

# Capstone Project

## Hotel Booking Analysis

### EDA analysis

Debashish Das  
Lucky Jain  
Vivek Katolkar

# Content :-

- Importing and loading data of Hotel booking analysis
- Data summary
- Data cleaning
- Data preparation
- Questions which we asked our Data
- Challenges
- Conclusion
- Thank & You

# Steps took to Analysis Hotel Booking Data



Step 1: Importing the necessary Libraries

Step 2: Mounting Google Drive and Creating a file path

Step 3: Importing Dataset From Drive

Step 4: Printing the information about a Data Frame including the index dtype and columns, non-null values and memory usage.

Step 5: We are going to use Pandas describe() view some basic statistical details like percentile, mean, std etc.

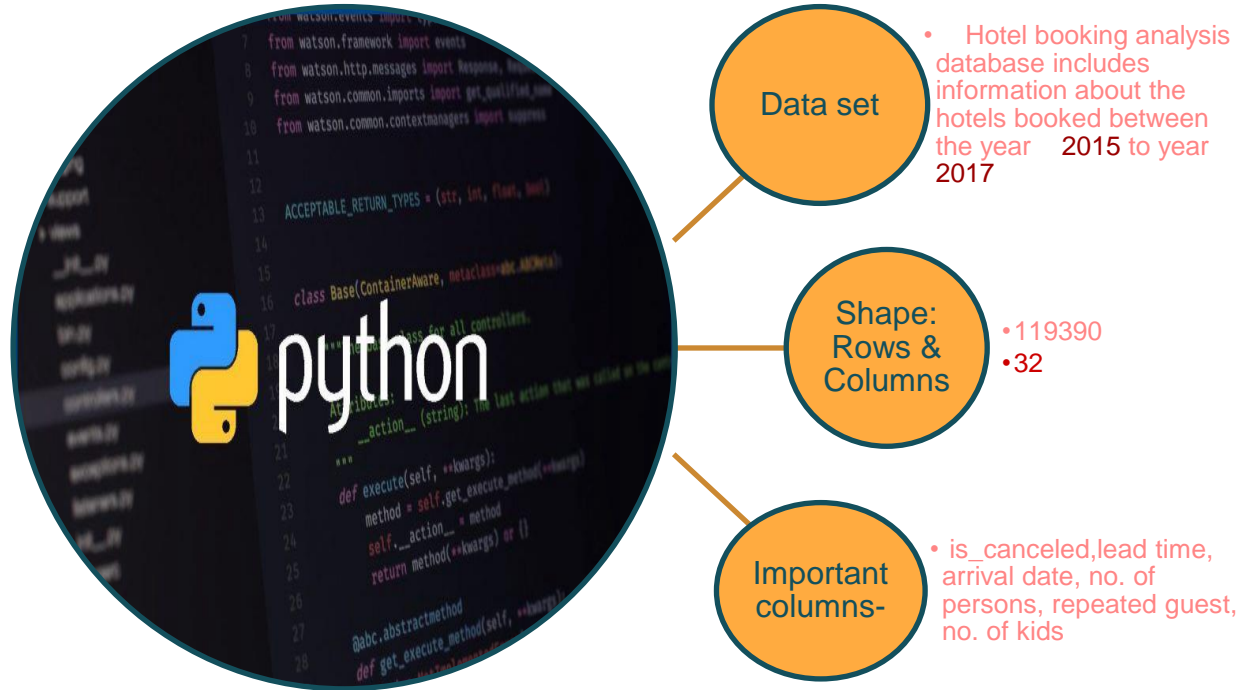
Step 6: Checking the sum of null values present in our dataset

