

FIRST CAPSTONE PROJECT

EDA- Hotel Booking Analysis

By

Ranjit R. Ghadge

- Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!
- This data set contains booking information for a city hotel and a resort hotel, and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. All personally identifying information has been removed from the data.
Explore and analyze the data to discover important factors that govern the bookings.

WORK FLOW FOR EDA

DATA CLLECTION & UNDERSTANDING:

Here we will going through each variable & will do proper understanding of each variable of data. & Import a data to further process.

DATA CLEANING & MANIPULATION:

Data Cleaning means the process of identifying the incorrect, incomplete, inaccurate, irrelevant or missing part of the data and then modifying, replacing or deleting them according to the necessity.

EDA & VISUALIZATION:

Exploratory Data Analysis (EDA) is a process of describing the data by means of statistical and visualization techniques in order to bring important aspects of that data into focus for further analysis.

- **Univariate Analysis:** The data we are analyzing is only one variable
- **Bivariate Analysis:** We are comparing two variables to study their relationship
- **Multivariate Analysis:** Same as Bivariate analysis but only we are comparing more than two variables
-

DATA COLLECTION & UNDERSTANDING

Data Input	Description
hotel	City and Resort hotel
is_canceled	indicating booking cancelled (1) or not cancelled (0)
lead_time	the time difference between booking date and actual check in
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 weekends night
stays_in_week_nights	no of week nights
adults	no of adults
children	no of children
babies	no of babies
meal	type of meal
country	customers country of origin
market_segment	Market segment type
distribution_channel	booking description channel
is_repeated_guest	if repeated guest (1) or no(0)

DATA COLLECTION & UNDERSTANDING

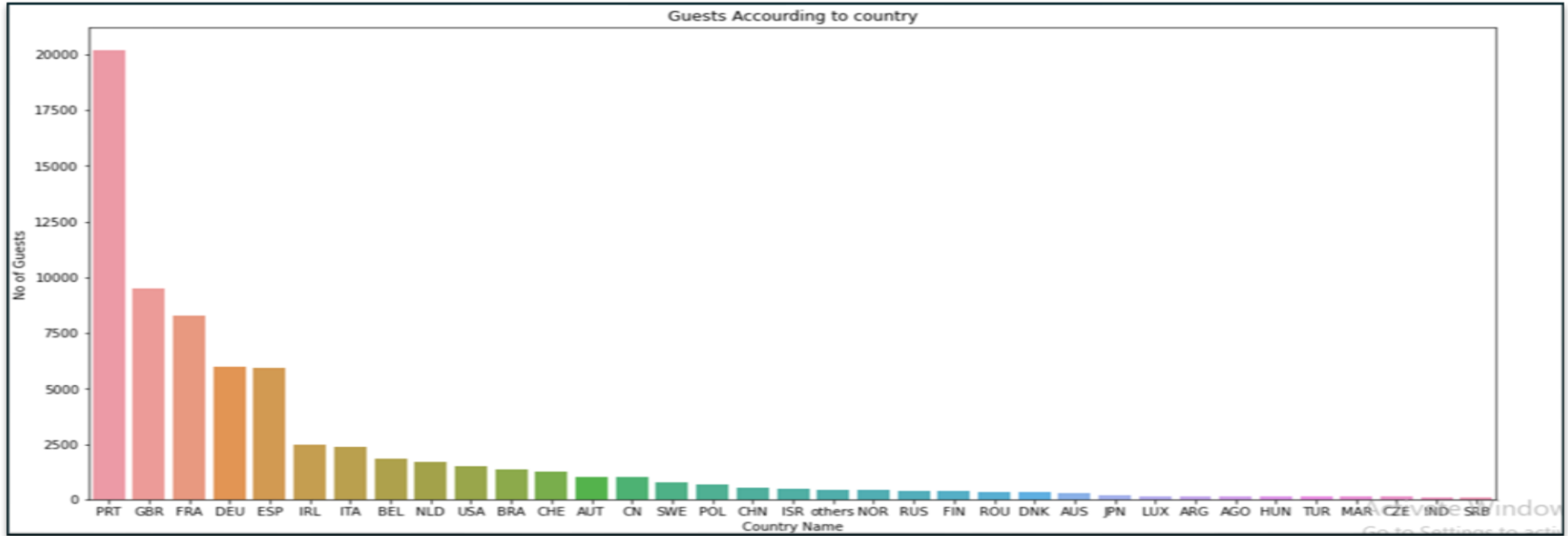
Data Input	Description
previous_cancellations	no of previous bookings those are cancelled by the customer before the current booking
previous_bookings_not_canceled	no of previous bookings not cancelled by the customer before the current booking
reserved_room_type	Type of reserved room
assigned_room_type	Type of assigned room
booking_changes	no of changes made in the booking from the moment the booking was entered till check in or cancellation
deposit_type	no deposit or refundable or non refundable
agent	ID of travel agent
company	ID of the company that made the booking
days_in_waiting_list	no of days the booking was in waiting list
customer_type	type of customer contract,group
adr	Average daily rate
required_car_parking_spaces	required car parking spaces
total_of_special_requests	no of special request
reservation_status	reservation last status
reservation_status_date	check out date

