Capstone project

Hotel Bookings Analysis

By- Ganesh Walimbe

Introduction

- Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, type of meals , days of week and many more. This makes analyzing the patterns available in the past data is more important to help the hotels plan better . Using the historical data, hotels can perform various campaigns to boost the business.
- I will be using the data available to analyze the factors affecting the hotel bookings. These factors can be used for reporting the trends and finding hidden patterns from dataset .my project will also help customers to find the best time to rent hotel room .

Methodology

- Problem statement
- Introduction to data
- Data pre-processing
- Performing Exploratory data analysis
- Conclusions

Problem statement

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.

Explore and analyze the data to discover important factors that govern the bookings

Introduction to data

The given dataset is containing information about the bookings of resort and city hotels from year 2015 to 2017 .And it is containing 119390 rows and 31 variables those variables are as follows

- **Hotel**- Resort Hotel or City Hotel
- **is_canceled**-Value indicating if the booking was cancelled (1) or not (0)
- l**ead_time**-Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
- **arrival_date_year**-Year of arrival date
- **arrival_date_month**-Month of arrival date
- **arrival_date_week_number**-Week number of year for arrival date
- **arrival_date_day_of_month**-Day of arrival date
- **stays_in_weekend_nights**-Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **Adults-**Number of adults
- **Children-**Number of children
- **Babies-**Number of babies
- **Meal**-Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner
- **Country-**Country of origin. Categories are represented in the ISO 3155–3:2013 format
- **market_segment-**Market segment designation. In categories
- **distribution_channel-**Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
- **is_repeated_guest-**Value indicating if the booking name was from a repeated guest (1) or not (0)
- **previous_cancellations-**Number of previous bookings that were cancelled by the customer prior to the current booking

Introduction to data

- **previous_bookings_not_canceled-**Number of previous bookings not cancelled by the customer prior to the current booking
- **reserved_room_type-**Code of room type reserved. Code is presented instead of designation for anonymity reasons.
- **assigned_room_type-**Code for the type of room assigned to the booking.
- **booking_changes-**Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
- **deposit_type-**Indication on if the customer made a deposit to guarantee the booking
- **Company-**ID of the company/entity that made the booking or responsible for paying the booking.
- **days_in_waiting_list-**Number of days the booking was in the waiting list before it was confirmed to the customer
- **customer_type-**Type of booking, assuming one of four categories
- **adr-**Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
- **required_car_parking_spaces-**Number of car parking spaces required by the customer
- **total_of_special_requests**
- Number of special requests made by the customer (e.g. twin bed or high floor)
- **reservation_status-**Reservation last status, assuming one of three categories: Canceled Check-Out,No-Show

Data preprocessing

Missing value treatment :

- Our dataset has missing values in four variables named country, children , agent and Company.
- I dropped company column because it has 94% missing values in it.
- After that i dropped the null values present in the dataset

Duplicate values treatment :

- There were 32001 duplicates present in the dataset
- So I dropped all the duplicate values from Dataset.

Creating new variables :

- I also created two variables first is total customers created by adding adults , children and babies variables and second is totat_stay_in_nights created by adding stays_in_week_nights and stays_in_weekend_nights variables .

```
percentage_missing_values=missing_values/len(data)
percentage_missing_values

hotel                             0.000000
is_canceled                       0.000000
lead_time                         0.000000
arrival_date_year                 0.000000
arrival_date_month                0.000000
arrival_date_week_number          0.000000
arrival_date_day_of_month         0.000000
stays_in_weekend_nights           0.000000
stays_in_week_nights              0.000000
adults                            0.000000
children                          0.000034
babies                            0.000000
meal                              0.000000
country                           0.004087
market_segment                    0.000000
distribution_channel              0.000000
is_repeated_guest                 0.000000
previous_cancellations            0.000000
previous_bookings_not_canceled    0.000000
reserved_room_type                0.000000
assigned_room_type                0.000000
booking_changes                   0.000000
deposit_type                      0.000000
agent                             0.136862
company                           0.943069
days_in_waiting_list              0.000000
customer_type                     0.000000
adr                               0.000000
required_car_parking_spaces       0.000000
total_of_special_requests         0.000000
reservation_status                0.000000
reservation_status_date           0.000000
total_customers                   0.000034
total_stay_in_nights              0.000000
dtype: float64
```

```
data.duplicated().value_counts()

False    87389
True     32001
dtype: int64
```

