

Capstone Project

Hotel Booking Analysis



Team Members:

Aayushi Sharma

Manzar Ansari

Shubham Chavhan

Content

1. Problem Statement
2. Data Summary
3. Exploratory Data Analysis
4. Challenges
5. Conclusion



Problem Statement

- Hotel industry is very volatile industry, and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. Booking cancelations are one of the major problems faced by the Hotel industry.
- To reduce this uncertainty and maximize the revenues , we have analyzed the patterns available in the past data to help hotels plan better.
- The main objective is to create an in-depth analysis to figure out the standard patterns of booking based on various factors

Data Summary

- Data set name: df

- Shape:

- Rows – 119390

- Columns – 32

- Columns:

- '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', 'babies', 'meal', 'country', 'market_segment', 'distribution_channel', 'is_repeated_guest', 'previous_cancellations', 'previous_bookings_not_canceled', 'reserved_room_type', 'assigned_room_type', 'booking_changes', 'deposit_type', 'days_in_waiting_list', 'customer_type', 'adr', 'required_car_parking_spaces', 'total_of_special_requests', 'reservation_status', 'reservation_status_date'



Exploratory Data Analysis



Handling Missing values

- We have 4 columns with missing values

- Percentage of missing value:

- Company – 94%
- Agent – 13%
- Country – 0.4%
- Children – 0.003%

company	112593
agent	16340
country	488
children	4
lead_time	0
arrival_date_year	0

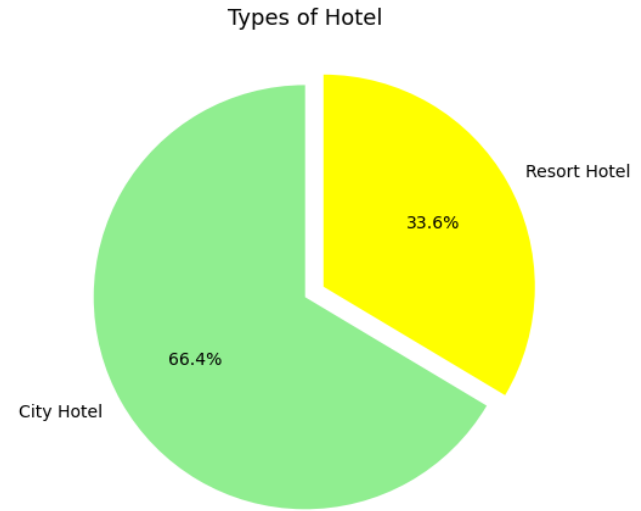
- We have deleted company and agent column

- Replaced Null values in Children with '0'
- Replaced Null values in Country with 'Unknown'

company	94.306893
agent	13.686238
country	0.408744
children	0.003350
lead_time	0.000000
arrival_date_year	0.000000

How Many types of hotels are we dealing with?

- There are 2 types of Hotel
 1. City Hotel with 66.4% demand
 2. Resort Hotel with 33.6% demand



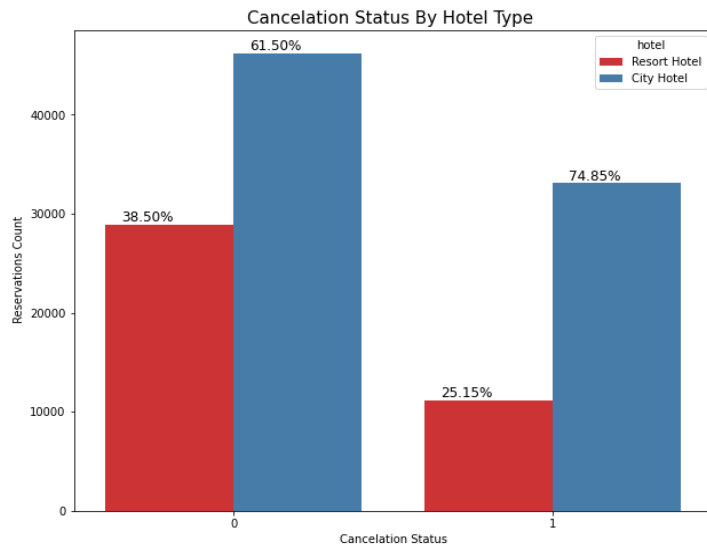
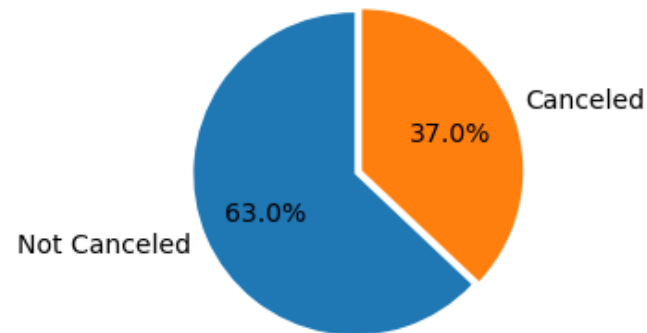
Overview of cancellations

