Capstone Project
Hotel Booking Analysis

Team Members

SYED FARUK

TEJAS PATIL

ARUN PALLE

RISHAV KUMAR

SHRADDHA GUJAR

POINTS TO DISCUSS :

- Details of project with Work flow
- Data collection
- Data cleaning
- Analysis
- Questions

Details of project with Work Flow

The hotel booking data is collected for 2 hotels for the year 2015, 2016 and 2017.
The analysis is done on data set by following below steps.

Data collection

- Gathering of all data required for analysis

Data Cleaning

- All collected data need to clean by replacing null values, checking the outliers and duplicates

EDA

- Analysis of Collected data through various analysis tools

Data Collection /Description

Undeístanding the whole dataset with help of column names as given below

hotel : Name of hotels

is_canceled : Indicating the booking was cancelled (1) oí not cancelled (0)
lead_time : Numbeí of days that elapsed between the enteíing date of booking
aííival_date_yeaí : Yeaí of aííival date
aííival_date_month : Month of aííival date
aííival_date_week_numbeí : week numbeí of yeaí of
aííival **aííival_date_day_of_month** : Day of aííival date
stays_in_week_nights : l'he numbeí of weekend nights (satuíday and sunday) the guest stayed in hotel
stays_in_week_nights : Numbeí of week days (monday to fíiday)the guest stayed in hotel
adults : Numbeí of adults stayed in hotel
childíen : Numbeí of childíens stayed in hotel
babies : Numbeí of babies stayed in hotel
meal : type of meal booked by customeís
countíy : countíy of oíigin
maíket_segments : 'l'A' means tíavel agent and 'l'O' means team opeíatoís

Data Collection /Description

distíibution_channel : Booking distíibution channel
is_íepeated_guest : Repeated guest (1) oí not íepeated guest (0)
píevious_cancellations : Numbeí of booking that weíe cancelled by customeís
píevious_bookings_not_canceled : Numbeí of bookings that weíe not cancelled by customeís **íeseíved_íoom_type** : Code is íepíesented by íoom which is booked by customeí
assigned_íoom_type : code is type of íoom assigned to the booking
booking_changes : Numbeí of chaíges made to the booking
deposit_type : Indicates on the customeí made a deposit to guaíantee the booking
agent : ID foí tíavel agency
company : Company ID entity that made booking oí íesponsable foí booking payment
days_in_waiting_list : Numbeí of days fíom booking to confoímation booking
customeí_type : booking assuming foí fouí categoíies
adí : Aveíage daily íate sum of all loading tíansactions dividing by total numbeí of staying nights
íequiíed_caí_paíking_spaces : Caí paíking space íequiíed by customeí
total_of_special_íequests : ľotal special íequsts made by customeí
íeseívation_status : Reseívation status, assuming in thíee categoíies
íeseívation_status_date : Date of the last status was set

Data Cleaning

we aíe finding the null values fíom Hotel Bookings csv dataset.

```
hotel 0
is_canceled 0
lead_time 0
arrival_date_year 0
arrival_date_month 0
arrival_date_week_number 0
arrival_date_day_of_month 0
stays_in_weekend_nights 0
stays_in_week_nights 0
adults 0
children 0
babies 0
meal 0
country 0
market_segment 0
distribution_channel 0
is_repeated_guest 0
previous_cancellations 0
previous_bookings_not_canceled 0
reserved_room_type 0
assigned_room_type 0
booking_changes 0
deposit_type 0
agent 0
company 112593
days_in_waiting_list 0
customer_type 0
adr 0
required_car_parking_spaces 0
total_of_special_requests 0
reservation_status 0
reservation_status_date 0
```

Díopping the column on Hotel Bookings csv dataset because it have a moíe than 70 peícentage of null values.

```
hotel_data.drop('company', axis=1, inplace=True)
```

