PROJECT
“Hotel Booking Analysis”

A PRESENTATION SUBMITTED TO

SUBMITTED BY

DEEPAK KUMAR SINGH

POST@RATHORE@GMAIL.COM

https://github.com/24deepak/Hotel-Booking-System

PROJECT

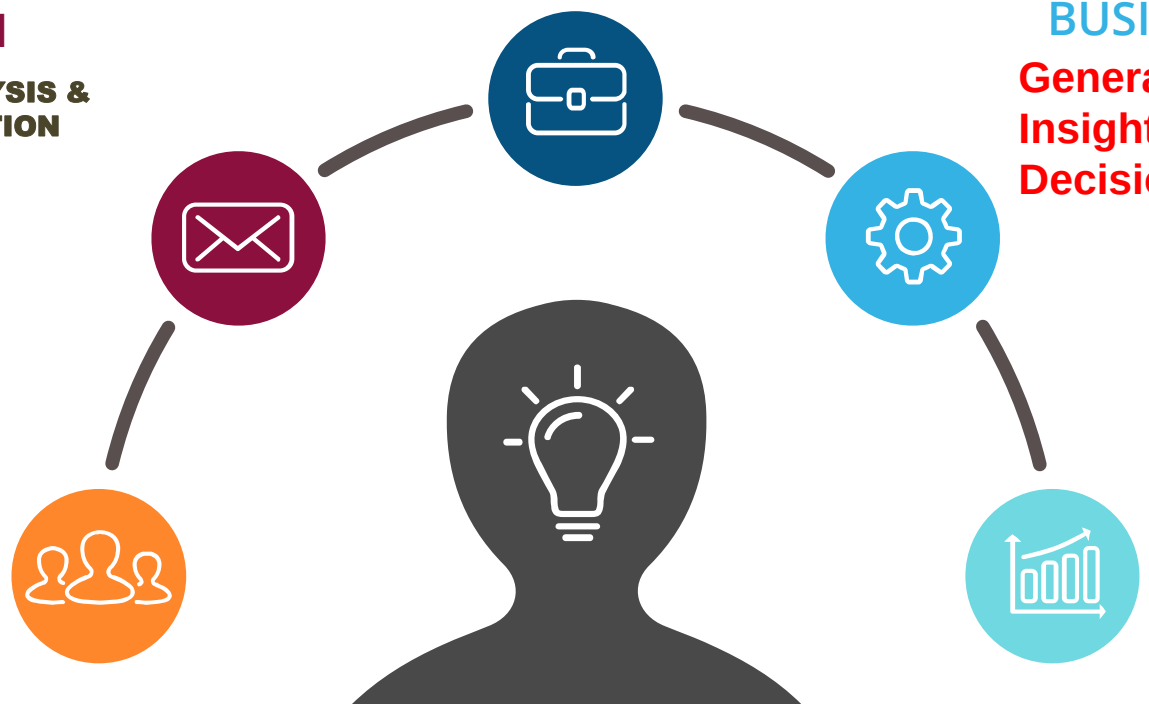
Hotel Booking Analysis

PLAN

DATA ANALYSIS & VISULISATION

BUSINESS

Generate Business Insights For Better Decision



TEAM

Deepak Kumar Singh

SUCCESS

Abstract

- This project is all about the "Room Booking" in all condition. After completing the project , one will able to understand the concept of Hotel booking System.

- This data describes two datasets with hotel Booking data. One of the hotel is a resort hotel and the other is a city hotel.

- Both datasets share the same structure, with 31 variables describing the 40,060 value of resort hotel and 79,330 value of city hotel. Each value represents a hotel booking.

- Both datasets comprehend bookings due to arrive between the 1st of July of 2015 and the 31st of August 2017, including bookings that effectively arrived and bookings that were cancelled.

- Since this is hotel real data, all data elements relevant hotel or costumer identification were deleted.

- Due to the availability of real business data for scientific and educational purposes, these datasets can have an important role for research and education in revenue management, AI learning, as well as in other fields.

Problem Statement

- ❖ Main target is conduct EDA on the named dataset and fetch out informative modern interpretation about well established general trends in hotel bookings system to make decision.
- ❖ As analyse the dataset, my main area to analyse why cancelation is taking place -----
- what factors like waiting time, price, meal etc involves for the cancelation
- and how we can optimize the cancelation in a better way.

ANALYSIS LANGUAGE/PLATEFORM UESD-

- **Python Programming language**
- **Python Library-**
- **- NumPy**
- **- Pandas**
- **-matplotlib**
- **-Google Colab**

Analysis/Visualisation

POINT OF EDA:

1.From which country the guest come most?
2.Which month has highest booking month and which one lowest booking month?
3.Year wise booking of hotel type.
4.How many bookings were cancelled?
5.Is the lead time directly connected with cancellation.?
6.How many customer are there who have repeatedly cancelled the bookings?
7.Which month has the highest number of cancelations and lowest cancelations month?
8.What is Bookings number by Market segment.?
9.Distribution by cancellation by market segments.
10.Deposit Vs Cancellations by market segments.?
11.How does the price per night vary over the year?

Dataset Variables

The dataset contain three years – (2015, 2016 and 2017) information.

