Capstone Project 1

Hotel Bookings Analysis

Individual project by Sharath S

Steps followed:
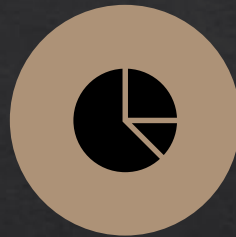
1.	2.	3.	4.	5.
DEFINING THE PROBLEM STATEMENT	UNDERSTANDING THE DATA	FINDING INSIGHTS USING THE DATA	VISUALIZING THE DATA USING DIFFERENT PLOTS	CONCLUSION

Problem Statement

- Using hotel data to understand what parameters to focus on to improve customer satisfaction and to help the hotel increase profits, while also using insights obtained from data to help find out where they are lacking and how to improve on their weaknesses.

Understanding the data

◈ Hotel: Type of hotel.

◈ is_cancelled: Whether the booking was cancelled or not.

◈ arrival_date_month: Month of arrival.

◈ stays_in_weekend_nights: Number of weekend nights stayed.

◈ stays_in_week_nights: Number of weeknights stayed.

◈ adults: Number of adults.

◈ children: Number of children.

◈ babies: Number of babies.

- meal: Type of meal plan chosen.
- country: country code
- distribution_channel: Distribution channel used by customers.
- is_repeated_guest: Is it their first time?
- reserved_room_type: Type of room reserved.
- deposit_type: Type of deposit paid.
- customer_type: Type of customer
- adr: Average daily rate.
- total_of_special_requests: Number of special requests made

