

Capstone Project

Hotel Booking Analysis

Team Dataloft

Mohan Vishe

Gaurav Yadav

Rahul Ray

Shambhuraj Desai

Contents

- Problem Statements
- Data Summary
- Data Cleaning & Manipulation
- Univariate Analysis
- Hotel Booking Analysis
- Timewise Analysis
- Deposit Type Analysis
- Preferred Room Type
- Monthly price Analysis
- Correlation Heatmap
- Distribution Channel Analysis
- Conclusion

Problem Statement



Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

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

Data Summary

The collected data had 1,19,330 records/observations and 32 columns/features. For this project we will be analysing Hotel Booking data. This data set contains booking information for a hotels and includes data such as when the booking was made, length of stay, annually number of bookings, countries, market segment, distribution channel, different types of meals, deposit types, different types of room, required car parking, special request and reservation status.

Data Description

Hotel: Resort Hotel & City Hotel

Is_Canceled: Cancelled (1) & Not Cancelled (0)

Lead_time: No. of days that elapsed between entering date of booking into property management system and arrival date.

arrival_date_year: Year of arrival date (2015-2017)

arrival_date_month: Month of arrival date (Jan - Dec)

arrival_date_week_number: Week number of year for arrival date (1-53)

arrival_date_day_of_month: Day of arrival date

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.

People - **Adults:** No. of Adults; **Children:** No. of Children; **Babies:** No. of Babies

meal: Type of Meal booked (BB), (HB), (FB), (SC)

Market_segment: Market segment Designation.(TA: Travel agents; TO: Tour operators)

distribution_channel: Booking distribution channel.(TA: Travel agents; TO: Tour operators)

is_repeated_guest: is a repeated guest (1) or not (0)

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 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 of the type of room assigned to the booked.

booking_changes: No. of changes 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, Non Refund, Refundable

agent: ID of the travel agency that made the booking.

company: ID of the company/entity that made the booking.

day_in_waiting_list: Number of days the booking was in the waiting list before it was confirmed to the customer

customer_type: (Contract, Group, Transient, Transient Party)

adr: average daily rate

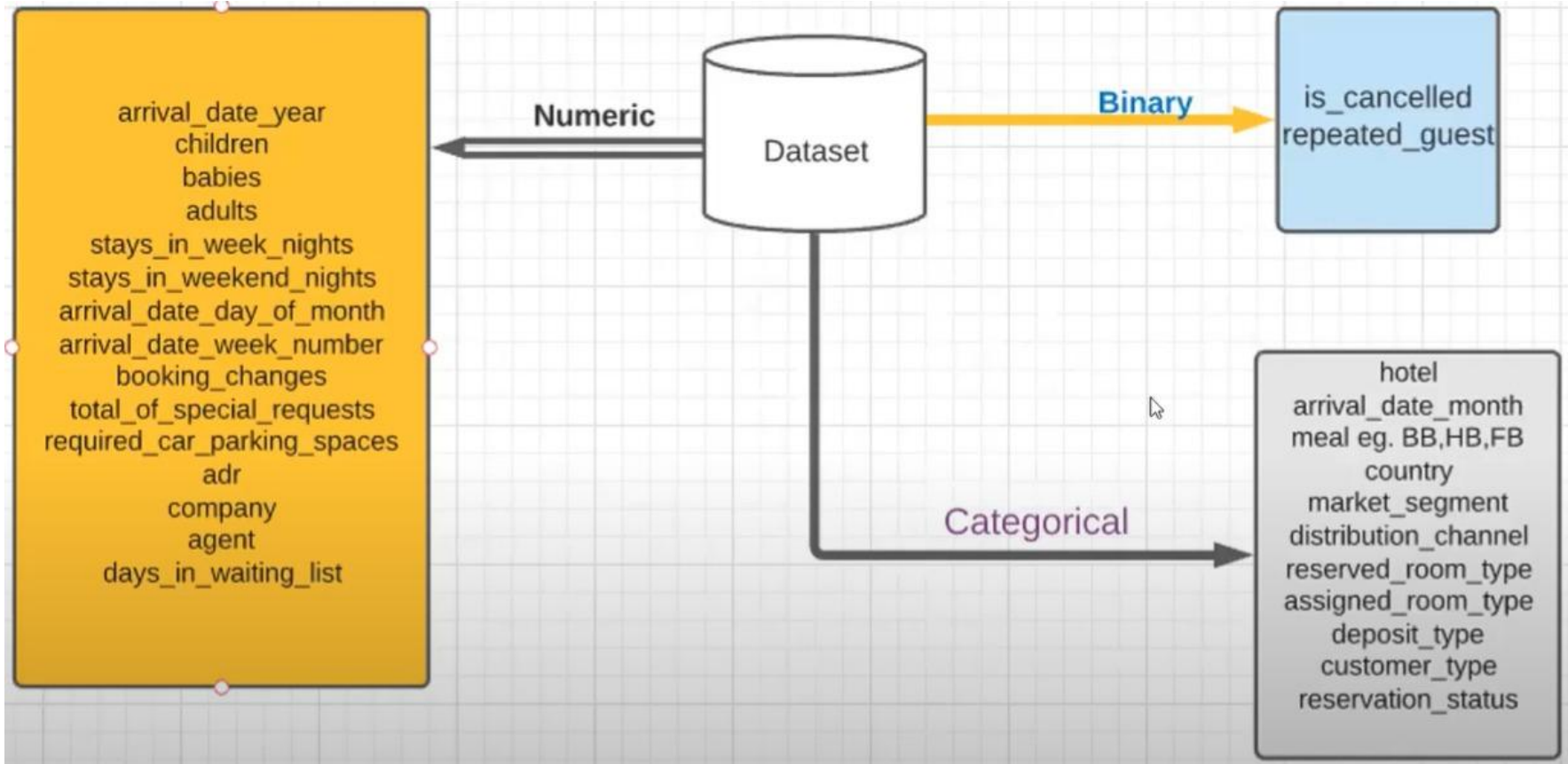
required_car_parking_spaces: No of car parking spaces required by the customer

total_of_special_requests: Number of special requests made by the customer

reservation_status: No. of reservation status bookings

reservation_status_date: No. of reservation bookings date of status

Types of Dataset



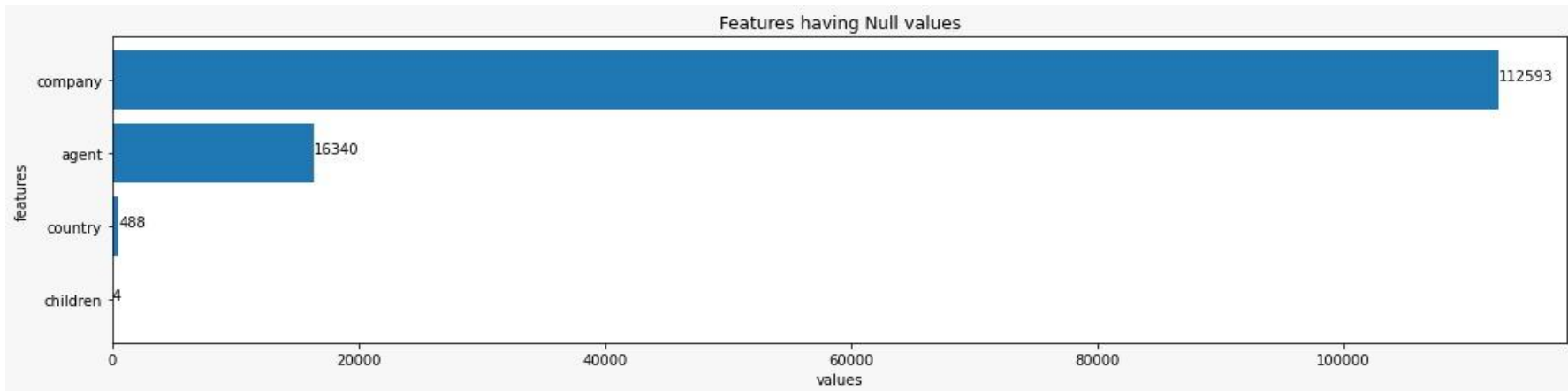
#Exploratory Data Analysis

Data Cleaning & Manipulation

In our data of 1,19,390 rows and 32 columns. There are missing values in columns namely countries, agent, company & children.

Missing Values

- In 'countries' we have 488 missing values.
- In 'agent' we have 16,340 missing values.
- In 'company' we have 1,12,593 missing values.
- In 'children' we have 4 missing values.