PROBLEM STATEMENTS



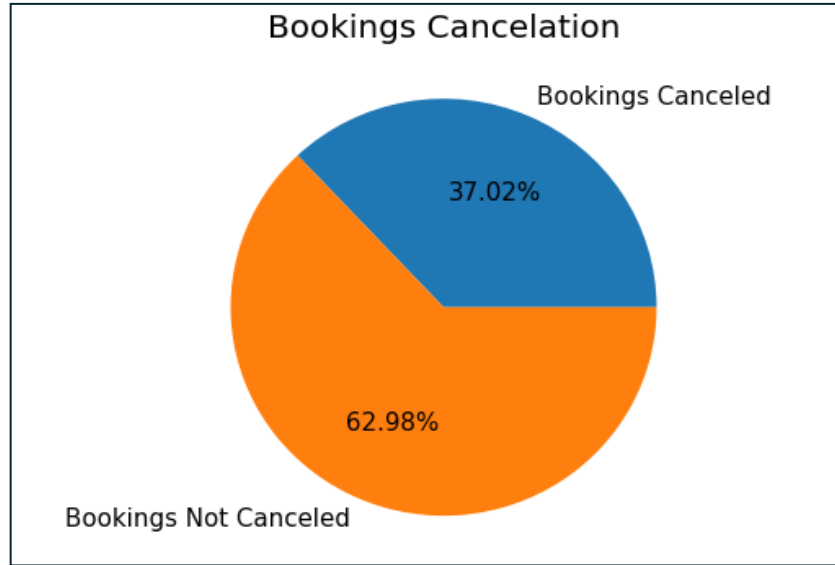
1. from where most guests are coming?
2. What percentage of bookings are cancelled
3. Which are the busiest months?
4. Which Year having highest booking?
5. Which agent made most booking?
6. Which type of food preferred by guest?
7. Most preferred room type by customers?
8. Which type of hotel having longer waiting time?
9. Comparison of no of adults VS Booking
10. When the best time of year to book a hotel room is?
11. Which type of hotel has highest ADR?
12. What is relationship of ADR with total no of people?
13. Which distribution channel has the highest cancellation rate?
14. The optimal length of stay in order to get the best daily rate?
15. Total Bookings vs. Length of Stay
16. Length of Stay: City Hotel vs. Resort Hotel
17. What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests?

