Capstone Project - 1

Exploratory data analysis- Hotel Booking Analysis

Sajal sinha

Mohammed arifuddin atif

Agenda

In this project, we will perform Exploratory Data Analysis on Hotel Booking Dataset and will get some meaningful insights.

What is Exploratory Data Analysis?

- Exploratory Data Analysis refers to the critical process of performing initial investigations on data so on discover patterns, to spot anomalies, to test hypothesis and to see assumptions with the assistance of summary statistics and graphical representations.

About Our Dataset

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.

Let’s know our data

Hotel: Type of hotel

is_canceled : if bookings were canceled

lead_time:

arrival_date_year: in which year did customer arrived

arrival_date_month: in which month did customer arrived

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_cancellation:

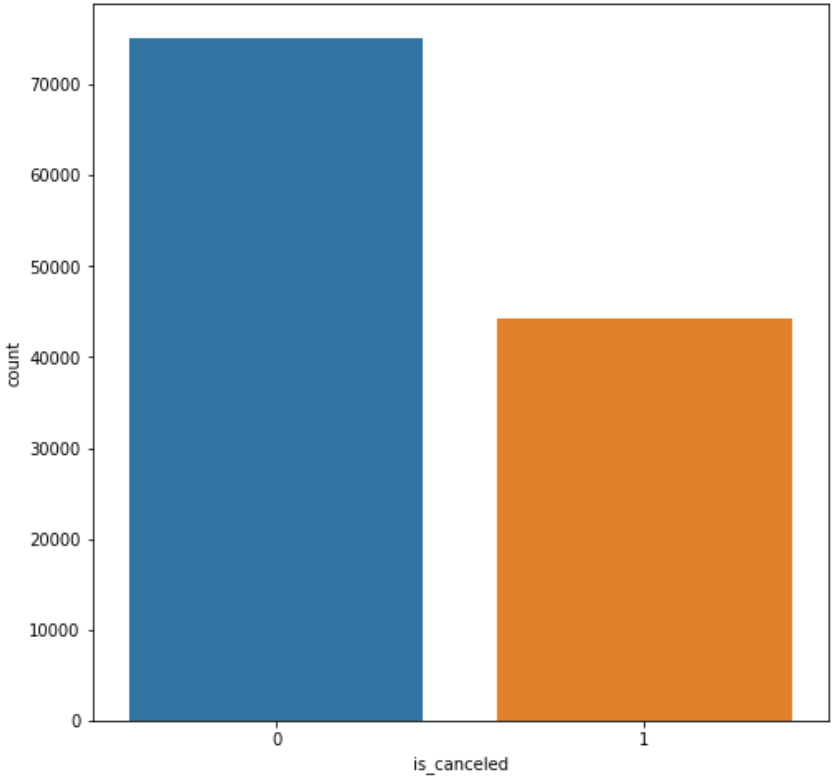
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count   Dtype  
---  ------                          --------------   -----  
 0   hotel                           119390 non-null  object 