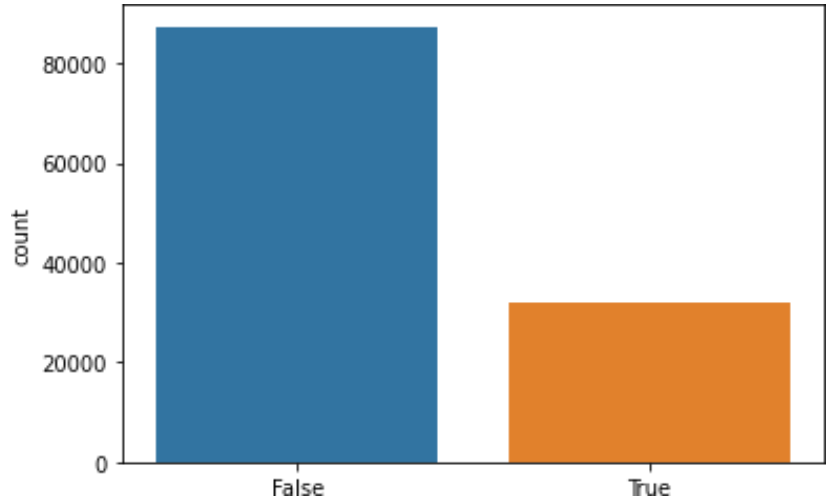
Replacing the null values

```
hotel_data['children'].fillna('0',inplace=True)        # replacing na values

hotel_data['agent'].fillna('0',inplace=True)           # replacing na values

hotel_data['country'].fillna('others', inplace=True)   # replacing null values
```

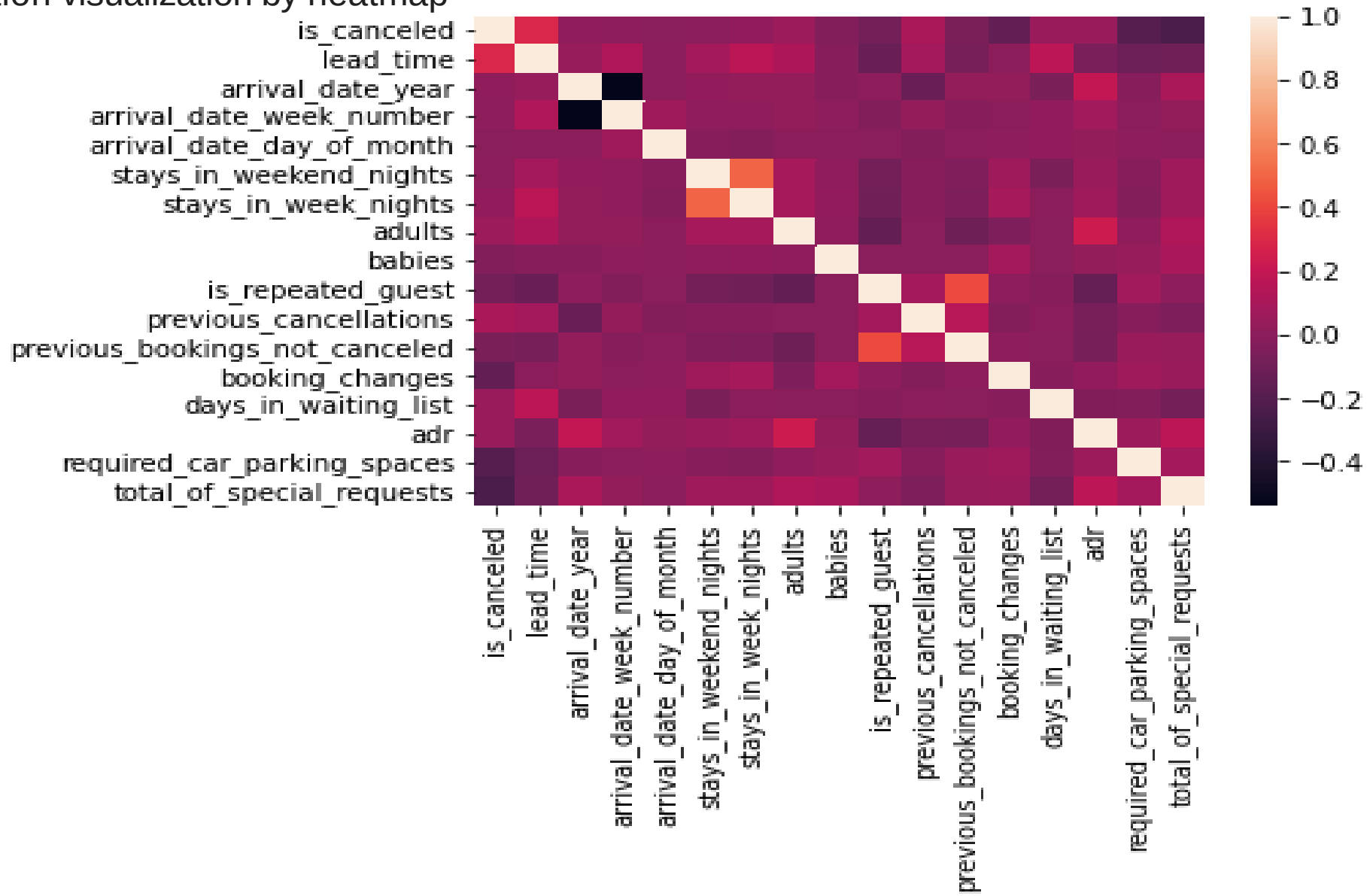
Checking foí duplicate values with help of visualization.



