

CAPSTONE PROJECT-1

Exploratory Data Analysis

Team Members

Jaya Vishwakarma

Priyvrat Sharma

Richa Pandya

Yash Sharma

HOTEL BOOKING ANALYSIS

Abstract: Our data set contains booking information for a city hotel and a resort hotel.

It includes information such as when the booking was made, length of stay, the number of adults, children or babies, cancellation count, revenue and other things.

In upcoming slide, we will analysis and get insight from the data.



POINTS FOR DISCUSSION

- Descriptive summary
 - Hotel wise analysis
- Distribution Channel wise analysis
 - Booking cancellation analysis
 - Timewise analysis
- Relation between deposit type and cancelation
 - Leading Hotel
 - Waiting time
 - Ordered meal at the hotel
 - Hotel Revenue
- ADR – Monthly & Highest ADR

PYTHON LIBRARIES USED

- NumPy
- Pandas
- Seaborn
- Matplotlib
- Plotly lib



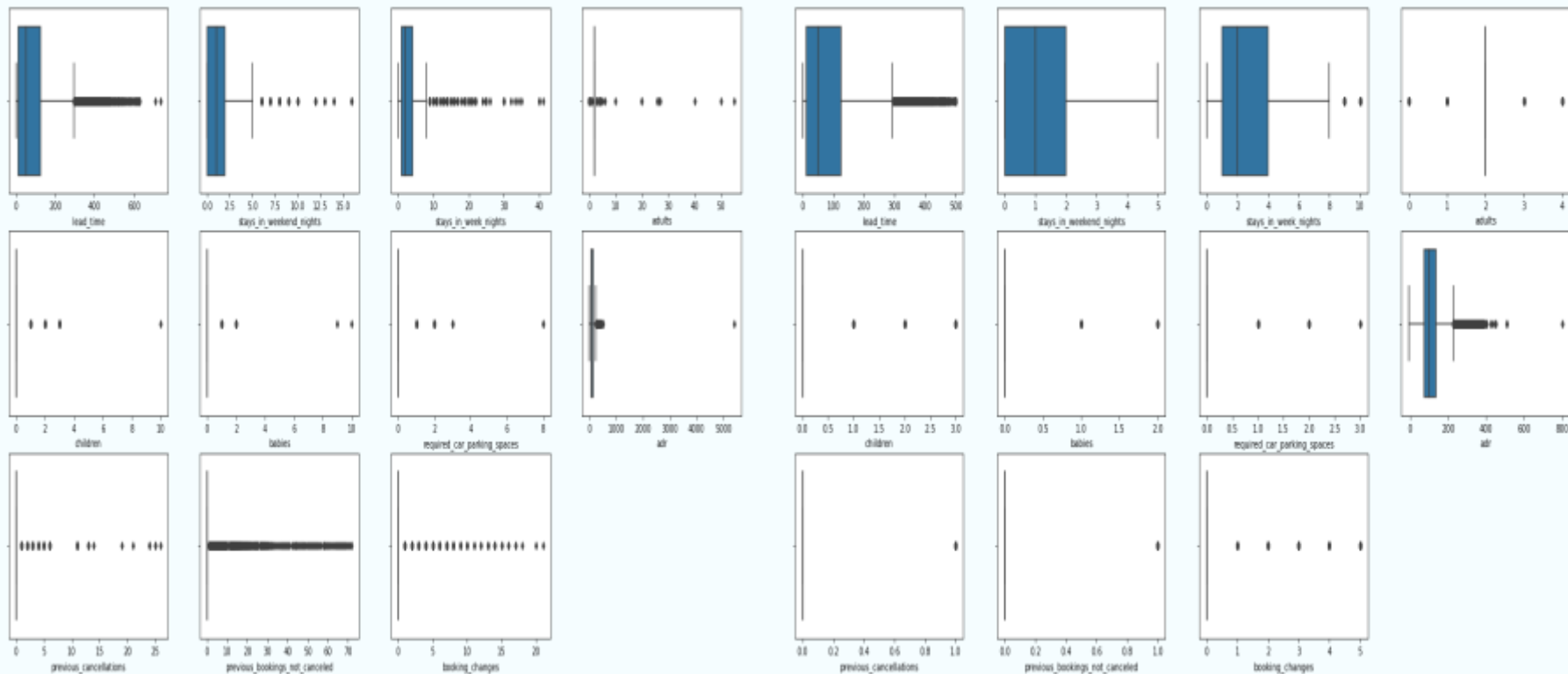
seaborn



PREPARING OUR DATA FOR DEEP ANALYSIS

- **Step 1.** Overview the whole data: number of rows and columns, D-type(Data type), null rows etc.
- **Step 2.** Looking at missing data .
- **Step 3.** Splitting the dataset into numerical and categorical(String).
- **Step 4.** Look at data distribution for numerical columns.
- **Step 5:** Look at Data Characteristics of text variables.
- **Step 6.** Looking at the duplicate & null values and dropping them.
- **Step 7:** Conclusion.