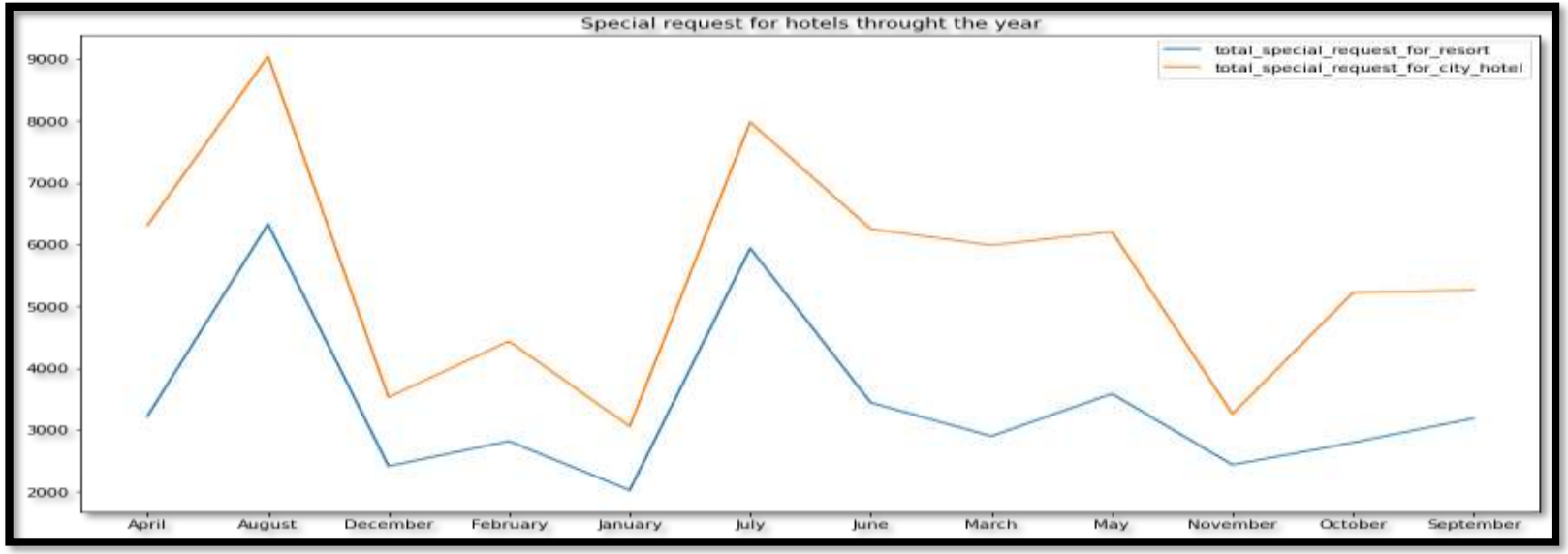
Exploratory data analysis

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

Objectives for performing EDA are :

- To find how many special request for hotels are requested throughout the year.
- Bookings cancelation analysis
- Booking from market segments
- To find Most preferred meals by customers
- Top Agents inviting customers to hotels
- Top countries with most customers visiting hotels
- Finding out in which months most customers are visiting hotels
- Room price per night variation over the year
- How long do people stay at hotels

