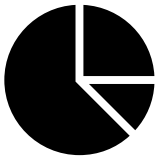
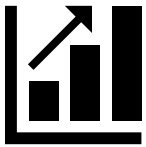
Capstone Project

Hotel Bookings Data Analysis

By-Aakash Yadav
Cohort-Seattle

Lets Analyse the hotel data

1. Defining Problem Statement.

2. Creating Questions for Solving the problem statement.

3. Cleaning the Data and getting the basic details about our Data.

4. Doing Exploratory Data Analysis for getting the Answers our questions and making final observations.

Steps:-

1. Data Processing:- In this step we had observed the data and for cleaning the data, removed the unnecessary features and handled the null values.

2. EDA- In this part, we do some exploratory data analysis(EDA) on the features selected in step 1 to see the trend.

3. Final Observations- After observing all the observation made during answering our questions and we had made a final observation.

Data Description

DataFrame contains the following features(Column)

1. **hotel** : Hotel(Resort Hotel or City Hotel)
2. **is_canceled** : Value indicating if the booking was canceled (1) or not (0)
3. **lead_time :** Number of days that elapsed between the entering date of the booking into the PMS and the arrival date
4. **arrival_date_year :** Year of arrival date
5. **arrival_date_month :** Month of arrival date
6. **arrival_date_week_number :** Week number of year for arrival date
7. **arrival_date_day_of_month :** Day of arrival date
8. **stays_in_weekend_nights :** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
9. **stays_in_week_nights :** Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
10. **adults :** Number of adults, **children :** Number of children and **babies :** Number of babies
11. **meal :** Type of meal booked. Categories are presented in standard hospitality meal packages:
12. **country :** Country of origin.
13. **market_segment :** Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”

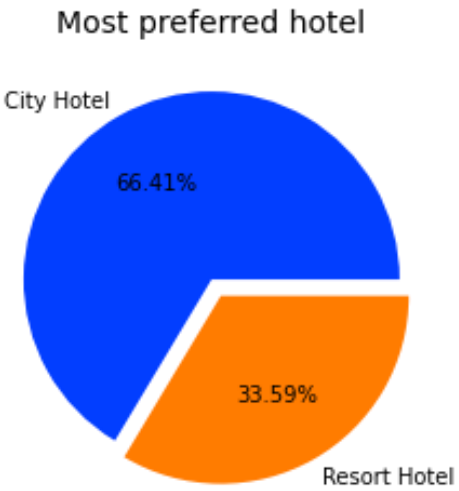
Data Description(Cont.)

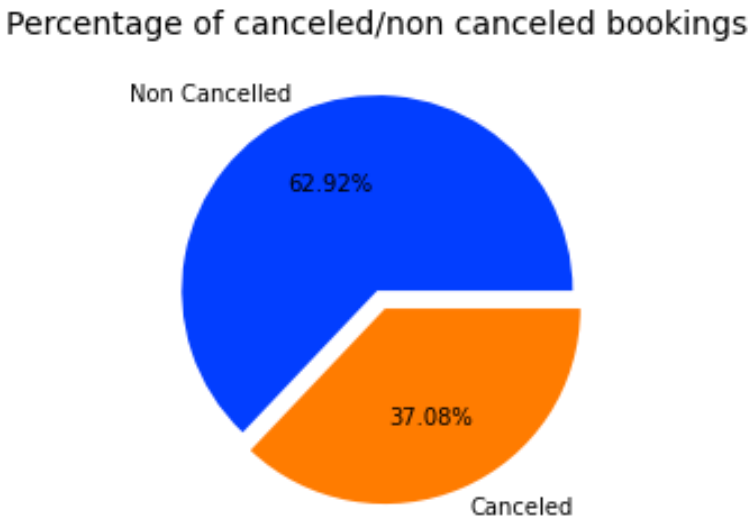
14. **distribution_channel :** Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”
15. **is_repeated_guest :** Value indicating if the booking name was from a repeated guest (1) or not (0)
16. **previous_cancellations :** Number of previous bookings that were cancelled by the customer prior to the current booking
17. **previous_bookings_not_canceled :** Number of previous bookings not cancelled by the customer prior to the current booking
18. **reserved_room_type :** Code of room type reserved. Code is presented instead of designation for anonymity reasons.
19. **assigned_room_type :** Code for the type of room assigned to the booking.
20. **booking_changes :** Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
21. **deposit_type :** Indication on if the customer made a deposit to guarantee the booking.
22. **agent :** ID of the travel agency that made the booking
23. **company :** ID of the company/entity that made the booking or responsible for paying the booking.

Data Description (Cont.)

24. **days_in_waiting_list :** Number of days the booking was in the waiting list before it was confirmed to the customer
25. **customer_type :** Type of booking, assuming one of four categories
26. **adr :** Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
27. **required_car_parking_spaces :** Number of car parking spaces required by the customer
28. **total_of_special_requests :** Number of special requests made by the customer (e.g. twin bed or high floor)
29. **reservation_status :** Reservation last status, assuming one of three categories
    Canceled : booking was canceled by the customer
    Check-Out : customer has checked in but already departed
    No-Show : customer did not check-in and did inform the hotel of the reason why
24. **reservation_status_date :** Date at which the last status was set. This variable can be used in conjunction with the Reservation Status to understand when was the booking canceled or when did the customer checked-out of the hotel

EDA & Observations

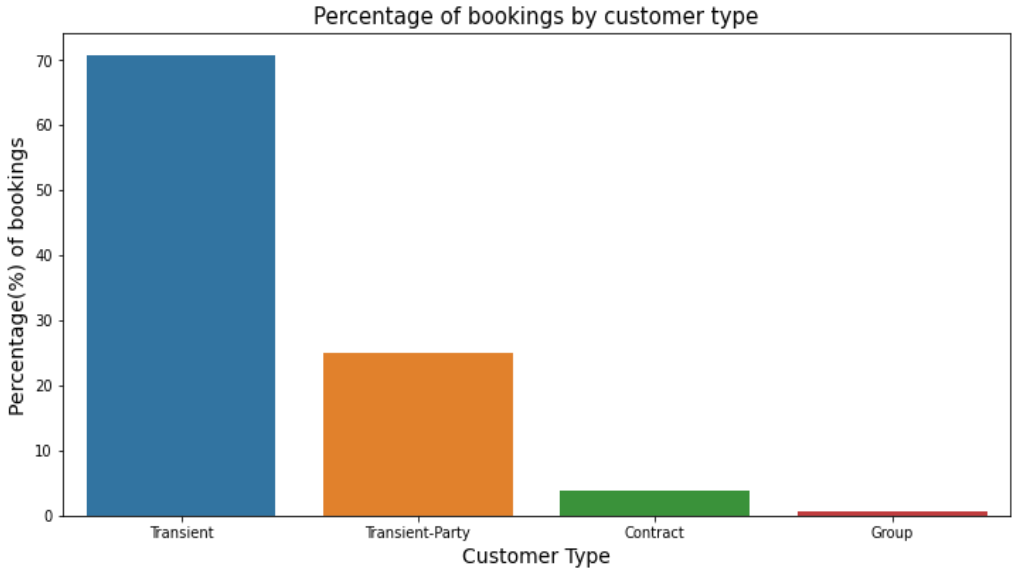


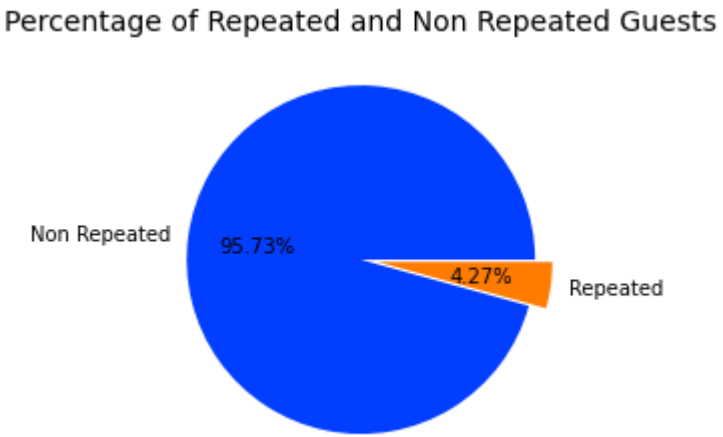


Observation: People prefer to book City Hotels more as compare to Resort Hotels.