Handling Missing Values:

```
[ ] # Remove columns company and agent
df=df.drop(columns=["company","agent"])

# for Null children value, replace with rounded median value
df['children'].fillna(round(HB.children.median()), inplace=True)

# Remove rows containing missing contry values
df = df.dropna(axis = 0)
```

Finally We have Dataset with no null or missing Values

▼ 2. Handling Datatypes:

As the number of children should be integer we should convert float to integer

```
[ ] ## converting datatype from float to integer
df["children"]=df["children"].astype("int32")
```


Univariate Analysis

While doing univariate analysis of given hotel booking dataset, we answered following questions:

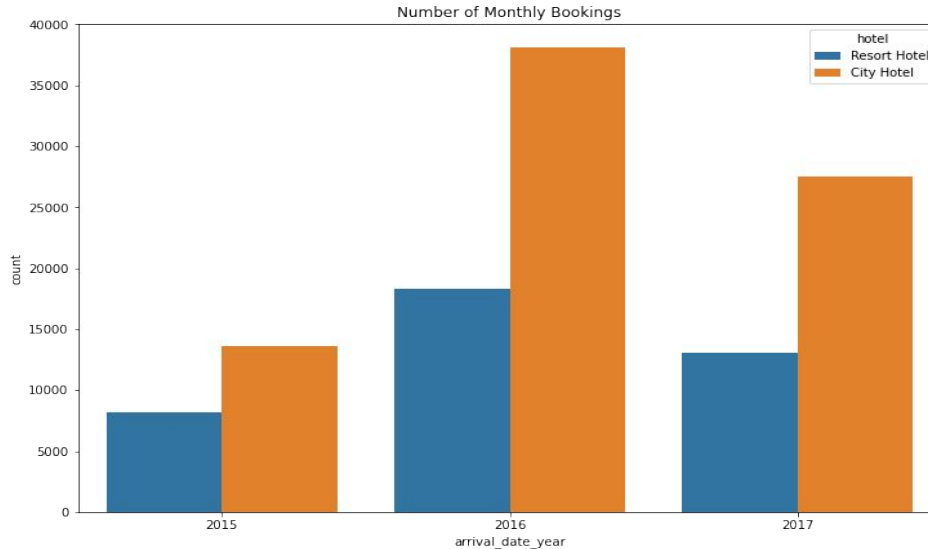
Questions

1. What is the percentage of hotels bookings?
2. What is the annually number and percentage of arriving visitors?
3. Which hotel has more cancellation of bookings?
4. Which months has the highest and lowest number of bookings?
5. Correlation between numerical feature.
6. See if the ADR increases as the number of people increases.
7. Which country has the highest number of booking?
8. Which Sector should be targeted for our advertisement?
9. Which months has cheaper booking rates?

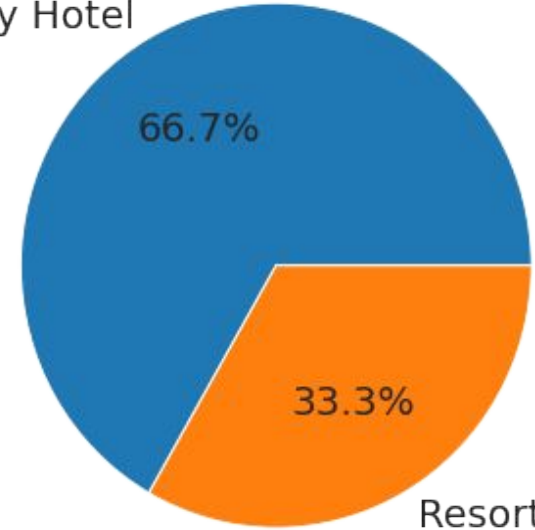
Hotel Booking Analysis

Hotel Comparison

- In hotels we have 2 types of hotels i.e. City Hotel & Resort Hotel.
- In 1,19,390 observations City Hotels are 79,306 and Resort Hotels are 39,596 bookings.
- Where City Hotel is 66.7 % and Resort Hotel is 33.3 %.
- In the hotels booking cancellation is 62.9% & 37.1%.



City Hotel

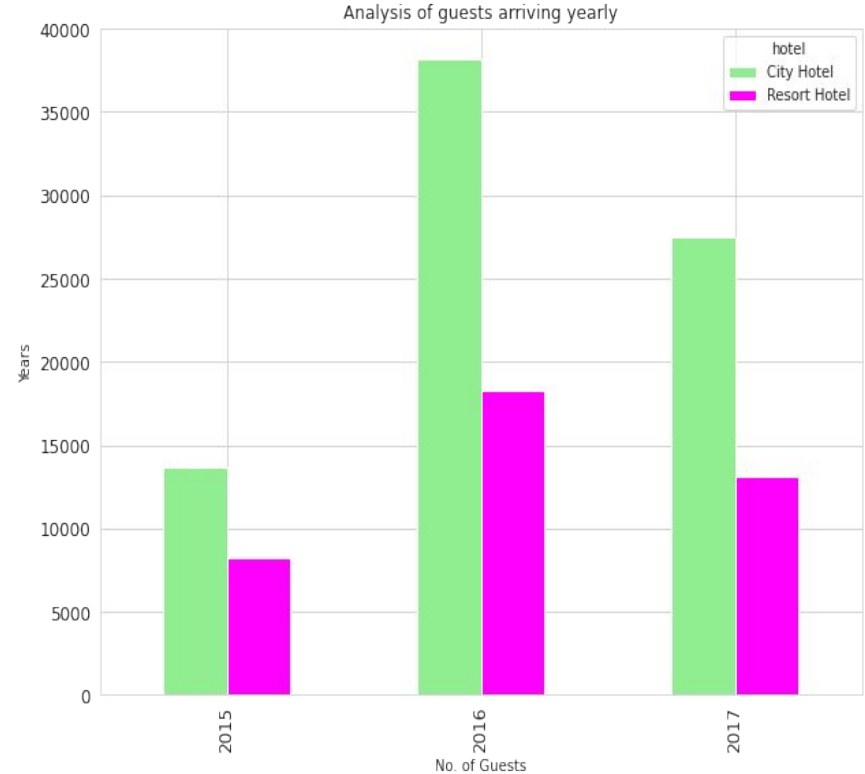
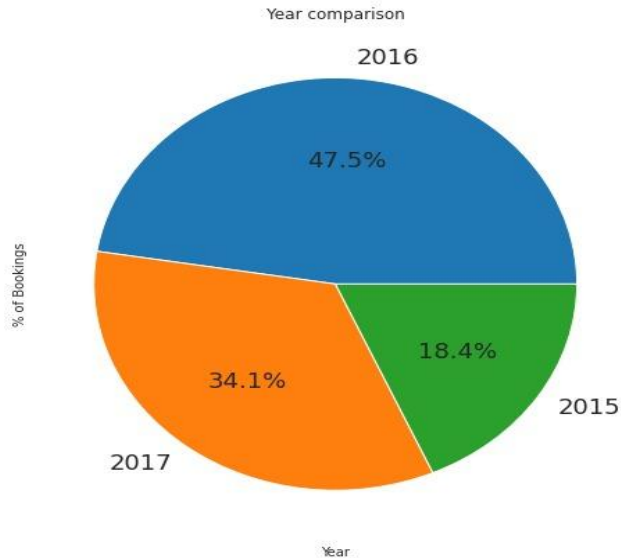


