AI

Capstone Project - 1
Exploratory Data Analysis
Hotel Booking Analysis

Team **DEBUG ENTITY**

Team Members:
Abhilash S
Harshitha K S
Surbhi Kumari

Contents :

- Introduction
- Data exploration
- Data cleaning
- Types of hotel
- Cancellation rate
- Deposit policies
- Bookings
- Customer segmentation
- Preferences
    - Meals
    - Room Type
    - car parking
- Guests from different countries
- ADR and Corelation Heat map
- Conclusion
- References

Hotel 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

Checking and replacing the Null Values

Missing Values

```
df.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
```

Data Cleaning

```
df["children"] = df["children"].fillna(0)
df["country"] = df["country"].fillna("Undefined")
df["agent"] = df["agent"].fillna(0)
df["company"] = df["company"].fillna(0)
df
```

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

119390 rows × 32 columns

✓ 0s   completed at 2:53 PM

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.
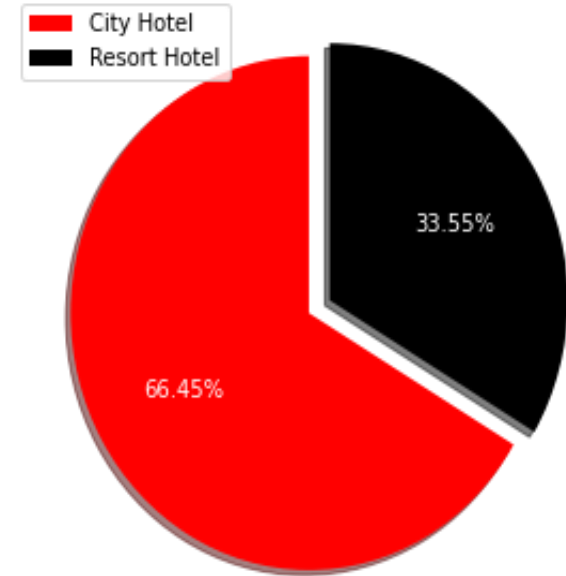
Cancelled Bookings