 1   is_canceled                     119390 non-null  int64  
 2   lead_time                       119390 non-null  int64  
 3   arrival_date_year               119390 non-null  int64  
 4   arrival_date_month              119390 non-null  object 
 5   arrival_date_week_number        119390 non-null  int64  
 6   arrival_date_day_of_month       119390 non-null  int64  
 7   stays_in_weekend_nights         119390 non-null  int64  
 8   stays_in_week_nights            119390 non-null  int64  
 9   adults                          119390 non-null  int64  
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64  
 12  meal                            119390 non-null  object 
 13  country                         118902 non-null  object 
 14  market_segment                  119390 non-null  object 
 15  distribution_channel            119390 non-null  object 
 16  is_repeated_guest               119390 non-null  int64  
 17  previous_cancellations          119390 non-null  int64  
 18  previous_bookings_not_canceled  119390 non-null  int64  
 19  reserved_room_type              119390 non-null  object 
 20  assigned_room_type              119390 non-null  object 
 21  booking_changes                 119390 non-null  int64  
 22  deposit_type                    119390 non-null  object 
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64  
 26  customer_type                   119390 non-null  object 
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64  
 29  total_of_special_requests       119390 non-null  int64  
 30  reservation_status              119390 non-null  object 
 31  reservation_status_date         119390 non-null  object 
dtypes: float64(4), int64(16), object(12)
memory usage: 29.1+ MB
```

Exploratory Data Analysis

Data cleaning and Manipulation

- While checking for null values, 4 columns had null values in them. They were children, country, agent, and company.
- As company column had more than 50% of null values, dropping it was more beneficial.
- Other columns were filled using mean, median and mode.
- Data Type of some columns were changed in order to get proper result.

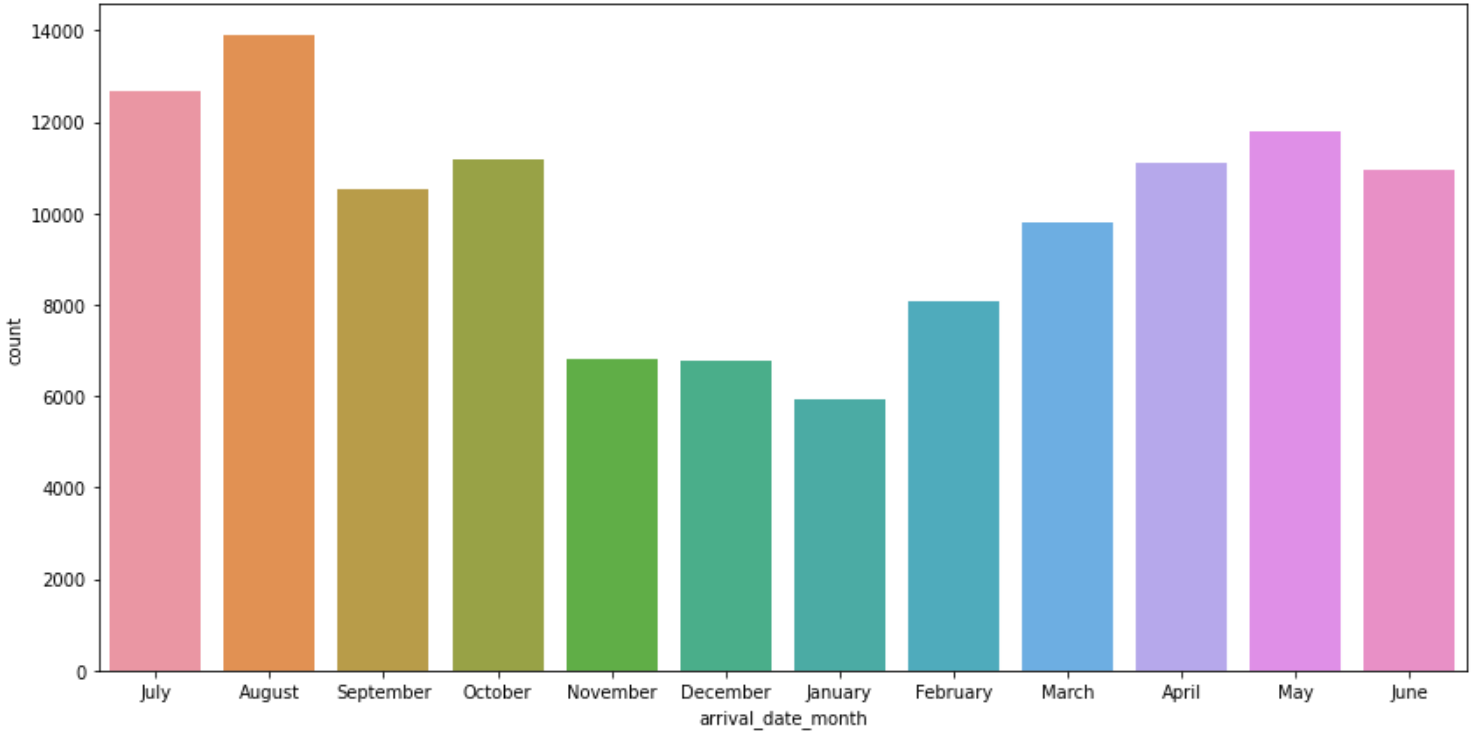
Count of bookings that were Canceled=1, Not canceled= 0



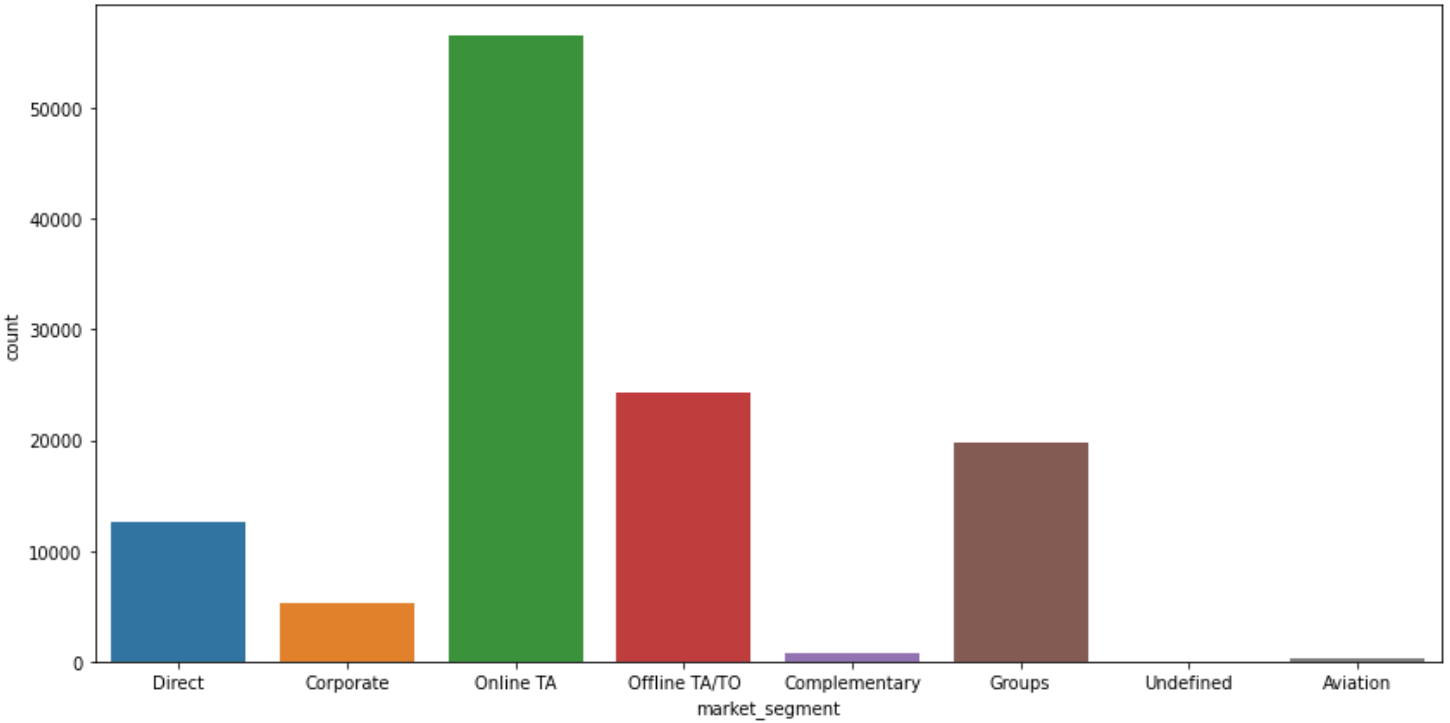
Top 3 countries which has most bookings