Resort Hotel

Timewise Analysis

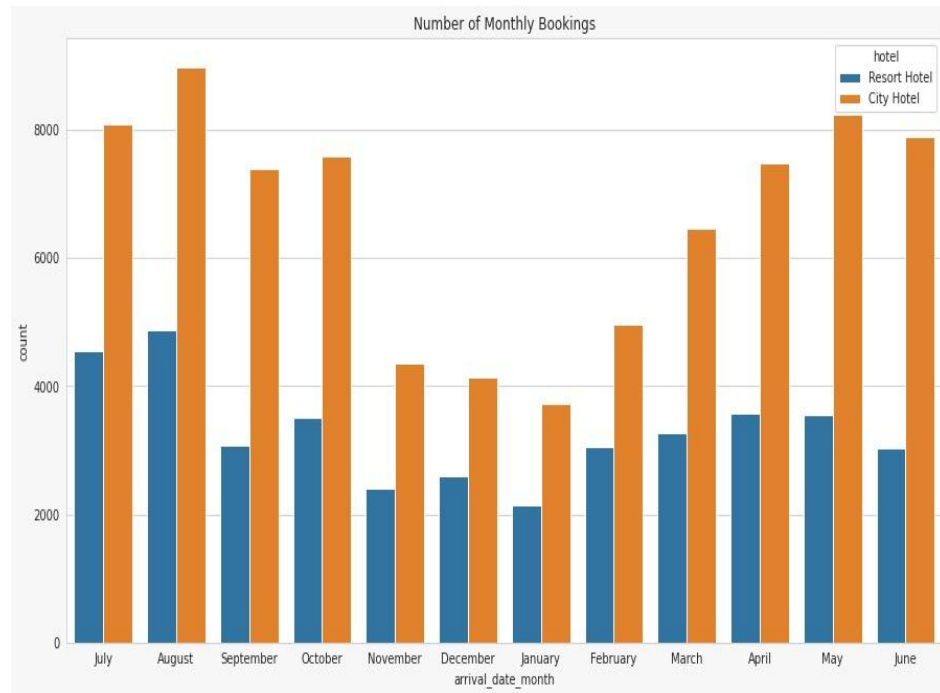
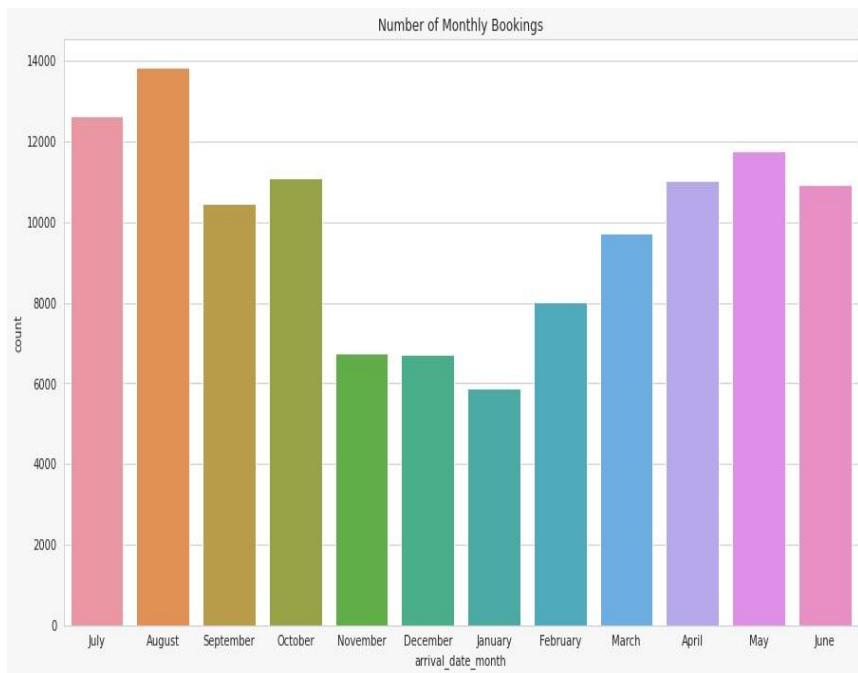
Yearly Analysis

- In year 2015 there are 21,863, In 2016- 56,435 & In 2017- 40,604 bookings.
- 2016 has more number of bookings than 2015 & 2017.
- We can see that after 2015 bookings increased in 2016 but then increased in 2017.
- In yearly comparison 2015 is 18.4%, 2016 is 47.5% and 2017 is 34.1 %.

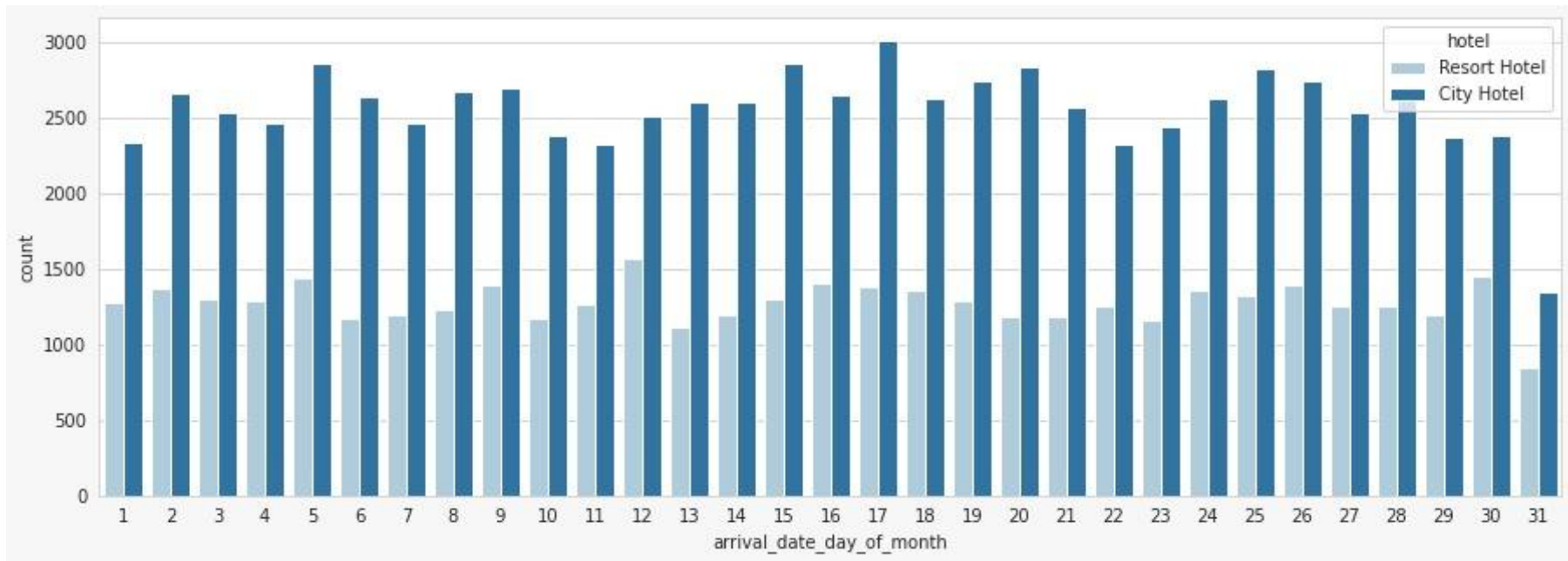


Monthly Analysis

- In months, August has the highest number of bookings with 13856 bookings.
- November, December and January has the lowest number of bookings with 6752, 6728 and 5874 bookings.

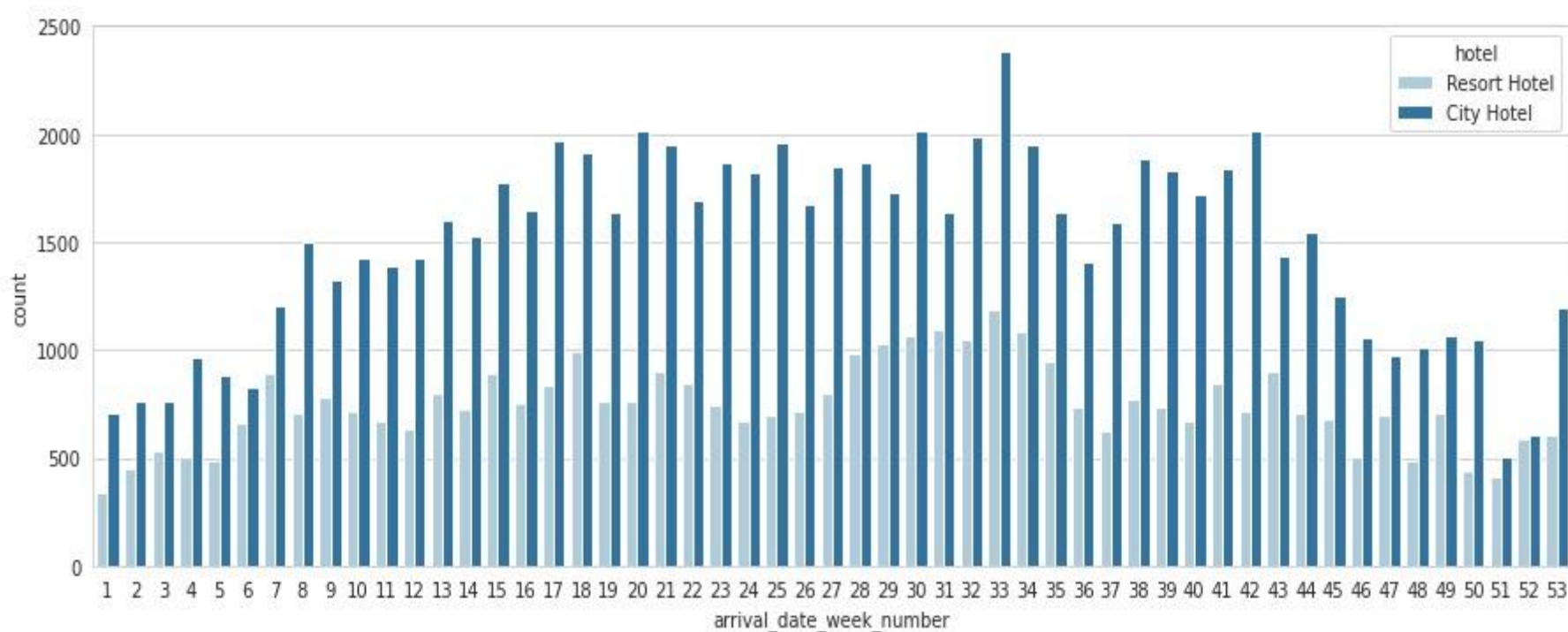


Daily Analysis



There is no variation seen when we see it on daily basis and for the date 31st, it is the lowest as only 7 months have the date 31.