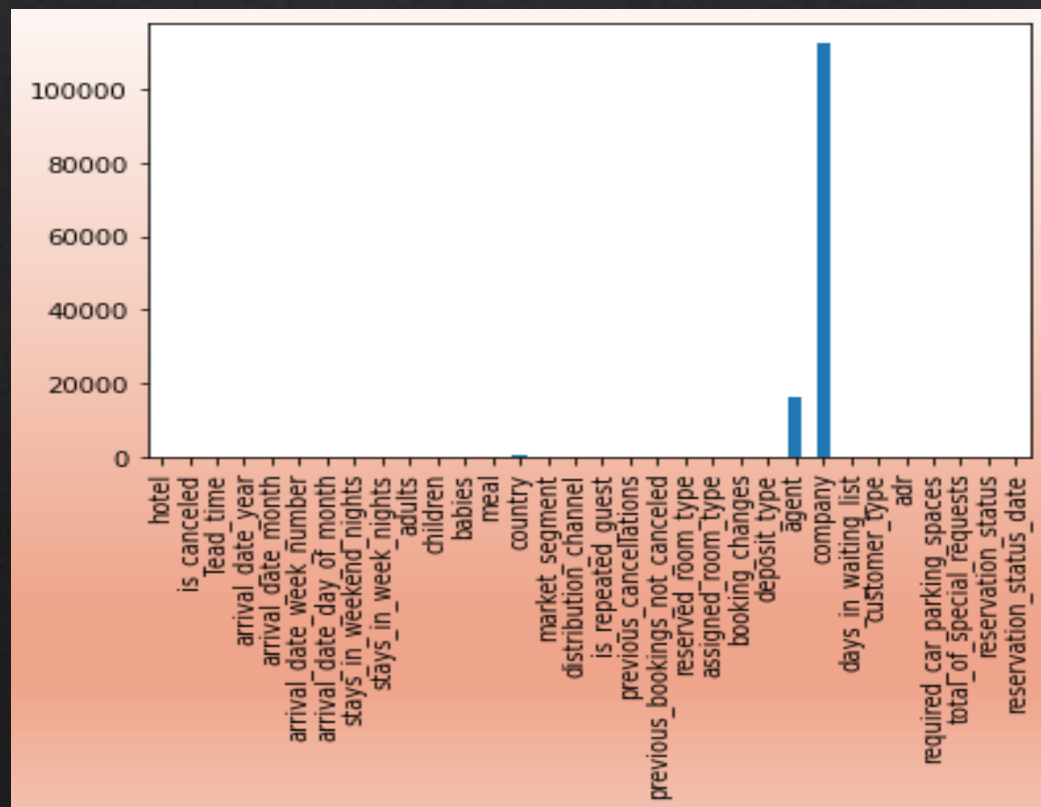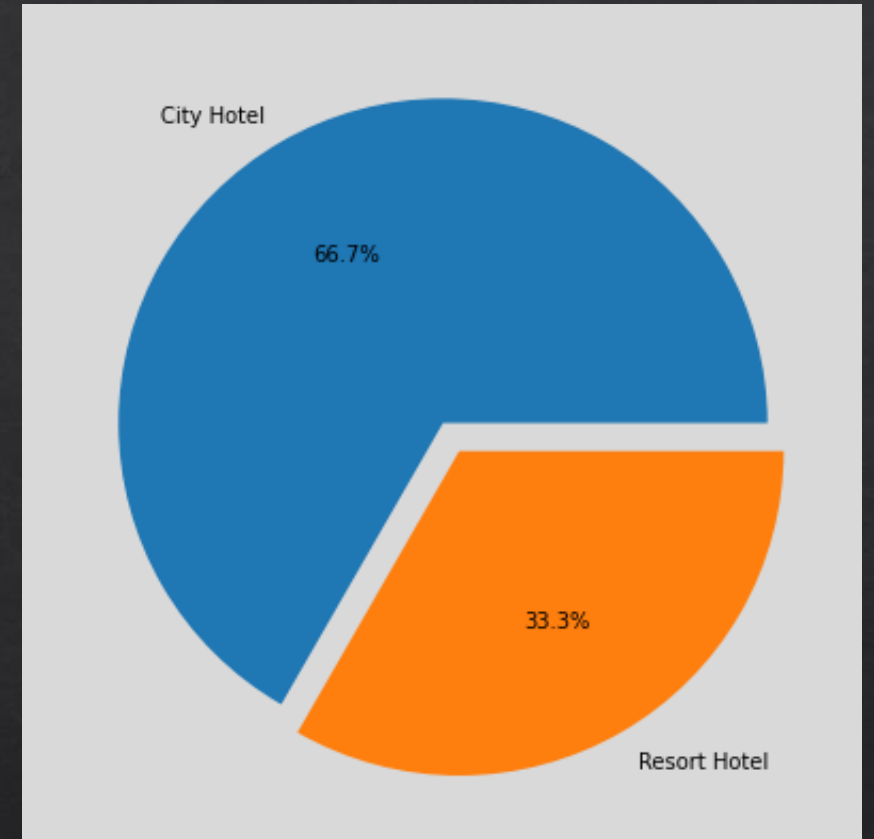# Visualizations to find insights
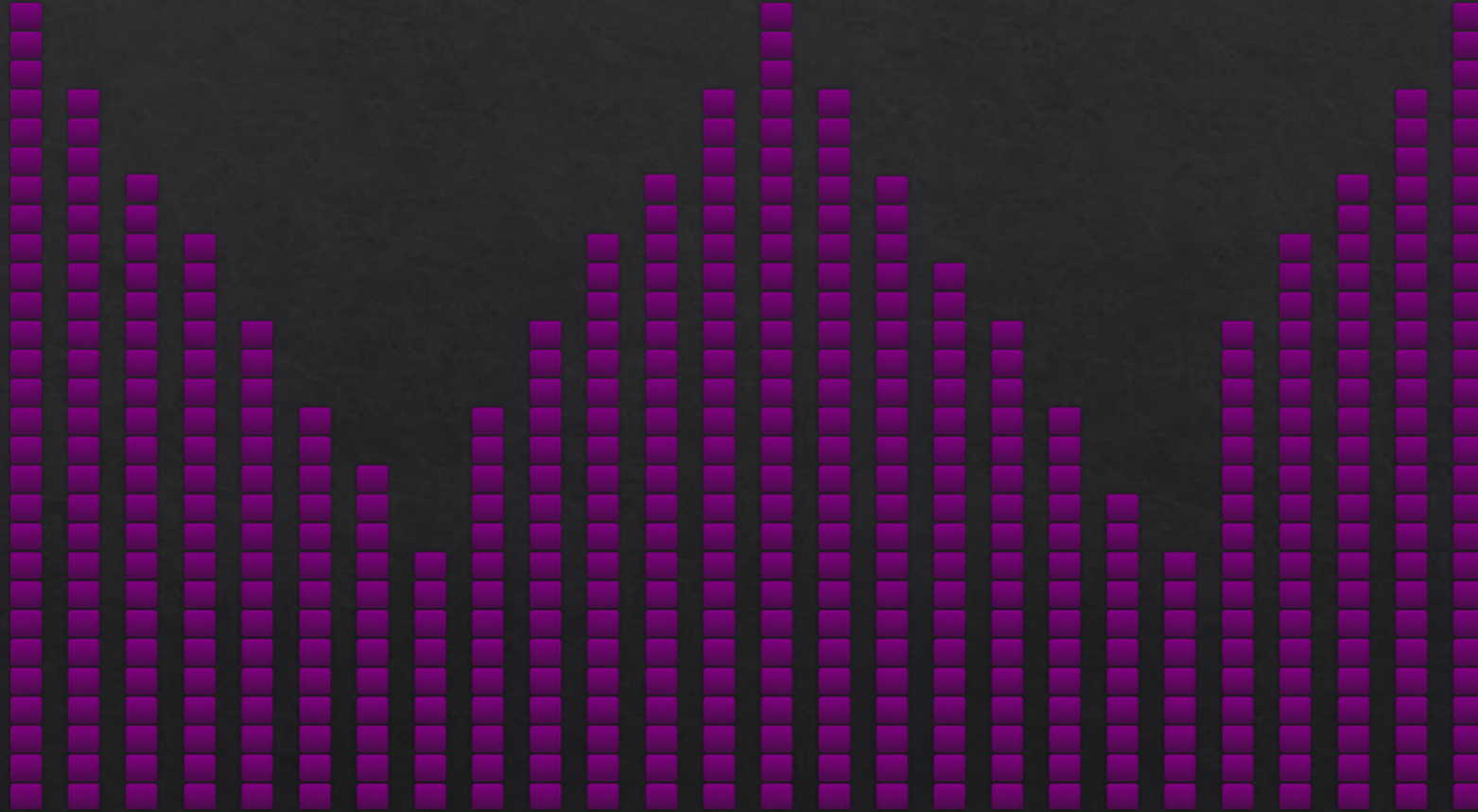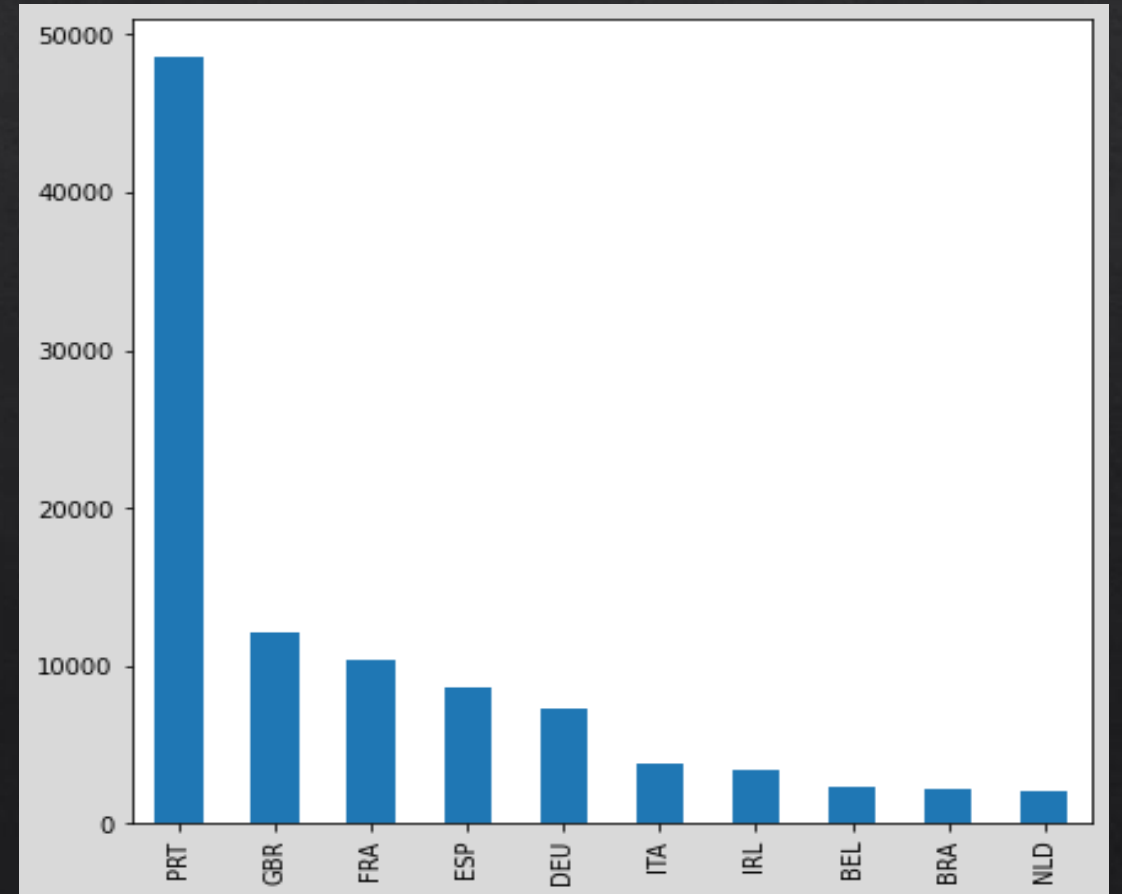
Visualizing the null values in the dataset:

