Capstone Project

Exploratory Data Analysis
Hotel booking

Sushmita Sardar
Sanjay Khatri
Vineeth Kumar .V
Lalan Jaiswal
Aman Jain

AI

Points for discussion:

Discussing the analysis of given data for hotel booking .

Doing analysis of the given data in multiple ways, that is .
1. Analysing and cleaning the given data.
2. Data Summary
3. Solving the questions using EDA
    a) Describing the analysis with graphs
    b) Observation
4. Correlation
5. Conclusion

Data Cleaning

Before doing the EDA we need to clean our data so that we can get correct outcome after doing EDA.

While cleaning the data we will follow fews steps:-

- Removing the null and NaN values.

- Dropping the duplicate rows.

Data summary

Hotel : Resort Hotel / City Hotel

is_canceled : if the booking was canceled- 1 / not - 0

lead_time : No of days 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 no of year for arrival date

arrival_date_day_of_month : Day of arrival date

stays_in_weekend_nights : No of weekend nights (Saturday / Sunday) the guest stayed / booked to stay at the hotel

stays_in_week_nights : No of week nights (Monday to Friday) the guest stayed / booked to stay at the hotel.

adults : No of adults

children : No of children

babies : No of babies

meal : Type of meal booked

country : Country of origin

market_segment : Market segment designation.
 In categories, "TA" - "Travel Agents" and "TO" - "Tour Operators" etc.

distribution_channel : Booking distribution channel. "TA" - "Travel Agents" and "TO" - "Tour Operators" corporate, direct etc.

is_repeated_guest : tells whether the booking name was repeated guest -1 / not- 0

previous_cancellations : No of previous bookings that were cancelled by customer prior to the current booking

previous_bookings_not_canceled : No of previous bookings not cancelled by customer prior to the current booking

reserved_room_type : Code of room type reserved.

assigned_room_type : Code for the type of room assigned.

booking_changes : No of changes made between booking till moment of check-in or cancellation.

deposit_type: Tells customer made a deposit to guarantee the booking

agent : ID of the travel agency that made the booking

company : ID of the company that made the booking

days_in_waiting_list : No of days the booking was in the waiting list before it got confirmed to customer

Analysing the given data we have solved following questions.

