

# Capstone Project

## Hotel Booking

### Analysis

#### Team Members

SYED FARUK

TEJAS PATIL

ARUN PALLE

RISHAV KUMAR

SHRADDHA GUJAR

# POINTS TO DISCUSS

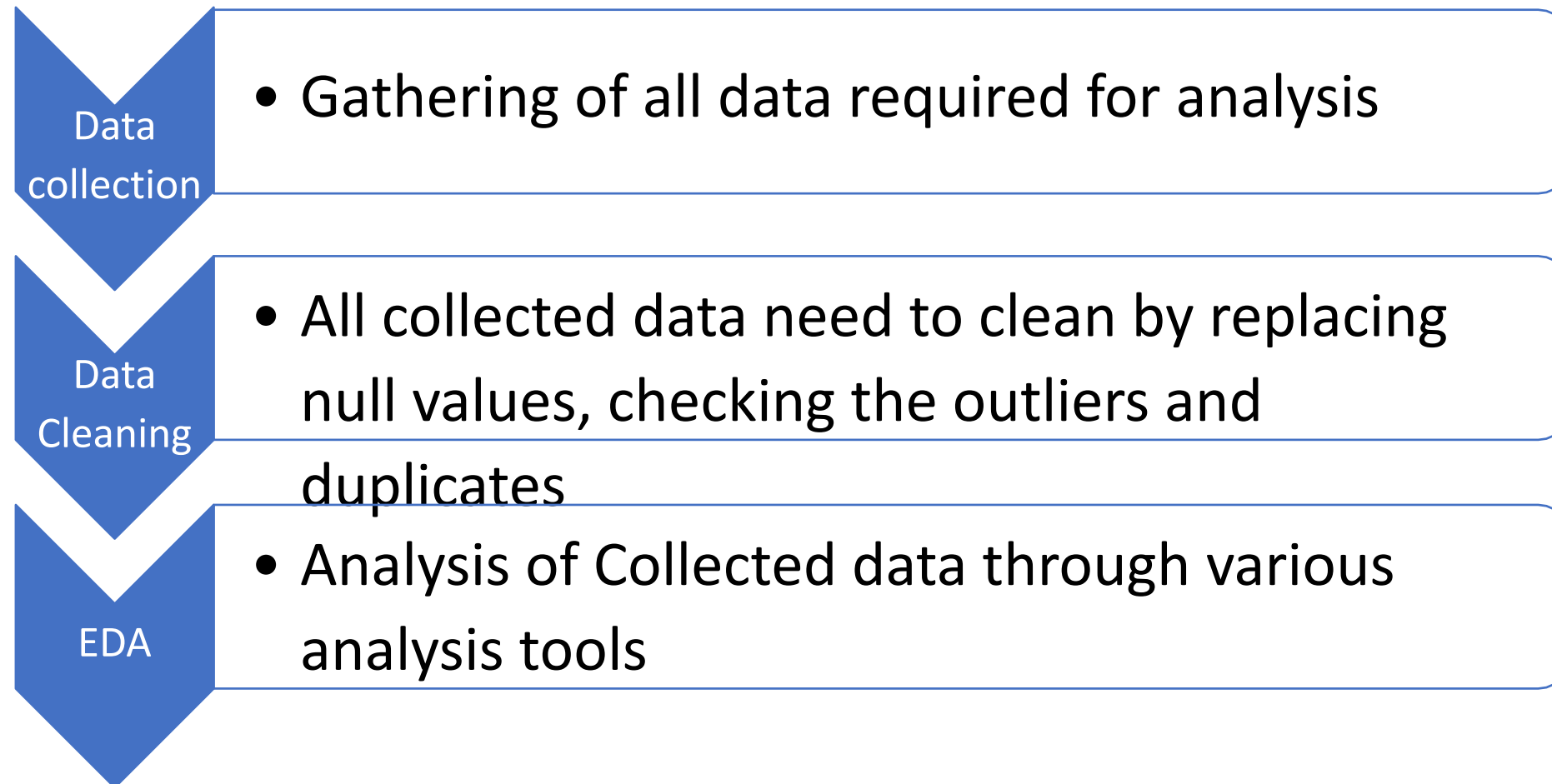
- 
- Details of project with Workflow
- Data collection
- Data cleaning
- Analysis
- Questions



