CAPSTONE PROJECT - 1

EDA on Hotel Booking Analysis

By

KISHAN SINGH

PROBLEM STATEMENT

- Using the provided dataset and various important library of python we want to know how we can attract customers and by using various planned approaches how to gain more profit.

- In this project we are finding important insights by exploring the given dataset of Hotel.

WORK FLOW :-

- The work flow is divided into following three steps.

DATA COLLECTION AND UNDERSTANDING

- hotel = Resort Hotel or city hotel
- is_cancelled = If the booking was cancelled (1) or not(0)
- lead_time = Number of days that elapsed between the entering date of the booking into the PMS and the arrival date.
- arrival_date_year = Year of arrival date
- arrival_date_month = Month of arrival date
- arrival_date_week_number = Week number of arrival date

- stays_in_weekend_nights = Number of weekend nights (saturday or sunday) the guest stayed or booked to stay at hotel.
- stays_in_week_nights = Number of week nights (Monday to Friday) the guest stayed or booked to stay at hotel.
- adults = Number of adults
- children = Number of children
- babies = Number of babies
- meal = kind of meal opted for
- country = country code
- market_segment = which segment the customer belongs to

- Distribution_channel = How the customer accessed the stay-corporate booking/Direct/TA.TO
- is_repeated_guest = Guest coming for first time or not
- previous_cancellation = Was there a cancellation before
- previous_bookings = count of previous bookings
- reserved_room_type = Type of room reserved
- assigned_room_type = Type of room assigned
- booking_changes = Count of changes made to booking

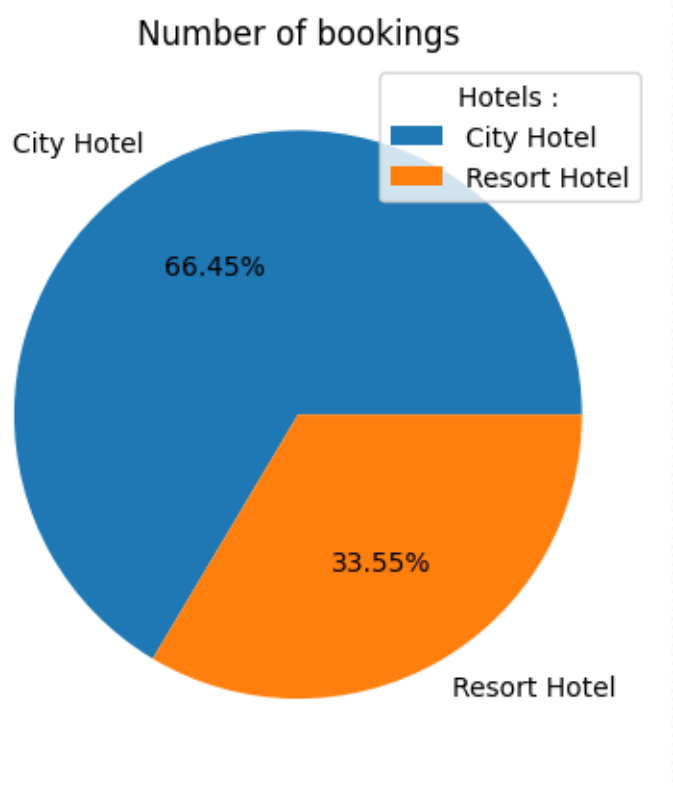
- Deposit_type = Deposit type
- agent = booked through agent
- days_in_waiting_list = No of days in waiting list
- customer_type = Type of customer
- required_car_parking = If car parking is required
- total_of_special_req = Number of additional special requirements
- reservation_status = status of reservation
- reservation_status_date = Date of specific status