- **is_canceled:**Binary variable indicating whether a booking was cancelled or not.
- **lead_time:**Number of days between booking date and arrival date
- **arrival_date_week_number, arrival_date_day_of_month, arrival_date_month :**Week number, day date, and month number of arrival date
- **stays_in_weekend_nights, stays_in_week_nights:** Number of weekend nights (Saturday and Sunday) and weeknights (Monday to Friday) the customer booked.
- **adults, children, babies:** Number of adults, children, babies booked for the stay

Dataset Variables (conti.)

- **is_repeated_guest:** Binary variable indicating whether the customer was a repeat guest.
- **previous_cancellations:** Number of prior bookings that were canceled by the customer.
- **previous_bookings_not_canceled:** Number of prior bookings that were not canceled by the customer
- **required_car_parking_spaces:** Number of parking spaces requested by the customer.
- **total_of_special_requests:** Number of special requests made by the customer.
- **avg_daily_rate:** Average daily rate, as defined by dividing the sum of all lodging transactions by the total number of staying nights.
- **booked_by_company:** Binary variable indicating whether a company booked the booking
- **booked_by_agent:** Binary variable indicating whether an agent booked the booking,

Dataset Variables (conti.)

- **hotel_City:** Binary variable indicating whether the booked hotel is a "City Hotel"
- **hotel_Resort:** Binary variable indicating whether the booked hotel is a "Resort Hotel"
- **meal_BB:** Binary variable indicating whether a bed & breakfast meal was booked
- **meal_HB:** Binary variable indicating whether a half board meal was booked
- **meal_FB:** Binary variable indicating whether a full board meal was booked
- **meal_No_meal:** Binary variable indicating whether there was no meal package booked

Dataset Variables (conti.)

- **market_segment:(Designation)**market_segment_Aviation, market_segment_Complementary,market_segment_Corporate, market_segment_Direct,market_segment_Groups, market_segment_Offline_TA_TO, market_segment_Online_TA, market_segment_Undefined Indicates market segment designation with a value of 1. "TA"= travel agent, "TO"= tour operators.
- distribution_channel_Corporate, distribution_channel_Direct, distribution_channel_GDS, distribution_channel_TA_TO, distribution_channel_Undefined Indicates booking distribution channel with a value of 1. "TA"= travel agent, "TO"= tour operators, "GDS" = Global Distribution System

Dataset Variables (conti.)

- reserved_room_type_A, reserved_room_type_B, reserved_room_type_C, reserved_room_type_D, reserved_room_type_E, reserved_room_type_F, reserved_room_type_G, reserved_room_type_H, reserved_room_type_L Indicates code of room type reserved with a value of 1. Code is presented instead of designation for anonymity reasons
- deposit_type_No_Deposit Binary variable indicating whether a deposit was made, deposit_type_Non_Refund Binary variable indicating whether a deposit was made in the value of the total stay cost, deposit_type_Refundable Binary variable indicating whether a deposit was made with a value under the total stay cost
- customer_type_Contract Binary variable indicating whether the booking has an allotment or other type of contract associated to it
- customer_type_Group Binary variable indicating whether the booking is associated to a group
- customer_type_Transient Binary variable indicating whether the booking is not part of a group or contract, and is not associated to other transient booking
- customer_type_Transient-Party Binary variable indicating whether the booking is transient, but is associated to at least another transient booking

Data Analysis & Visualization

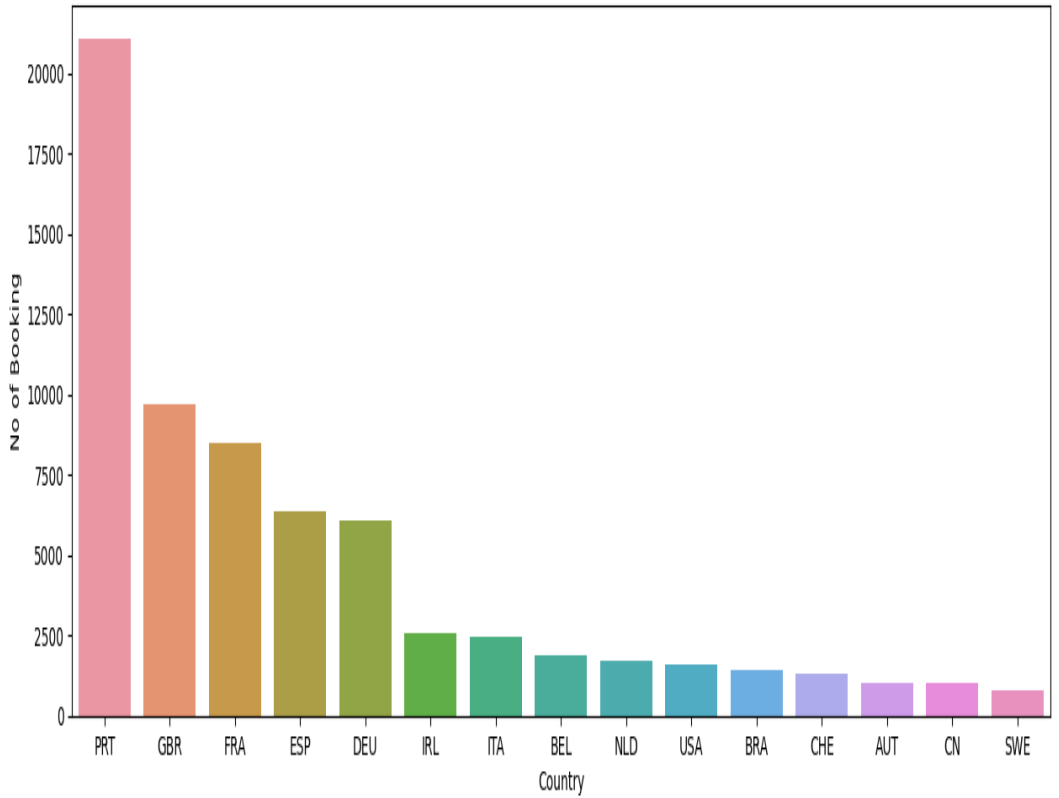
1.from which country the guest come?