Special request for hotels throughout the year

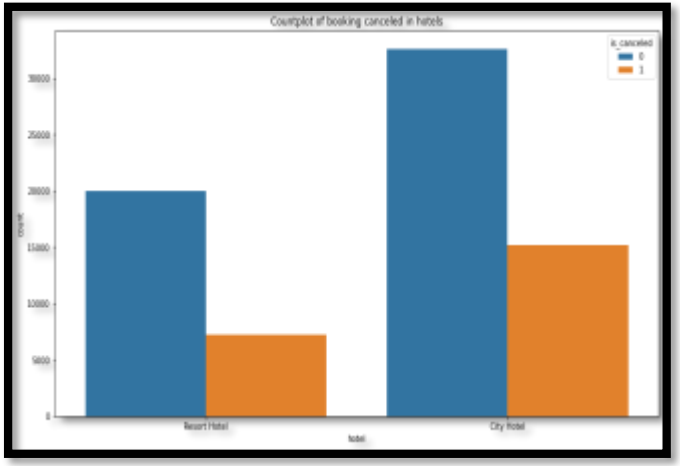


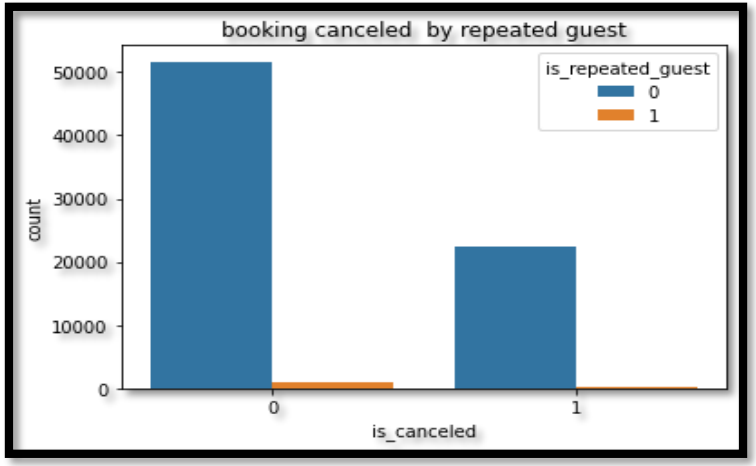
From lineplot we can see that ,

Special request are high the month august and July for both resort and city hotel .

Special request are low in the month of January , November and December.

Bookings cancelation analysis

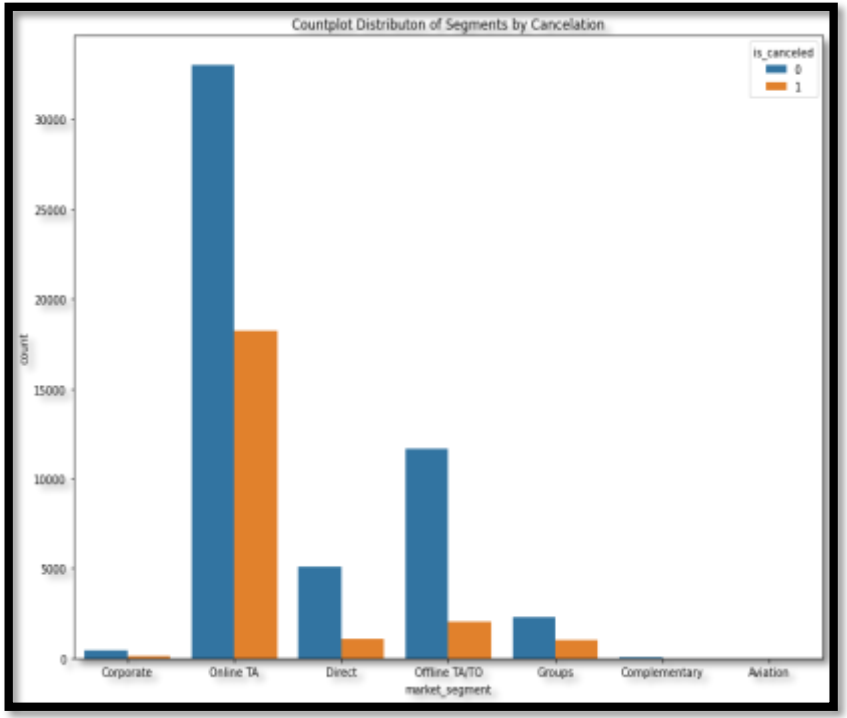




From above plots we can see that,

- There are more number of booking cancelation in city hotel as compared to resort hotel.
- It is seen that if customers are repeated customers then they are more likely to not cancel any of the bookings.

