

Capstone Project

Hotel Booking(EDA)

Team Members:



Anannya Sagar Das
*AlmaBetter Data Science
Trainee*



Itisha Jain
*AlmaBetter Data Science
Trainee*

Exploring the dataset

```
#Looking at the first 5 rows  
df.head(5)
```

	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
0	Resort Hotel	0	342	2015	July	27	1	0	0
1	Resort Hotel	0	737	2015	July	27	1	0	0
2	Resort Hotel	0	7	2015	July	27	1	0	0
3	Resort Hotel	0	13	2015	July	27	1	0	0
4	Resort Hotel	0	14	2015	July	27	1	0	0

5 rows × 32 columns

Checking for data types and null values

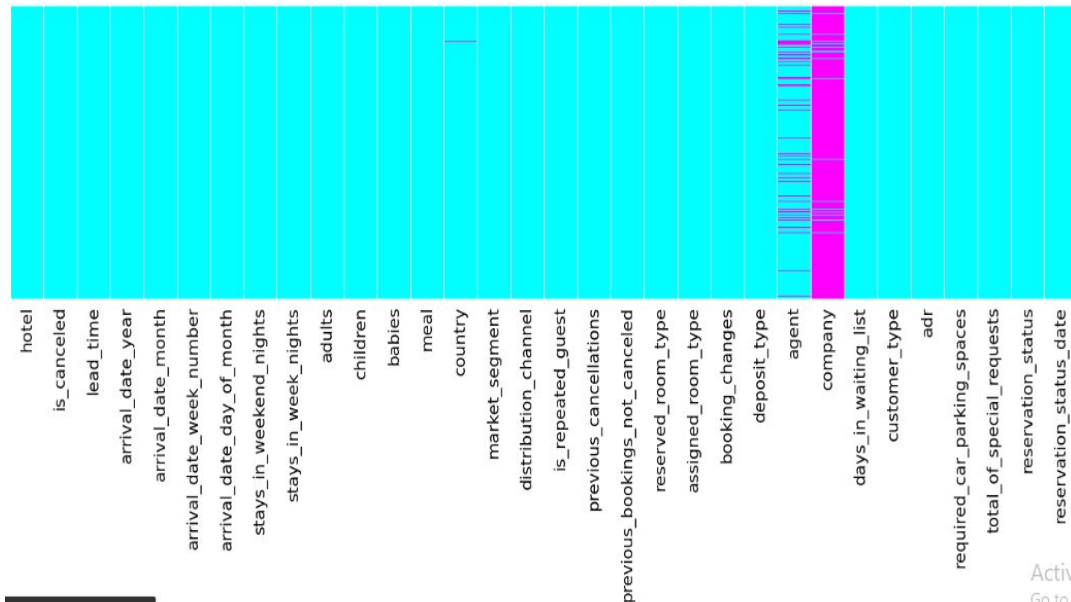
```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 119390 entries, 0 to 119389
```

```
Data columns (total 32 columns):
```

#	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

```
dtypes: float64(4), int64(16), object(12)
```



#treating nulls

```
df1.drop(['company'],axis=1,inplace=True)
df1.children.fillna(df.children.mode()[0],inplace=True)
df1.country.fillna(df.country.mode()[0],inplace=True)
df1.agent.fillna(0, inplace=True)
```

#Changing data type of Children and Agent column

```
df1[['children','agent']]=df1[['children','agent']].astype('int64')
```

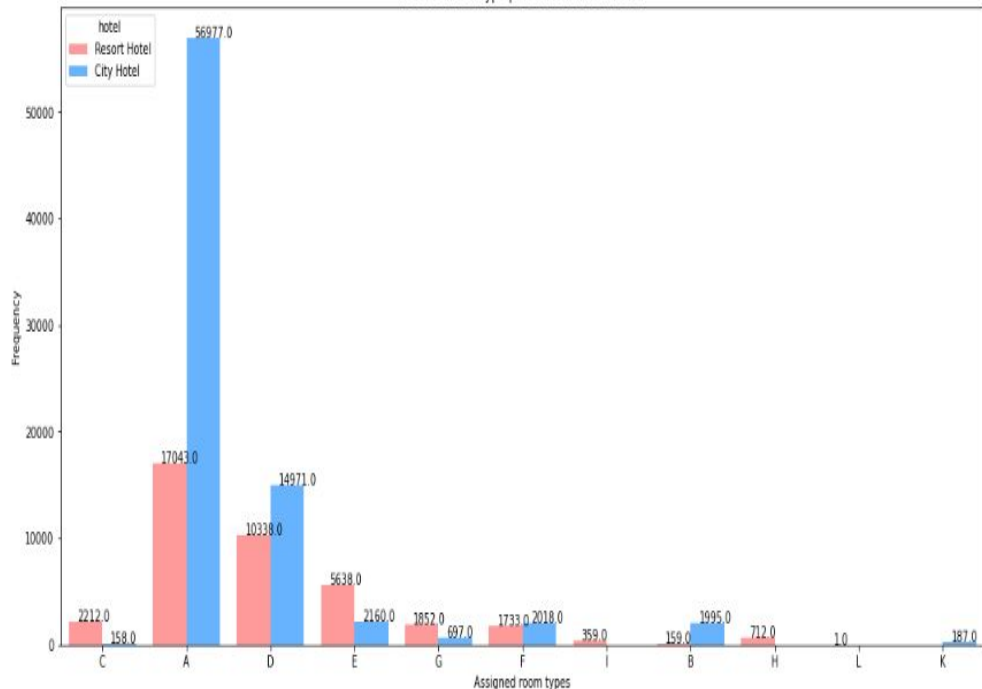
Hotel Type

Percentage of Resort Hotel vs City Hotel



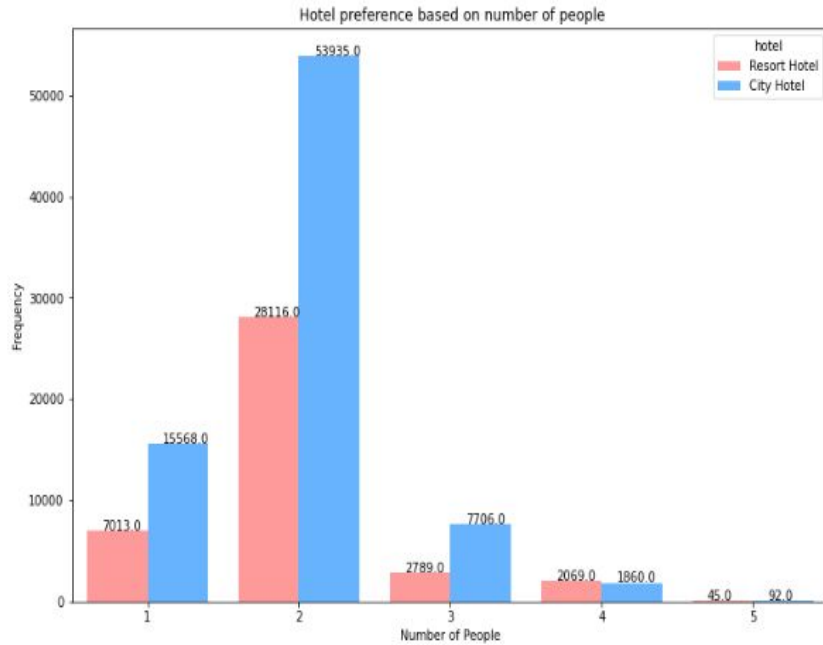
Bookings for City Hotel is more than Resort Hotel by 33%.

Accommodation type preference for both hotels

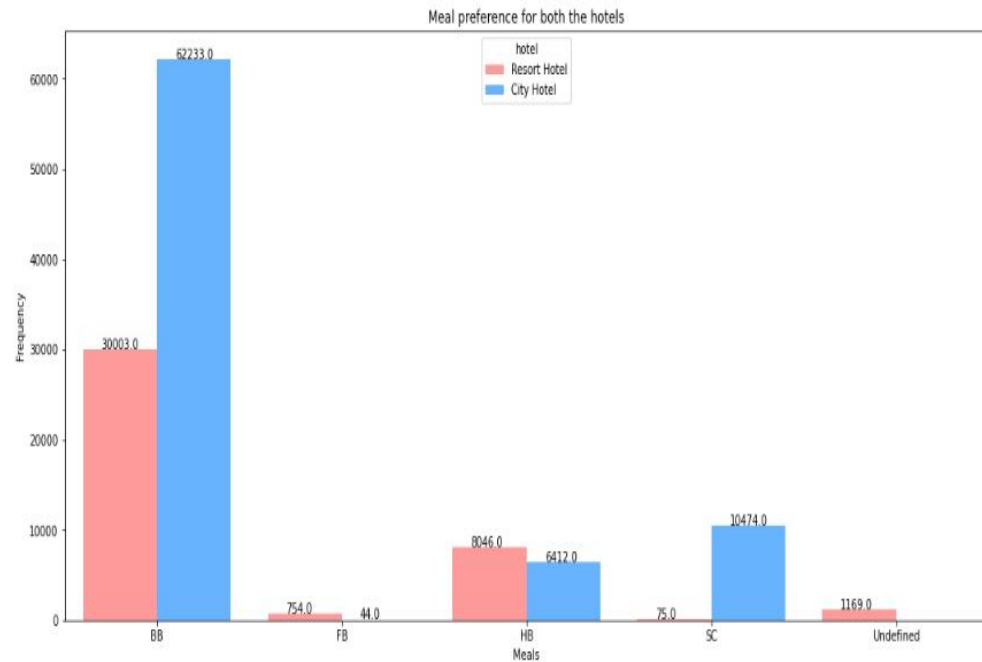


The most preferred room type is A, followed by D for both the hotels.

