Capstone Project

Hotel Booking Analysis

Mohammed Ghousuddin

Points to discuss:

- Agenda
- Data summary
- Hotel wise Analysis
- Hotel Bookings month wise
- Hotel Bookings year wise
- Most Busiest months of the hotel
- Average price per room
- Correlation Heatmap

Agenda

The analysis of given hotel bookings data set from 2015-2017.

I’ve done analysis of given data in following ways:

- Hotel wise Analysis
- Hotel Bookings month wise
- Hotel Bookings year wise
- Most Busiest months of the hotel
- Average price per room
- Correlation Heatmap

Data Summary:

Given data set has different columns for hotel bookings.

1. **hotel:** type of hotels
2. **is_canceled:** canceled or not
3. **lead_time:** no. of days before actual arrival in the hotel
4. **arrival_date_year:** year of booking
5. **arrival_date_month:** month of booking
6. **arrival_date_week_number:** week number of the year in which booking
7. **arrival_date_day_of_month:** arrival month date
8. **stays_in_weekend_nights:** no. of weekends guest stayed
9. **stays_in_week_nights:** no. of weekdays guest stayed
10. **adults**
11. **children**
12. **babies**
13. **meal:** BB – Bed & Breakfast
    HB – only two meals including breakfast meal
    FB – breakfast, lunch, and dinner
    SC : Self-catering
    Undefined: Rooms only packages without meals

Data Summary(contd..)

15. **market_segment: TA: Travel agents**
    TO: Tour operators
16. **distribution_channel**
17. **is_repeated_guest**
18. **previous_cancellations:** cancellation in past
19. **previous_bookings_not_canceled:** not cancelled in past
20. **reserved_room_type**
21. **assigned_room_type**
22. **booking_changes**
23. **deposit_type**
24. **agent**
25. **company**
26. **days_in_waiting_list**
27. **customer_type**
28. **adr:** average daily rate
29. **required_car_parking_spaces**
30. **total_of_special_requests**
31. **reservation_status**
32. **reservation_status_date**

Libraries use

- import numpy as np
- import pandas as pd
- import seaborn as sns
- import matplotlib.pyplot as plt
- %matplotlib inline

Data Cleaning

Data Cleaning is one of he crucial steps while analyzing any dataset. It is generally performed beforehand, to get quality insights.

Null/Missing Value Treatment

In the real scenarios, data always comes with imperfection or partialness. These are called Null-values in the data. Treating them well should be our first priority, before deploying any operation on them.

These null values can be treated by either dropping all the null value rows or filling them with average or most preferably with the mode value of the column. I will prefer the later process by using fill.na() command.

Removing unwanted columns

During Picking up Right variables, I decide to drop a column which is not a part of our problem statement. This is generally done by subsetting a dataframe, under the cases of data with billions and millions of rows and thousands of columns.

In this case, working on a considerably small dataset with just 29 columns.

Data Manipulation is the modification of information to make it easier to read or more structured.

Data Visualization is the graphical representation of information and data. By using visual elements like charts, graphs and maps.