# Details of project with Workflow



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 /Description

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

**hotel** : Name of hotels

**is\_canceled** : Indicating the booking was cancelled (1) or not cancelled (0)

**lead\_time** : Number of days that elapsed between the entering date of booking

**arrival\_date\_year** : Year of arrival date

**arrival\_date\_month** : Month of arrival date

**arrival\_date\_week\_number** : week number of year of arrival

**arrival\_date\_day\_of\_month** : Day of arrival date

**stays\_in\_week\_nights** : The number of weekend nights (Saturday and Sunday) the guest stayed in hotel

**stays\_in\_week\_nights** : Number of week days (Monday to Friday) the guest stayed in hotel

**adults** : Number of adults stayed in hotel

**children** : Number of children stayed in hotel

**babies** : Number of babies stayed in hotel

**meal** : type of meal booked by customer

**country** : country of origin

**market\_segments** : 'I'A' means travel agent and 'I'O' means team operator

## Data Collection /Description

**distribution\_channel** : Booking distribution channel

**is\_repeated\_guest** : Repeated guest (1) or not repeated guest (0)

**previous\_cancellations** : Number of booking that were cancelled by customer

**previous\_bookings\_not\_cancelled** : Number of bookings that were not cancelled by customer

**reserved\_room\_type** : Code is represented by room which is booked by customer

**assigned\_room\_type** : code is type of room assigned to the booking

**booking\_changes** : Number of changes made to the booking

**deposit\_type** : Indicates on the customer made a deposit to guarantee the booking

**agent** : ID of travel agency

**company** : Company ID entity that made booking or responsible for booking payment

**days\_in\_waiting\_list** : Number of days from booking to confirmation booking

**customer\_type** : booking assuming for four categories

**ad** : Average daily rate sum of all loading transactions dividing by total number of staying nights

**required\_car\_parking\_spaces** : Car parking space required by customer **total\_of\_special\_requests** : Total special requests made by customer

