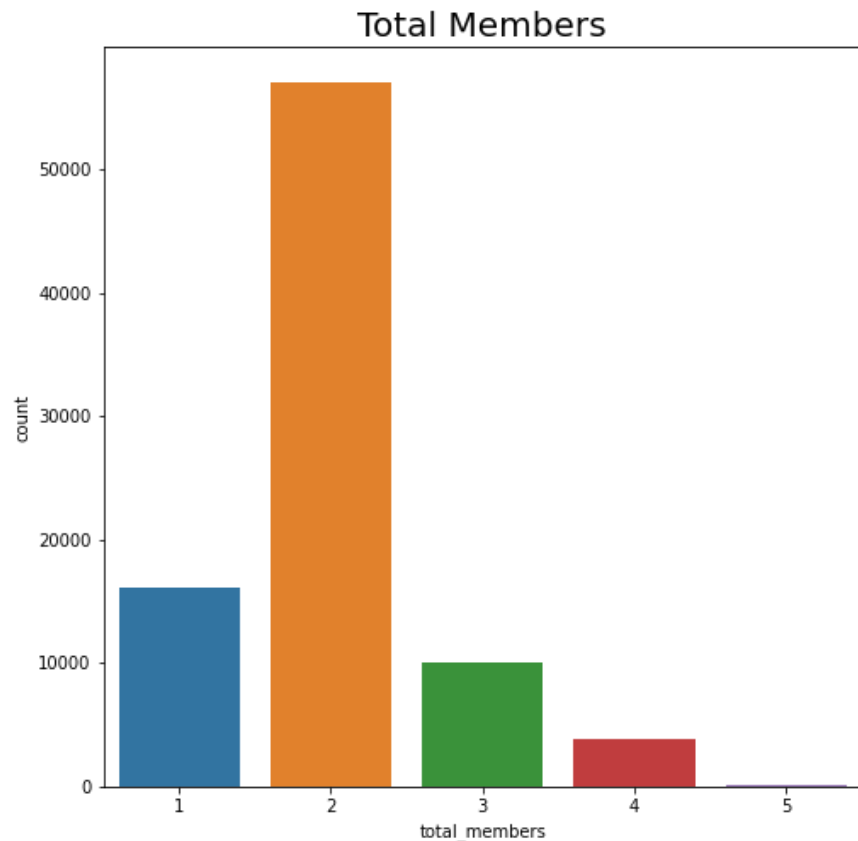
- Number of special requests made by the customer (e.g. twin bed or high floor)
- When travelling with babies travellers tend to make more special request

	total_of_special_requests	babies	children	adults
babies	0.103940	1.000000	0.019668	0.024028
children	0.044531	0.019668	1.000000	0.029586
adults	0.144755	0.024028	0.029586	1.000000



# Total Members

- Total members per reservation can be seen.
- Majority visits are by Pair.



# Cancellations

- City Hotel Cancellation Rate is 30.1%

```
city_canceled_per = round(city_canceled / df_copy[df_copy['hotel']=="City Hotel"].shape[0] * 100,2)
print(f"The cancellation rate of City Hotel is {city_canceled_per}%")
```

The cancellation rate of City Hotel is 30.1%

- Resort Hotel Cancellation Rate is 23.48%

```
resort_canceled_per = round(resort_canceled / df_copy[df_copy['hotel']=="Resort Hotel"].shape[0] * 100,2)
print(f"The cancellation rate of Resort Hotel is {resort_canceled_per}%")
```