AI

- 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.
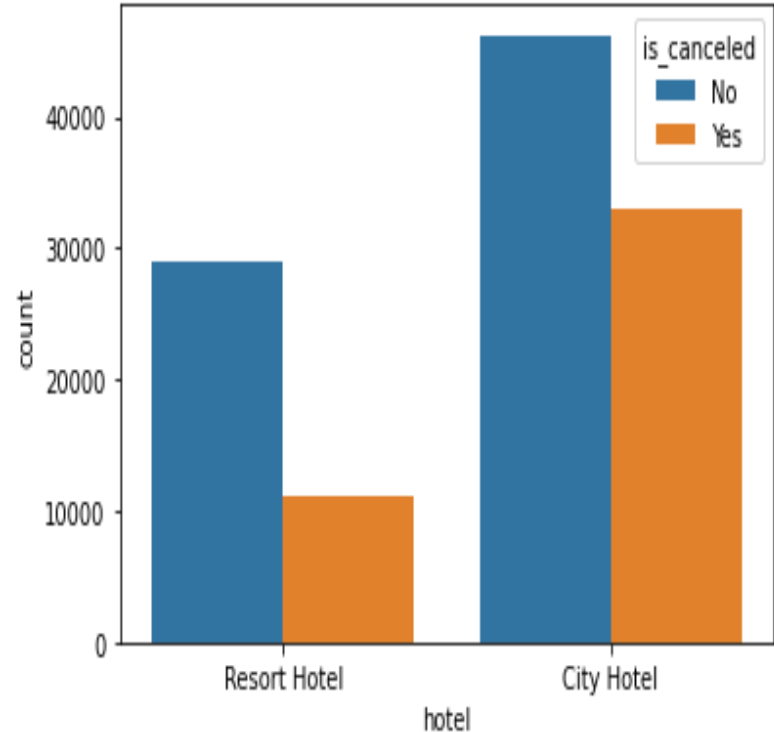
# Number of Booking cancelled by the same customers repeatedly
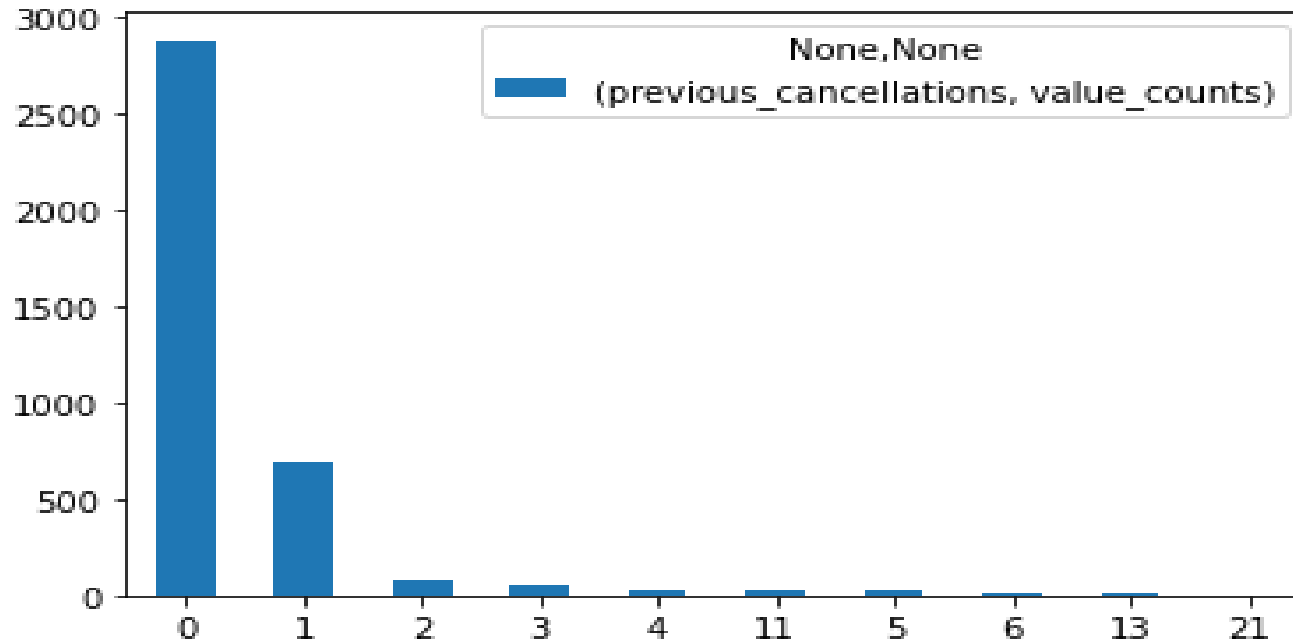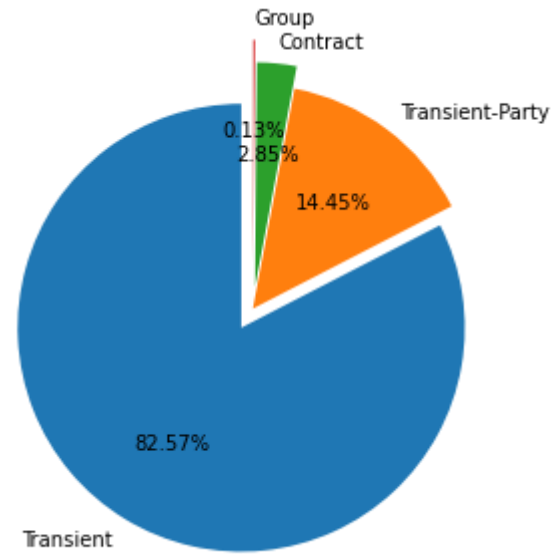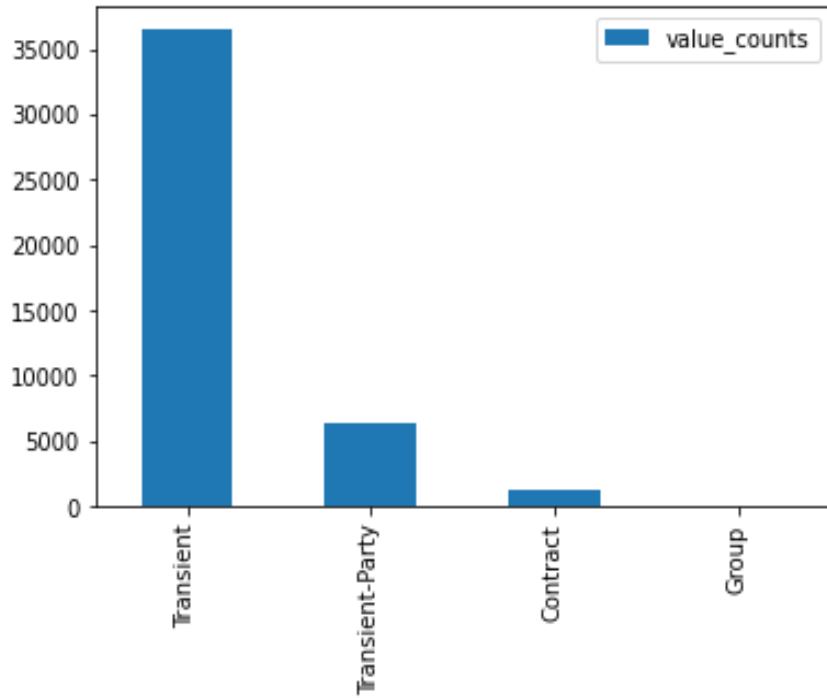