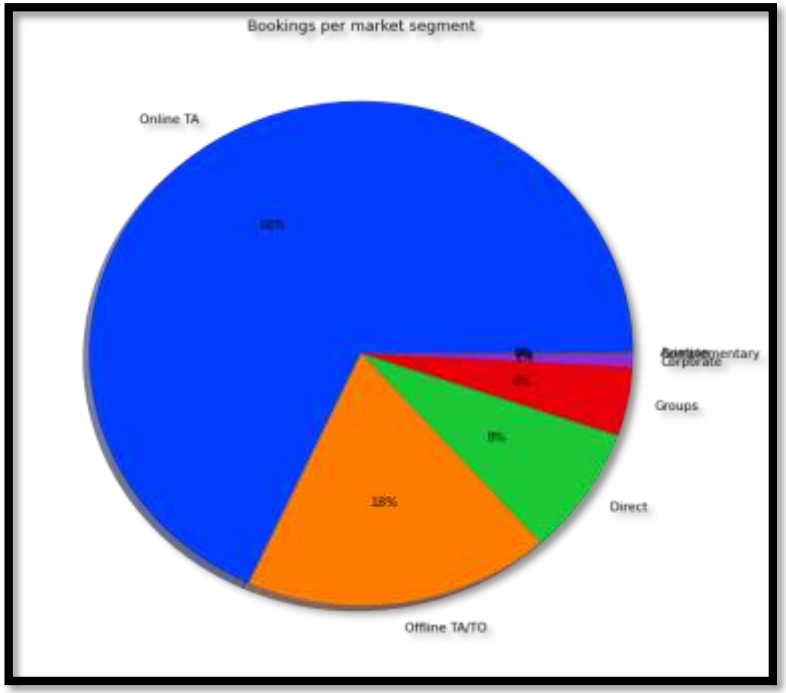
Bookings cancelation analysis



From the countplot we can say that ,

- Maximum number of bookings cancellation are happening from online TA market segment followed by offline TA/TO segment.
- Least amount of bookings cancellation are happening from corporate segment.

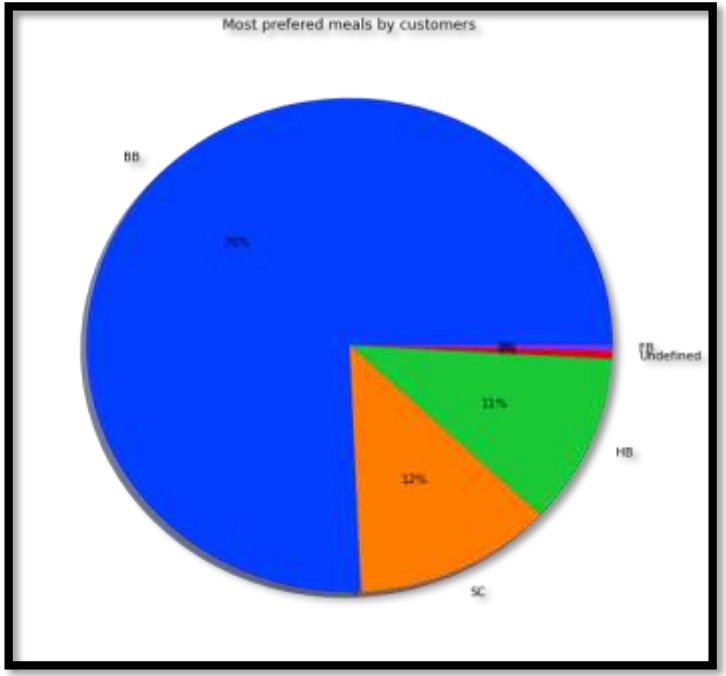
Booking from market segments



From pie chart we can see that

- 68% of the hotel bookings are done in online TA market segments.
- 18% of hotel bookings are done in offline TA/To market segment .
- Direct bookings are 8%
- And remaining bookings are done by groups and corporate segments.