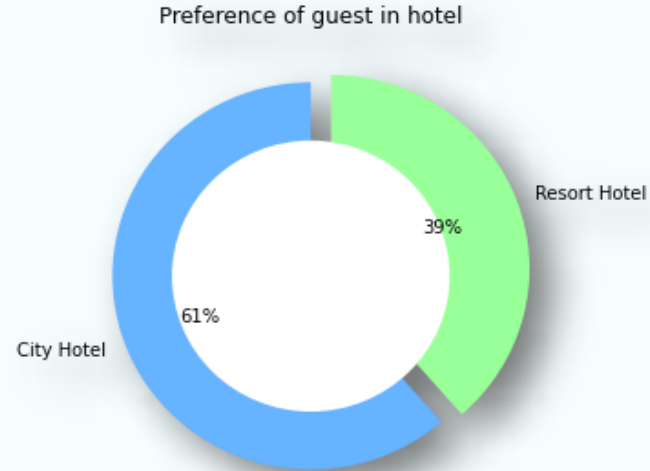
OUTLIERS



MOST PREFERRED HOTEL

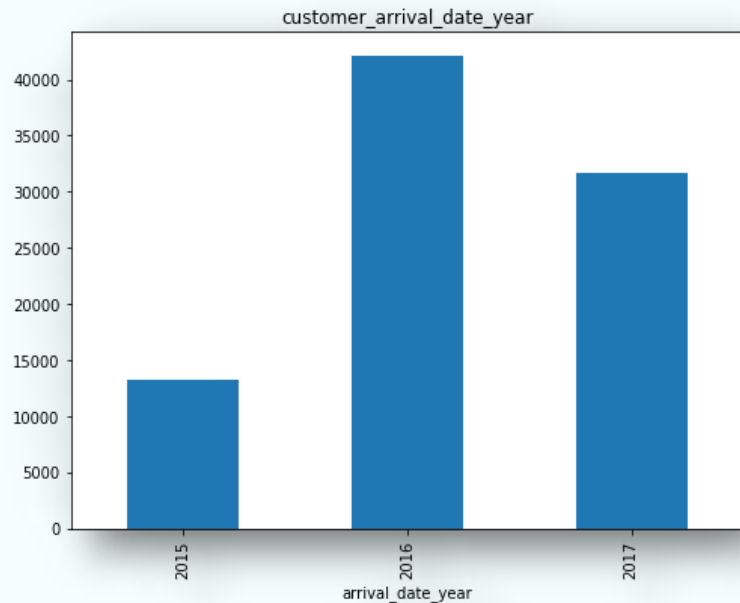
City hotels are more preferred by customer as compare to resort hotel.

61% customers booked for city hotel **and 39%** booked for resort.

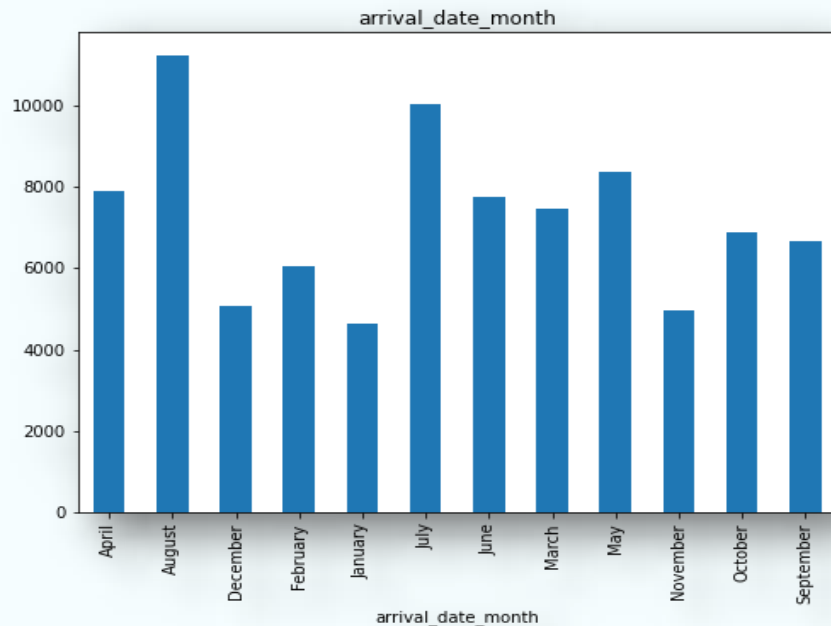


CUSTOMER ARRIVAL

Yearly

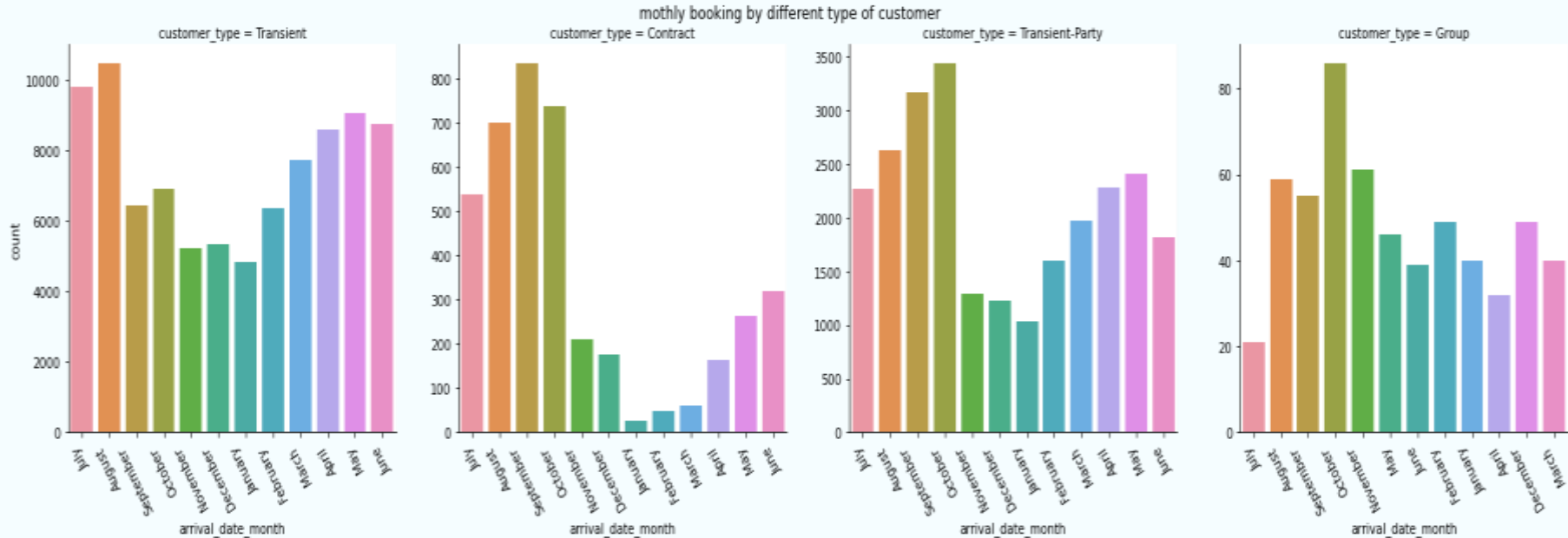


Monthly



Maximum customer arrive in to the hotel in **year 2016**.
In **august** month, **maximum** number **people** checked in.