- Number of Bookings canceled : **37%**
- Number of bookings Not canceled: **63%**

- Resort Hotel:**
- Booking canceled: **25.15%**

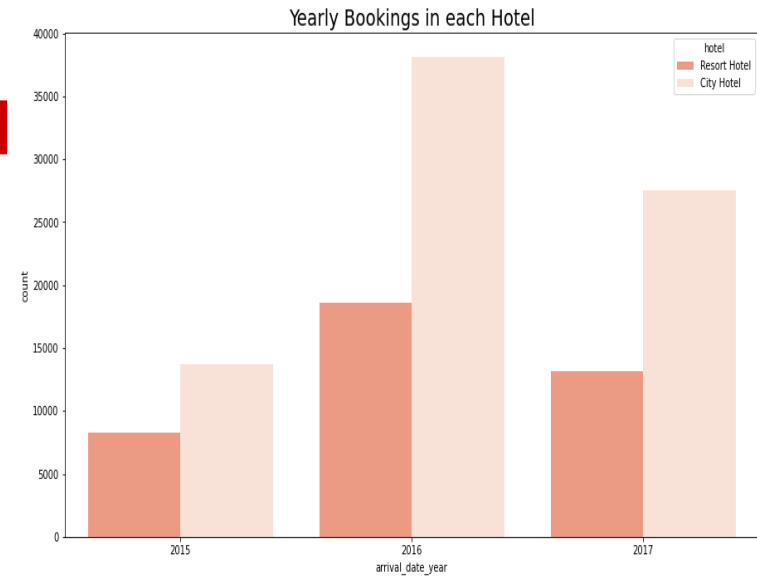
- City Hotel:**
- Booking canceled: **74.85%**

Proportion of Canceled & Not Canceled Bookings

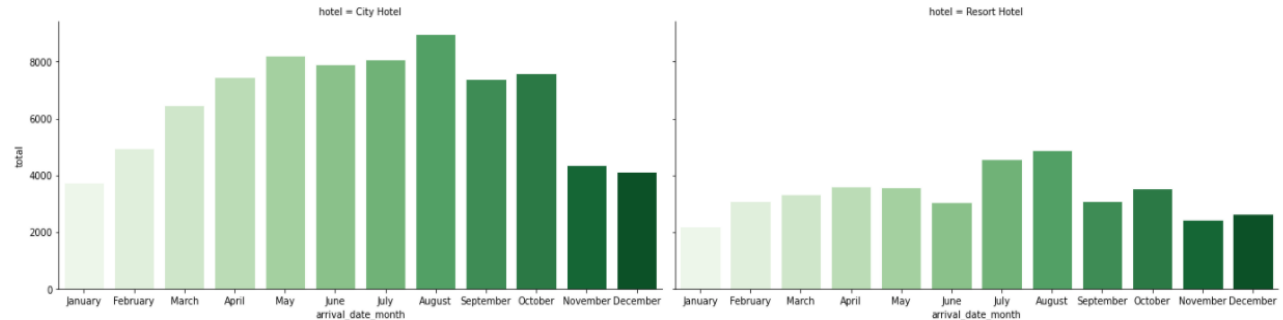


Overview of Arrival Period

- 2016 has highest number of Bookings
- May to August is the peak season
- August is the busiest month



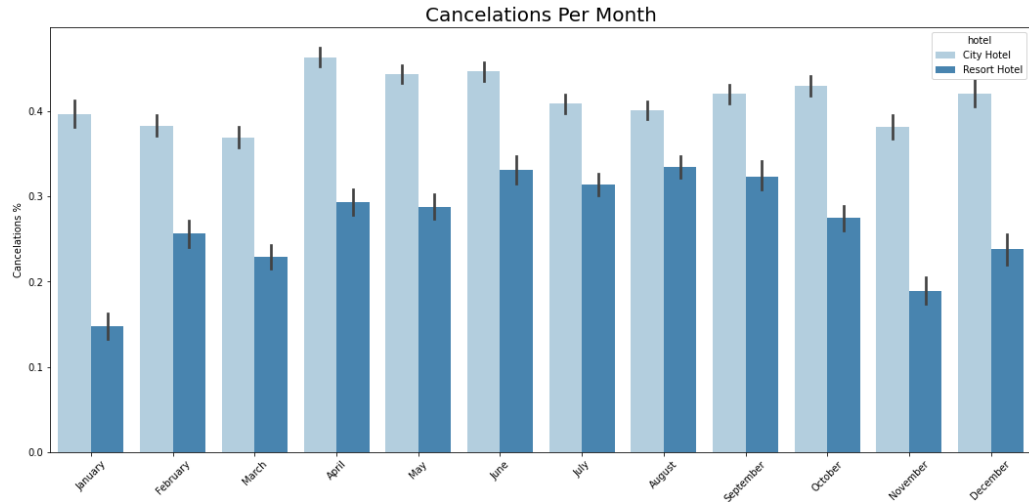
Most Occupied(Busiest) Month with Bookings