Why did you pick the specific chart?

For showing segments of information, Vertical bar charts are useful to compare different categorical or discrete variables, as long as there are not too many categories to compare. here only two category.

What is/are the insight(s) found from the chart?

It is clear that from the chart that Portugal has most no booking. the reason may be that the hotels are in Portugal



Data Analysis & Visualization (cont.)

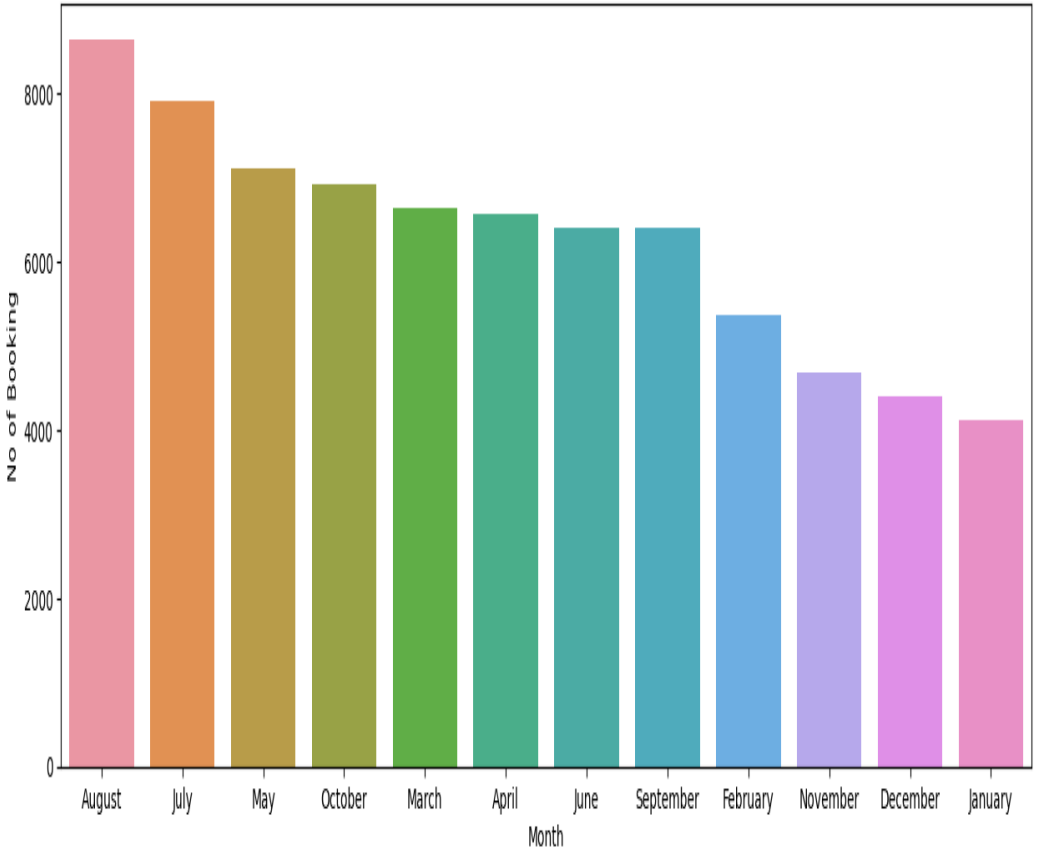
2.Which month has top booking month and which one lowest booking month?

Why did you pick the specific chart?

For showing segments of information, Vertical bar charts are useful to compare different categorical or discrete variables, as long as there are not too many categories to compare. here only two category.