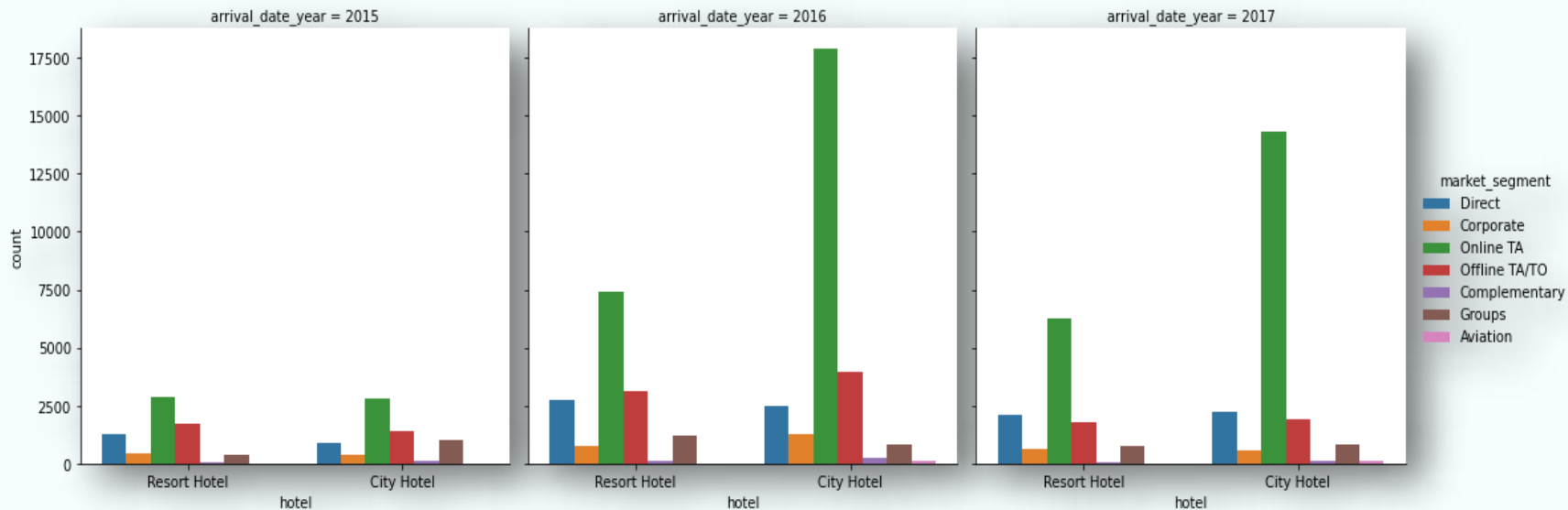
ANALYSIS BASED ON MARKET SEGMENT



Here, we can see the most active month as indicated by the appearance of each type of customer.

Highest bookings - Transient customer booked for **august** month, **contract** booked for **September**, **transient party** and **group** both booked for **October** month.

ANALYSIS BASED ON MARKET SEGMENT



Aviation and complementary bookings are negligible compared to the **online and offline TA**

Corporate, direct and group bookings together add up to **online TA**.

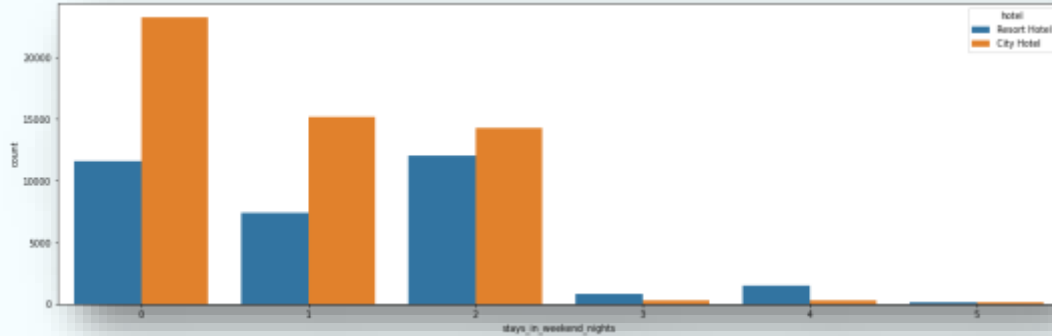
COUNTRY RECORD



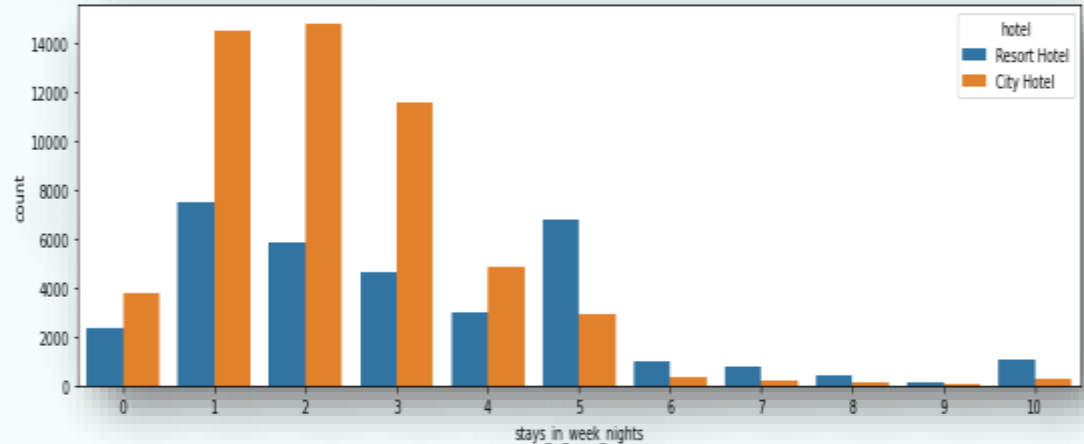
Most of the customers from **European countries** like **Portugal, Great Britain, France** and **Spain**. **Around 40%** of all bookings are from Portugal followed by Great Britain(**10%**) & France(**8%**).