Pie chart of type of hotels.
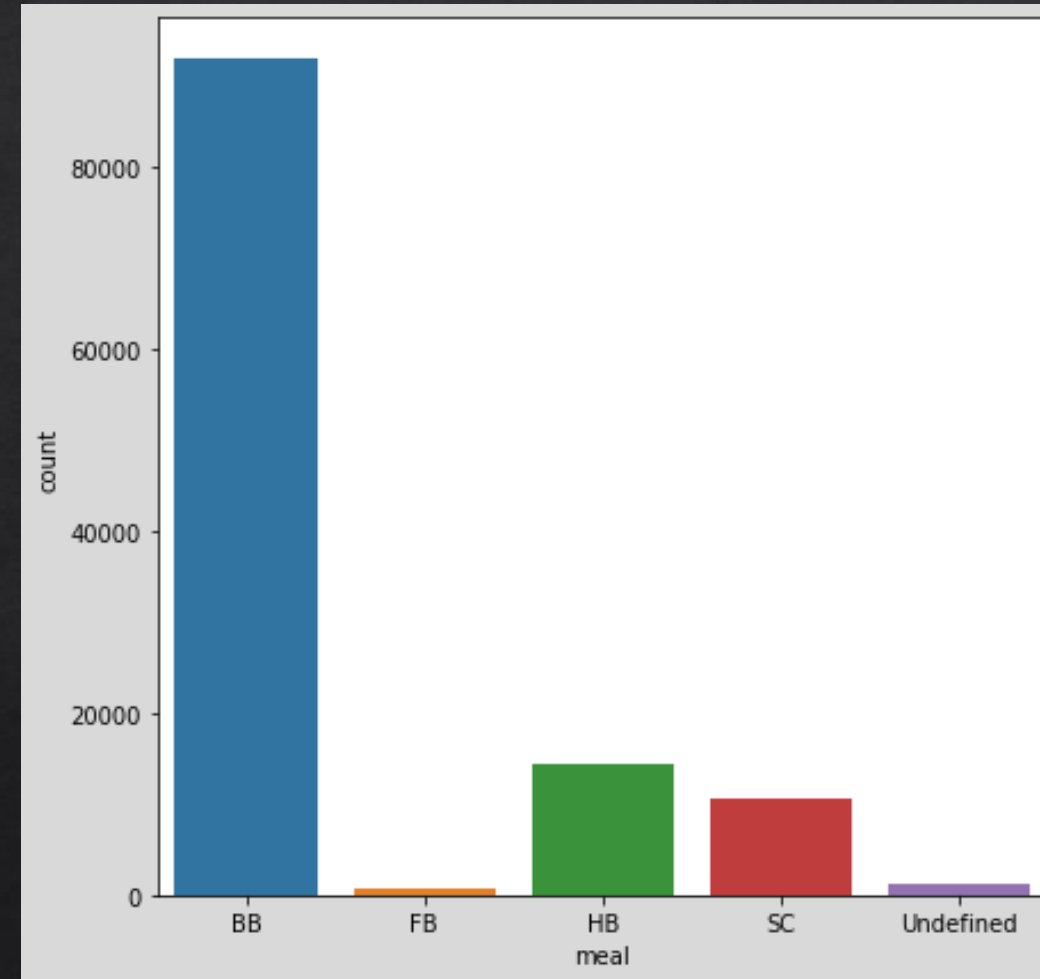
Bar plot of top 10 countries of booking.