Overview Cancellations per month

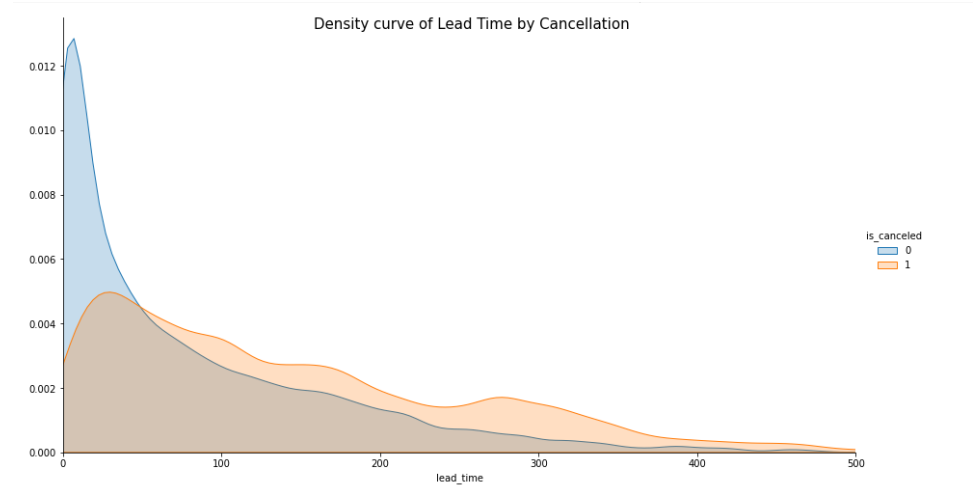
- **City Hotel**
- Cancellation per month is 40% throughout the year
- Highest in April

- **Resort Hotel**
- Cancellation highest in Summer(June, July, August)
- Lowest in winter (November, December, January)



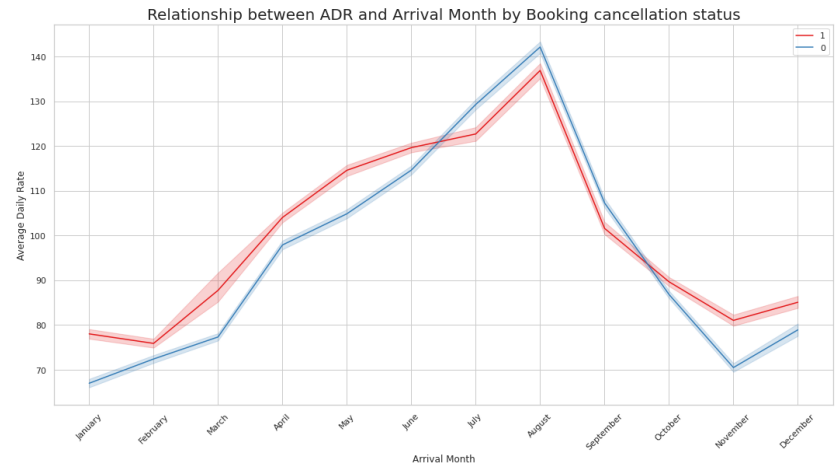
Relation between Lead time and Cancellation

- As Lead time increases,
- Cancellation increases



Overview of ADR

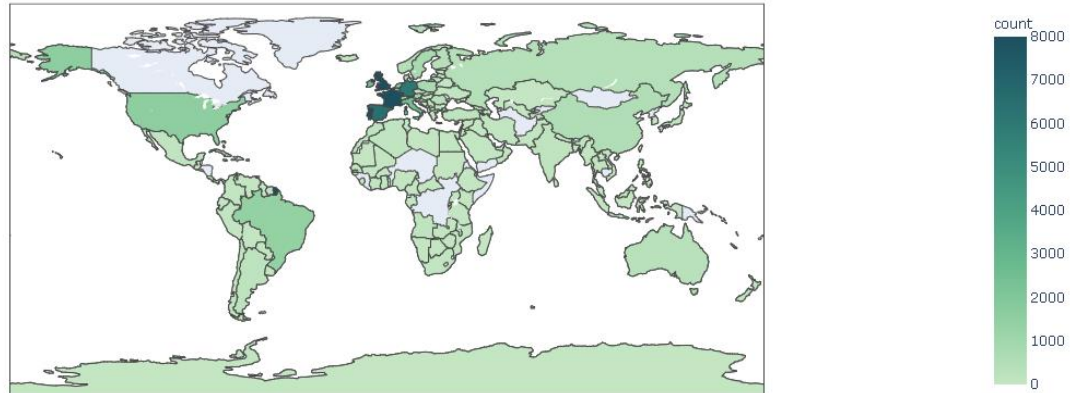
- **Resort Hotel:**
- ADR is more expensive during
- July & August
- **City Hotel:**
- ADR is slightly more during
- April, May & June
- Busiest Month : August
- Highest ADR: August
- Due to Highest ADR in August
- maybe it could be one of the
- reasons for more canceled bookings in
- August.