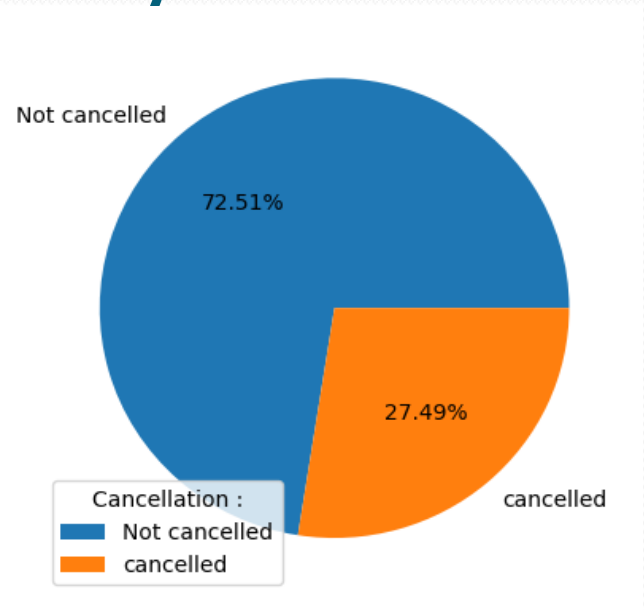
DATA CLEANING AND MANIPULATION

- The dataset 119390 rows and 32 columns.
- The dataset have 31994 duplicate values so duplicate values are dropped by using **drop_duplicates**()
- The dataset have 4 columns that have missing values.
- The column children has 4 missing values.
- The column country has 452 missing values.
- The column agent has 12193 missing values.
- The column company has 82137 missing values.

- Now I have replaced all null values of column children to zero(0).
- All null values of column country is replaced by “others”.
- All null values of column agent is replaced by zero(0).
- All null values of column company is replaced by “No company”.
- As data type of column ‘children’ is float. I replaced data type from float to int.
- The new column “total_people” is added by summing columns “adults”, “children” and “babies”.