CUSTOMER STAYS

Weekend Nights

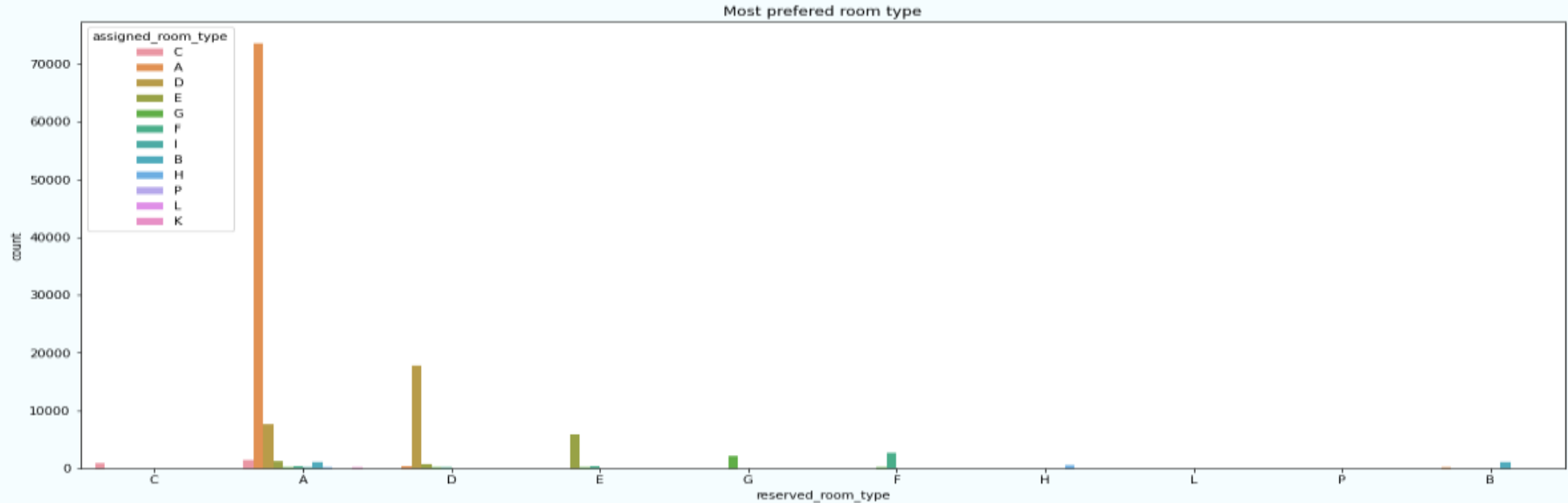


Weekdays Nights



We observe that weekday night stays were more in City Hotel as compare to weekend night.

ROOMS ASSIGNMENT

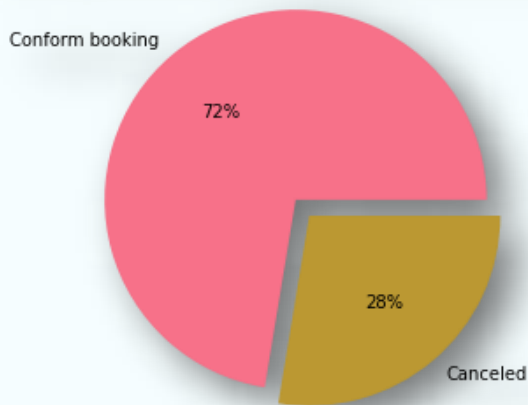


A- type room is **most preferred** by the customer and they are getting same room assigned as per their reservation, followed by D-type and E-type. While **B,L** and **P** are the **least** preferred room type.

CANCELLATION

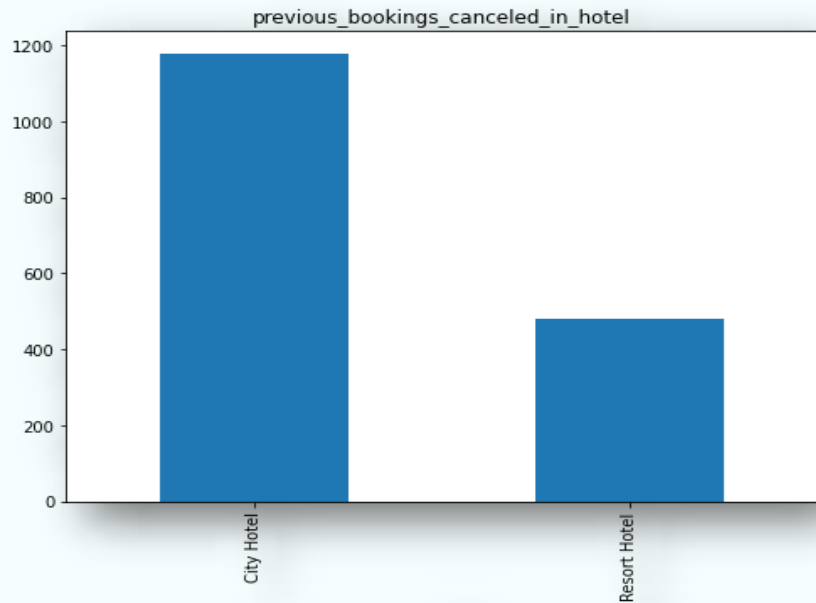
Ratio between cancelled and confirmed bookings?

Percentage ratio between cancelled and confirmed bookings



Confirmed bookings are more
than the cancelled bookings.

Highest cancellation between hotels?

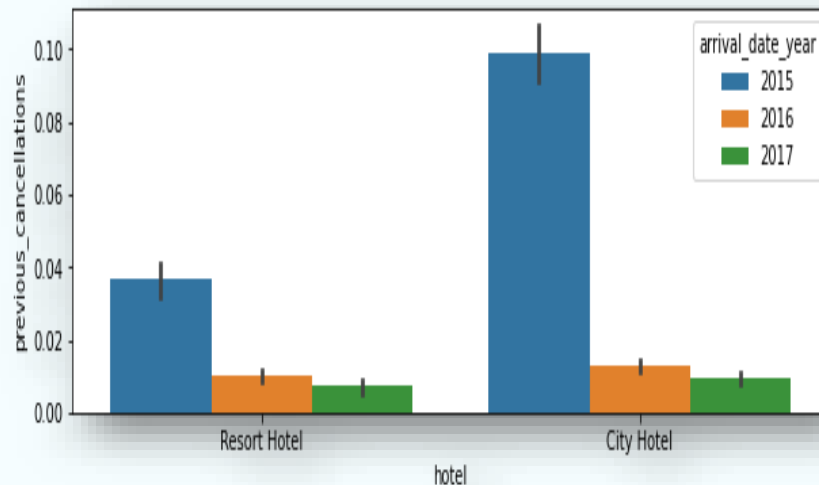


City hotel has higher cancellation
than the resort hotel.

