Capstone Project - 1
Hotel Booking Analysis

by-

Team GodSpeed:

Ranjith K

Shreedarsh M

Sachin S Panchal

Rohan A G

Kshipra Parihar

Contents :

AI

- **Phase 1**

1. Data Exploration
2. Data Pre-processing
3. Data Cleaning
4. Type of Hotels
5. Cancellation rate
6. Deposit Policies

- **Phase 2**

Number of Bookings across various factors to observe the trends

- **Phase 3**

1. Top 10 countries
2. Meals Preferred
3. Customer Segmentation
4. Average Daily Rate
5. Car parking space
6. Repeated Guests

Hotel Business Industry

- The hotel industry is one of the most important components within the service industry, catering for customers who require overnight accommodation. It is closely associated with the travel industry and the hospitality industry, although there are notable differences in scope.

- Overall, sales from hotel accounts 87.4% of industry revenue (in India 53.9% and 8% total employment rate)

Exploring the dataset

```python
#Data Exploration
data.head()
```

	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	children
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	0.0
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	0.0
2	Resort Hotel	0	7	2015	July	27	1	0	1	1	0.0
3	Resort Hotel	0	13	2015	July	27	1	0	1	1	0.0
4	Resort Hotel	0	14	2015	July	27	1	0	2	2	0.0

```python
#Data Exploration
data.tail()
```

	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	chil
119385	City Hotel	0	23	2017	August	35	30	2	5	2	
119386	City Hotel	0	102	2017	August	35	31	2	5	3	
119387	City Hotel	0	34	2017	August	35	31	2	5	2	
119388	City Hotel	0	109	2017	August	35	31	2	5	2	
119389	City Hotel	0	205	2017	August	35	29	2	7	2	

```python
#Basic Informations of our dataset
data.info()
```

Checking the nullvalues

```
[ ]  #Checking the number of null values in each columns
     data.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
     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
     reservation_status_date         0
     dtype: int64
```

Replacing the Null Values

```
[ ]  data["children"].fillna(data["children"].mean(),inpla
     data["country"].fillna("Absent",inplace = True)
     data["agent"].fillna(data["agent"].mean(),inplace = T
     data["company"].fillna(data["company"].mean(),inplace
```

```
[ ]  #Ensure there is no null value in our dataset.
     data.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                           0
     company                         0
     days_in_waiting_list            0
     customer_type                   0
     adr                             0
     required_car_parking_spaces     0
     total_of_special_requests       0
     reservation_status              0
     reservation_status_date         0
     dtype: int64
```

City Hotel and Resort Hotel

- It is Observed that City Hotel has more number of bookings compared to Resort Hotel.

- The obvious assumption for this trend would be because of monetary reasons.

- City hotels would be more affordable than resort hotels due to accessibility, reliability and lavishing factors.

City Hotel

66.4%

33.6%

Resort Hotel

Checked-in and Cancelled Bookings

- It is Observed that out of 119390 bookings made, 75166 of them have checked in and 44224 of them has cancelled their bookings.

- In percentage, it is about 63% of check-in bookings and 37% of bookings were cancelled.

- High cancellation rate will adversely effect the hotel industry which reduces the yield of profit.

Booking canceled and not canceled
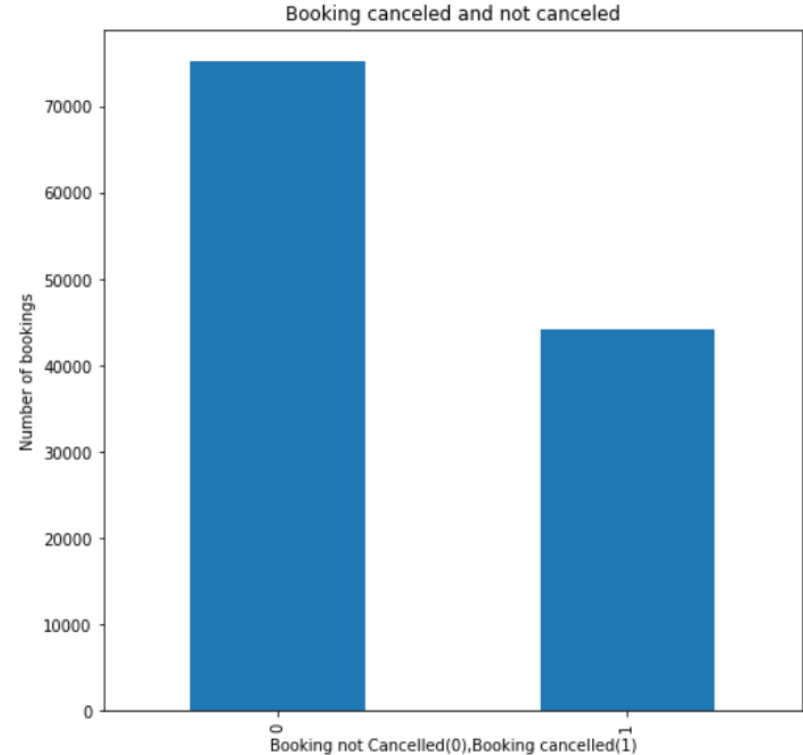
Deposit Policies of Hotel

- Most number of the hotel bookings are made in 'No Deposit, type of transaction which is the reason for the high cancelation rate.

- 'No Deposit' type of transaction means without paying any advance money for the booking.

- In order to reduce cancelation rate, we need to find suitable alternative for the 'No Deposit' type of the transation.