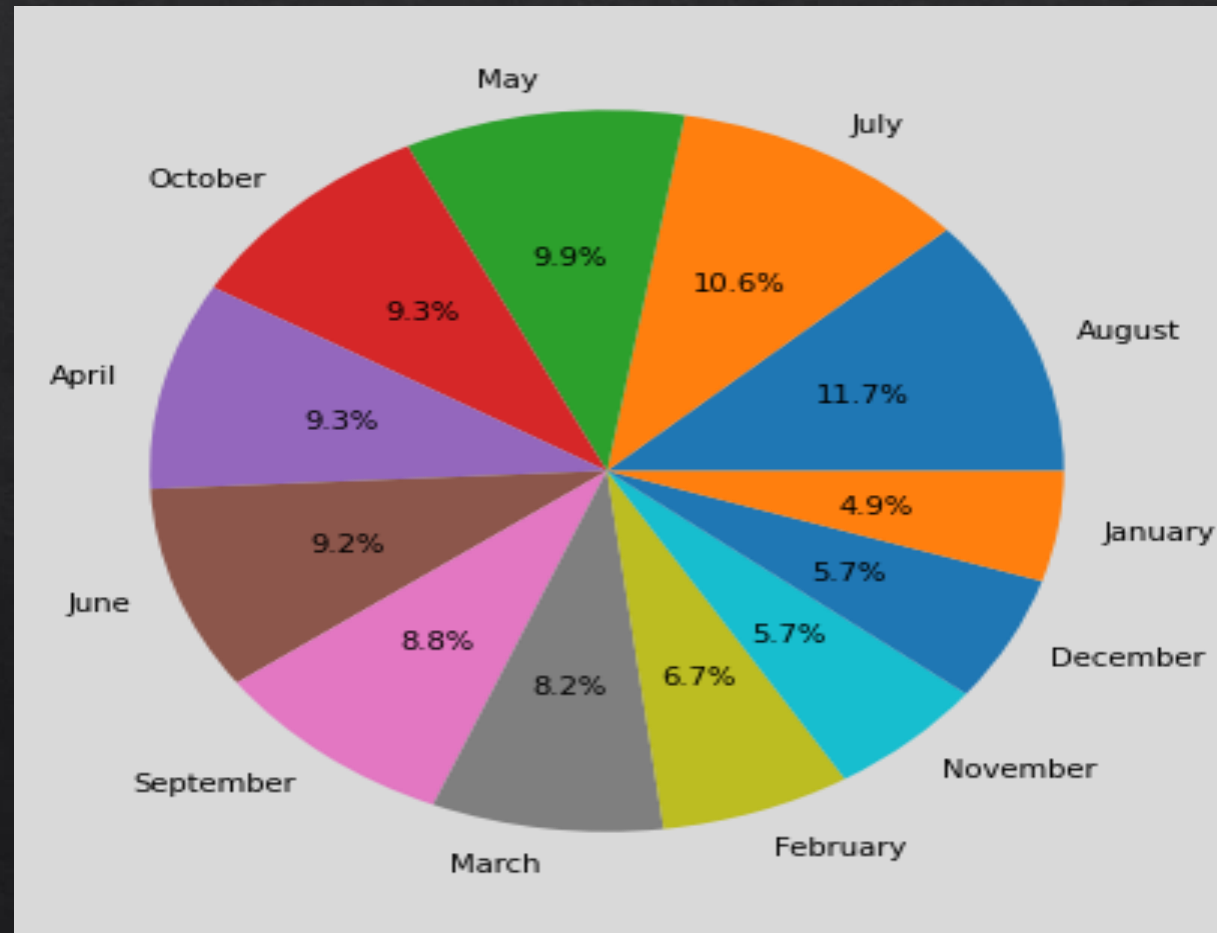# Countplot of Meal plan preference of customers.