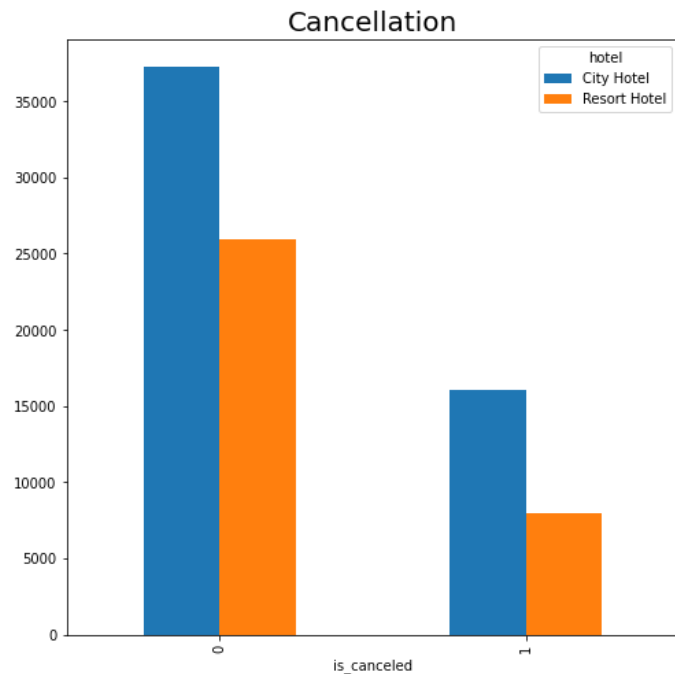
The cancellation rate of Resort Hotel is 23.48%

- Total Cancellation Rate is 27.52%

```
total_canceled_per = round(total_canceled / df_copy.shape[0] * 100,2)
print(f"The cancellation rate of both Hotels is {total_canceled_per}%")
```

The cancellation rate of both Hotels is 27.52%

- Total Non – Cancellation Rate is 72.48%

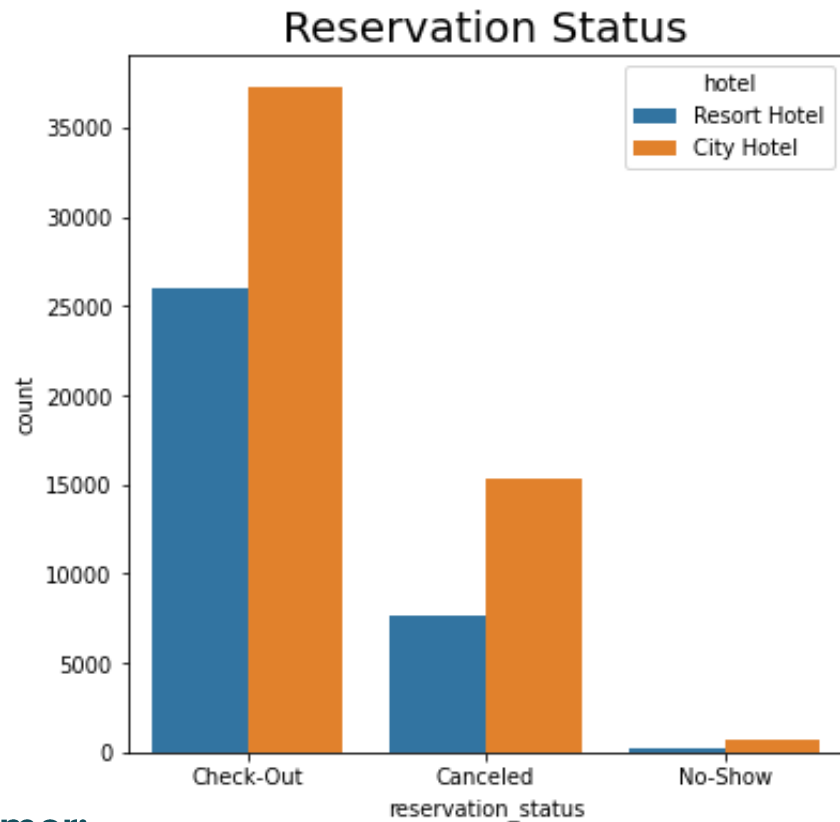


**0- Not – Cancelled ; 1- Cancelled**



# Reservation Status

- Reservation last status as per booking.
- That could explain why cancellation rate was actually 50% of non-cancellation rate.