- ❑ Which hotel have the higher number of bookings?
- ❑ What is the percentage of cancellation.
- ❑ Calculate the ADR with respect to distribution channel.
- ❑ Find out the booking trends on total stay by the customers.
- ❑ What was the average ADR between hotels.
- ❑ Which is the most preferred room types by the customers.
- ❑ Market segment that has the highest cancellation rate.
- ❑ Which meal type is most preferred by the customers.
- ❑ Which hotels makes more revenue.
- ❑ The percentage of car space is required.
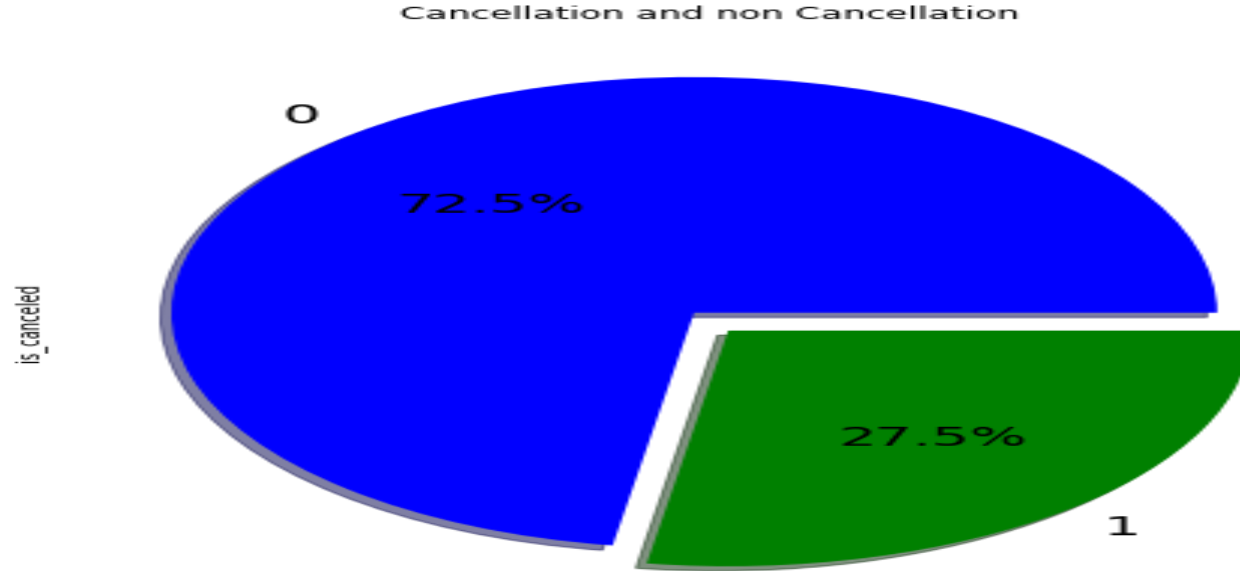
Which hotel have the higher number of bookings? **AI**
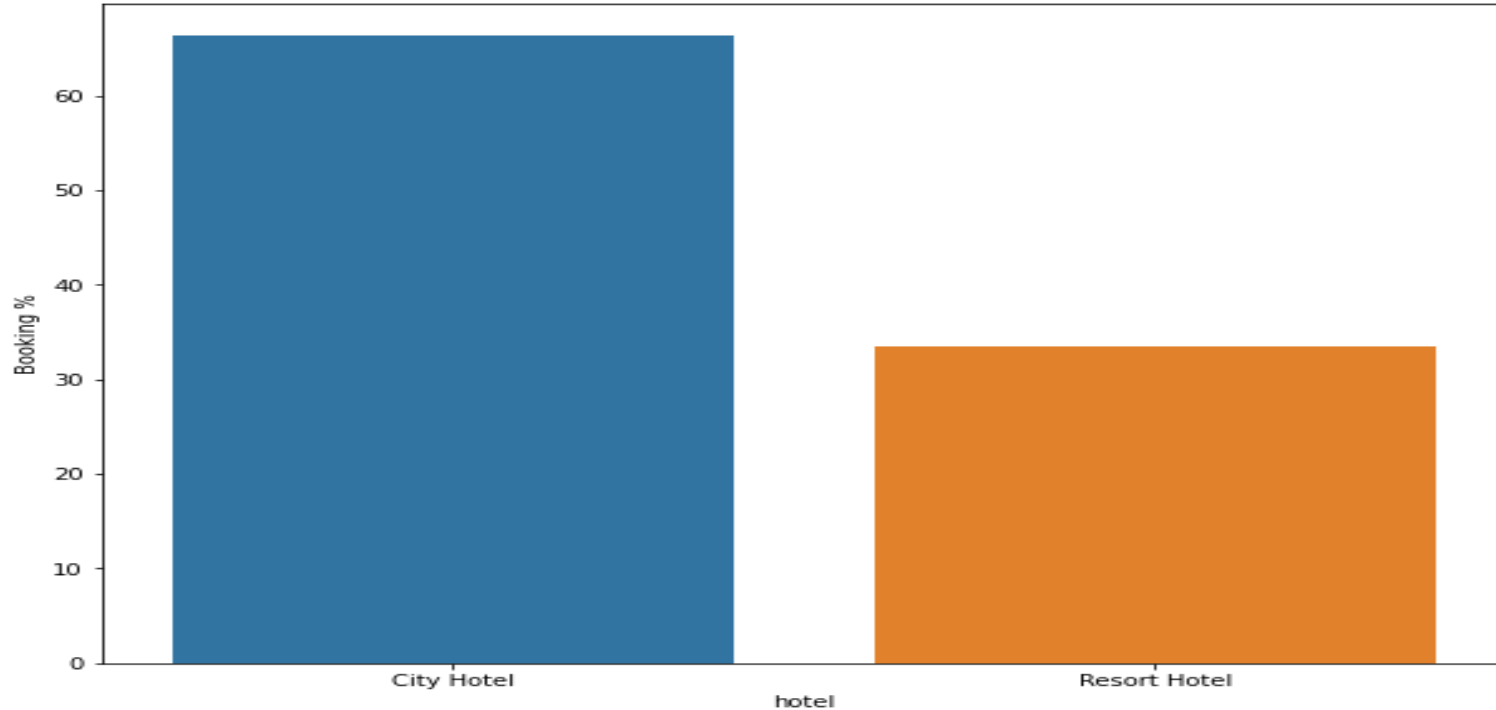
Observation:
•From above graph we can see that city hotel have maximum number of booking than Resort hotel