1. FROM WHICH COUNTRY MOST OF GUEST COMING ?



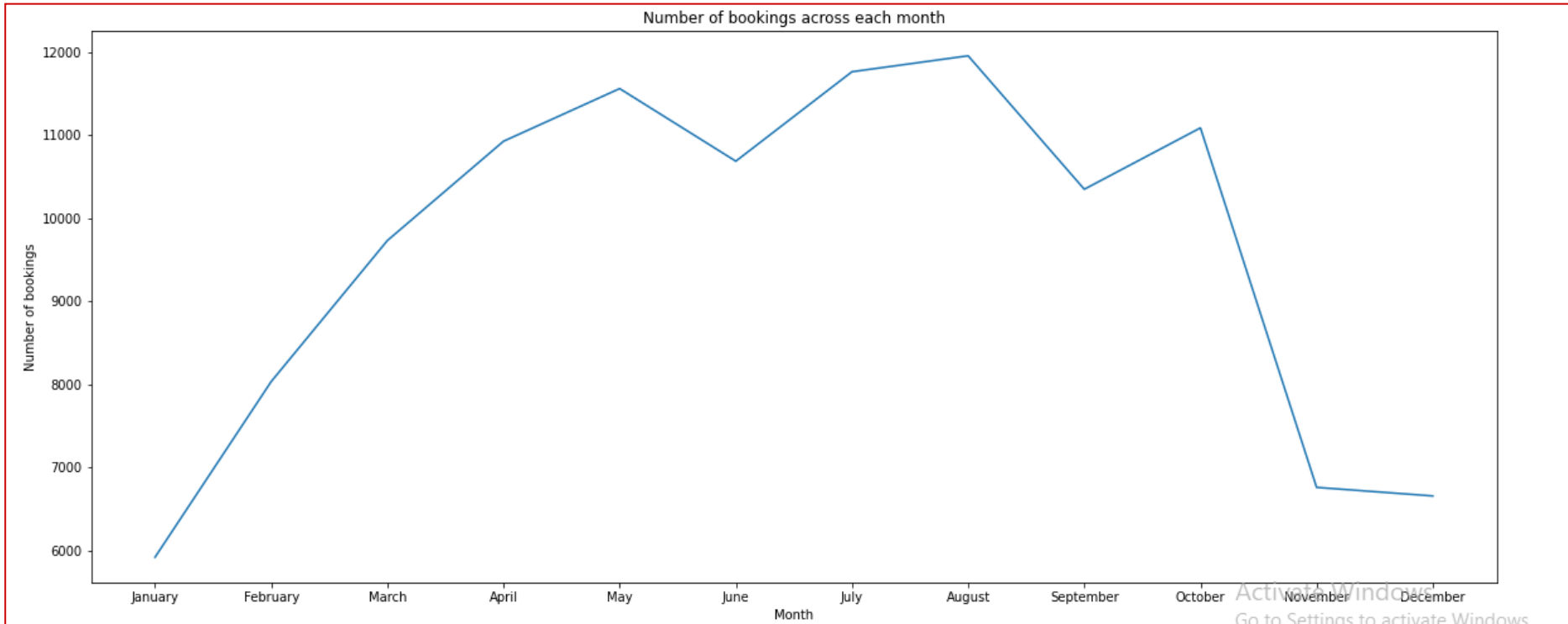
Most of guest are coming from PRT, GBR & FRA.

