CAPSTONE PROJECT

EDA On Hotel Booking Analysis

By - Ambuj Kumar Dubey

POINTS TO DISCUSS

Objective

Data Summary

Exploratory Data Analysis(EDA)

Conclusion

OBJECTIVE

The main objective behind this project is to explore and analyse data to discover important factor that govern the bookings and give insights to hotel management , which can perform various campaigns to boost the business and performance.

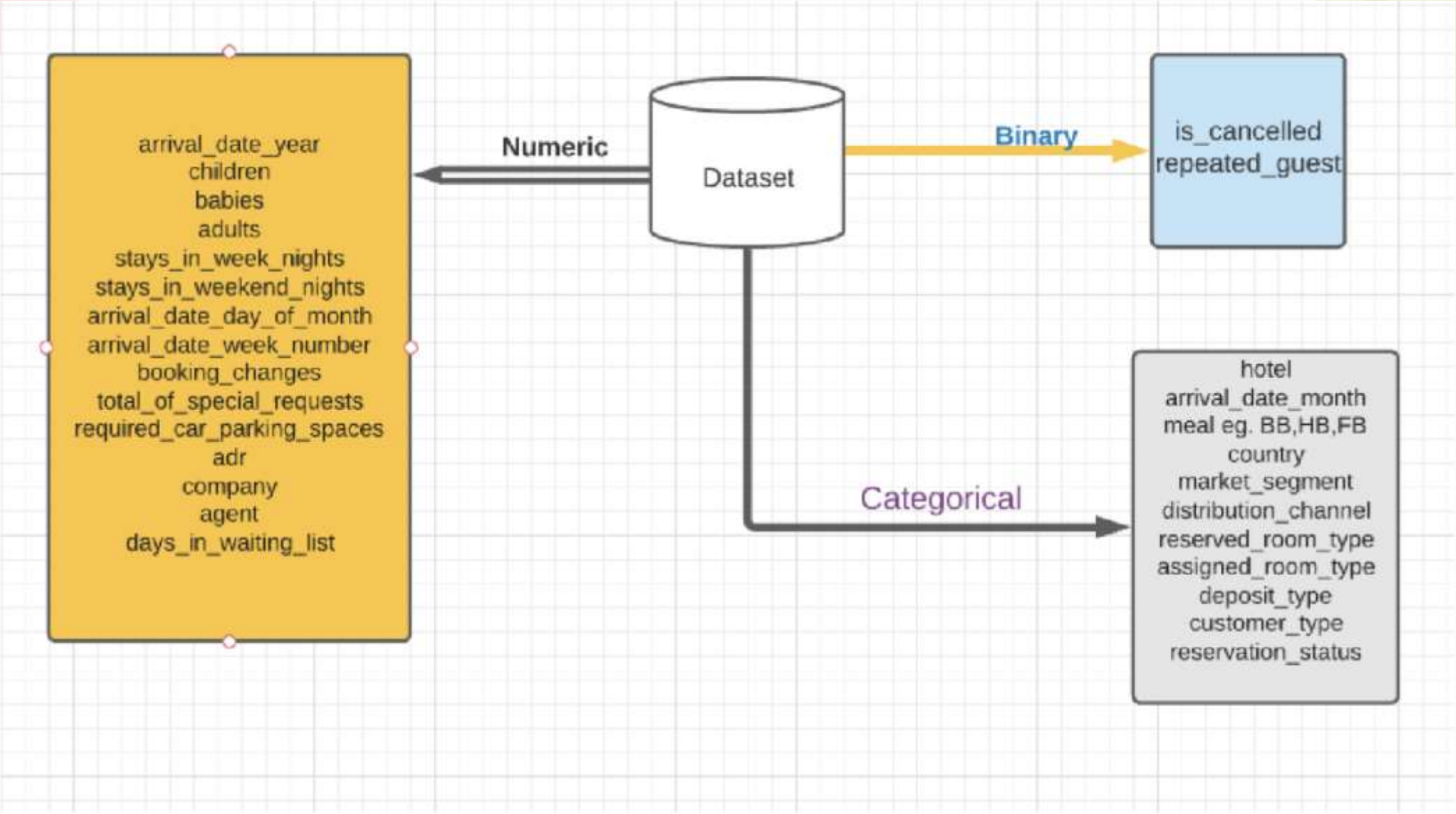
For this project, we will analyze Hotel Booking data. The data set contains booking information for a city hotel and a resort hotel.It also includes information such as when the booking was made , length of stay , the number of adults , children , or babies, and the number of available parking spaces.

DATA SUMMARY

The data set has 32 columns of variables crucial for hotel booking. Some variables are as follows:

- hotel: The category of hotel (City or Resort).
- is_canceled: Whether the booking is canceled or not (0 for no canceled and 1 for canceled).
- lead_time: time (in days) between booking transaction and actual arrival.
- stays_in_weekend_nights: No. of weekend nights spent.
- stays_in_week_nights: No. of weeknights spent in a hotel.
- arrival_date_year: Year of arrival
- arrival_date_month: Month of arrival
- meal: Type of meal chosen(BB,FB,HB,SC,Undefined)

- market_segment: What segment via booking was made and for what purpose.
- distribution_channel: Medium through which booking was made.
- is_repeated_guest: Whether the customer has made any booking before(0 for No and 1 for Yes).
- previous_cancellations: No. of previous canceled bookings.
- previous_bookings_not_canceled: No. of previous non-canceled bookings.
- days_in_waiting_list: No. of days on waiting list.
- customer_type: Type of customer(Transient, Group, etc.)
- adr: Average Daily rate.
- required_car_parking_spaces: No. of car parking asked in booking.
- country: Country of origin of customers.


arrival_date_year
children
babies
adults
stays_in_week_nights
stays_in_weekend_nights
arrival_date_day_of_month
arrival_date_week_number
booking_changes
total_of_special_requests
required_car_parking_spaces
adr
company
agent
days_in_waiting_list
Numeric
Dataset
Binary
is_cancelled
repeated_guest
Categorical
hotel
arrival_date_month
meal eg. BB,HB,FB
country
market_segment
distribution_channel
reserved_room_type
assigned_room_type
deposit_type
customer_type
reservation_status

EXPLORATORY DATA ANALYSIS

PERCENTAGE OF BOOKINGS IN CITY HOTEL AND RESORT HOTEL

INFERENCE:

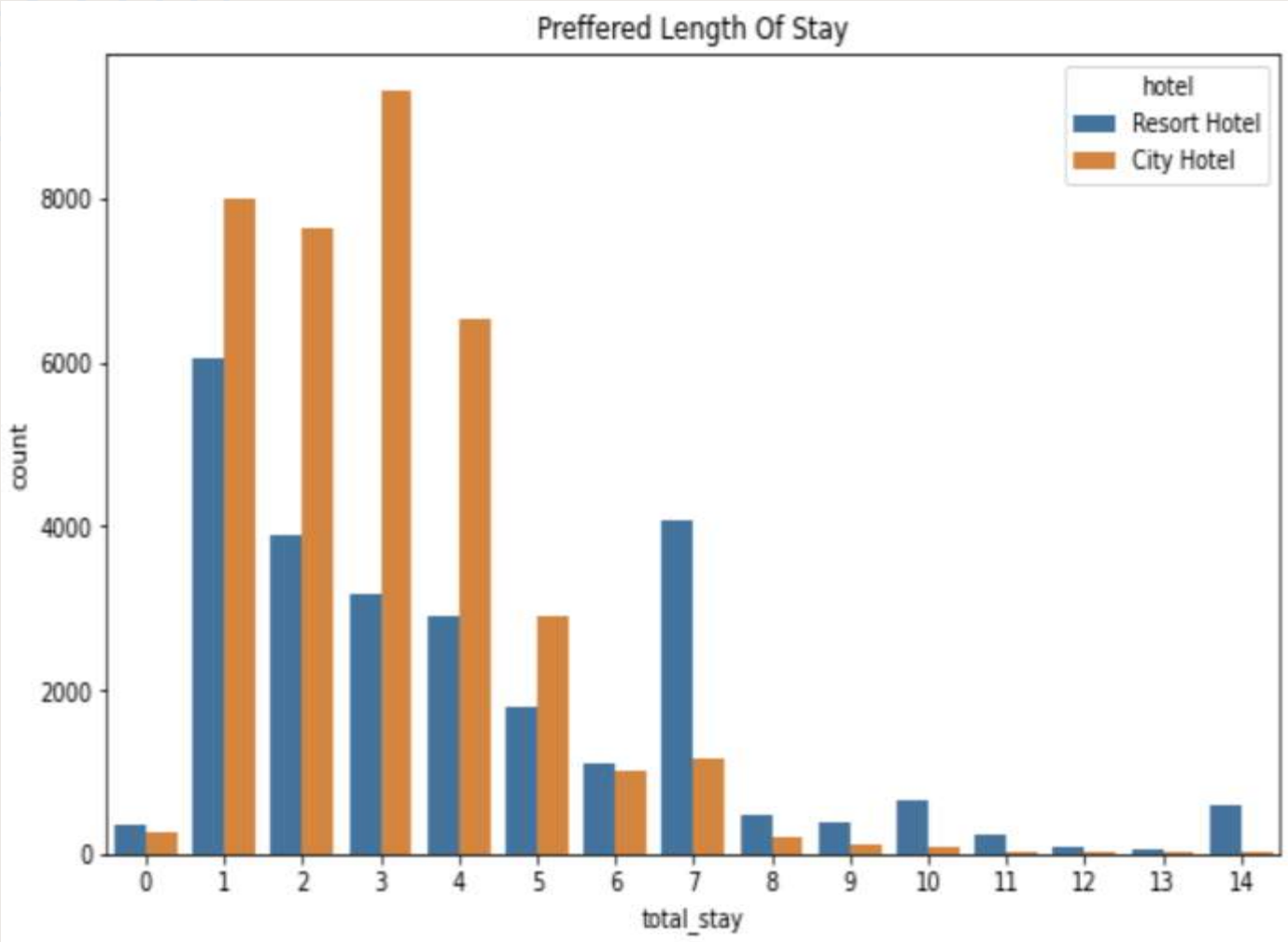
The city hotel has 60% bookings while the resort hotel has 40% bookings.