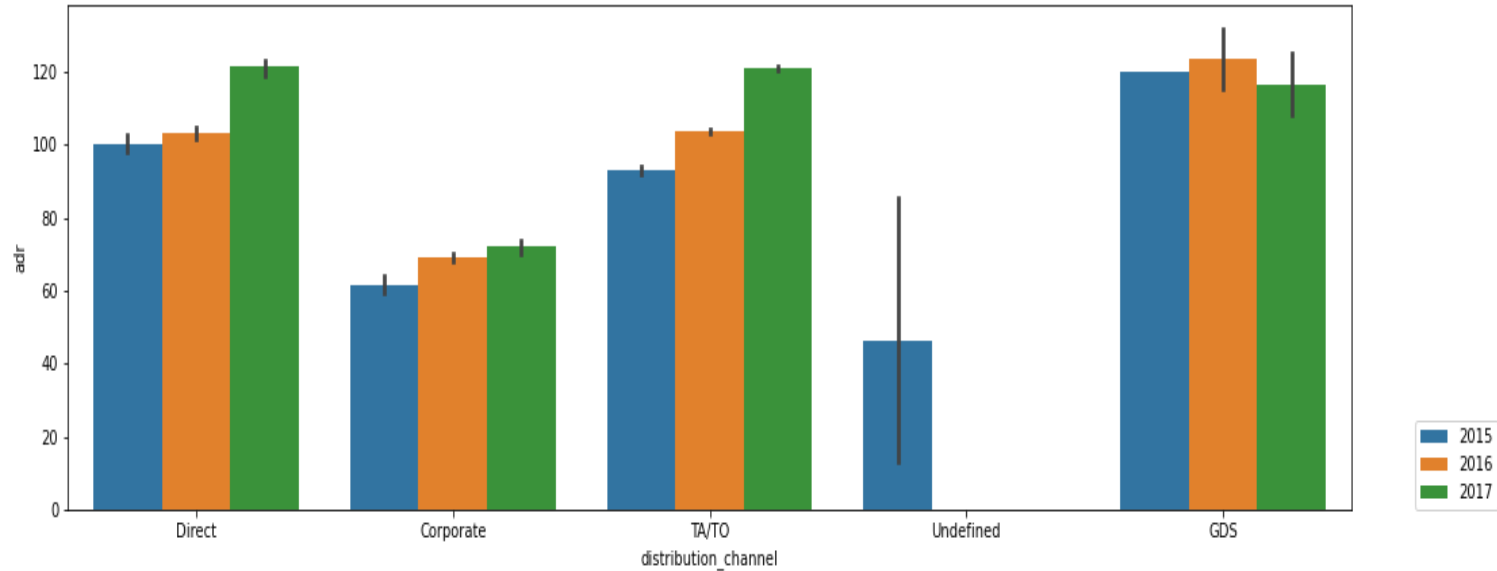# Percentage of Booking cancellation.

Cancellation and non Cancellation

Observation:
From the above piechart we see that:-
0 = Non cancelled 1 = cancelled
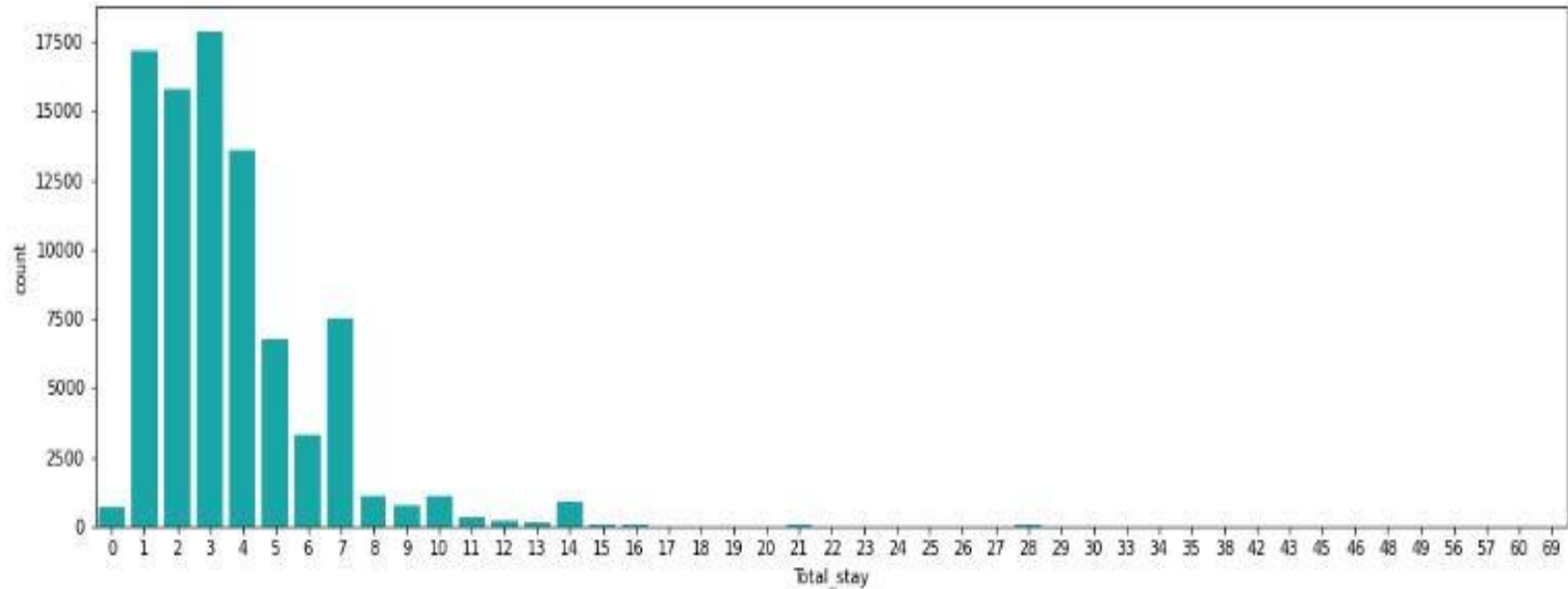27.5 % is cancelled

