CAPSTONE PROJECT -1

Hotel Booking Analysis

Team Members

Abhash Jain

Mounika Dontula

Karan Rawat

Points to Discuss:

- Agenda
- Data summary
- Data wrangling
- Data analysis
- Data visualisation
- Summary

Agenda

To discuss the analysis of given hotel bookings data set from 2015-2017.

We'll be doing analysis of given data set in following ways :

- Booking wise analysis.
- Guest wise analysis.
- Type of visitors.
- Month wise analysis.
- Room wise analysis.

By doing this we'll try to find out key factors driving the hotel bookings trends.

Data Summary

- The data set contains booking information of city hotel and resort hotel the dataset has a shape of (119210,32) which means the dataset contains 119210 rows and 32 columns.
- The data set contains booking information of city hotel and resort hotel. It contains the information like hotel type , when the booking was made , room type , revenue ,length of stay , lead time etc. Among other thing personal information has been deleted from the database.
- The picture represents the data of booking.

```
x(['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', 'agent',
  'company', 'days_in_waiting_list', 'customer_type', 'adr',
  'required_car_parking_spaces', 'total_of_special_requests',
  'reservation_status', 'reservation_status_date'],
```

Data Wrangling

Data Cleaning

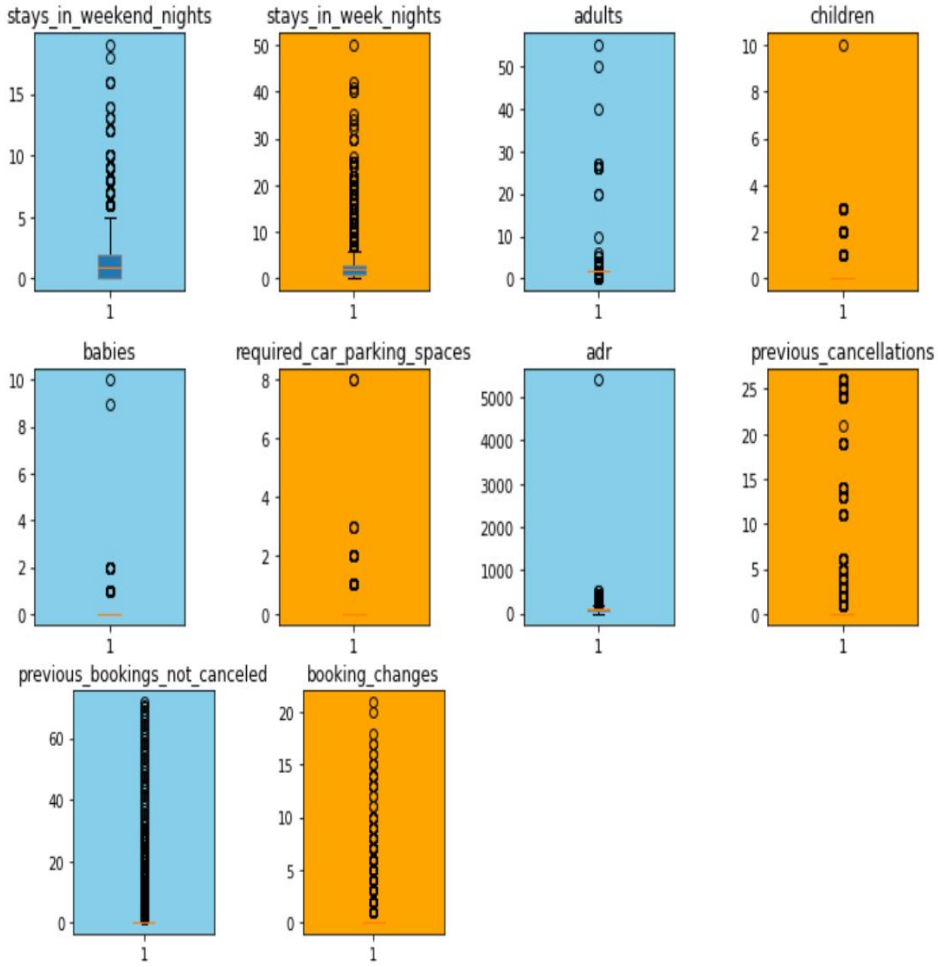
- The Data file consists of some null values “Nan”.
- Replacing those null values with zero, median and mode.
- Checking for outliers.