Step 7 : Create New columns

Step 8: Taking Necessary Columns Only

Step 9: Answering Question

# Data summary:-



# Data Cleaning :-

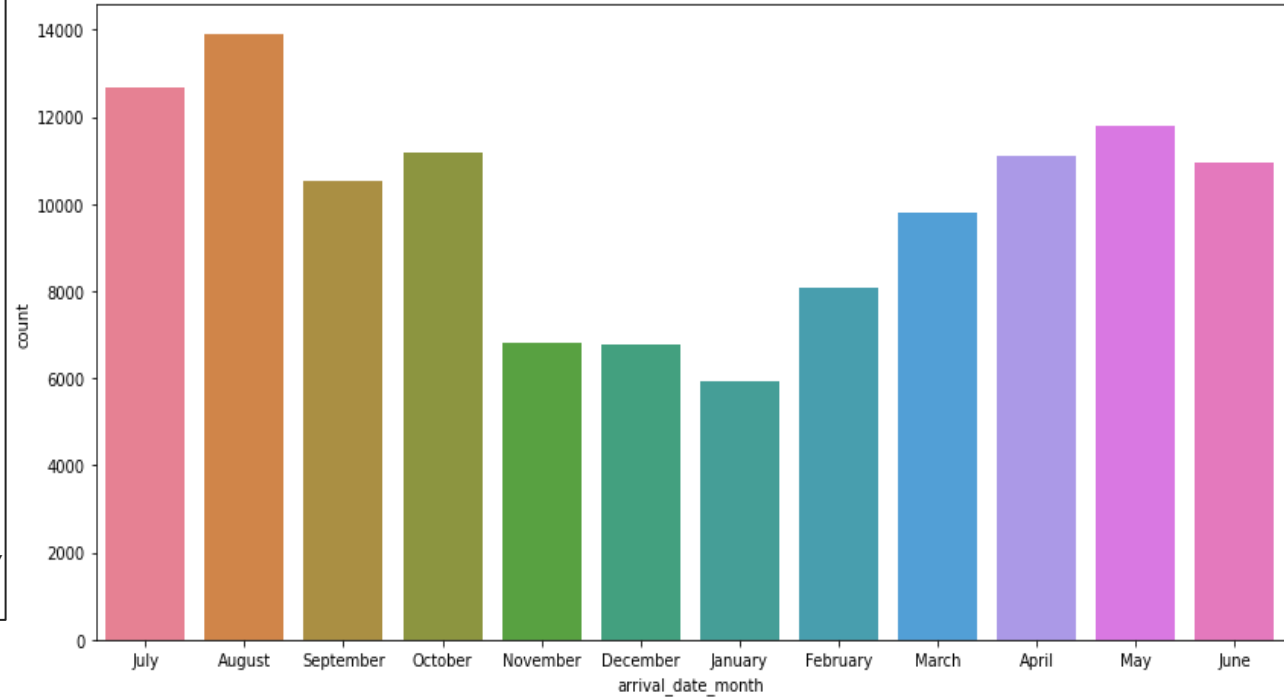
Cleaning data is crucial step before EDA as it will remove the ambiguous data that can affect the outcome of EDA.

While cleaning data we will perform following steps:

- 1) Remove duplicate rows
- 2) Handling missing values.
- 3) Convert columns to appropriate datat ypes.
- 4) Adding important columns

# How many confirmed bookings are there in a month ?

```
August :13877  
July :12661  
May :11791  
October :11160  
April :11089  
June :10939  
September :10508  
March :9794  
February :8068  
November :6794  
December :6780  
January :5929  
Name: arrival_date_month,  
dtype: int64
```