HOTEL WITH LONGER WAITING TIME

INFERENCE:

City hotel has slightly longer waiting time.Hence,it is most busier than resort hotel.

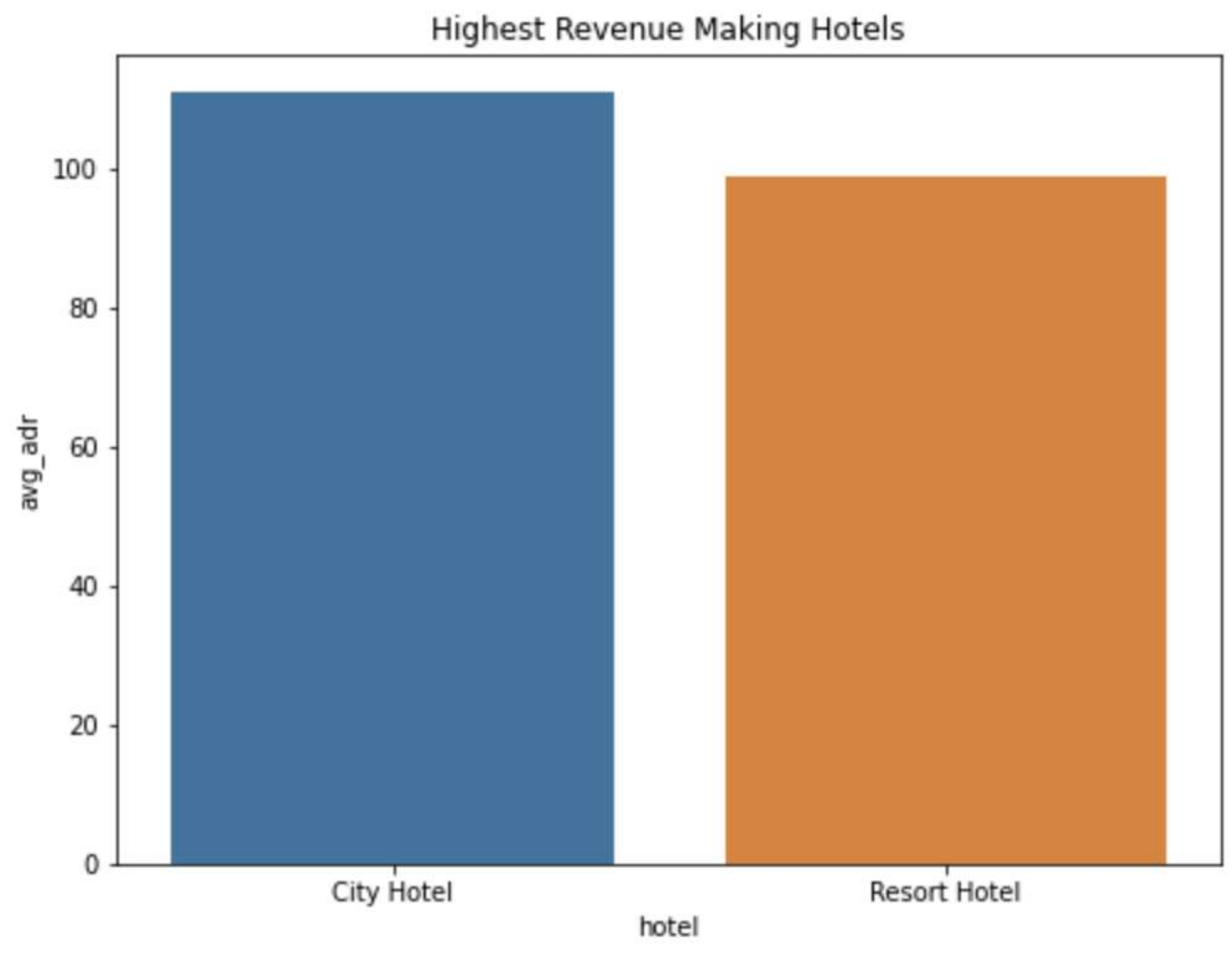
PREFFERED LENGTH OF STAY IN EACH HOTEL



INFERENCE:

Preffered length of stay in city hot el is around 3 days and in resort h otel is around 2 days.