Data Preparation

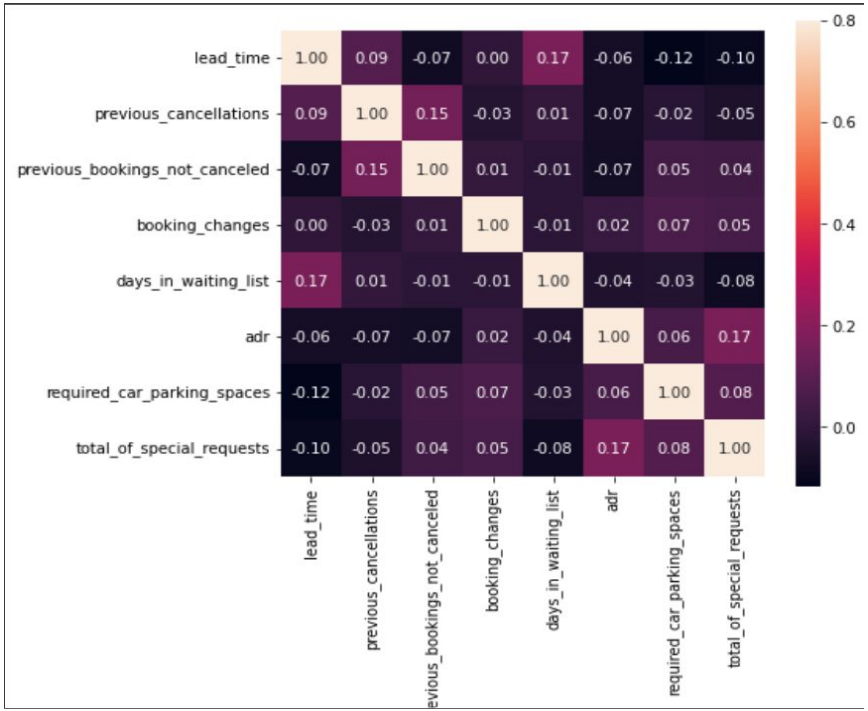
- Data file consists of different types of data.
- Data type: integer , float , objects.
- Dropping some of the rows which don’t have values.

Checking for outliers

- In the dataset there are binary features like 'is_canceled', 'is_repeated_guest' which are mapped to float data type. There are outliers also, as we can see mean and median difference is quiet large for most of the features.
- We have selected certain columns to check for outliers.
- We made a subplot with 3 rows and 4 columns.
- Plotting each feature's boxplot to check outliers.
- Removing left out blank subplots.



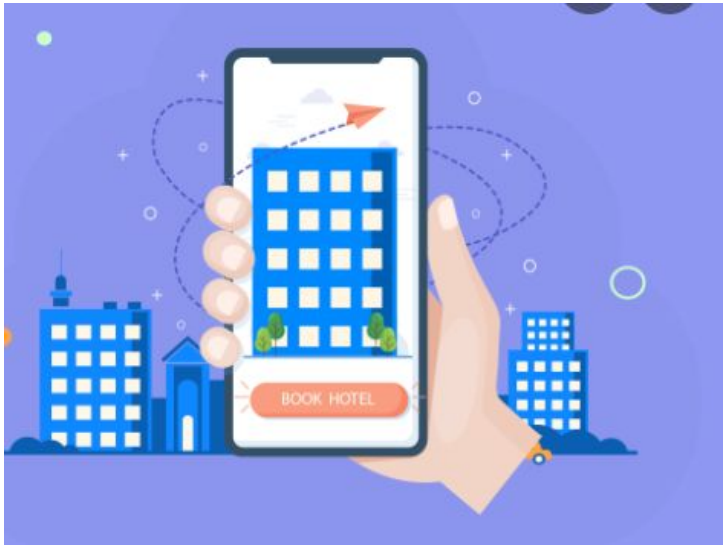
Correlation With Heatmap



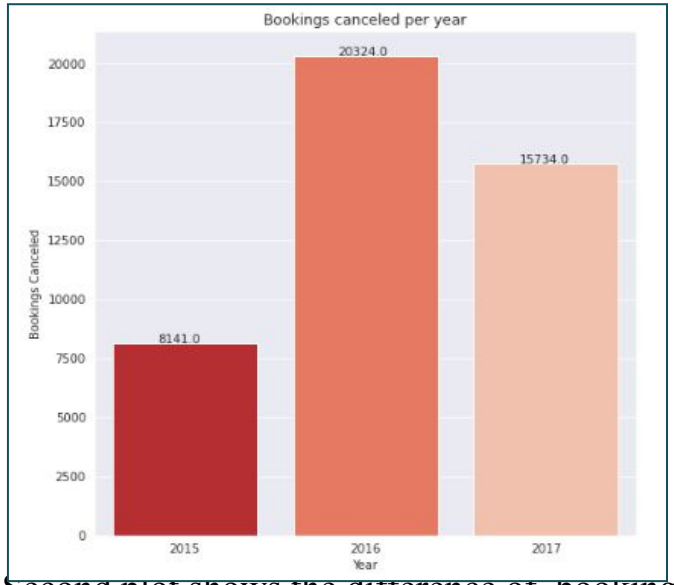
- We can see that days_in_waiting_list is slightly corelated with lead_time.
- adr is corelated to total_of special_requests.