**reservation\_status** : Reservation status, assuming in three categories

**reservation\_status\_date** : Date of the last status was set

# Data Cleaning



we are finding the null values from 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
```

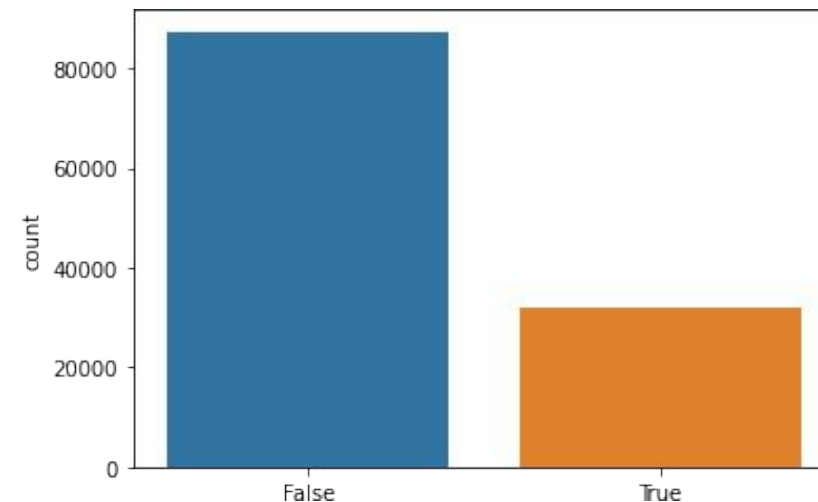
Dropping the column on Hotel Bookings csv dataset because it have a more than 70 percentage of null values.