Exploratory Data Analysis

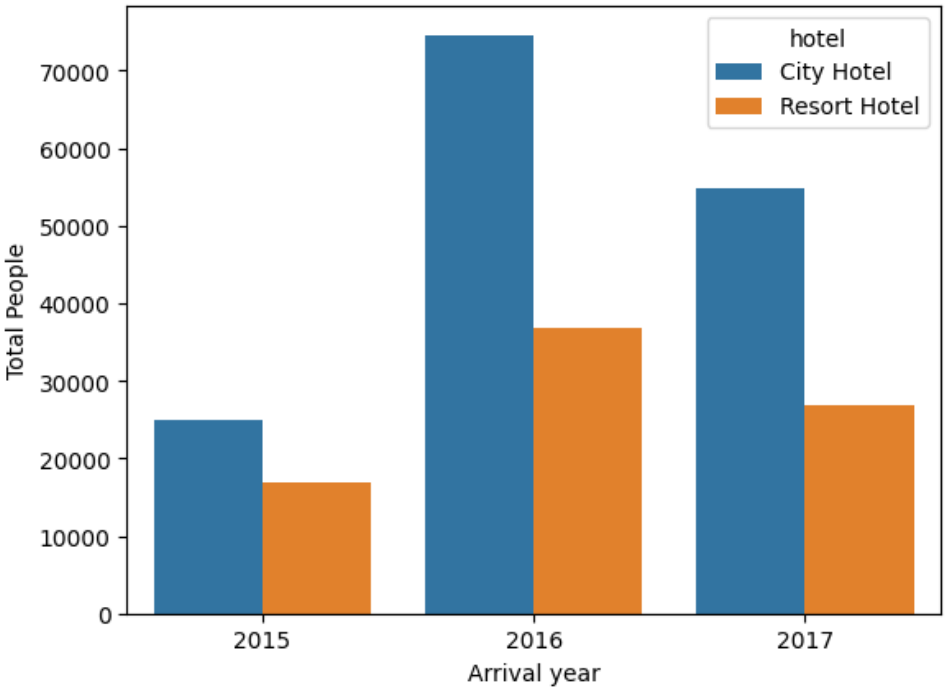


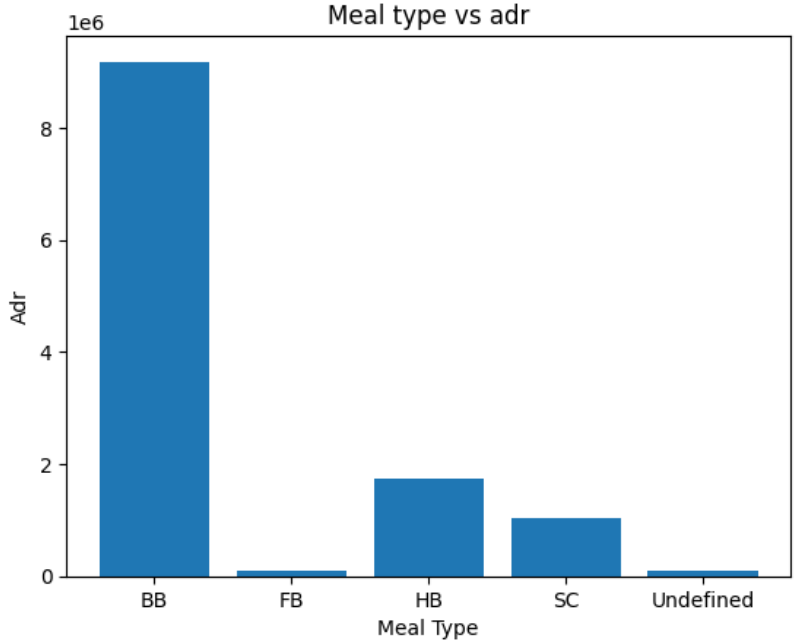


Conclusion :-

(1) **Number of bookings is more in city hotel than resort hotel.**

(2) **Almost 28% orders got cancelled.**