Observation: 37.08% bookings got canceled by the guests.

EDA & Observations (Cont.)



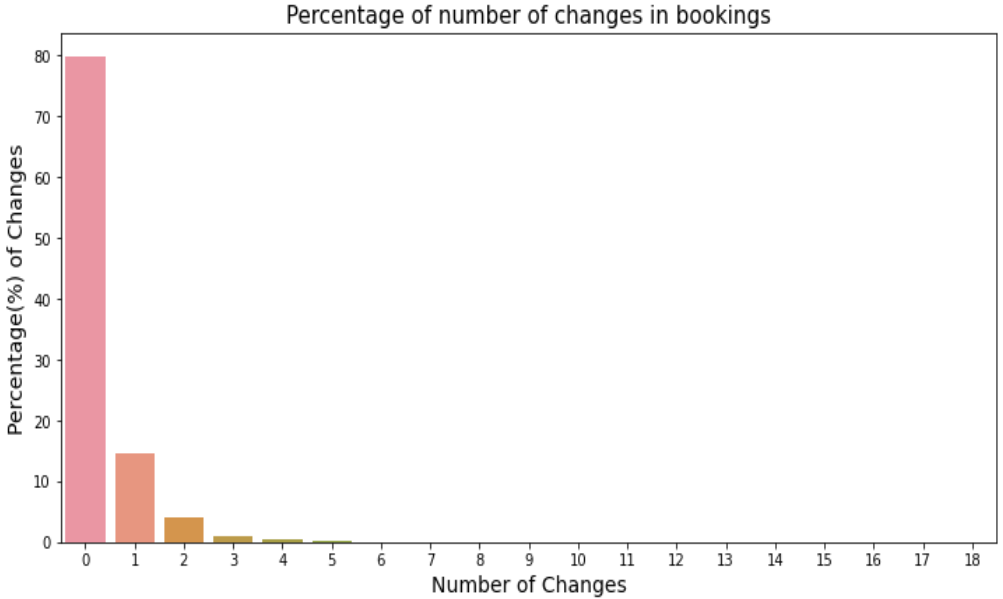


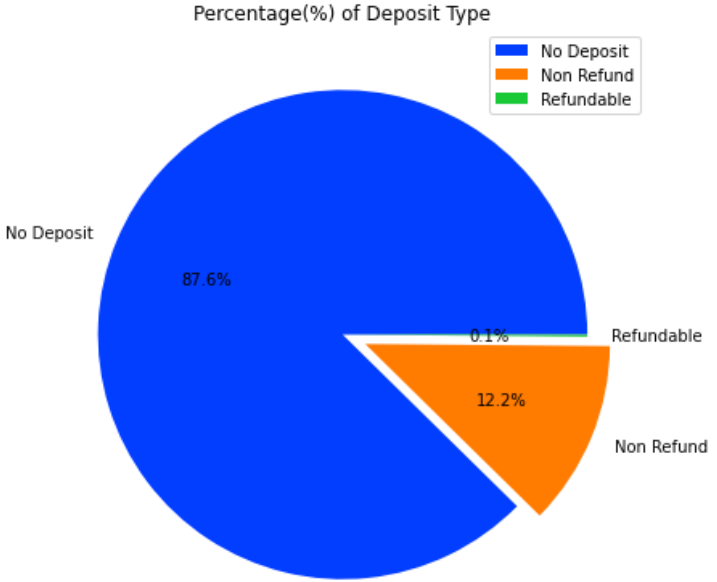
Observations:

•Booking by the Transient Customer is very high as compare to other type of customers(approx. 70% of all the bookings).

•Repeated guest are very few, which only 4.27%.

EDA & Observations (Cont.)

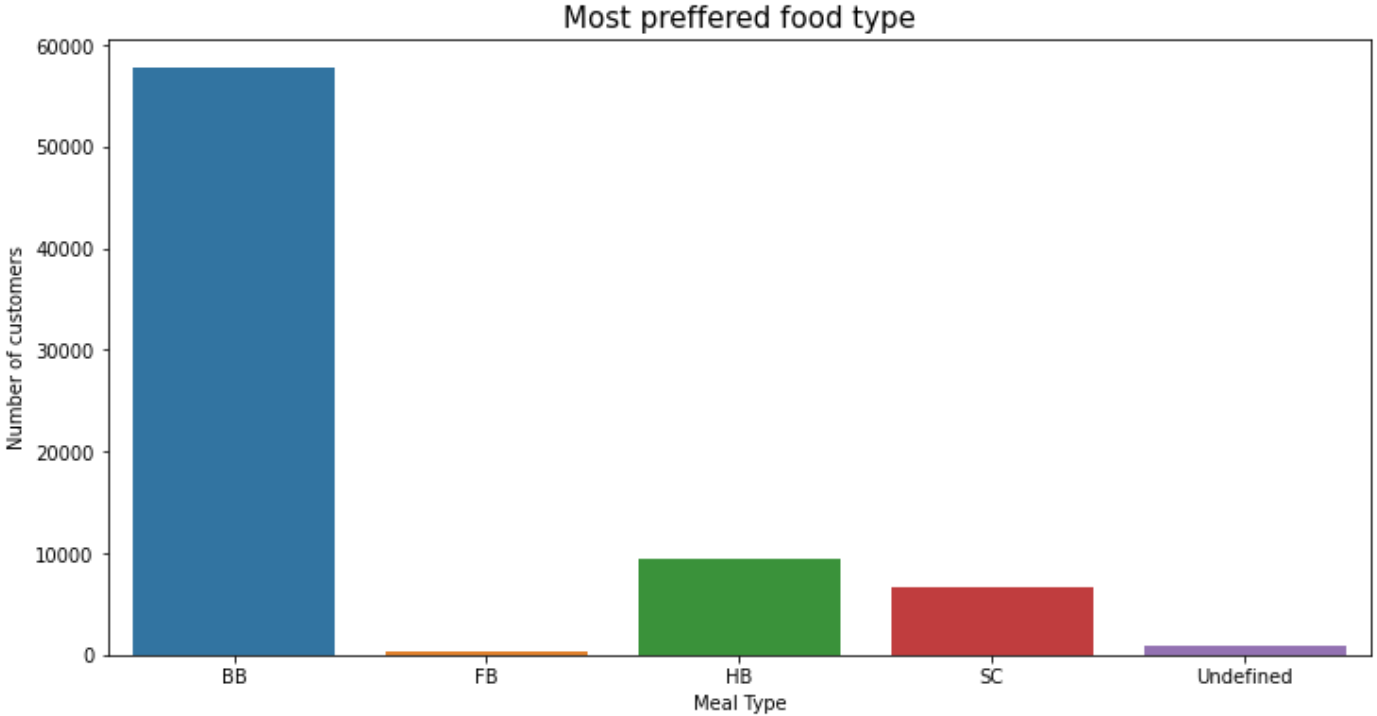




Observations:

- Almost 85% of the bookings were not changed by guests.
- Some time guests made 1 or 2 changes in their bookings. Otherwise, its very rare.
- 87.6% of guests prefer "No deposit" type of deposit.