Types of Deposits
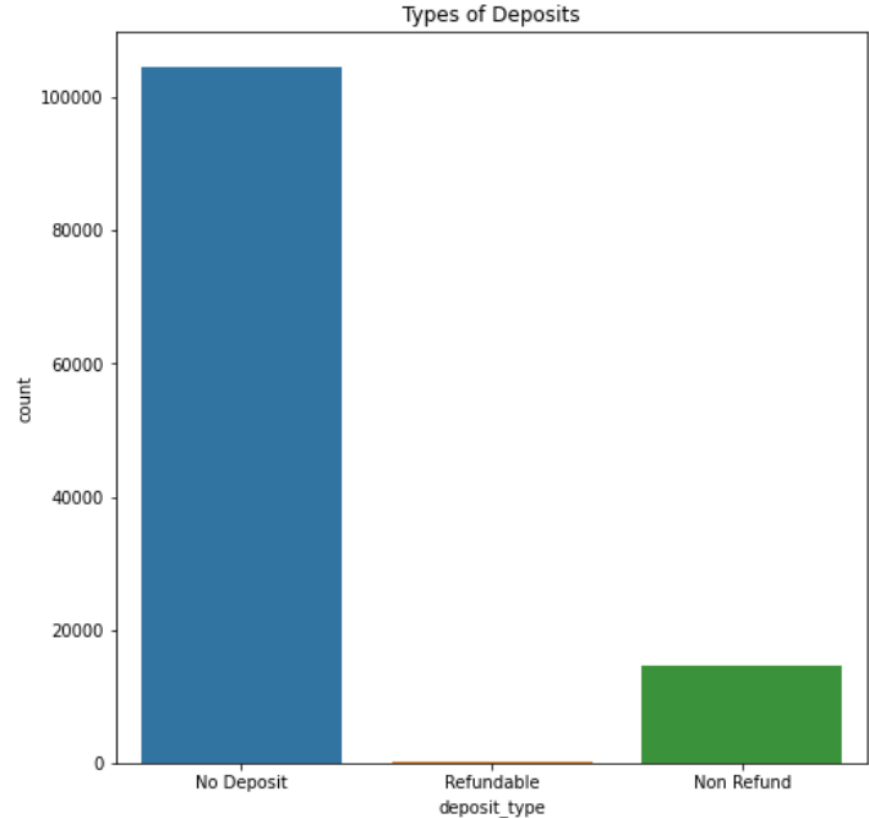
Number of Bookings Across Different Years

- Most Number of the bookings are done in the year 2016 following 2017 and 2015 for the dataset given comprising the data for these 3 years.