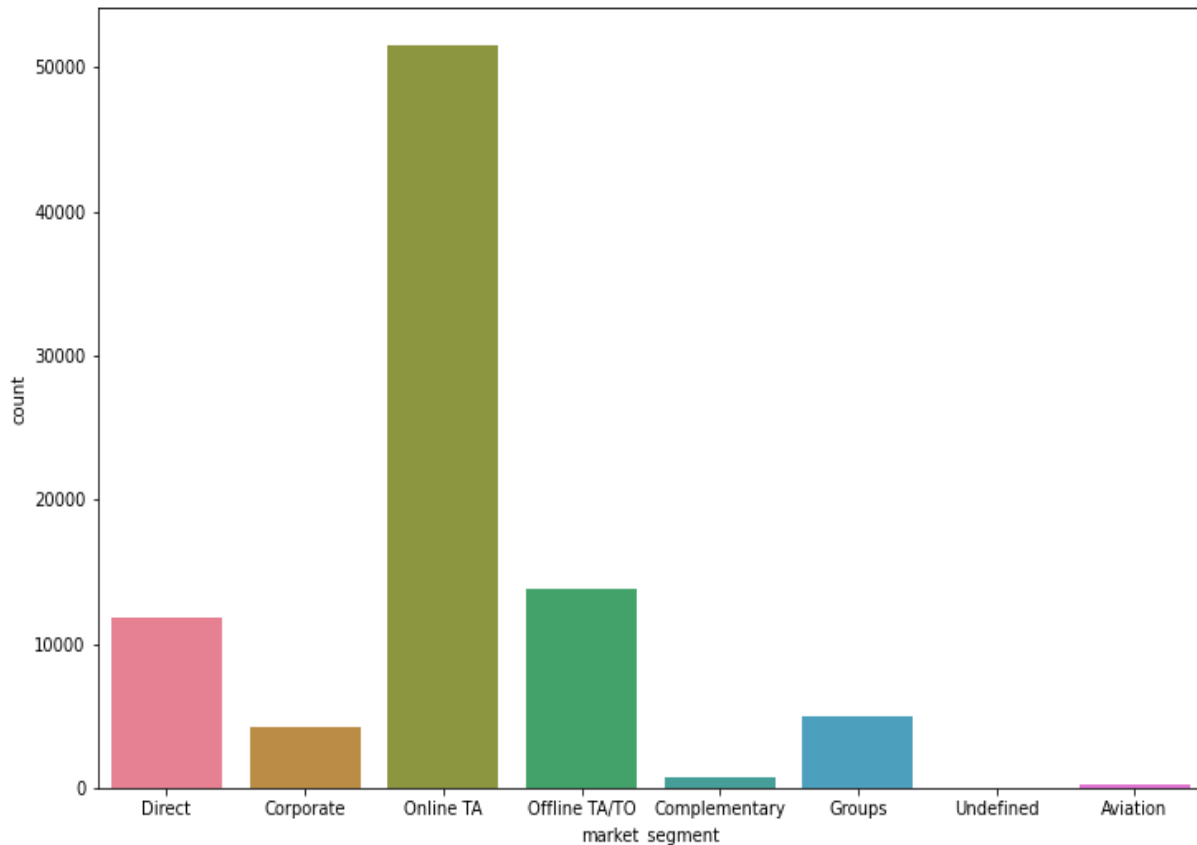


- August Month 13877 is high confirmed booking hotel.
- July month having 2nd Highest 12661 booking confirmed
- May month having 3<sup>rd</sup> highest 11791 booking conform.

# Which month get most visitors?

```
Online TA 0.591001
Offline TA/TO 0.158833
Direct 0.135045
Groups 0.056426
Corporate 0.048149
Complementary 0.007933
Aviation 0.002591
Undefined 0.000023
```

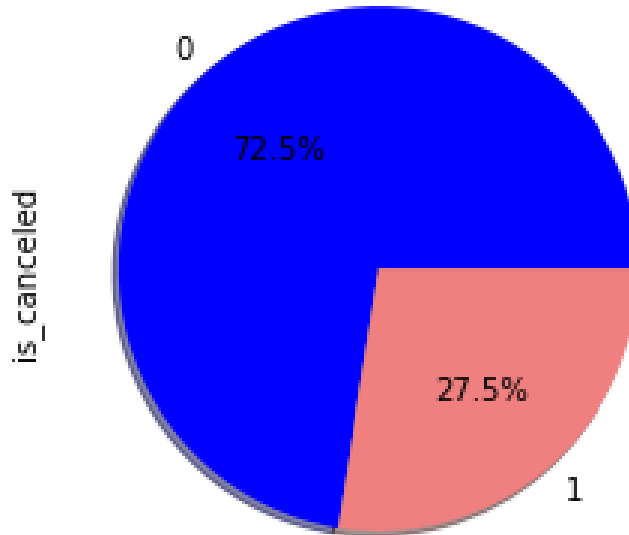
- Around 59% of bookings are made via Online Travel Agents,
- Almost 16% of bookings are made via Offline Travel
- Agents and less than 15% are Direct bookings without any other agents