EDA & Observations (Cont.)

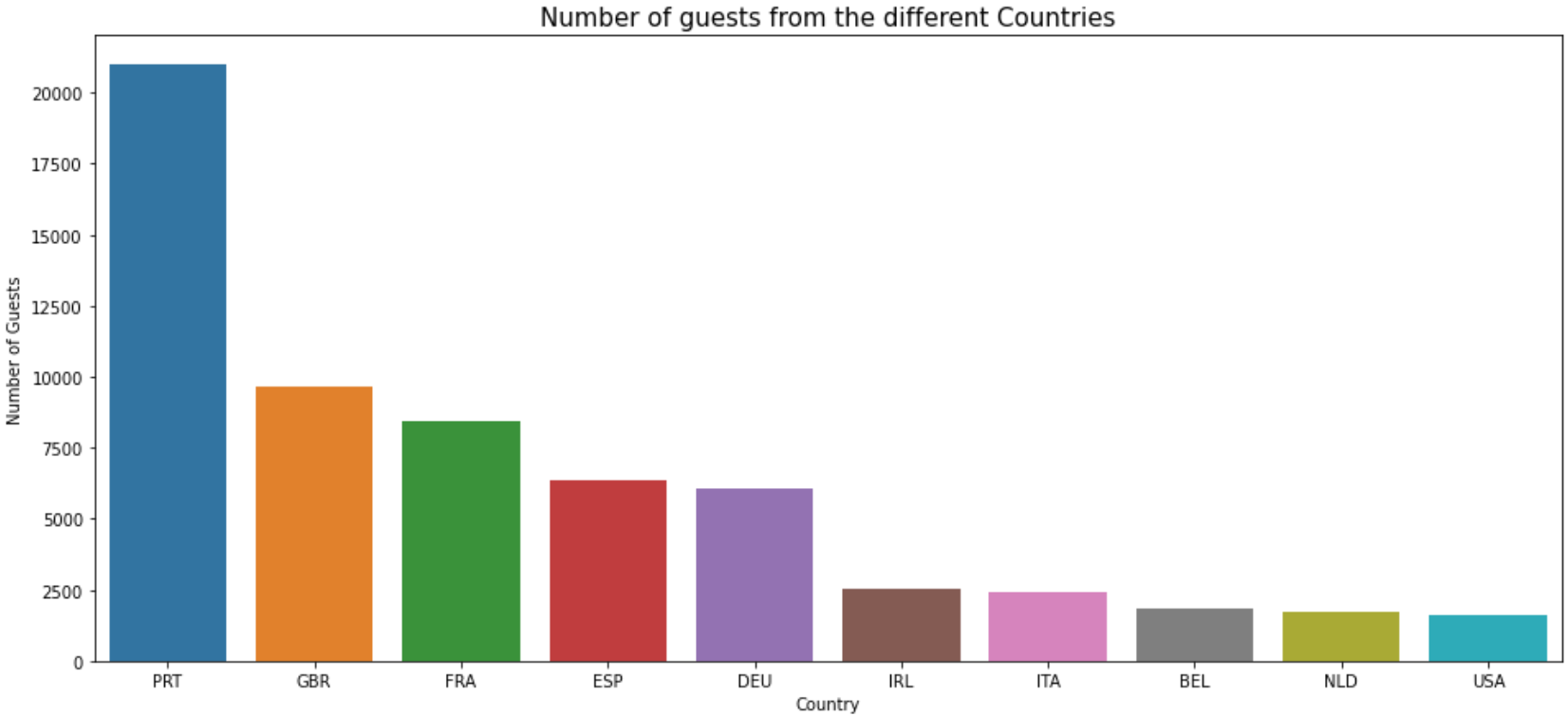


Observations:

•Most preferred food by the customers is BB(Bed and Breakfast).

•HB(Half Board) and SC(Self catering) are apprx. equally preferred.

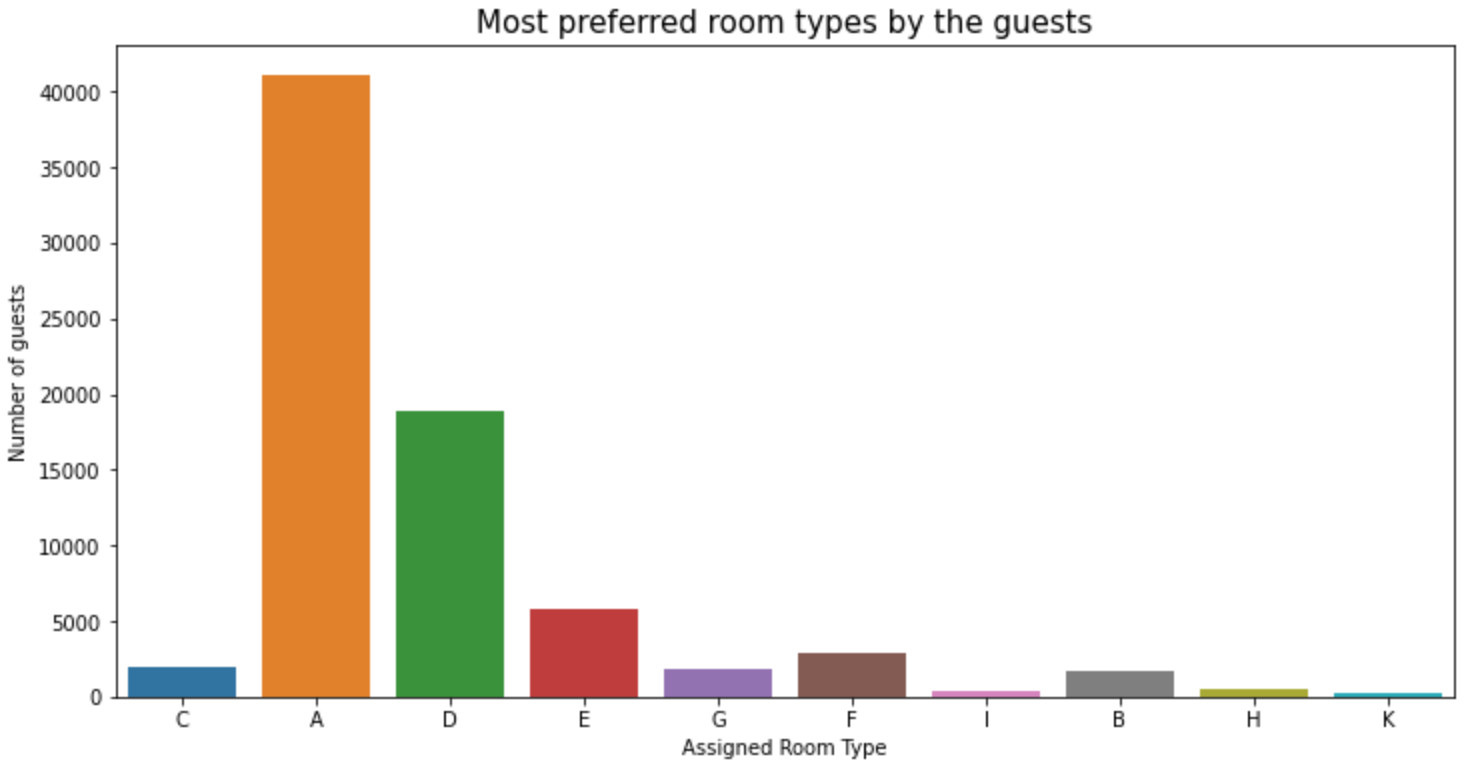
EDA & Observations (Cont.)