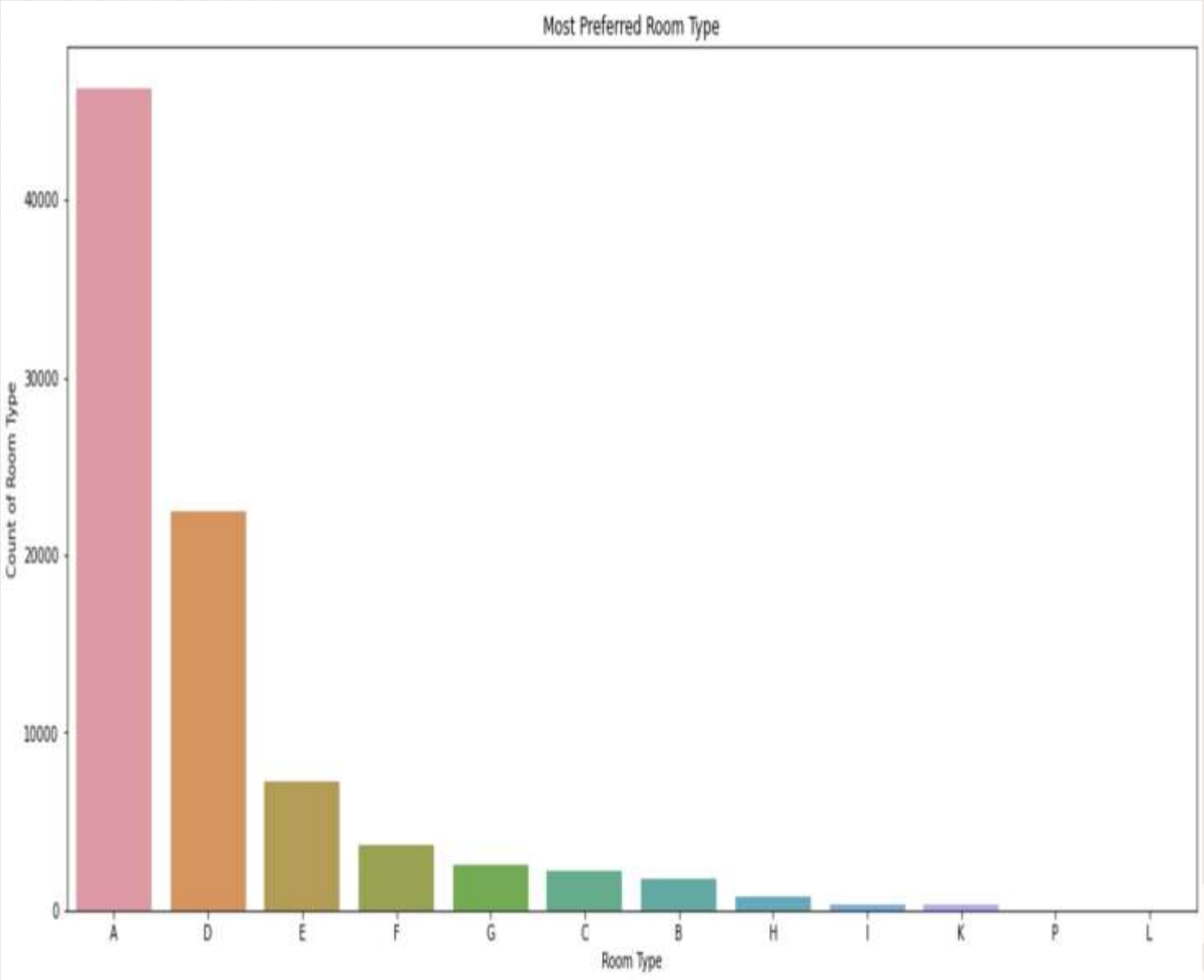
HOTEL MAKING HIGHEST REVENUE



INFERENCE:

City hotel seems to be making slightly more revenue.

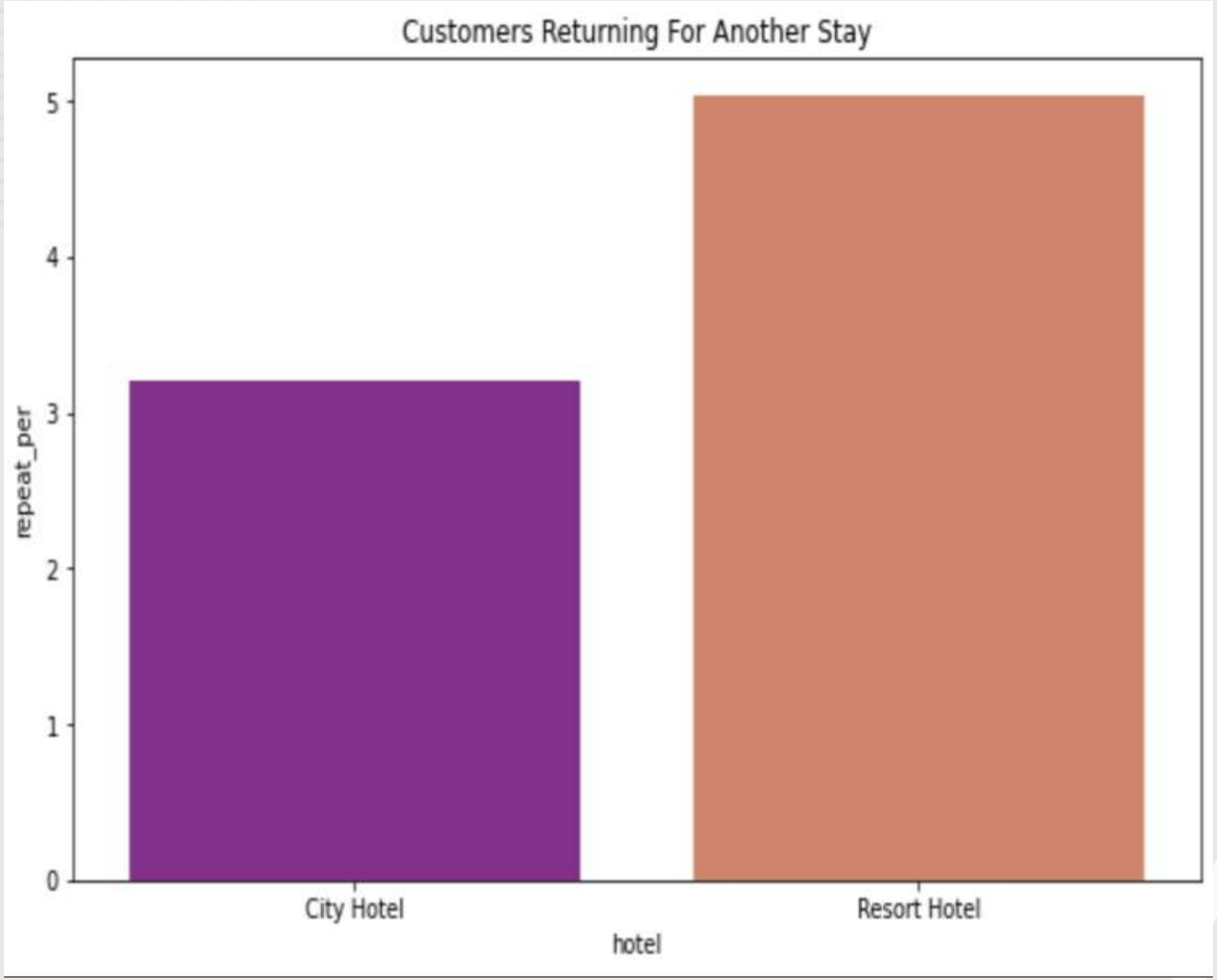
ROOM TYPE PREFERRED BY THE CUSTOMER



INFERENCE:

Most preffered room type is 'A'.

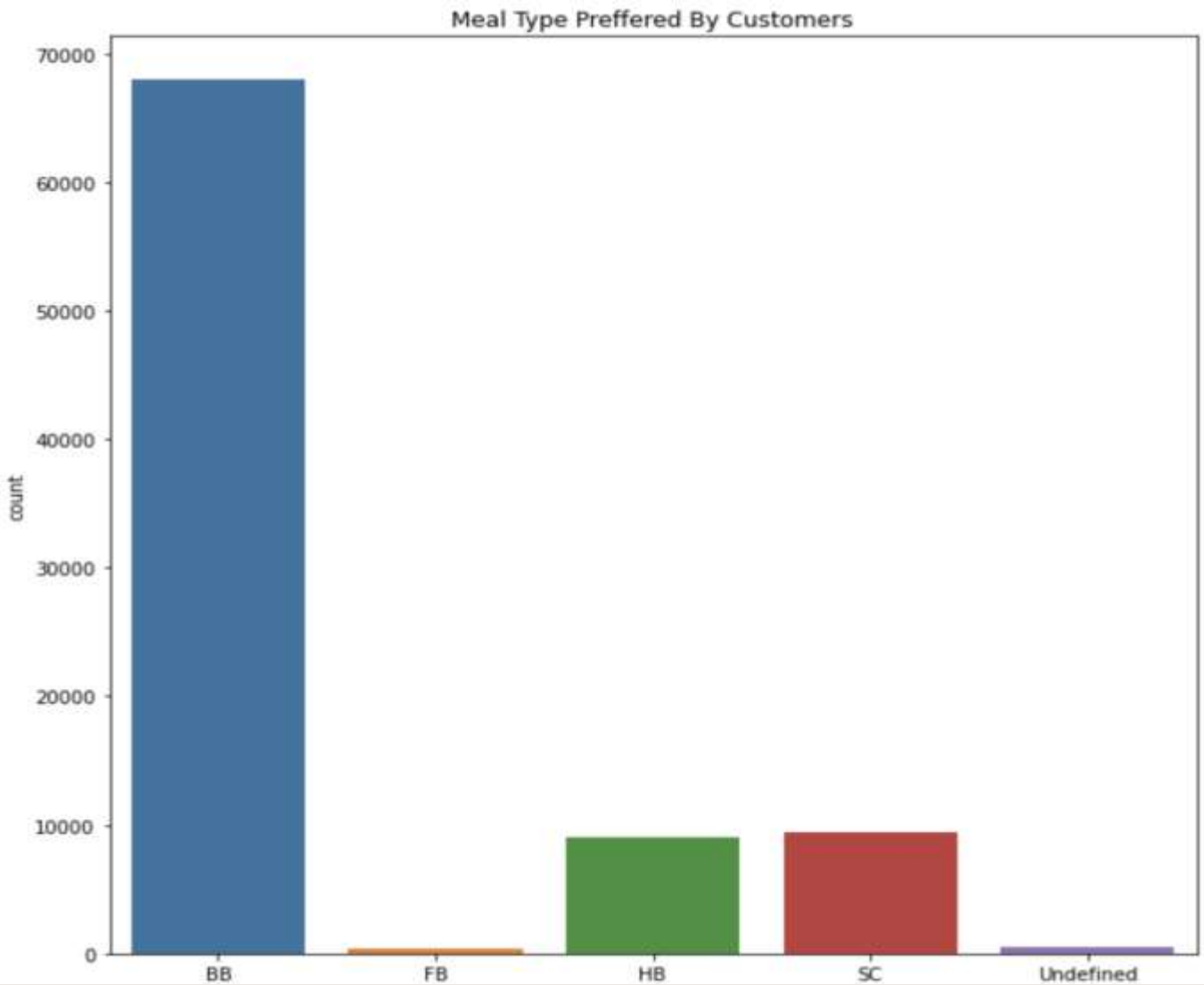
HOTEL WITH HIGHER CHANCE OF CUSTOMERS RETURNING FOR ANOTHER STAY



INFERENCE:

The resort hotel has higher chanc e of customers returning for another stay.