It is observed that Transient customers are more likely to cancel the booking.
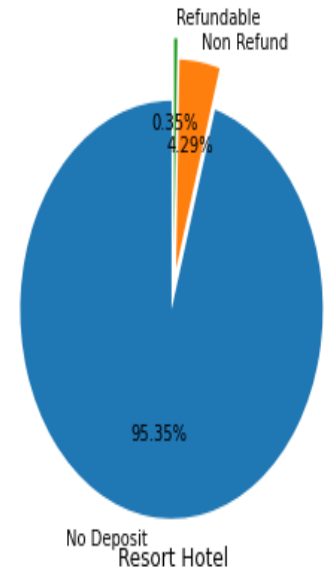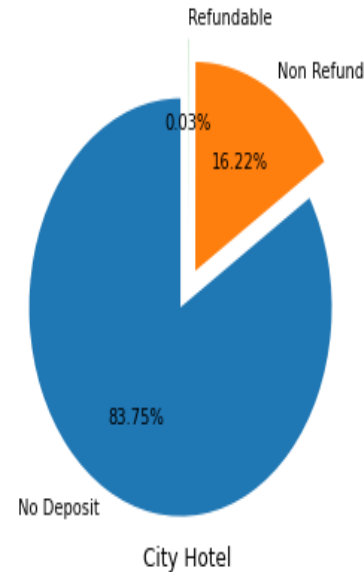
Total Number of Customers across Various Market Segments