Hotel Type

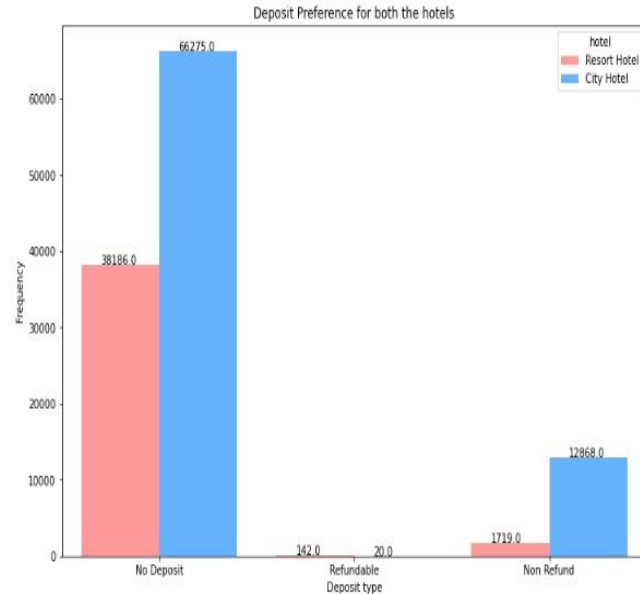
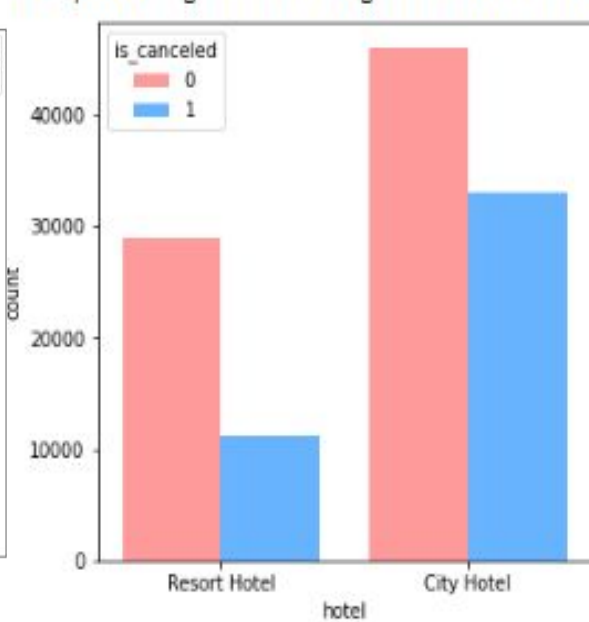
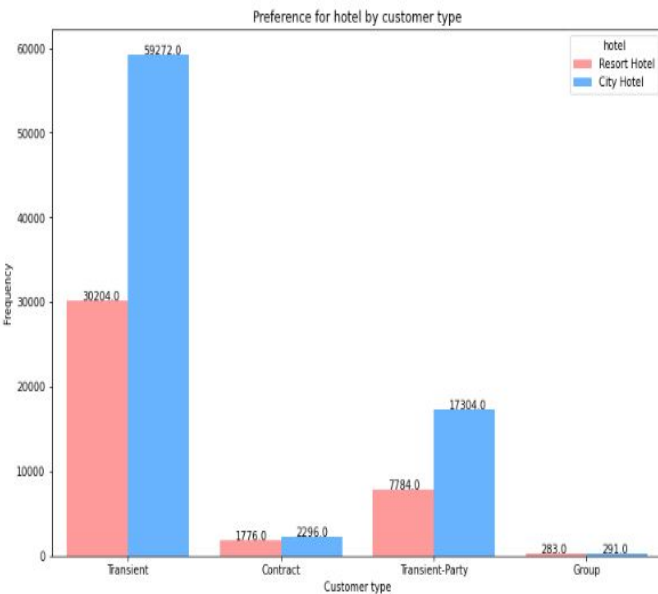


Most of the bookings were for two people, except for the group of four people, everyone prefers City Hotel more than Resort Hotel.



For both the hotels, BB (Bed and breakfast) is the most preferred meal.

Hotel Type



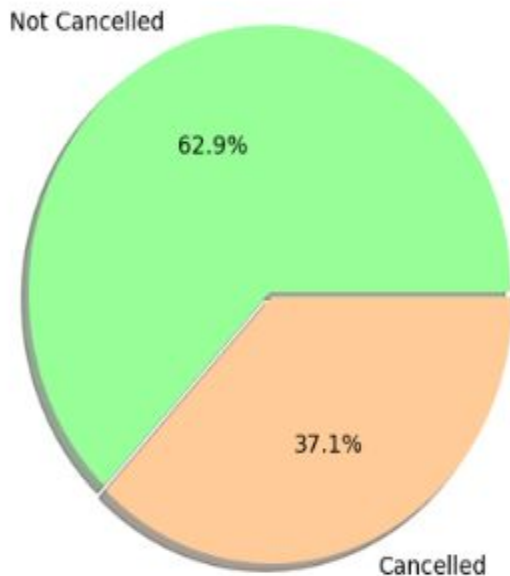
The majority of customers being transient and every customer type prefers city hotel over resort hotel.

City Hotel has cancellation rate of 41.79 % and Resort Hotel has a cancellation rate of 27.77%.

Customers do not prefer a deposit to be made while booking the hotel.

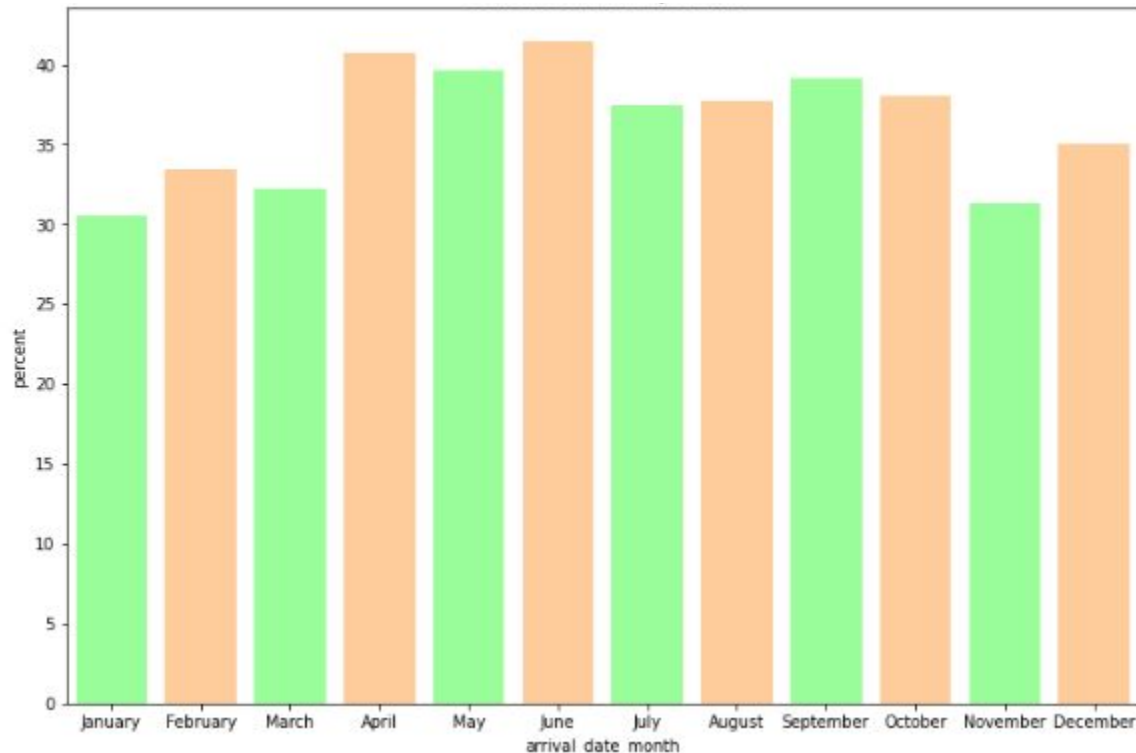
Cancellation

Proportion of Cancelled & Not Cancelled Bookings



Bookings got cancelled 37% of the time. While guests did checked-in (did not cancel the booking) almost 63% of the time.