2. WHAT PERCENTAGE OF BOOKING ARE CANCELLED ?



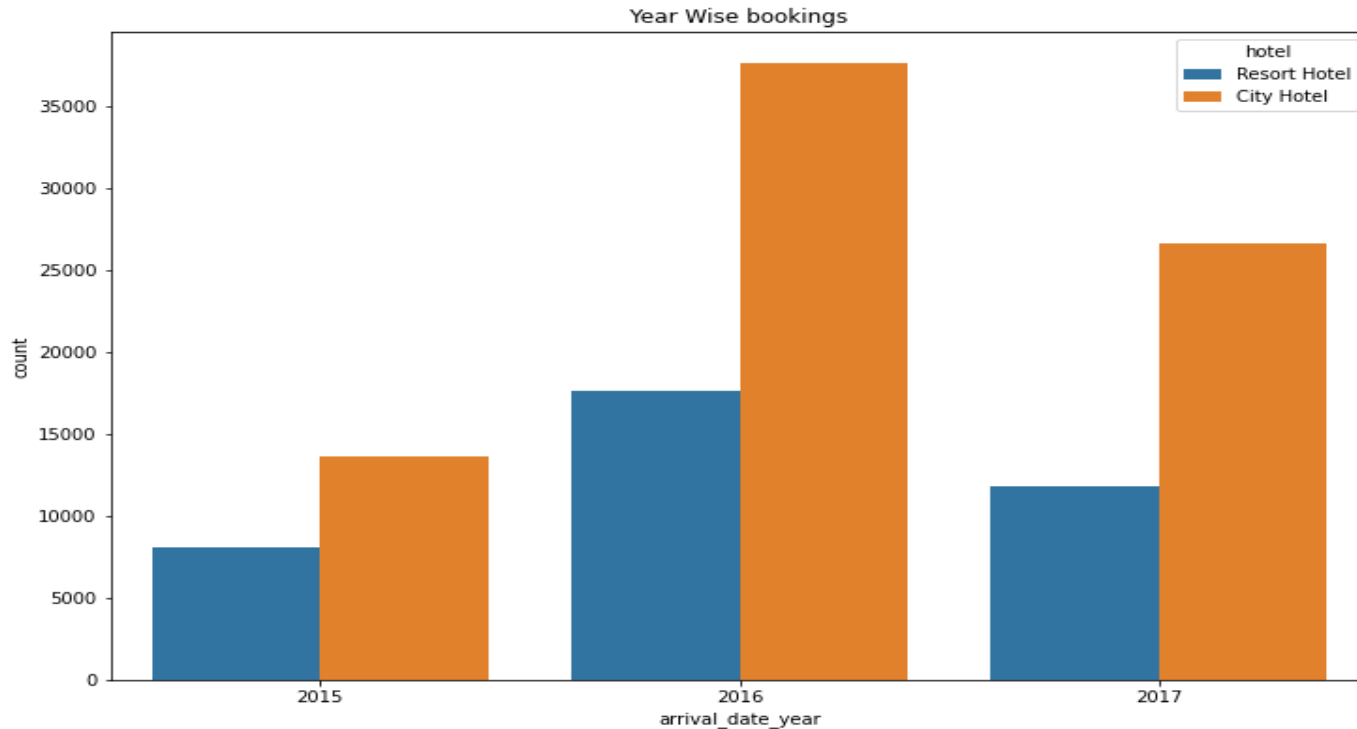
- 37.02% of booking got cancelled
- 62.98% of booking not cancelled.