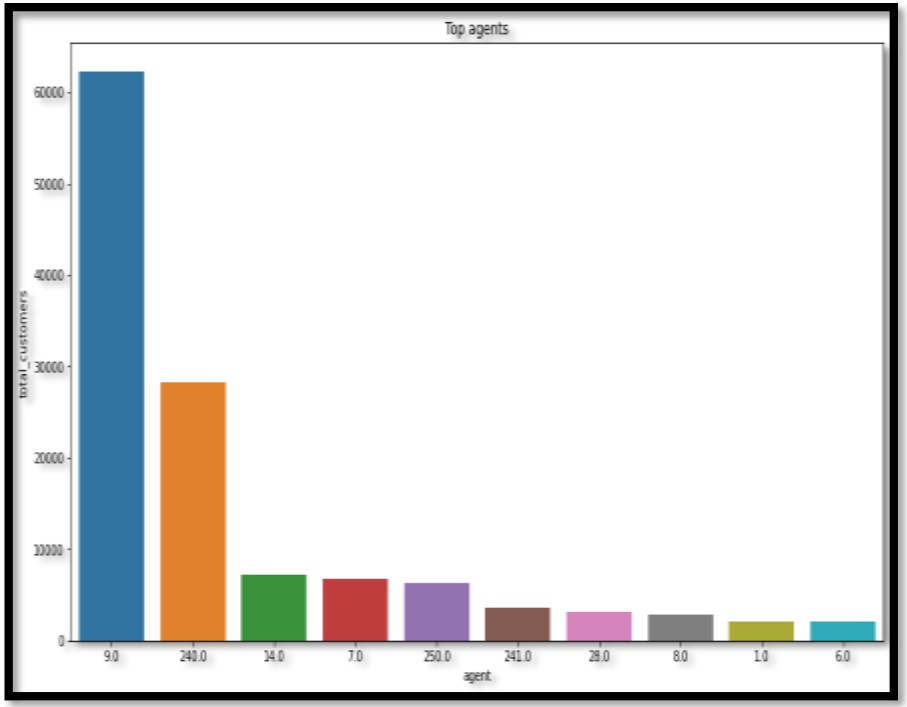
Most preferred meals by customers



From pie chart we can see that ,

- 76% of customers prefer to have meal BB-Bed and breakfast.
- 12%of the customers prefer to have meal SC-no meal package.
- 11% of customers prefer to have meal HB-half board
- Remaining customers prefer to have FB-full board and undefined meal.

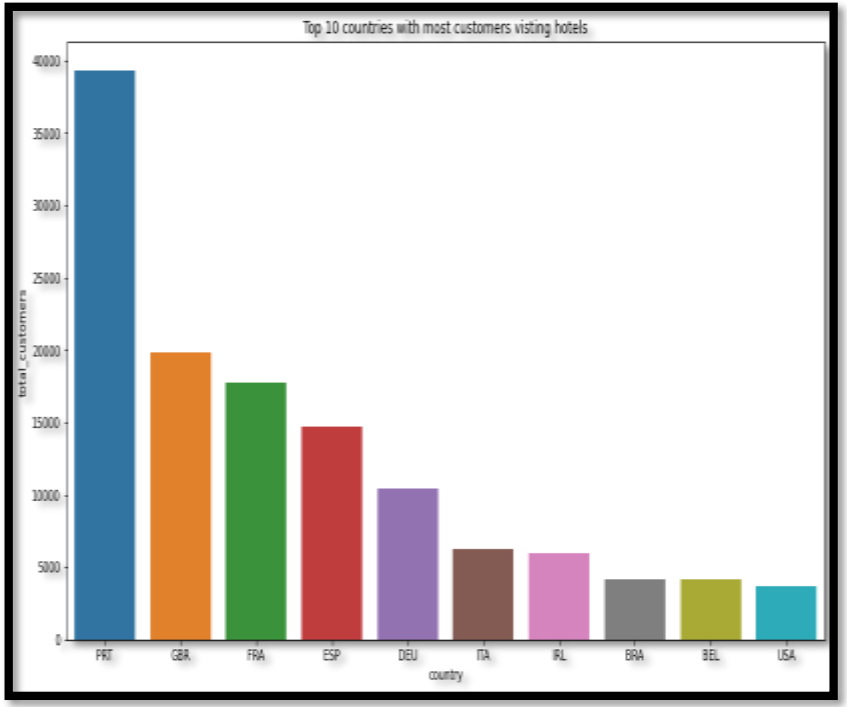
Top Agents



Top agents inviting customers to the hotels with agent id are:

1. Agent id 9
2. Agent id 240
3. Agent id 14
4. Agent id 7
5. Agent id 250
6. Agent id 241

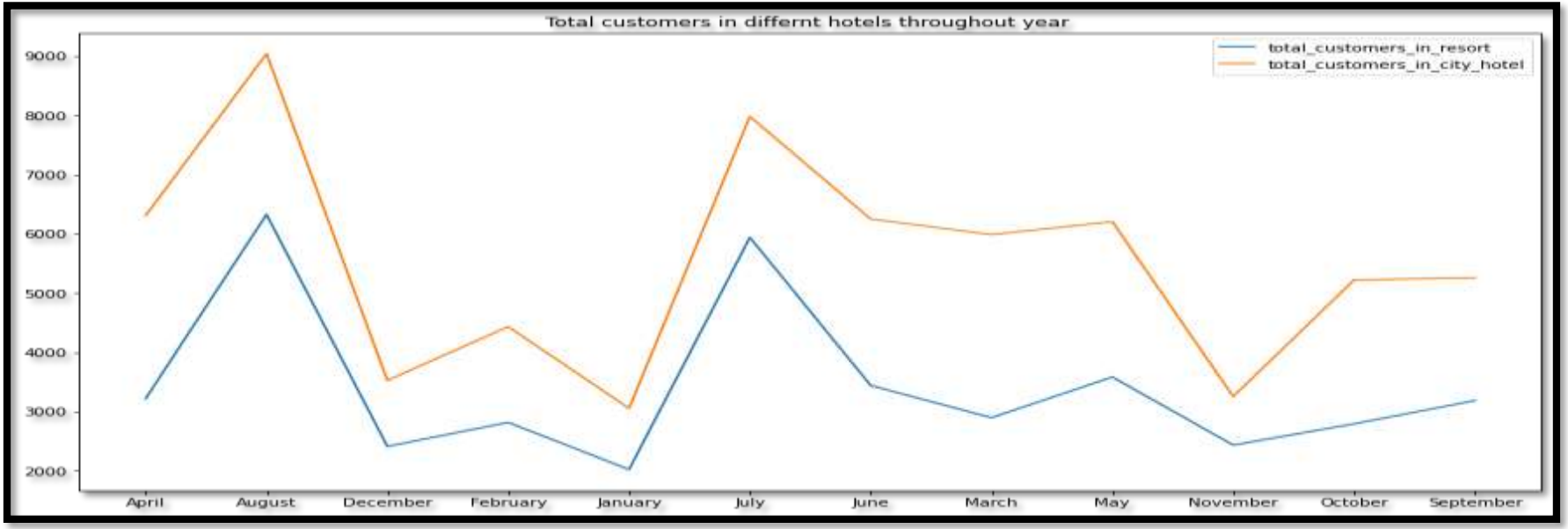
Top countries with most customers visiting hotels



From barplot we can see that,

- Most of the customers are coming from country PRT-Portugal to visit hotels
- Followed by GBA-Great Britain :Alderney ,FRA-France , ESP-Spain and DEU-Germany ,etc

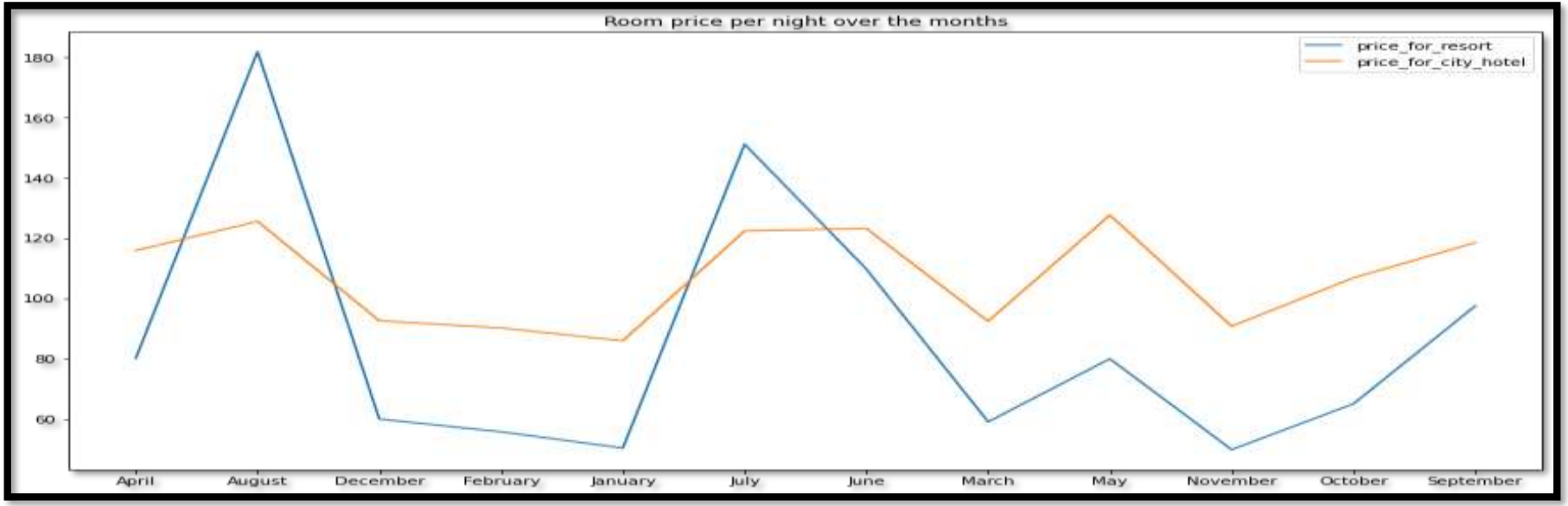
Finding out in which months most customers are visiting hotels