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

## Replacing the null values

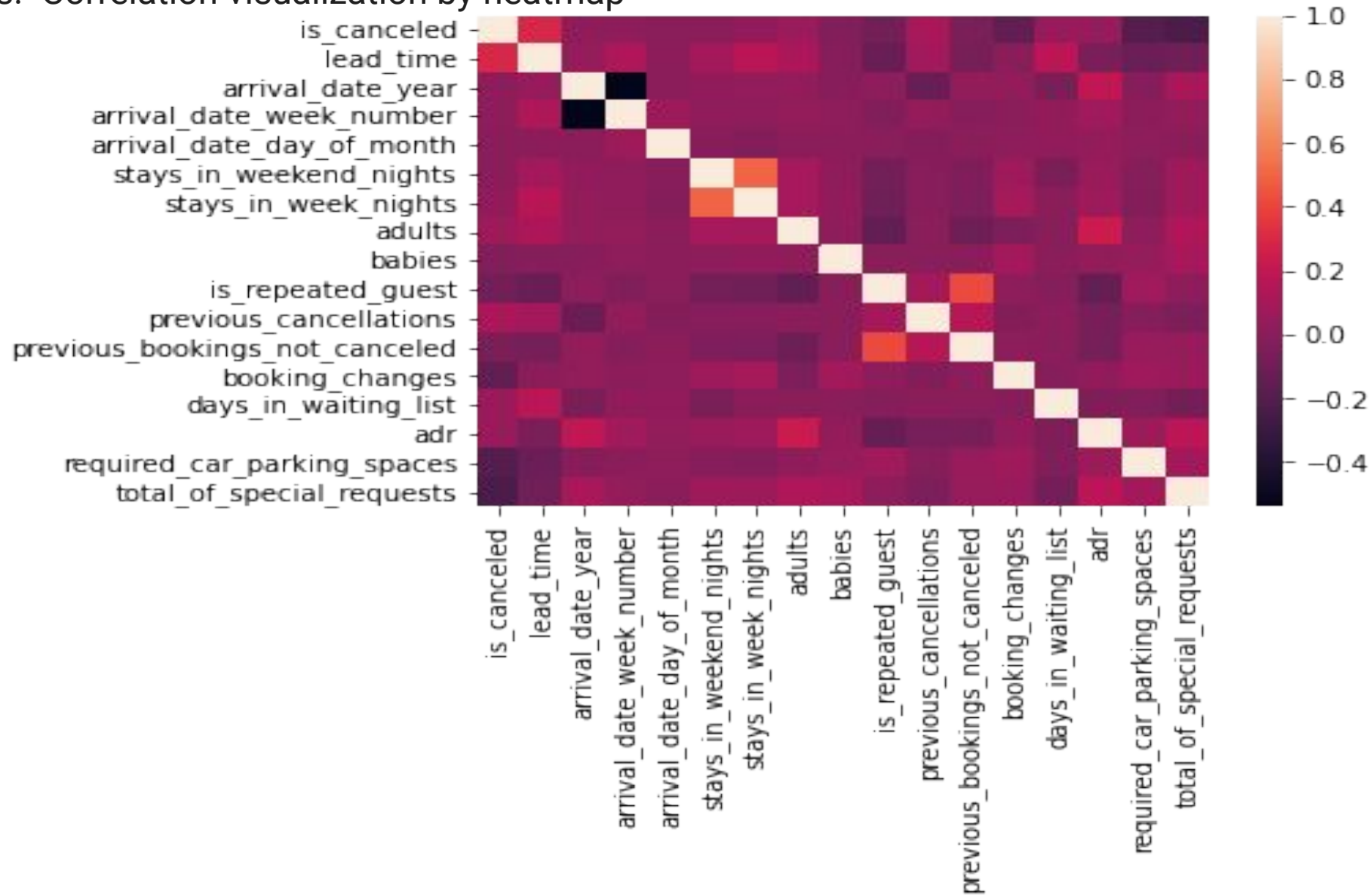
```
hotel_data['children'].fillna('0',inplace=True) # replacing na values
hotel_data['agent'].fillna('0',inplace=True) # replacing na values
hotel_data['country'].fillna('others',inplace=True) # replacing null values
```

Checking for duplicate values with help of visualization.



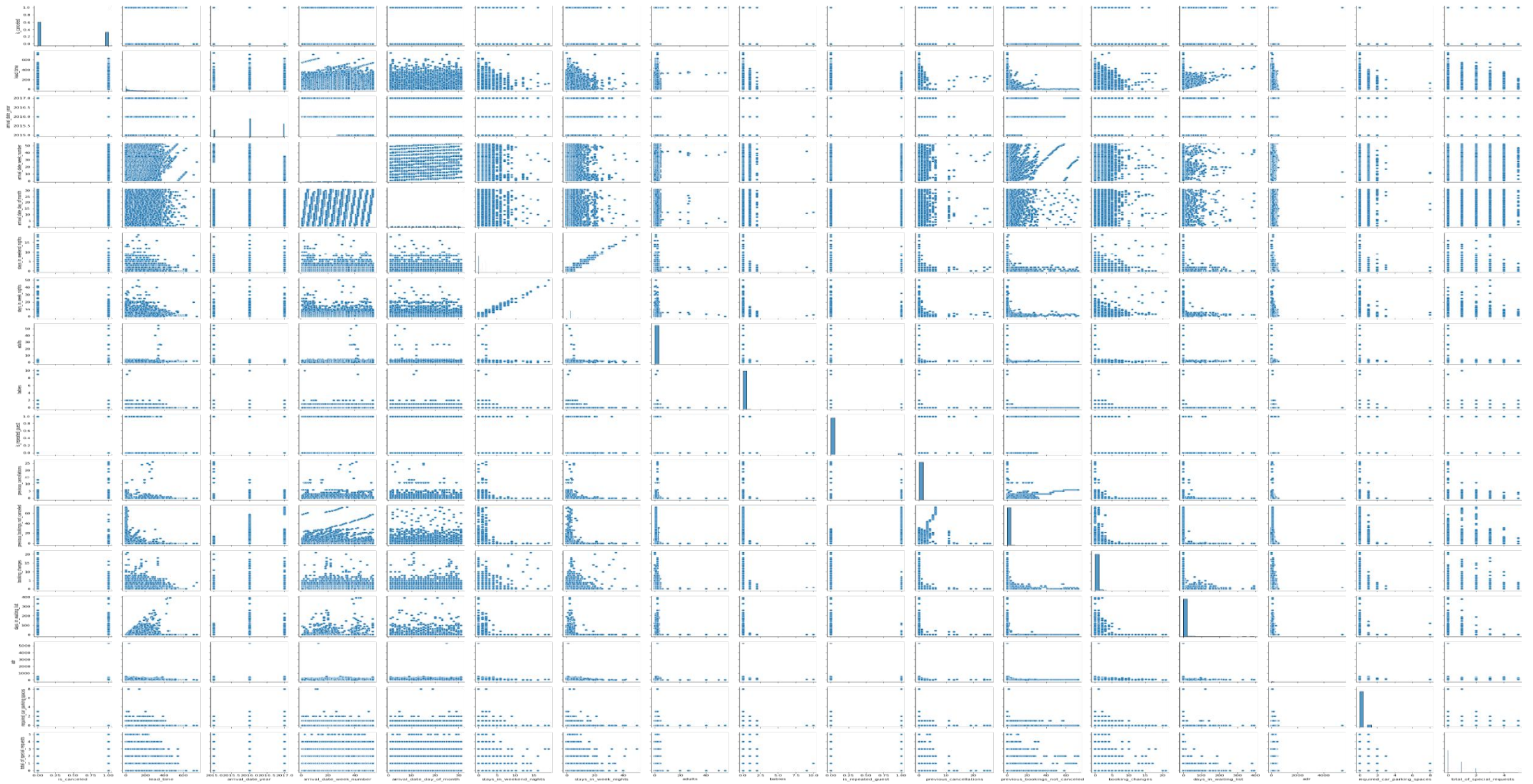