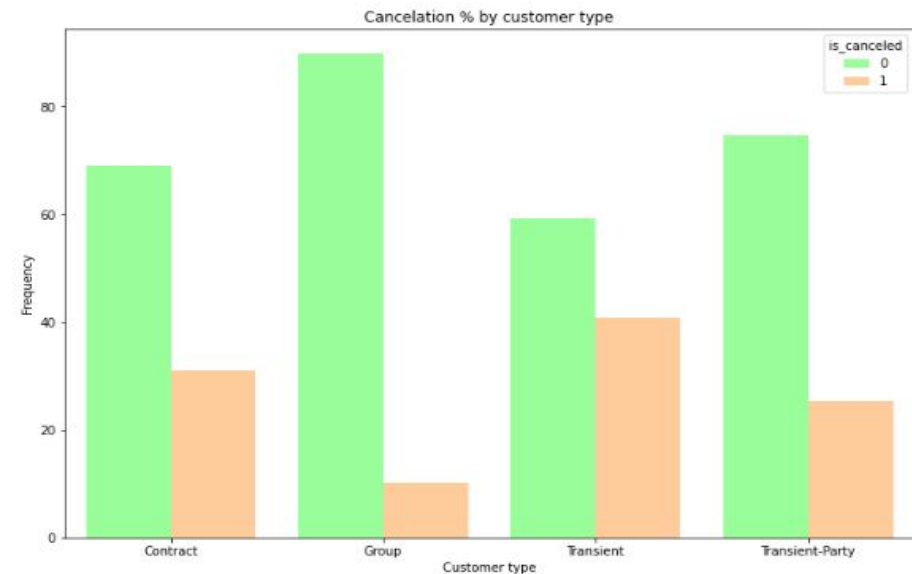
Cancellation ratio by month



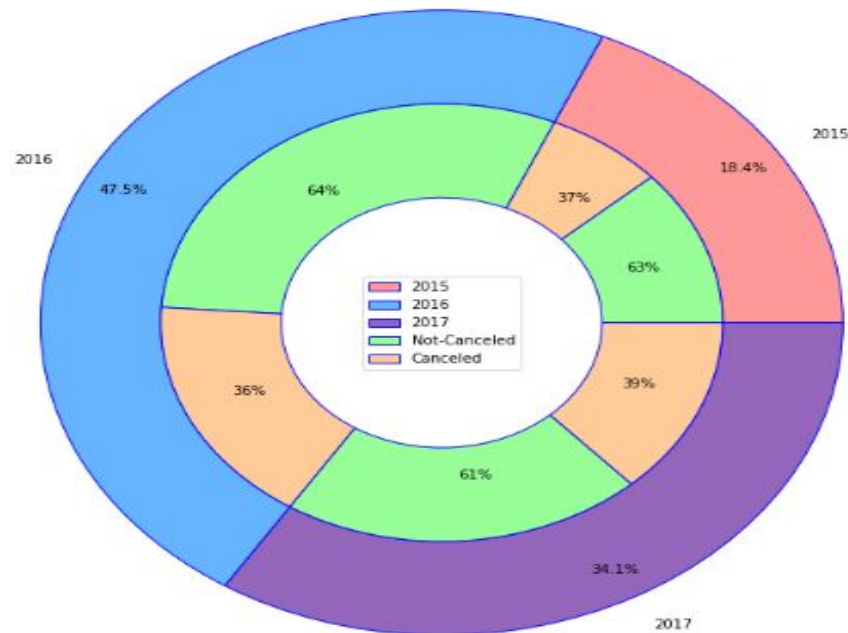
The Cancellation ratio is higher for the month of June followed by April and September and it is lowest for the month of January.

Cancellation

Yearwise bookings and Cancellation percentage*

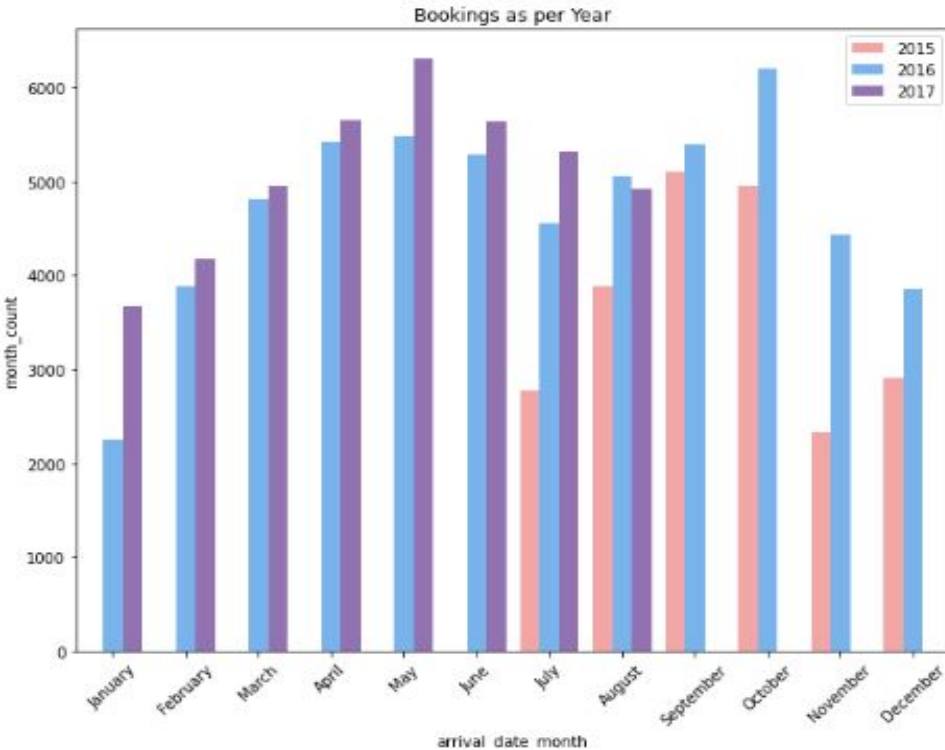


Transient customer types have the highest cancellation percentage.



Most of the data we have is for the year 2016 and the cancellation rate is ranging between 36%-39%.

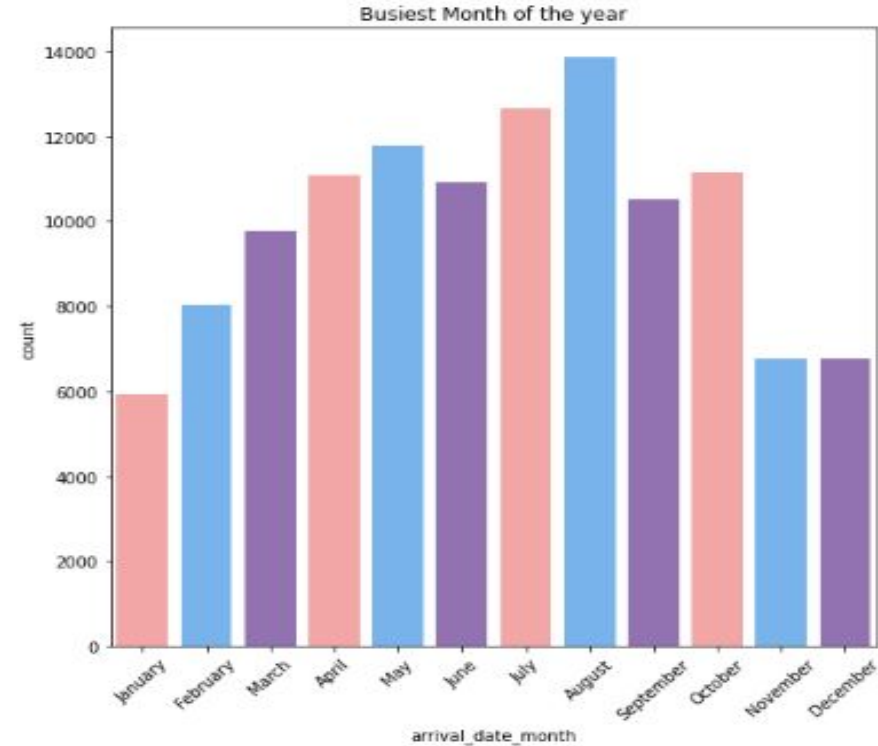
Bookings



For the year 2015, the busiest month for the year was September.