Calculate the ADR with respect to distribution channel?
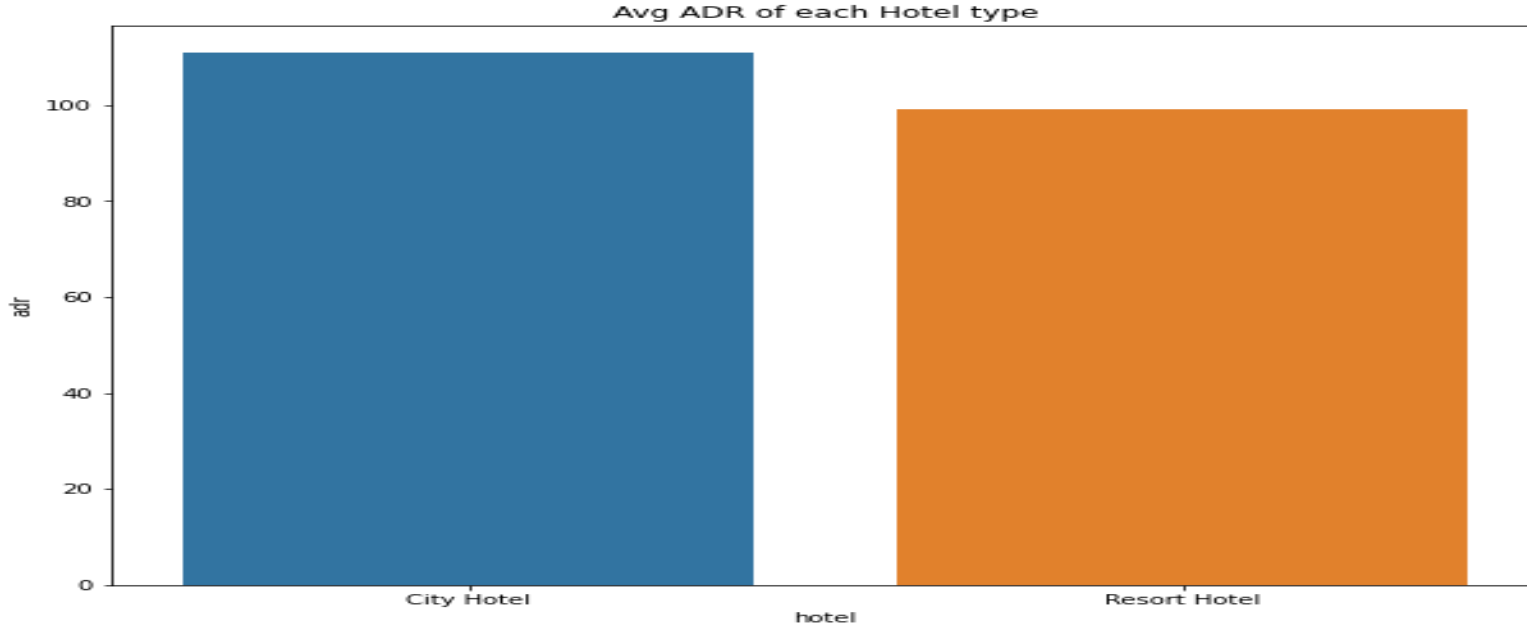

Observation:
- Max ADR comes from GDS(overall) except for year 2017.
- Min ADR comes from Corporate and undefined channel in the 3 years

# Find out the booking trends on total stays by the customers?

Most of the people stayed less than 5 days in the hotel and only few people stayed beyond 10 days.

Average ADR between hotels.