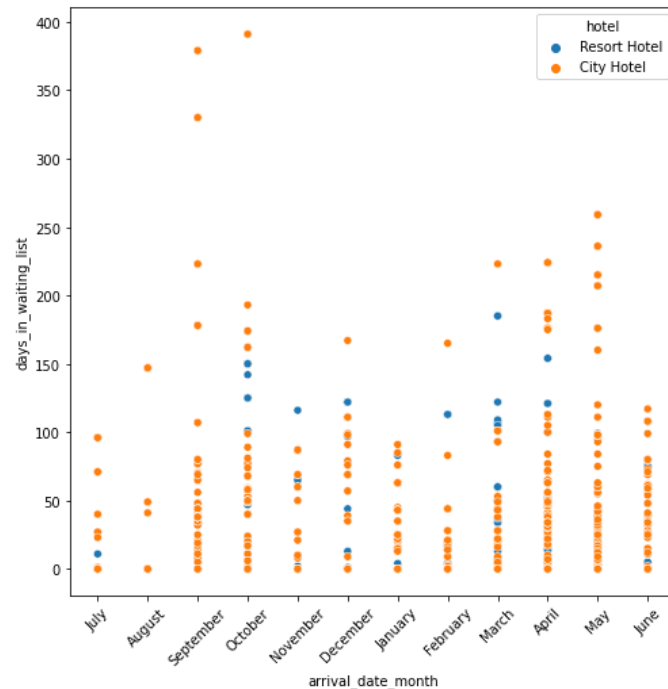
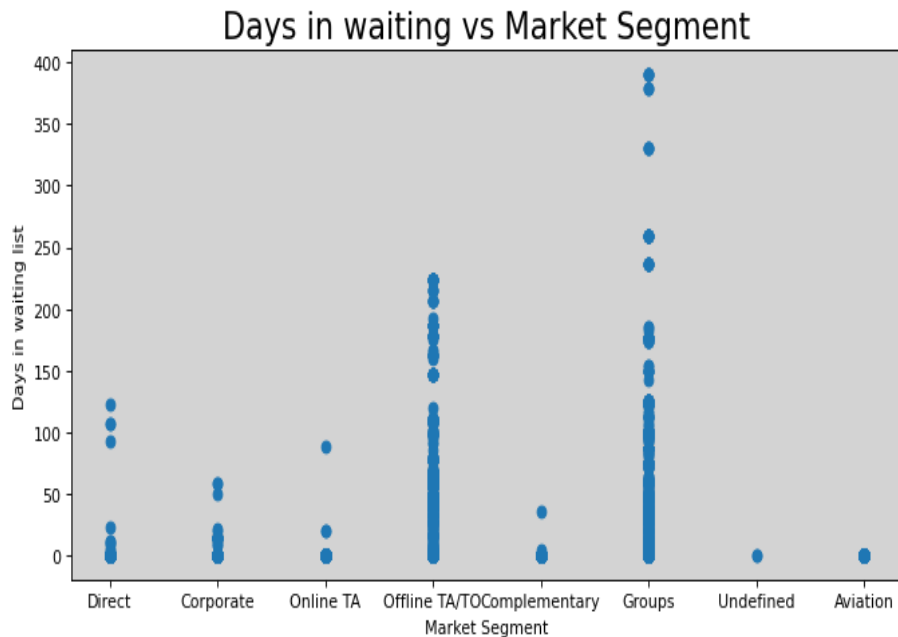
Canceled — booking was canceled by the customer;

Check-Out — customer departed;