- 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.
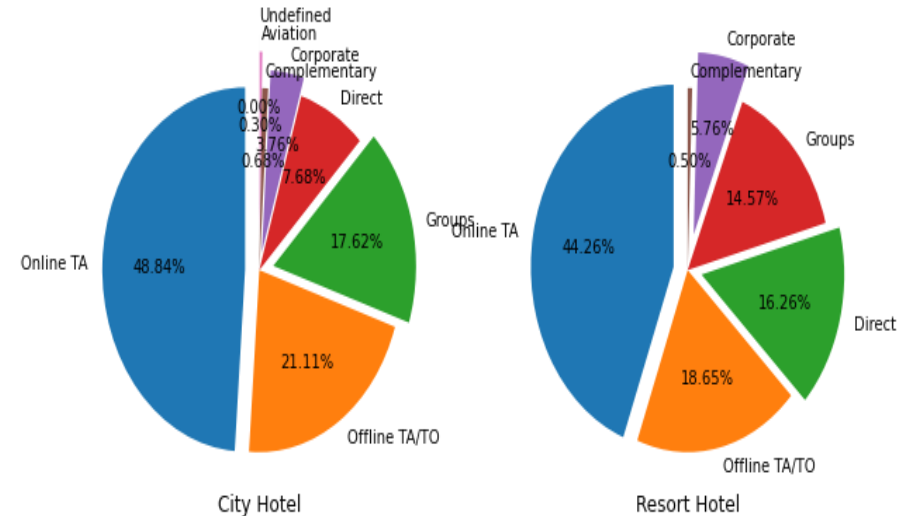
Deposit Policies of Hotel

- 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 the cancelation rate, we need to find suitable alternative for the 'No Deposit' type of transaction.