```
Most common countries:
PRT    49078
GBR    12129
FRA    10415
Name: country, dtype: int64
----------------------------
```

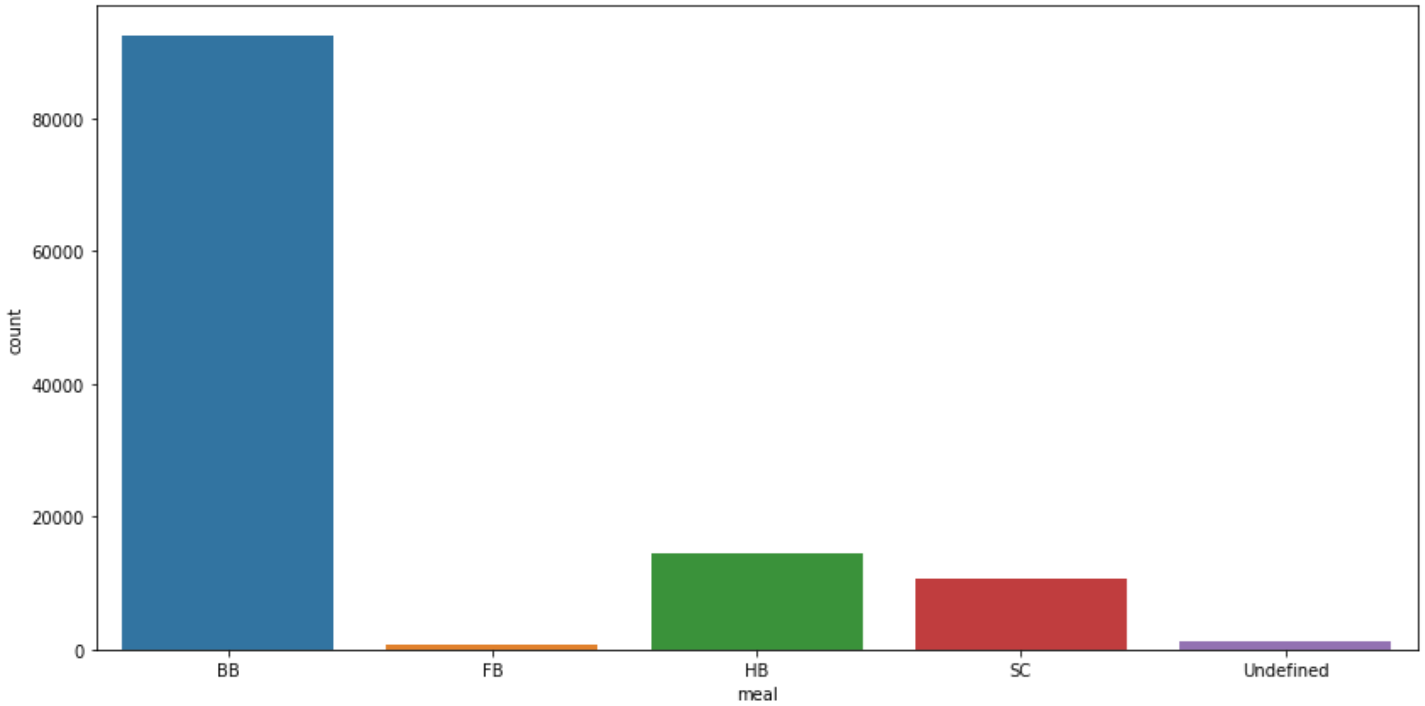
Number of Booking in each month



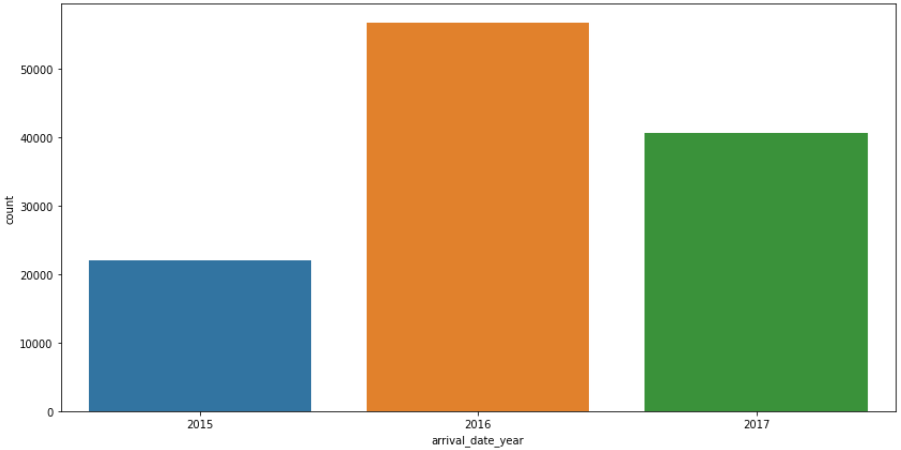
Type of market segment customers belong.