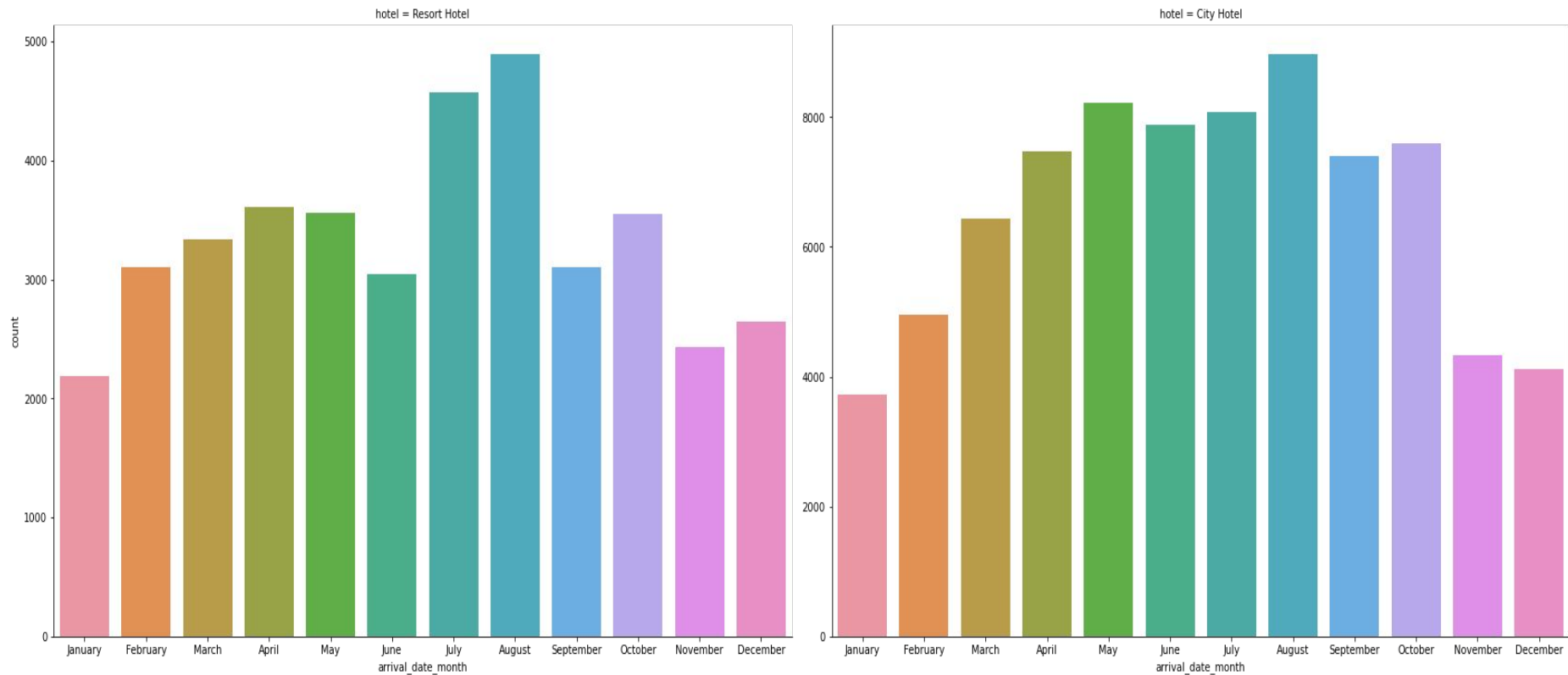
For the year 2016, the busiest month for the year was October.

For the year 2017, the busiest month for the year was May.



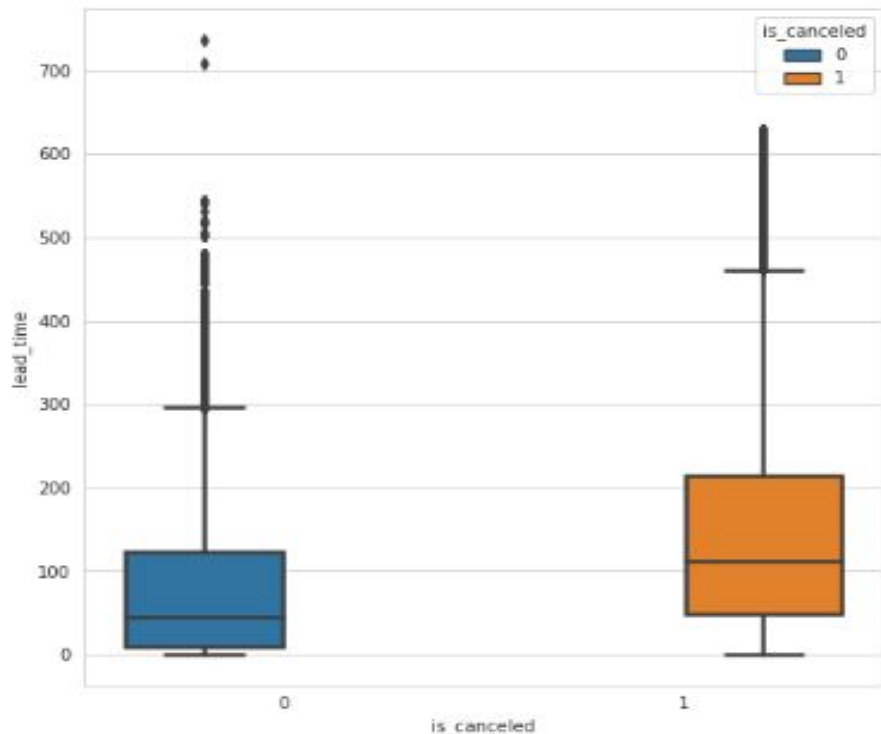
On an overall basis, the busiest month of the year is August and December is the most non-busiest month.

Bookings

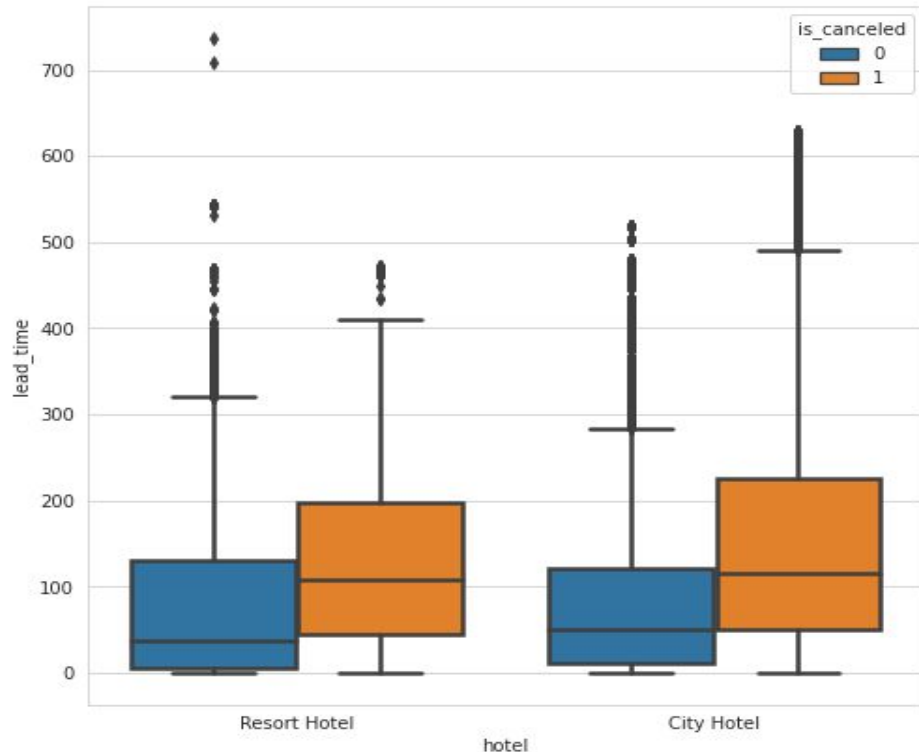


1. The busiest month for both the Hotels was August and in January there were the least bookings.
2. For both, the hotel's booking pattern was the same for all the months except for June and March.

Lead Time

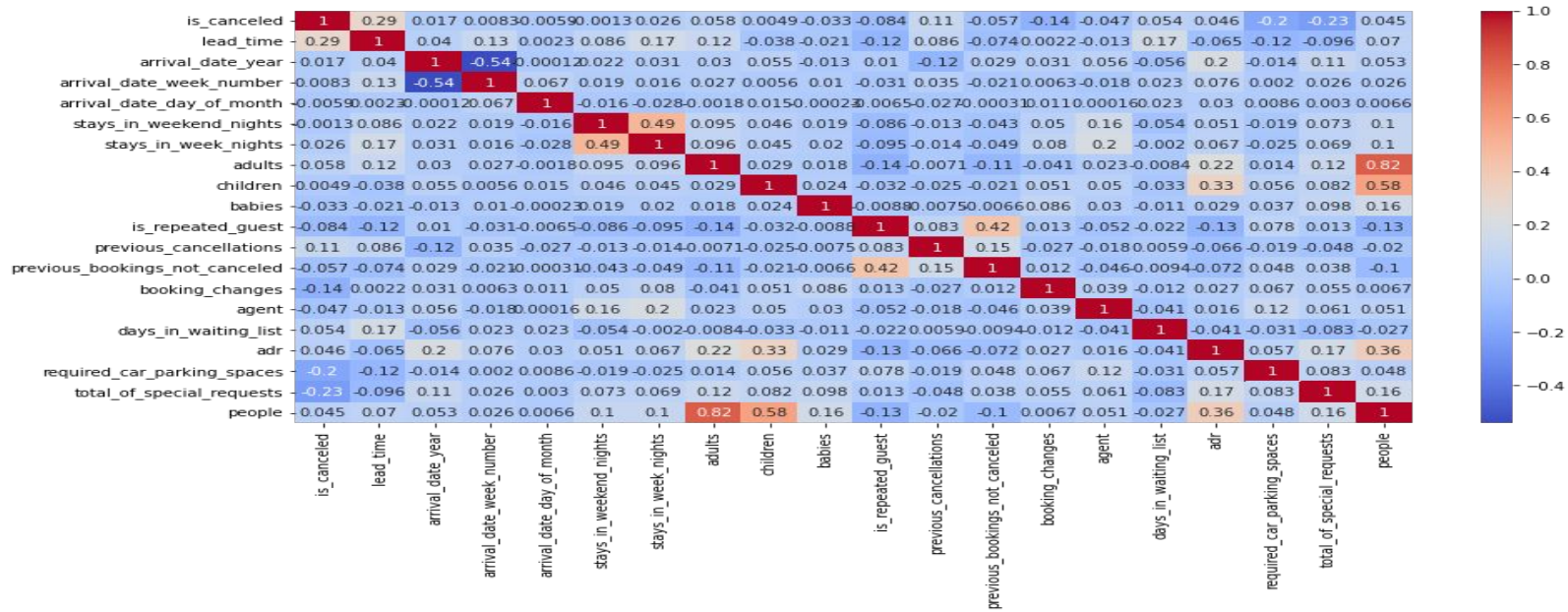


Cancelled Bookings had higher lead time.