Booking wise analysis

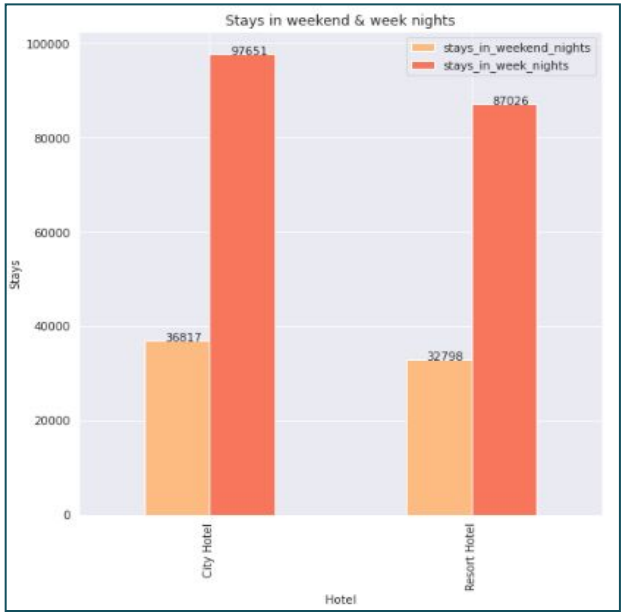
- How many booking cancelled each year?
- What is the booking difference between weekends and week days night?
- From which market segment bookings done the most?
- What is the booking percentage difference between city hotel and resort hotel?
- What are the bookings percentage each year?



Booking wise analysis...

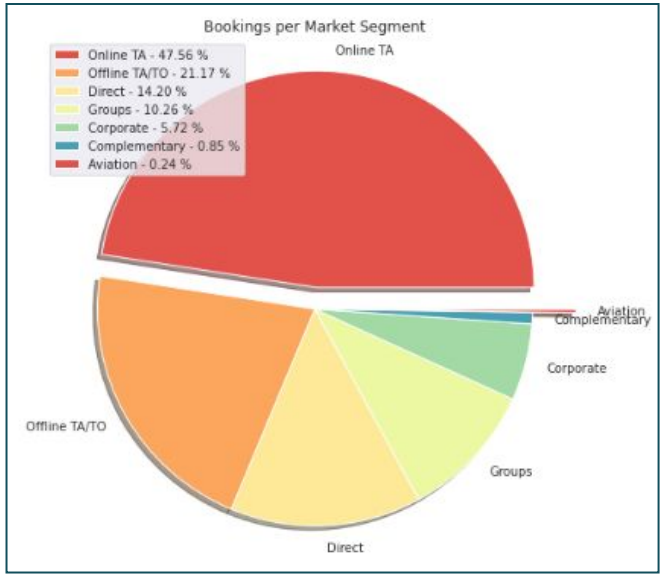


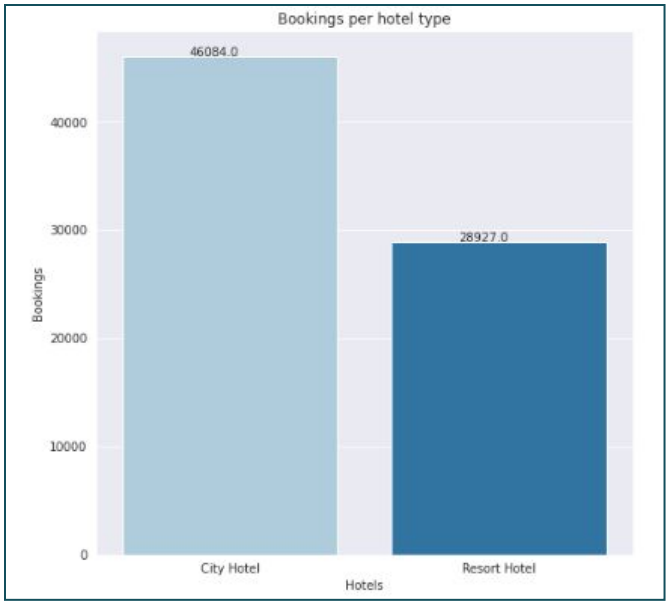
- The 37% of bookings got cancelled.
- In 2015, the 8141 bookings were cancelled.
- In 2016, the 20324 bookings were cancelled.
- In 2017, the 15734 bookings were cancelled.