Observation:

•Most of the guests are coming from Portugal. More then 25000 guests are from Portugal.

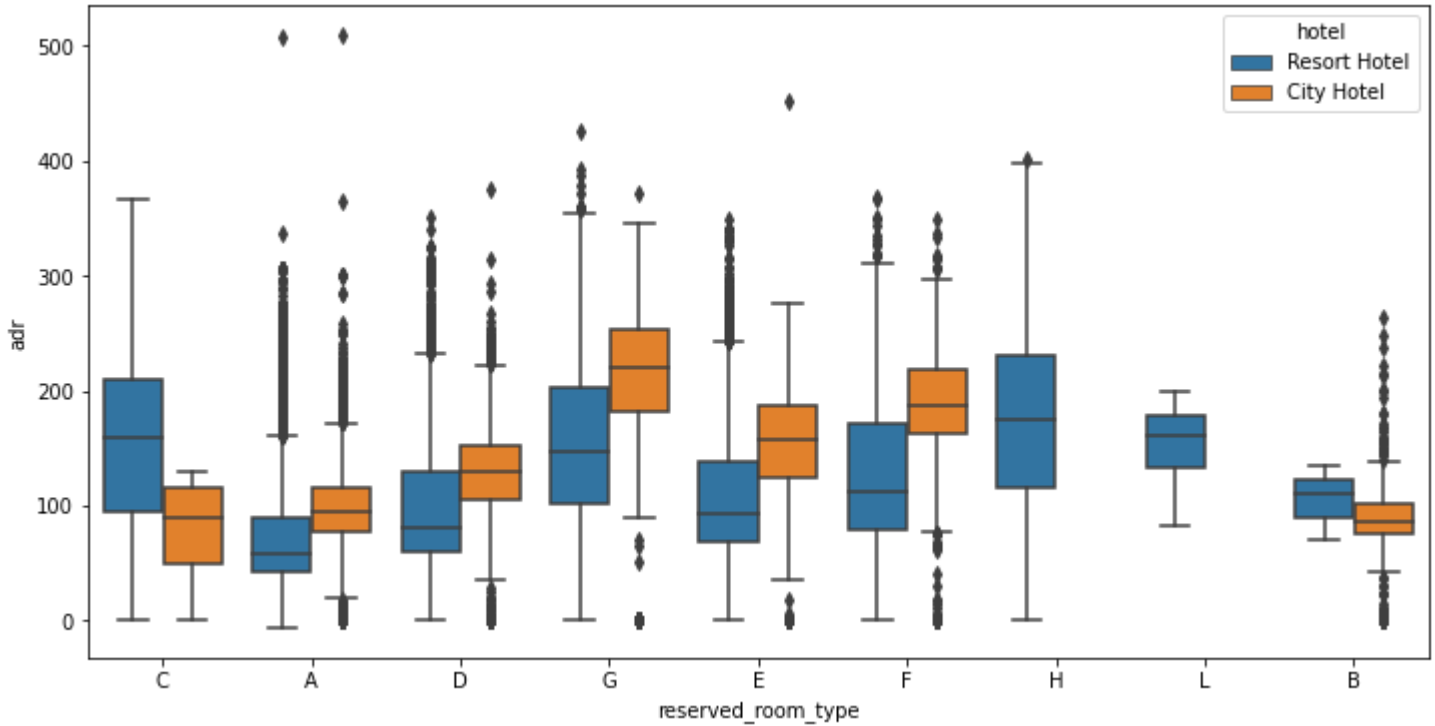
EDA & Observations (Cont.)



Observation:

•The most preferred room is A.

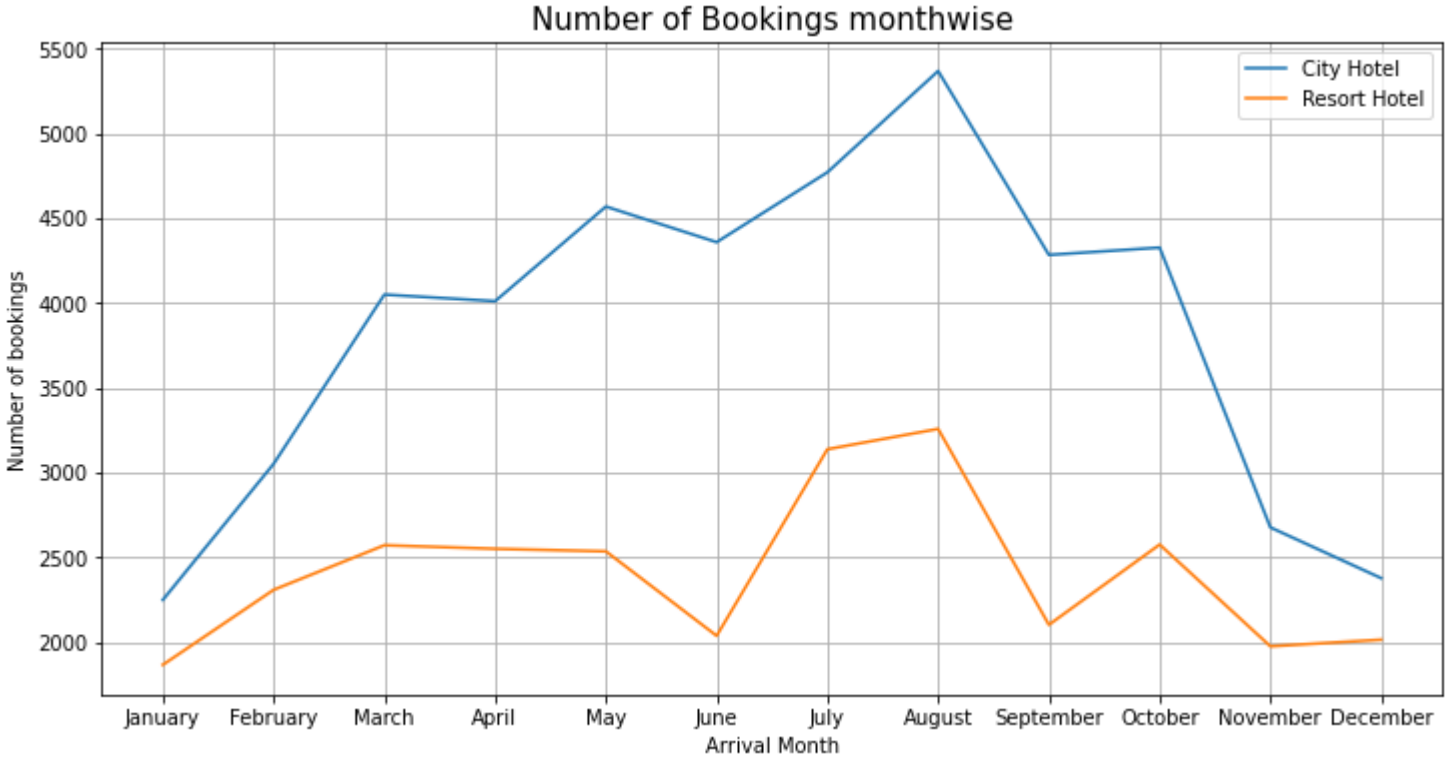
EDA & Observations (Cont.)