No-Show — customer did not check-in and did inform the hotel of the reason why

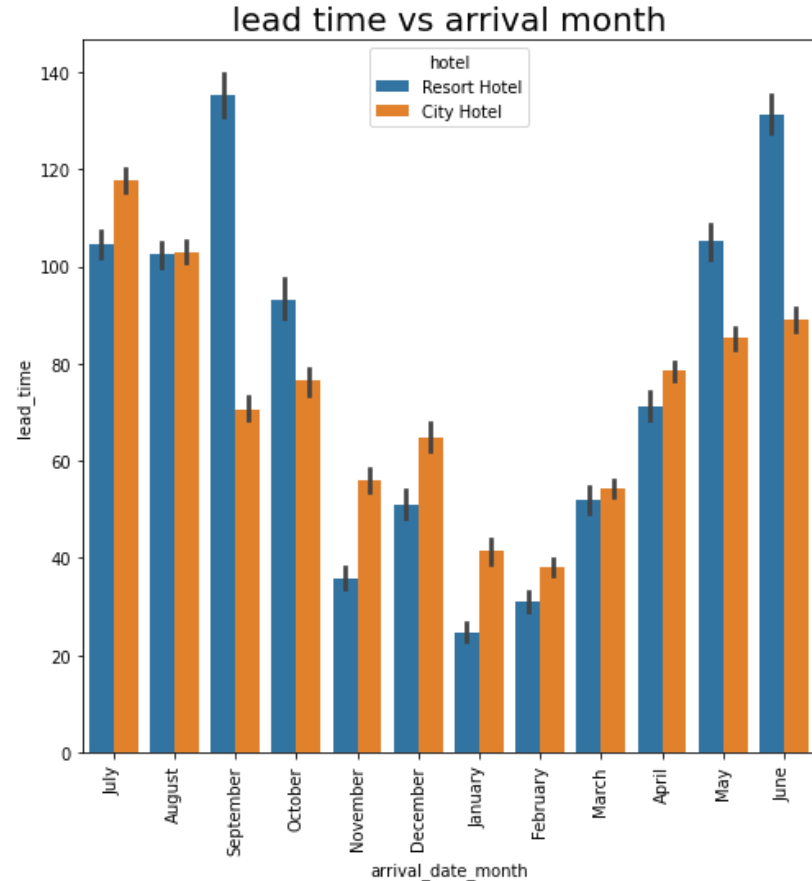
# Days in Waiting List

No of days the booking was in the waiting list before it was confirmed to the customer.