# How many cancelled are there after booking?

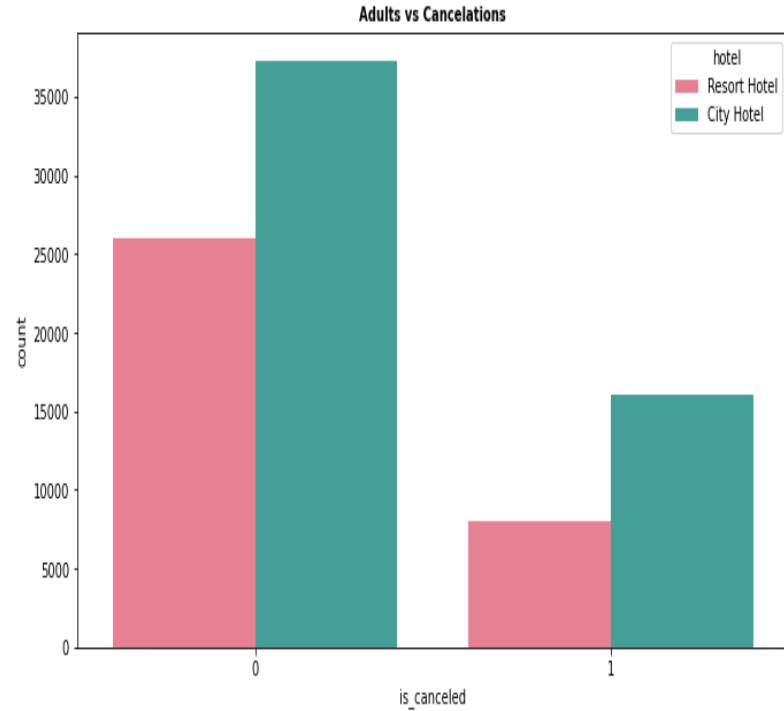
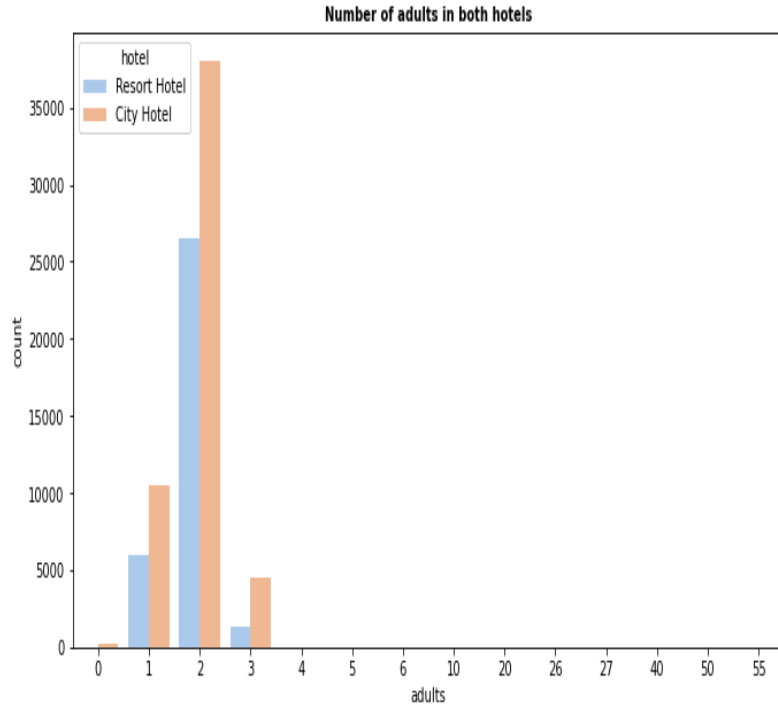
```
Total Bookings cancelled
0 63221 1 24009
  Name: is_cancelled,
  dtype: int64
Cancellation percentage 0
0.724762 1 0.275238
  Name: is_cancelled,
  dtype: float64
```

- According to the pie chart, 72.5% of bookings were not cancelled and 27.5% of the bookings were canceled at the Hotel.