- When, sub plotted, we can see that City hotel has high demand rather than Resort hotels due to its reliability and price difference compared toresort hotels.
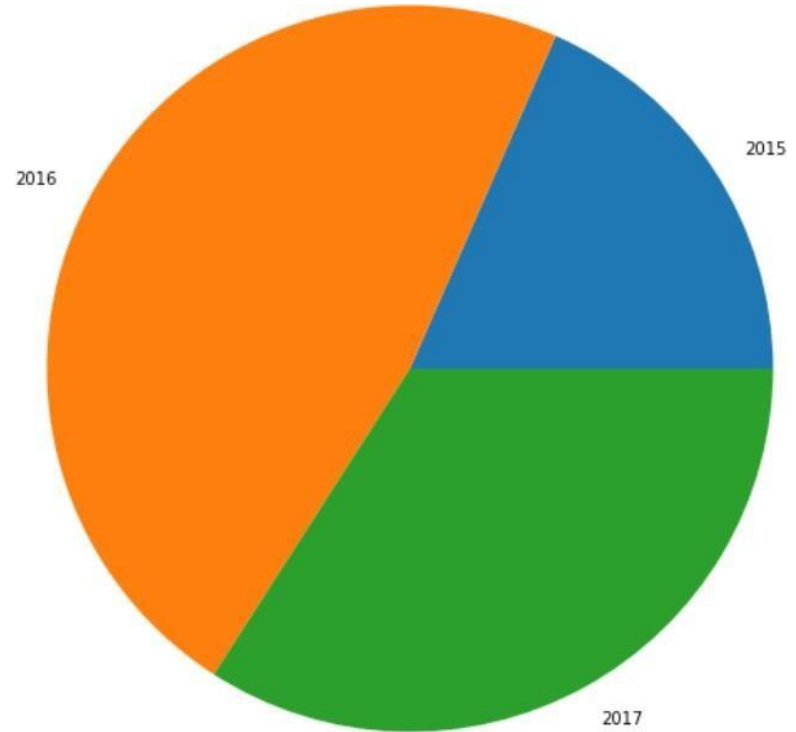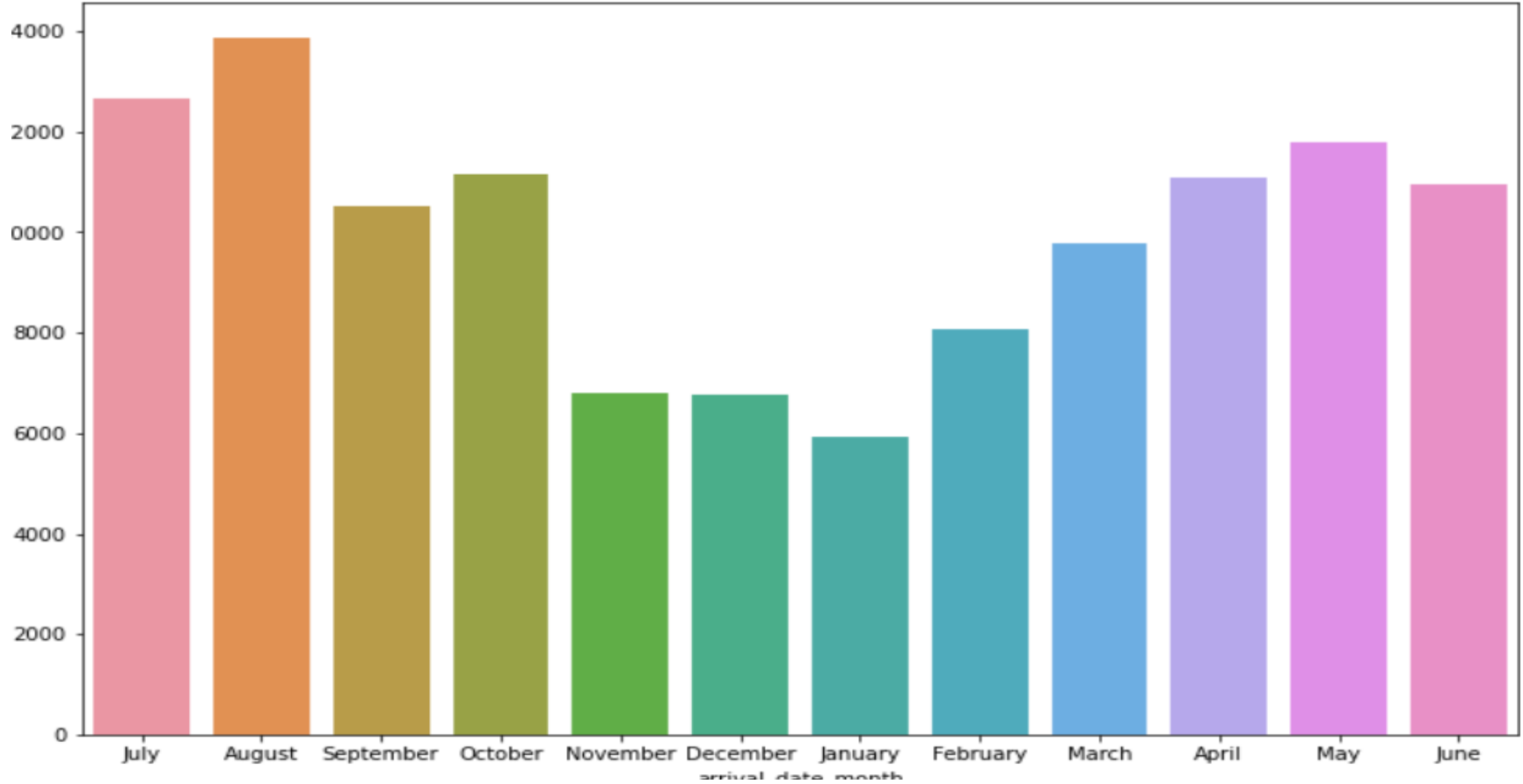
Number of Bookings Across Different Years
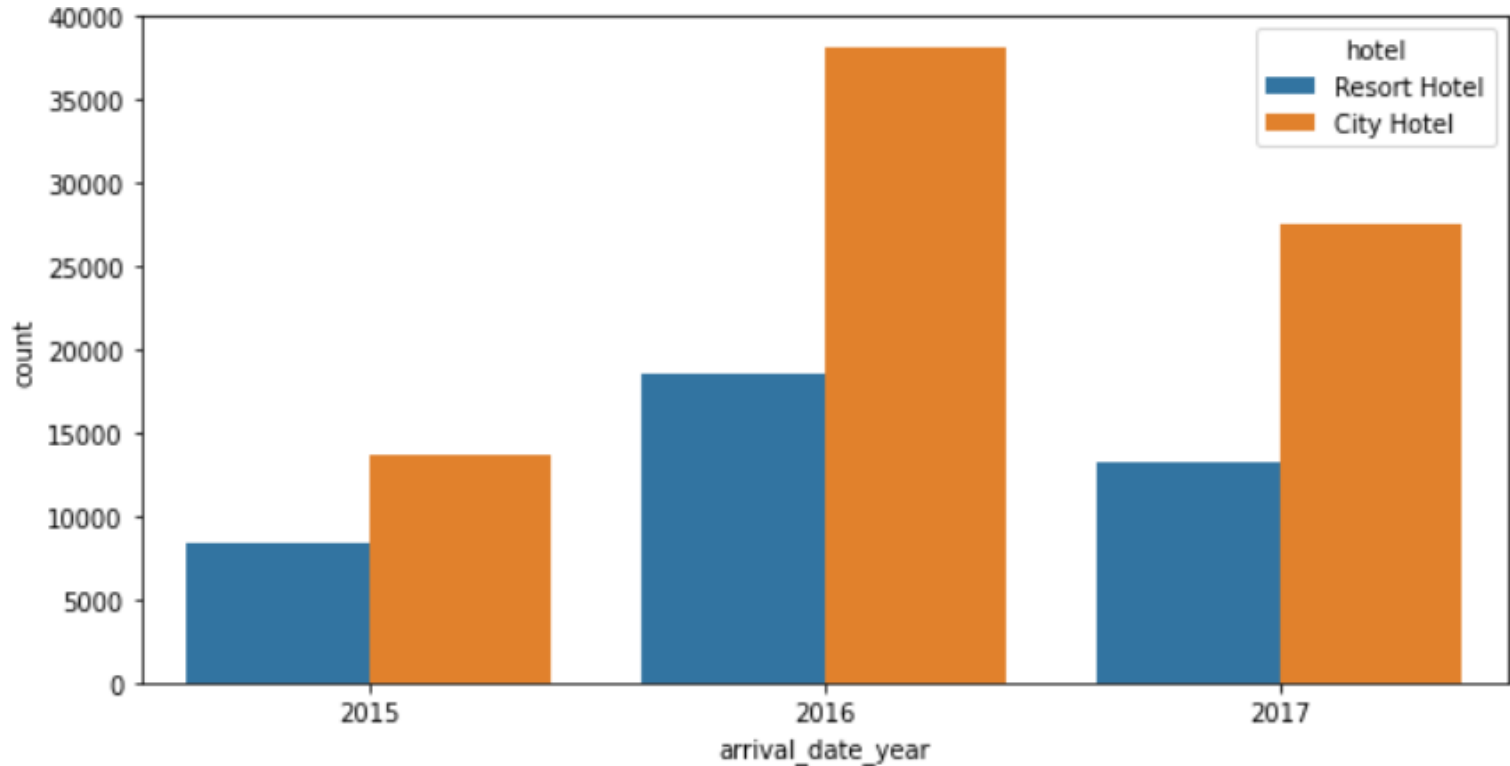
Number of Adults and Children
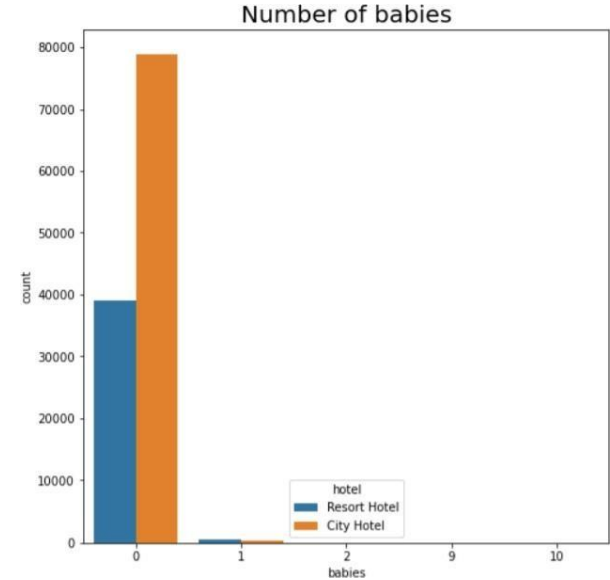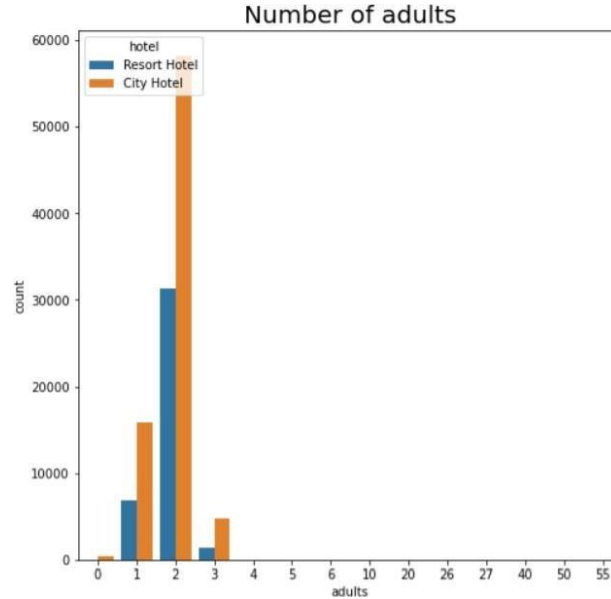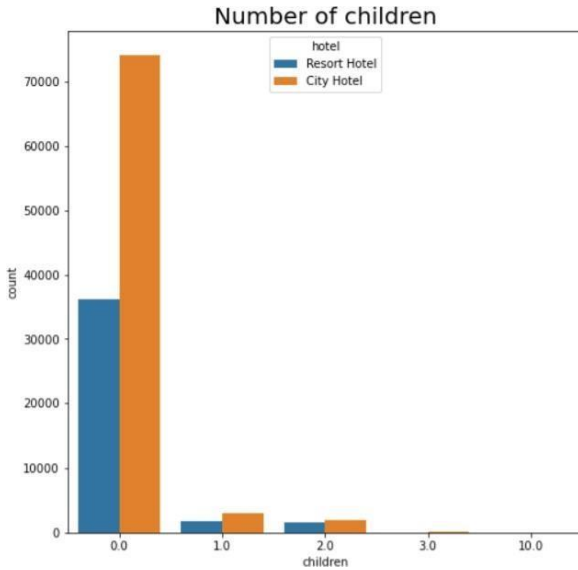
- On an average, two adults checks in at a time. The trend is same for both resort and city type hotel.
- From the graphs above we get the insight that usually customers would be adults, if the children comes along it would be usually one or two.
- The count of babies are even less.