Which countries should be targeted for promotion?

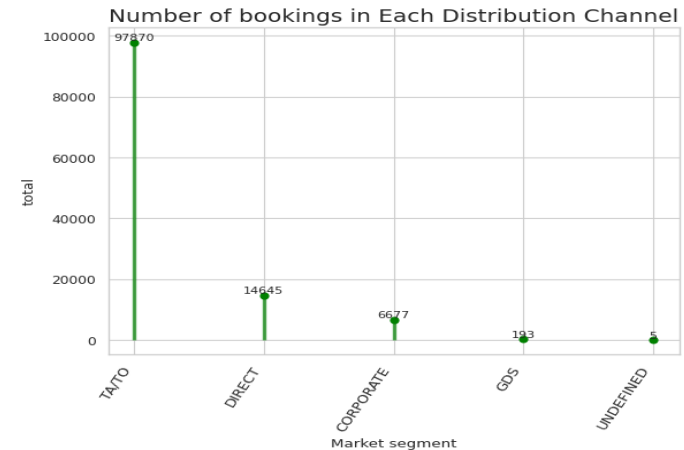
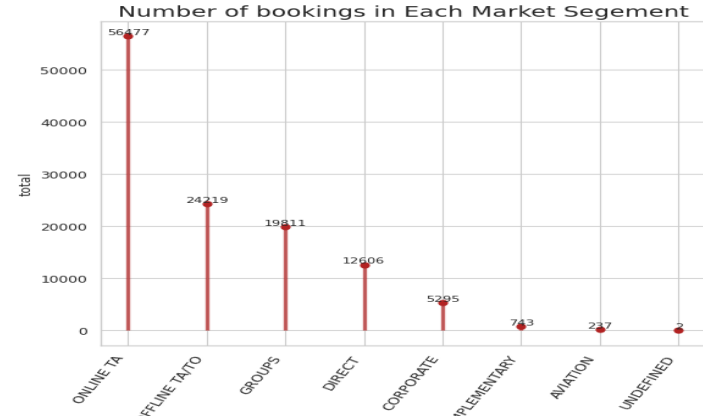
- We have a huge number of visitors from western Europe,
- namely France, UK and Portugal being the highest.
- So, this can be the target region for promotion.

Home country of visitors



Overview of market segment and Distribution channel

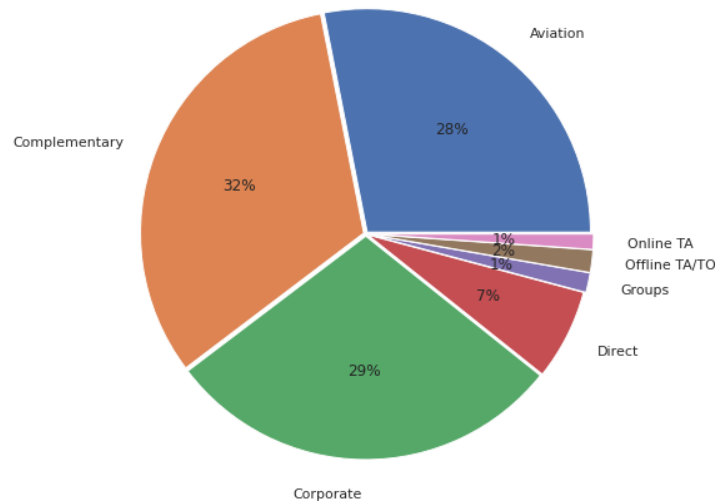
- Majority of the bookings were made through online travel agent and the travel agents whether online/offline have the highest distribution rates



Does market segment have any impact on Repeated guests?