## Correlation analysis

This analysis checks the correlation between columns. Correlation visualization by heatmap -



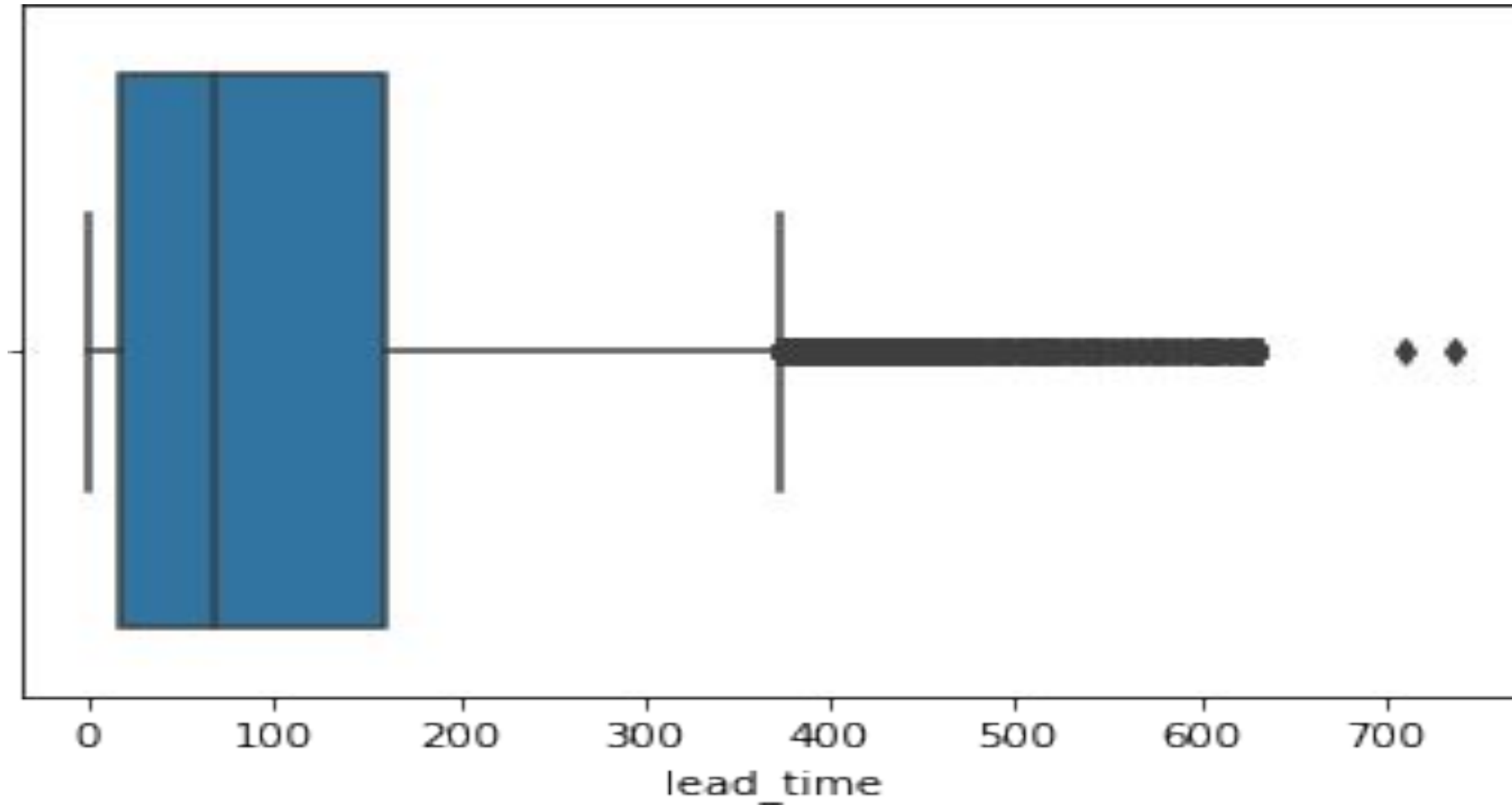


# Relationship analysis between all columns with help of pair plot.





# Checking outlier with help of box plot.



Q&A

# Some important questions

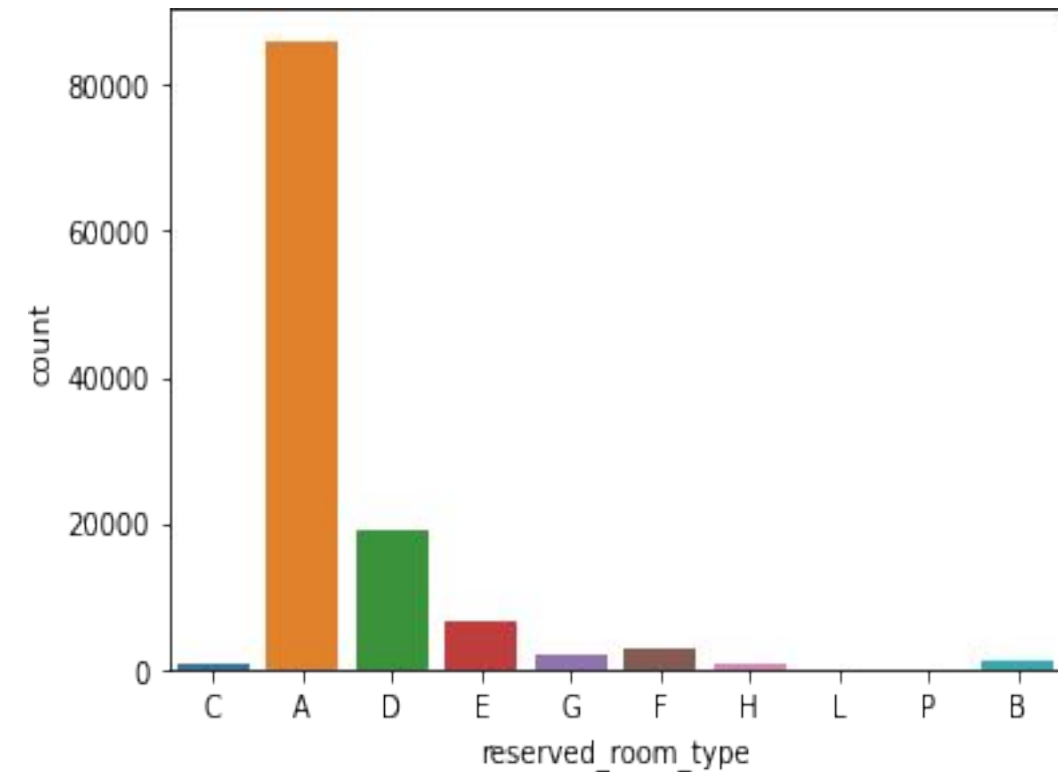
Some other analysis are also done which are as follows