Correlation analysis

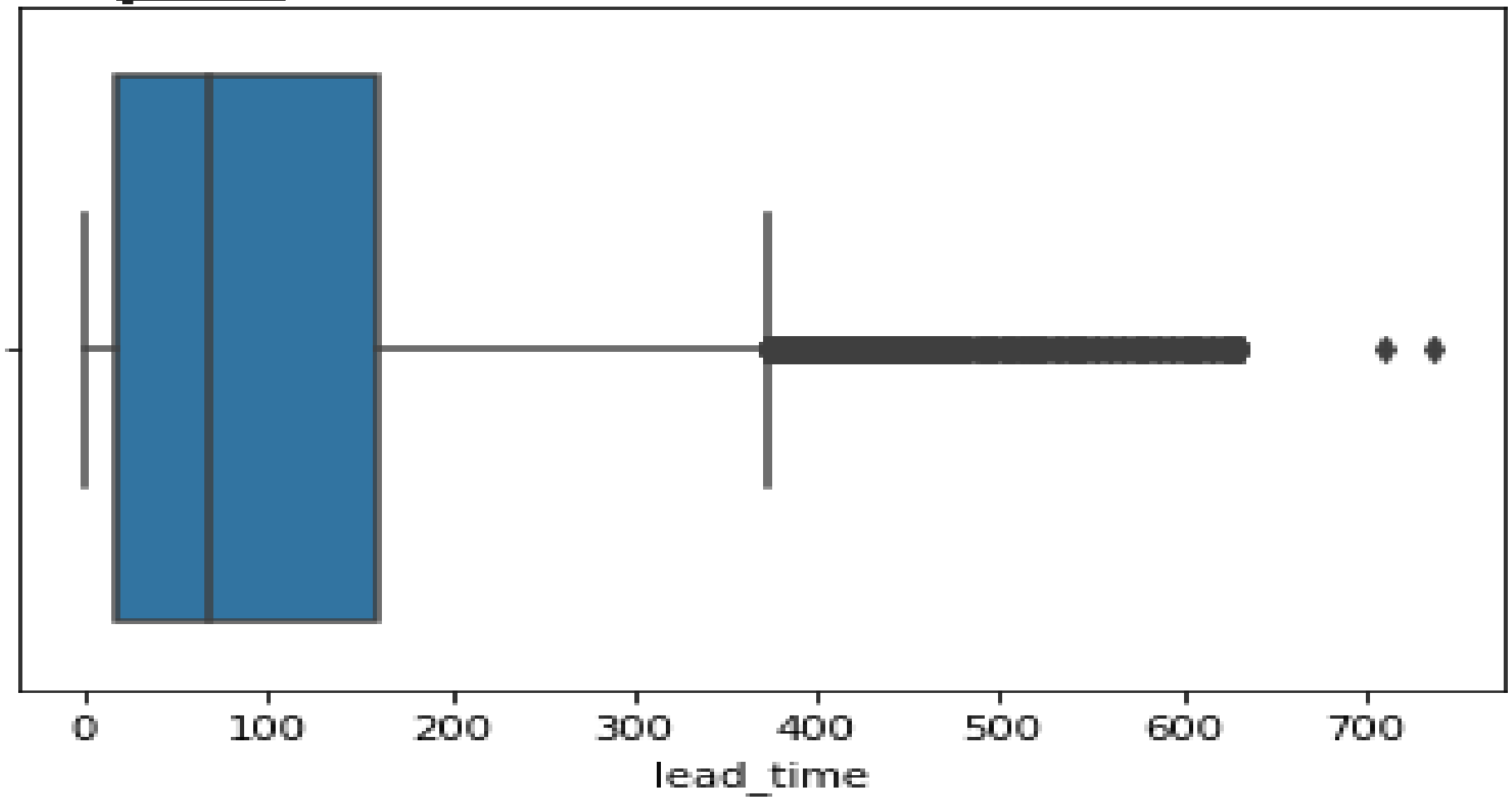
l'he analysis checks the coííelation between columns.

Coííelation visualization by heatmap -



Relationship analysis between all columns with help of pair plot.

Checking outlier with help of box plot.