Arrival Date week analysis



In City hotel bookings are volatile as compare to resort hotel bookings.

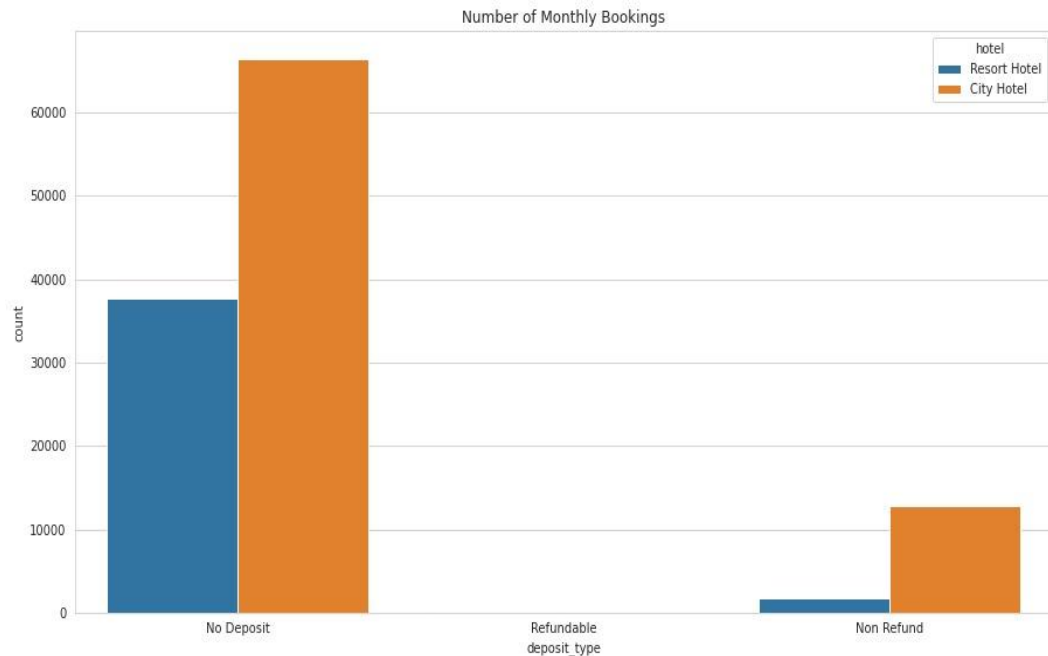
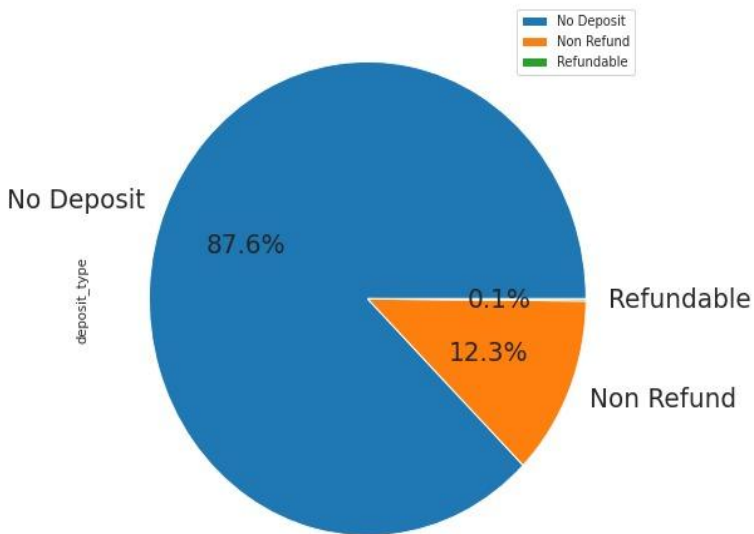
In City hotel, from 8th week to 42th week shows high bookings.

Deposit Type Analysis

There are three types of deposits namely “No Deposit”, “Non Refund”, & “Refundable”.

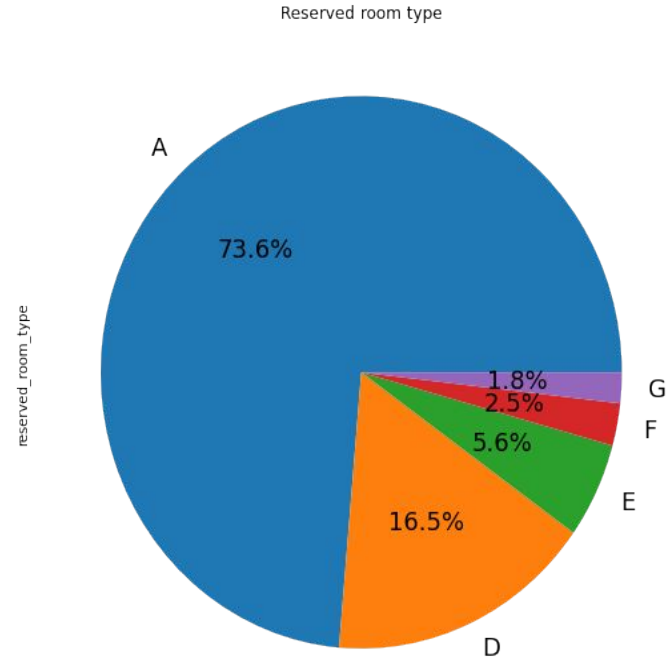
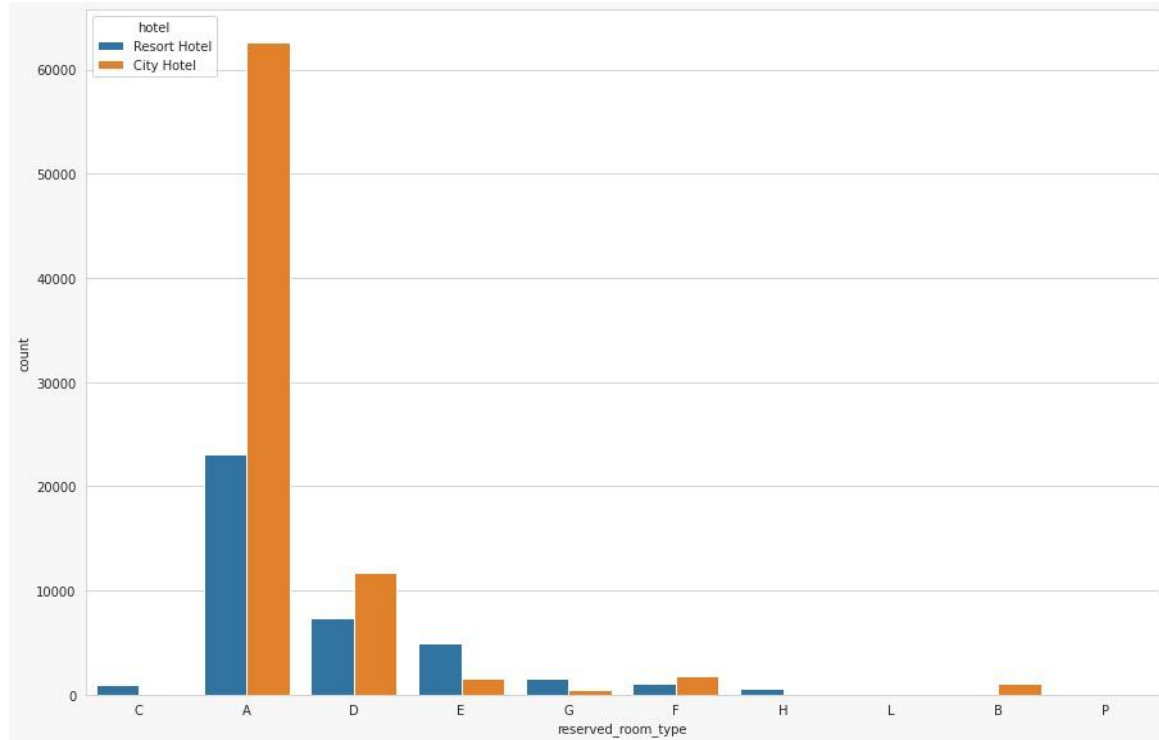
“No Deposit” type has the highest number of bookings with 1,04,167 while “Refundable” is least.