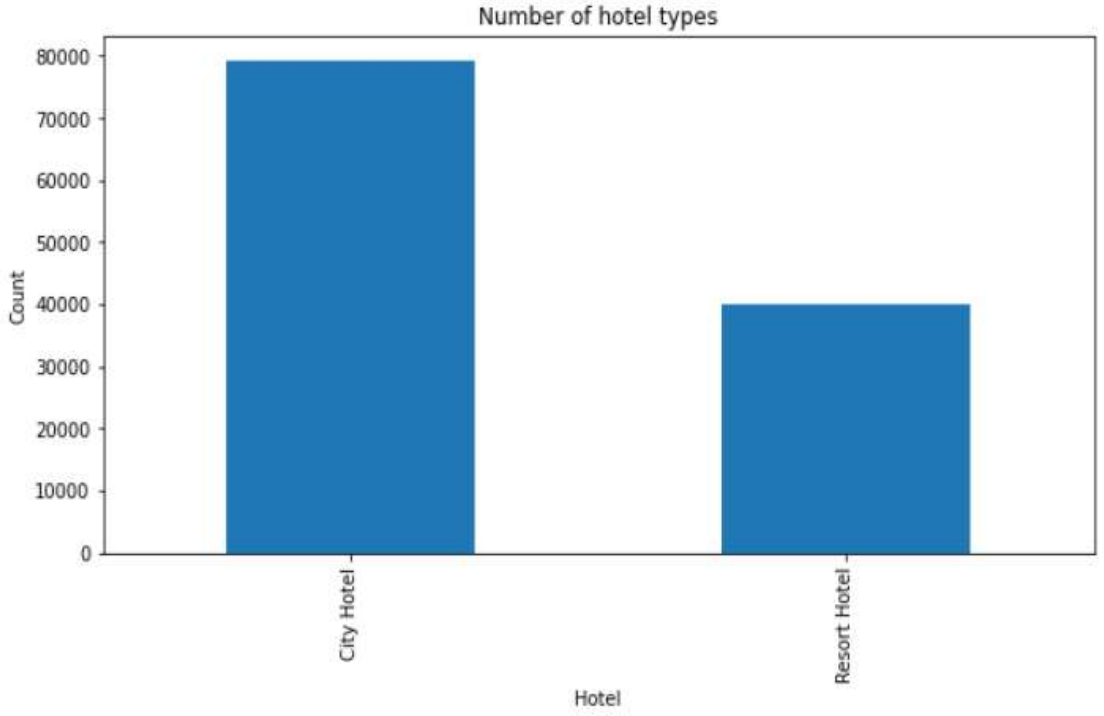
Primary Data Visual Comparison using Matplotlib

Moving ahead, I would have to see the shares of each hotel in the analysis. Let's use the Data Visualization tool i.e., Matplotlib (alternatively seaborn can also be used), for plotting the shares of each hotel data.

Mainly performed using Matplotlib and Seaborn library and the following graph and plots had been used:

- Bar Plot.
- Seaborn plot
- Line Plot.
- Box Plot
- Heatmap.