From line chart we can see that ,

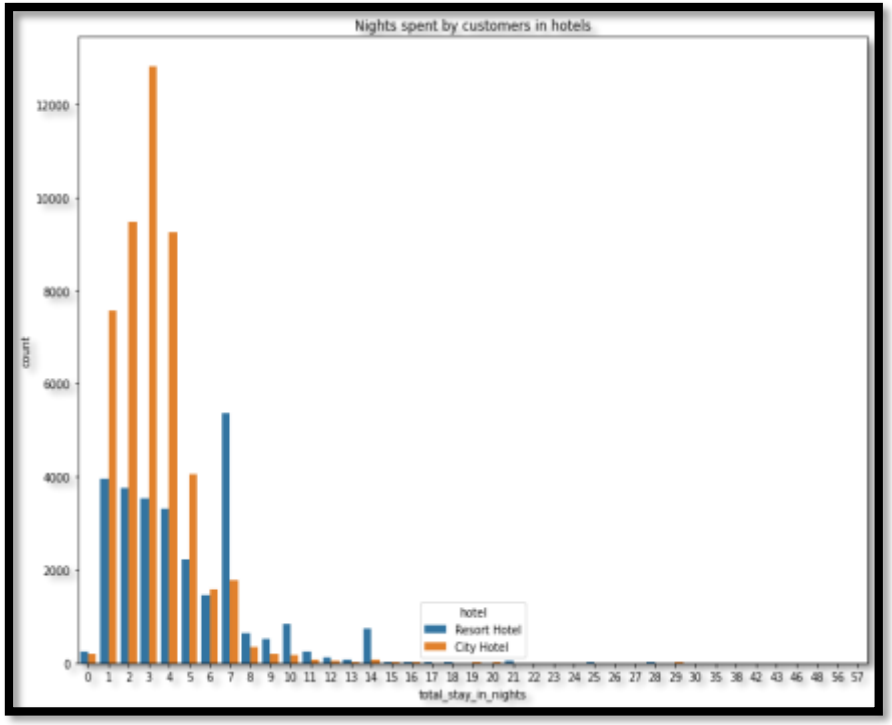
- Most customers are visiting resorts and city hotels in the month of July
- And August

Room price per night variation over the year



- From the line plot we can see that the room price per night is high in august and July for resort .price is low in December ,January and November
- For city hotel room price per nights high in the month of august and may and price per night is low in December ,January and March.

How long do people stay at hotels



From countplot we can see that
Most of the customers are like to spend four nights in city hotels
And seven nights in resort
Maximum nights spent by customers in hotels is 21 nights

Conclusions

- Customers can expect high number of special request for hotels in the month August and July.
- If the customers is repeated customers then they are more likely to not cancel the hotel bookings.
- Maximum number of bookings cancellation are happening from Online TA market because it is easy to book and cancel reservation through online travel agents.
- Maximum number of bookings for hotel are done via online travel agent market segment and minimum bookings are done via group and corporate segments.
- Most popular meal category of hotels is bed and breakfast category

Conclusions

- Top agents inviting customers to hotels with agent id are :agent id 9,agent id 240 ,agent id 14,agent id 7 and agent id 250,etc
- The busiest months for bookings of hotels are July and August because people are have holidays in this time of year .
- Room price per night in resort is high in month of August and July and low in month of January, November and December.
- Room price per night in city hotel is high in month of May and August and low in month of January and March ,December.
- Customers like to spend four nights in city hotels and seven nights in resorts.

Thank you