“No Deposit” is 87.6% and “Non Refund” is 12.3% and “Refundable” is 0.1% in deposit type.

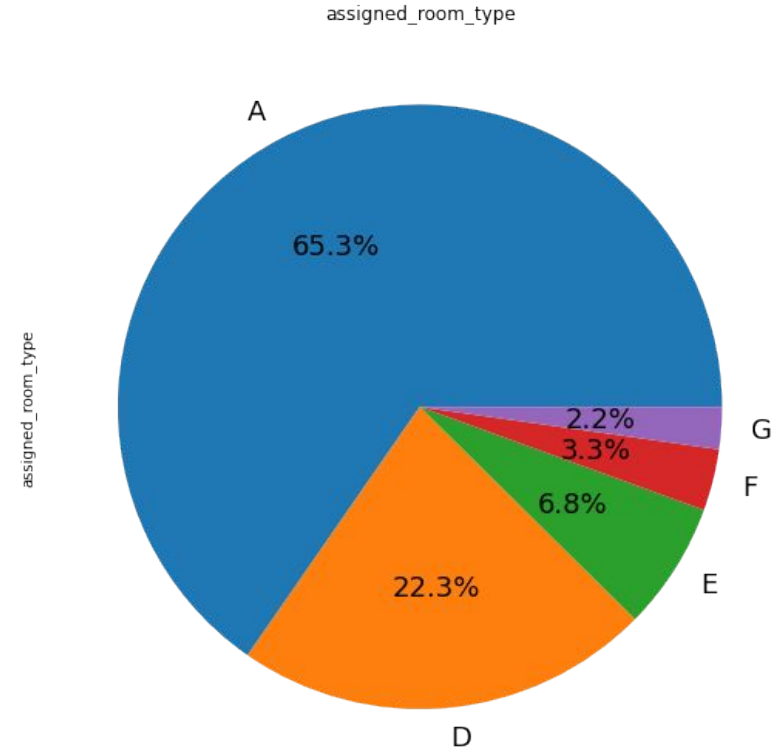
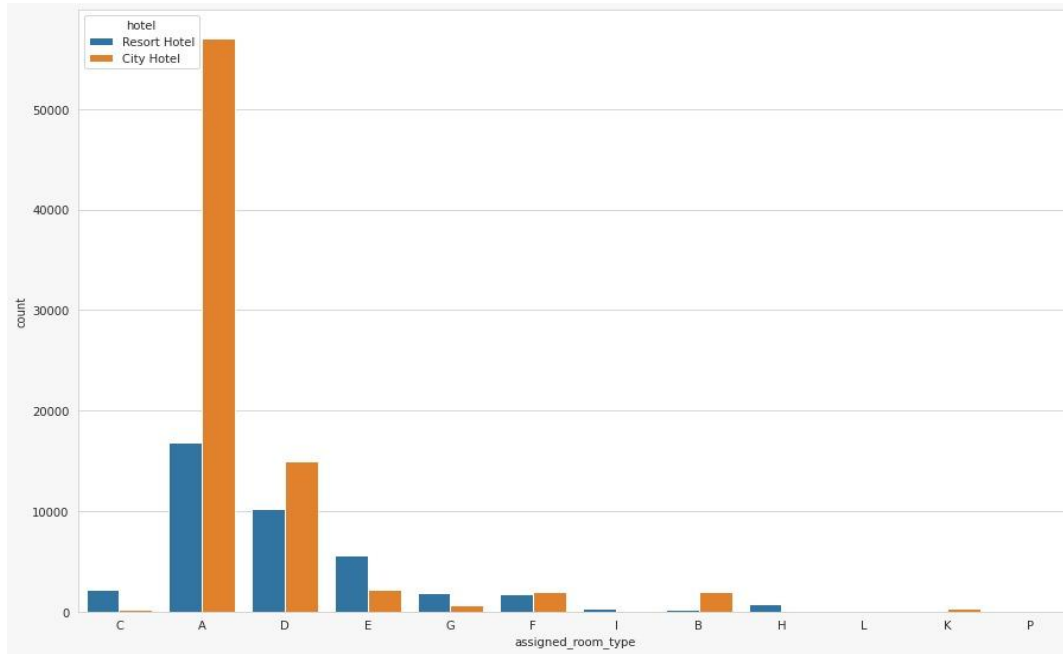


Reserved Room Type

- Type A rooms are most Reserved room type by the customers while booking.
- A rooms are most demanded by customers.



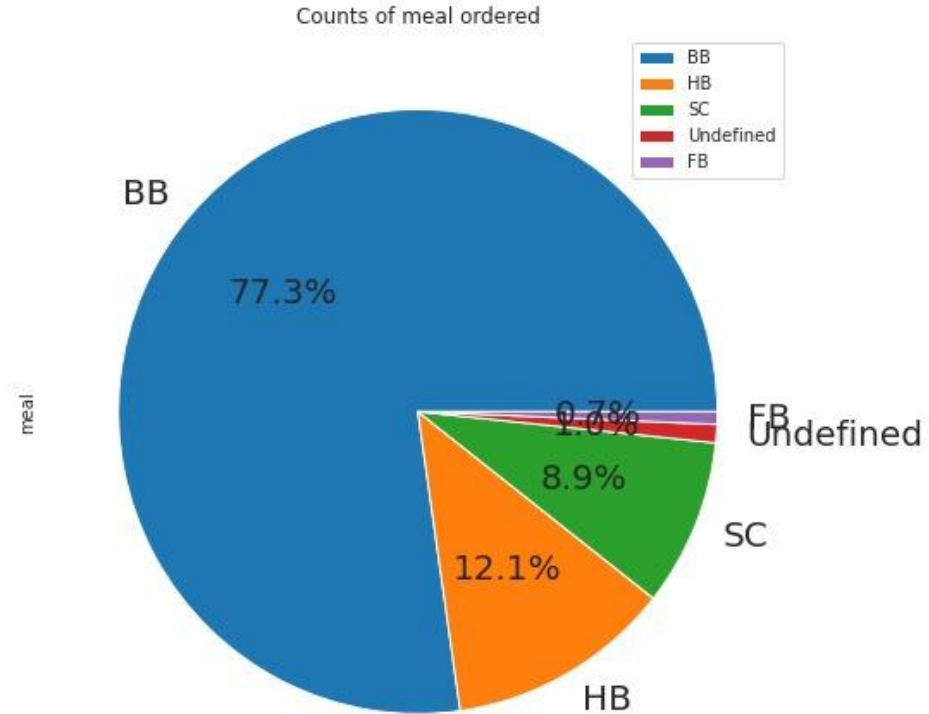
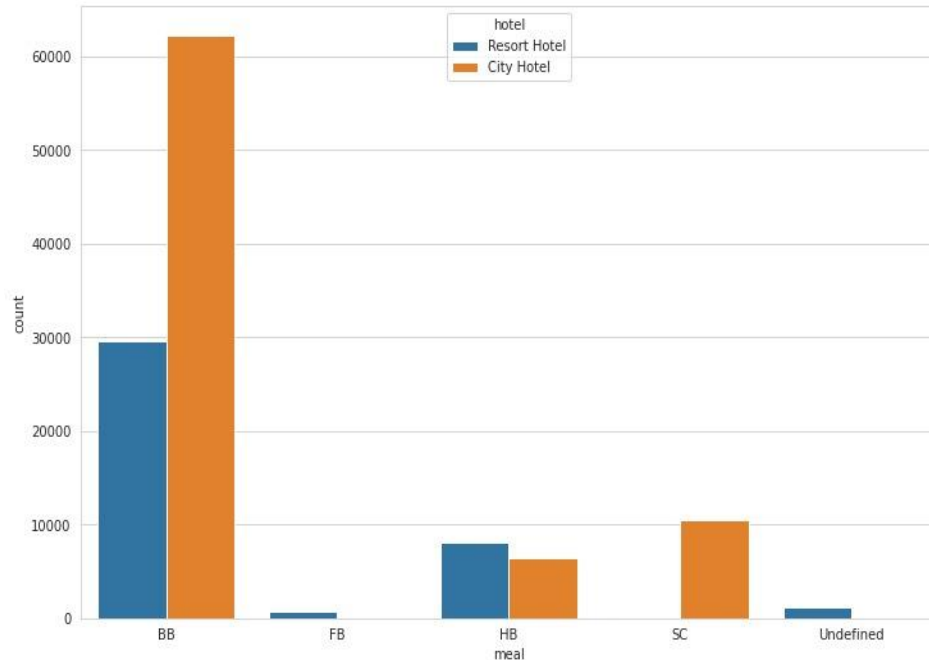
Assigned Room Type



Type A room are also most assigned room type in both the Resort and City hotel.
Type A room is 65.3%, D is 22.3%, E is 6.8%, F is 3.3%, G is 2.2% in assigned rooms.