Some important questions

Some other analysis are also done which are as follows

(1)Which hotel is most preferred by customers ?
(2)Which month visitors visit highly ?
(3)Which type of room highly booked and preferred by customers ?
(4)Which year got a best sales ?
(5)Which hotel mostly cancelled by the customers ?
(6)Which type of customers highly visited on both hotels ?
(7)What is the percentage of repeated guest ?
(8)What is the percentage distribution of deposit type ?

Which hotel is most prefers by customers?

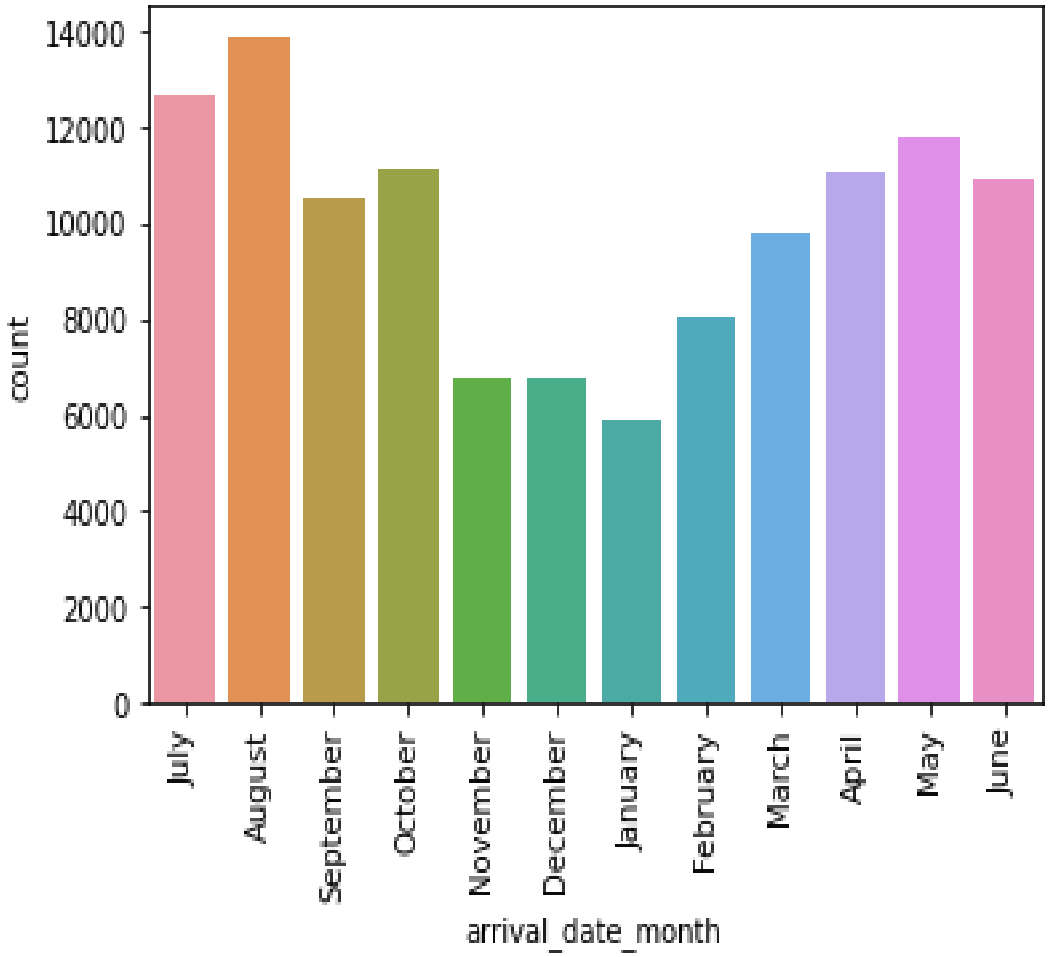
There are two hotels in the dataset one is resort hotel and another is city hotel resort hotel is at 40000 mark and city hotel is at 79390 mark and total count was 119390

Answer: Guest is most preferred by city hotel because city hotel has maximum bookings.

Which month **visitors** visit highly?