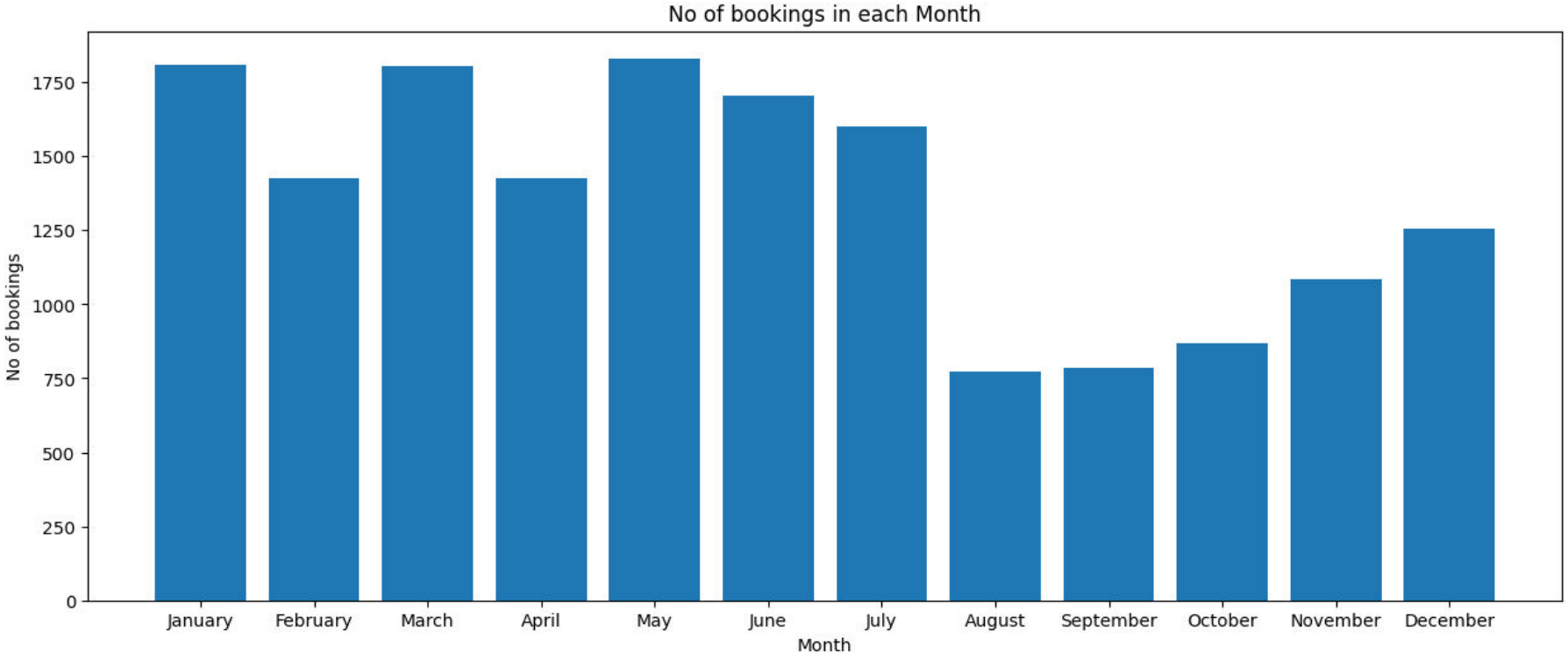


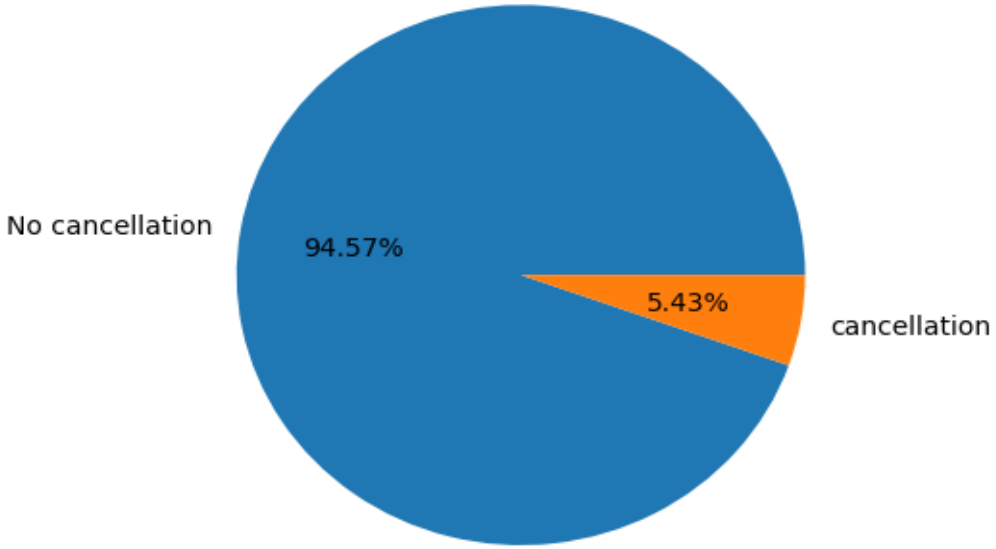


Conclusion :-

(1) In 2015,least number of guest were arrived. Also in 2016, maximum number of guest were arrived.

(2) 'BB' meal type has greatest adr and 'FB' has least adr.