Preferred Meal Type

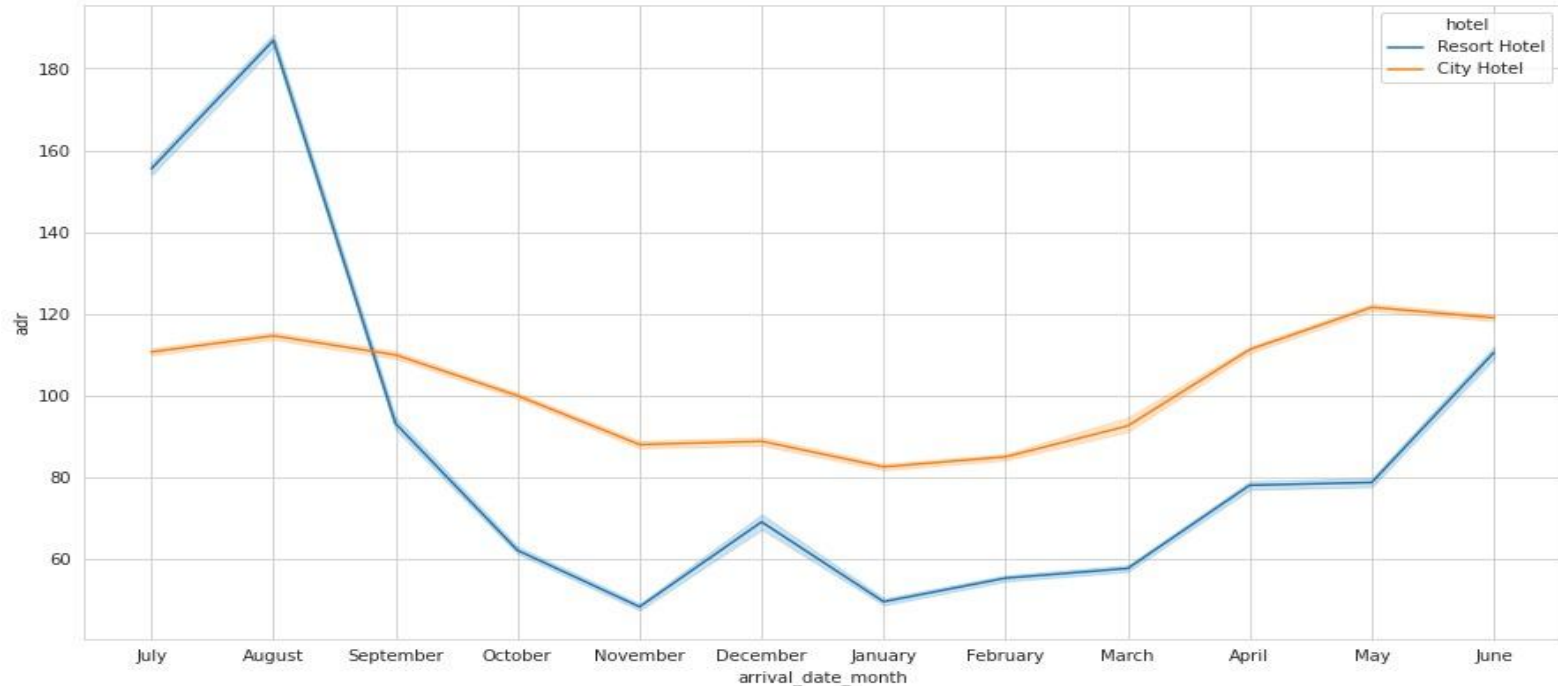
Around 77.3% of guests preferred BB (Bed and Breakfast).
HB (Half Board) is 12.1% and SC is 8.9% booked by guests.
In both the hotels BB is the most preferred meal type.



Monthly Price Analysis

For both city and resort hotel, November to January have cheaper average monthly rates.

This confirms with our monthly analysis as August has the highest number of bookings and November, December and January has the lowest number of bookings.

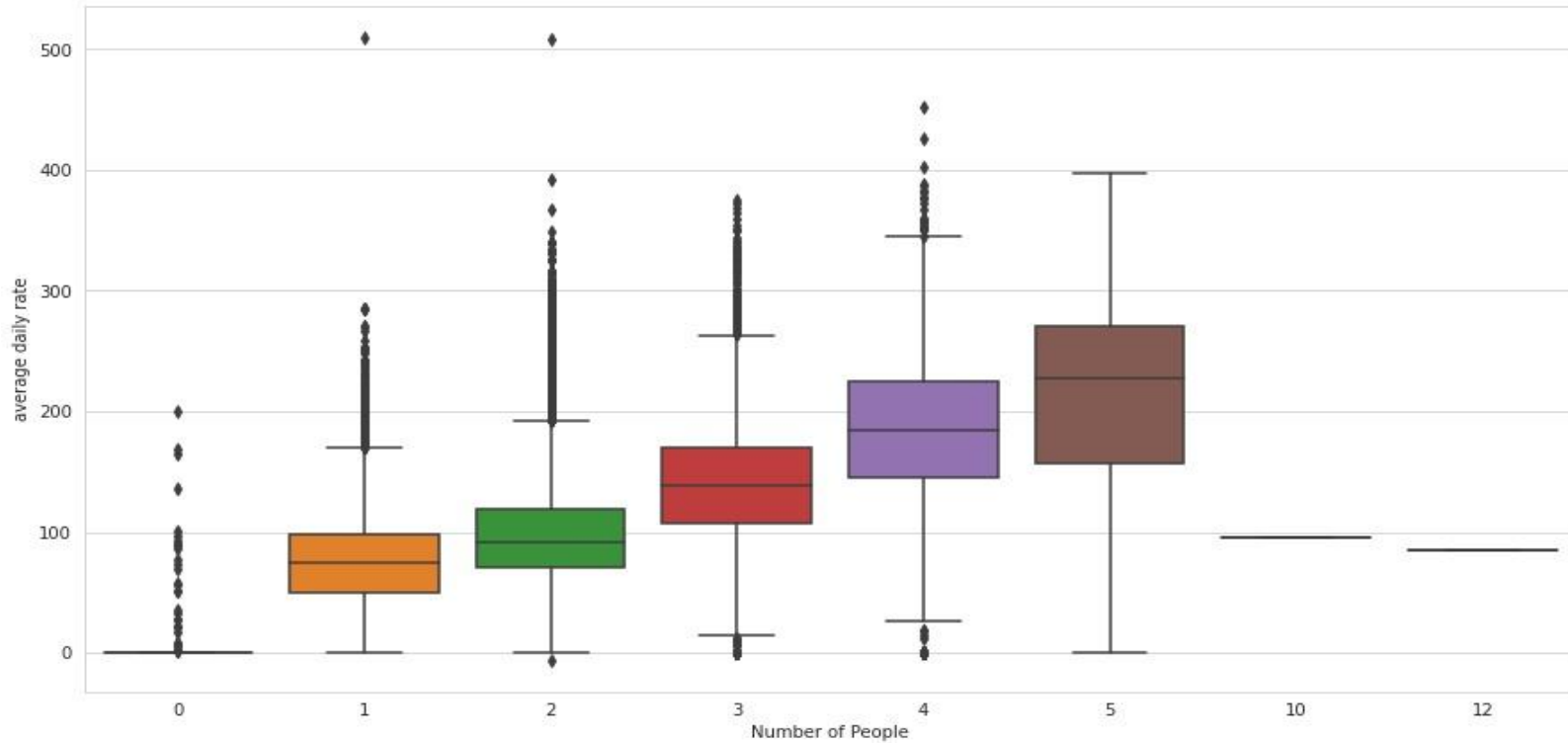


Correlation Heatmap

ADR is correlated with people as number of people increases earnings increase and so does ADR.
Total stay is correlated with lead time because people who stay longer plan ahead time.