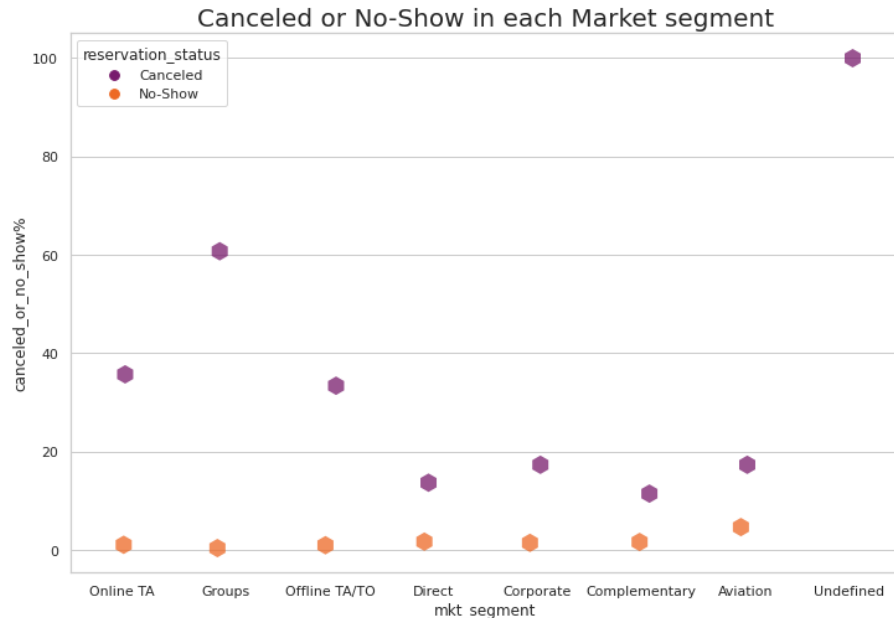
- Most number of repeated guests are from Complementary
- Complementary segment followed by Corporate
- Corporate

Percentage of Repeated guests in each market segment



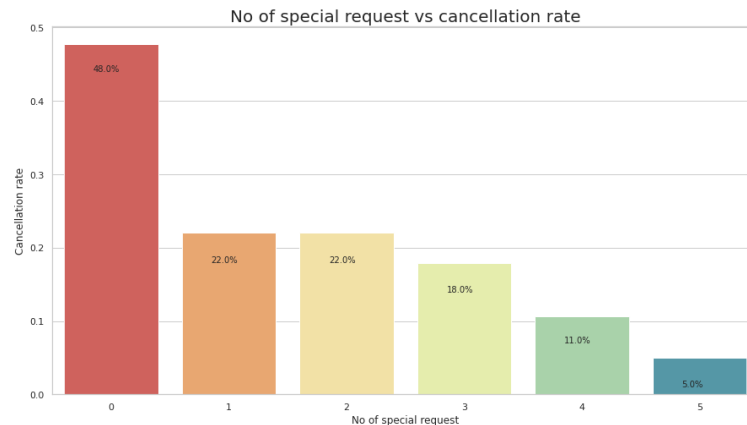
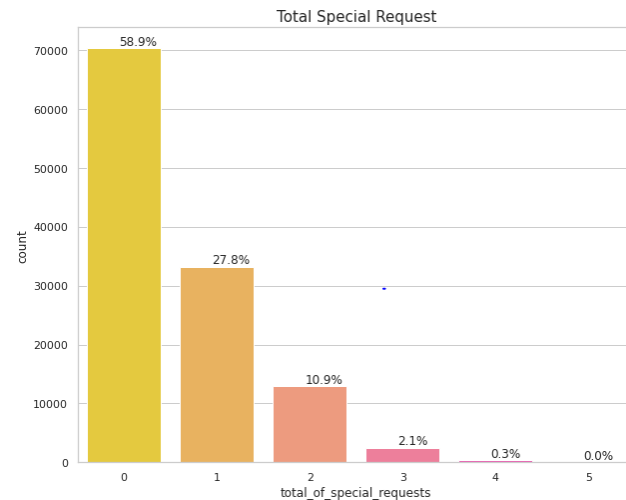
Does market segment and reservation status affect each other?

- Undefined is an outlier
- Highest rate of cancellation:
Group market segment
- Least rate of cancellation:
Direct and Complementary