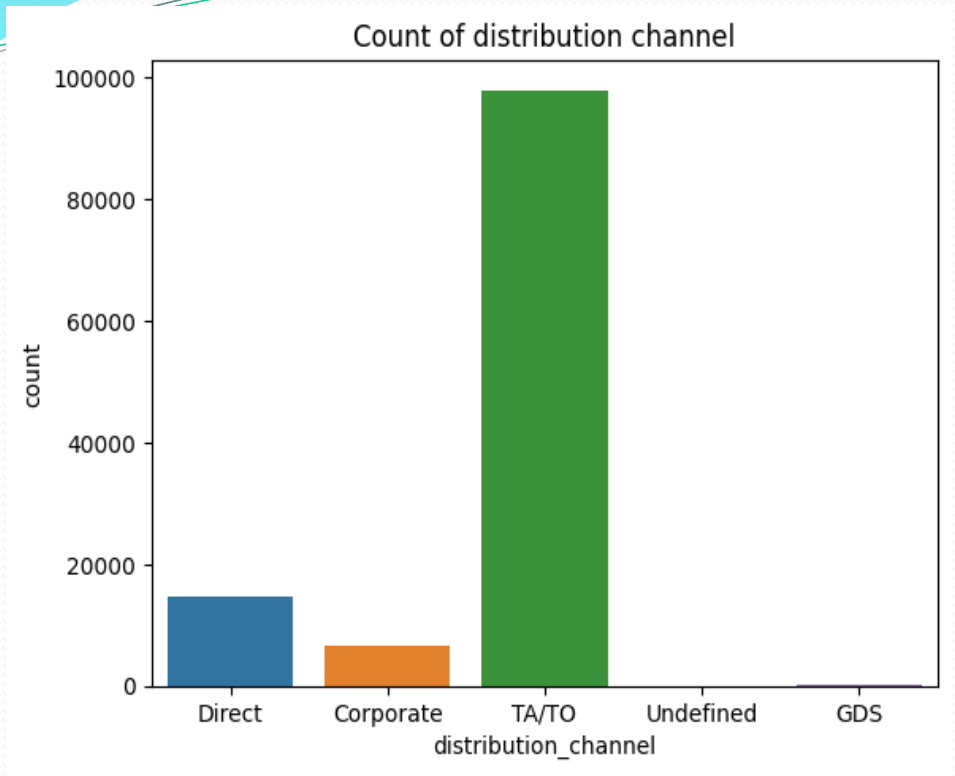


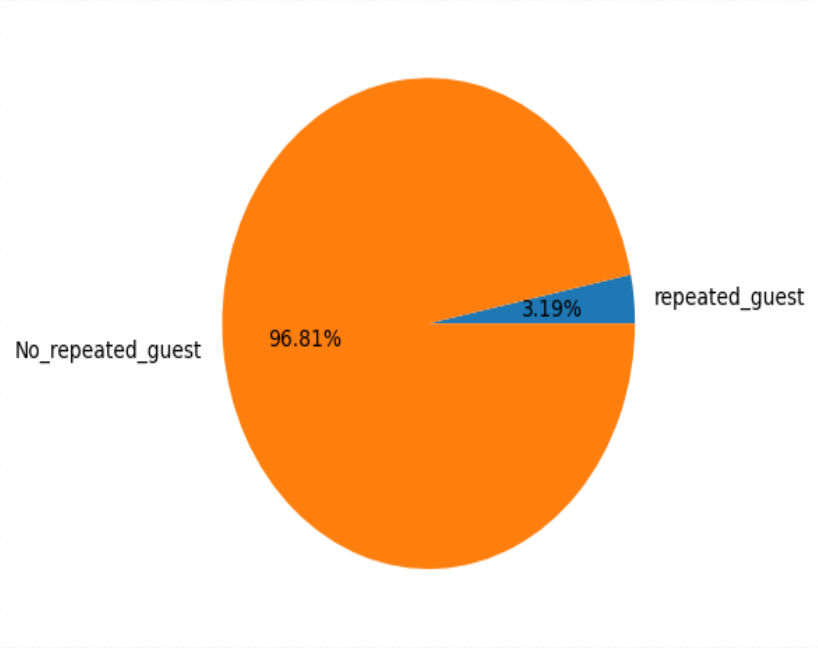


Conclusion :-

(1) Only 5.43% cancellation was done.

(2) In january, March and May month maximum bookings were done by guests. Also in August and september least bookings were done.