Avg ADR of each Hotel type

Observation:
City hotel has the highest ADR. That means city hotels are generating more re
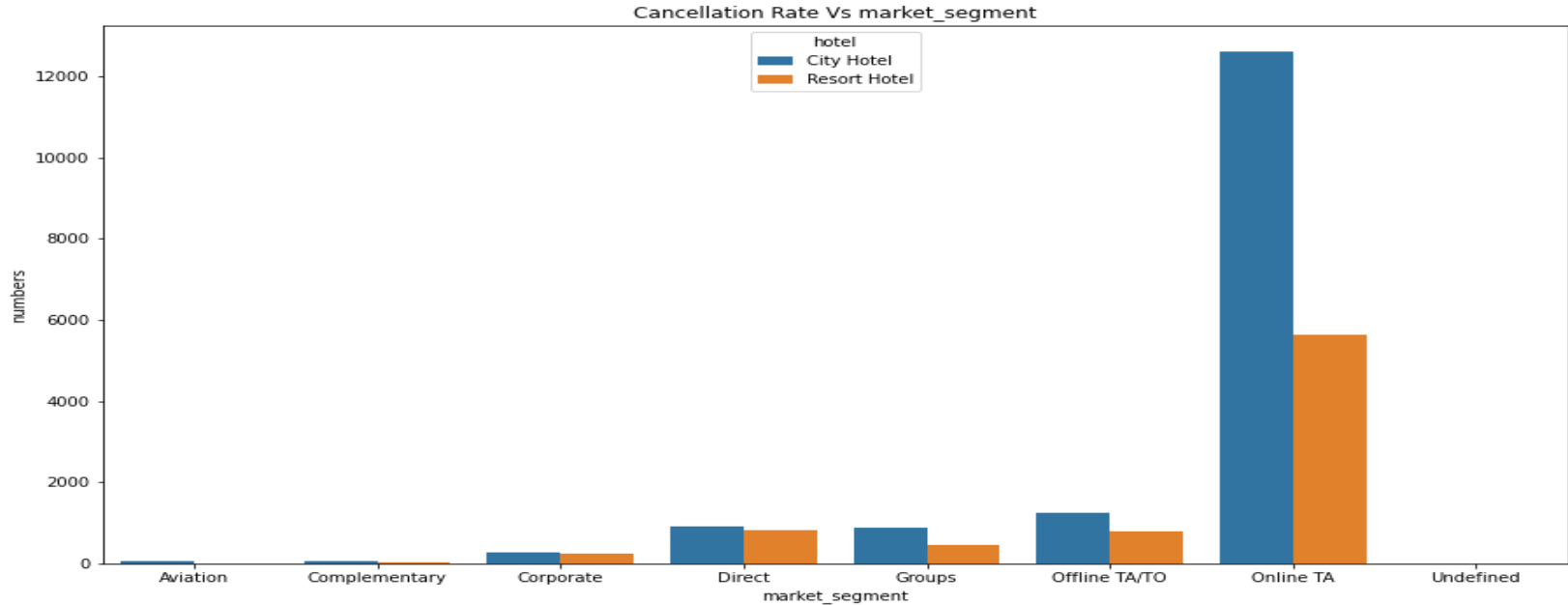venues than the resort hotels. More the ADR More is the revenue.

The most preferred room types by the customers.
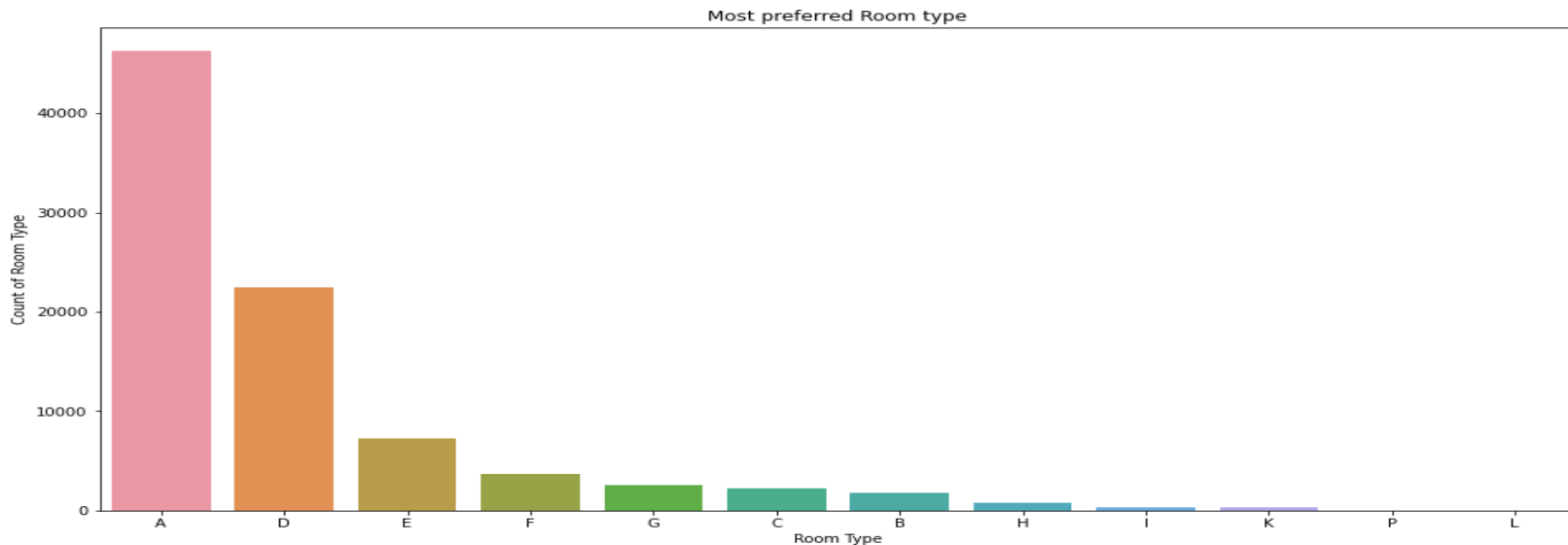
Obervation:
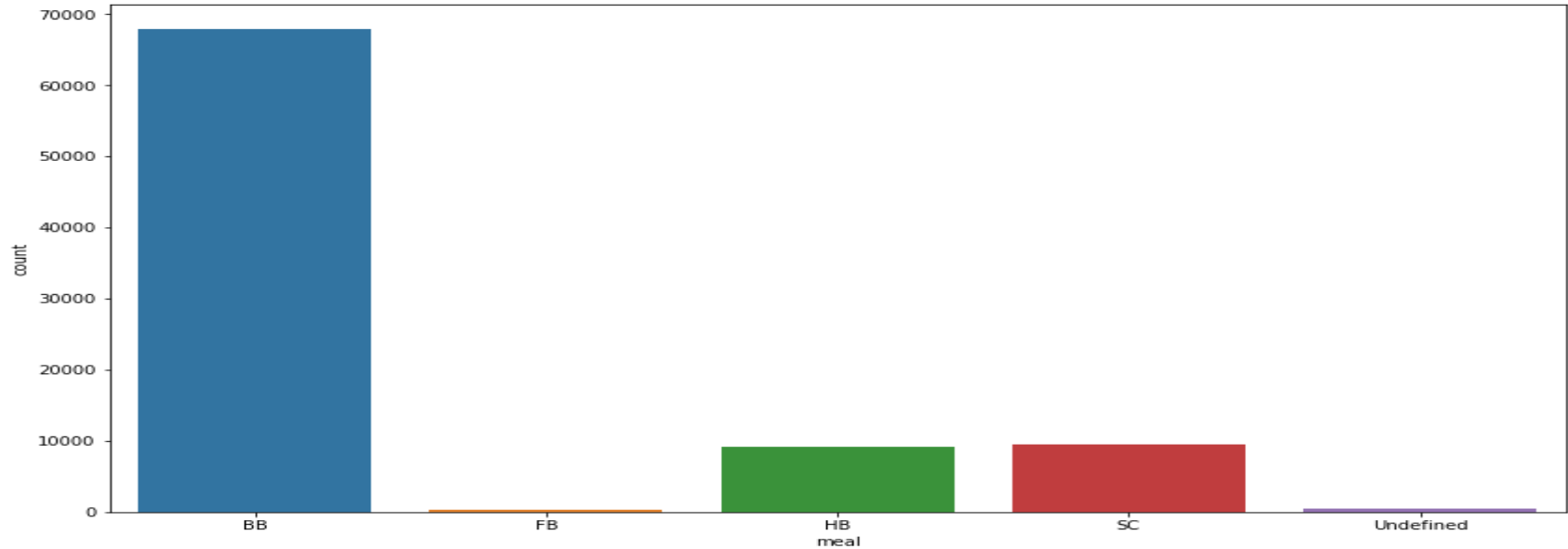The most preferred Room type is "A".