Observations

- The figure shows that the average price per room depends on its type.
- For Resorts G,H and C type of Rooms are most costly.
- For City Hotels, G type rooms are most costly. They are more expensive then the resort hotels.

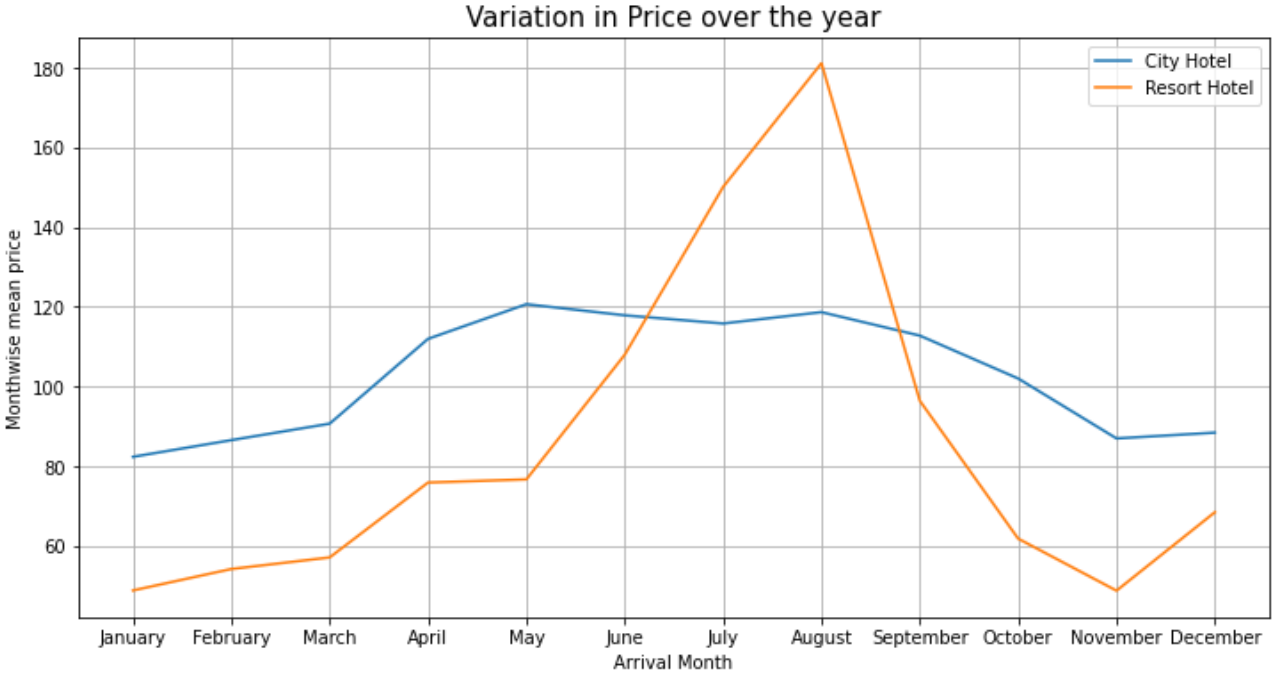
EDA & Observations (Cont.)



Observation:

•July and August months had the most Bookings. Summer vacation can be the reason for bookings.

EDA & Observations (Cont.)

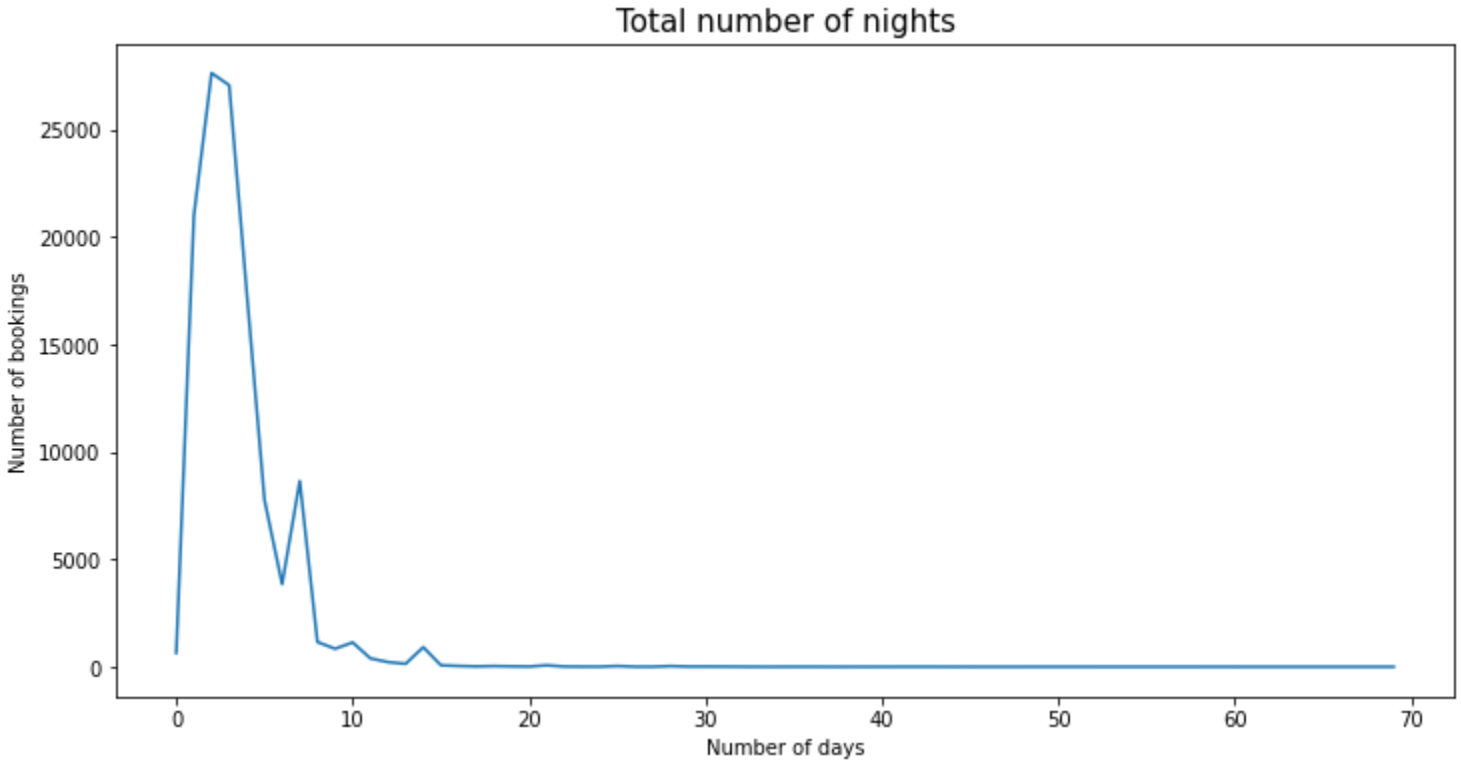


Observations:

•For Resort hotel ADR is high in the month June, July, August as compared to City Hotels. May be Customers/People wants to spend their Summer vacation in Resorts Hotels.

•From 9 observation, it is cleared that the, both type of hotels have the fewest guests during the winter. Therefore, the best time for guests to visit Resort or City hotels is January, February, March, April, October, November and December as the average daily rate in this month is very low.

EDA & Observations (Cont.)