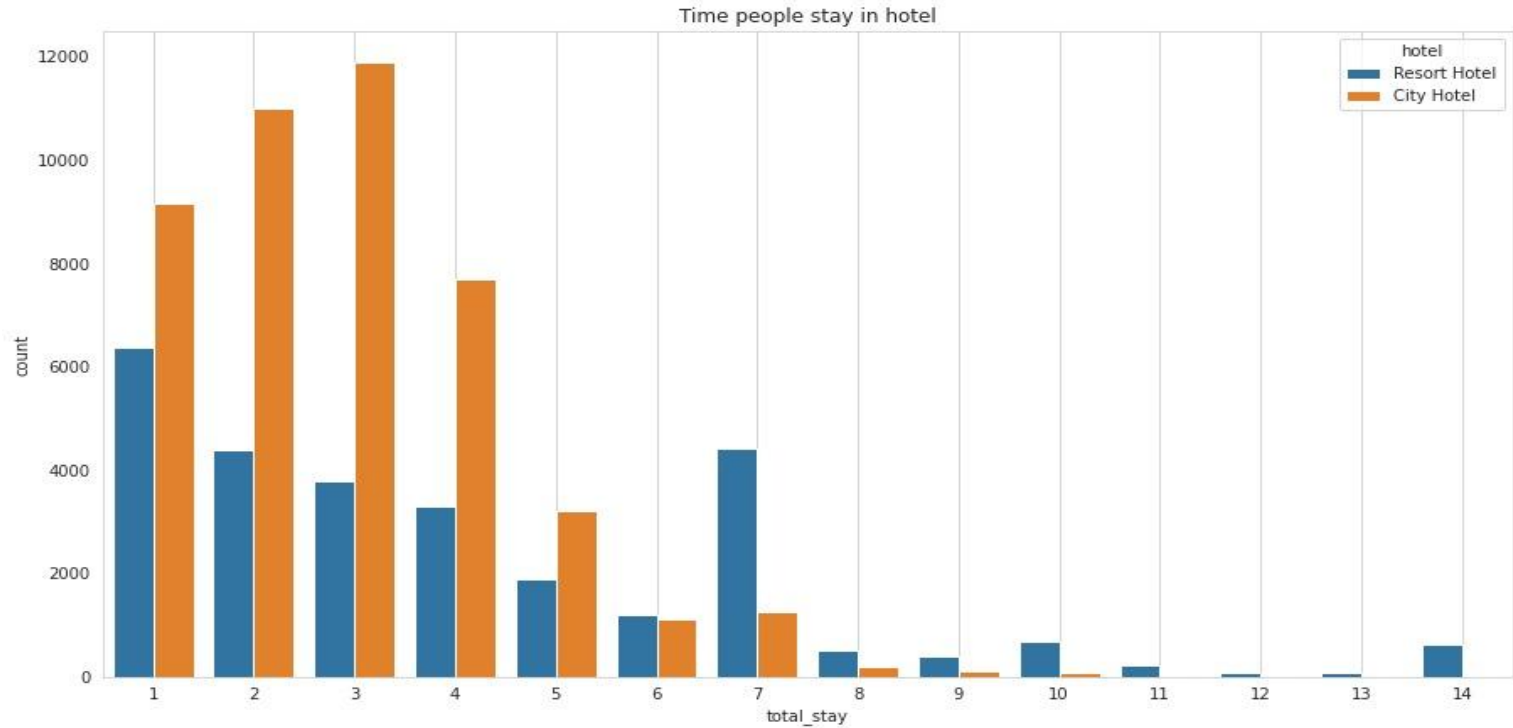
ADR Plot



For the plot, we can see that the median value is increasing as the population increases.

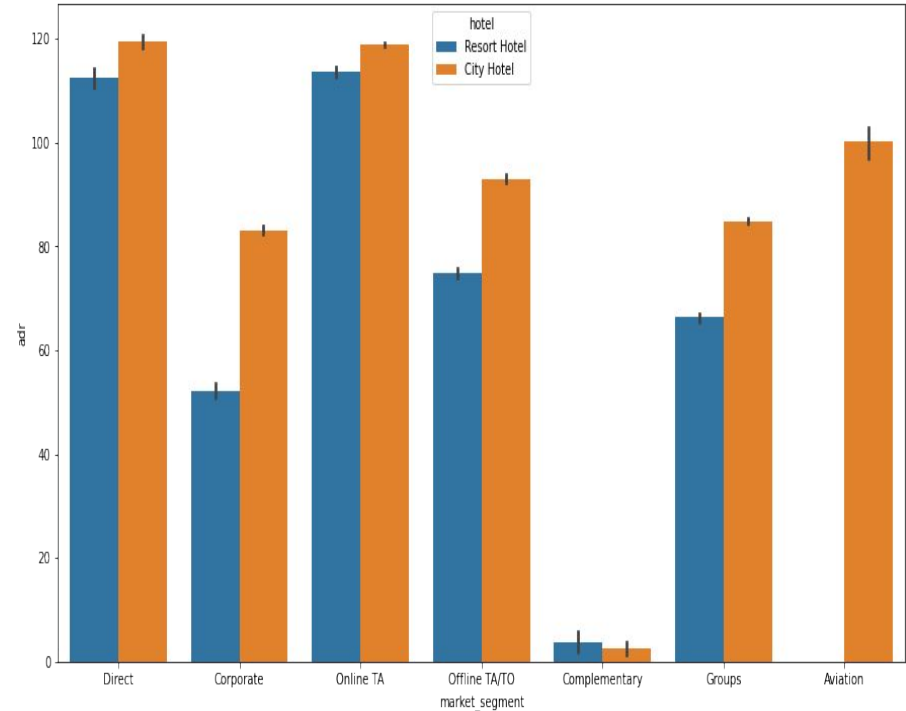
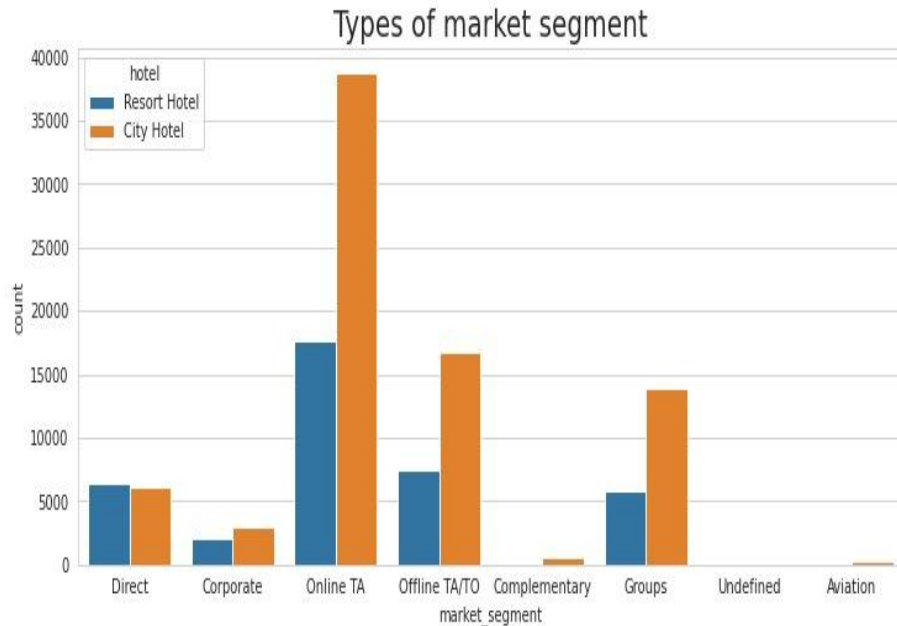
Total Stay

Most guests in city hotel on an average stay for 3 days.
Resort hotel shows a decrease in number of days except for 7 days.
Most guests in resort hotel on an average stay for a single day.