Market segment that has the highest cancellation rate

Cancellation Rate Vs market_segment

Observation:
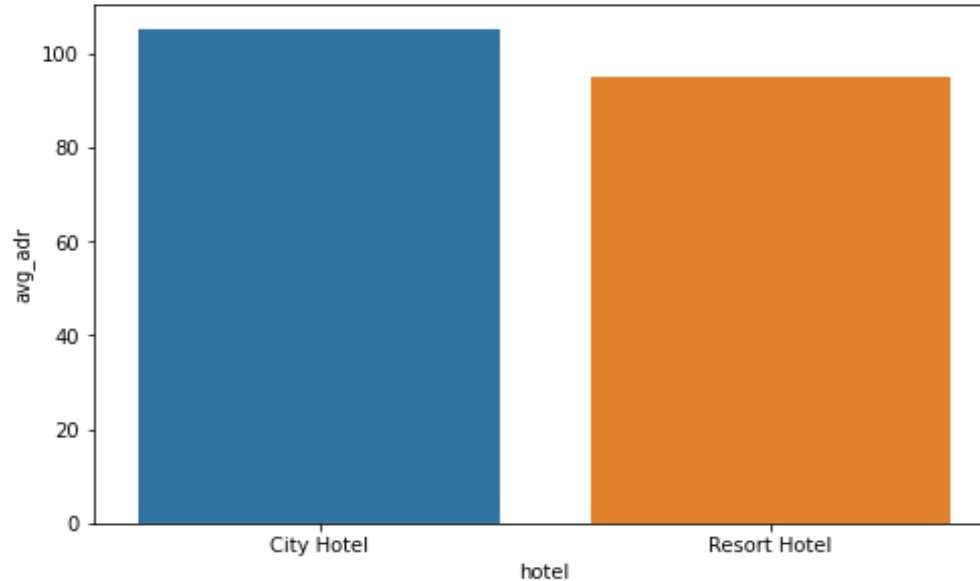•Online TA/TO has highest cancellation rate

Meal type that is most preferred by the customers.