Number of booking done across different months

Number of booking done across different months

- From the graph which we have interpreted, we can see that most number of the bookings are done at month of August, and then July.

- Whereas January has the lowest number of bookings.

- From Customer point of view, January would be the best month to book a hotel as Average daily rate would be less at this point of time.

- We can also assume that most bookings are done during Autumn season due to the ideal weather conditions without much fluctuations.

Arrival Per day

Graph showing number of arrival per day
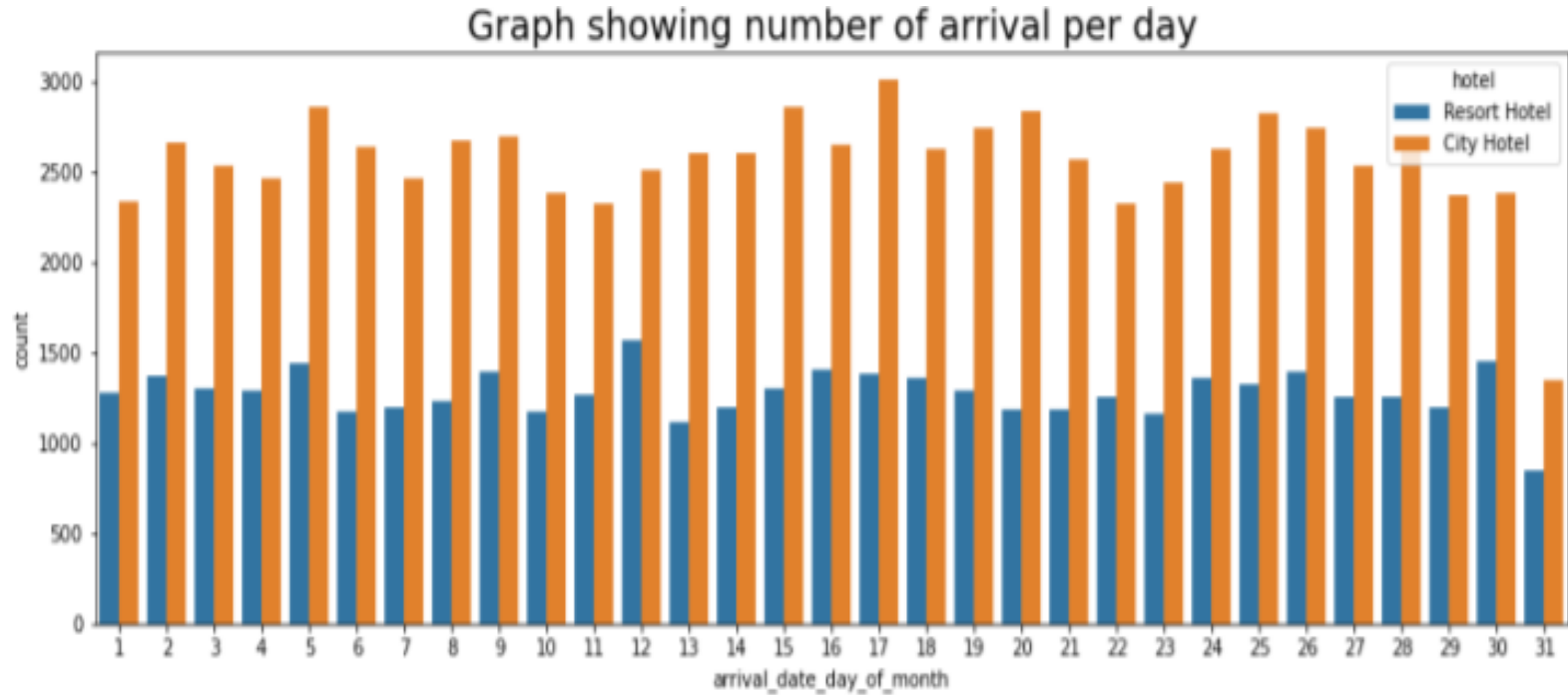
Night Stay Duration