- Second plot shows the difference of booking between weekends and week days night stays in hotel.
- Total guests stays in week nights are 184677 in that 97651 are from city hotel and 87026 are resort hotel.
- Total guests stays in weekend nights are 69615 in that 36817 stays in city hotel and 32798 stays in resort hotel.

Booking wise analysis...



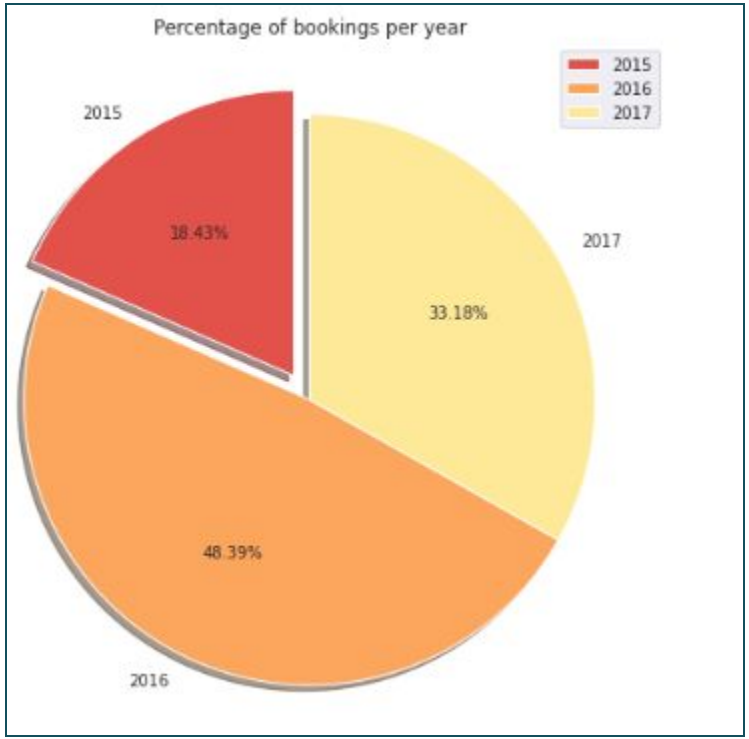


- This plot shows the bookings per market segments.
- We can see that most of the booking are done through online TA i.e. 47.56%
- The least booking done through complementary and aviation.

- This plot shows the booking percentage between city hotel and resort hotel.
- More than 60% of the population booked the City hotel i.e. 46084 .

Booking wise analysis...

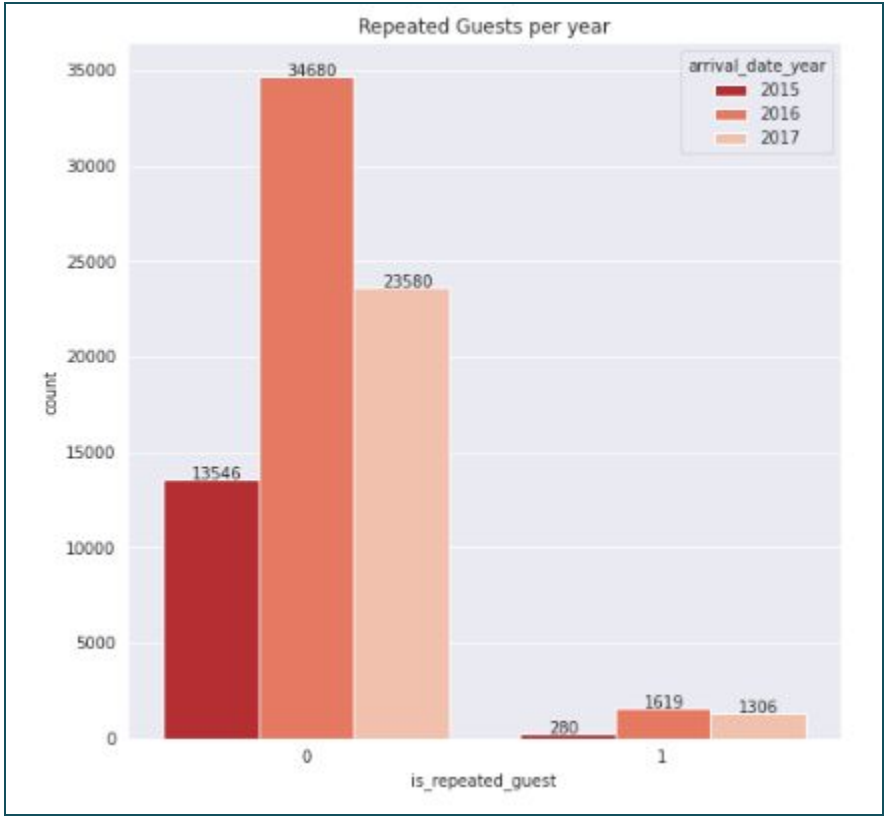
- This figure shows that the percentage of booking per year.
- More bookings were made in year 2016, compared to the previous year. But the bookings decreased by almost 15% the next year.