Observation:
•BB is the most preferred meal by the customers
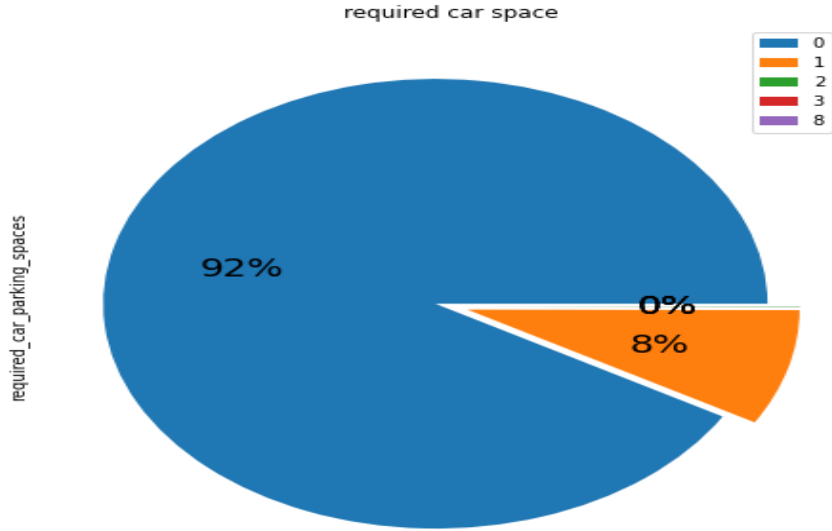
The Hotels that makes more revenue.
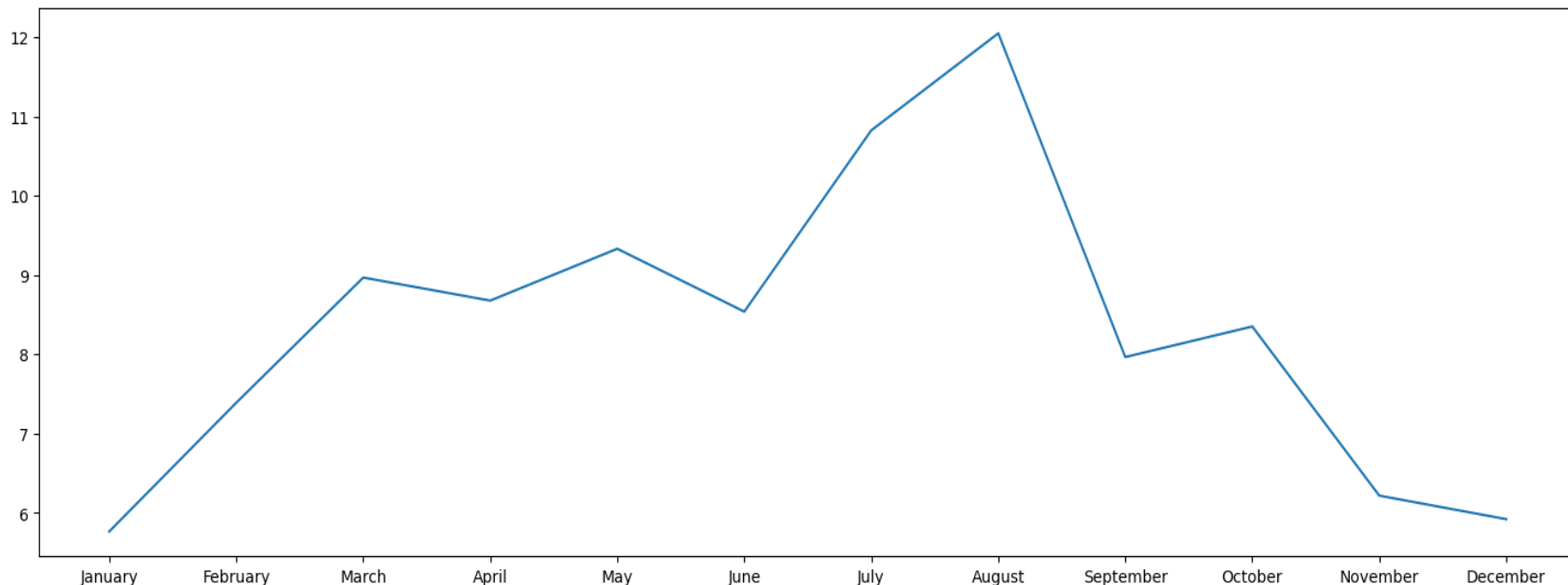

Observation:
•City hotel makes more revenue

The percentage of car space is required.

required car space

92%

0%

8%

Observation:
92 % guests did not required the parking space. only 8 % guests require
d only 1 parking space.

Busiest booking month of a calendar year

Observation:

•Max Booking are done from the period of July to Mid-August in a calendar Year

•Min Booking are done from the period of January - February and November - December

•Booking are low in the start and end of a calendar year and it gradually increase and peaks at middle of a calendar year and then drops down.