City Hotel — Refundable, Non Refund 0.03%, 16.22%, No Deposit 83.75%

Resort Hotel — Refundable, Non Refund 0.35%, 4.29%, No Deposit 95.35%

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 to resort hotels.
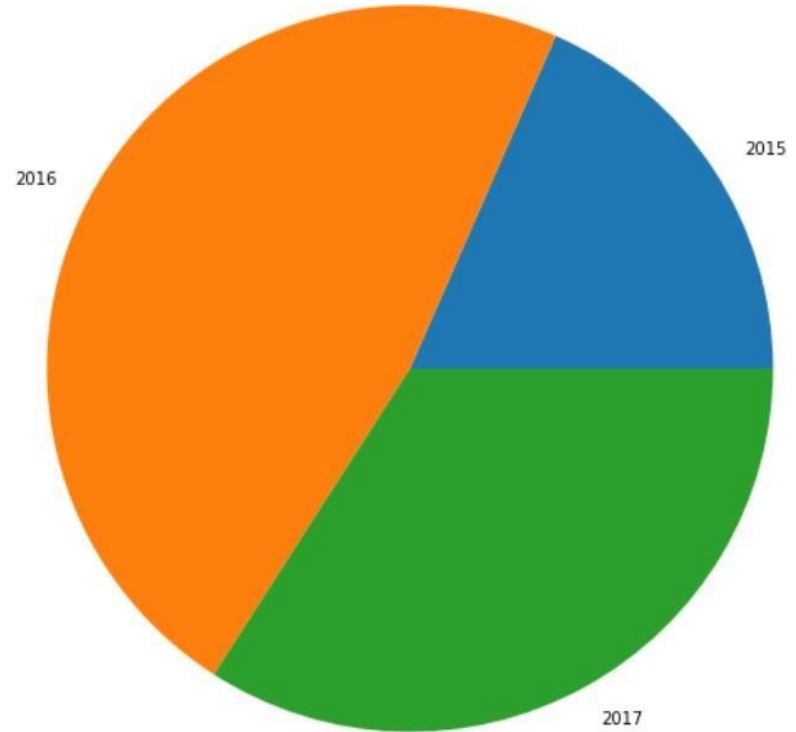
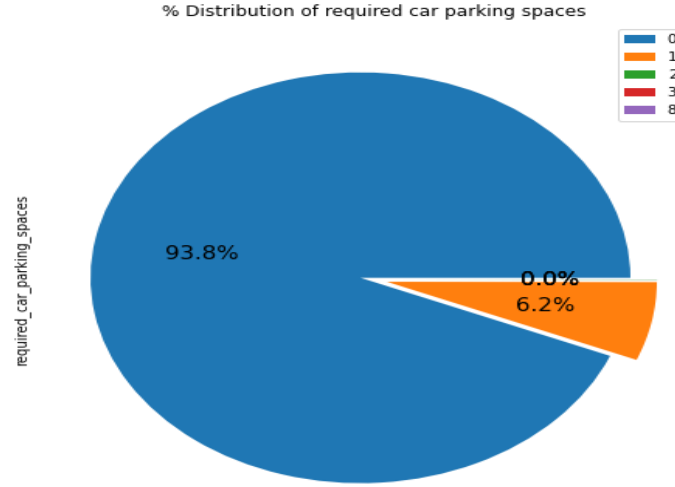# Number of Bookings Across Different Years
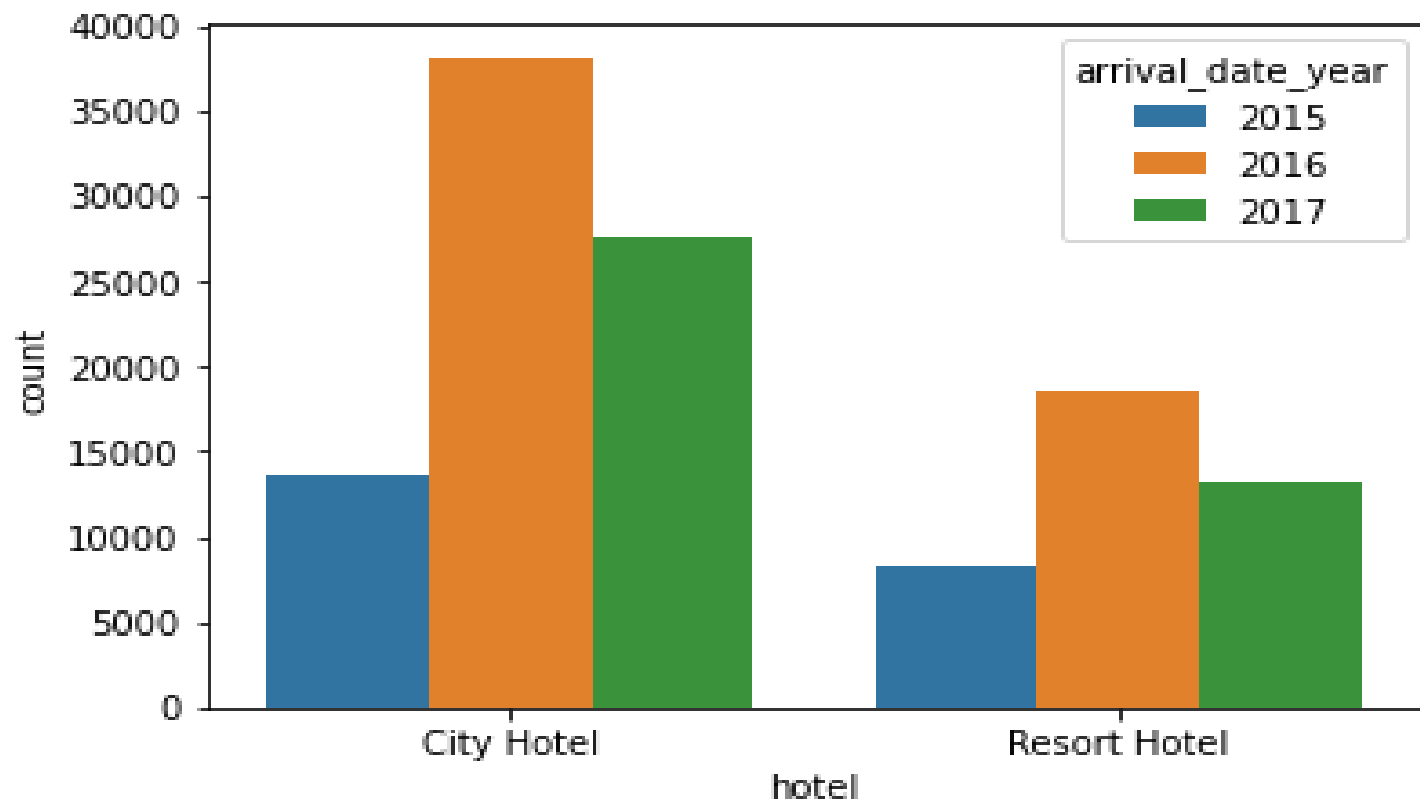