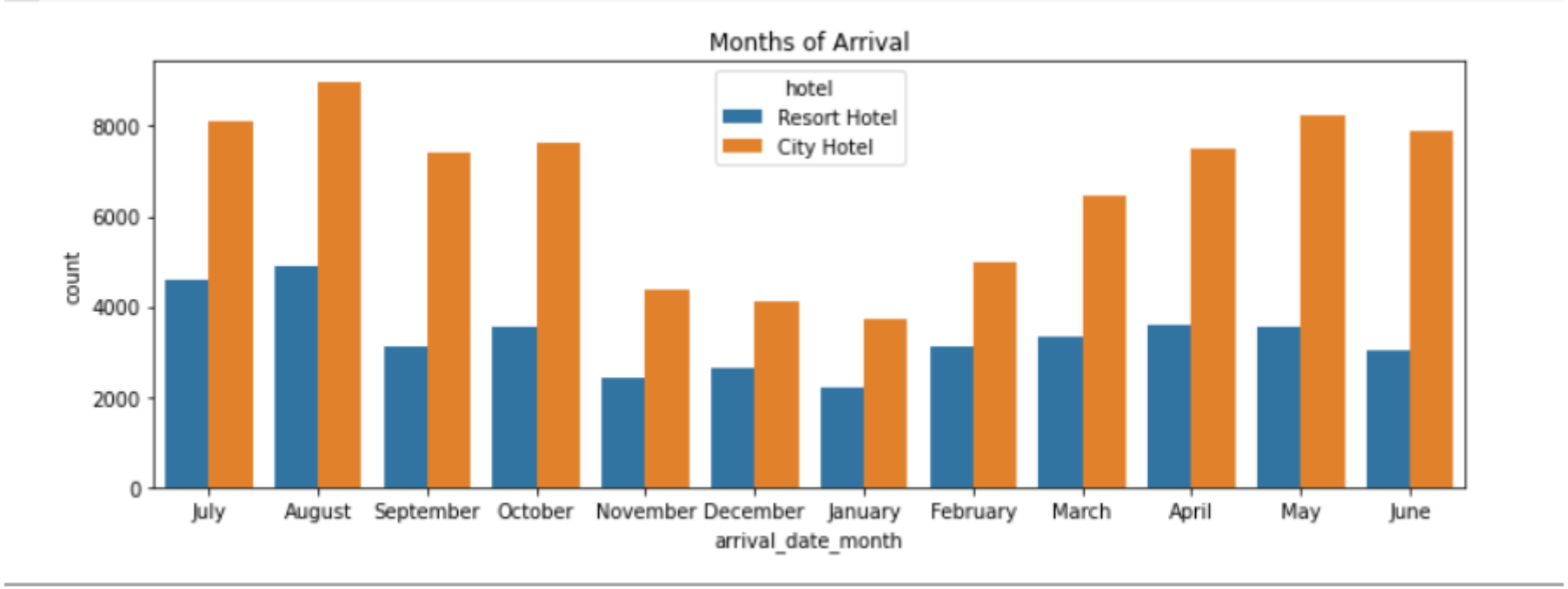
Types of Hotel plot



Observation

City hotel have greater bookings than Resort hotel

Analyzing month – wise bookings against hotel type



Observation

City hotel have greater bookings than Resort hotel in every month

Analyzing year – wise bookings against hotel type

Observation

More than double bookings were made in the year 2016, compared to the previous year.

But the bookings decreased in next year 2017

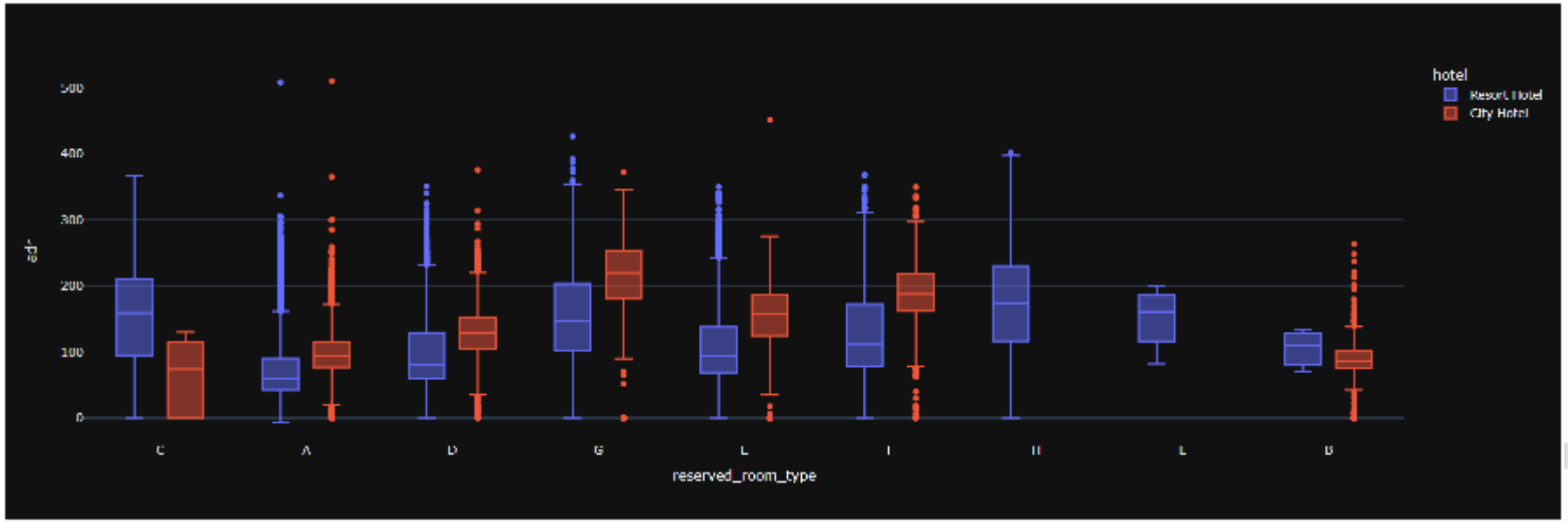
Analyzing most busiest months

Observation

Both city and resort hotel bookings are peak in August and lowest in January.