Guest wise analysis

How many guests repeated each year?

- This figure shows that are there any repeated guests through years.
- 1 means guest repeated and 0 means guest not repeated.
- In the year 2015 - 280 guests were repeated.
- In the year 2016- 1619 guests were repeated.
- In the year 2017- 1306 guests were repeated.
- Highest guests were repeated in the year 2016.

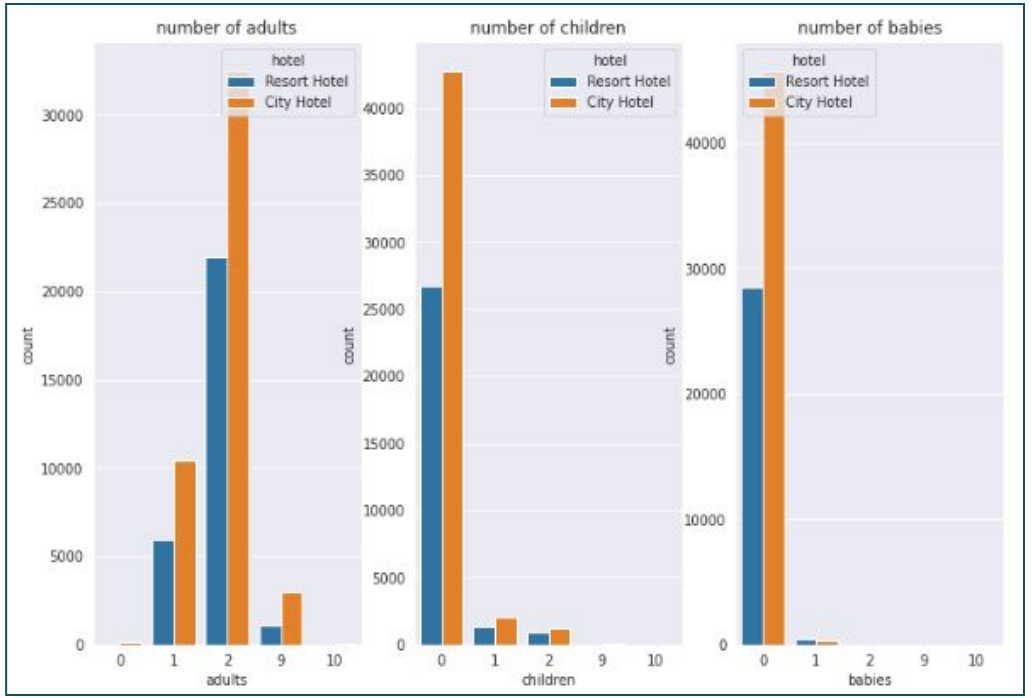


Type of visitors wises analysis

- Which is the most booked accommodation type?
- From which country visitors comes the most?

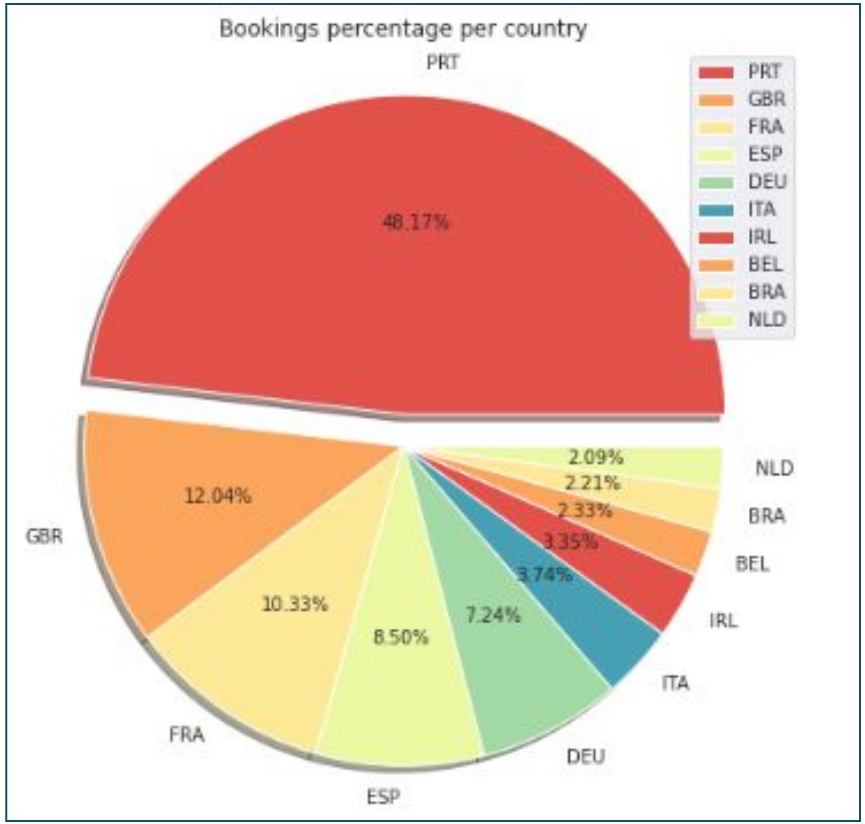
Type of visitors wises analysis...