Night Stay Duration
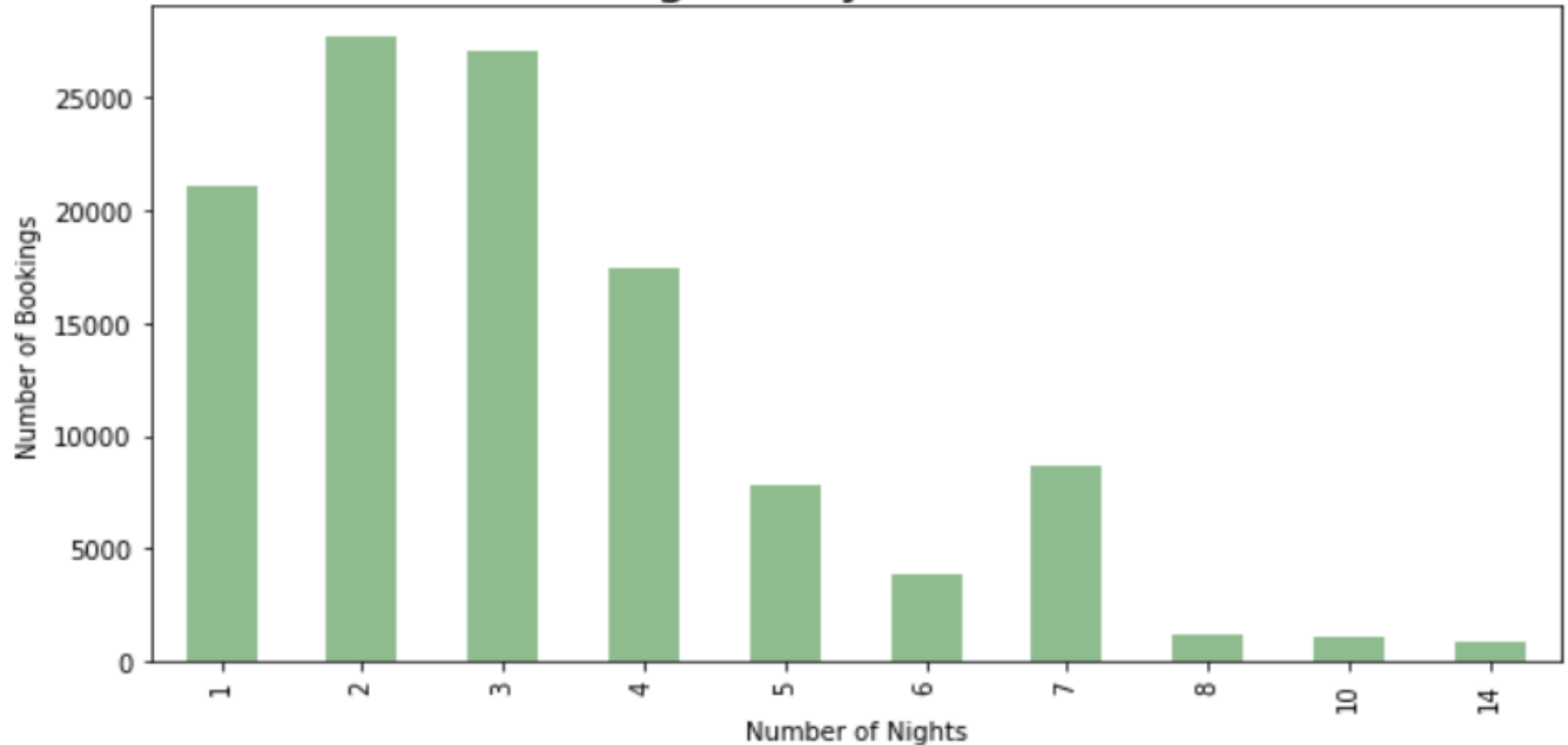
Night Stay Duration

- By combining the two columns 'stay_in_week_nights' and 'stay_in_weekend_nights' we got total number of night stays.

- From the graph we can easily interpret that most number of customers prefers to stay for one, two and three days of which most customers prefer for 2 night stays.

- Very few number of customers are interested to stay for more than 5 days.

Top 10 Countries with highest number of Bookings

Total number of customers from different states
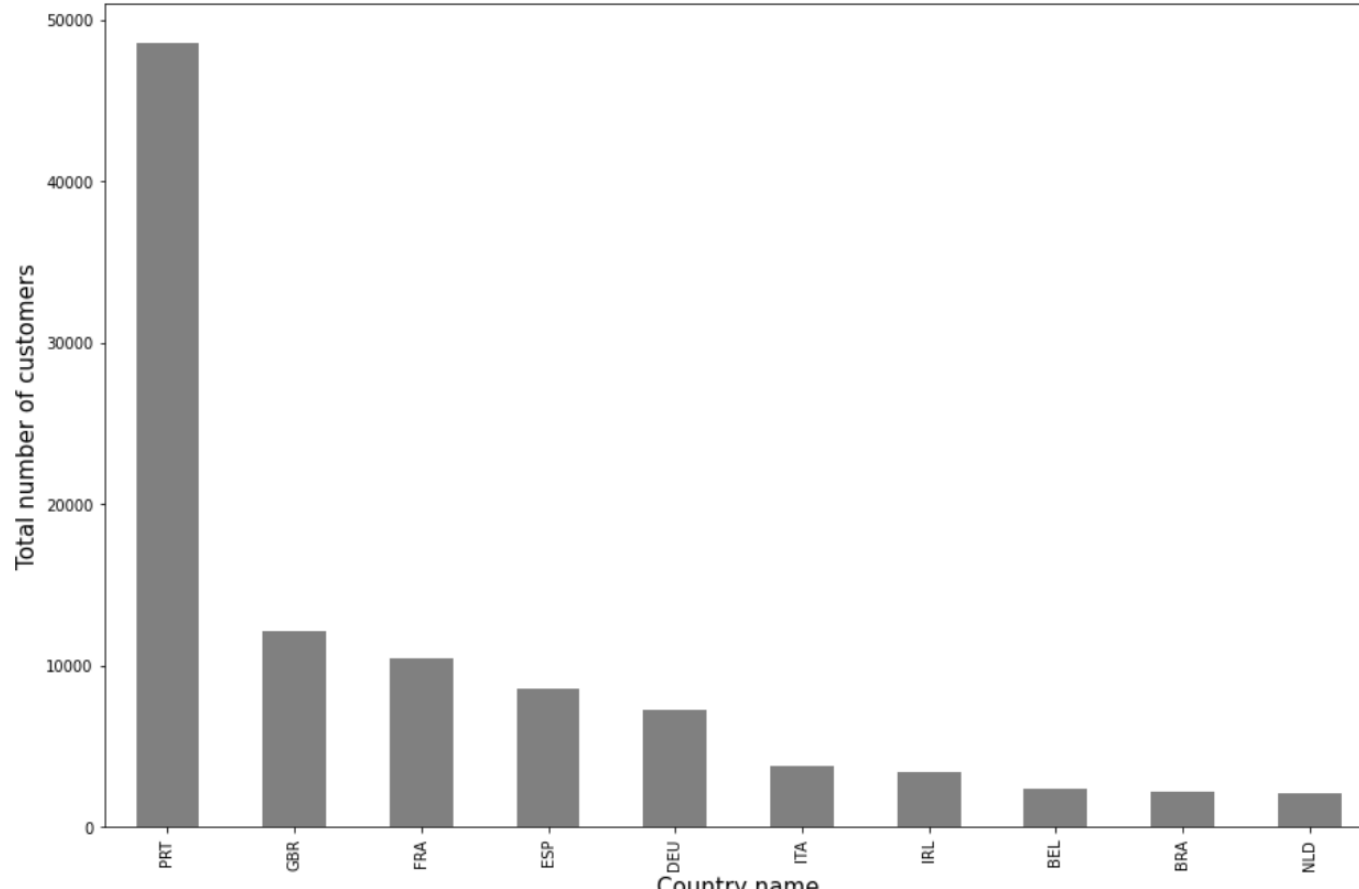
Top 10 Countries with highest number of Bookings

- After analyzing the dataset, we found that Portugal tops the position with 48590 customers followed by UK with 12129, France with 10415, Spain with 8568 and Germany with 7287 customers.