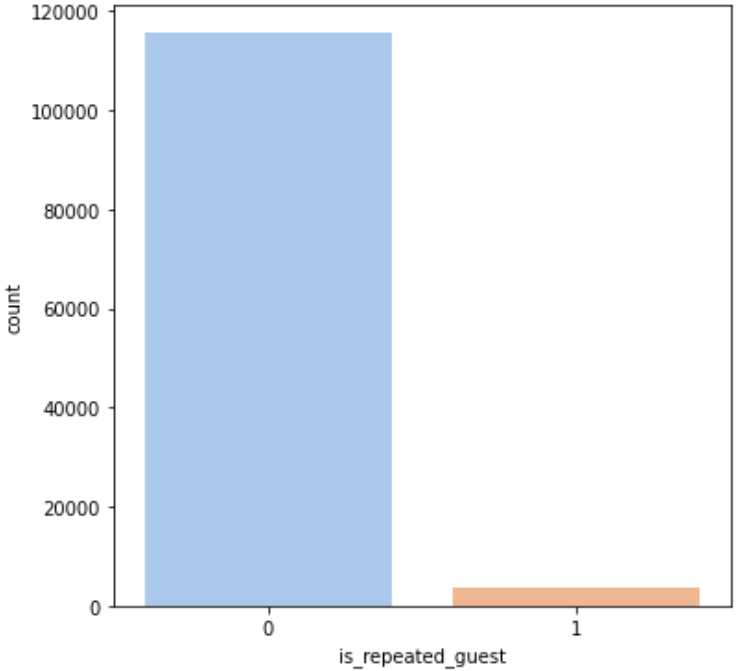
Type of Meal customers preferred.



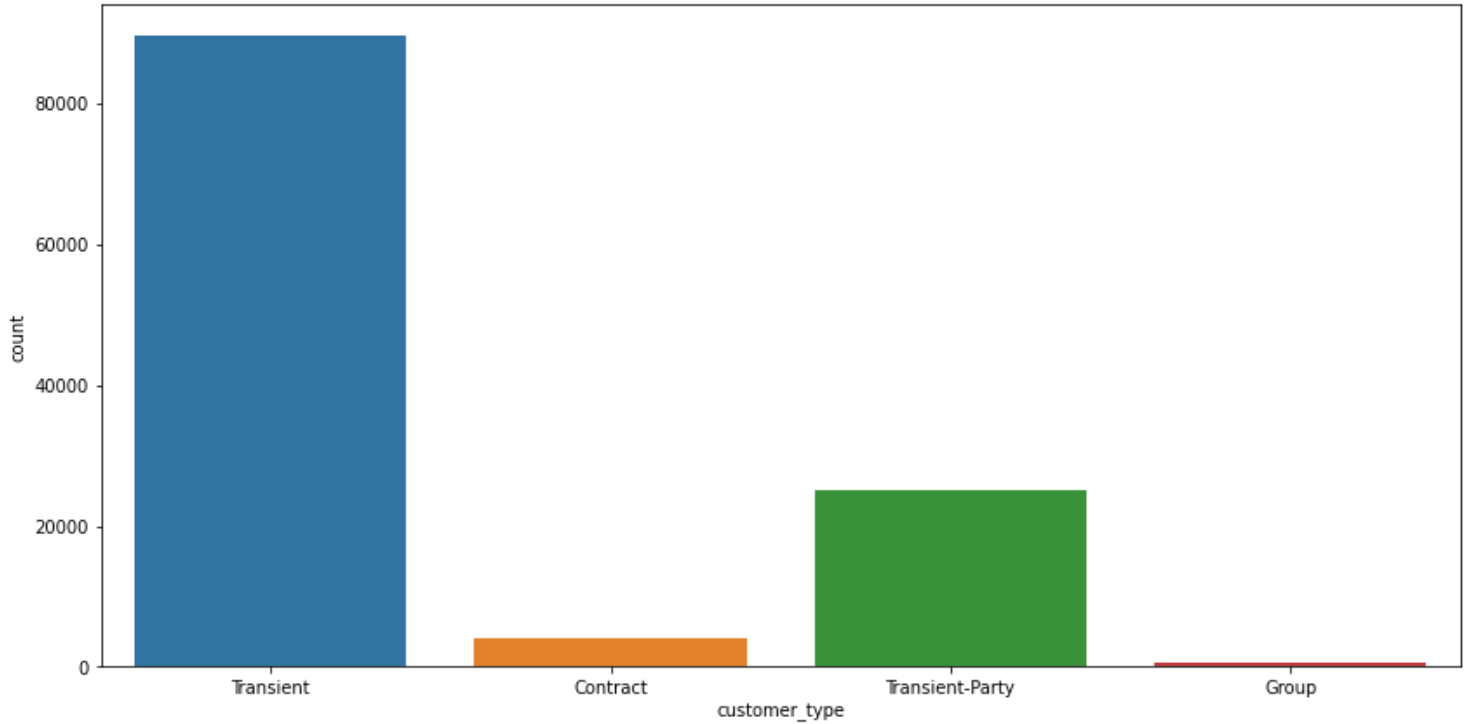
- Count of customers each year



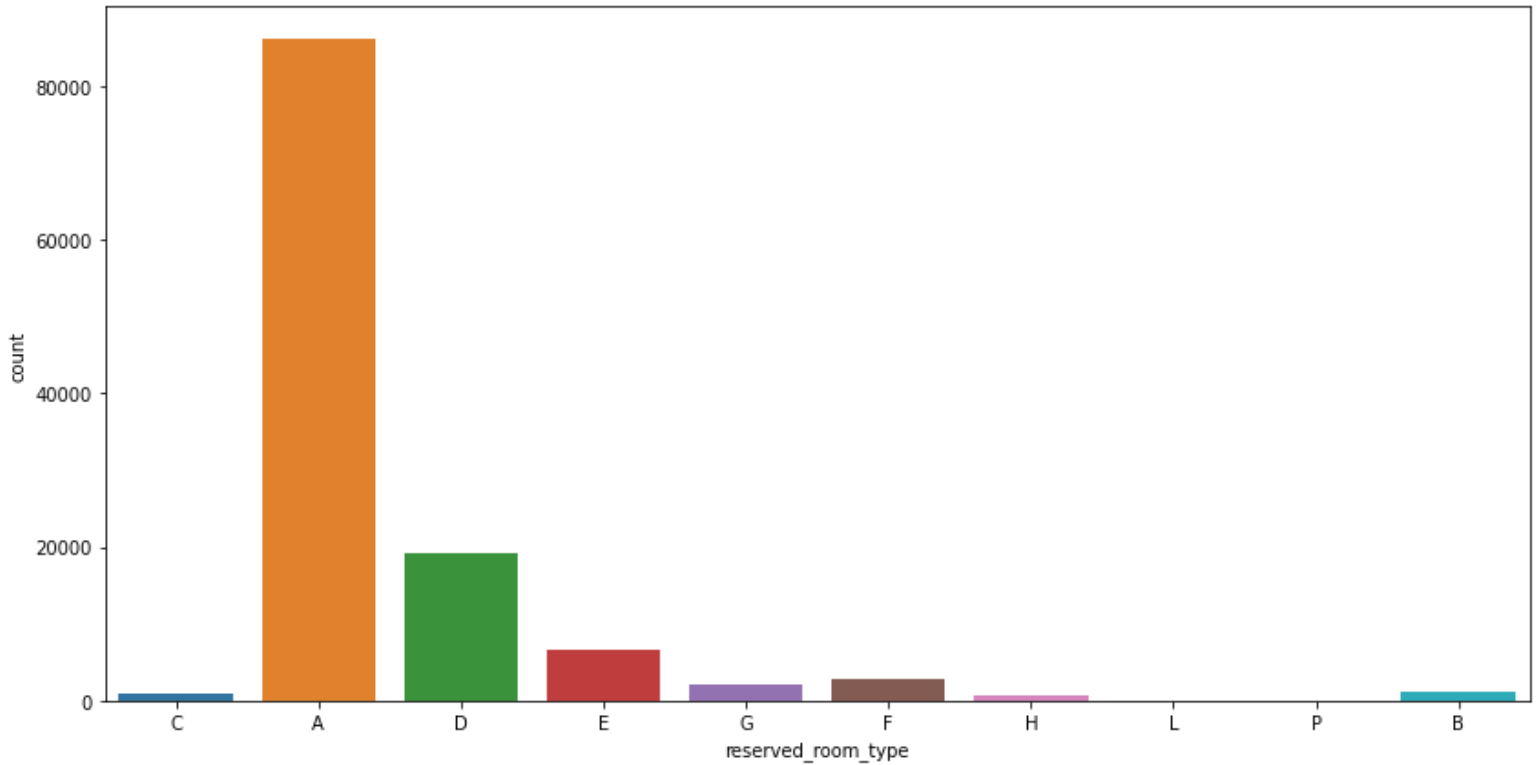
- Count of Customers that were repeated.



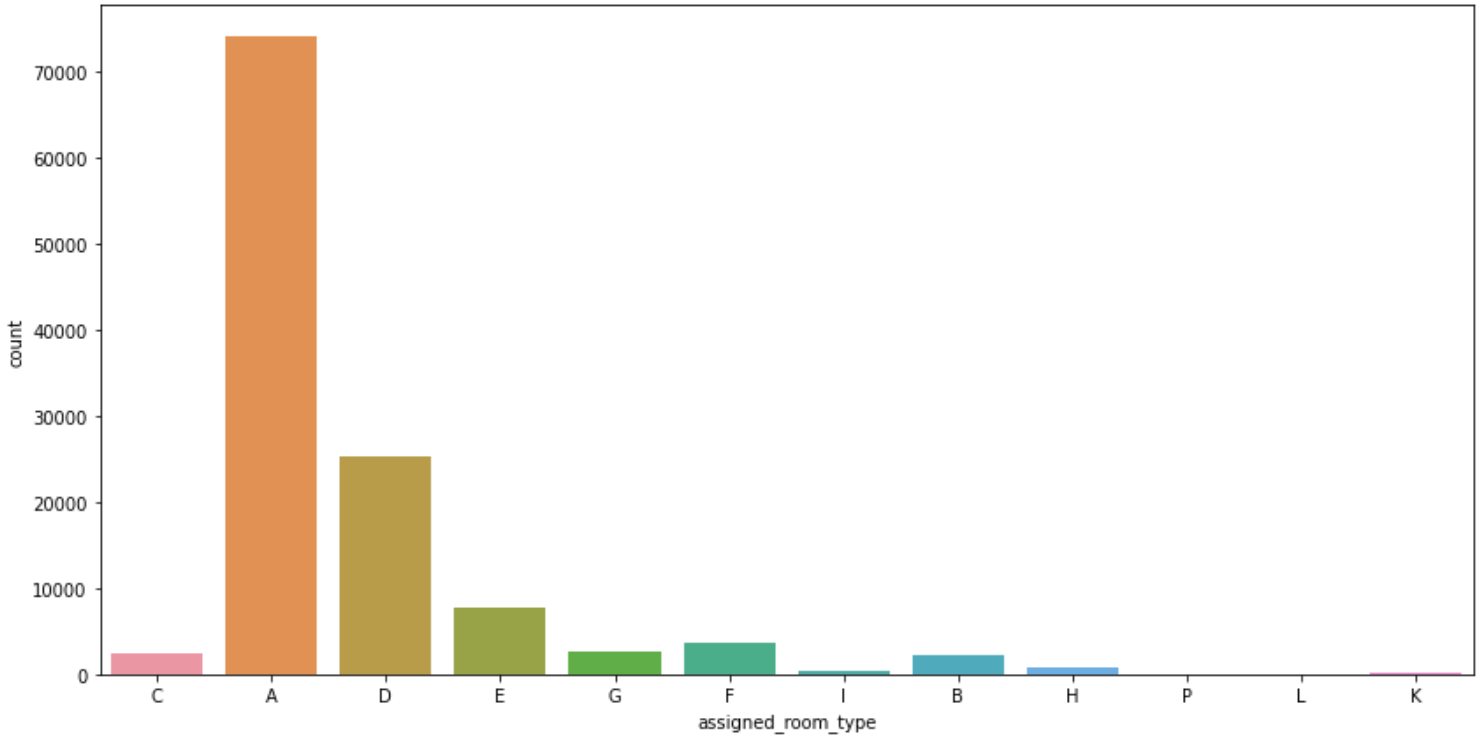