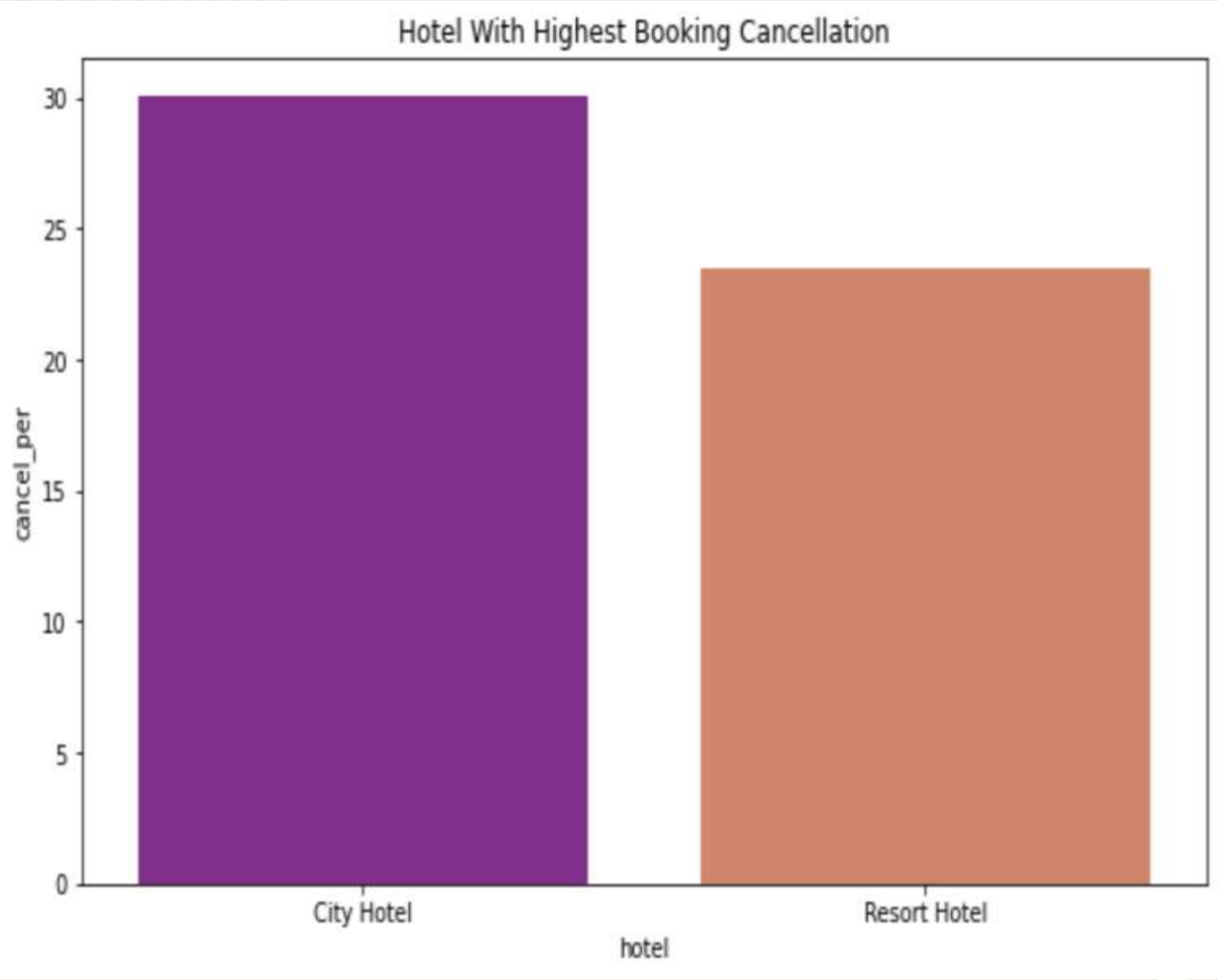
MOST PREFFERED MEAL TYPE BY CUSTOMERS



INFERENCE:

Most preferred meal type is BB (Bed and breakfast).

HOTEL WITH HIGHEST BOOKING CANCELLATION



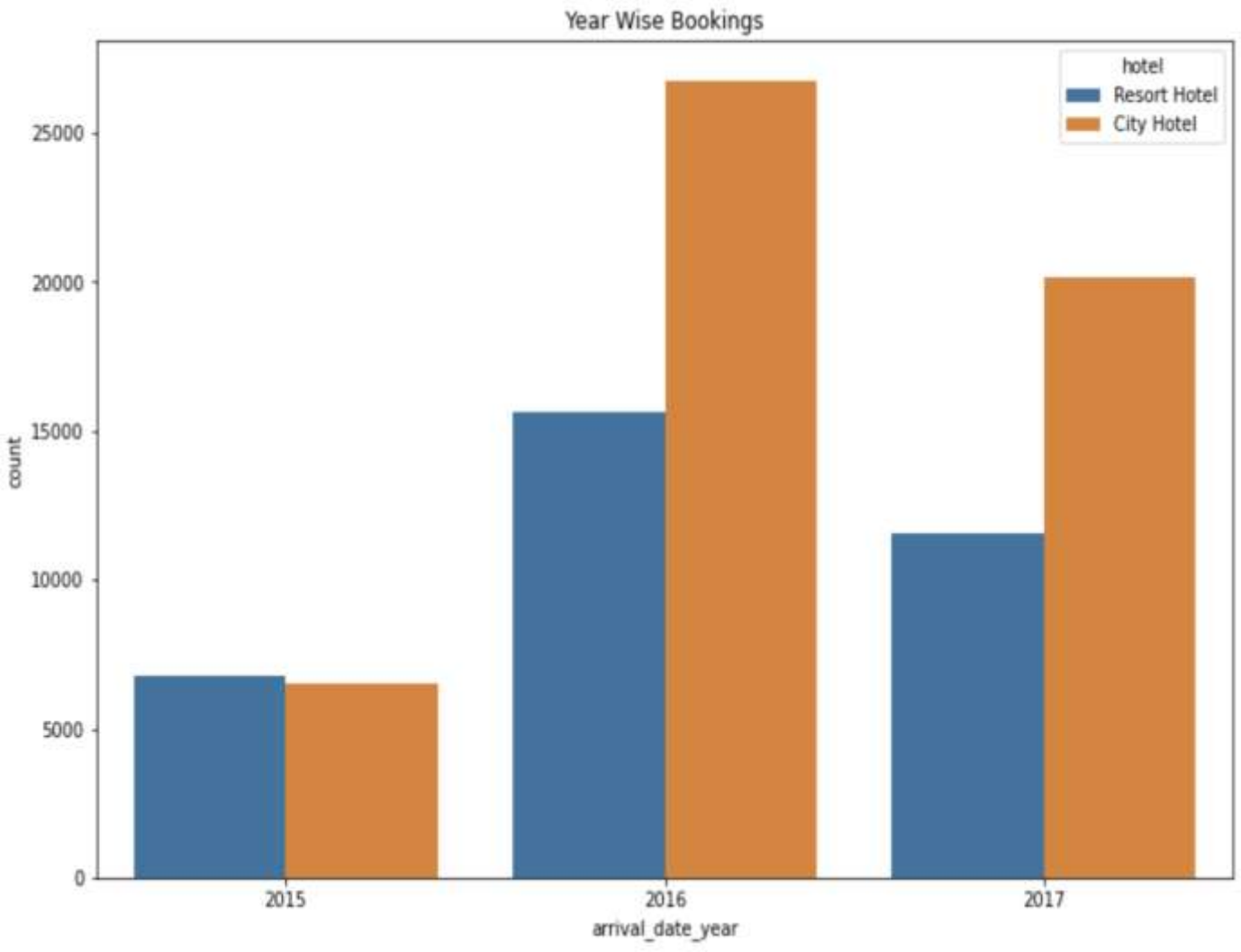
INFERENCE:

The rate of cancellation of booki ng is highest in city hotel.