- This graph shows the type of visitors those are adults , children and babies.
- Most of the visitors travel in pairs. They mostly prefer city hotel.
- Visitors with children are very few and they mostly prefer city hotel
- Visitors with babies prefer resort hotel for their comfort.



Type of visitors wises analysis...

- This pie chart shows the booking percentage with respect to country code.
- We can see in the chart majority of the visitors are from country Portugal..
- The countries UK , France, Spain and Gremany also holds a great portion in bookings .
- The approx. 70% comes from these 5 countries.
- The least visitors are from country Netherlands.

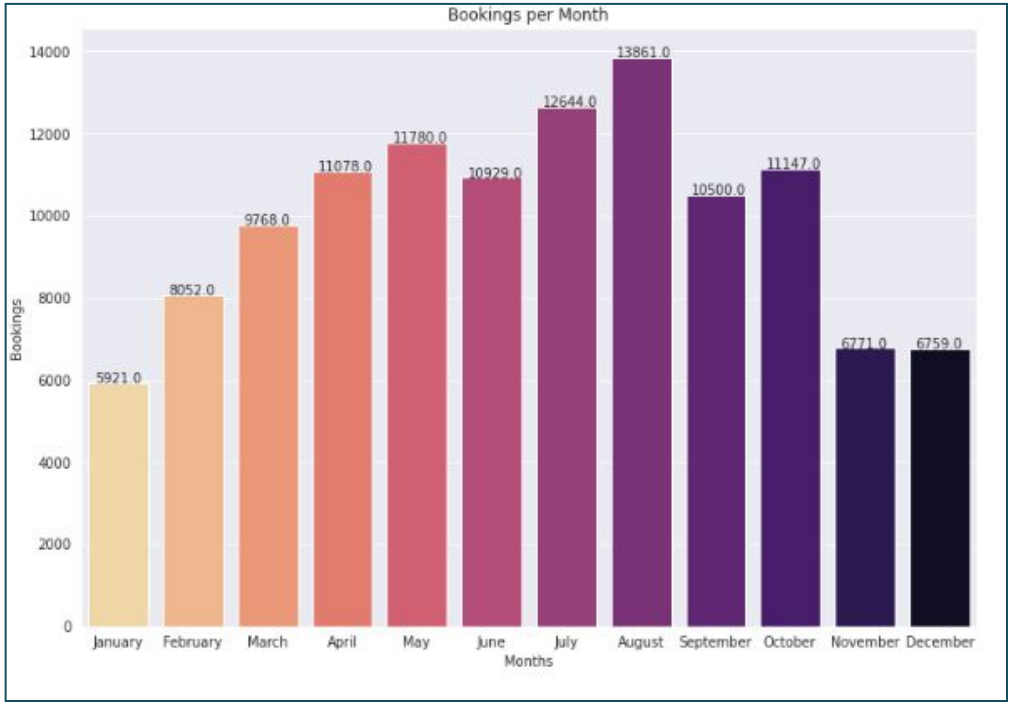


Month wise analysis

- Which is the most occupied month for hotels?
- What is the average daily rate for each month per hotel type?