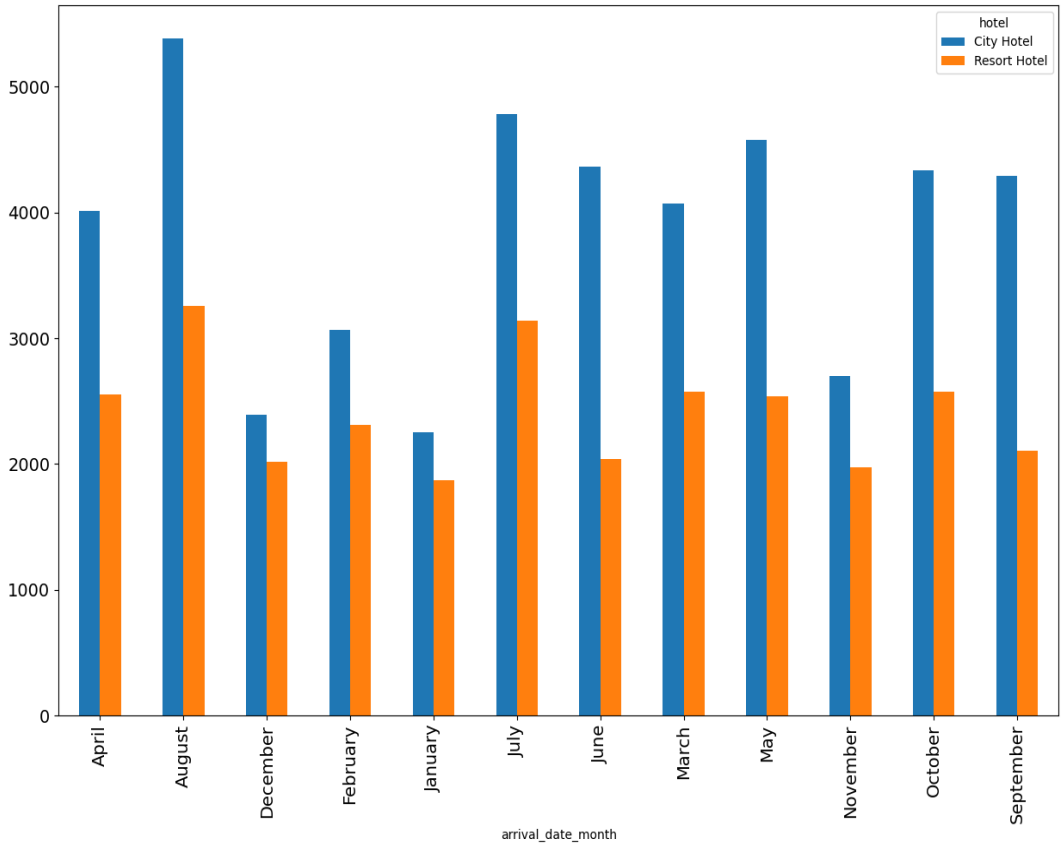
What is/are the insight(s) found from the chart?

As we can see most bookings were made from July to August. And the least bookings were made at the start and end of the year.



Data Analysis & Visualization (cont.)

2.Which month has top booking month and which one lowest booking month?



Why did you pick the specific chart?

For showing segments of information, Vertical bar charts are useful to compare different categorical or discrete variables, as long as there are not too many categories to compare. here only two category.

What is/are the insight(s) found from the chart?

As we can see most bookings were made from July to August. And the least bookings were made at the start and end of the year.

Data Analysis & Visualization (cont.)

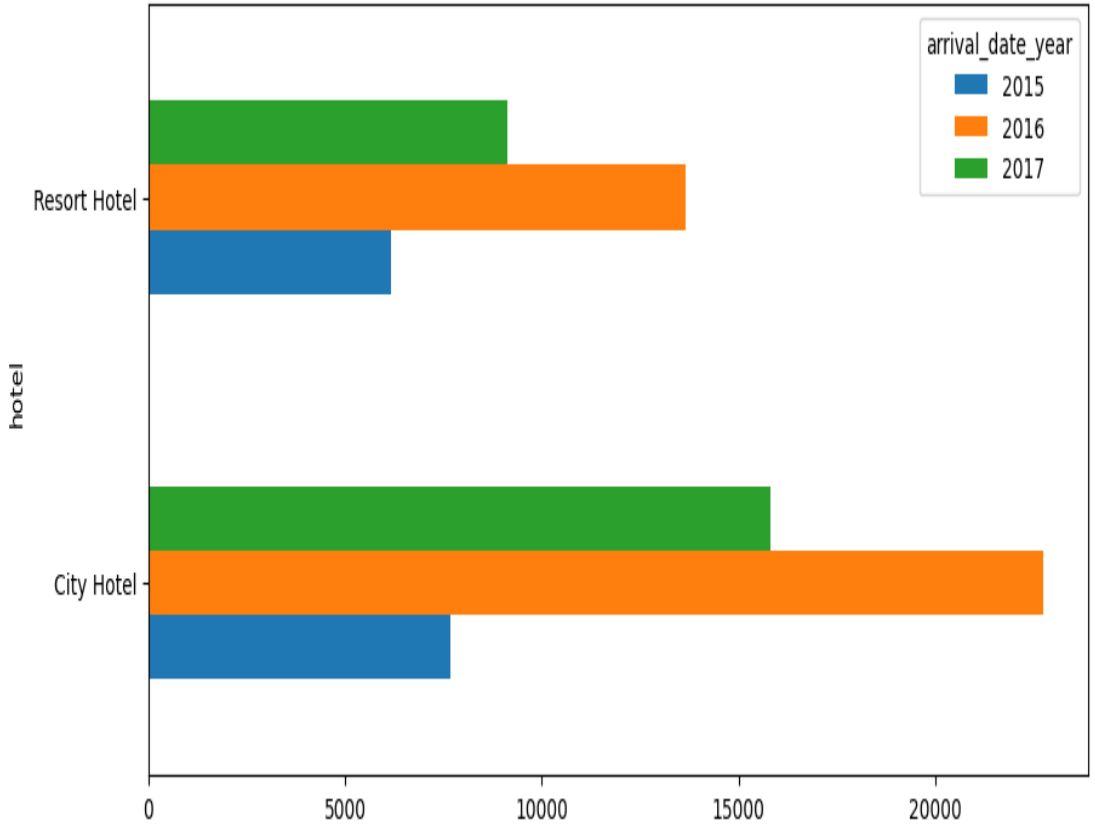
3. Year wise booking of hotel type.