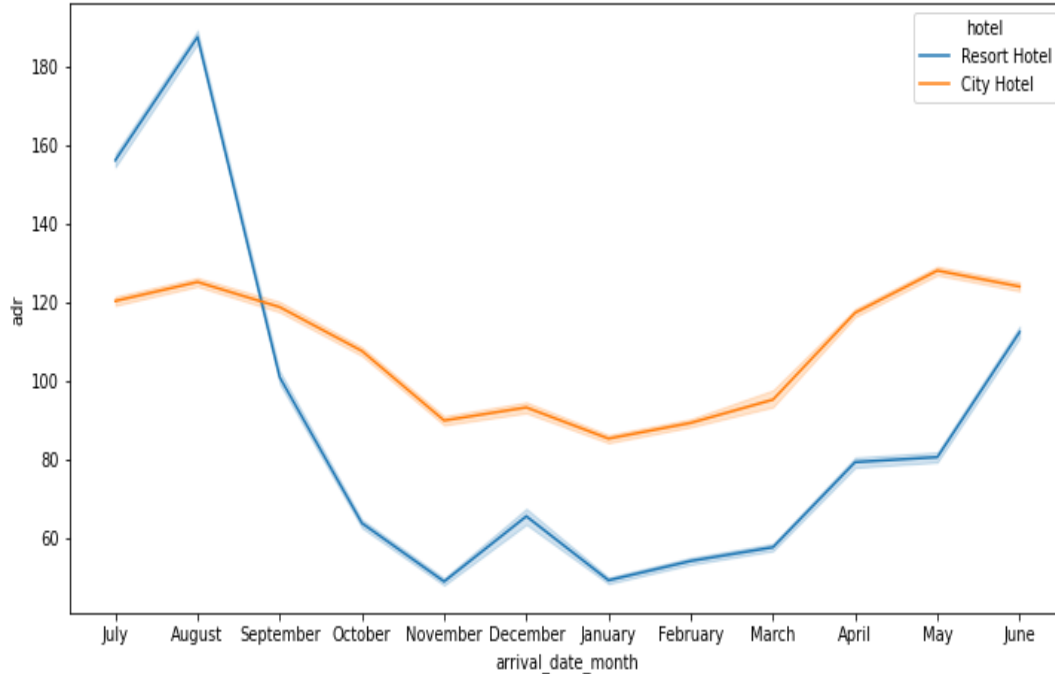


# Which type of hotels are preferred by adults ?



Adult most prefer city hotel, adult cancelled are less.

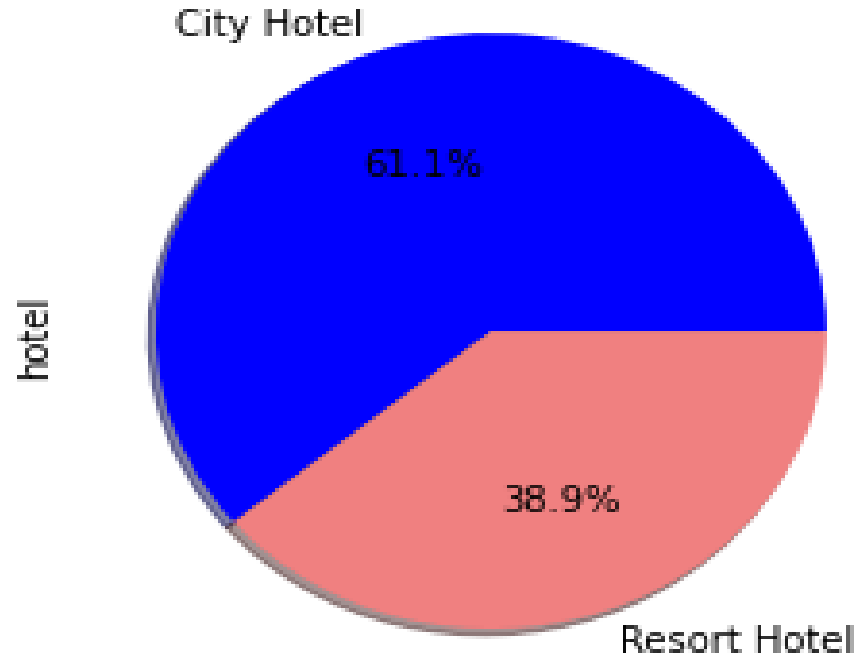
# Which month are most expensive hotels ?



- For resort hotels, the average daily rate is more expensive during August, July and September.
- For city hotels, the average daily rate is more expensive during August, July, June and May.

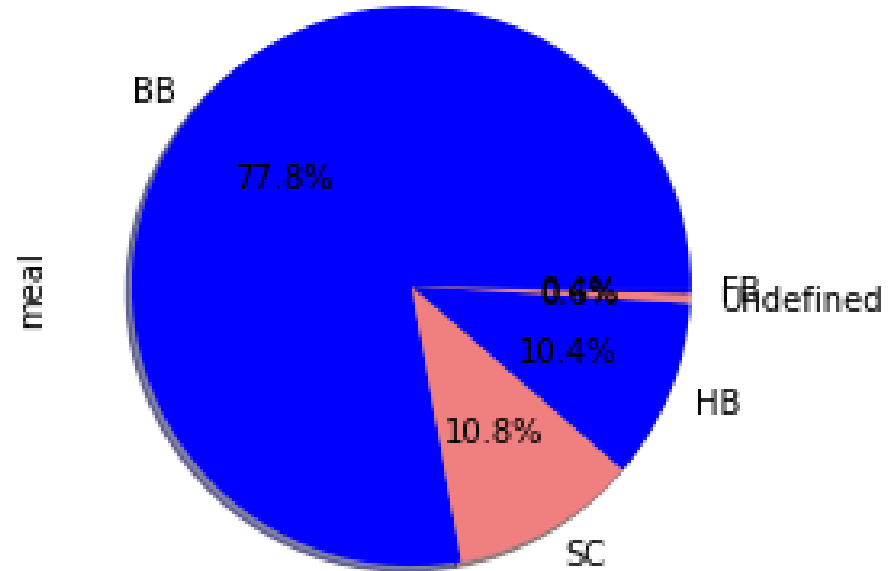
# Which type of hotel most booking percentage?