- Netherlands sits back with least number of customers among the top 10 Countries.
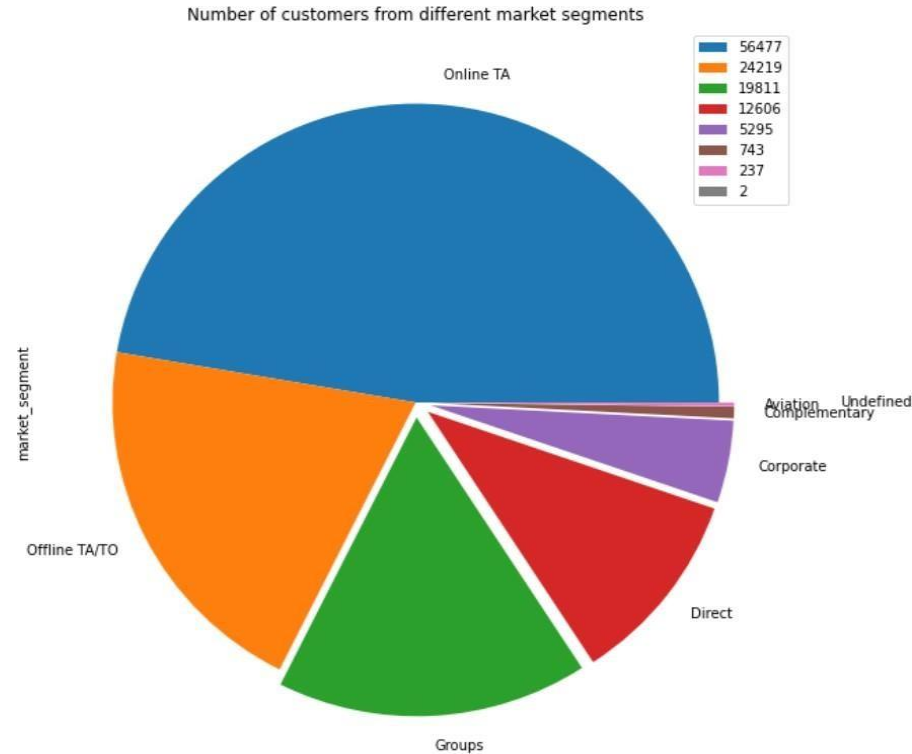
Percentages of Meals Prefered

- BB means bed and breakfast.
- HB means half board, in which breakfast and dinner are included.
- FB means full board, in which breakfast, lunch and dinner are included.
- SC means self-catering (no meals are included).
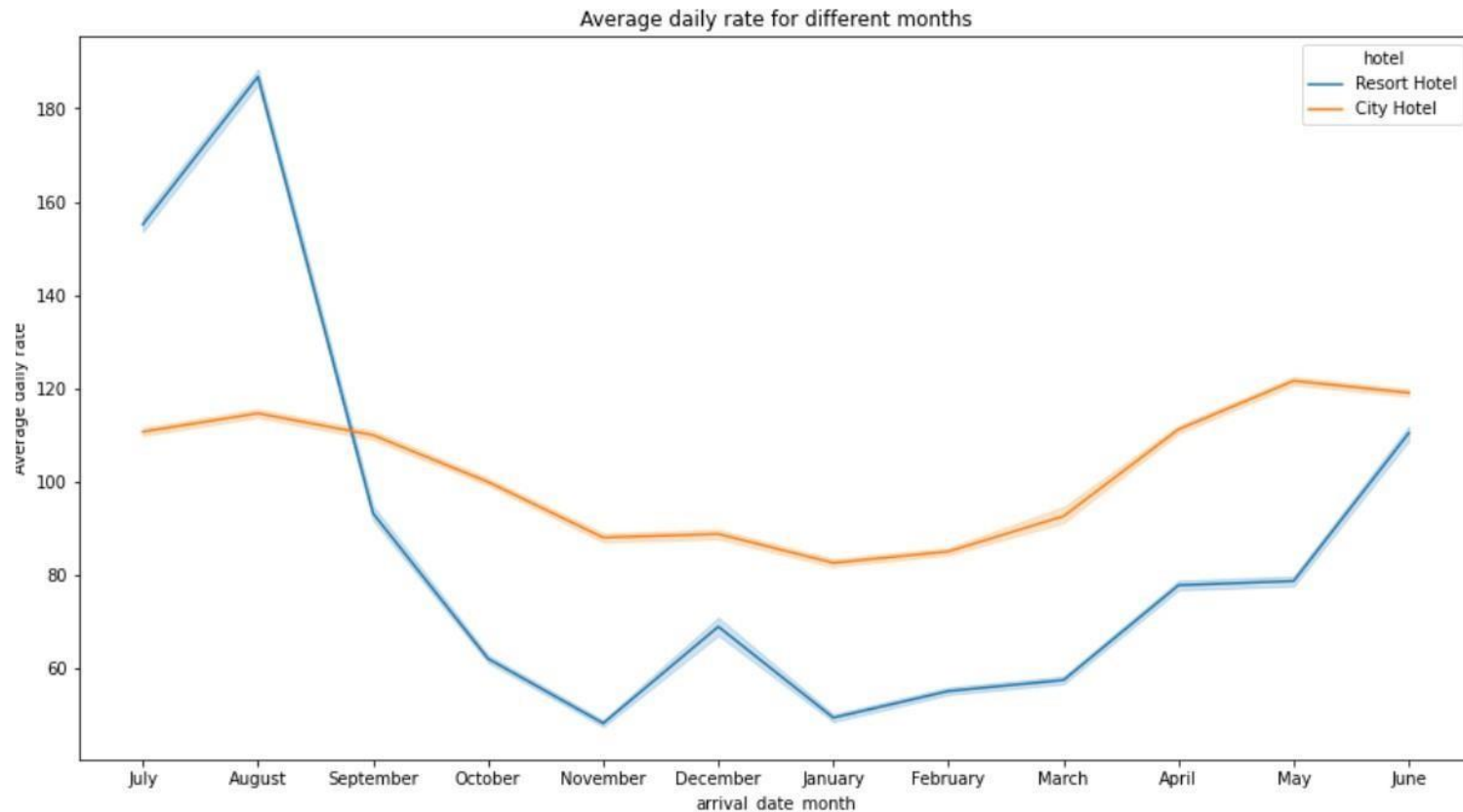
PERCENTAGE OF MEAL PREFERENCE
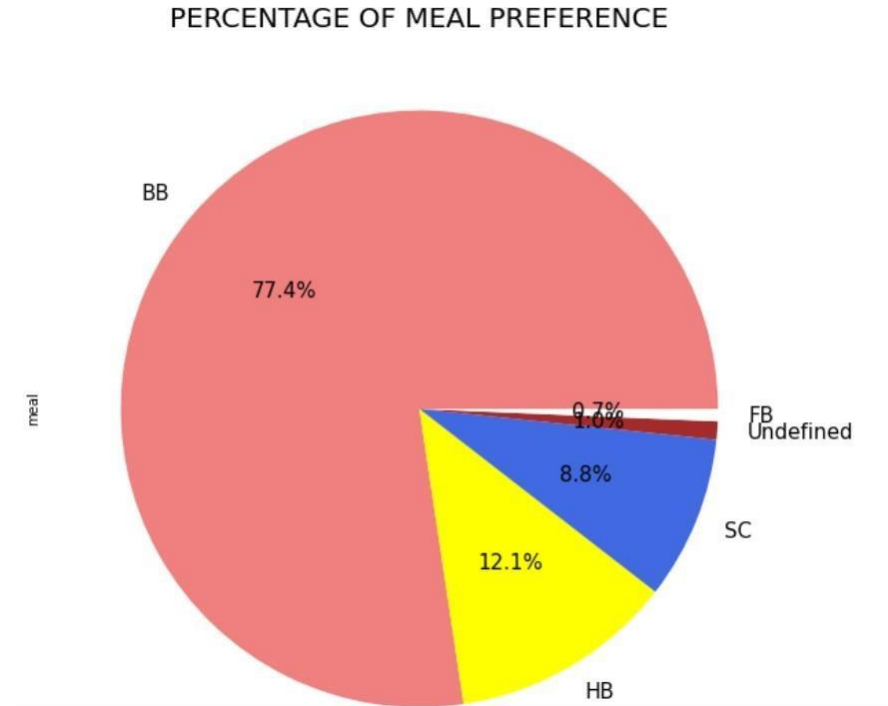
Total Number of Customers across Various Market Segments