Improvement in the lead time of bookings can lead to fewer cancellations for Resort Hotel as there lies much difference in cancelled and non-cancelled bookings.

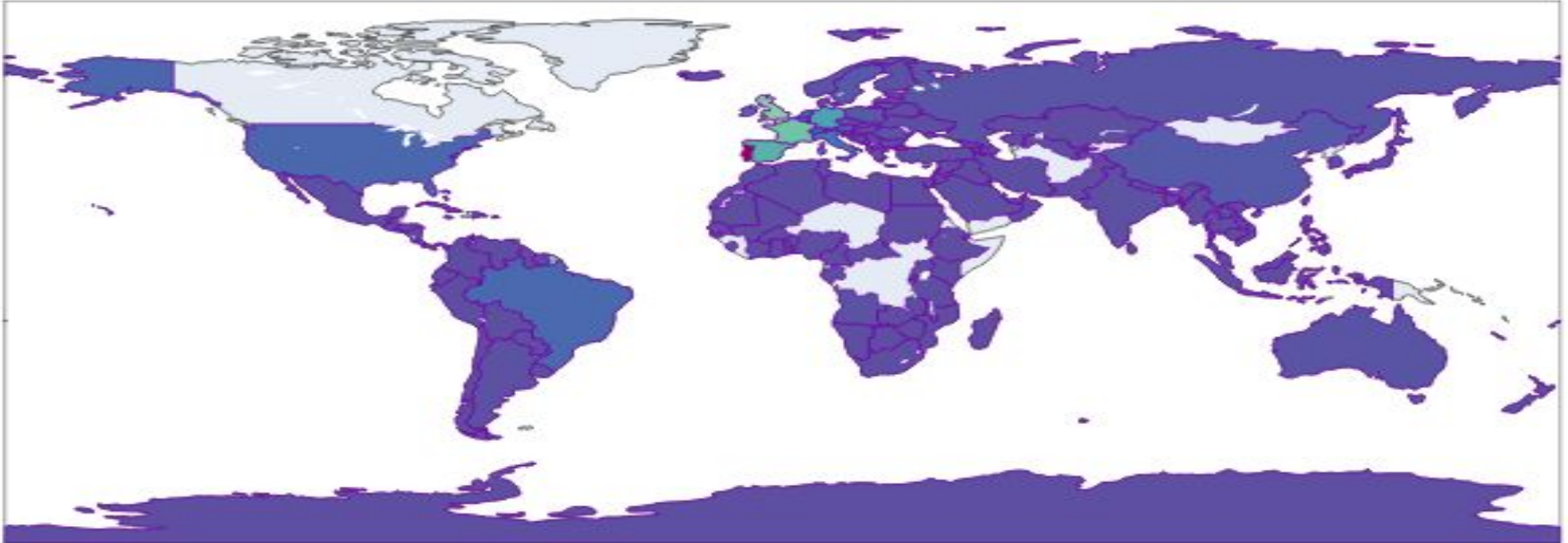
Correlation



1. stays_in_week_nights and stays_in_weekend_nights are positively correlated.
2. is_repeated_guest and previous_bookings_not_canceled are positively correlated. May be repeated guests are not more likely to cancel their bookings.
3. total_of_special_requests and is_canceled are negatively correlated implying higher the number of special request lower the chances of cancellations.
4. Adults and Children are positively correlated to ADR that means more the people more will be the ADR.
5. arrival_date_week_number and arrival_date_year are negatively correlated.

Country

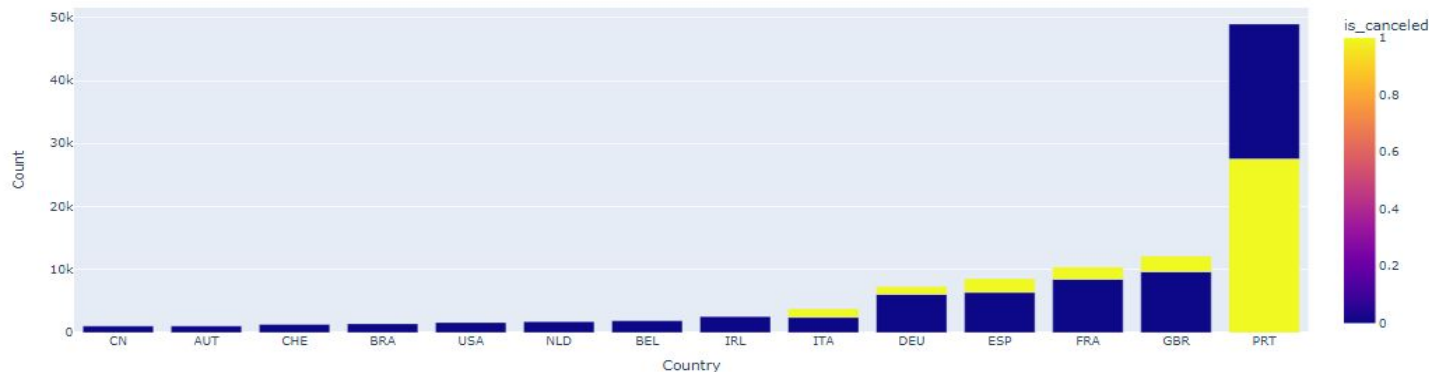
Number of Records by Countries



Most guests come from Portugal(PTR), Great Britain(GBR), France(FRA), Spain(ESP), Germany(DEU), Italy(ITA), Ireland(IRL), Belgium(BEL).

Country

Number of Cancellations by Country



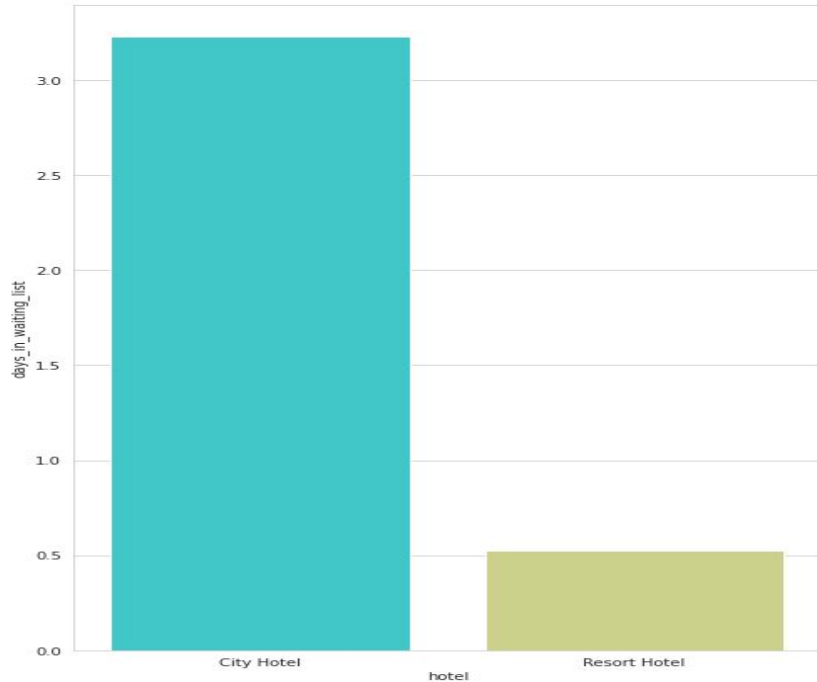
The cancellation rate in Portugal is around 60% and in other countries, it's very minimal.