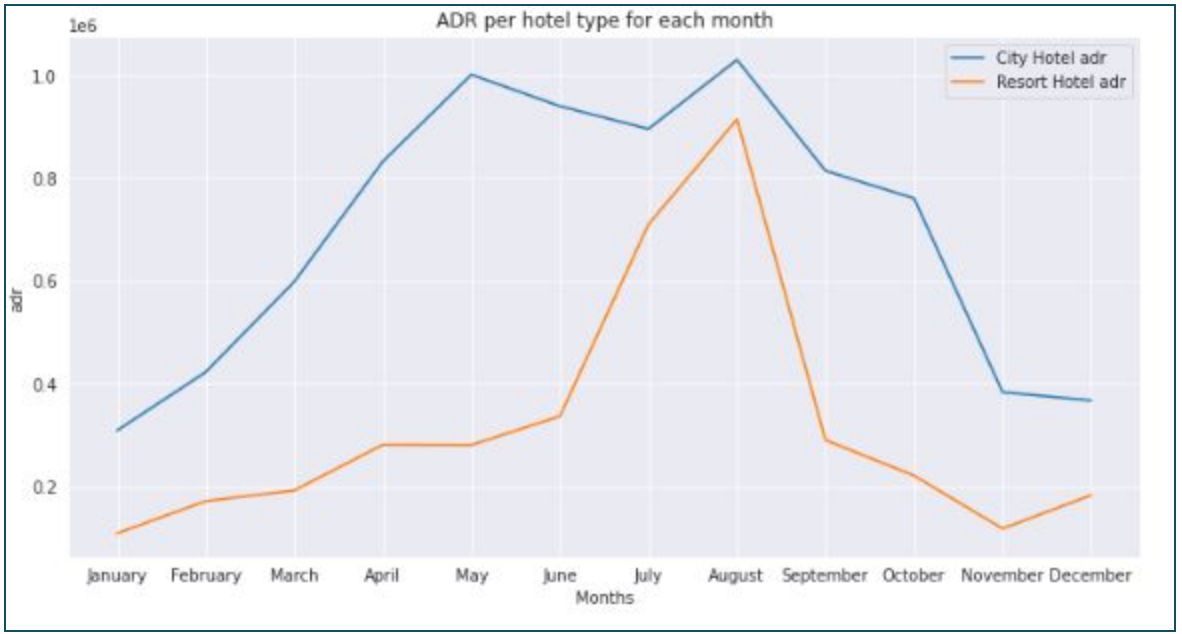
Month wise analysis...

- This bar chart shows booking for each month.
- As we can see in below chart most of the bookings were made from July to August.
- And the least bookings were made at the start and end of the year.



Month wise analysis...

- This line chart shows ADR for hotels for each month.
- The ADR for City Hotel is highest for the months May and August.
- The ADR for Resort Hotel is highest for the August month.
- The ADR for City Hotel is more expensive than Resort Hotel for each month.

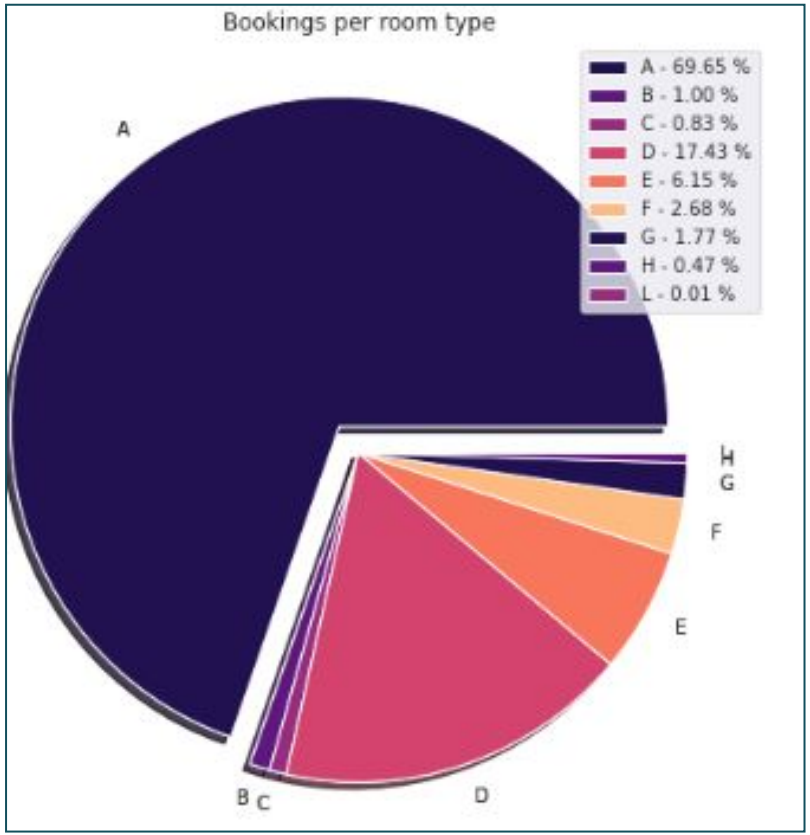


Room wise analysis

- Which room type have the most demand?
- How many rooms wrongly assigned to with respect to booked room type by each hotel?
- Which room type generates highest ADR?