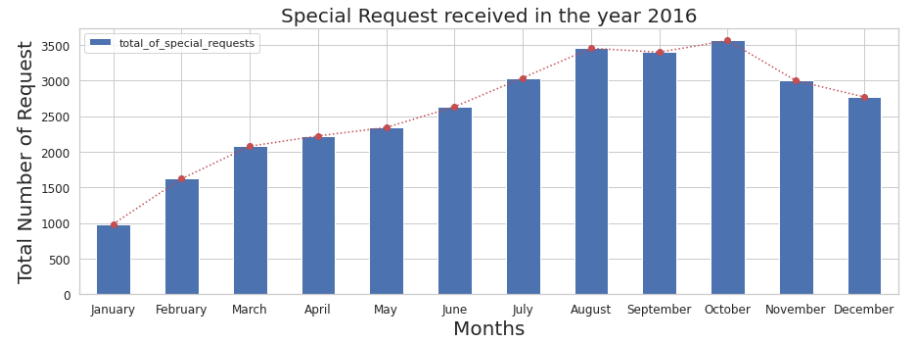
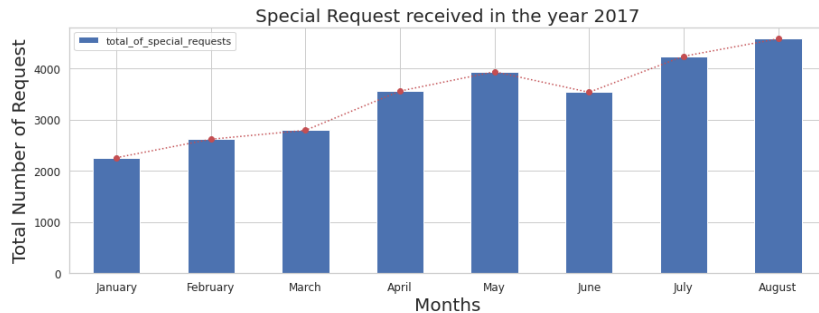
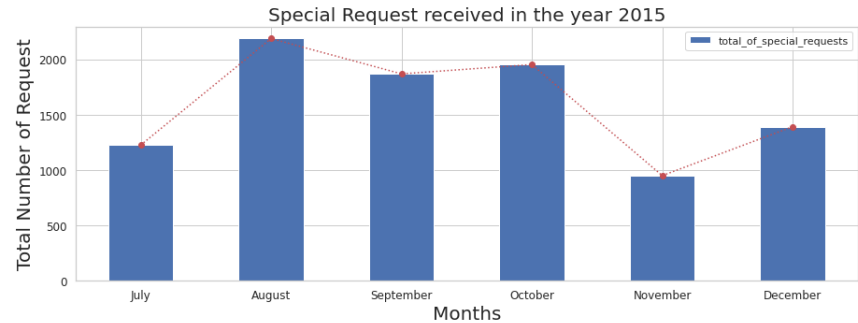
Overview of special request

- **58.9%** guests do not have any special request
- Cancellation rate is much lower if there are special requests



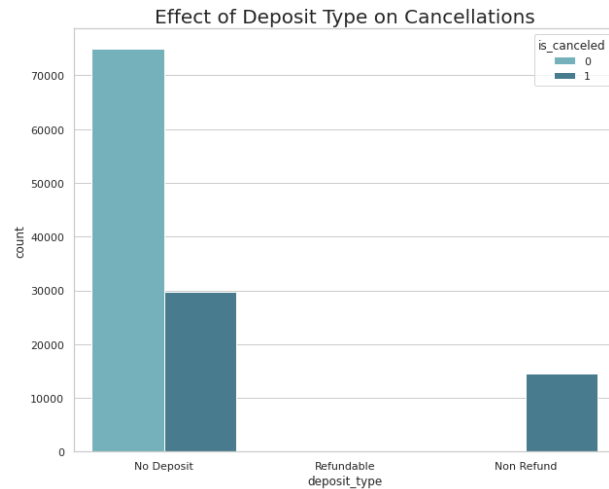
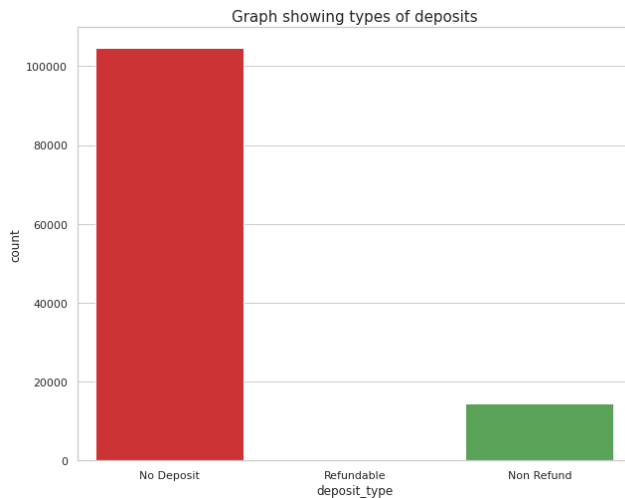
Special request is highest in which month?