Meals preference
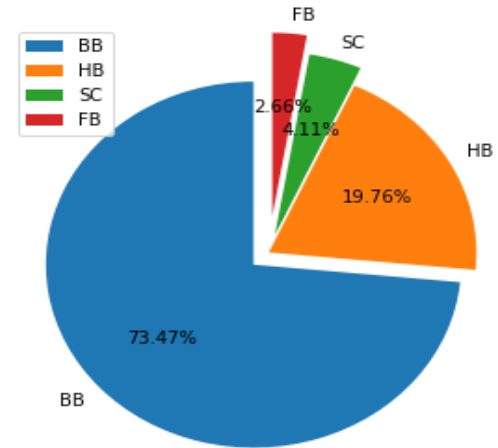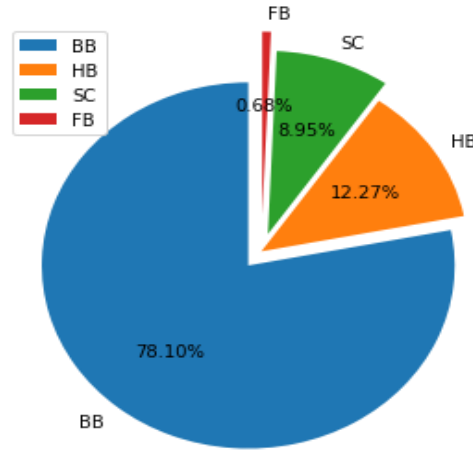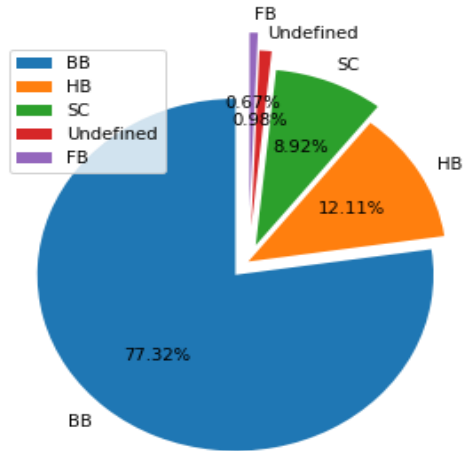
As per the above Graph, we can conclude that guests for BB (i.e. Bed & Breakfast) is more than others i.e. Half Board(HB), Full Board(FB), Self Catering(SC)
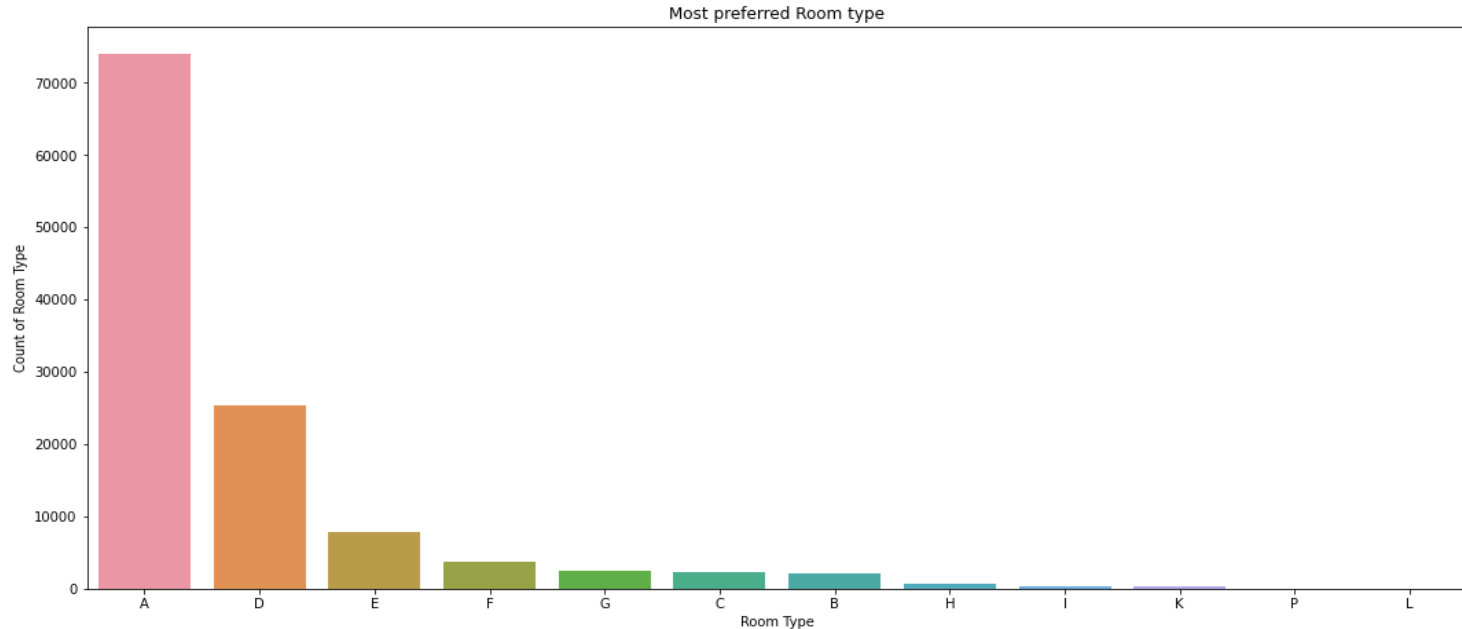
Car Parking Space Analysis