Corelation of columns

Co-relation of the columns
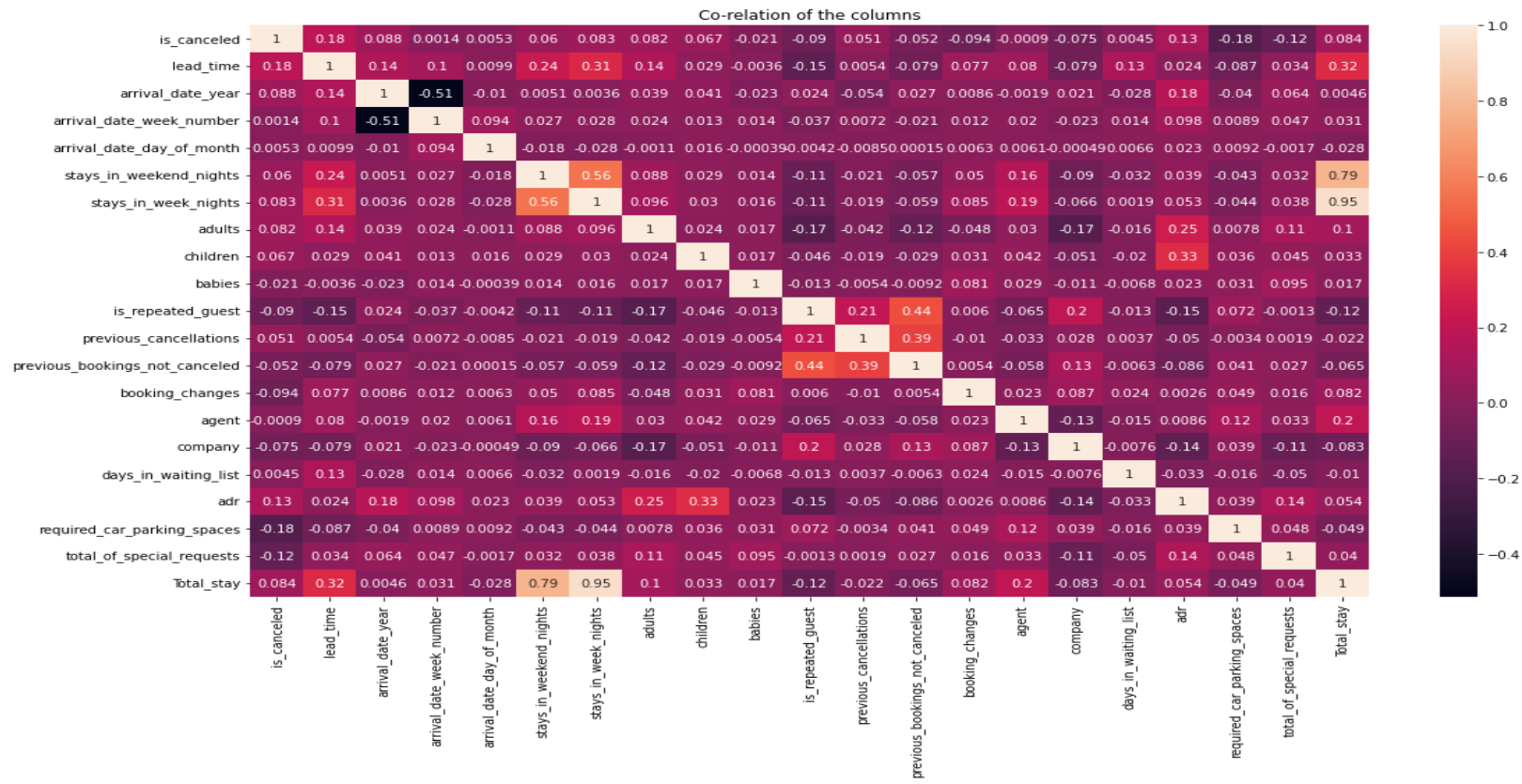
Observation

- lead_time and total_stay is positively corelated.it means stay of cutsomer is proportional to lead time.
-
  is_canceled and same_room_alloted_or_not are negatively corelated. That means customer is unlikely to cancel his bookings if he don't get the same room as per reserved room.
-
  Adults,childrens and babies are corelated to each other. more customer more the adr.

- Total stay length and lead time have slight correlation. This may means that for longer hotel stays people generally plan little before the the actual arrival.

Conclusion

AI

- Maximum guests are booking city hotel over resort hotel.
- The percentage of cancellation is 27.5% and the percentage of non-cancellation is 72.5%.
- Since GDS gives better ADR, it would be the best among distribution channel as it gives higher revenue even with lower occupancy.
- Most of the people stayed less than 5 days in the hotel.
- City hotel has the highest ADR. That means city hotels are generating more revenues than the resort hotels. More the ADR More is the revenue.
- After analysing we see that the most preferred Room type is "A".
- Online TA/TO has highest cancellation rate.
- BB is the most preferred meal by the guests.
- After analysing we see that city hotel makes more revenue than resort hotel.
- 91.6 % guests did not required the parking space.
- Max. Booking happens July to Mid-August. Marketing and booking offer provided in these months can yield higher booking.
- Lead time and total stay is positively corelated.it means stay of customer is proportional to lead time.
- Is canceled and same_room_alloted_or_not are negatively corelated. That means customer is unlikely to cancel his bookings if he don't get the same room as per reserved room.
- Adults,childrens and babies are corelated to each other. more customer more the adr.
- Total stay length and lead time have slight correlation. This may means that for longer hotel stays people generally plan little before the the actual arrival

Thank you