- 61.1% of people prefers booking .
- 38.9% of people prefers resort hotel booking



# Which type of most prefer meal ?

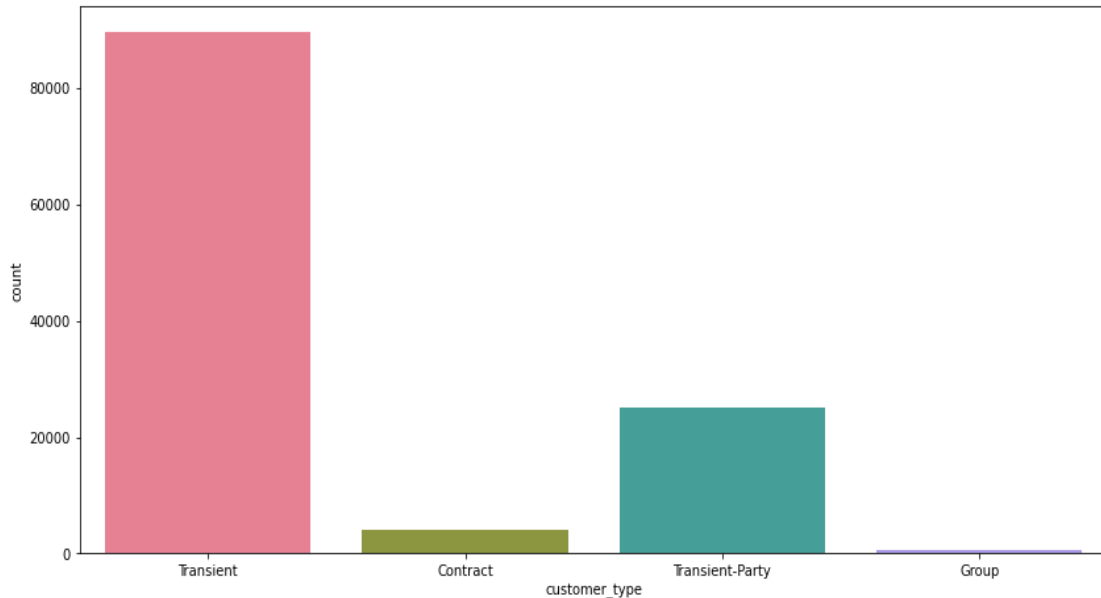
- 77.8% people prefer (BB — Bed & Breakfast) meal
- 10.8% people prefer( SC /undefined )meal
- 10.4% people prefer Half board (breakfast and one other meal — usually dinner)



# Which customer type are having most booking ?

```
Transient 0.823822  
Transient-Party 0.134036  
Contract 0.035939 Group 0.006202 Name:  
customer_type, dtype: float64
```

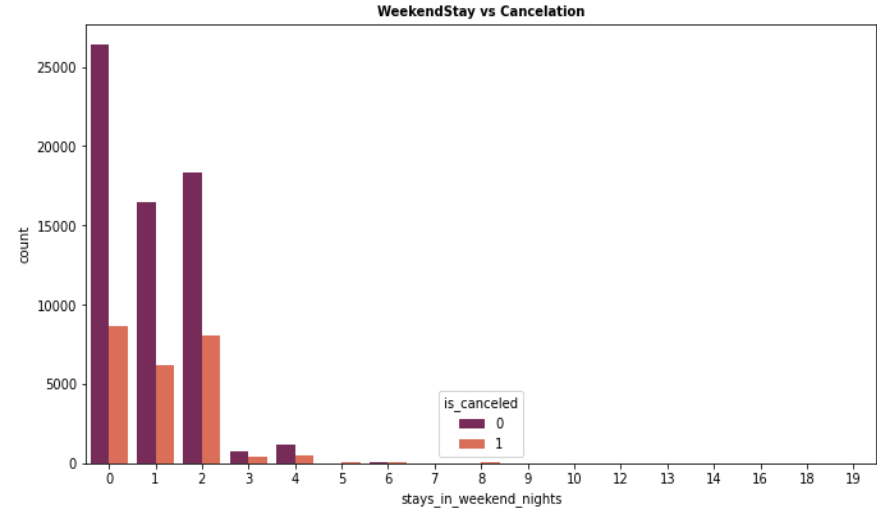
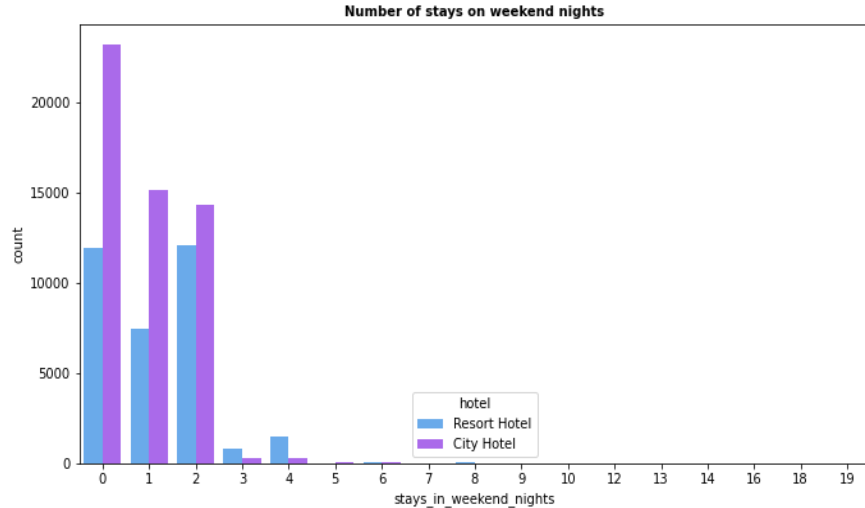
- 82% of Transient are having customer type most booking.
- 13% Transient –party are having customer type most booking