- Most special request were received
- In August, September & October



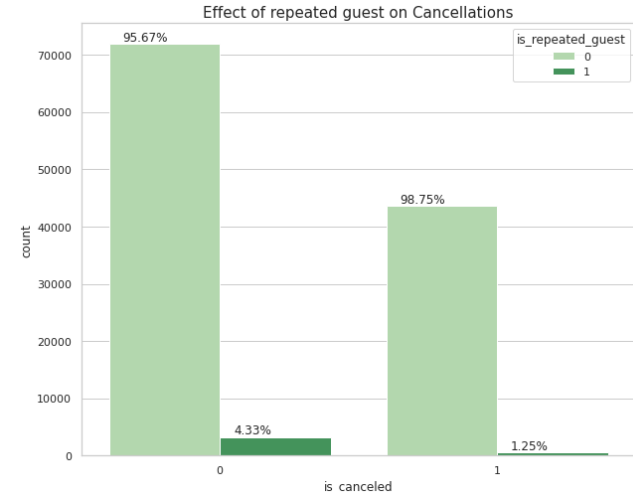
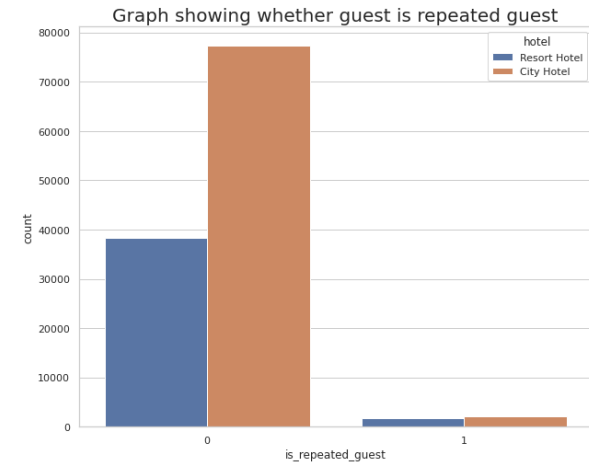
Overview of deposit type

- Majority of the bookings do not require deposit
- Most of the bookings were cancelled by guests with No Deposit



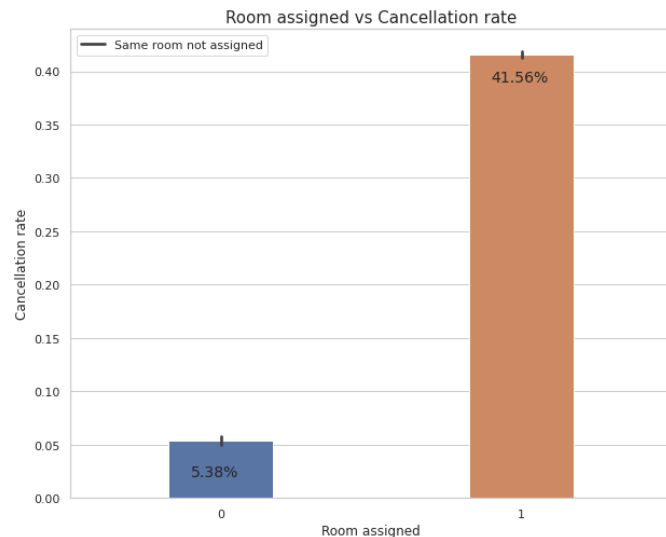
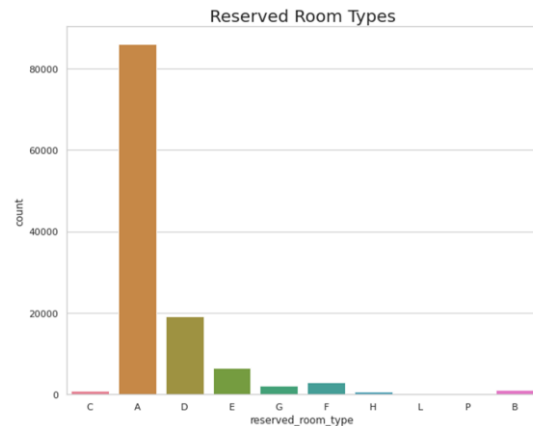
Overview of repeated guests

- Number of repeated guests are too low
- Most repeated guests do not cancel their reservation



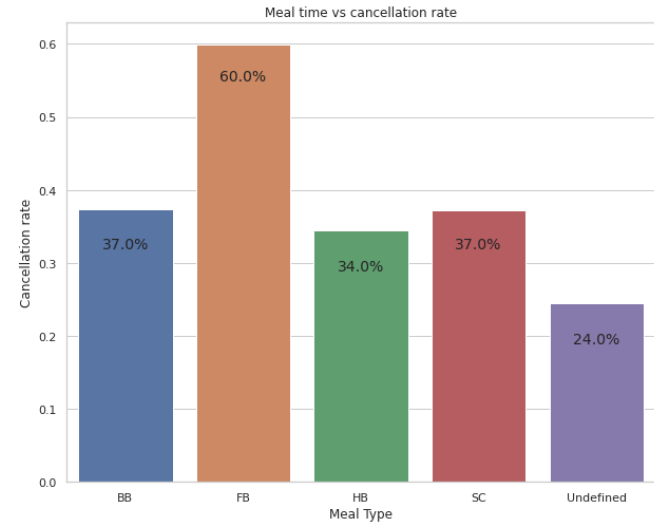
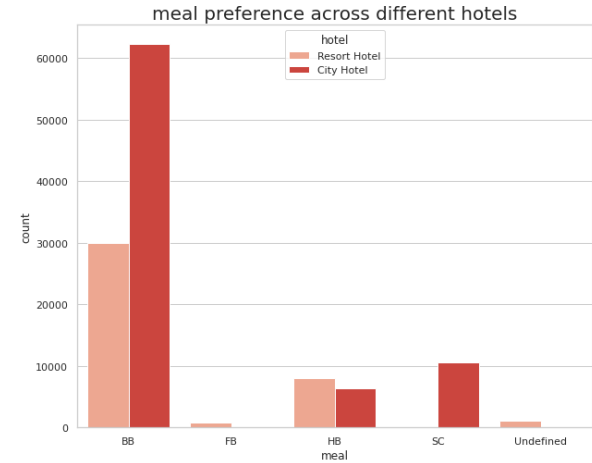
Overview of Room type