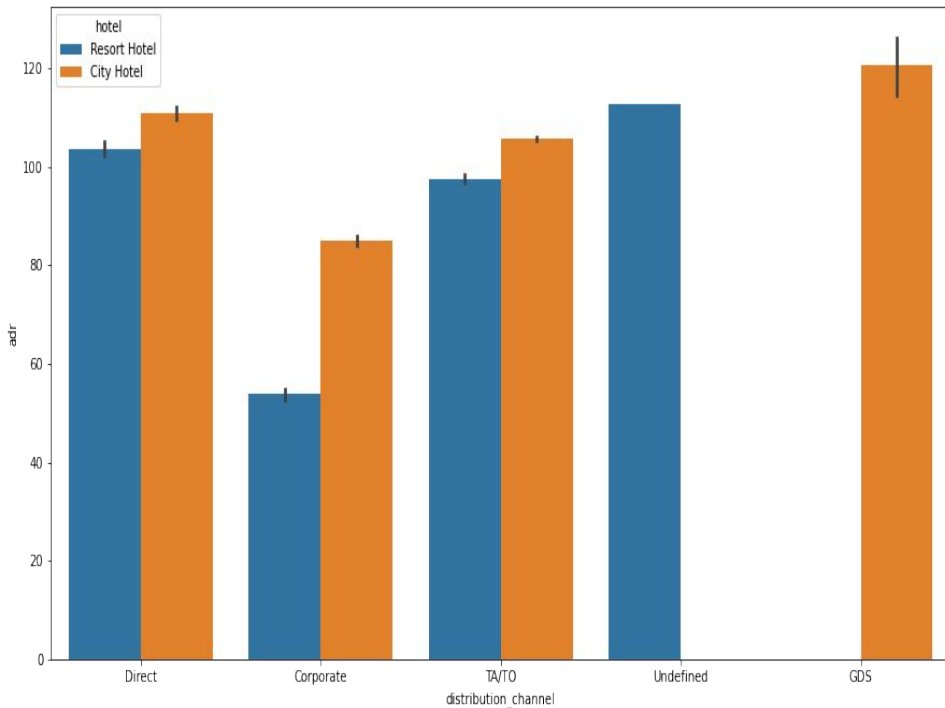
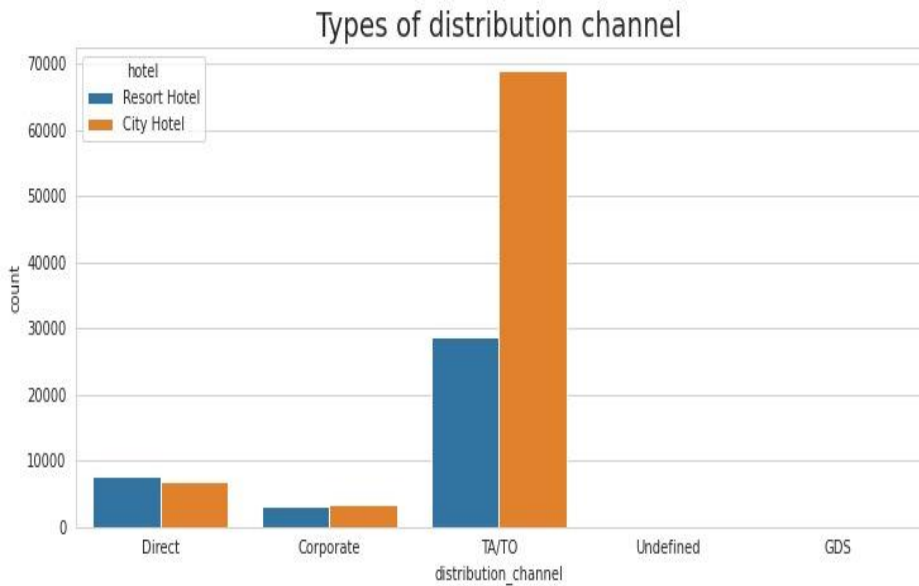
Market Segment Analysis

- The majority of market segment involved travel agencies(Online/Offline).
- In market segment there are 56,403 online TA and 24,160 Offline TA/TO number of bookings.
- We have to focus on targeting on these agencies website and work with them since majority of the visitors tend to reach out to them.



Distribution channel analysis

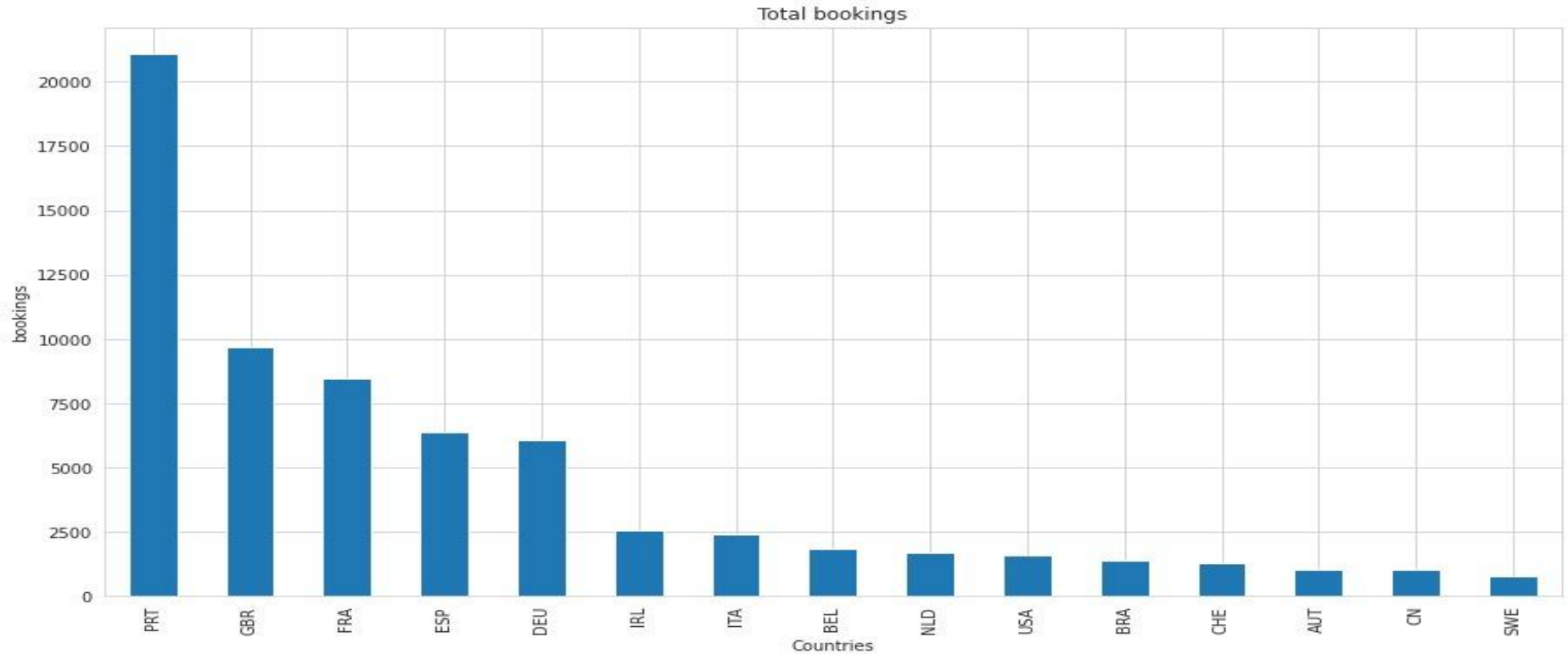
- The majority of distribution segment involved travel agencies(Online/Offline).
- In Distribution channel there are 97,730 TA/TO and 14,483 direct number of bookings.



Which country has the highest number of bookings



Highest number of bookings are from Countries: Portugal (PRT) followed by Great Britain (GBR), France (FRA) and Spain (ESP).



Conclusions

1. There are less number of feature with the missing values and outliers so dealing with them is easy.
2. After the peak in 2016 there is fall in bookings in year 2017. After 2015 bookings increased in 2016 then decreased in 2017.
3. City hotels has higher number of bookings so we have to focus on them and May to Aug period to be targeted as there is peak of summer period.
4. 1/3 of bookings where canceled Which is worse.

5. As booking rate for country PRT,GBR,FRA is high so we have to focus on advertisement in that country.
6. As average daily rate increases with increase in number of people so we have to encourage large group people/families to book by giving offers and discount.
7. We can say that more than 93% of customers not prefer for the parking, so we can commit that most of the customers reach the hotel by transport.
- 8.The hotel should be near stops of public transport
- 9.We should give less space for parking

10. From above we see that the majority of the distribution channels and market segments involve travel agencies (online or offline). We have to focus on targeting these travel agencies website and work with them since majority of the visitors tend to reach out to them.

11. For both city and resort hotels, November to January have cheaper average monthly rate

Thank You!