Strategy to Counter Cancellation:

- **Set Non-refundable Rates, Collect deposits, and implement more rigid cancellation policies.**
- **Encourage Direct bookings by offering special discounts**
- **Monitor where the cancellations are coming from such as Market Segment, distribution channels, etc.**

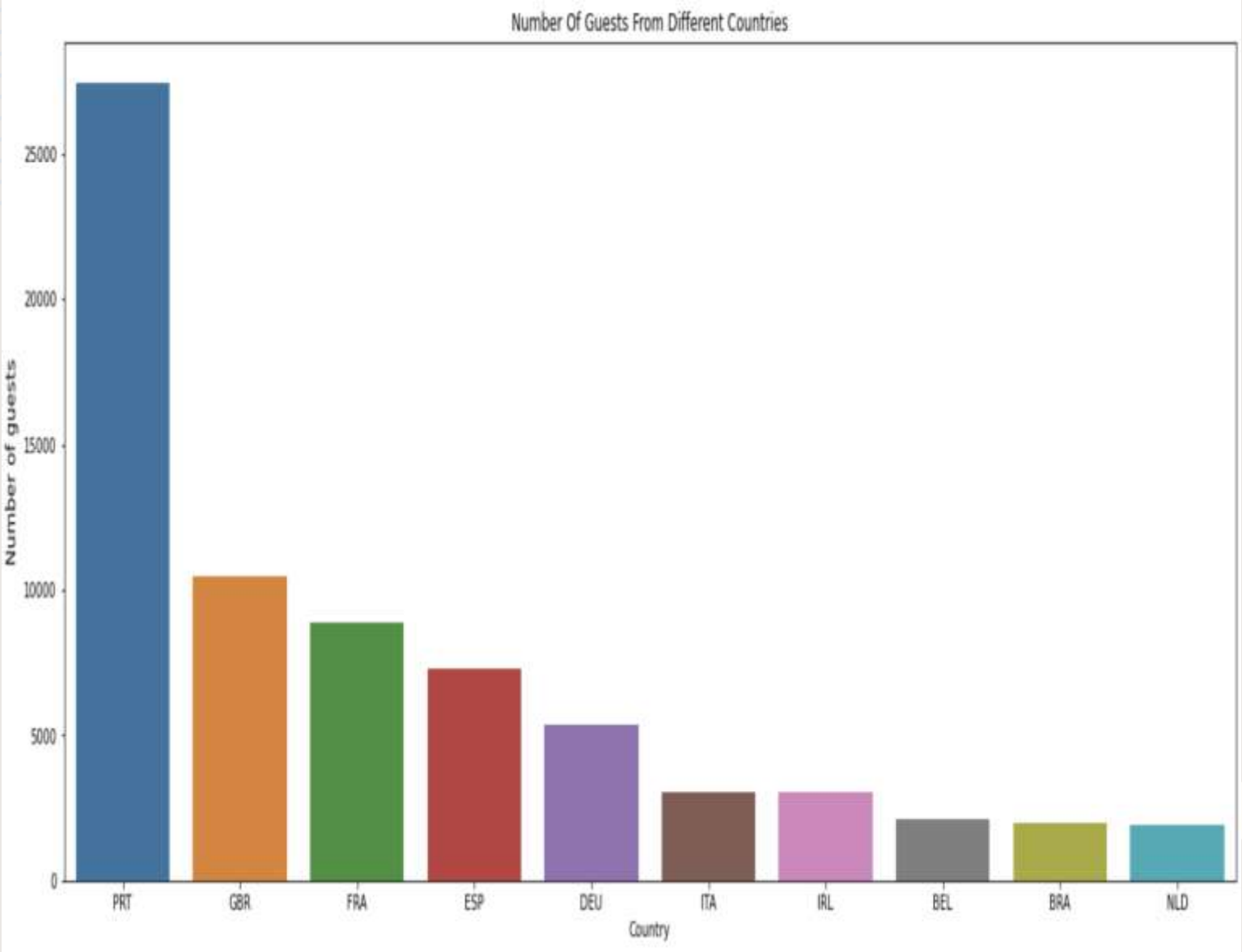
YEAR WITH HIGHEST NUMBER OF BOOKINGS



INFERENCE:

City hotels had the most of the bookings.
2015 had less 7000 bookings.
2016 had the higest bookings.

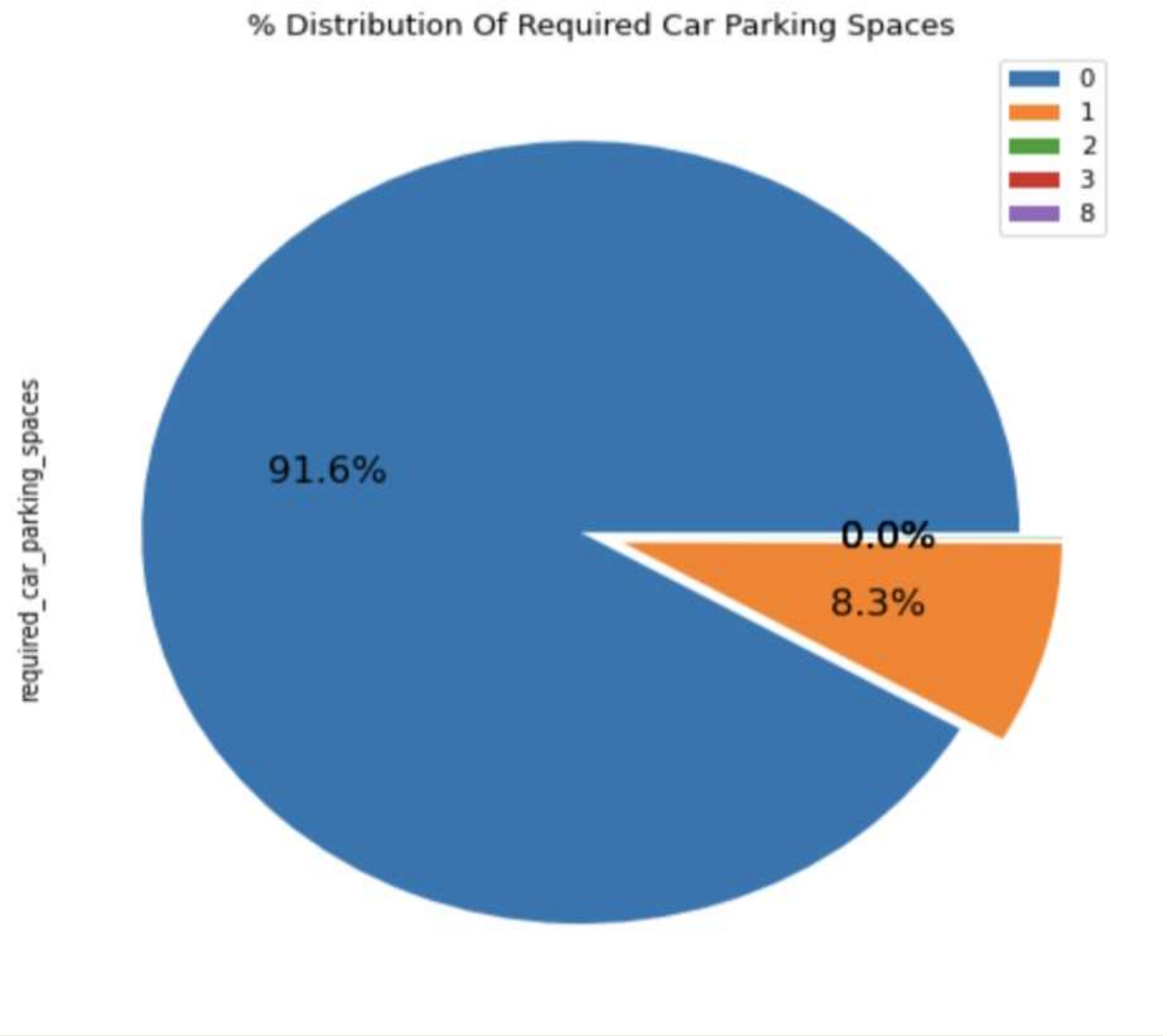
NUMBER OF GUESTS FROM DIFFERENT COUNTRIES



INFERENCE:

Most of the guests are coming fr omportugal i.e more 25000 guest s are from portugal.