% Distribution of required car parking spaces

- 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.

Preferred room type by guests :

Maximum number of guest Preferred 'A' type rooms.


Most preferred Room type

# 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 10Countries.
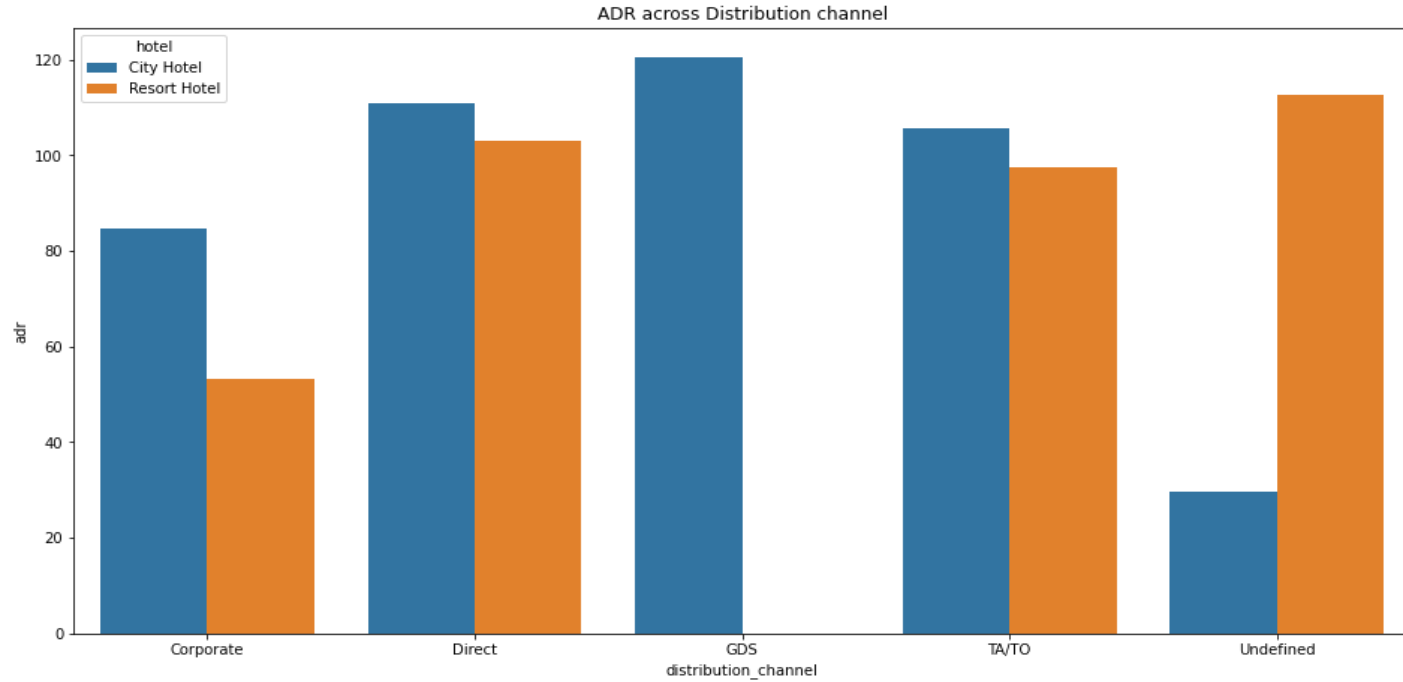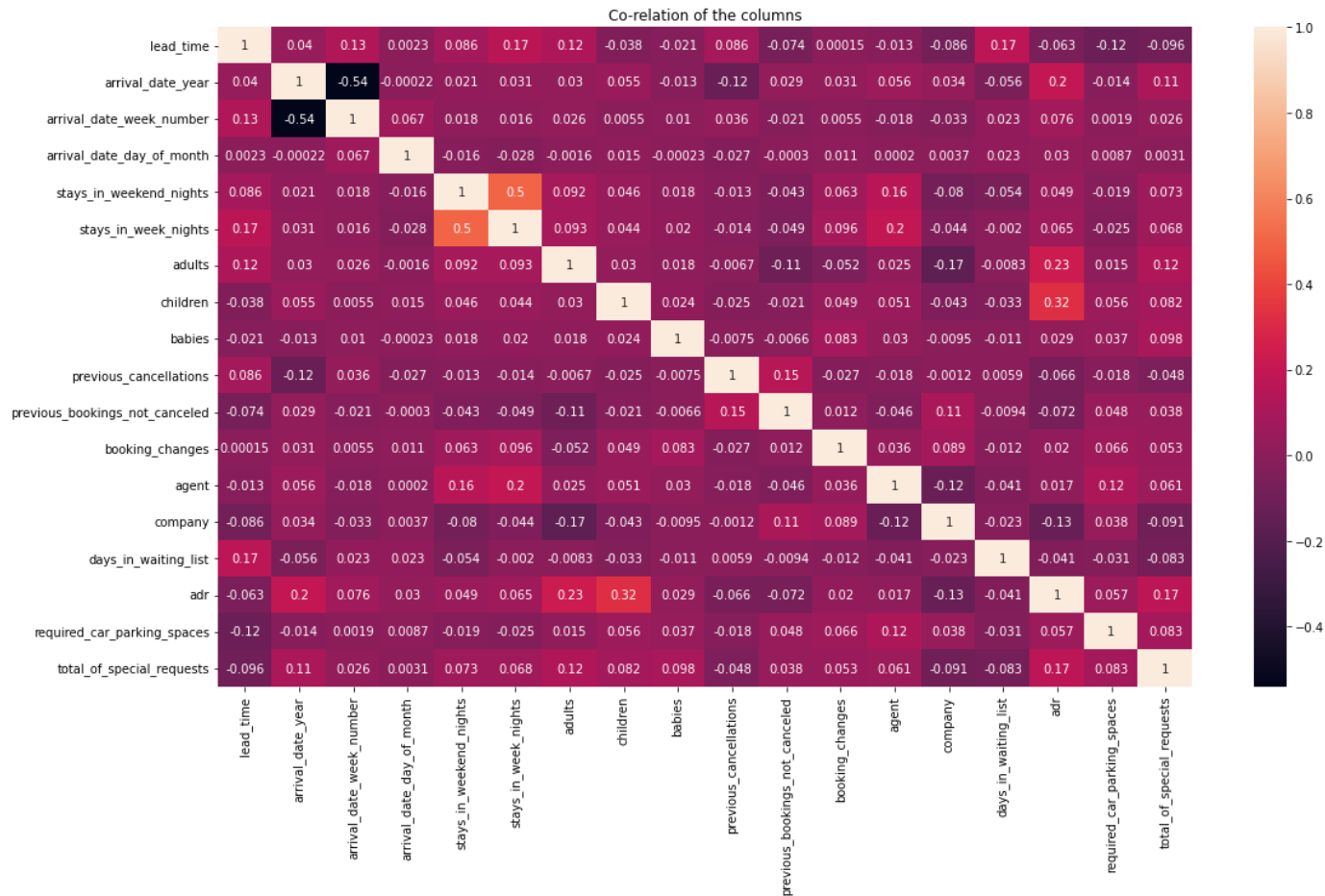
Top 10 Countries with highest number ofBookings