Count of customers based on customer type



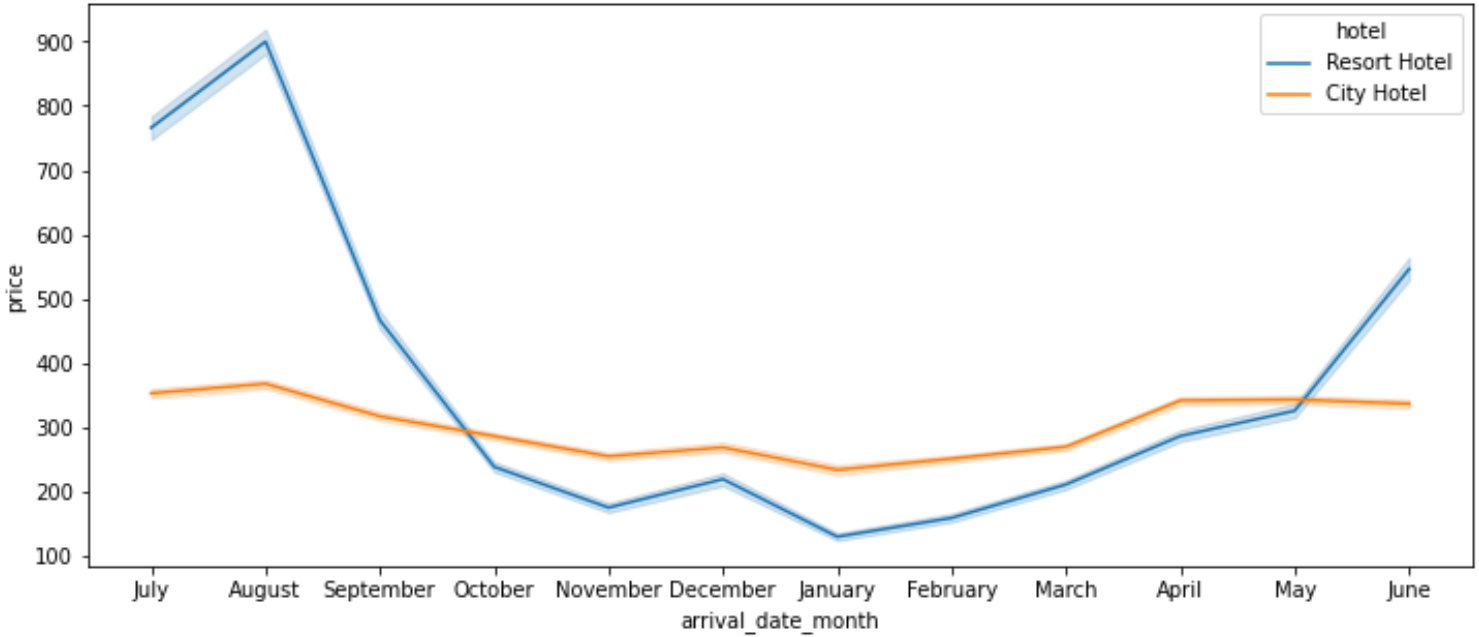
Type of rooms which were most reserved.



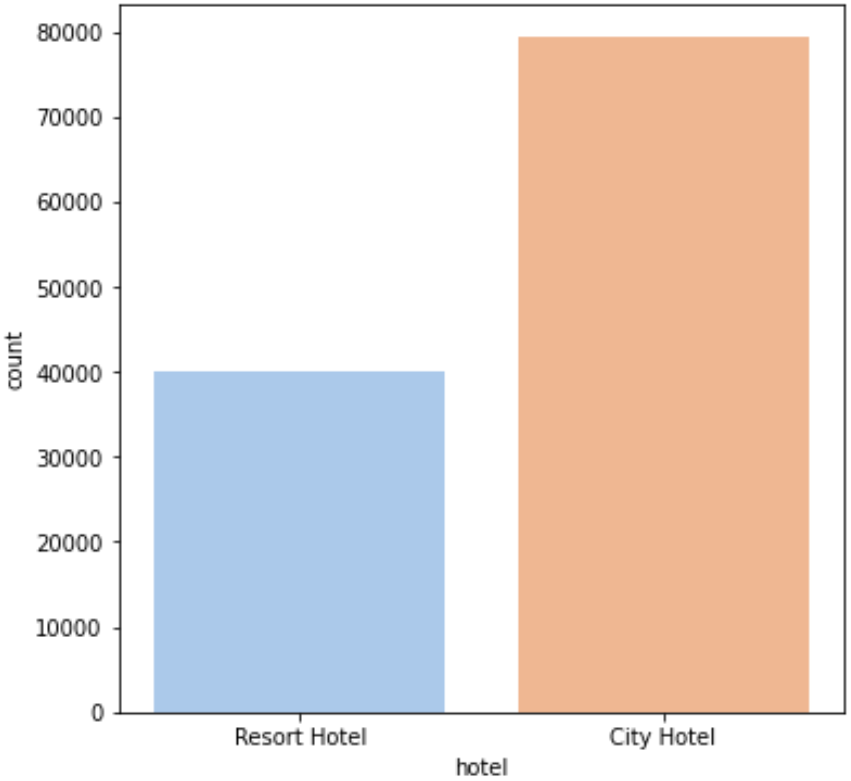
Type of rooms that were assigned the most.



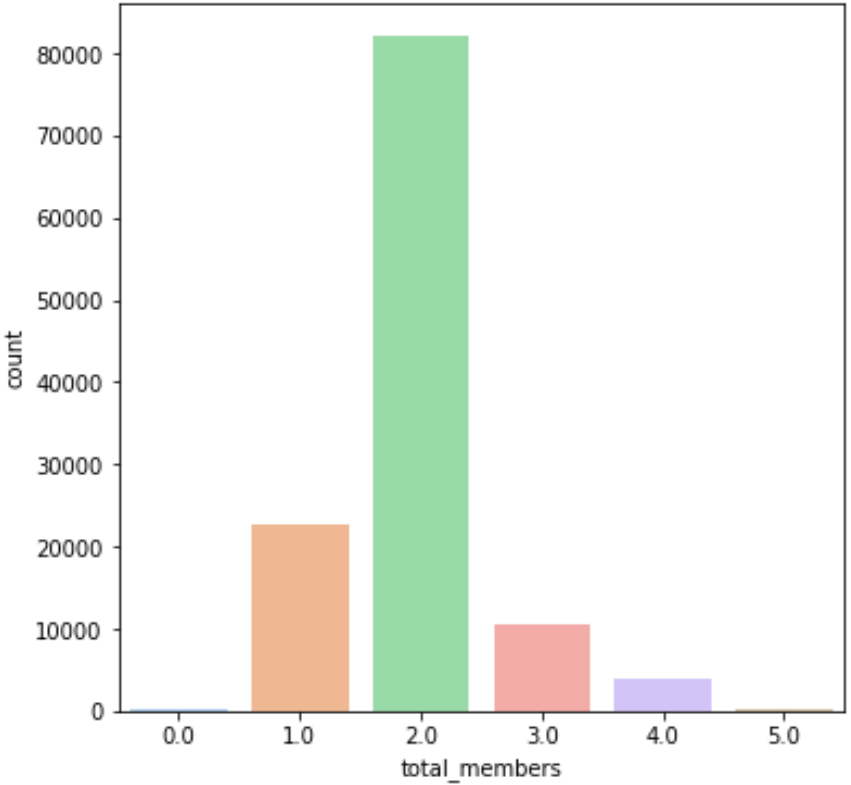
Average Daily Rate (per person)



Count of customer



Family members of customers.



Conclusions

- **Portugal, Great Britain, and France are native place of most customers.**
- **August is the MOST busiest month.**
- **Most of our customers were brought in by Online Travel Agents.**
- **Bed and Breakfast is most preferred meal package.**
- **In 2016, we had most customers considering overall customers whereas city hotel had more bookings compared to resort hotels.**
- **Most of the time, people come in pairs.**
- **Room Type A is MOST favourite room type among customers.**