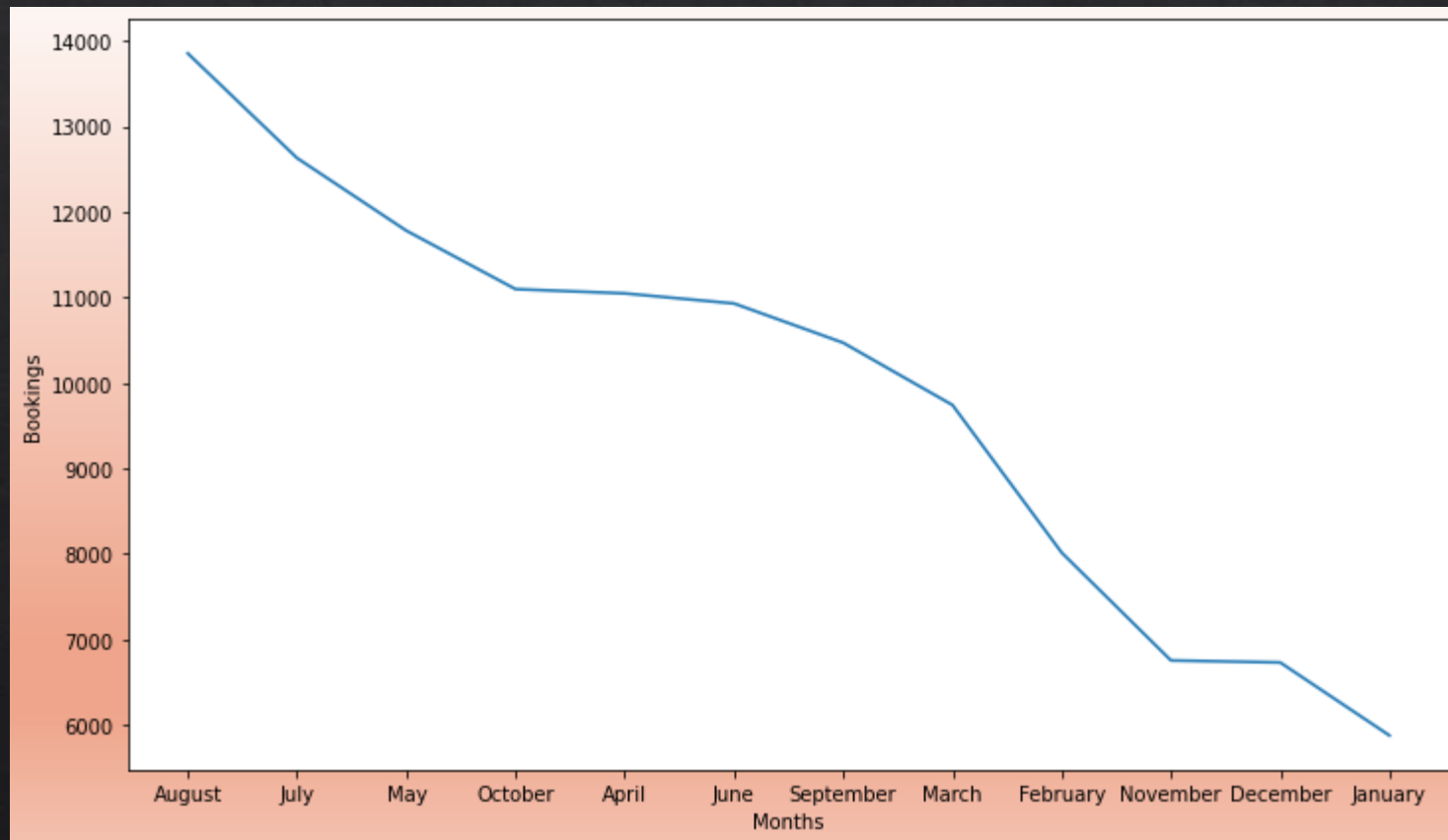**BB- Bed and Breakfast**

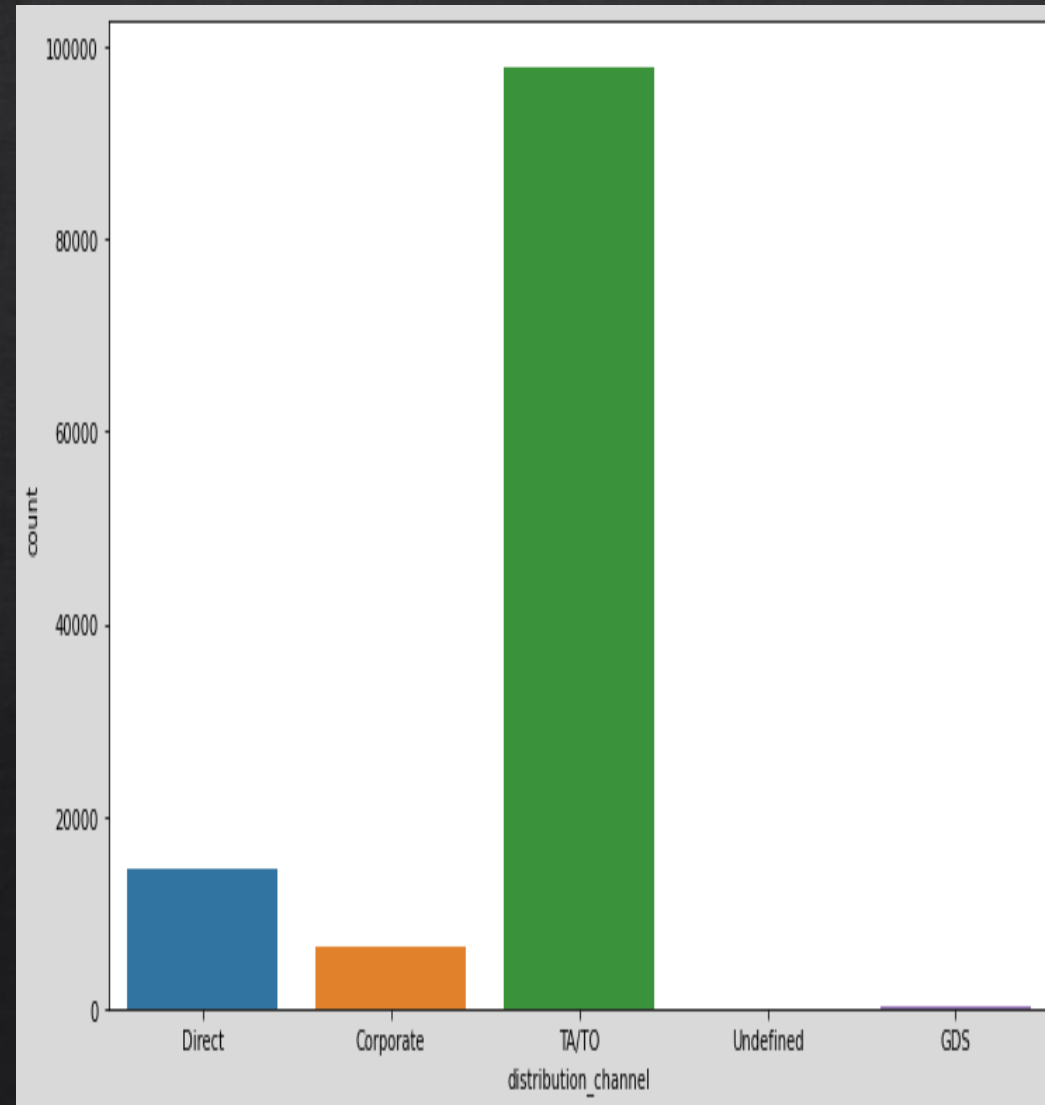**FB- Full Board**

HB- Half Board

SC- Self Catering

Pie chart of busiest months

Line graph of amount of bookings per month

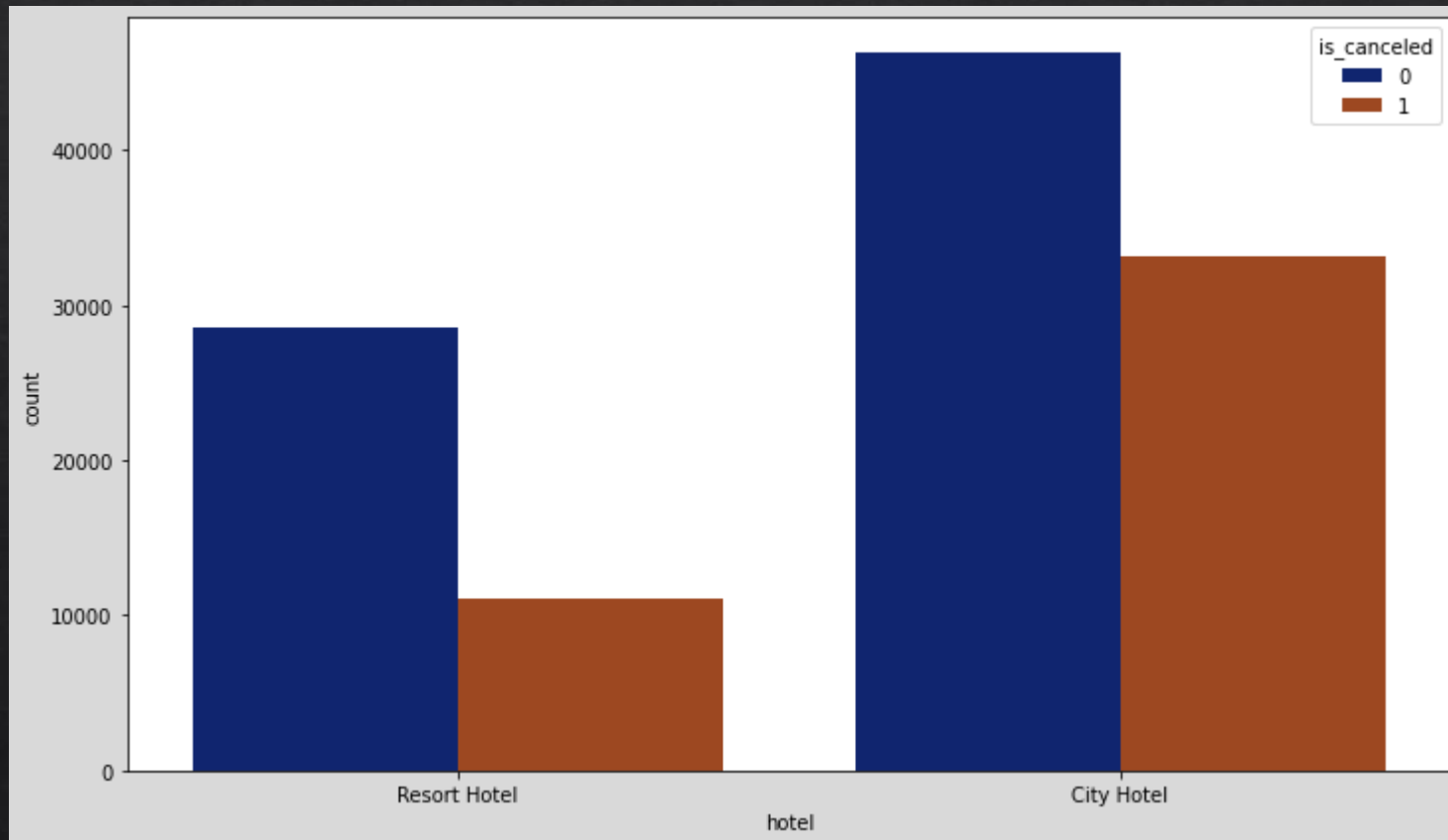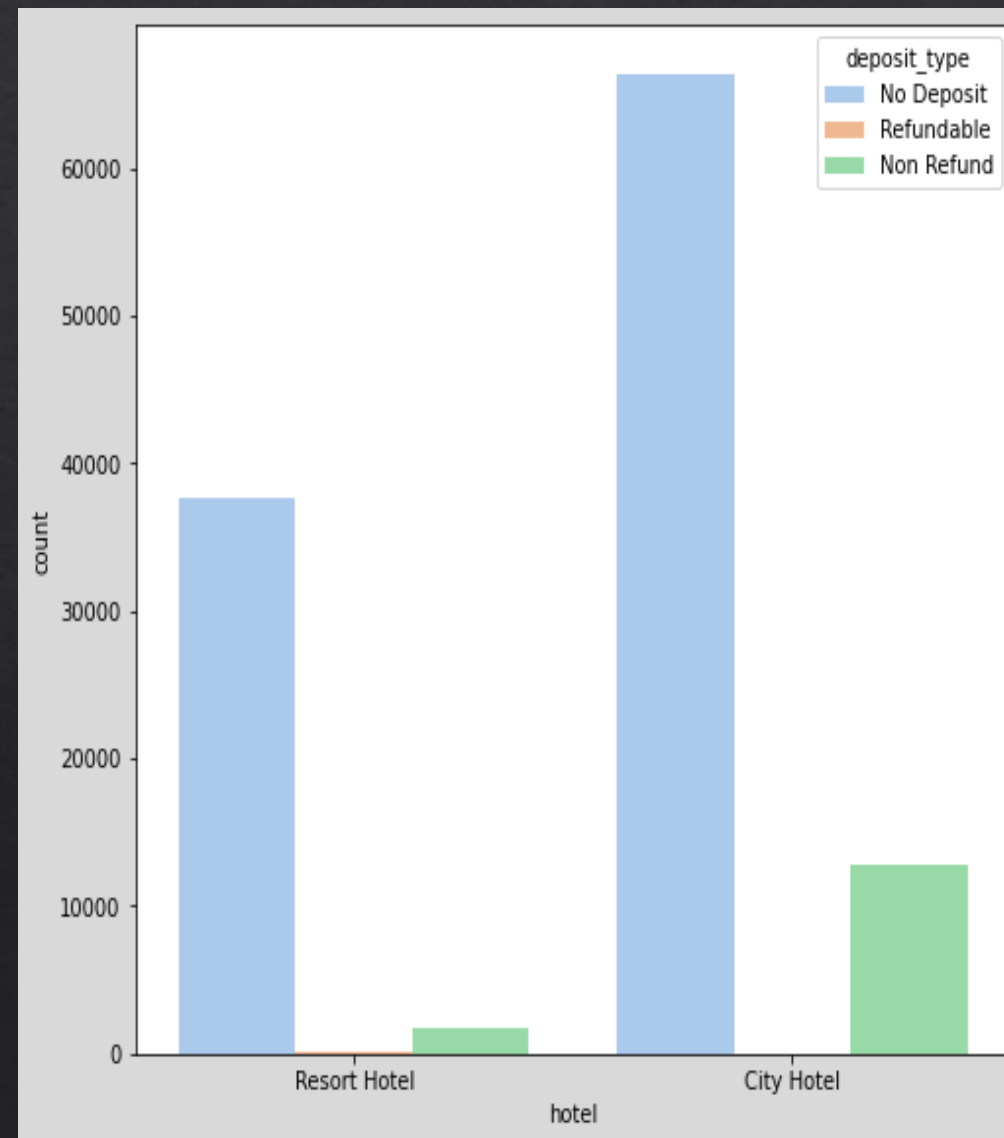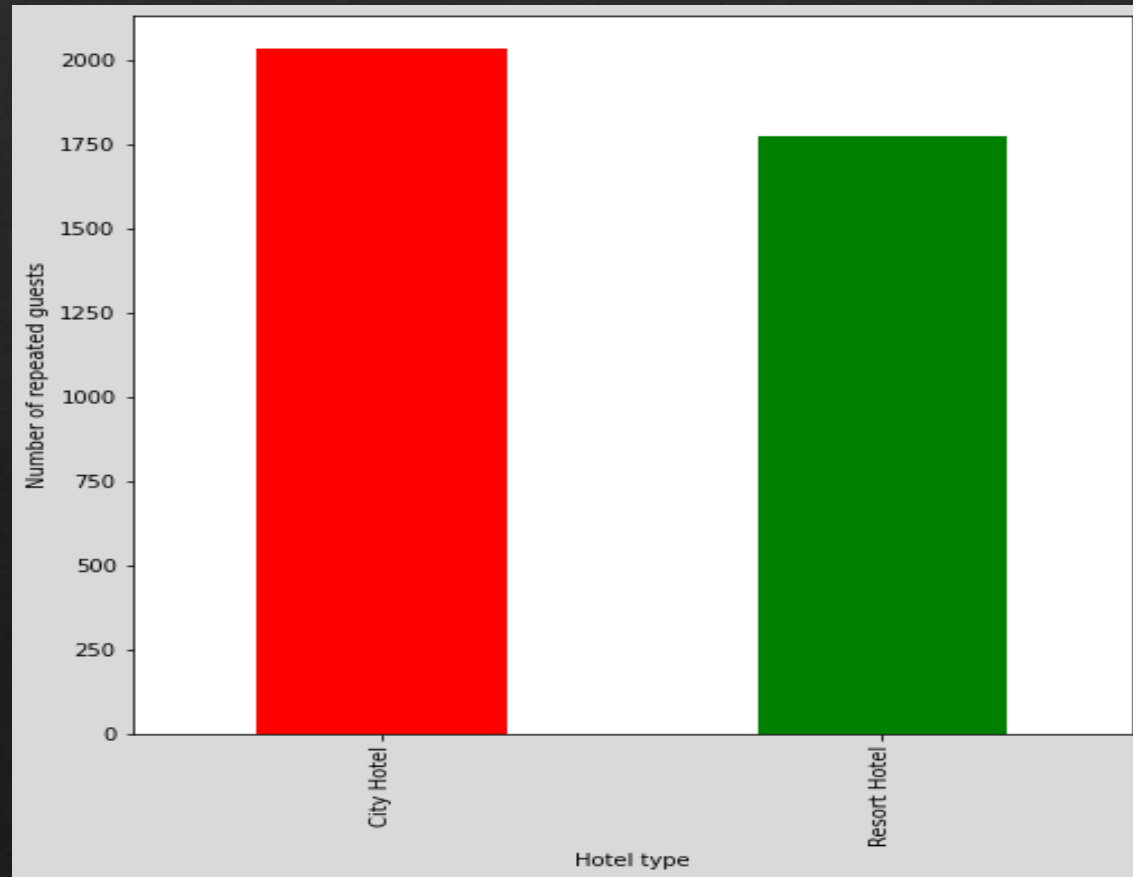# Countplot of distribution channels
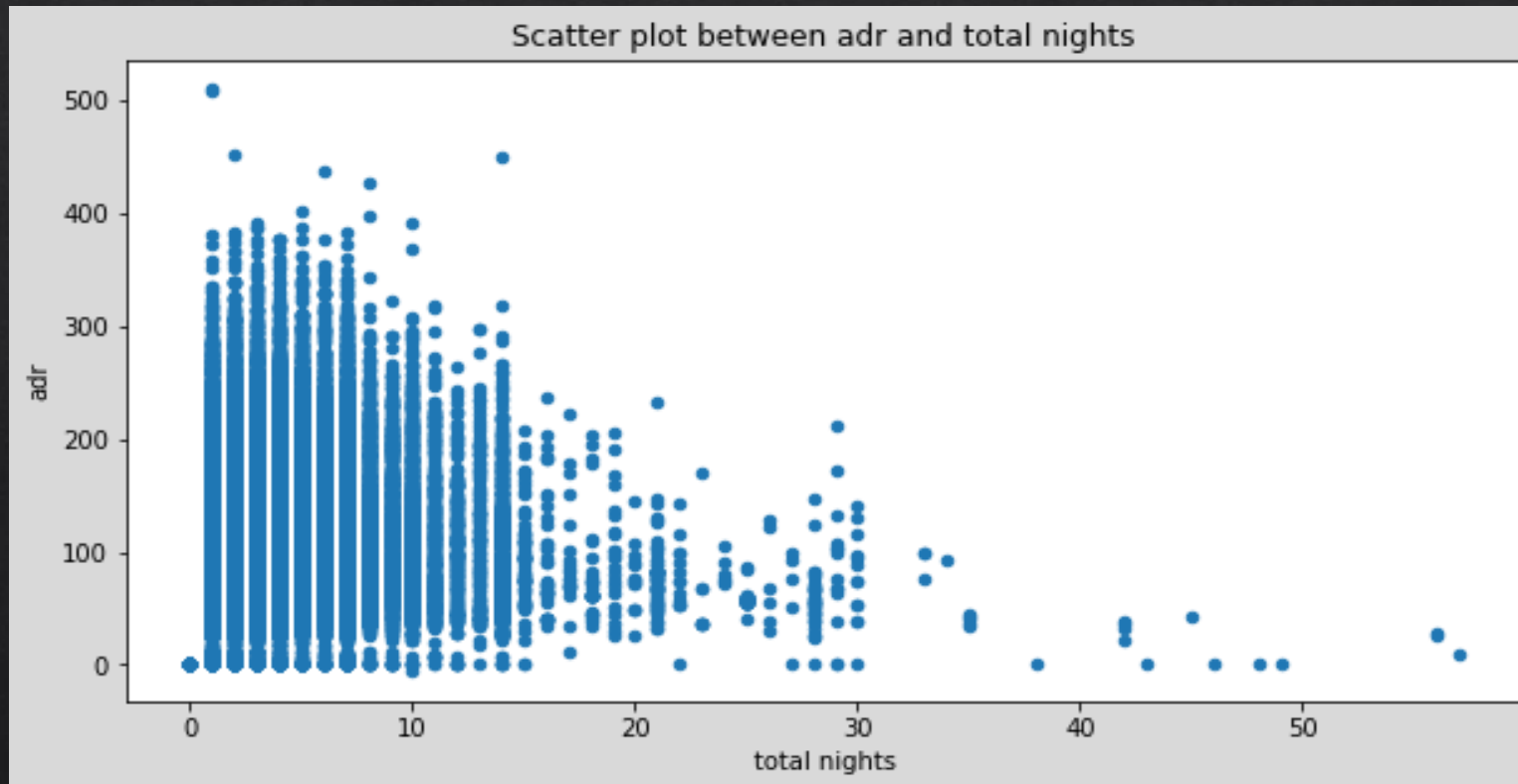