- (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 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.

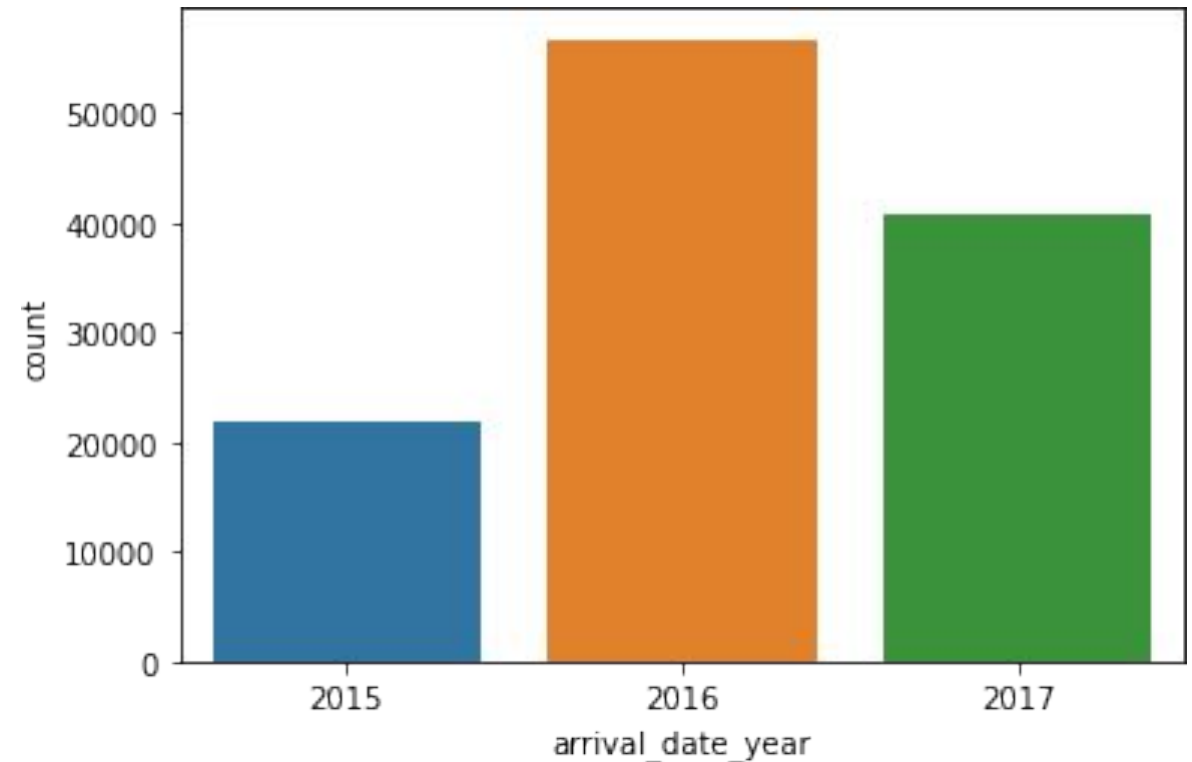


## Which year got a best sales on both hotels?

AI

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 is most preferred 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



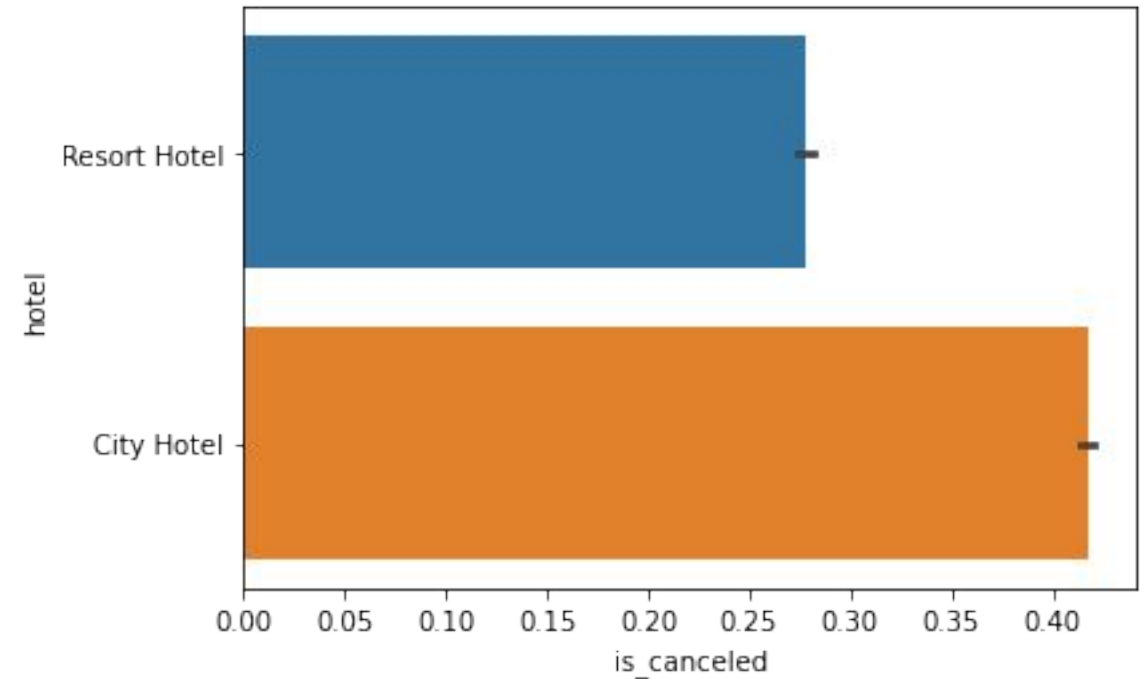