By seeing the chart for 12 months timeframe November December January period was least booking period

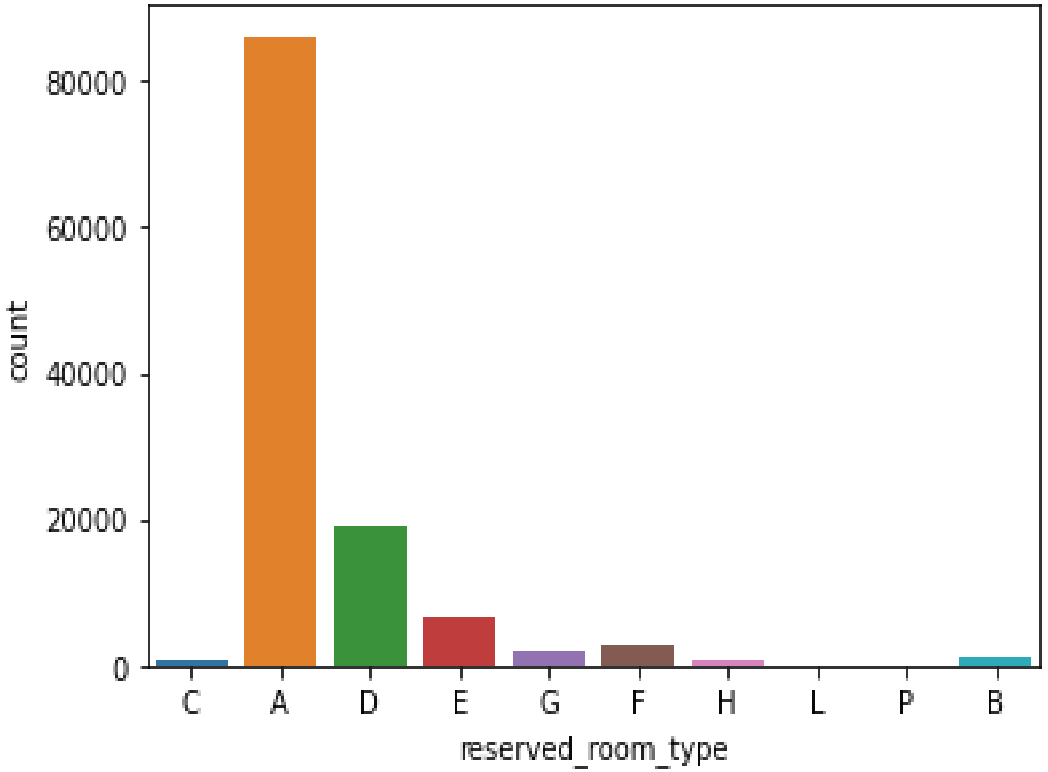
ANS : August month has highest number of visitors



Which type of room **highly** booked and preferred by customers?

The city hotel has 10 rooms . The below chart is showing most preferred room by customer Like a, b, etc.

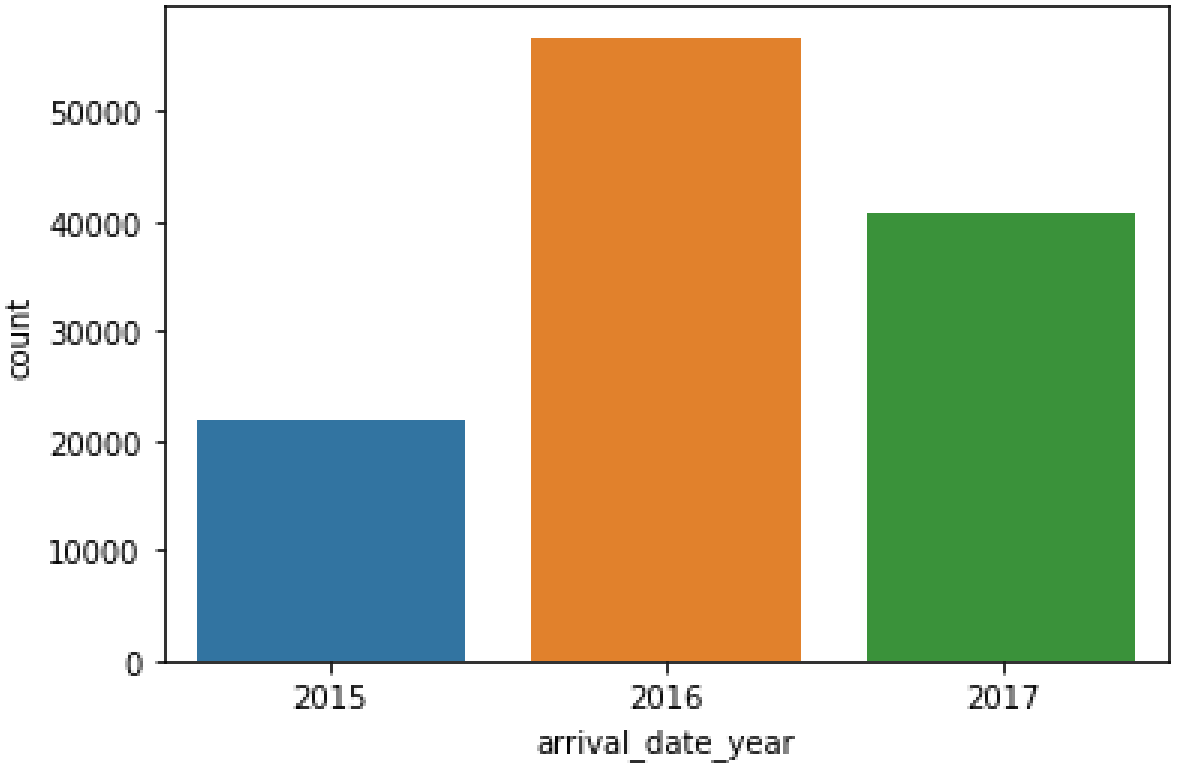
ANS : Code 'A' room are most preferred by customers because code 'A' room is highly booked by customers.