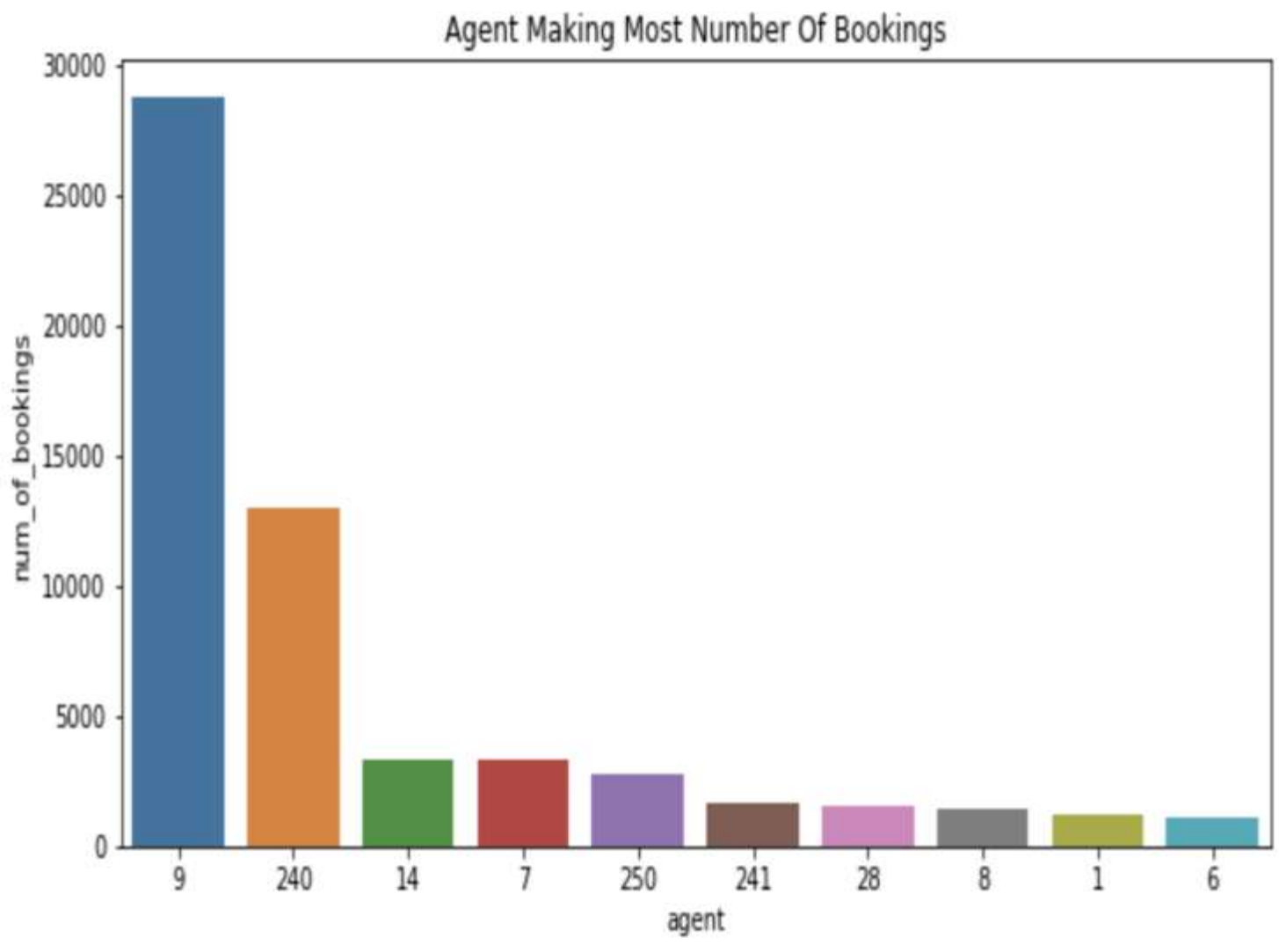
PERCENTAGE DISTRIBUTION OF REQUIRED CAR PARKING SPACES



INFERENCE:

91.6 % guests did not required the parking space.
Only 8.3 % guests required only 1 parking space.

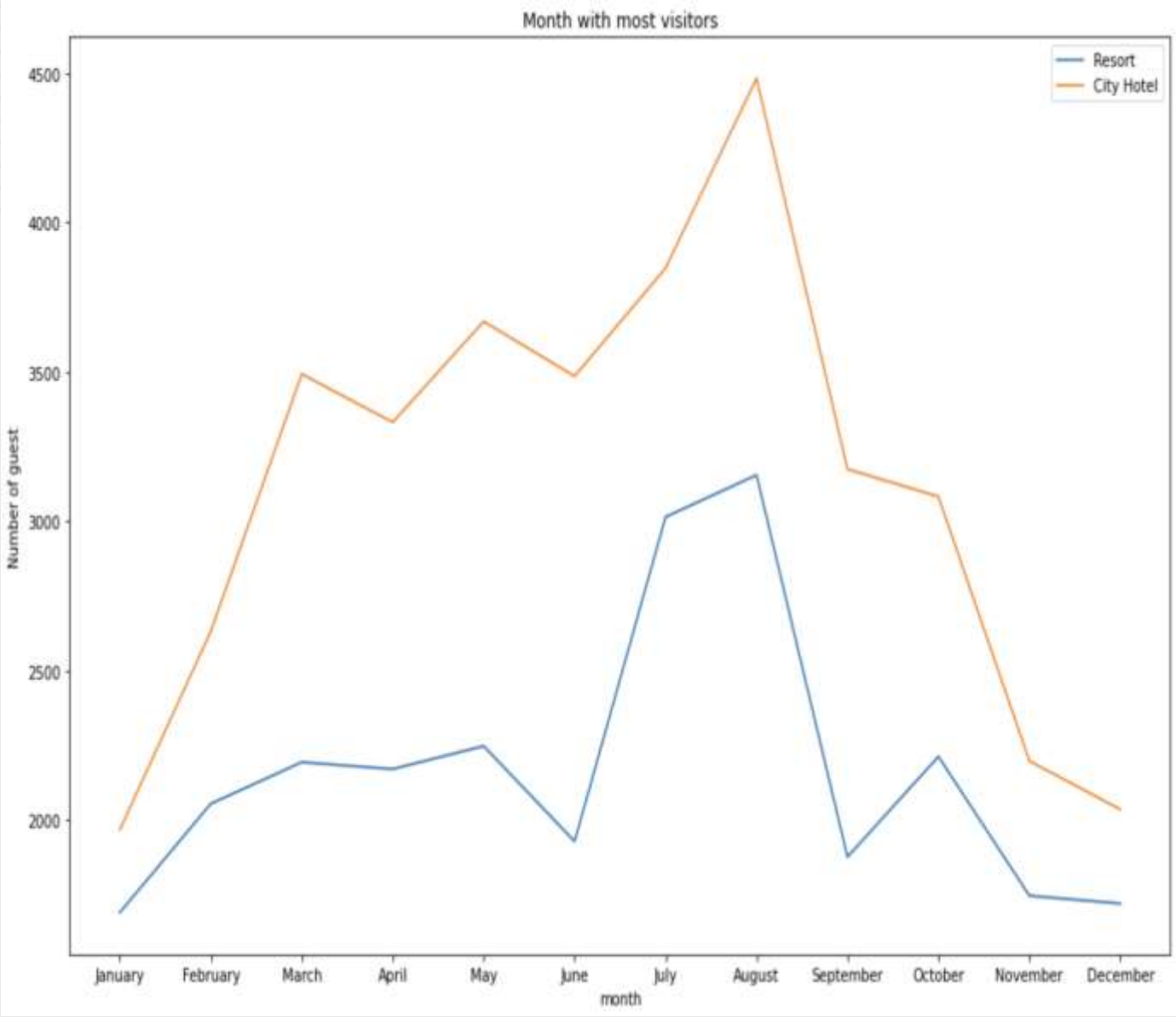
AGENT MAKING MOST NUMBERS OF BOOKINGS



INFERENCE:

Agent no. 9 has made most no. of bookings.