Why did you pick the specific chart?

barh(y axis) creates a horizontal bar graph with one bar for each element in y . If y is an m-by-n matrix, then barh creates m groups of n bars. barh(x , y) draws the bars along the vertical axis at the locations specified by x . thats why i am using this

What is/are the insight(s) found from the chart?

Insights are clear , city hotel is more no of booking then resort hotel. strange thing in 2016, both hotel have most no of booking in compare to 2015 and 2017.



Data Analysis & Visualization (cont.)

4.How many bookings were cancelled in each type hotal?

.

Why did you pick the specific chart?

as it is also categorial compare with 2 categories , I used bar chart.

What is/are the insight(s) found from the chart?

resort hotel is less cancelation as compare to city hotel.

Data Analysis & Visualization (cont.)

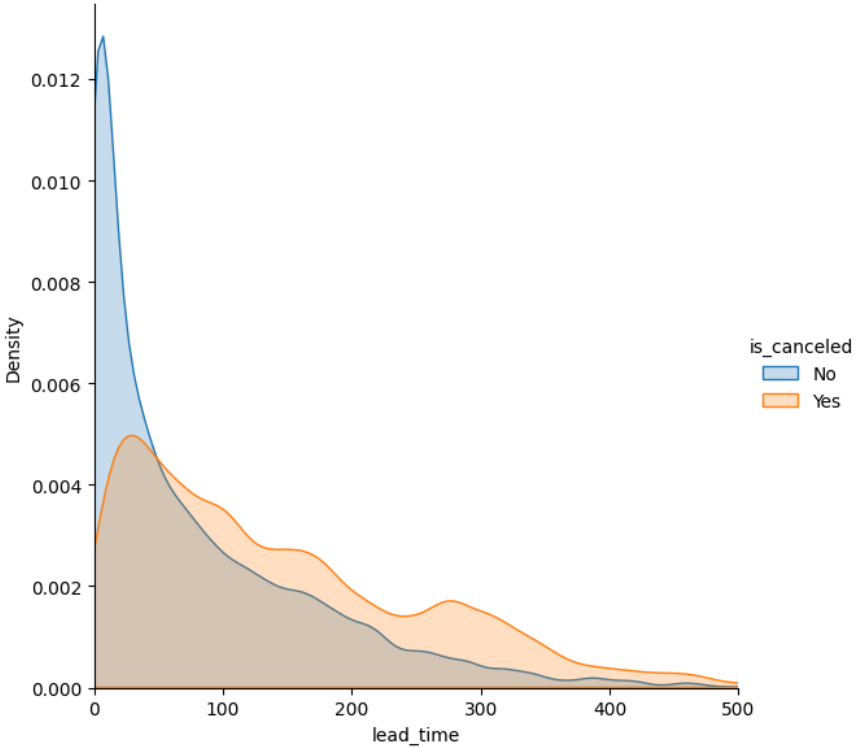
5. Is the lead time directly connected with cancellation.?

Why did you pick the specific chart?

FacetGrid class helps in visualizing distribution of one variable (cancelation)as well as the relationship between multiple variables (lead time)separately within subsets of your dataset using multiple panels.

What is/are the insight(s) found from the chart?

When the lead time about to 60 or more, bookings frequently cancelled.



Data Analysis & Visualization (cont.)

6. How many customer are there who have repeated the bookings?

Why did you pick the specific chart?

Here I need to count of observation in each category, that's why used count plot. Count plot provides counts of categorical values, count of observations in each category.

What is/are the insight(s) found from the chart?

City hotel is more repeated guest than the resort hotels

Data Analysis & Visualization (cont.)