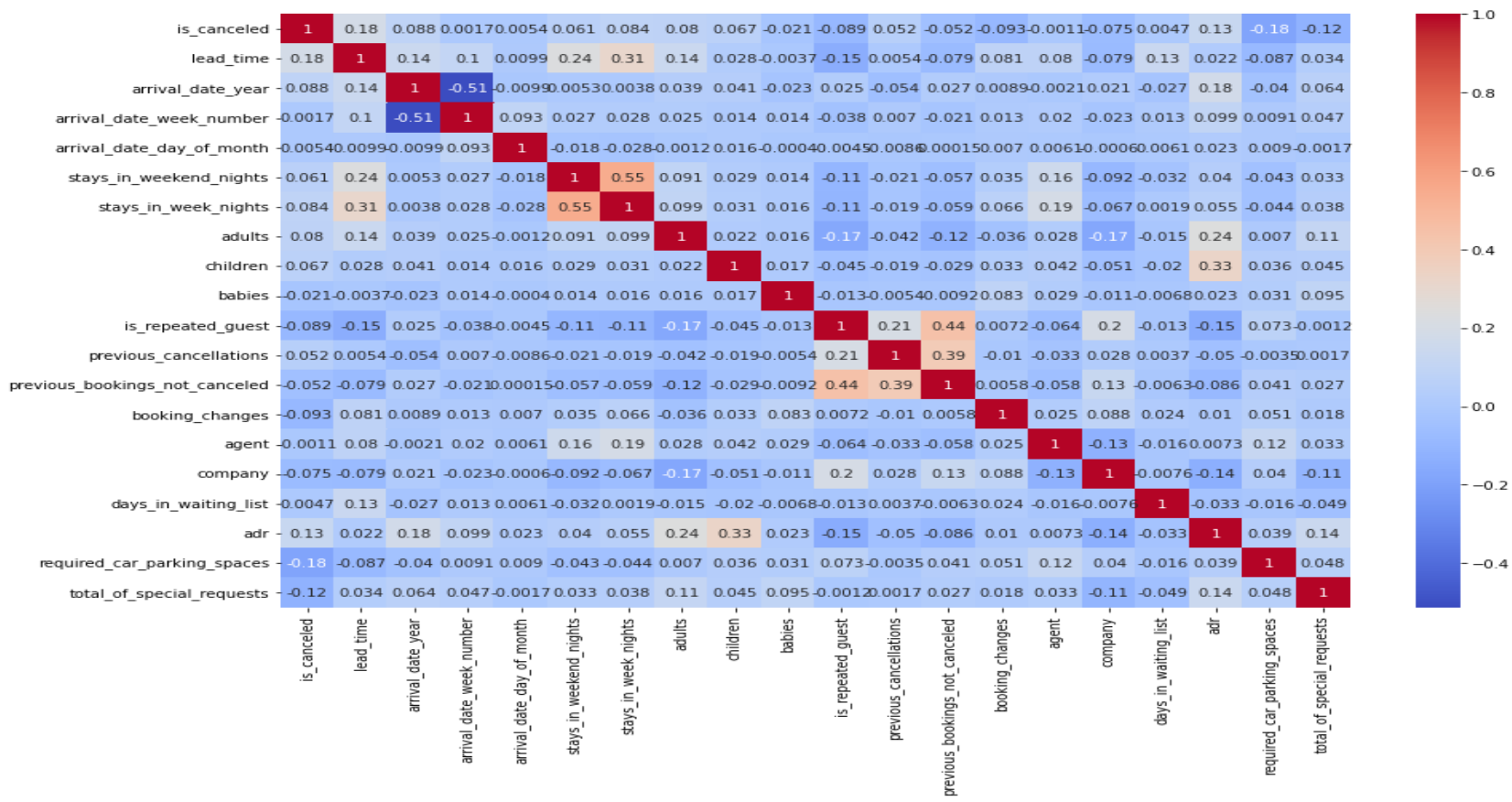


# Lead Time

- No of days that elapsed between entering date of booking into Property Management System and Arrival Date.
- From May to October the bookings were made before 3 to 4 months as it was the peak time.