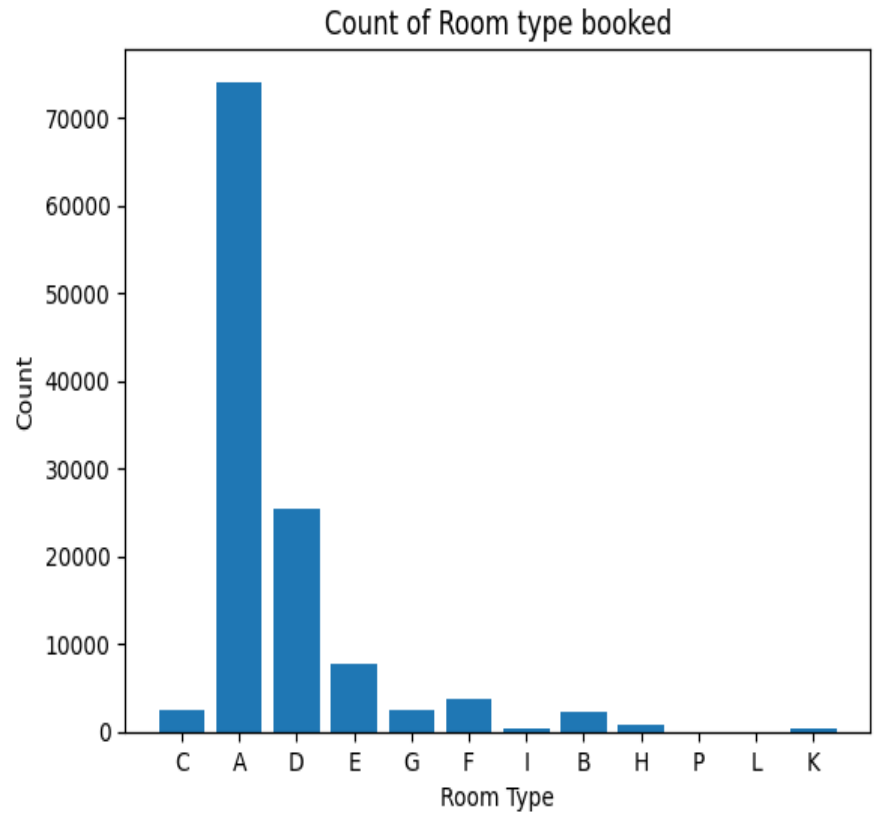


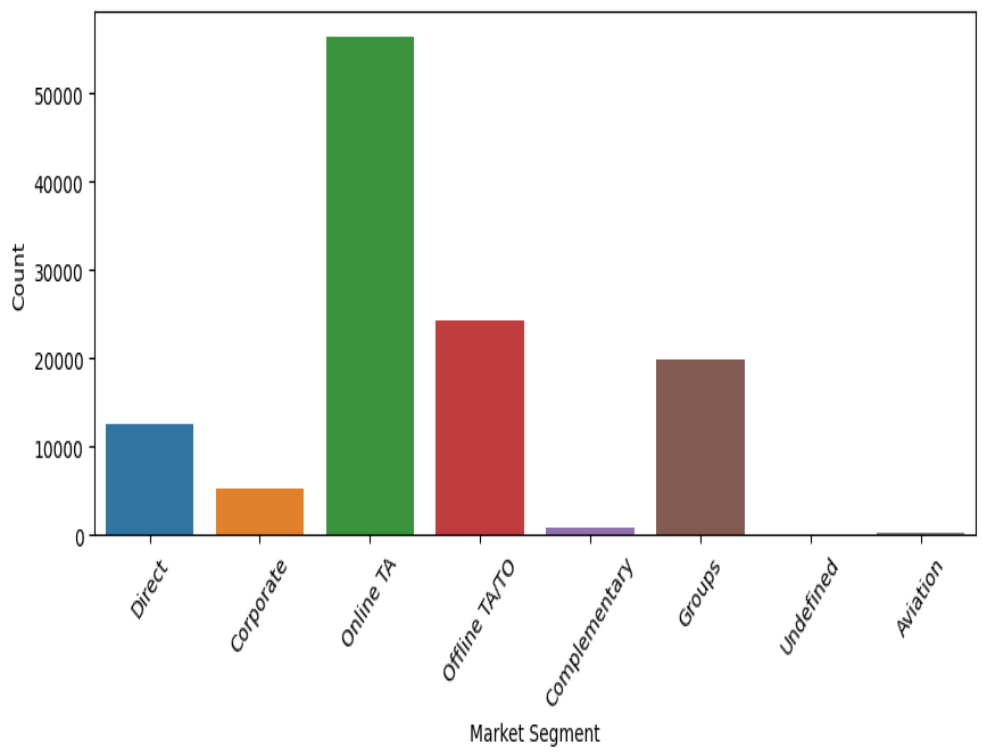


Conclusion :-

(1) **TA/TO distribution_channel has maximum count whereas GDS has least count.**

(2) **Very less guests are repeated according to this chart. So Attractive offers should be given to guests to increase percentage of repeated guest.**