Cancellation with arrival date-year of hotel respectively?

In year **2015**, the count of **cancellation** is **higher** than 2016 and 2017.

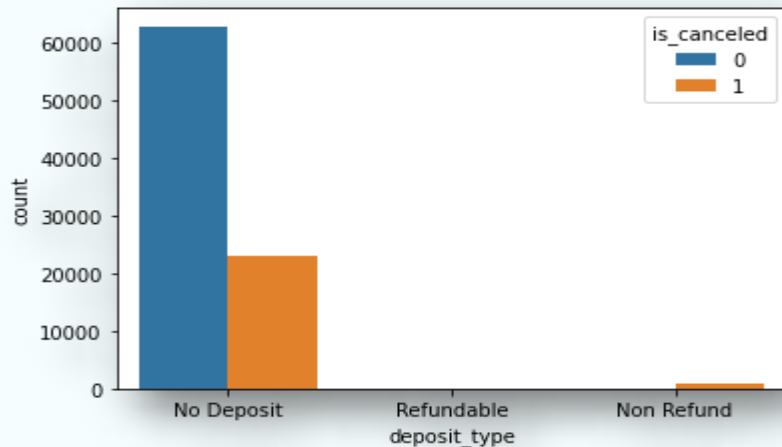
That too from **city** hotel.



Relation between deposit type and cancellation?

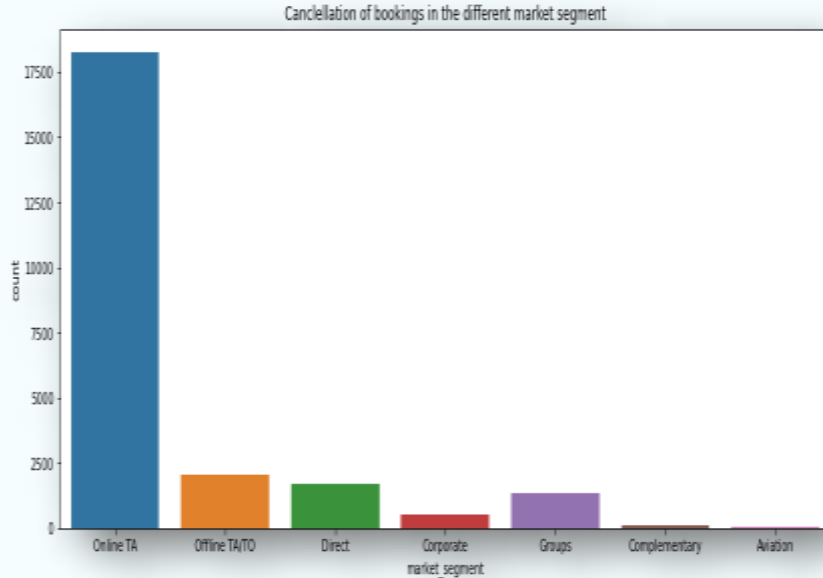
Refundable is **least preferred** deposit type amongst all.

Cancellation is **higher** in hotel which offers no **deposited**/no advance payment.

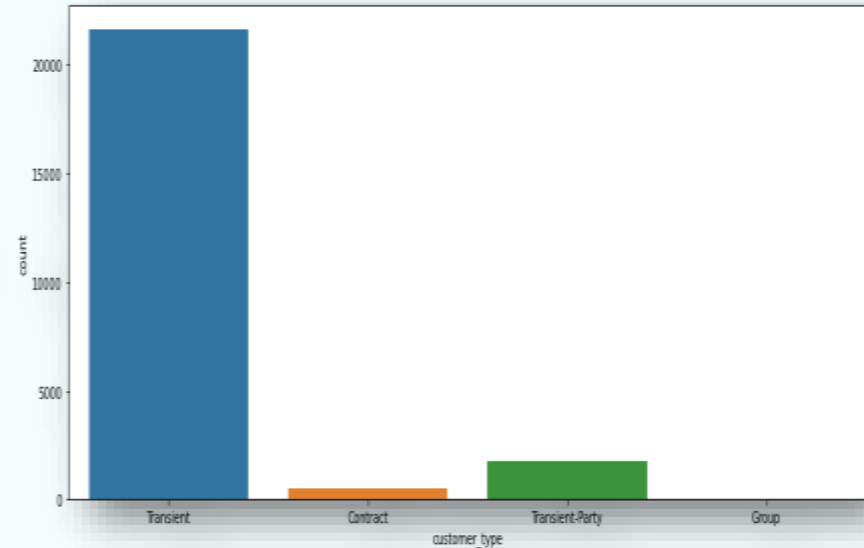


CANCELLATION ANALYSIS BASED ON:

Market Segment



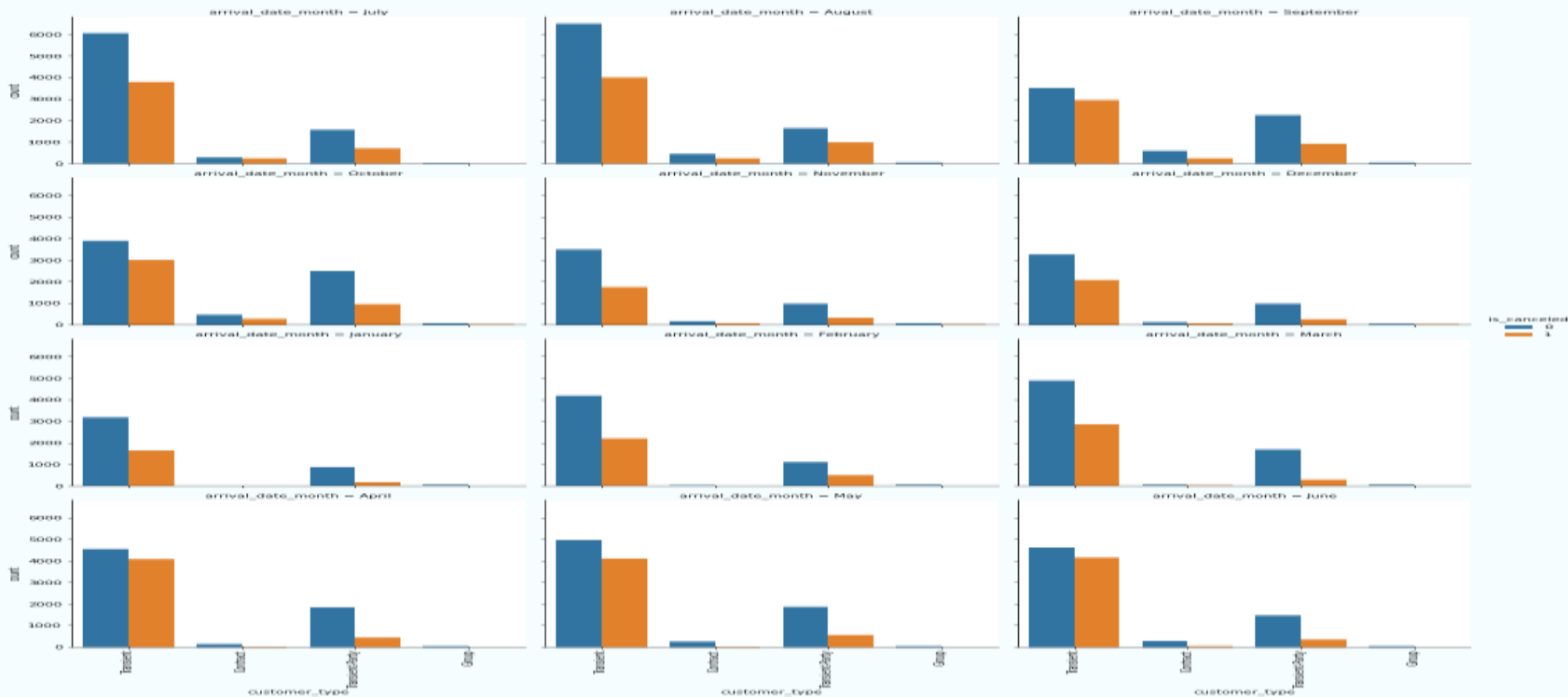
Customer Type



Transient customer from **online booking** platform **cancelled** their booking more **frequently**. **Aviation** and **complementary** bookings are **negligible** compared to the online and offline TA

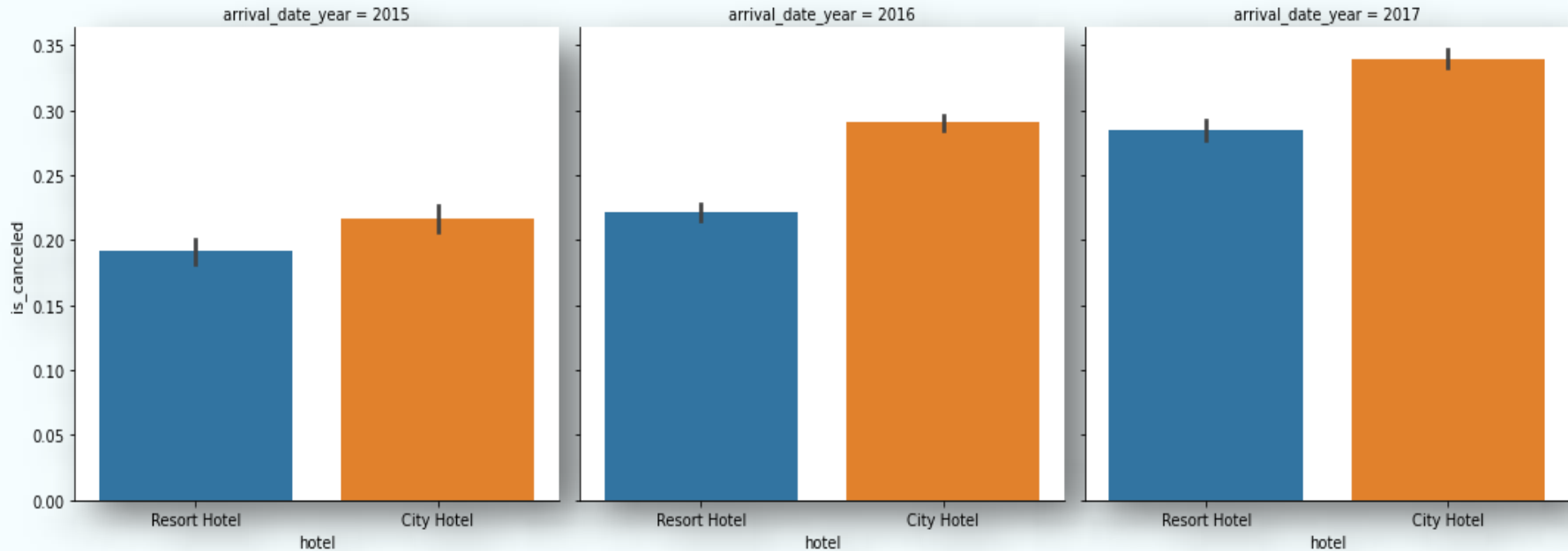
CANCELLATION: MONTHLY ANALYSIS

Monthly cancellation with customer type



Cancellation rate is **higher** by the **transient** customer in the month of **august**.

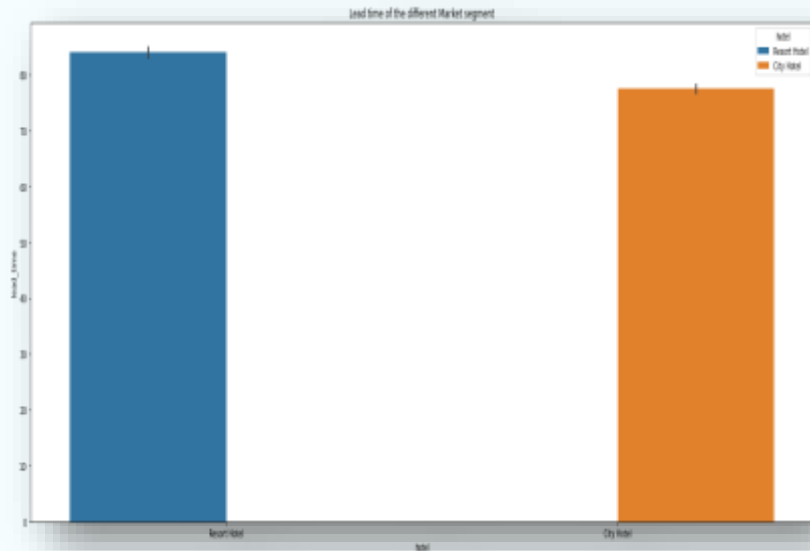
REPEAT CUSTOMER



Maximum repeat customers arrived in **city** hotel.