Answer: Guest is most preferred by city hotel because city hotel has maximum bookings.



## Which 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

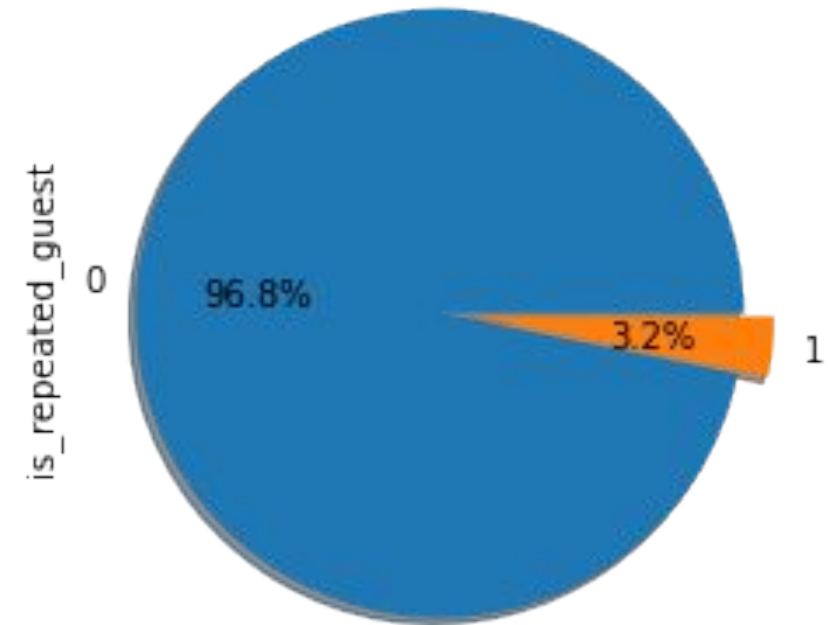


## What is the percentage of repeated guest?

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

Ans: 3.2 percentage of guest are repeated

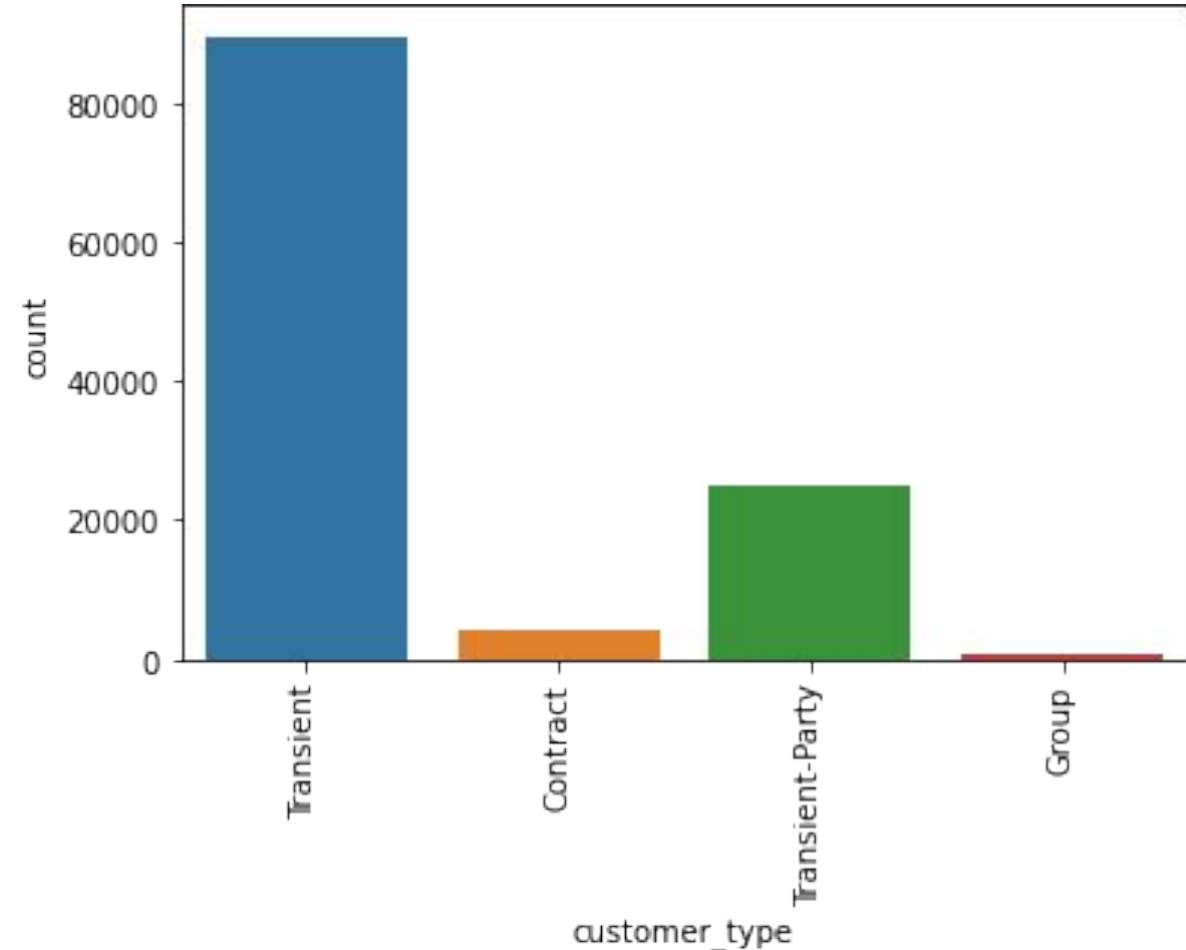
(%) percentage of repeated guest on both hotels