- Online Travel agency segment gives the high amount of leads for the hotel booking than that of any other sources of Marketsegments.

- We can report that we need to target our marketing area on online TA websites or apps and focus majorly on online TA.

- The following majority market segments are offline travel agencies, groups and direct customers.

Number of customers from different market segments

Color	Value
blue	56477
orange	24219
green	19811
red	12606
purple	5295
brown	743
pink	237
gray	2

Online TA
Offline TA/TO
Groups
Direct
Corporate
Aviation Undefined
Complementary
market_segment

Average Daily Rate for Different Months


Average daily rate for different months

Average Daily Rate for Different Months

- The average daily rates are high during the months of may, June, July and august for City hotels.

- The average daily rates are high during the months of July, august and September for Resort Hotels.

- Overall ADR for both city and resort hotels are more expensive between may and September.

Average Daily Rate for Different Years

AVERAGE ADR OF MONTH & YEAR

# Car Parking space analysis

CAR PARKING SPACE ANALYSIS
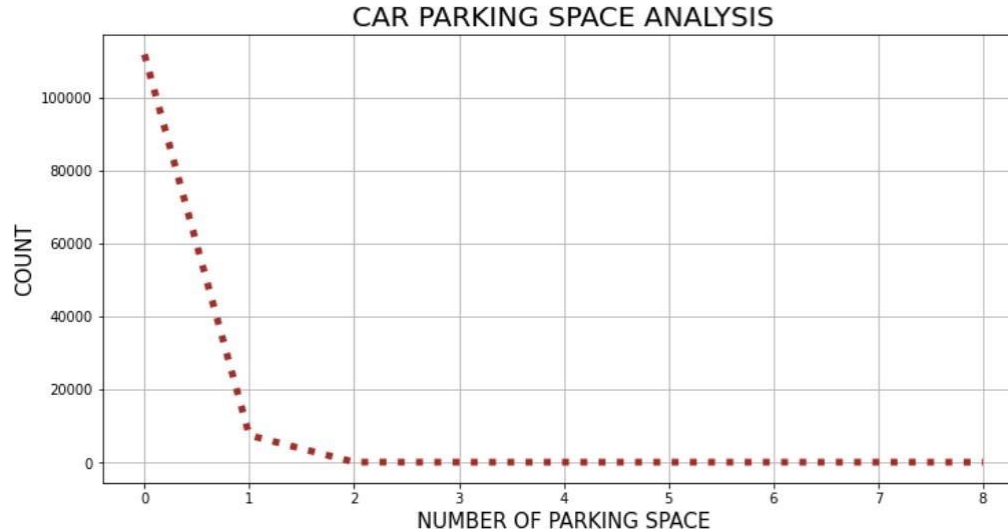
- Most of the hotels have zero car space parking.
- Among which have car space parking has one car spacing and the trend decreases drastically as number of car spacing increases.
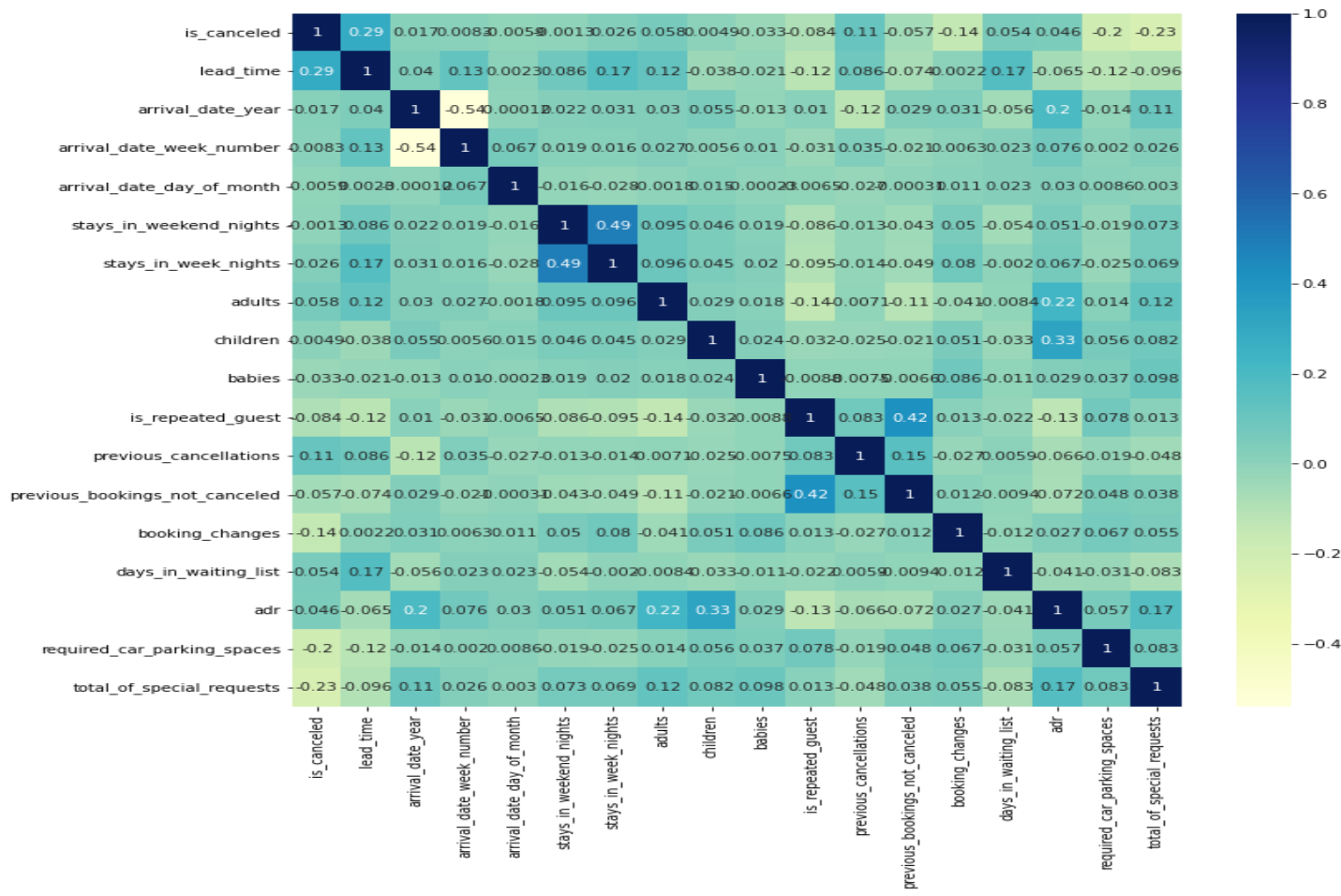
Is-Repeated Guests Counts