Hotel type by Country

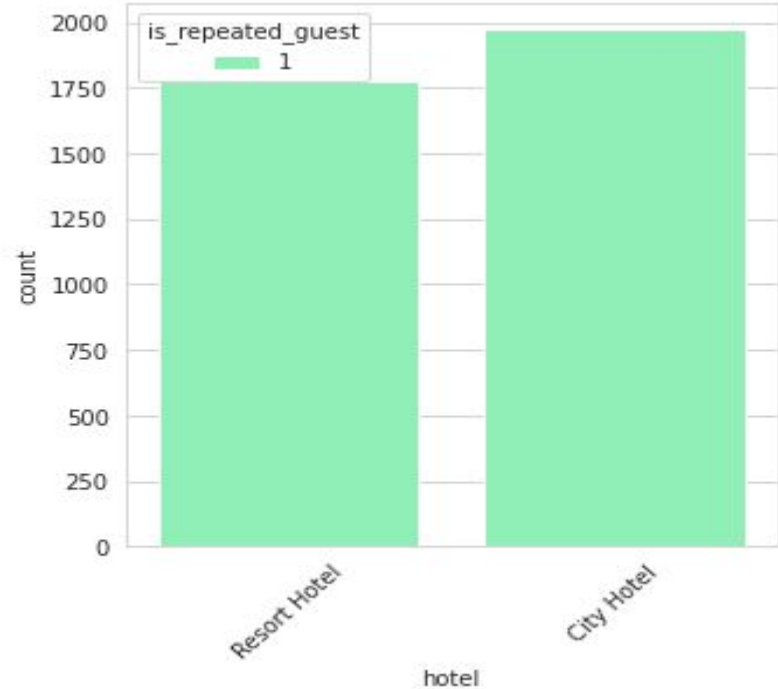


All countries prefer the City Hotel over the Resort hotel except Great Britain.

Hotel type

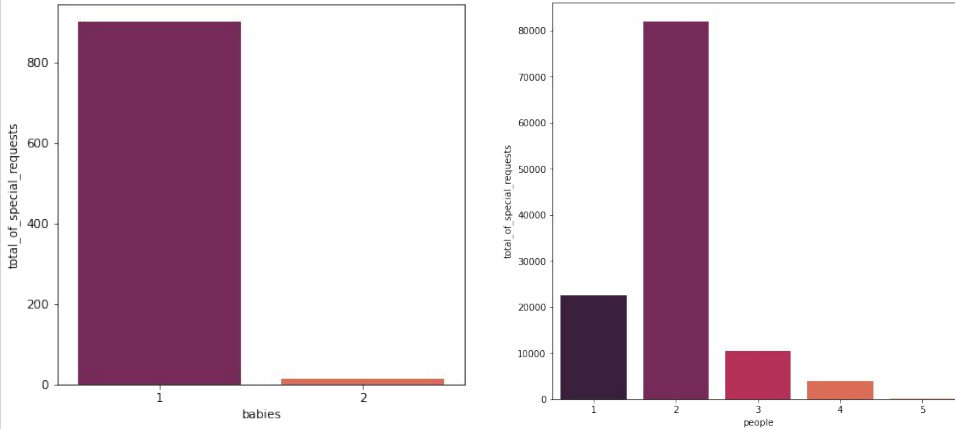


So the City Hotels has longer waiting period than the Resort Hotels. Thus we can say that City Hotels are much busier than the Resort Hotels



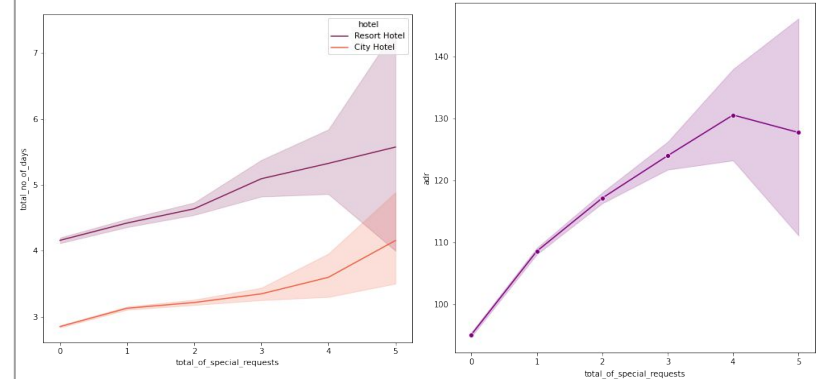
City Hotel has more repeated guest than Resort Hotel.
Guests prefer City Hotel over Resort Hotel.

Special Requests

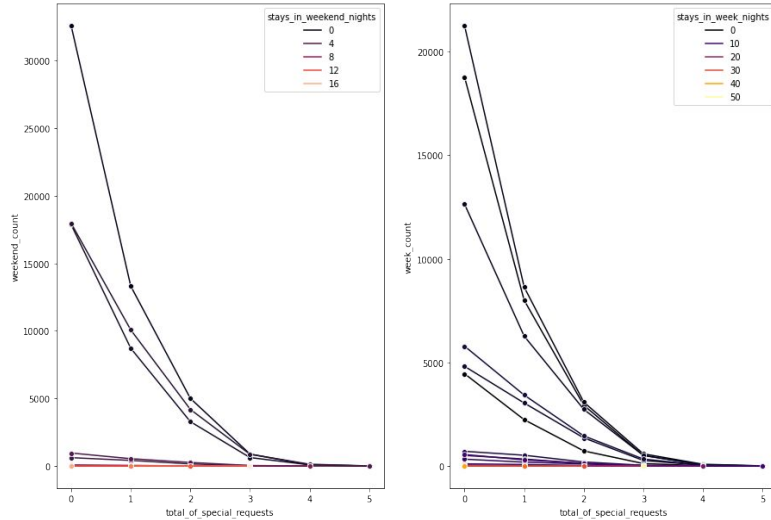


1. From the above graph it can be inferred that guests in groups of 2 have the most number of requests. And guests with only one baby make more special requests.

2. With the increase in the number of stays the special requests tend to increase. And the average daily rate increases with more special requests.

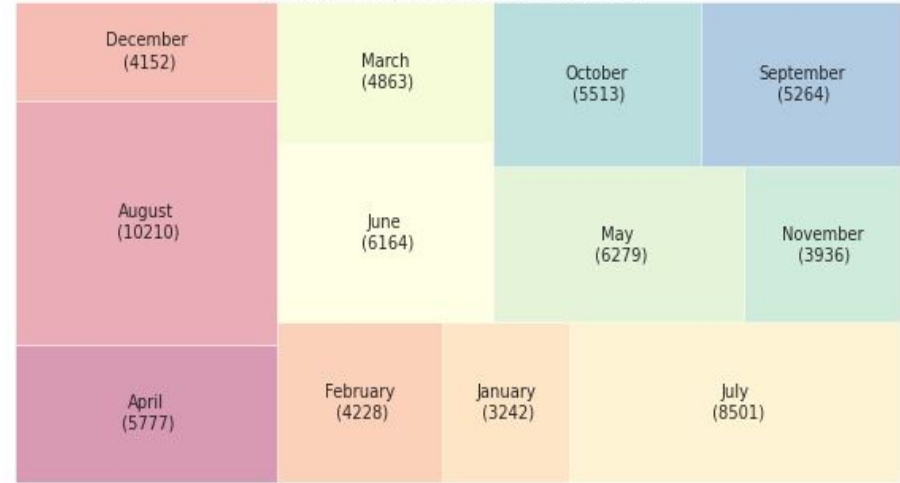


Special Requests



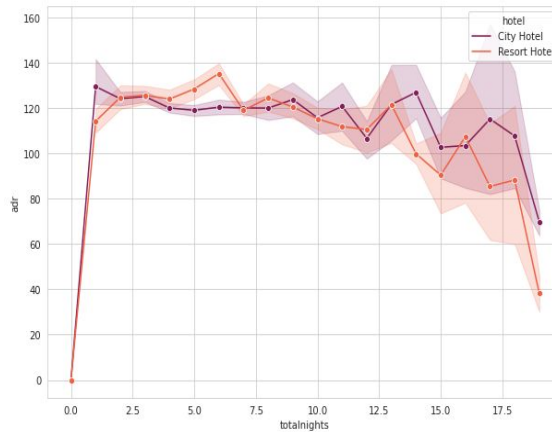
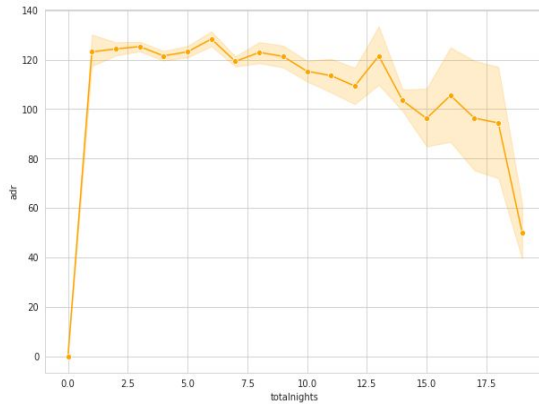
3. The trend is same for weekends as well as weekdays i.e the number of special requests rise with increase in number of stays.

Monthly Special Requests over the Years

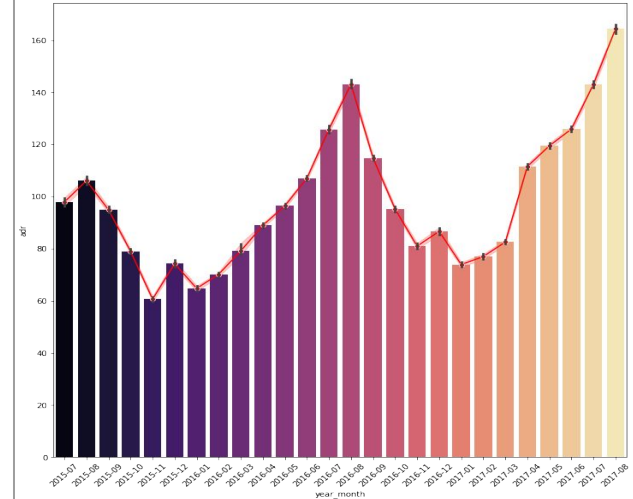


4. From the above treemap, it is visible that the highest number of requests is for the month of August followed by July, May, June, April and so on.