- Most preferred Room type is 'A'
- When same room was not assigned, the cancellation rate was very low.
- It implies assigning a better room
- drastically reduce the cancellation rate



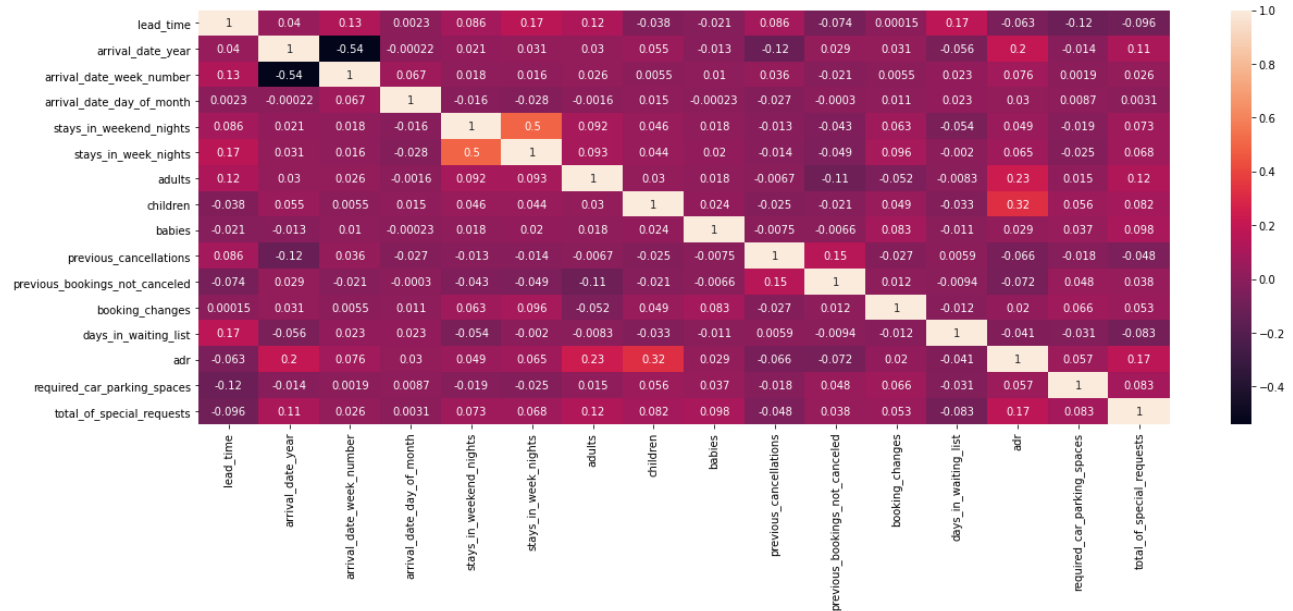
Overview of meal type

- Most preferred meal is Bed & Breakfast
- Full Board booking has highest rate of Cancellation (60%)



Correlation using HeatMap

- arrival_date_week_number and arrival_date_year are 54% negative correlated
- adr and children are positively correlated by 32%
- stays_in_week_nights and stays_in_weekend_nights is 50% positively correlated



Challenges

- 1. There were lot of Null values that had to be handled carefully
- 2. There were huge number of columns, so it was difficult to draw relationship between various columns and come on a conclusion

Conclusion and Summary

- Maximum booking is for city hotels
- Peak season is between May to August.
- Most special request were received In August, September & October and cancelations are low if there is a special Request
- Majority of the guests are from Western Europe
- Number of repeated guests is too low, but most repeated guests do not cancel bookings. Majorly repeated guests are from Complementary market segment
- Majority of the visitors travel in pair
- Most of the guests stay up to 4 days
- Bed & Breakfast option is the most popular meal, Full Board option has highest cancellation
- 37% of the bookings were Canceled
- If lead time < 100 days there are fewer chances of booking getting canceled
- Room type 'A' is most preferred but changing the Room type decreases cancellation rate
- Grouped market segments have a very high cancellation rate
- Most of the bookings were canceled by guests with no deposit