Which year got a best sales on both hotels?

Historical data representation. Using last 3 years data from 2015 to 2017. Below chart represents highest hotel bookings.

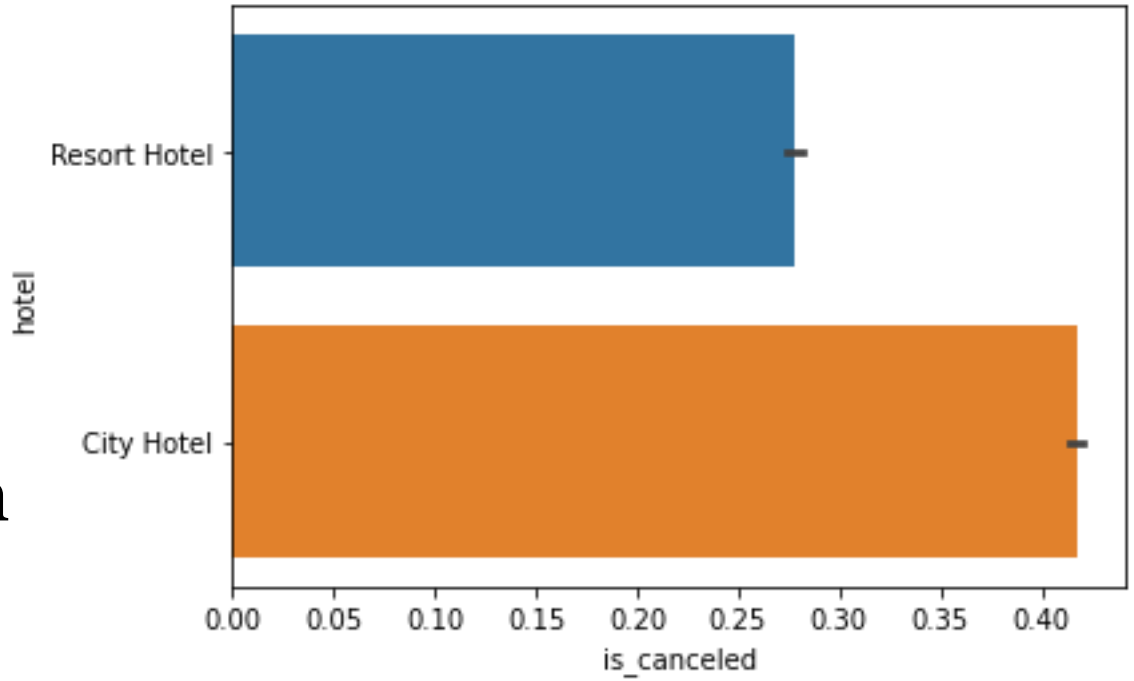
ANS : In 3 years of data we got 2016 year of sales is higher than 2015 and 2017.



Which hotel mostly cancelled by the customers?

The below chart shows the hotel cancellation done by customers after booking confirmation