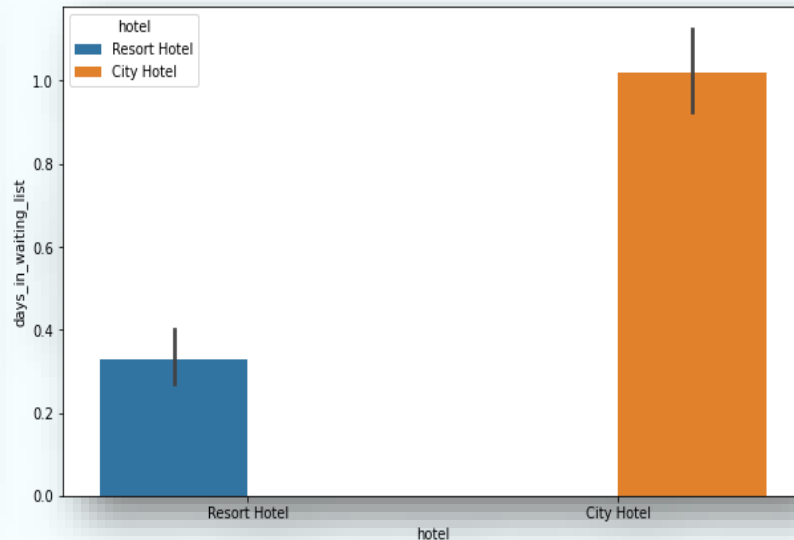
In year **2017**, **maximum** repeat customer noticed.

LEADING TIME



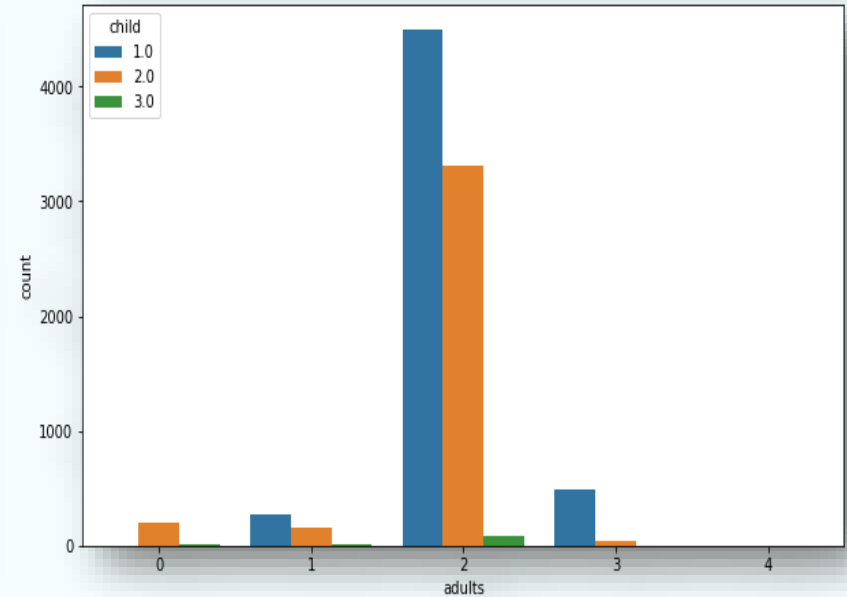
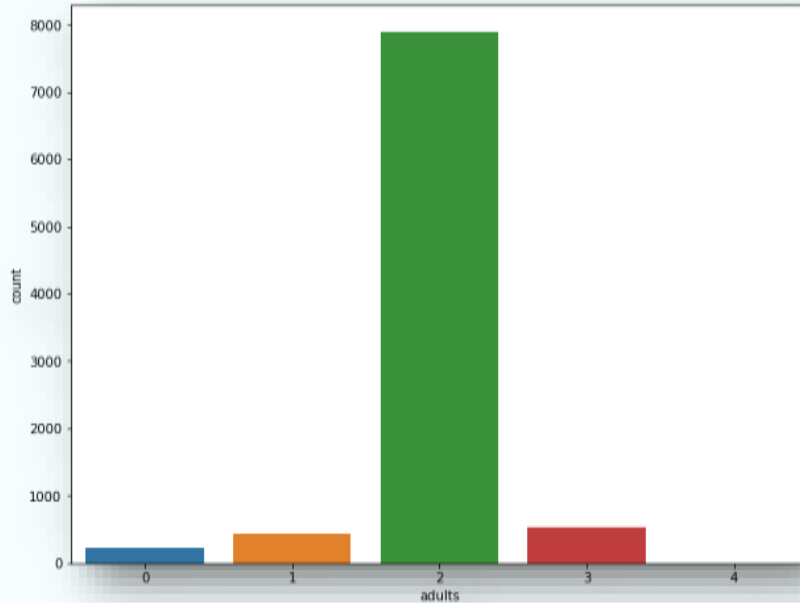
Resort hotel has **higher lead** time over the city hotel.