# Questions which we are going to ask our Data

- How many confirmed bookings are there in a month ?
- How many confirmed bookings are there in a month ?
- Which month get most visitors?
- Which type of hotels are preferred by adults.
- Which month are most expensive hotels?
- Which type of hotel most booking percentage?
- Which type of most prefer meal?
- Which customer type are having most booking ?
- Which hotel type most booking in weekend night and cancelled?
- How many family member per reservation ?
- How many customer have booked and then cancelled?
- Which months are having most expensive hotels?

# Which hotel type most booking in weekend night and cancelled?

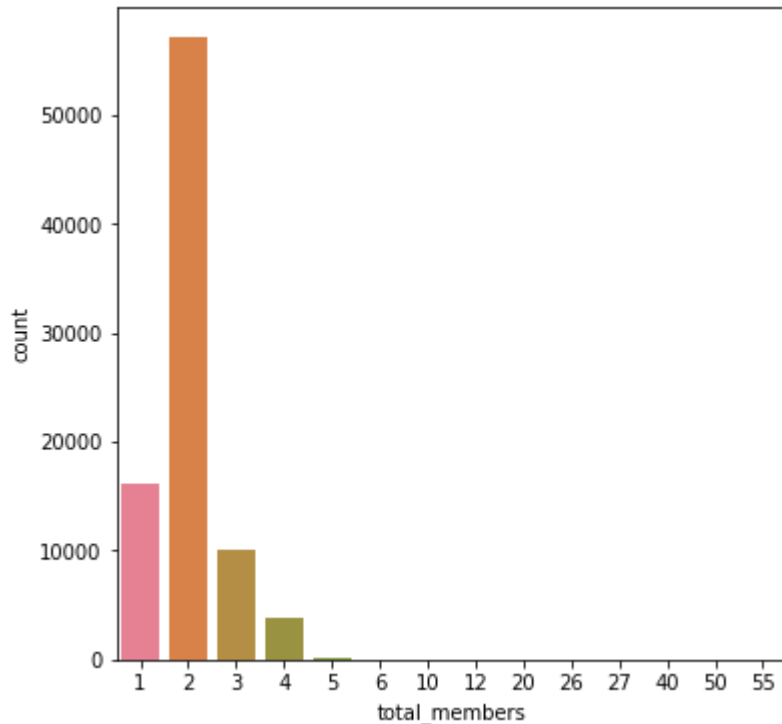


- In the first graph we can see that most of the weekend nights were booked in City Hotel

Second plot shows most of weekend nights which were booked were not canceled

# How many family member per reservation ?

- 65% of 2 people in family are having most reservation.
- 18% of 1 people are having most reservation.
- 11% of 3 people in family are having most reservation