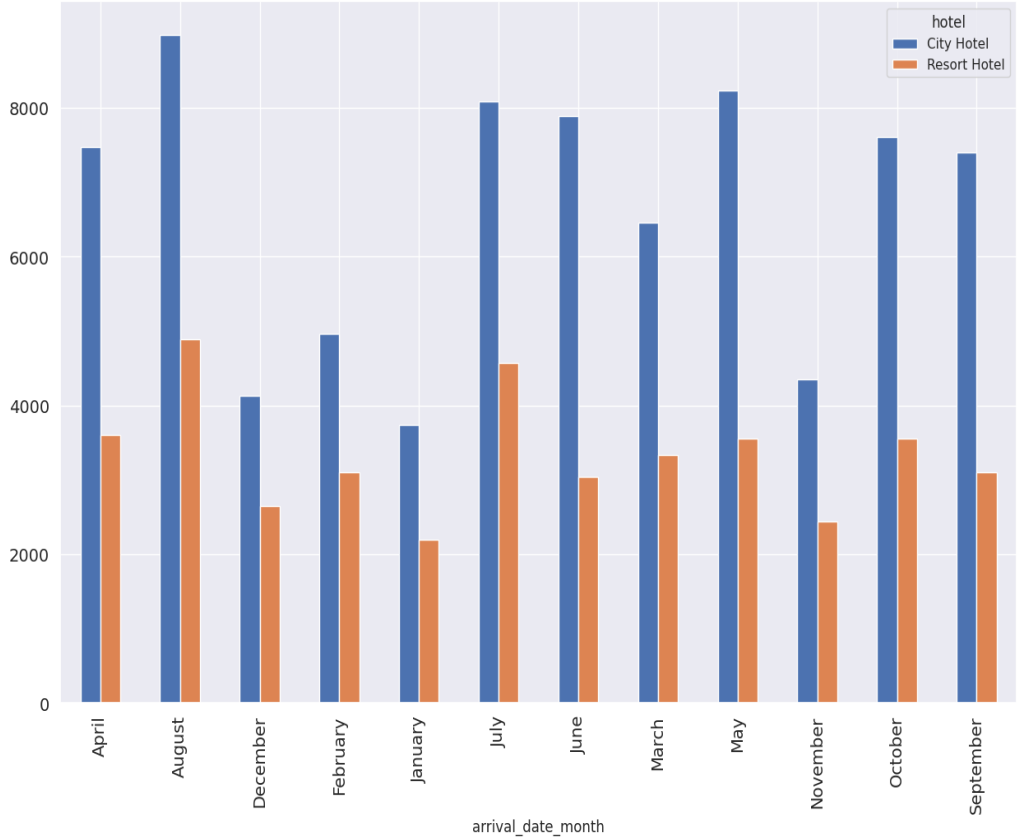
7.Which month has the highest number of cancelations and lowest cancelations month?

Why did you pick the specific chart?

Here I need to count of observation in each category, that's why used count plot. Count plot provides counts of categorical values, count of observations in each category.

What is/are the insight(s) found from the chart?

City hotel is more repeated guest than the resort hotels

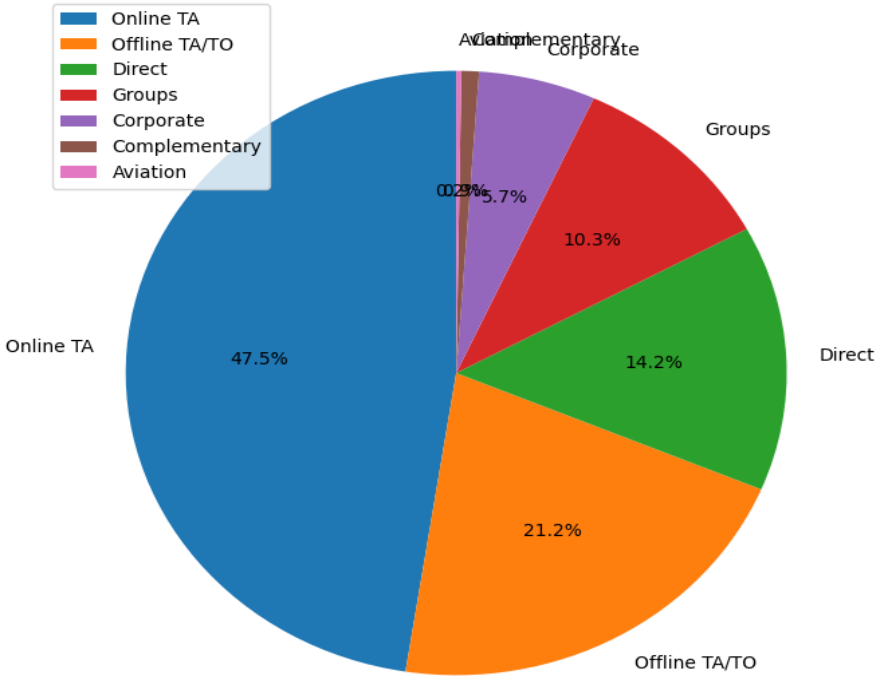


Data Analysis & Visualization (cont.)

8. What is Bookings number by Market segment.?

Why did you pick the specific chart?

pie chart helps organize and show data as a percentage of a whole. True to the name, this kind of visualization uses a circle to represent the whole, and slices of that circle, or “pie”, to represent the specific categories that compose the whole. Here Market Segment is whole Circle and Specific categories are plotted using **pie.**



What is/are the insight(s) found from the chart?

47.5% booking online, then offline 21.2%, Then Direct is 14.2. These are the major Categories by market Segment to booked the hotel.

Data Analysis & Visualization (cont.)