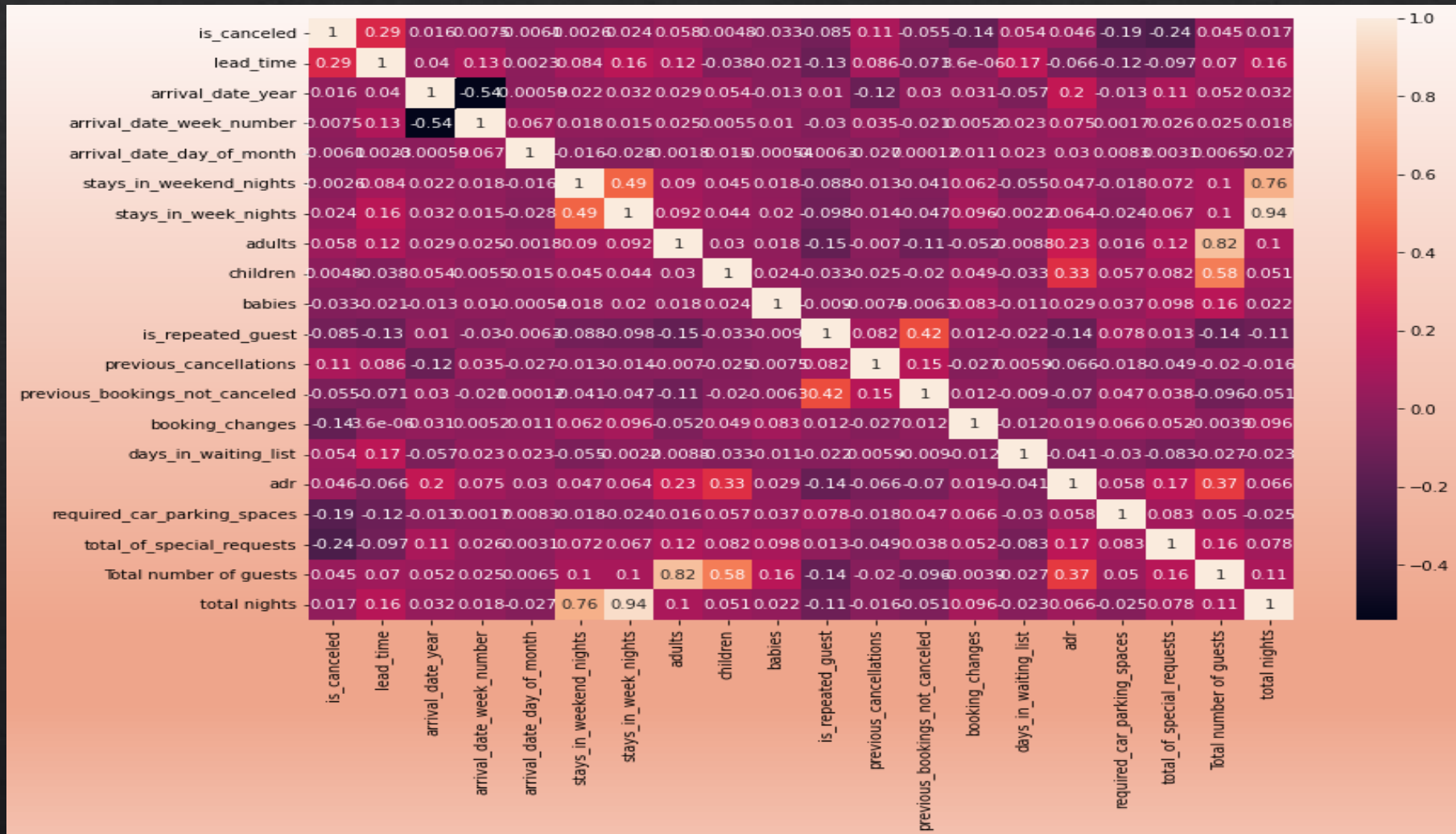# Countplot of cancellation by hotel type

Countplot of deposit types

Bar plot of number of repeated guests.

Scatterplot between adr and Total number of nights.

Correlation heatmap

Conclusions

Here is what I learnt after completing the analysis.

1. There is a big difference in the number of city hotels, over resort hotels. (66.7% of people prefer city hotel as compared to 33.3% who prefer res ort hotel.

2. Customers from the country of Portugal are higher in number when compared to customers from other countries.

3. Most people prefer the bed and breakfast meal plan.

4. August is the month with the highest number of bookings and January has the least number of bookings.

5. TA/TO distribution channel is the most successful distribution channel, with D irect distribution channel in the second place.

6. City hotels have a higher percentage of cancellation.

7. Both the hotels have a high number of people who do not pay deposits.

8. From the chart, we can see that there are very few repeated guests for both hotel types.

9. There are very few correlations among the variables.

Challenges faces

Challenges faced:

1. There were quite a few null values especially in the columns 'company' and 'agent'.

2. There were very few correlations between the different variables which made it hard to derive insights based on their relationships.

3. Pair plot computation time was a bit high.