Average Daily Rate

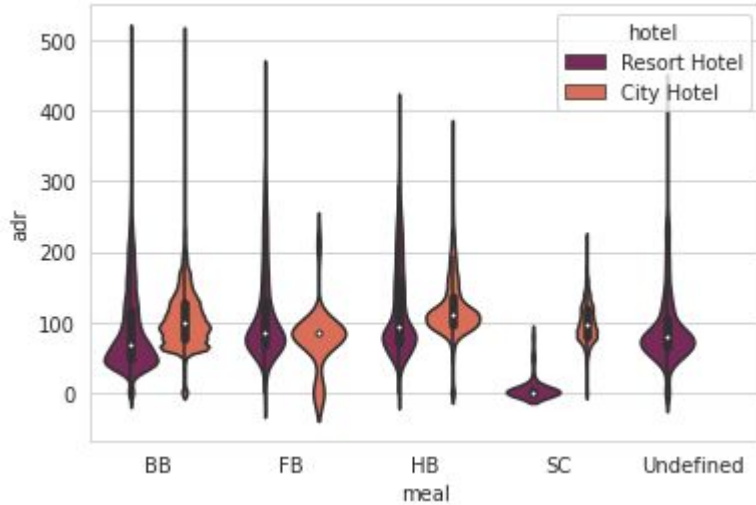


1. From above plot we can say that as the stay increases adr is decreasing. Thus for longer stays customer can get good adr .For stays longer than 13 days Resort Hotel is cheap.

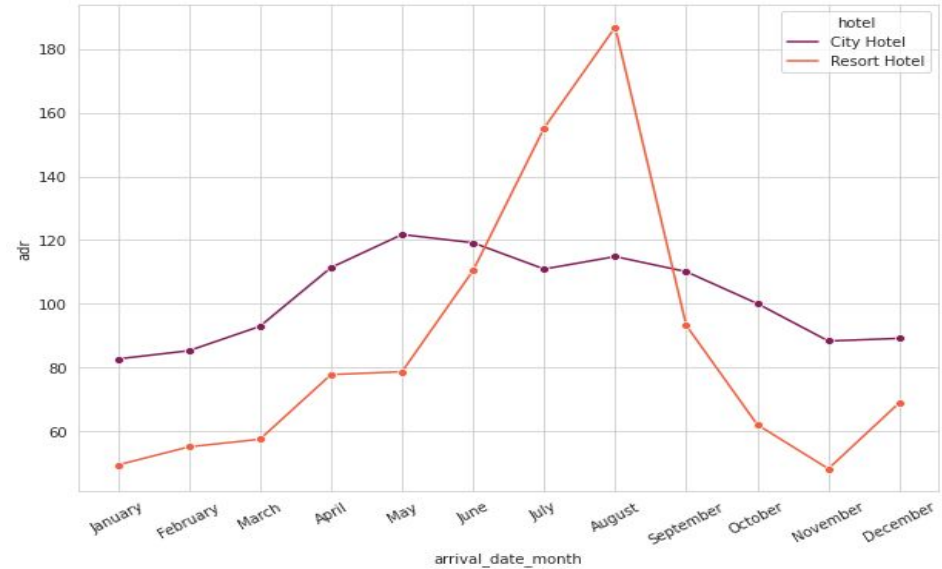


2. If we look at the monthly pattern for all three years, it is observed that August 2017 recorded the highest average daily rate

Average Daily Rate



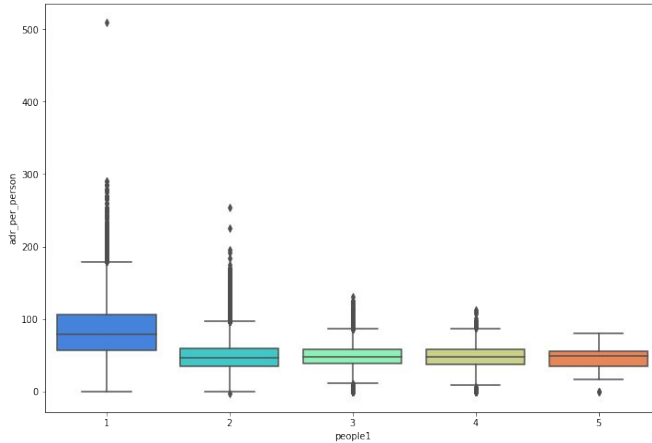
4. The highest adr was for HB meal types for both the hotels.



5. For Resort hotel, ADR is high in the months June, July, August as compared to City Hotels. May be Customers/People want to spend their Summer vacation in Resort Hotels.

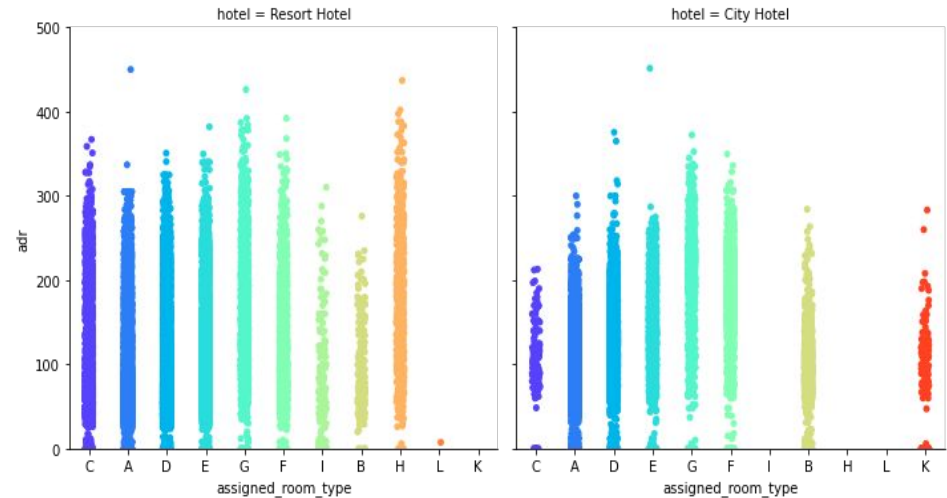
The best time for guests to visit Resort or City hotels is January, February, March, April, October, November and December as the average daily rate during these months are very low.

Average Daily Rate

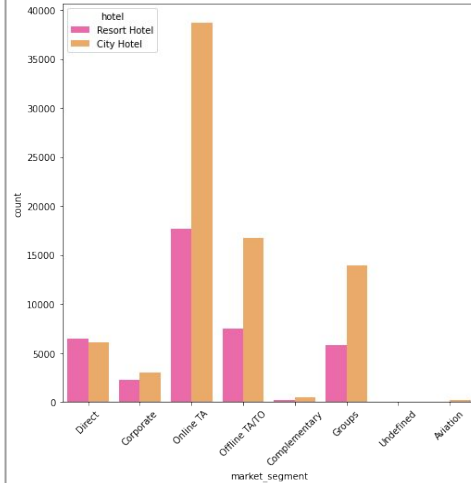


6.The highest average daily rate per person was for bookings with only one guest. And for bookings with more than one guest the average daily rate was comparatively lower.

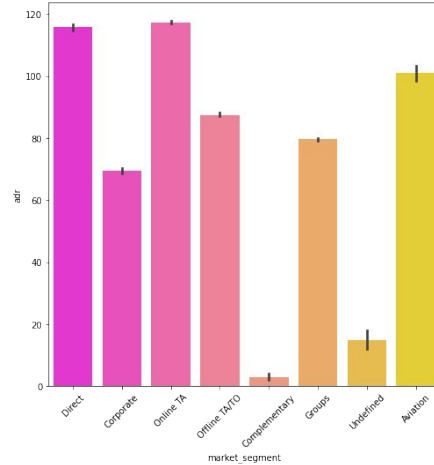
7.In case of City Hotel room types K,L,I have the lowest adr whereas for Resort Hotel room types I,H,L have the lowest.



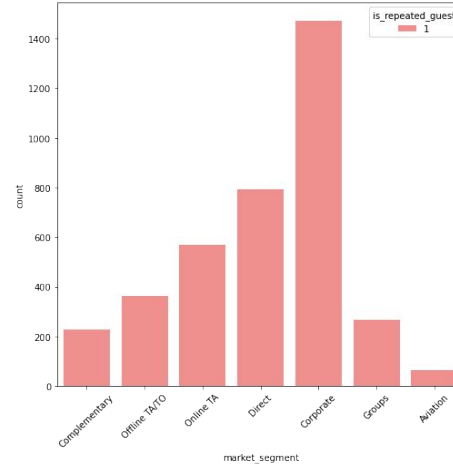
Market Segment



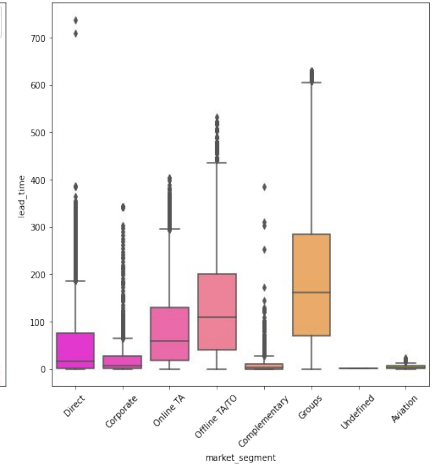
1. The most number of bookings were via Online TA of the market segment for both hotel types.