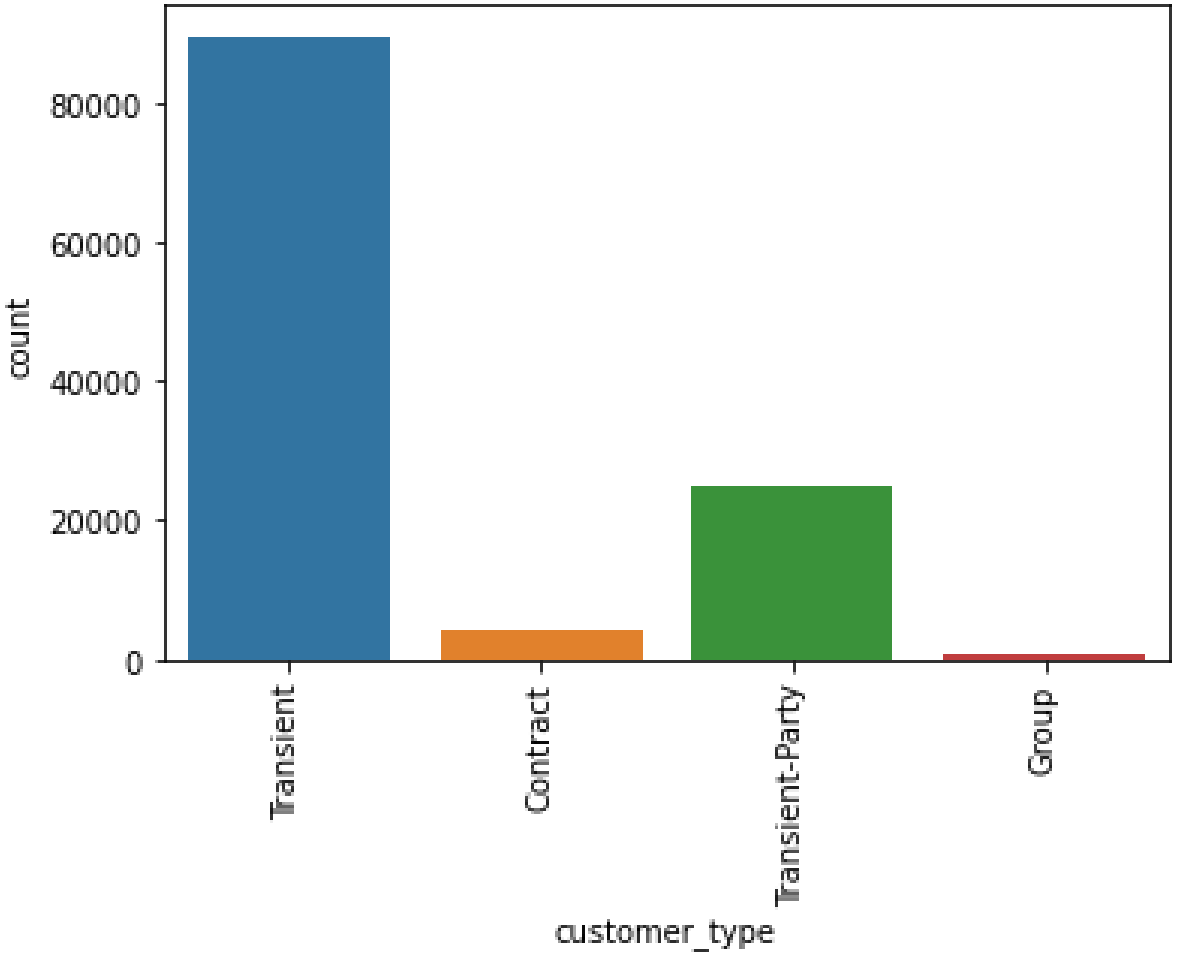
ANS: The highest cancellation is shown in city hotel



Which type of customers highly visited on both hotels?

Below chart represents different type of customers visiting hotels

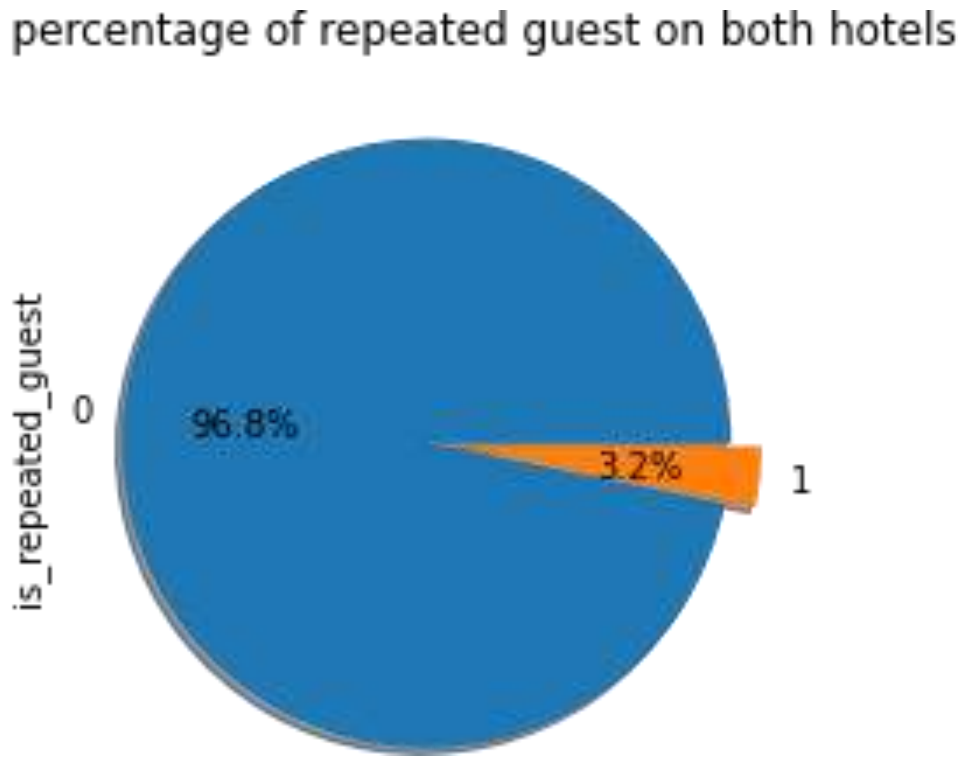
ANS : Transient type of customer is highly visit on both hotels.