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.

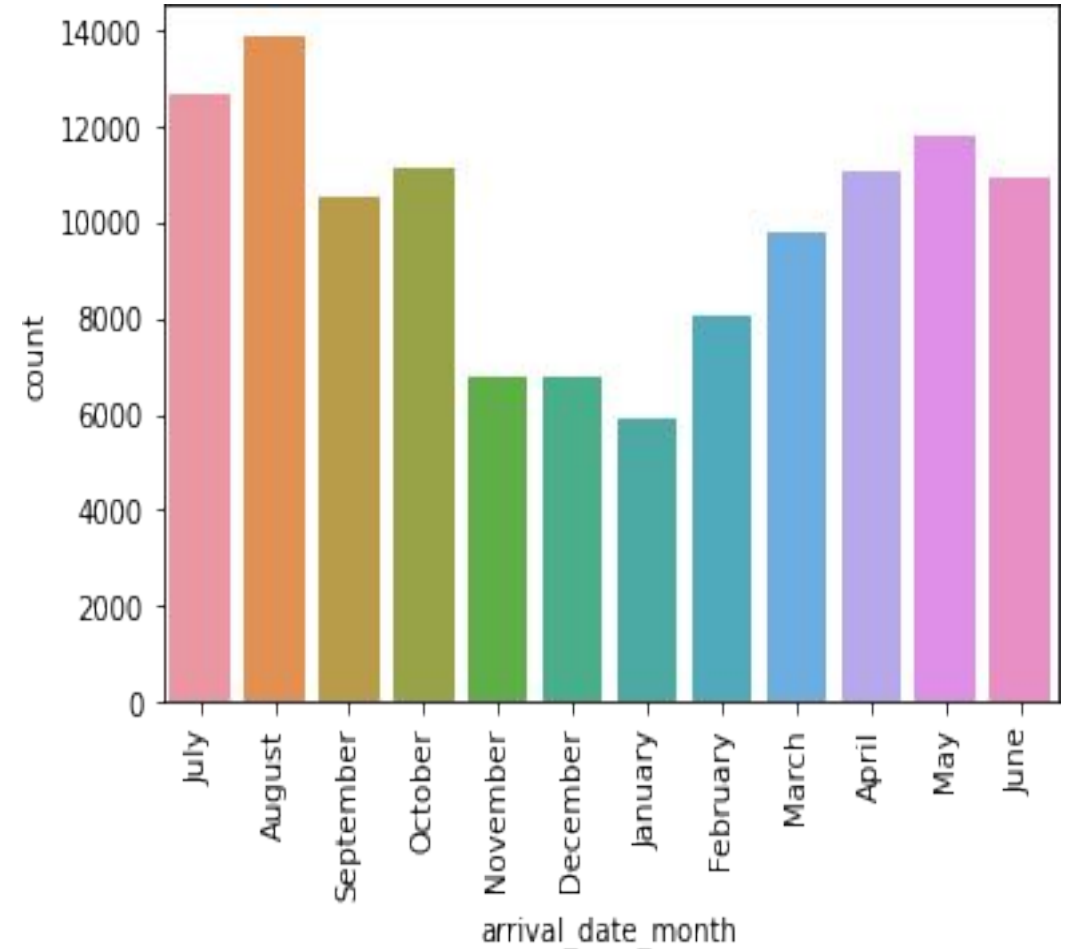


Which month visitors visit highly?

AI

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

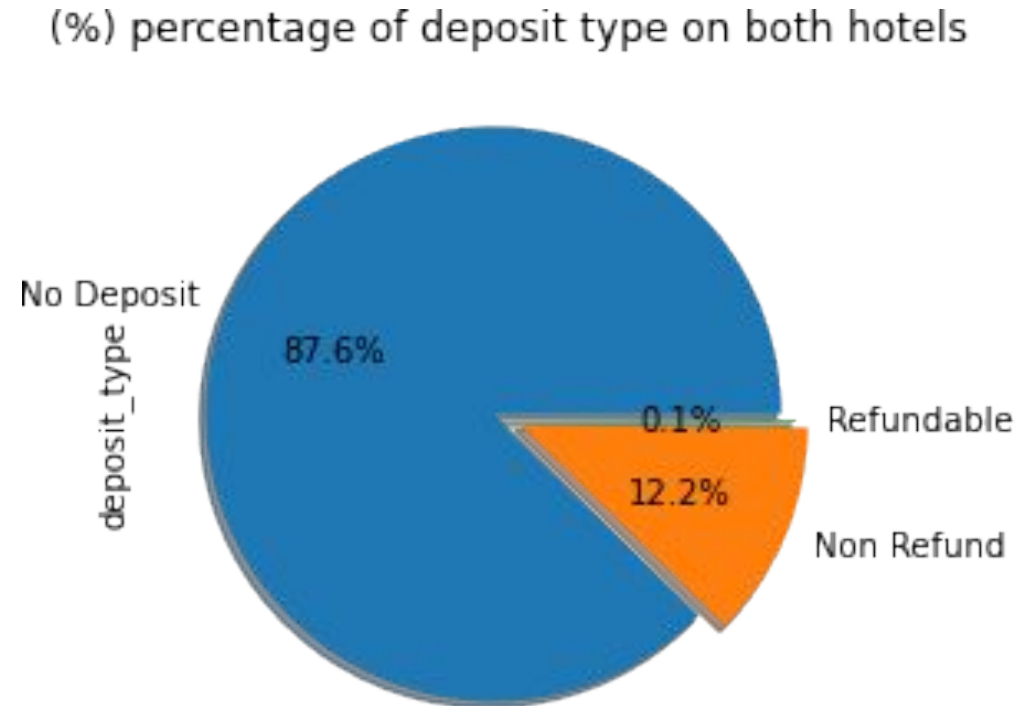
ANS : August month has highest number of visitors



## 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.





## CONCLUSION

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.



**Thank You!**