2. And the lowest adr was for bookings via complementary type whereas Direct, Online TA, aviation was high

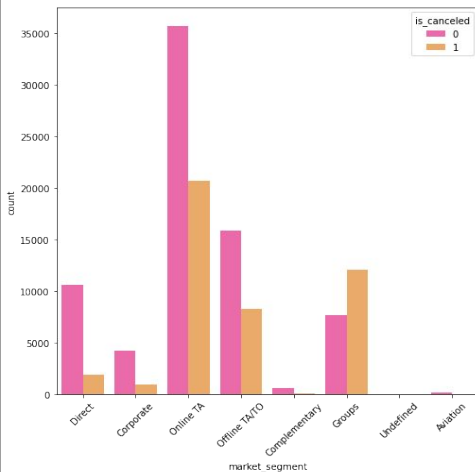


3. The guests who repeated their bookings chose to book via the Corporate segment, Direct and Online TA.

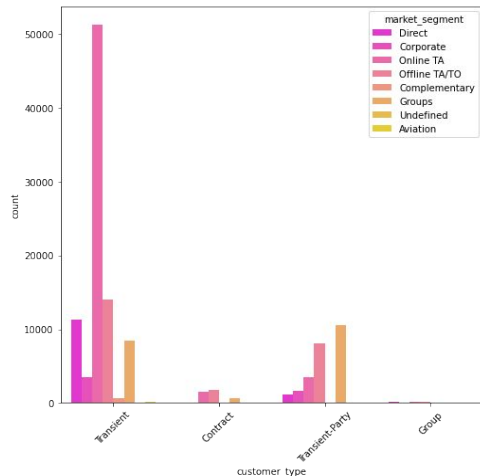


4. The lead time for the Group segments was the highest.

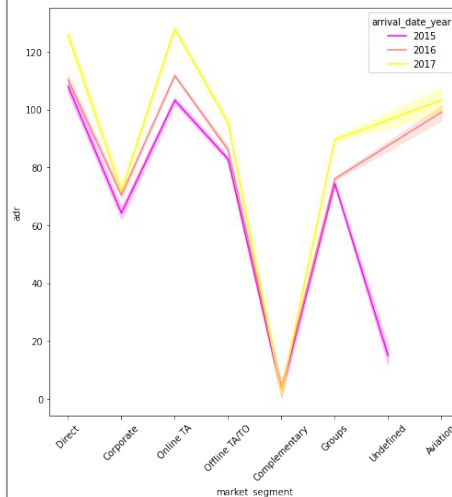
Market Segment



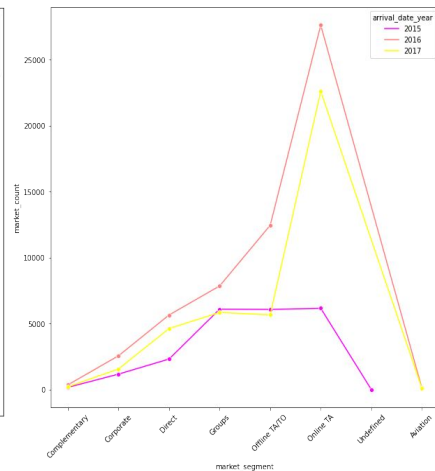
5.The most number of cancellations were for the Online TA segment.



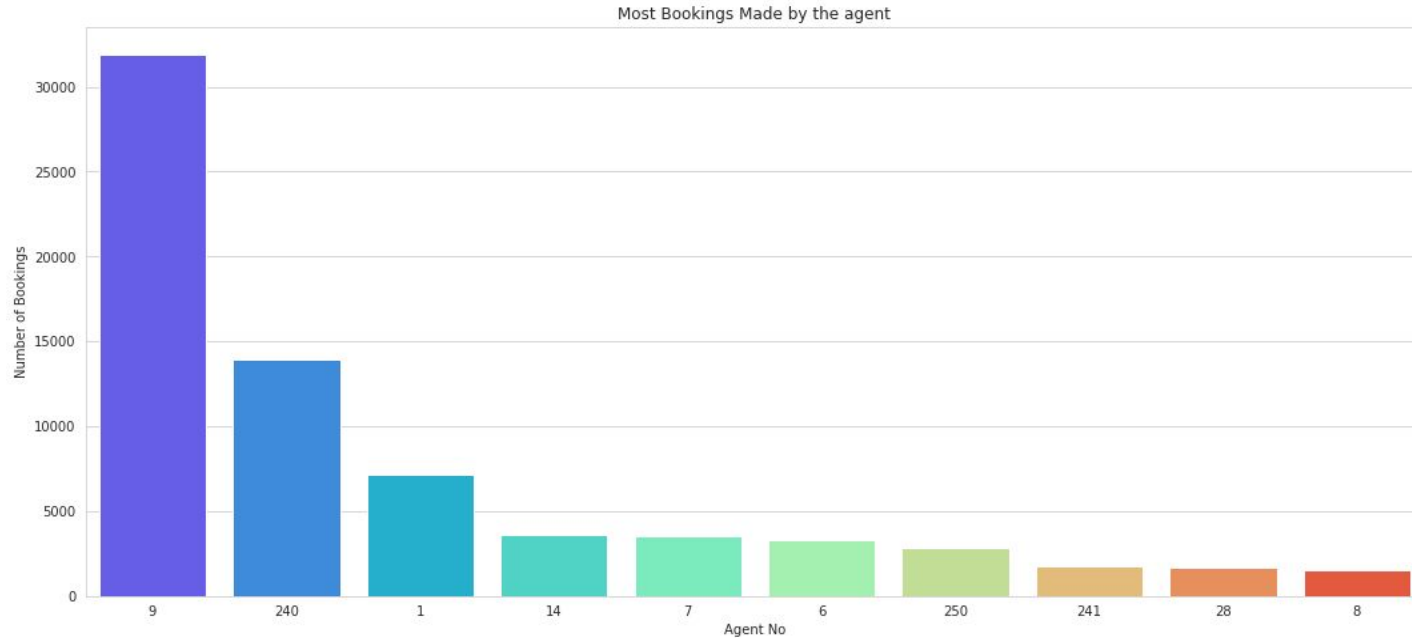
6.The most number of bookings were of Transient customer types for the Online TA market segment.



7.We observed that from 2016 there was a rise in bookings via Online TA.



Agent



Agent ID No.9 made most of the bookings followed by Agent ID No. 240, Agent ID No.1 and so on.

Conclusion

Milestones

- ❖ The best time to book a hotel is during the months from October to February because during this time period the average daily rate is comparatively lower compared to other months. Additionally, guests should select room types I, L and make the bookings via Corporate, Groups and if possible use complementary channels. Further for best deals the bookings should be made for more than one guests.
- ❖ The optimal length of stay in order to get the best daily rate is for one day or more than seven days and for bookings with more than one guest.
- ❖ The highest amount of special request is received for the month of August followed by the months of July, May, June, April and so on. These request are mainly from guests coming in groups of two (mostly couples) or from those who has a baby with them. Also, the special requests are high for guests with longer stays.

Further Inferences

1. For each year there were more bookings for City Hotel than Resort Hotel.
2. As bookings, Cancellation ratio was higher for City Hotel.
3. For both the hotels "A" room type was most preferred room followed by "D".
4. For both the hotel types, BB (Bed and Breakfast) was the most preferred meal.
5. Non Refundable bookings are really less in Resort Hotel as compared to City Hotel.
6. All of the customer types prefer City hotels over Resort Hotels.
7. The most number of bookings were for 2 people and based on the number of people who come as a family for the stay, all prefer City Hotel over Resort Hotel except when the number of people is 4.
8. Bookings got canceled 36% of the time. While booking guest did checked-in (did not cancel the booking) almost 64% of the time.
9. Transient customer types have the highest cancellation percentage.
10. Cancellation ratio is higher for the month of June followed by April and September and it is lowest for the month of January.
11. Most of the data we have is for year 2016 and cancellation rate is ranging between 36%-39% .
12. For 2015-September, 2016-October, 2017- May was the busiest month of the year. On overall basis busiest month of the year is August and December is the most non-busiest month.
13. Bookings which got cancelled had higher lead time.
14. Optimal length of the stay for both hotel is 7 days.
15. Most of the bookings were from Portugal(PTR), Great Britain(GBR), France(FRA), Spain(ESP), Germany(DEU), Italy(ITA), Ireland(IRL), Belgium(BEL) but we can also see a higher cancellation rate of bookings from Portugal around 6% .

THANK YOU