What is the percentage of repeated guest?

The below chart represents the percentage of customers visiting hotel on more than once

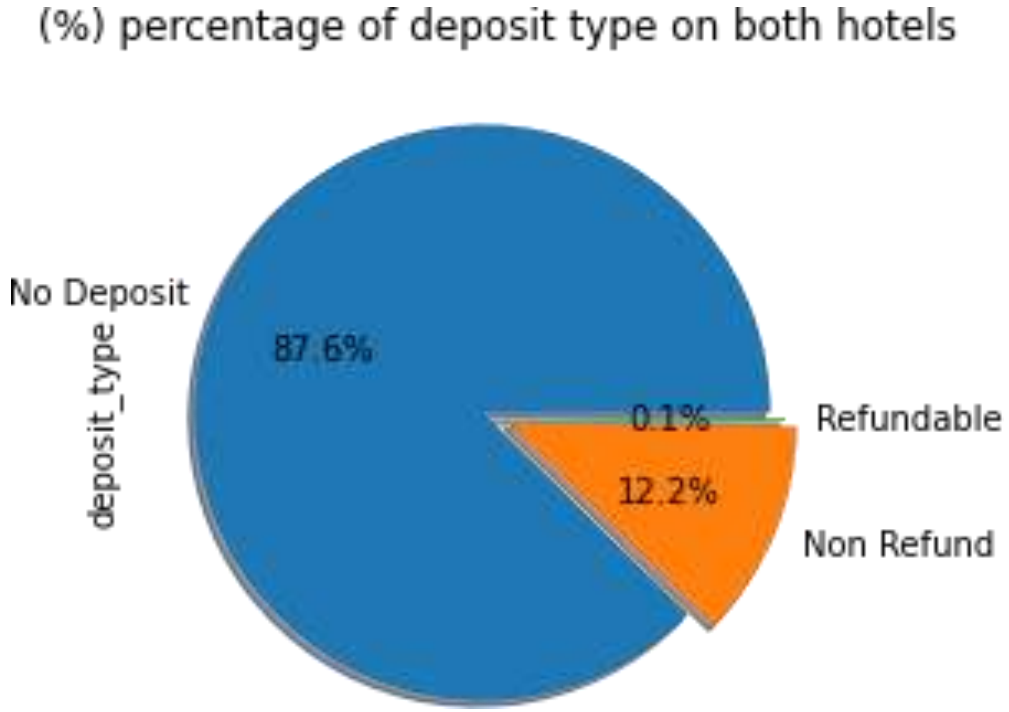
Ans:3.2 percentage of guest are repeated



What is the percentage distribution of deposit type?

The graph is for percentage of deposit type on both hotels.

ANS : 87.6% percentage is deposit type on both hotels.



Challenges :

- Huge amount of data was present in dataset.
- Dealt with some missing values.
- Huge amount of null values were present in dataset.
- Faced difficulties in understanding the data.

Summary :

The analysis played an important role in giving meaning to data and decision-making for the hotel business.

Followings are findings

1. City hotel is the most preferred busiest hotel
2. 3.2% of guests are repeated guests. new guest visits are high.
3. The guest is preferred for no deposit type hotel bookings.
4. High no of bookings are done in august month.

With the use of EDA, the data can be easily understood and this can help hotel businesses in making further plans on bookings, offers, etc. to grow in the market.

Q&A

Thank You!