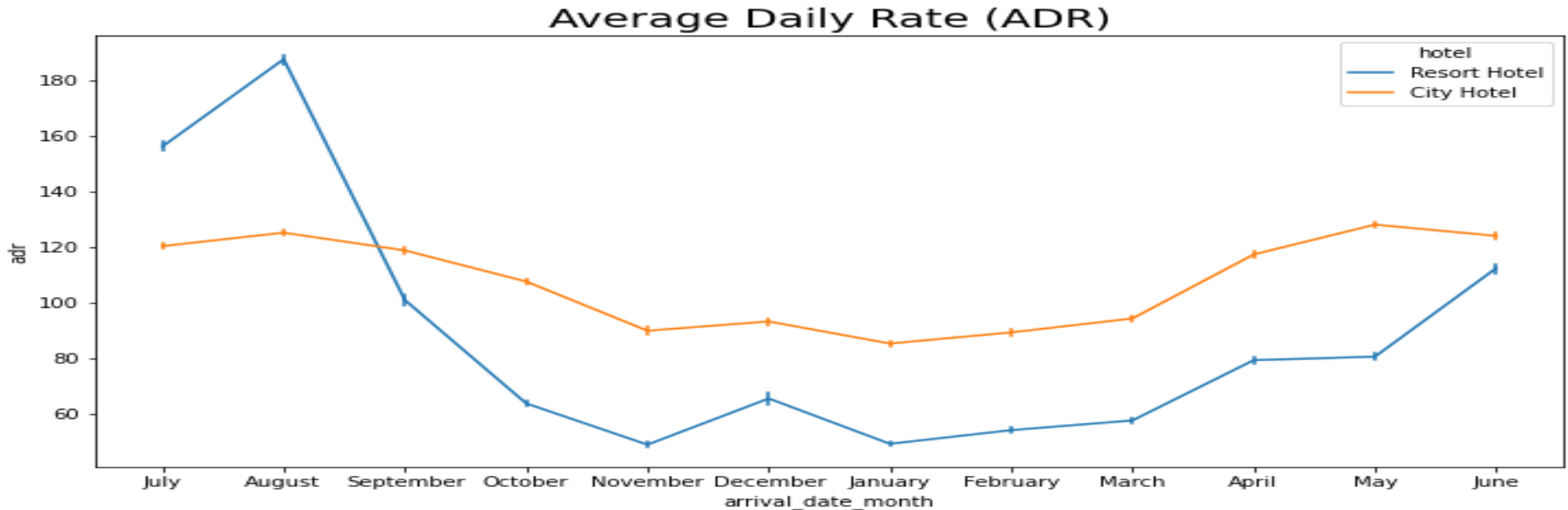


# Heatmap



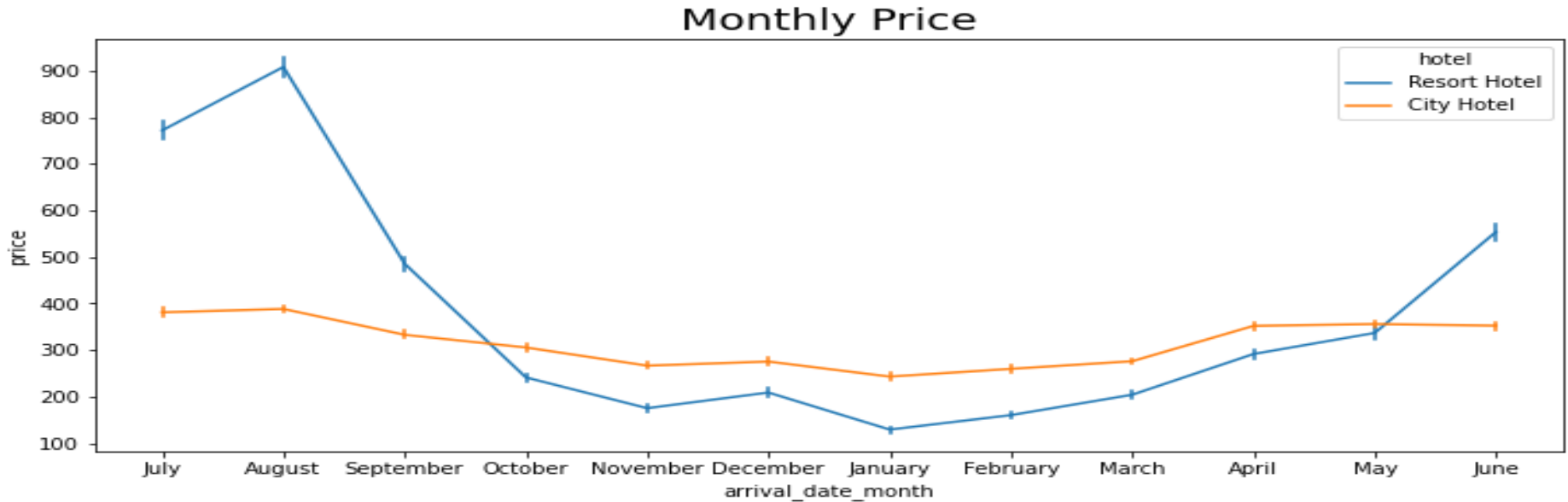
# Average Daily Rate

- For Resort Hotels, the average daily rate is more expensive during July, August and September.
- For City Hotels, the average daily rate is more expensive during May, June, July and August.



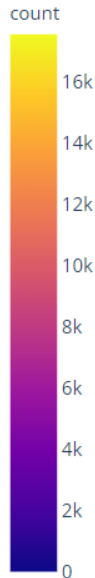
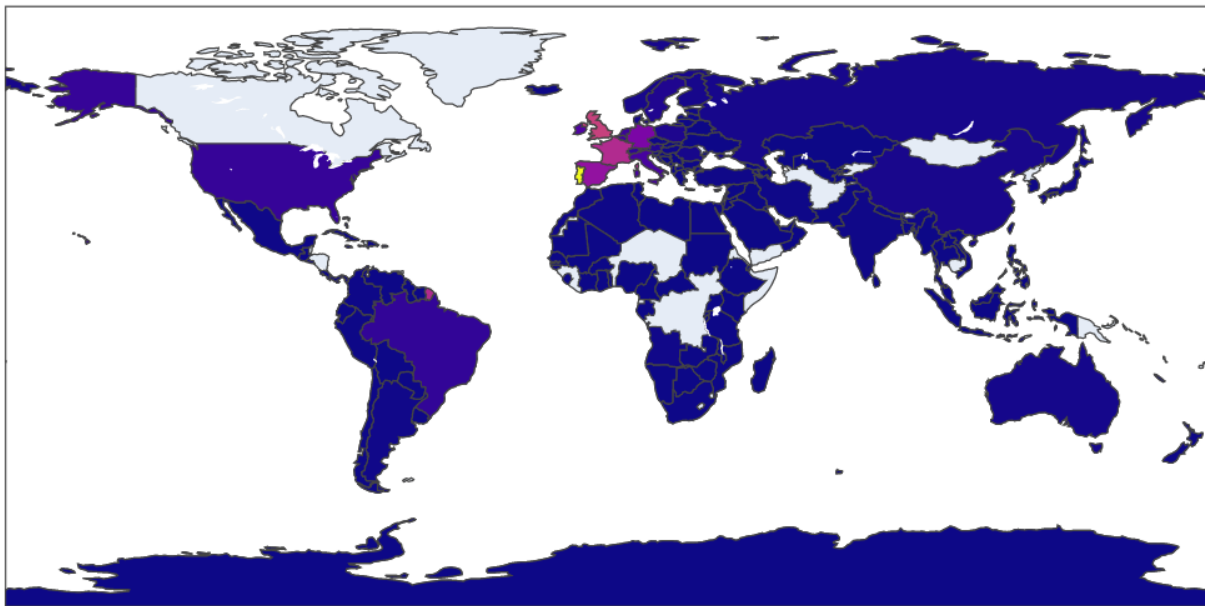
# Monthly Price