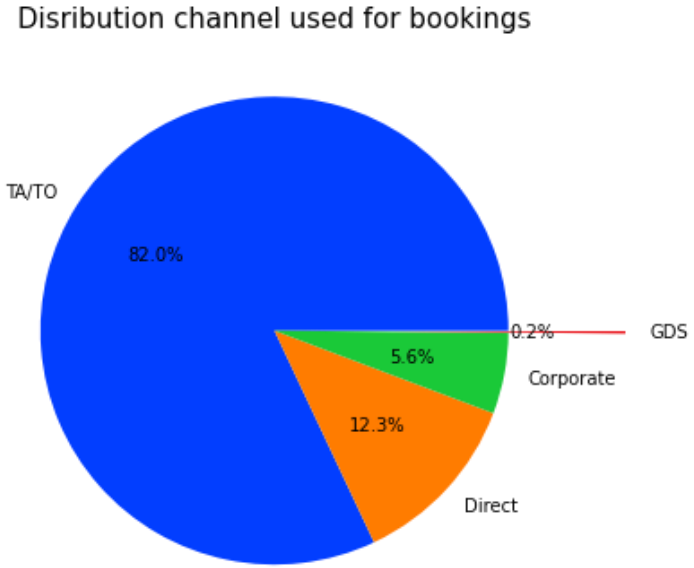


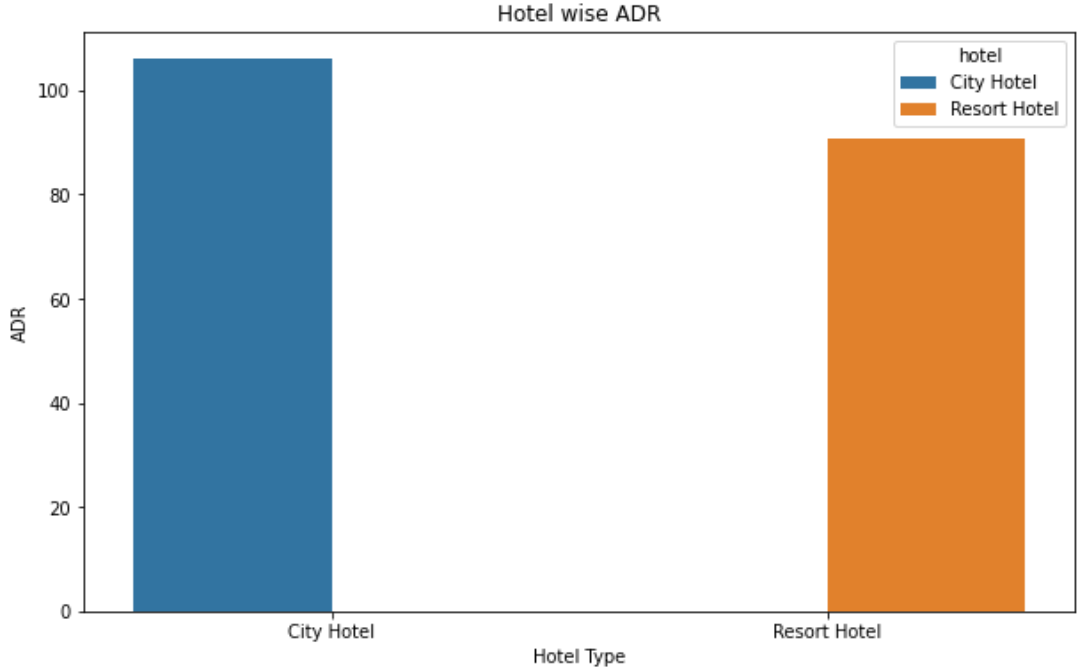
Observation:

•Guests booked hotels mostly for 0 to 3 nights.

•Very few guests booked hotels for more than 8 days.

EDA & Observations (Cont.)



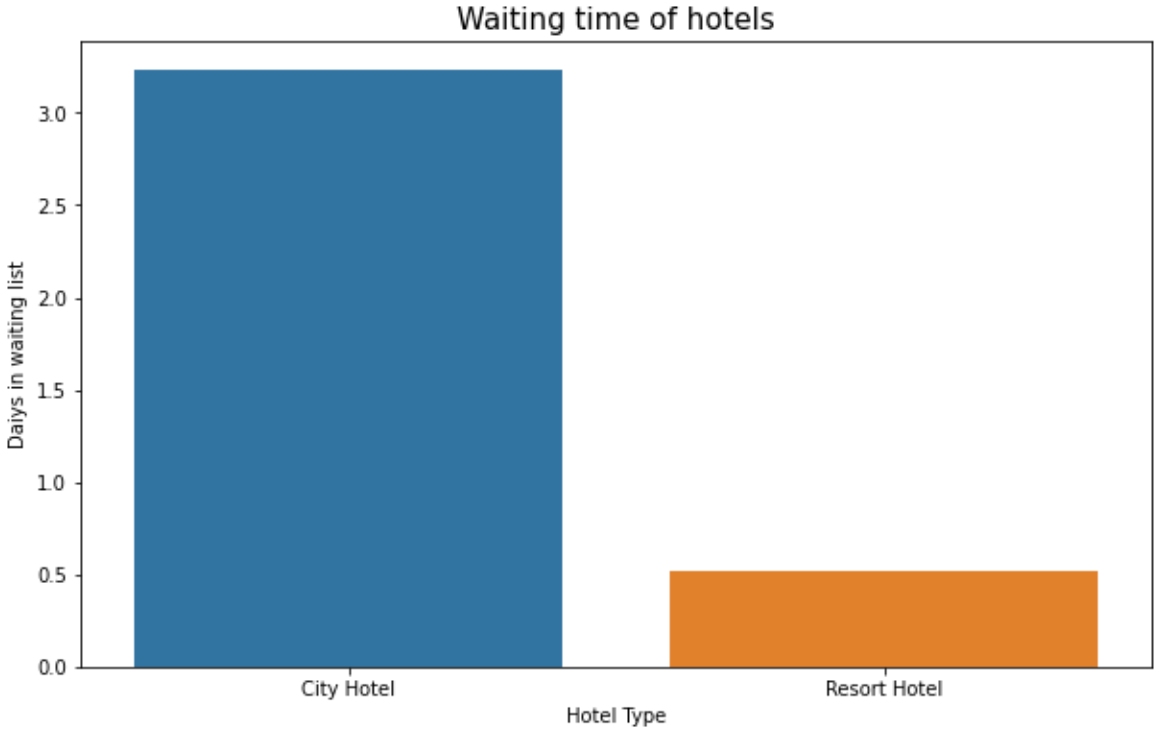


Observations:

•TA/TO is mostly (82%) used for booking hotels

•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.