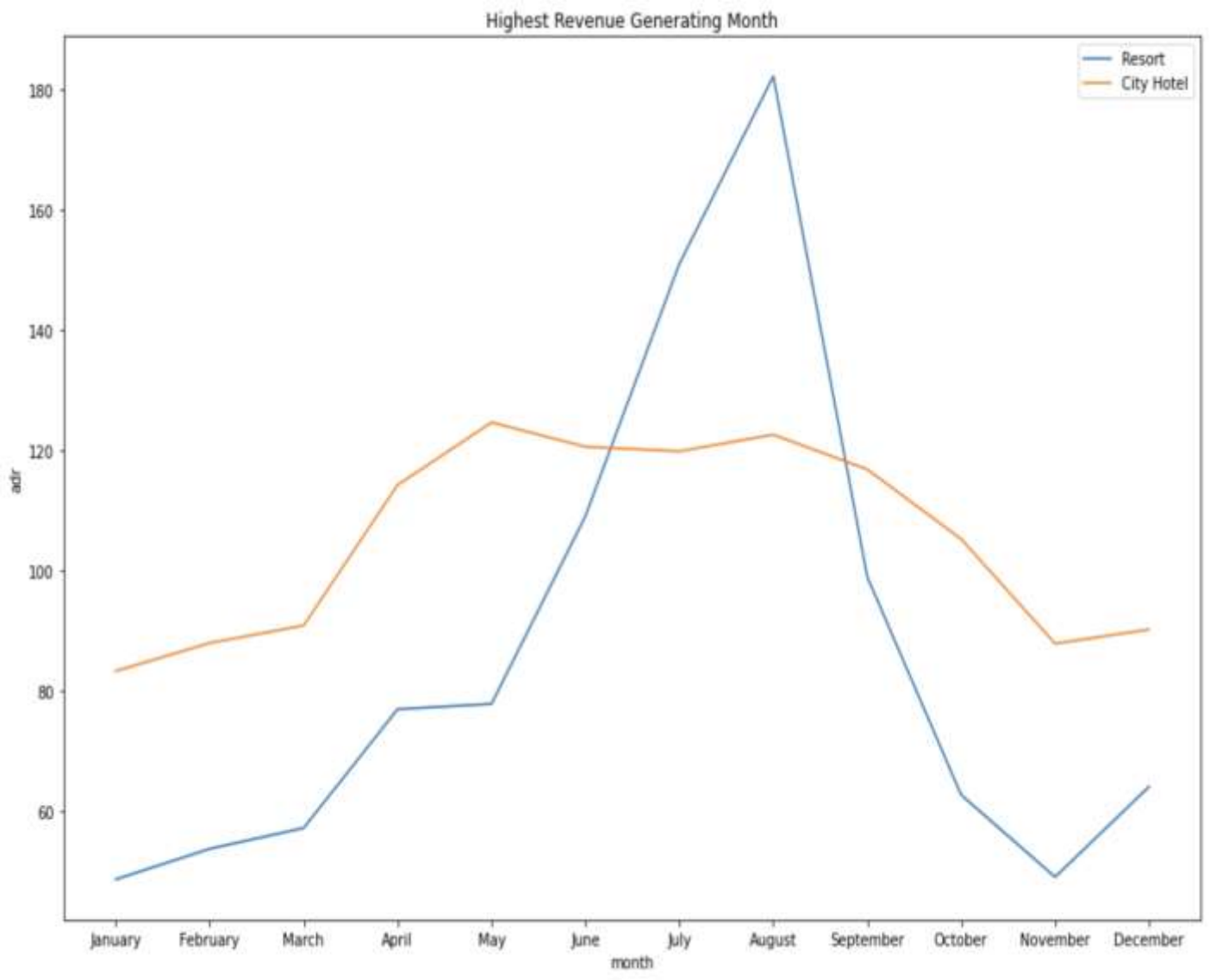
THE MONTH WITH MOST VISITORS



INFERENCE:

The month of august welcomes the m ost number of guests.

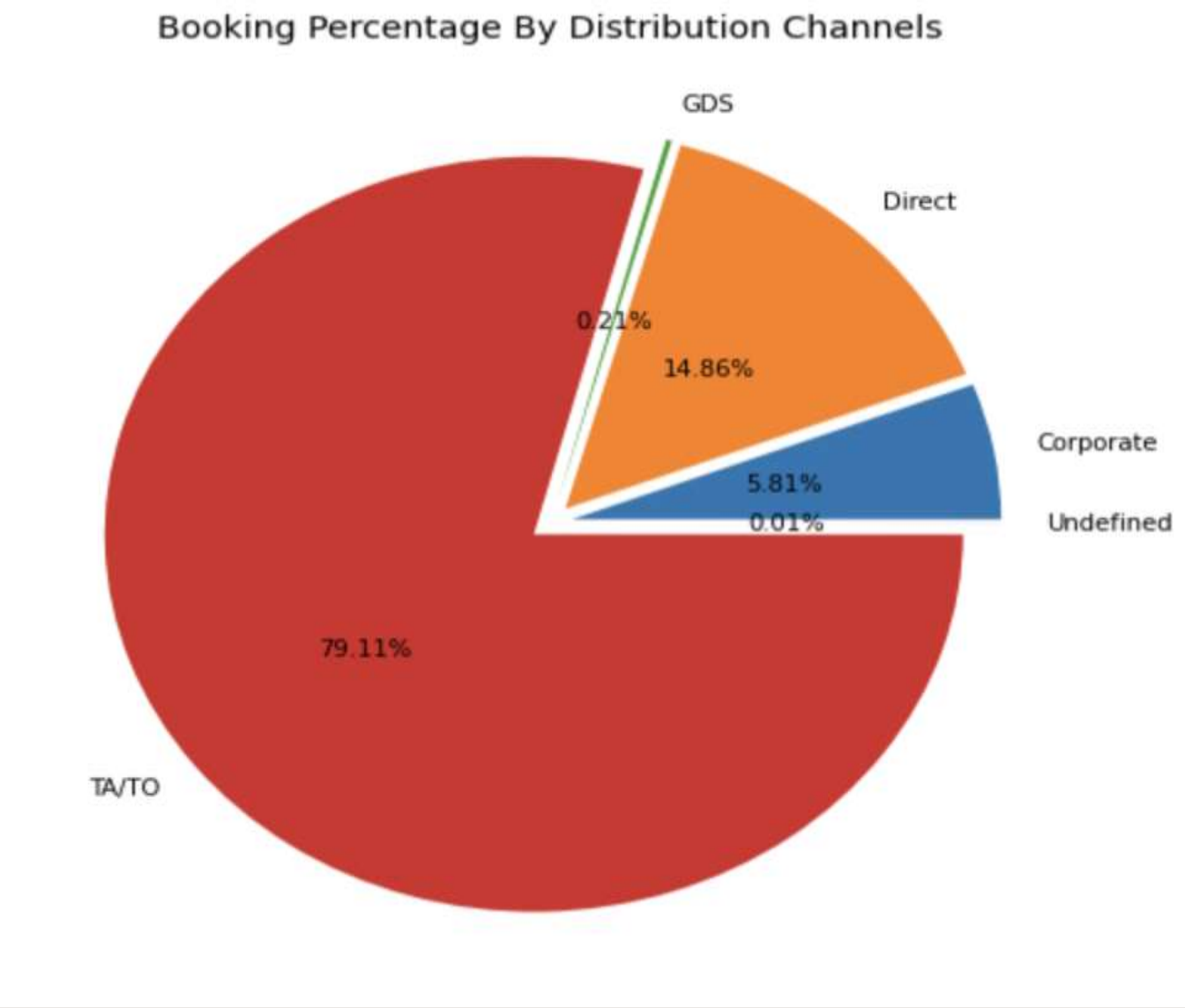
HIGHEST REVENUE GENERATING MONTH



INFERENCE:

Resort in the month of August generates highest revenue.