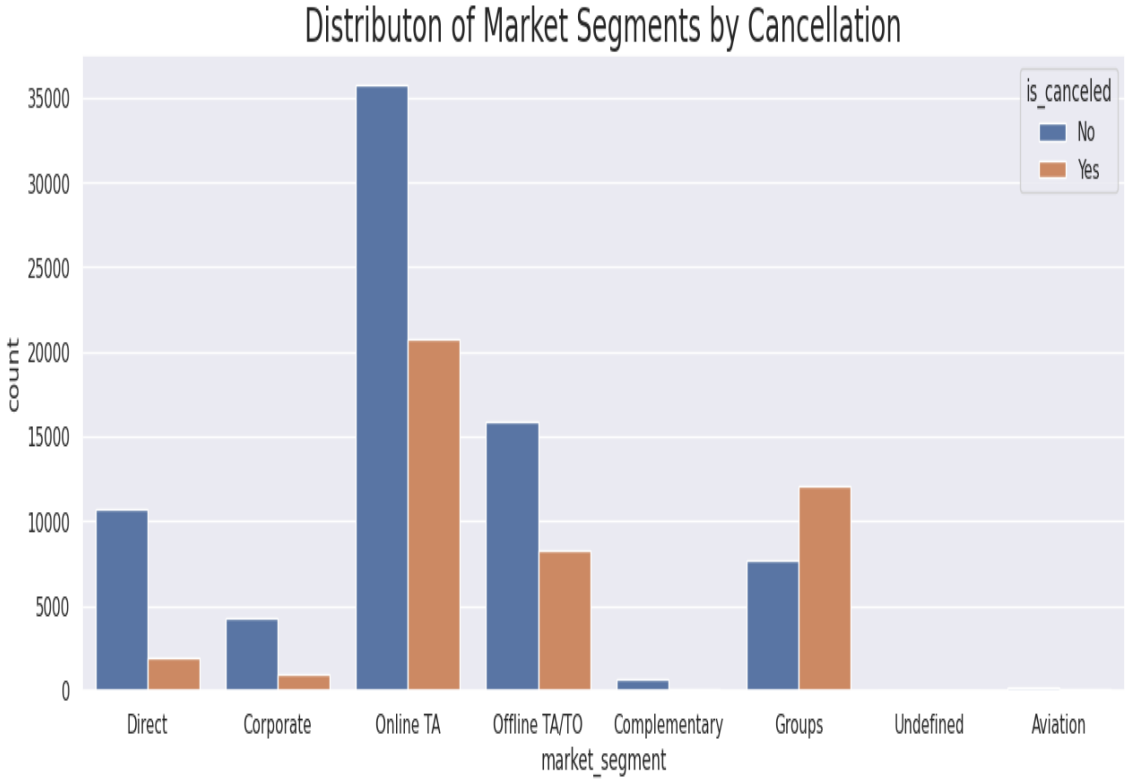
9. Distribution by cancellation by market segments.

Why did you pick the specific chart?

Here I need to count of observation in each category, that's why i used count plot. Count plot provides counts of categorical values, count of observations in each category. Here i need to count of cancelation of booking by each market category.

What is/are the insight(s) found from the chart?

The cancellation rate for groups is greater than 50%. The cancellation rate for offline TA/TOs (Travel Agents/Tour Operators) and online TAs is greater than 34%. the cancellation rate in the direct segment is so low is another issue that caught my attention. In the event when people are speaking one-on-one, I believe that a mutually trusting relationship has been built at this time.



Data Analysis & Visualization (cont.)

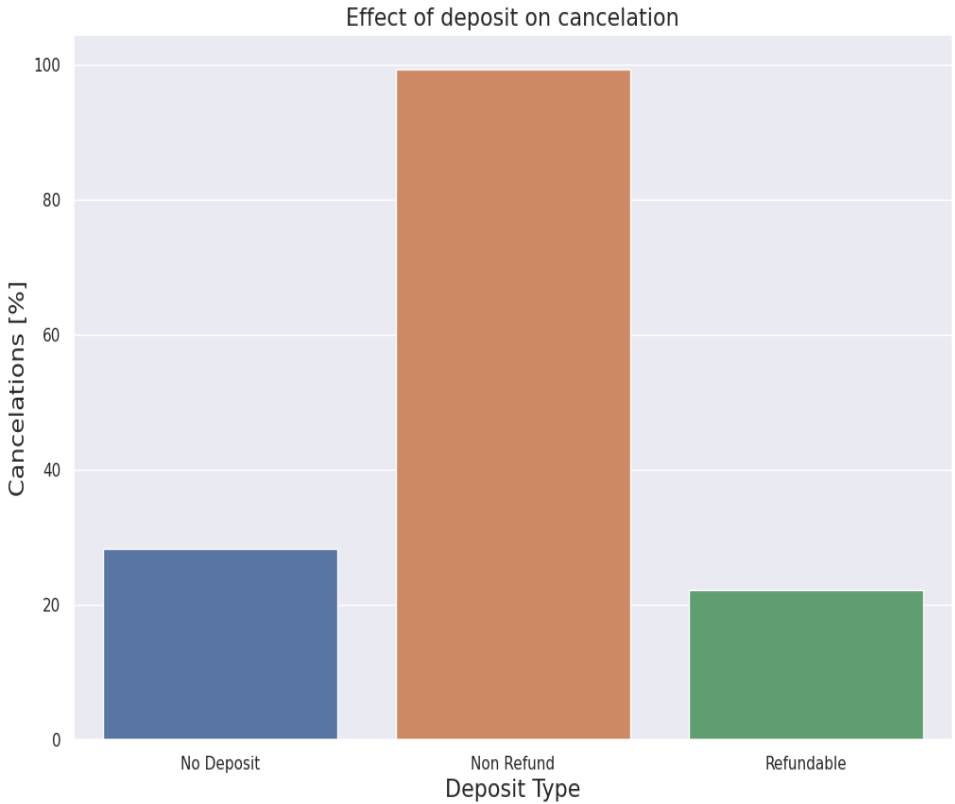
10. Deposit Vs Cancellations effect.?

Why did you pick the specific chart?

categorial analysis is here that's why I choose this one

What is/are the insight(s) found from the chart?

Non refund Deposit type had most cancelation.