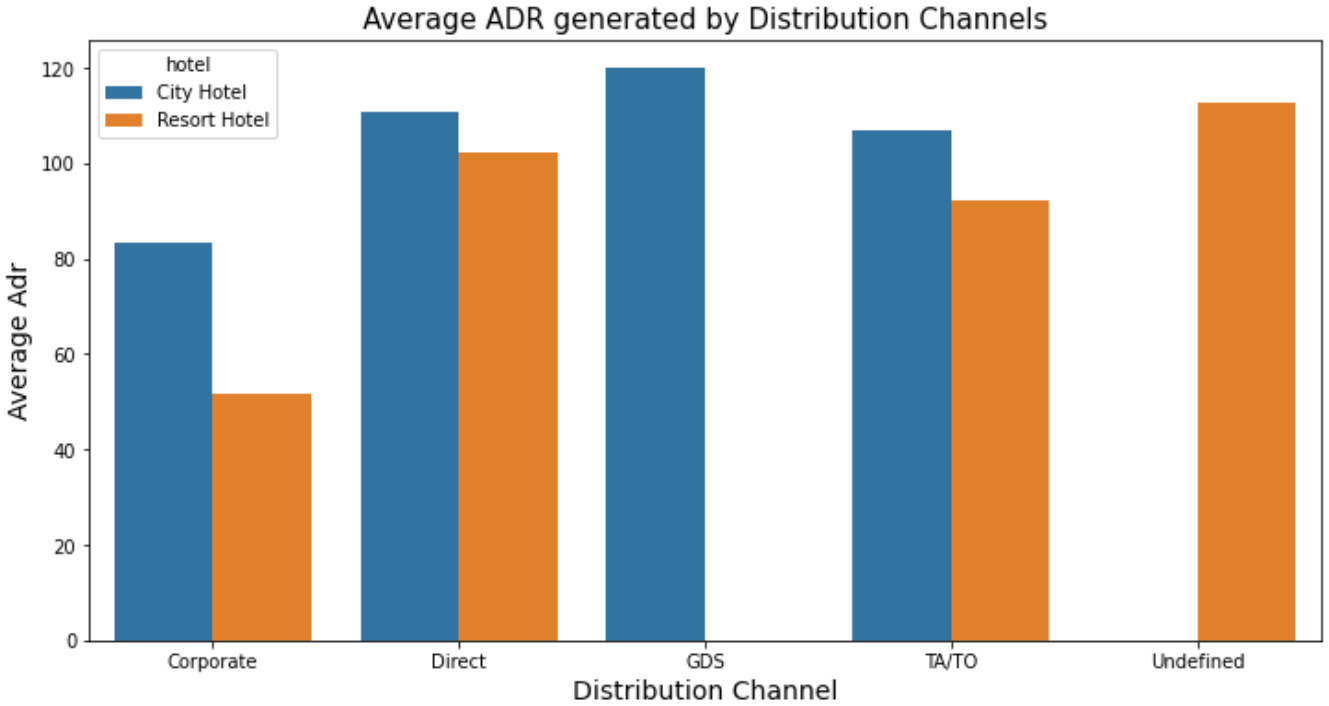
EDA & Observations (Cont.)



Observation

•City hotel has the higher waiting time.

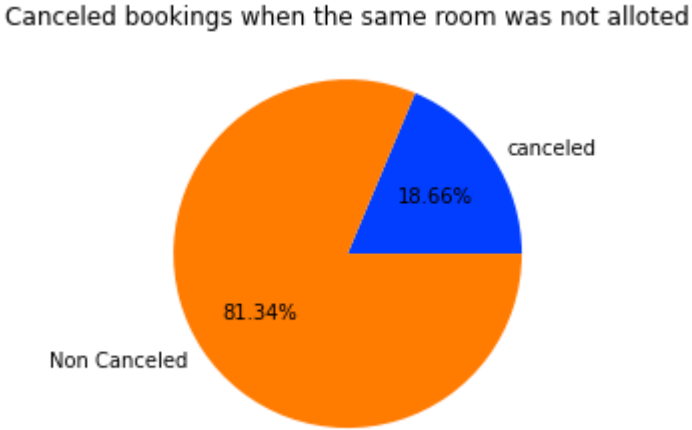
EDA & Observations (Cont.)



Observation:

- 'Direct' and 'TA/TO' has almost equally contributed in ADR in both types of hotels.
- GDS has highly contributed in ADR in 'City Hotel' type.
- GDS need to increase Resort Hotel Bookings, for increasing its ADR.
- Resorts made high ADR by Undefined mode of booking also.

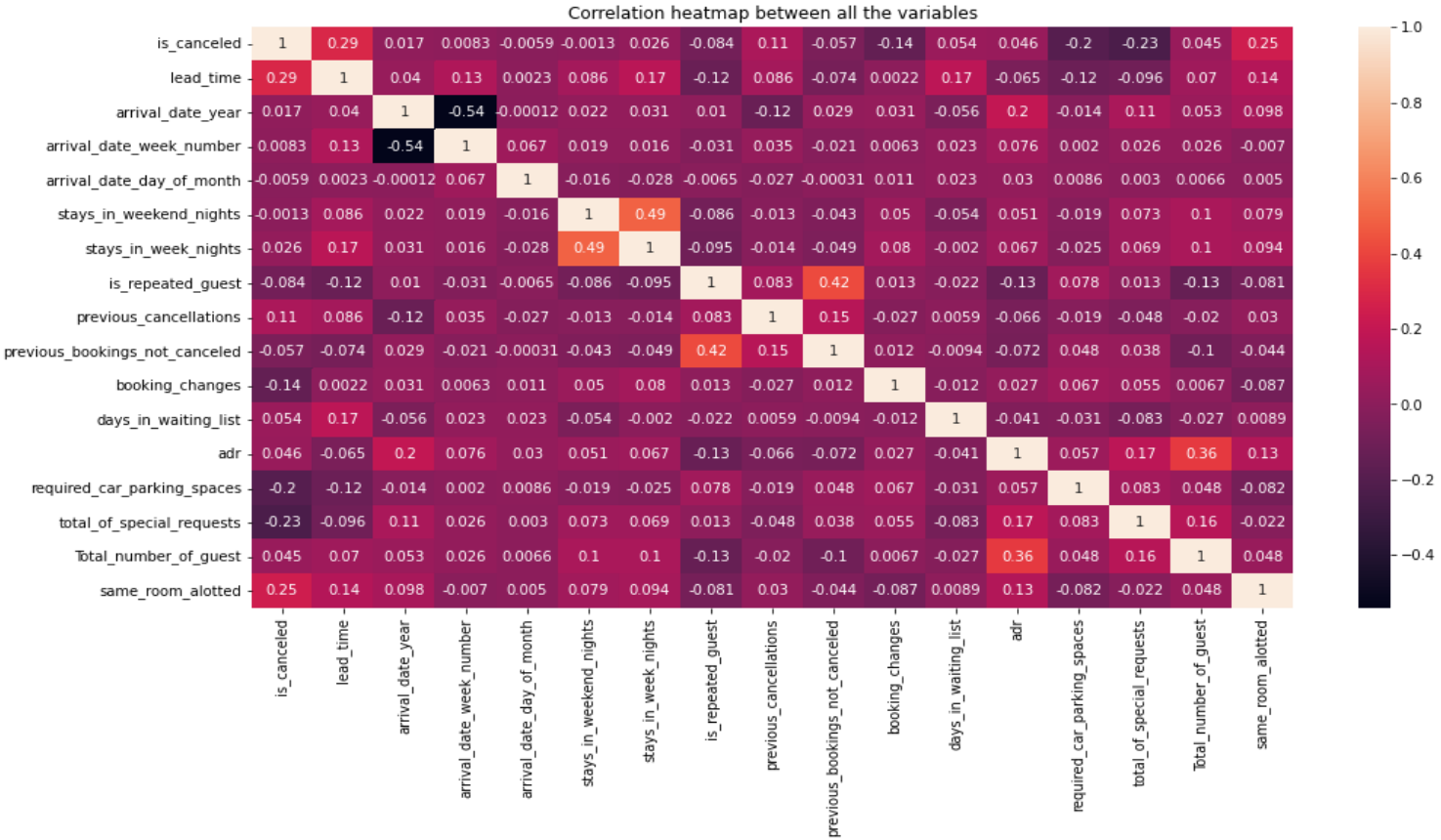
EDA & Observations (Cont.)