# How many customer have booked and then cancelled?

```
Check-Out 0.724762
```

```
Canceled 0.263625
```

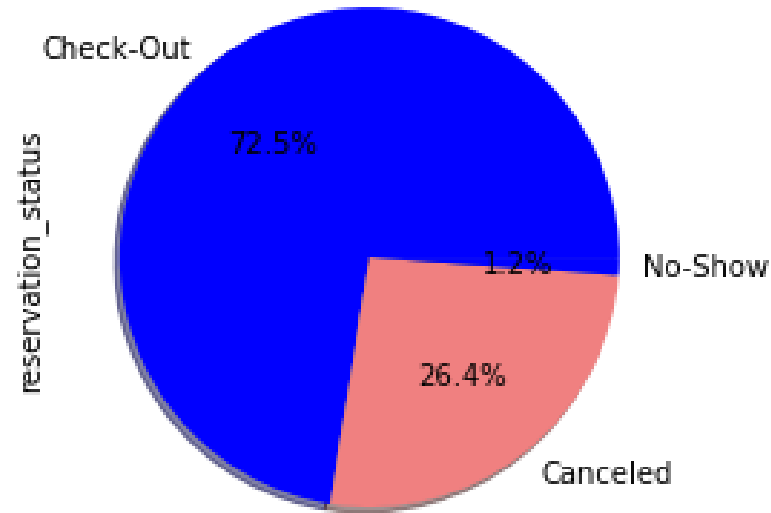
```
No-Show 0.011613
```

```
Name:
```

```
reservation_status,
```

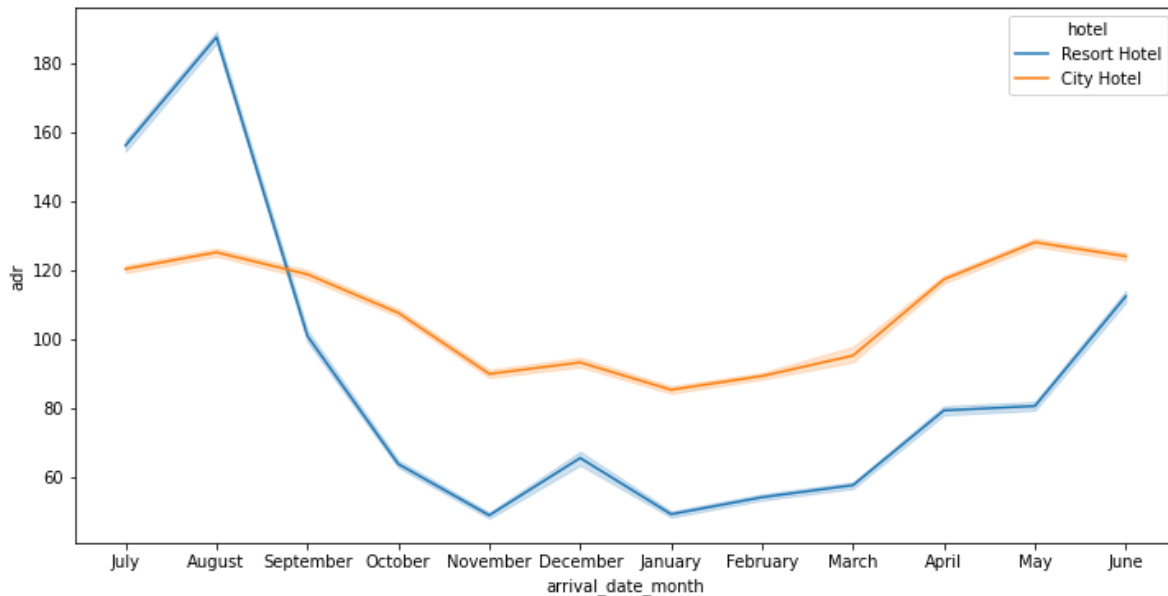
```
dtype: float64
```

- 26 % of after booking canceled
- 72% of customers are check-out



# Which months are having most expensive hotels?

- For resort hotels, the average daily rate is more expensive during august, July and September.
- For city hotels, the average daily rate is more expensive during august, July, June and may.



# CHALLENGES :-

- The name of the countries was not in the proper format ,because of which we are not able to plot the geomean plot.
- column has lots of duplicate value a.
- Company, agent and country column has lots of null value values a.
- There were many rows with almost similar data
- create some new column

## Conclusion :-

- Majority of the guests are from Western Europe. We should spend a significant amount of our budget on those area .Encourage Direct bookings by offering special discounts
- Majority of the hotels are booked by city hotels. Definitely need to spend the most targeting fund on those hotel.
- The number of repeated guests are too low. we should target our advertisement on guests to increase returning guests.
- The majority of reservations converts into successful transactions.
- We have also realize that the high rate of cancellations can be due to high no deposit policies.
- We should also target months between May to Aug. Those are peak months due to the summer period.

### Strategies to Counter High Cancellations at the Hotel

- Set Non-refundable Rates, Collect deposits, and implement more rigid cancellation policies.
- Monitor where the cancellations are coming from such as Market Segment, distribution channels, etc.

Thank You