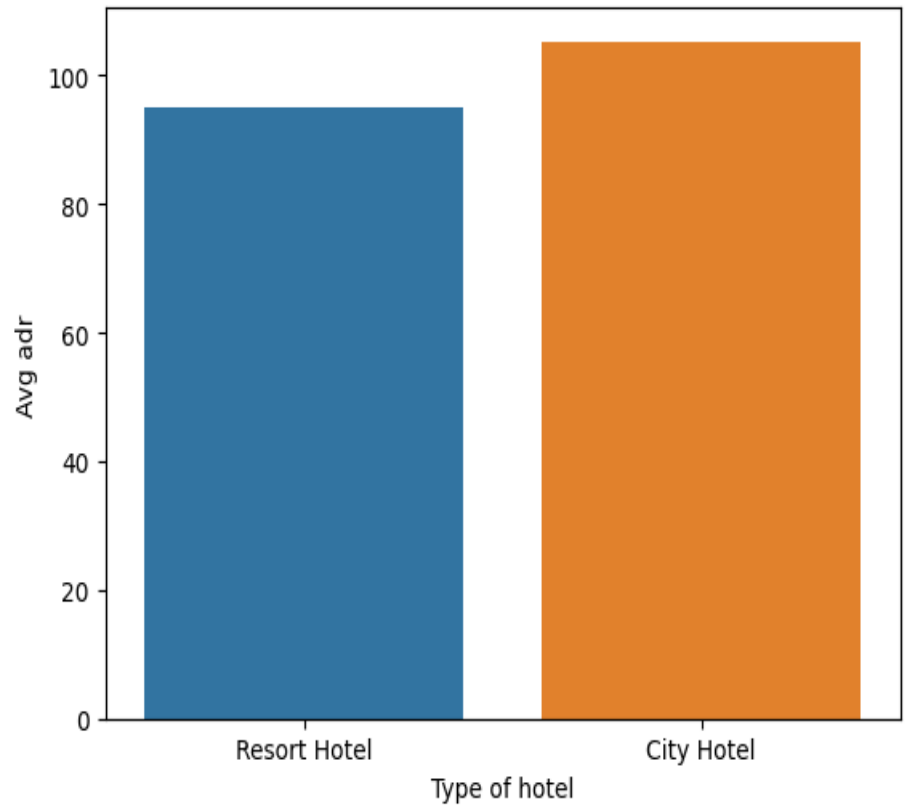


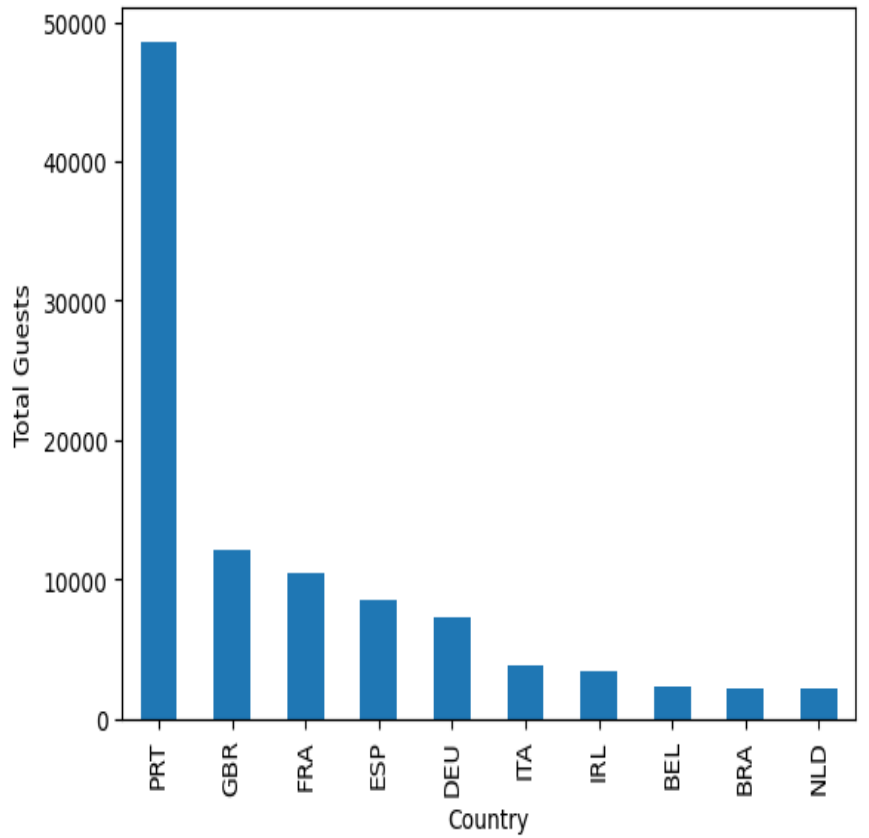


Conclusion :-

(1) Room type A is used by most of guests.

(2) Online TA market segment was used by most of the guests.