Observations:

- Maximum customer do not made any special requests alongwith their bookings.
- Number of special request is higher in case of city hotel, because the number of bookings is also higher.
- In the end, we can say that, very few customers made more than 2 special requests.
- It is noticed that the majority(81.34%) customer do not canceled their booking, when they don't get the desired room.
- 18.66% bookings canceled due to this, so hotels need to take little care about this.

EDA & Observations (Cont.)



EDA & Observations (Cont.)

Observation:

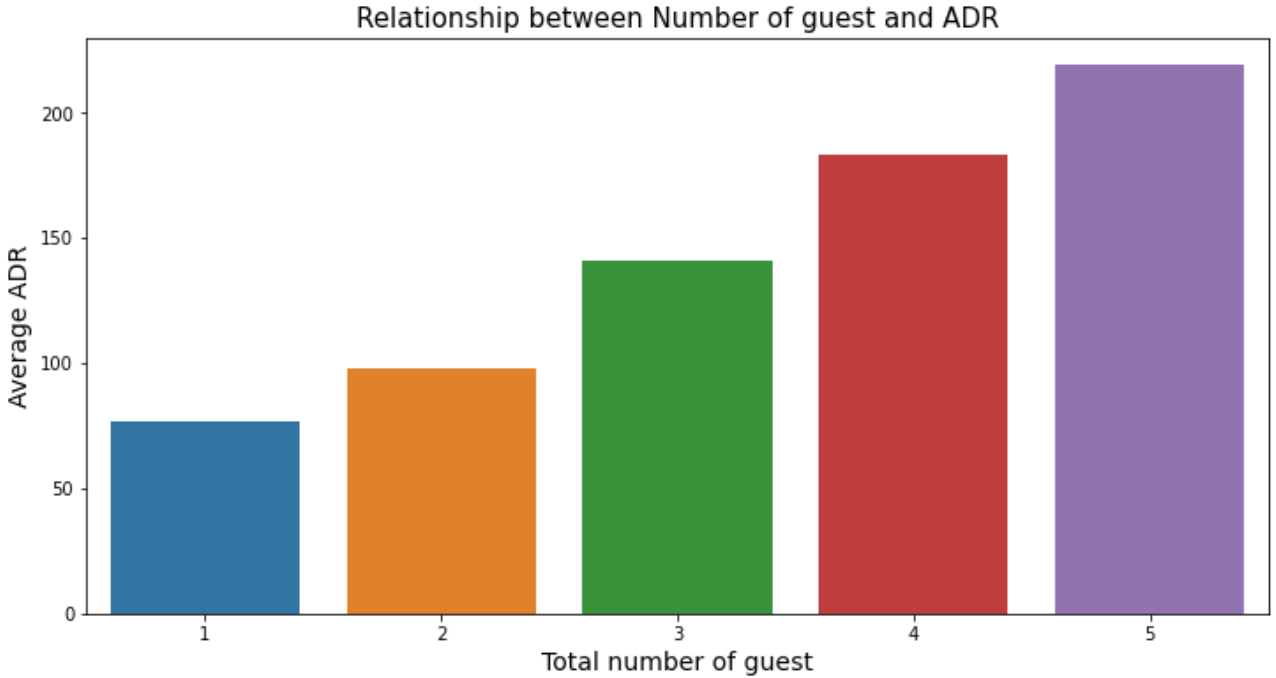
•is_canceled and same_room_alloted_or_not are positively corelated. That means customer is sometimes cancel his bookings if he don't get the same room as per reserved room.

•lead_time and total_stay is positively corelated. That means more is the stay of customer more will be the lead time.

•total number of guest is positively corelated to each other. That means more the people more will be adr.

•is_repeated guest and previous bookings not canceled has strong corelation. May be repeated guests are not more likely to cancel their bookings.

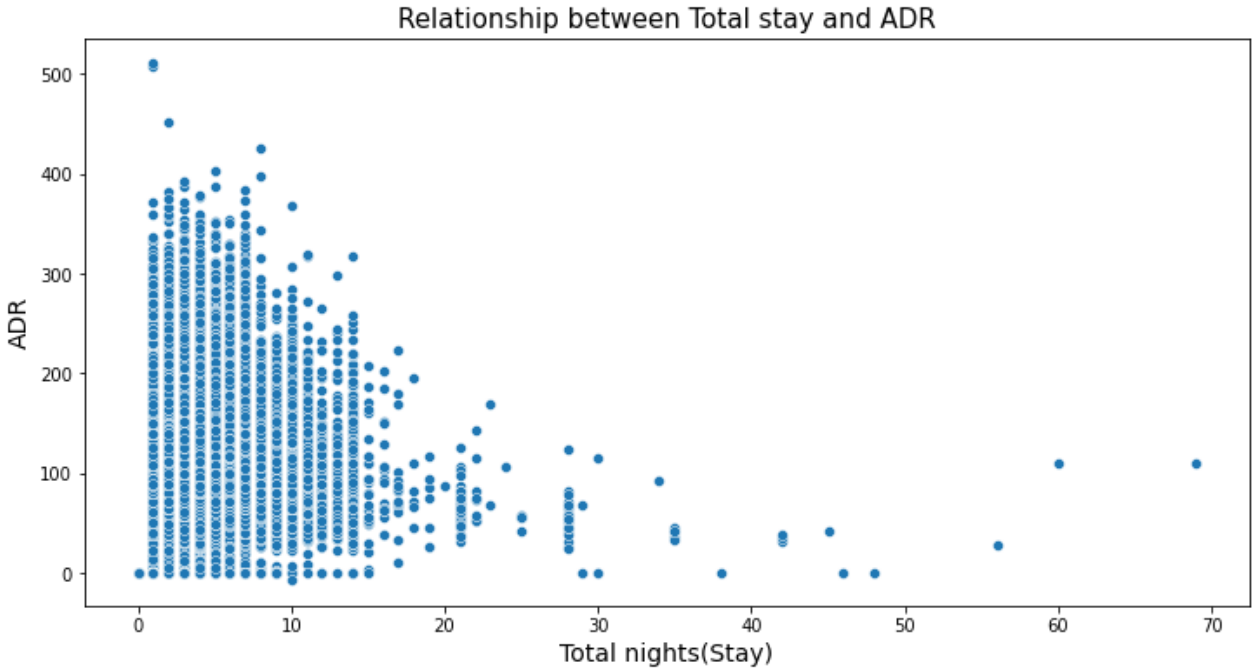
EDA & Observations (Cont.)



Observation:

•As the total number of people increase adr also increases. Thus adr and Total people are proportional to each other.

EDA & Observations (Cont.)



Observation:

•As the total stay increases adr is decreasing. Thus, for longer stays customer can get good adr(Price).

Final Conclusion

Above observations are very useful for Hotel team for the growth of their business. Alongwith above observations, I made the some important conclusion for the Hotels and Customer given below:

1. Best months for planning a trip are October to February, because prices for both the hotels are lessor as compare to other months due to less bookings.

2. Guest numbers for the Resort hotel go down slightly from June to September, which is also when the prices are highest. Thus, these months should be avoided for booking.

3. Very large number of customers are cancelling their bookings, so hotels need to make strict cancellation policy, like they can use non-refund options.

Final Conclusion(Cont.)

4. Cancellations are high when done through agents compared to direct booking. Hotels need to do marketing and give special incentives for direct booking as these may establish personal one to one relationship promoting customer loyalty.

5. The number of repeated guest is very low, in order to retained the guests management should take feedbacks from guests and try to improve the services.

Thank You!