Data Analysis & Visualization (cont.)

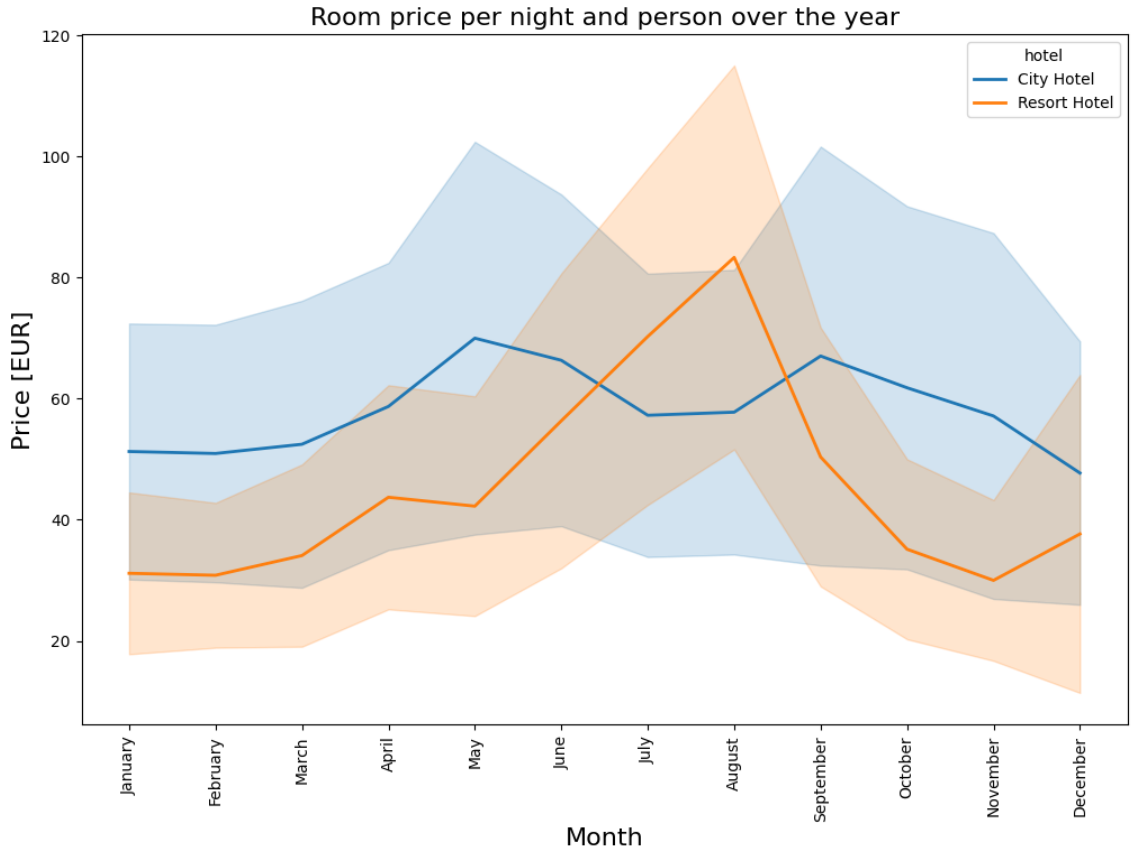
11. How does the price per night vary over the year?

Why did you pick the specific chart?

The important use of line graph is to track the changes over the short and long period of time. It is also used to compare the changes over the same period of time for different groups. It is always better to use the line whenever the small changes exist.

What is/are the insight(s) found from the chart?

shows that the prices in the Resort hotel are much higher during the month of August. The price of the city hotel varies less and is most expensive during may and September where there is spring and autumn season respectively.



Challenges

- Big dataset, study very carefully with extra mindset.
- There was a lot of duplicate data.
- Variables presented in different datatype format, so analysis need more effort
- Choosing appropriate visualization techniques to use was difficult.
- Choosing appropriate chart for prestation was difficult.
- Some variable value need more clarification for more common understanding i.e. for country name
- Huge number of null values were present in the given dataset.

Conclusion

- high number of bookings are from Portugal, probably because the hotel is located in Portugal itself, followed by second country is the United Kingdom which is approx. 75% behind.
- -more Repeated guests do not cancel their reservations. Of course there are some exceptions. Also most of the customers are not repeated Customers.
- -Customers from Aviation Segment do not seem to be staying at the resort hotels and have a relatively lower day average.
- -The weekends and weekdays averages are roughly equal.
- -Customers in the Aviation Segment are likely to arrive shortly due to business.
- -The cancellation rate for groups is greater than 50%.
- -The cancellation rate for offline TA/TOs and online TAs is greater than 33%.
- -Direct segments have a lower rate of cancellation.
- -When the lead time exceeds about 60, guests frequently cancel their bookings (cancellation rate is higher after this point).
- -Additionally, people want their vacation or work schedules to be calculated across 100 days, or 50% of the data.
- -City hotels receive more guests throughout the year. Resort hotels appear to be slightly closer to city hotels in the summer when comparing proportions.
- -The City Hotel sees an increase in visitors in the spring and fall when rates are also the greatest. Less people arrive in July and August, when the prices are still lower.
- -From June through September, when costs are at their greatest, fewer guests stay at the Resort hotel. The winter season draws the fewest visitors to both hotels.