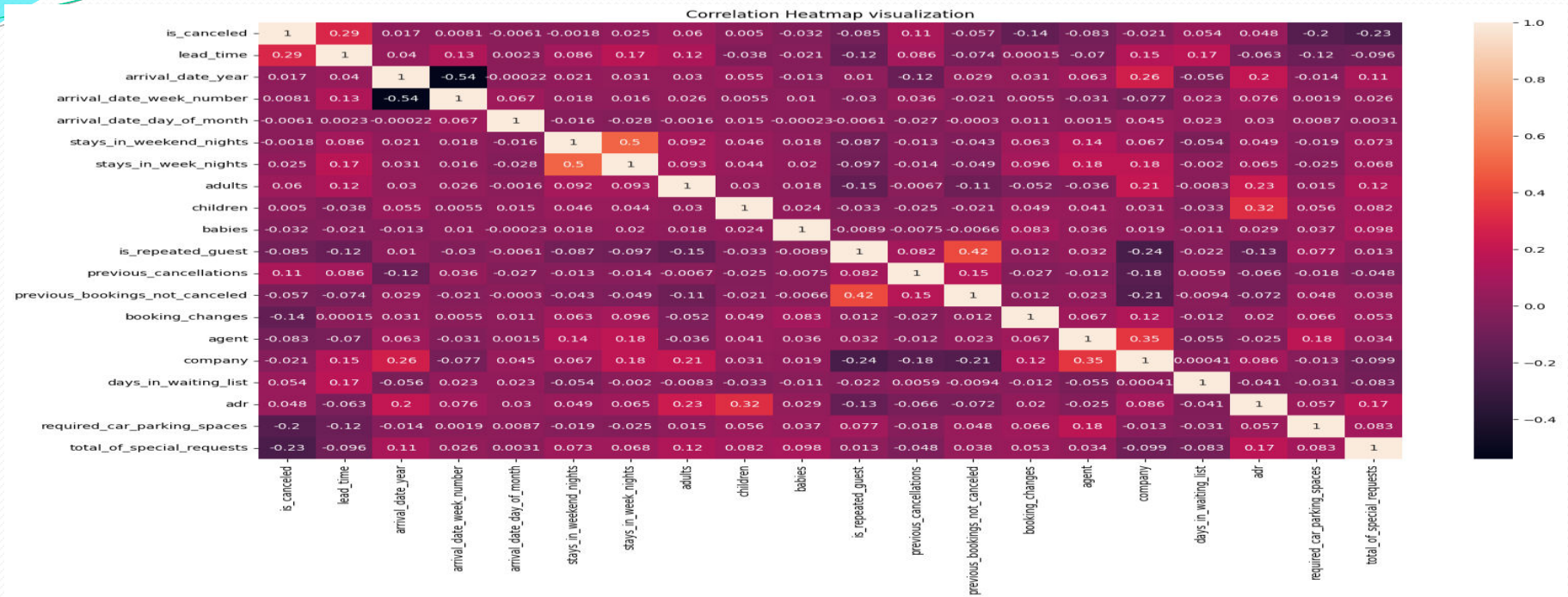


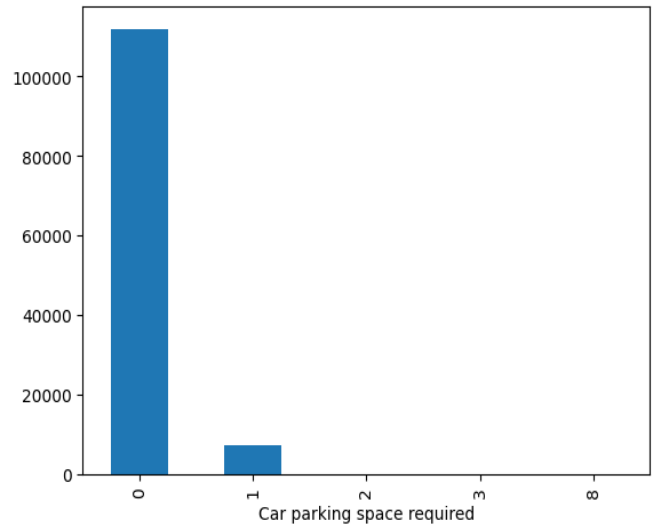


Conclusion :-

(1) City hotel has more average adr than resort hotel.

(2) "PRT" country code has maximum number of guests.





Conclusion :-

(1) Car parking space requirement is very less.

(2) Highest correlation value is 95% positive and lowest correlation value is -51% negative.

Solution to Business Objective

- Client can increase revenue as shown in bar chart of room type most preferred and month which have more guests.
- Meal type, required car parking spaces etc factors should be considered to better planning of increasing profit and growth.
- So using all these outcomes, Client can plan their strategy to get more revenue and reduce complaints regarding their products and enhancing business growth.

Solution to Business Objective

- As repeated guest are very less, feedback of guests should be taken to increase number of repeated guest in hotel.
- Also attractive offers should be given to old guests to increase percentage of repeated guests.

Conclusion

- Number of bookings is more in city hotel rather than resort hotel.
- Almost 28% bookings got cancelled.
- In 2015, number of arriving guests are less as compared to year 2016.
- 'BB' meal type got maximum adr so by 'BB' meal type we are getting maximum revenue.
- Only 1.93% cancellation was done by guests.
- In january, March and May month maximum bookings were done by guests. Also in August and september least bookings were done.

Conclusion

- 'TA/TO' distribution_channel has maximum count.
- 3.19% guests are repeated which is very less so focus on problems faced by guests.
- Room type 'A' has greatest number of bookings.
- Online TA market segment was used by most of the guests.
- City hotel has more average adr than resort hotel so cit y hotel is earning more revenue than resort hotel.

Conclusion

- In 'PRT' country, there are maximum number of guests
- Car parking space requirement is very less.
- Highest correlation value is 95% positive and lowest correlation value is -51% negative.