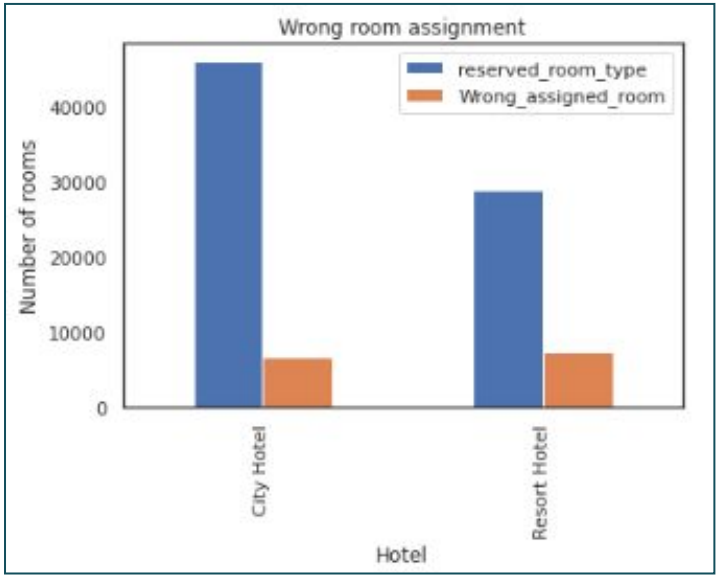
Room wise analysis...

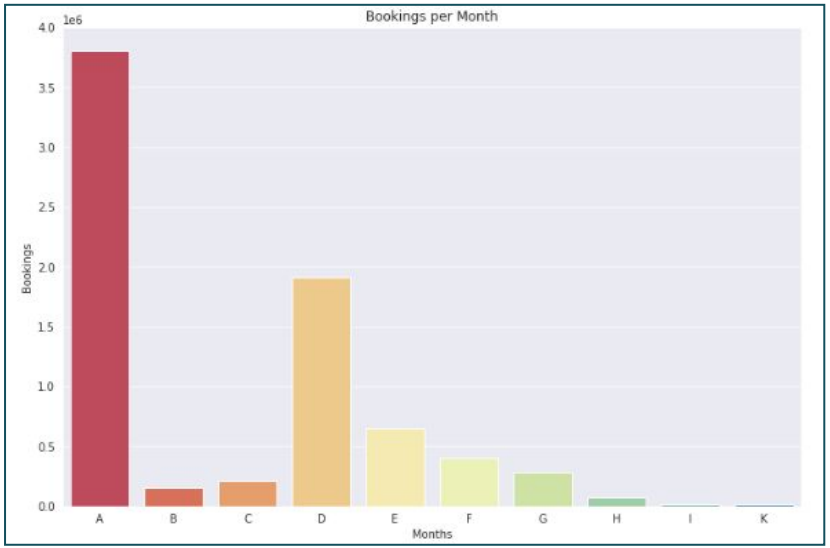
- This pie chart shows the highest booking for room type.
- We can clearly visualise that room type A had more demand compared to other .
- After A , D gets the leading .
- Room type H has the least booking rate.



Room wise analysis...

- The bar chart on the right side shows the wrong room assigned with respect to booked room type.
- In resort hotel 7334 room assigned wrong i.e. 25.4% of total r eserved room type in resort hotel.
- In city hotel 6661 room assigned wrong i.e.14.5% of total res erved room type in city hotel.





- The chart on the left side defines the room type which generates high ADR.
- We can see that room type a has the highest ADR.
- Next to A type room D room type has highest ADR.

Summary

- More than 60% of the population booked the City hotel.
- Total bookings got cancelled 37% of total booking. Most of the booking cancelled for City hotel during the year of 2016 and 2017 that is 61% of total booking cancelled.
- Most of the bookings were made in the year 2016 compared to other years.
- Most bookings were made by online TA market segment
- More repeated guest bookings were made in 2016, compared to the previous year. But the bookings decreased by almost 15% the next year.
- Most of the wrong room assigned was with resort hotel compared to city hotel.
- Majority of visitors travel in pairs.
- The visitors from Portugal has highest booking rate prior to countries UK, France, Spain and Germany also holds a great portion in bookings.
- Most bookings were made from July to August. And the least bookings were made at the start and end of the year.

Challenges

- Data was present in wrong data type format.
- Choosing appropriate visualization techniques to use was difficult.
- A lot of null values were there in the dataset.

Thank You