- Very few number are seem to be loyal to the hotels.

- Only around 3755 bookings are of repeated guests from all the bookings.

- The management team has to deal with this by giving discounts for next visit or free coupons or something to increase the number of guests.

(0 = Not repeated & 1 = Repeated)
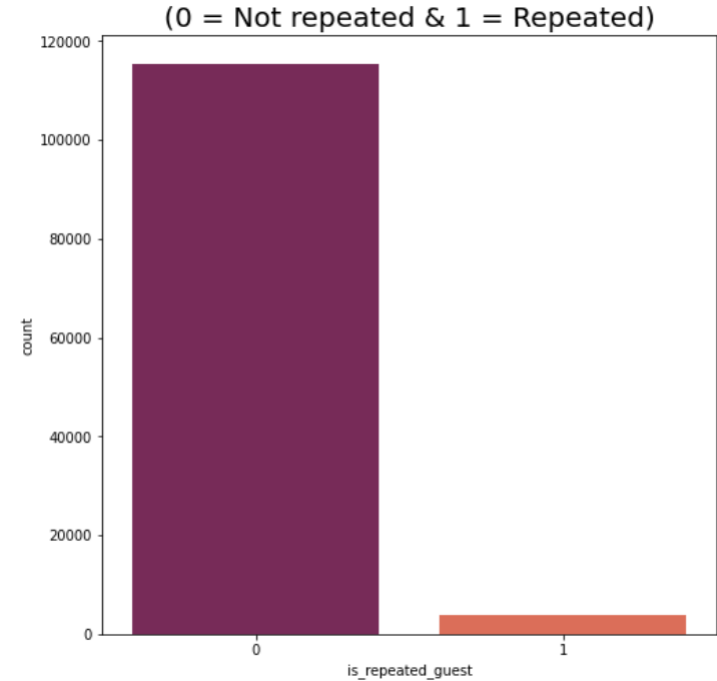
Correlation Heatmap

Conclusion

1.'City hotels' and 'Resort hotels' are two types of hotels present in the dataset, out of which, 'City hotels' are more preferred by the customers than the latter. (66.4% customer prefers 'City hotels' whereas 33.6% customerprefers 'Resort hotels')

2.Out of 119000 customer dataset, 75166 customers checked in hotel while 44224 customers cancelled their bookings, that is about 37% of the booking got cancelled whereas 63% customers checked in the hotel.

3.Majority of the deposit type is of **'No deposit'** type, which itself concludes the high rate of cancellation rate.

4.From the given dataset, we can see 2016 is the year in which hotel bookings are the highest.

5.We can also see the trend in the middle of the year as those season has less weather condition and holidays are more during that season. We can also infer that winter season has the lowest number of bookings around the globe and we can assume it because of the weather condition.

6.Out of all months, 'August' witnessed highest number of hotel bookings whereas 'January' witnessed the least.

7.City hotels are of in high demand compared to resort in all aspects due to its reliability of majority of the population.

Conclusion

8.Most of the customers prefer to stay for one, two and three nights and majority its two and three night which tops.

9.Among the top 10 countries with respect to number of bookings, most of them are European countries which belongs to the list of which Portugal stands first with 48.5k bookings.

10.For resort hotels, ADR is more expensive during June, July, August and September whereas for city hotels, ADR is more expensive during May, June July and August. So, Overall ADR of both cities are more expensive between May and September.

11.Its observed that average ADR is incrementing every year from 2015 to 2017, which clearly states that hotel business is scaling up every year.

12.Online Travel agencies provides high amount of leads and customers following Offline TA, Groups, Direct, etc., So, we an conclude that we need to focus our marketing area on online TA since majority of the visitors tend to reach out to them.

13.After analyzing the meal data, its found that 77.3% of customers prefers BB(Bed &

Breakfast). 14.Most of the hotels have 0 to 1 car parking space.

 15.Given that we do not have repeated guests, we should target our advertisement on guests to increase returning guests.

References

1) https://pandas.pydata.org/

2) https://matplotlib.org/

3) https://seaborn.pydata.org/

4) Geek for geeks

Thank You

Team GodSpeed