- Prices of Resort Hotel are much higher.
- Prices of City Hotel do not fluctuate that much.



# Home Country of Visitors

- Portugal - 27,355
- Great Britain- 10,424
- France - 8,823
- Spain – 7244
- Germany - 5,385.



# Questions

1. Which are the Five most common and uncommon countries of origin?
2. Which are the months of highest and least occupation?
3. What is the most popular meal package?
4. Which is the most reserved room type?
5. How many bookings changes have been done during the studied period?
6. How many people have been registered in the hotel?
7. What is the most common customer type?



# Question and Answers

## Question 1 –

**Which are the Five most common and uncommon countries of origin?**

## Answer 1 –

As we can see, Portugal tops the list with 27,355 of the cases, followed by Great Britain with 10,424; France with 8,823; Spain with 7244 and Germany with 5,385.

```
df_copy.country.value_counts().head(5)
```

PRT	27355
GBR	10424
FRA	8823
ESP	7244
DEU	5385

Name: country, dtype: int64

## Question 2 –

**Which are the months of highest and least occupation?**

## Answer 2 –

The month of highest occupation is august with 11.65% of the reservations. The month of lest occupation is January with 4.94% of the reservations.

```
df_copy.arrival_date_month.value_counts()/100
```

August	112.42
July	100.43
May	83.44
April	79.00
June	77.56
March	74.89
October	69.21
September	66.82
February	60.83
December	51.12
November	49.73
January	46.85

Name: arrival\_date\_month, dtype: float64

### Question 3 –

**What is the most popular meal package?**

### Answer 3 –

**The Bed & Breakfast option is the most popular, with a frequency of 77.85%.**

```
df_copy.meal.value_counts(normalize=True)*100
```

```
BB          77.848217
SC          10.765792
HB          10.409263
Undefined    0.564026
FB           0.412702
Name: meal, dtype: float64
```

## Question 4 – Which is the most reserved room type?

### Answer 4 –

The “A” room type is the most popular among the clients, with 64.69% of the reservations.

```
df_copy.reserved_room_type.value_counts(normalize=True)*100
```

```
A    64.697925
D    19.919752
E     6.919638
F     3.232833
G     2.350109
B     1.141809
C     1.047805
H     0.683251
L     0.006878
Name: reserved_room_type, dtype: float64
```

## Question 5 –

How many bookings changes have been done?

## Answer 5 –

There were 23,033 registered changes in the bookings during this period.

```
df_copy.booking_changes.sum()
```

```
23033
```

## Question 6 –

**How many people have been registered in the hotel?**

## Answer 6 –

**1,76,939 people have been registered in the hotel.**

```
df_copy.total_members.sum()
```

```
176639
```

## Question 7 –

**What is the most common customer type?**

## Answer 7 –

**Transients are the most common customer type, they represent 82.38% of the total customers.**

```
df_copy.customer_type.value_counts(normalize=True)*100
```

Transient	82.382208
Transient-Party	13.403646
Contract	3.593947
Group	0.620199

Name: customer\_type, dtype: float64

# Inferences and Conclusion

- The Majority Of Guests Come From Western Europe Countries.
- The Majority Of Reservations Are For City Hotels.
- The Number Of Repeated Guests Is Too Low. We Should Target Our Advertisement On Guests To Increase Returning Guests.
- The Majority Of Reservations Converts Into Successful Transactions.
- We Also Realize That The High Rate Of Cancellations Can Be Due High No Deposit Policies.
- We should also target months between May to Aug. Those are peak months due to the summer period.



Thank You