From March to October it seems to be busiest for the year. Bookings falls down from November to January

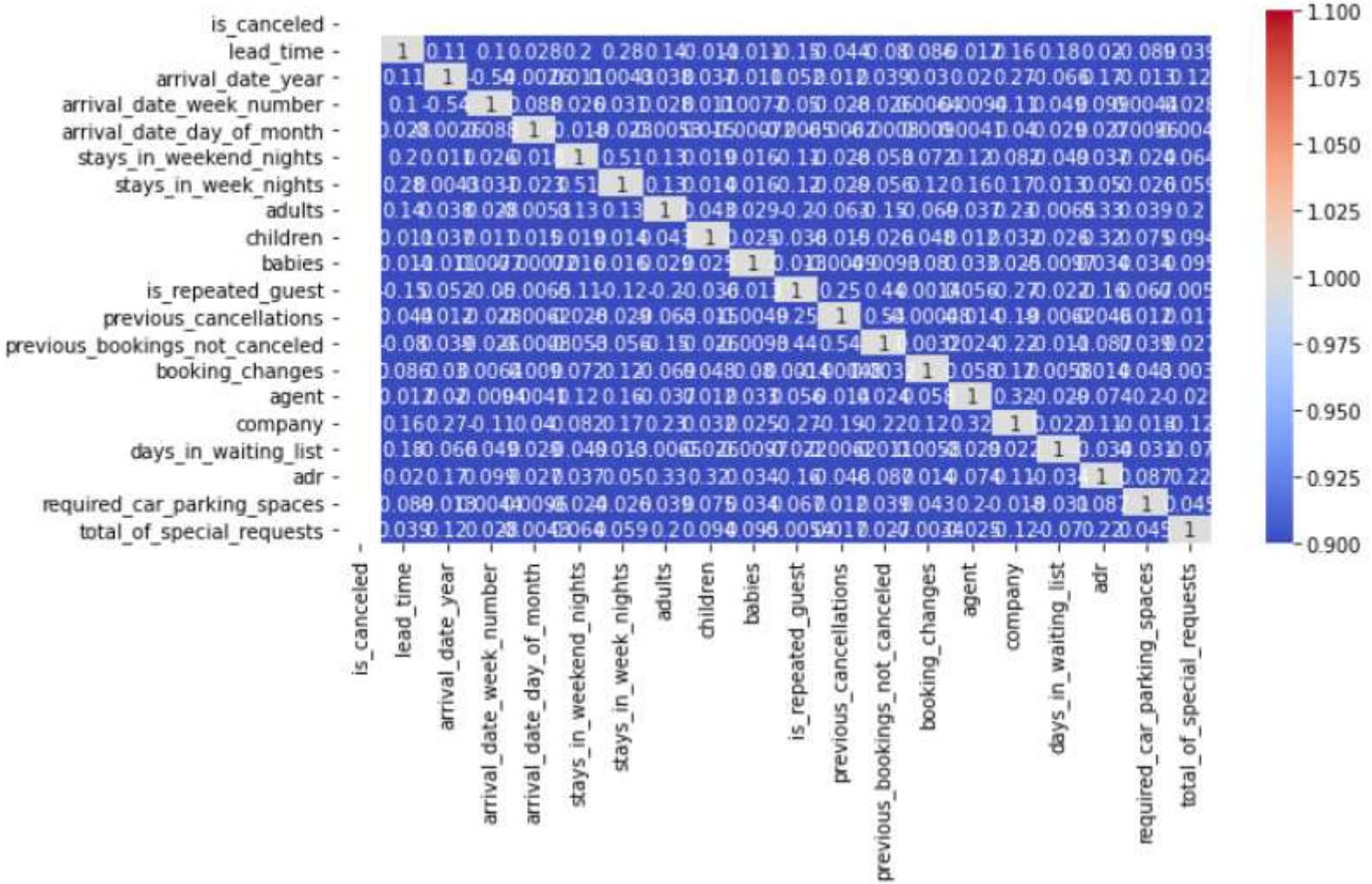
Analyzing average price per room depends on its type



Observation

The average price per room depends on its type and the standard deviation shown using box plot

Correlation Heatmap



Challenges

1. Dataset contains a lot of duplications.
2. Against few columns having a lot of Null values.
3. Few dataset columns with wrong datatype format.

Conclusions:

1. City hotel bookings has twice the amount of reservations as compared to Resort hotel booking over same period year
2. People are booking city hotels more than Resort hotels. In which month the people book the hotel
3. More than double bookings were made in the year 2016, compared to the previous year. But the bookings decreased in next year 2017
4. Both city and resort hotel bookings are peak in August and lowest in January. From March to October it seems to be busiest for the year. Bookings falls down from November to January.
5. The average price per room depends on its type and the standard deviation
6. Correlation Heatmap

Any Questions

Thank You