Number of guests from diffrent Countries

ADR Across Distribution Channel
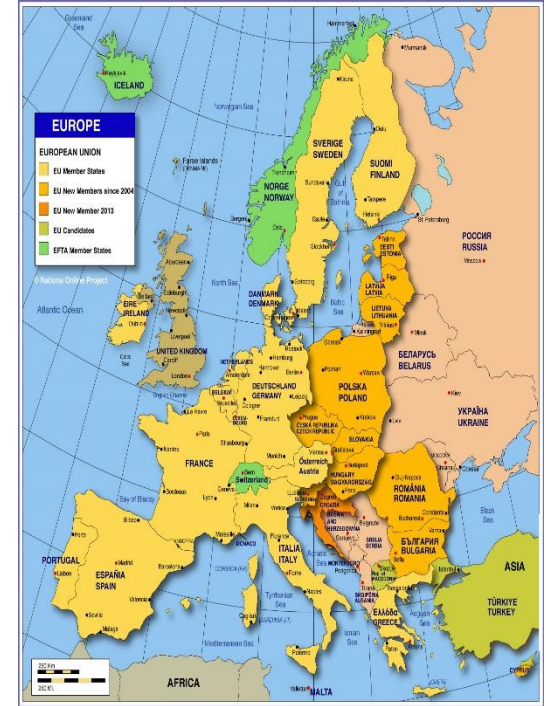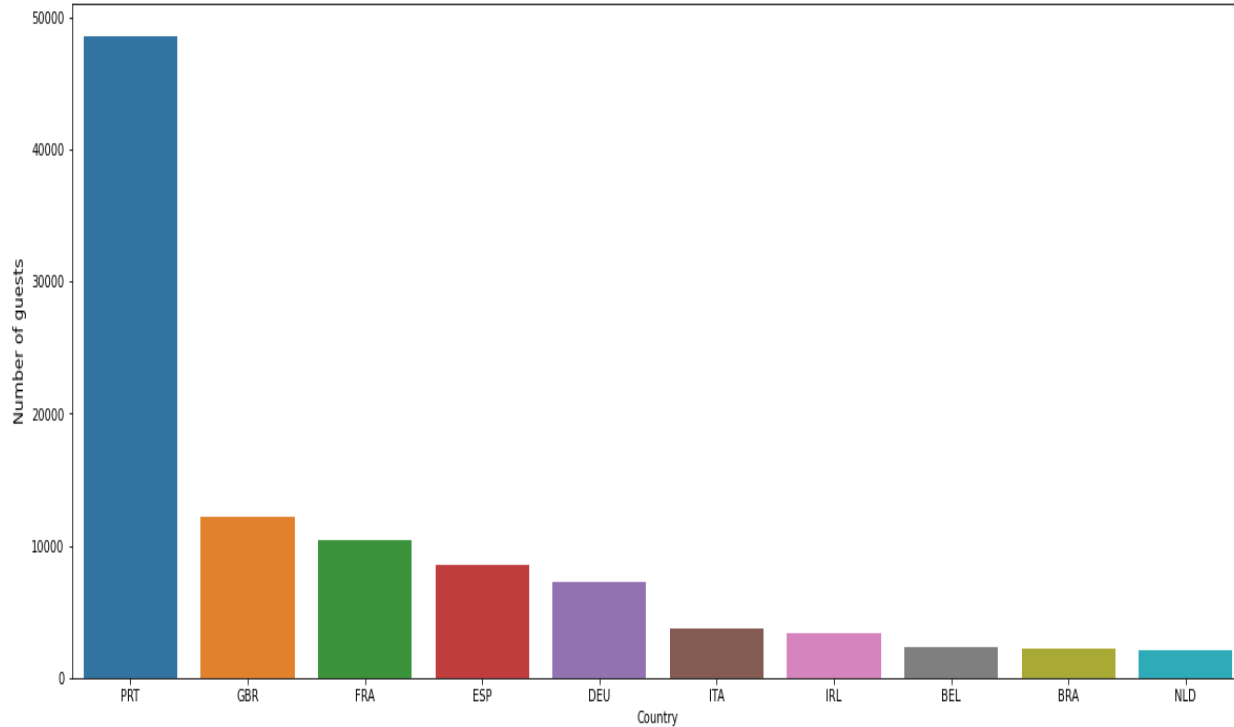
ADR across Distribution channel

Correlation Heatmap

Co-relation of the columns

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

7.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

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

9.Most of the hotels have 0 to 1car parking space.

10.Among the top 10countries 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.

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.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

5) Almabetter Study Materials

Thank You

DEBUG ENTITY