WAITING TIME



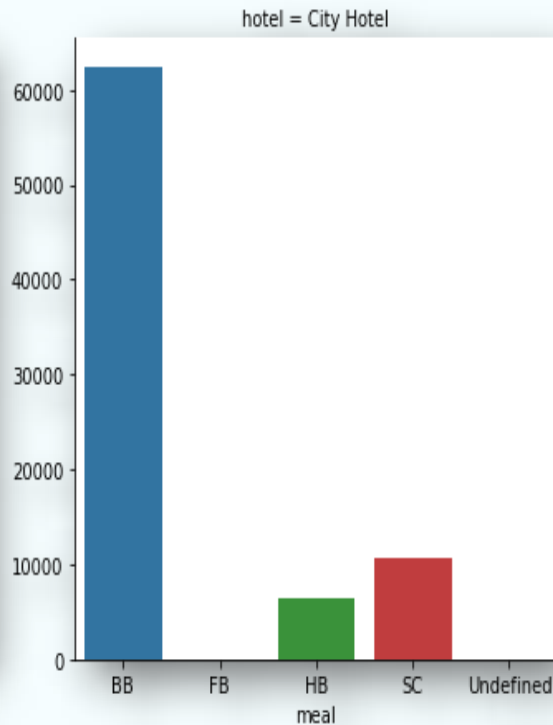
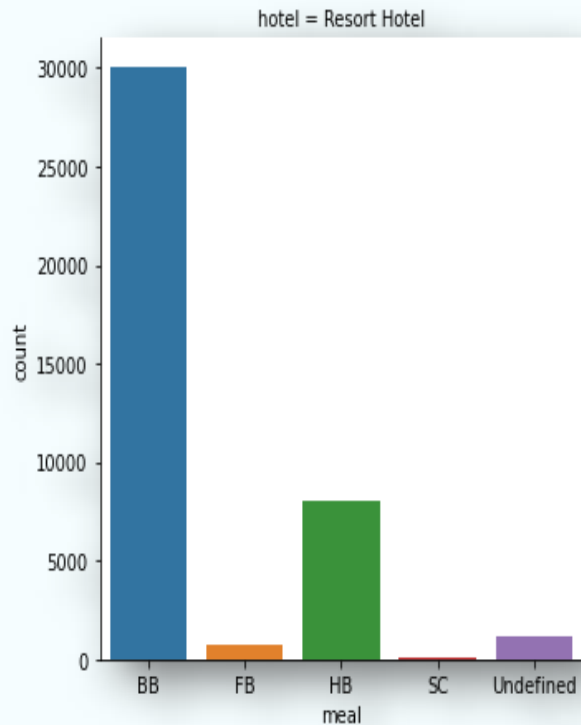
City hotel has **longer waiting** time than resort hotel.

GUEST ARRIVAL: ADULTS & KIDS



Most visitors arrived in **pair** with no children/babies and **preferred city** hotel over resort hotel.
Visitors who have 1 or 2 children also preferred city hotel

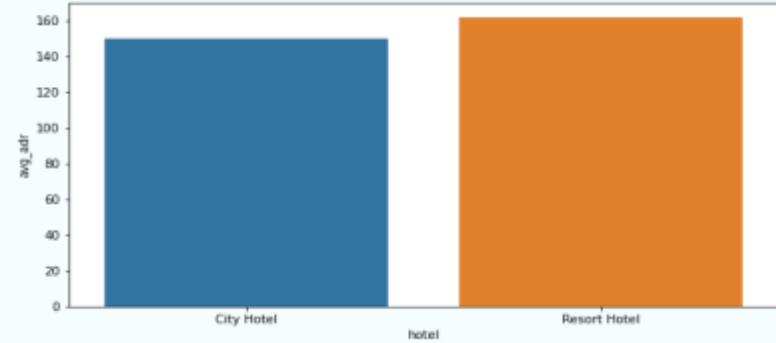
MOST ORDERED MEAL AT THE HOTEL



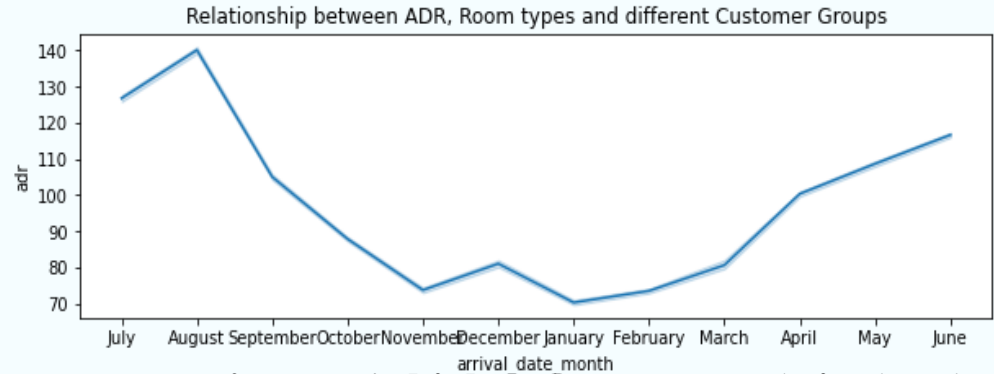
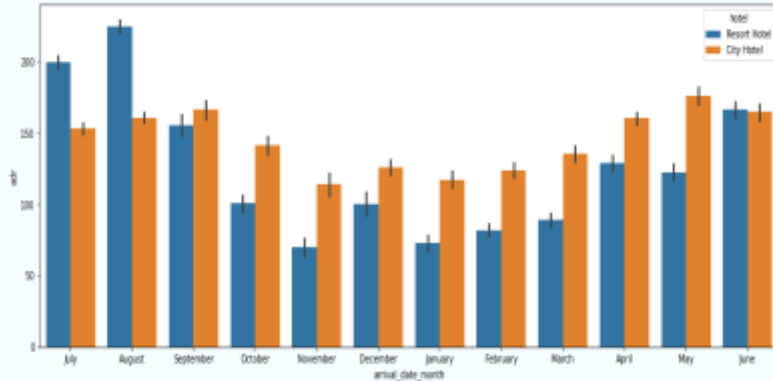
Out of the meals, **BB (Bed & Breakfast)** is the **most ordered** meal which is around 77.2%, followed by HB(Half Board), SC(no meal package), Undefined and FB (Full Board).

REVENUE

Resort hotel has generated **more revenue** as compare to city hotel.



HIGHEST ADR



Highest ADR is observed in **august** month. However **resort's** ADR is **highest** in **august** and **city** hotel ADR is **highest** in **may** month

CONCLUSION

- There are more city hotel than resort hotel. Booking rate is also higher in city hotel.
- August is the most demanded month and January is the least. Hotel management should come up with interesting offers/deals in the months when they have least bookings.
- Repeat customers are more in city hotel and the number is increasing every month. Resort management needs to improve the hotel facilities to make more repeated guest. They should take feedback seriously for improvement.
- Resort hotel is likely to make more revenue. Attractive scheme can play a vital role.
- We have seen that most of the guests were from western Europe countries. Portugal is the most active country.
- Since, cancellation is higher in online-ta. Hotel management should work on their terms & policies.

THANK YOU!