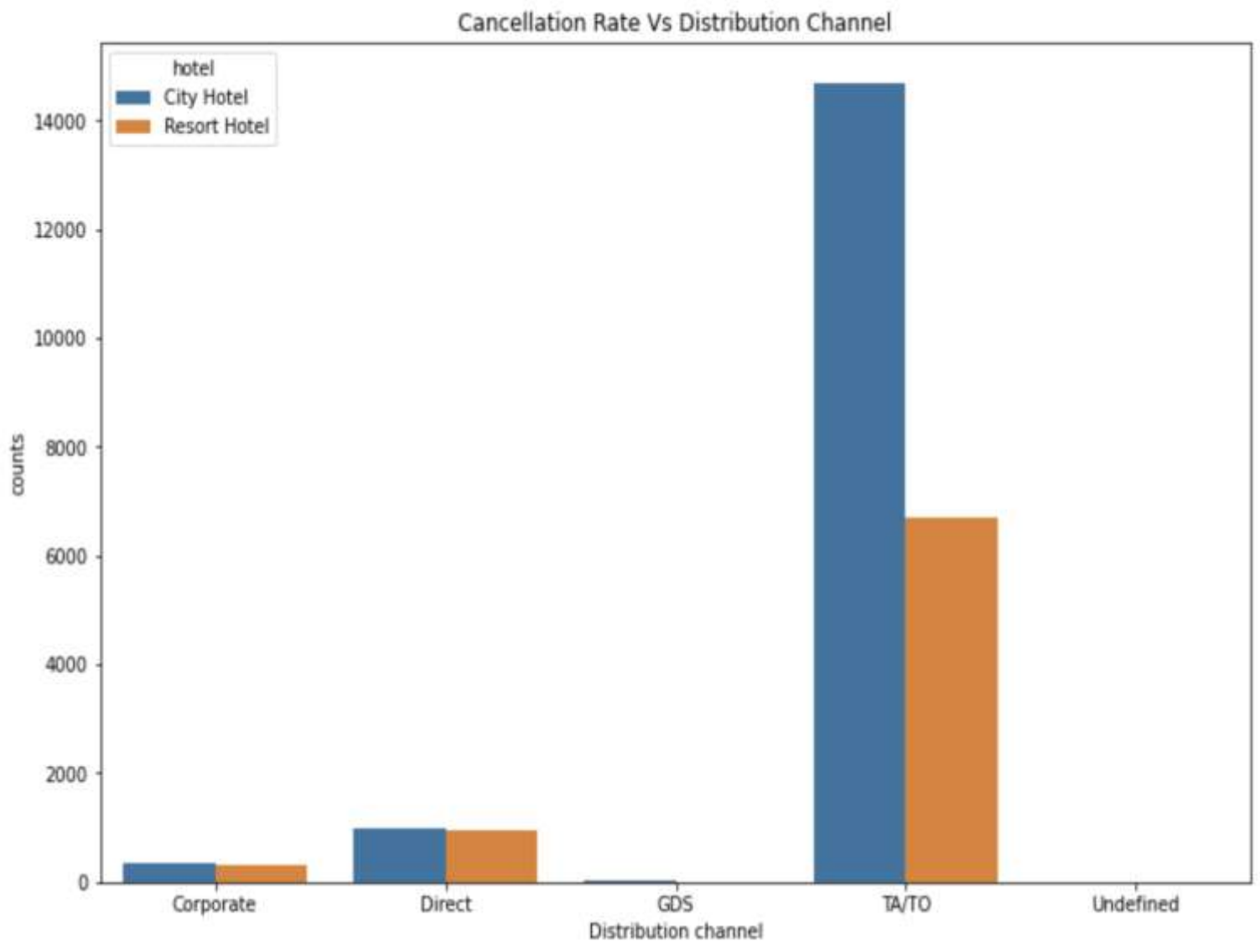
MOST COMMON CHANNEL FOR BOOKING HOTEL



INFERENCE:

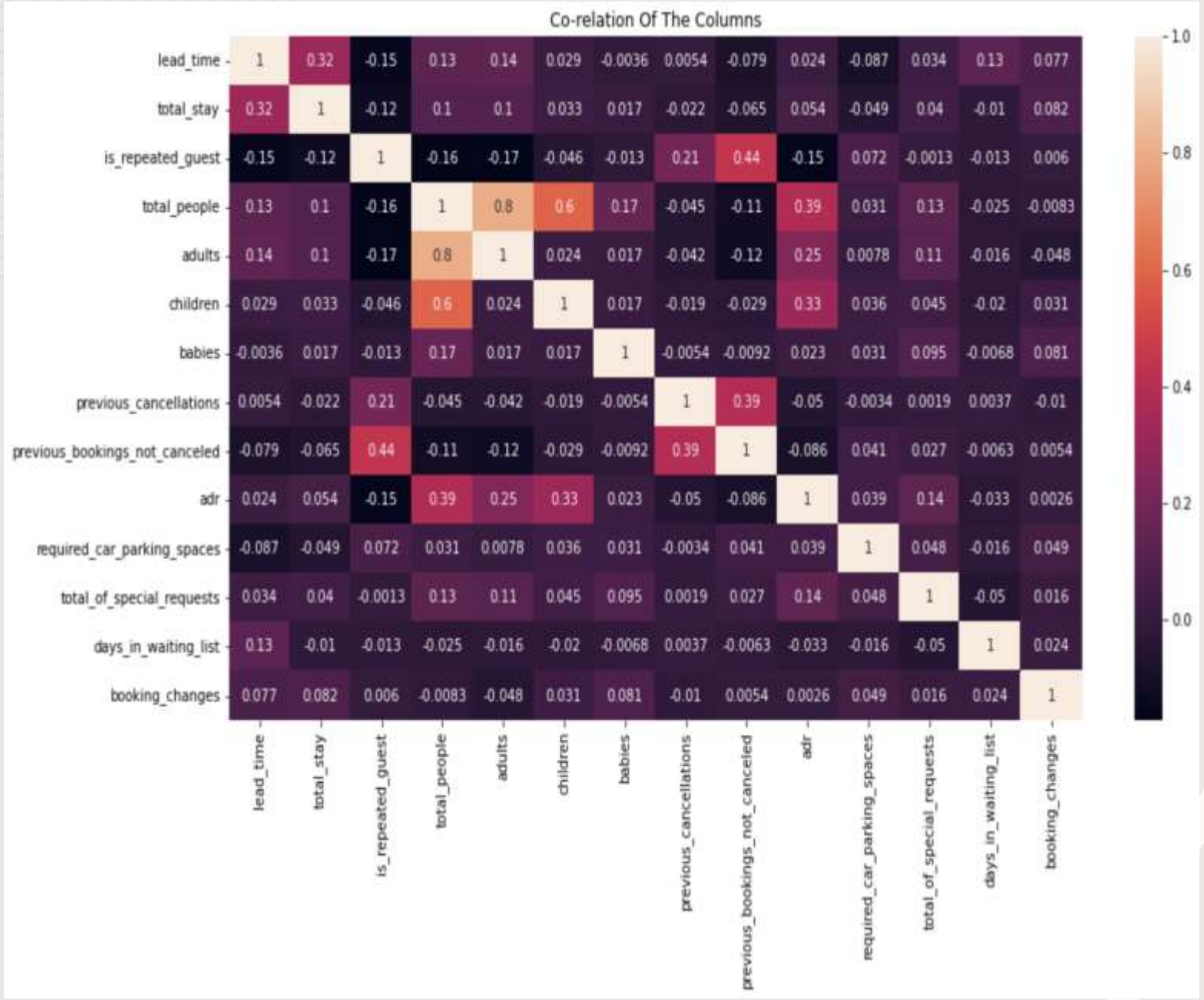
TA/TO is the most common channel fo r booking hotels.

DISTRIBUTION CHANNEL WITH HIGHEST CANCELLATION PERCENTAGE



INFERENCE:

TA/TO has highest booking cancellation percentage.Bookings made via TA/TO is 30% likely to get cancelled.



INFERENCE:

1.'total_people' and 'adult' are slightly positive correlated which means that most people staying in the hotel are a dults.

2.'total people' and 'adr' are slightly correalted which means more people more the adr.

3.'total_people' and 'children' are slig htly positive correlated which means t hat the most people staying in the hot el are also
children but is usually less than adults .

4.'is_repeated_guest' and 'previous_b ookings_not_canceled' are positively correlated which means that the repe ated guests
are mostly those who have not cancel led their bookings earlier.

CONCLUSION

- Majority of the hotels booked are city hotel. Definitely need to spend the most targeting fund on those hotel.
- We can also say that the high rate of cancellations can be due to no deposit policies.
- We should also target months between May to Aug. Those are peak months due to the summer period.
- TA/TO should generate better deals for customers as it is the distribution channel with highest booking cancellation.

THANK YOU