3. WHICH ARE MOST BUSY MONTHS ?



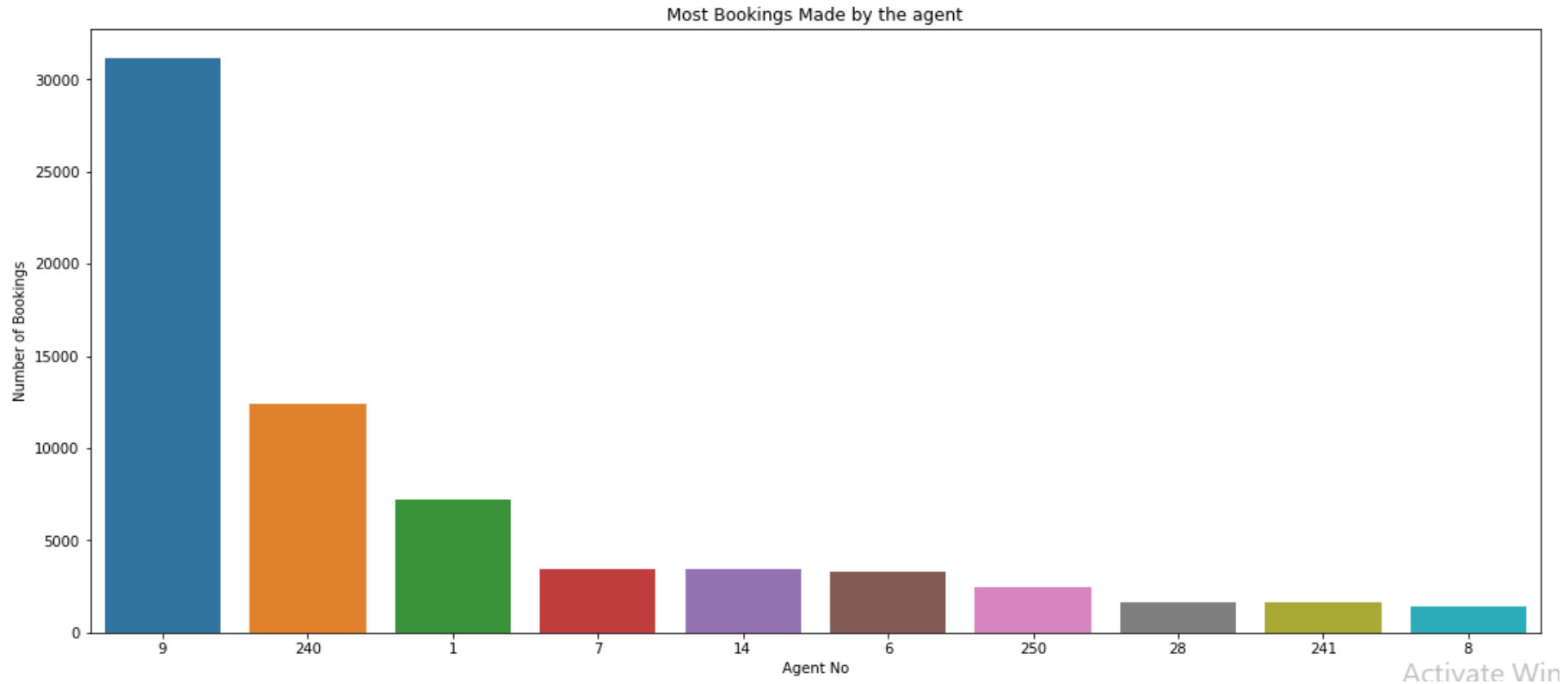
- From above plot graph August & July months in a year having highest booking.
- Very less no of bookings are done in December & January month.

4. WHICH YEAR HAVING HIGHEST BOOKING?



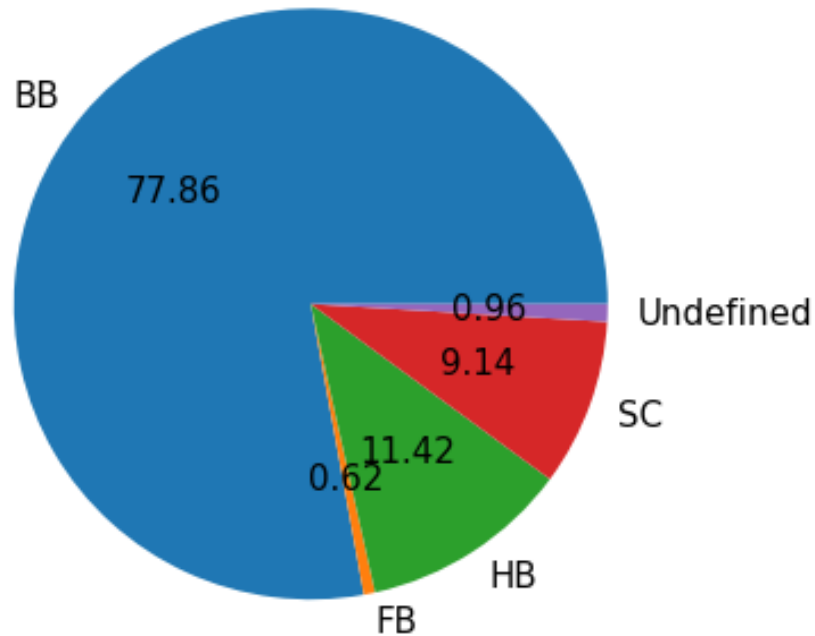
- The year 2016 had a high number of bookings for both City Hotels and Resort Hotels.
- Hotel bookings for cities are almost always higher than those for resorts.

5. WHICH AGENT MADE MOST BOOKING?



- 9 No. agent make highest no of booking as compare to all other agents.
- Agent no 9,240 & 1 make highest no of booking.

6. WHICH TYPE OF FOOD PREFERD BY GUEST?



BB: Bed & Breakfast

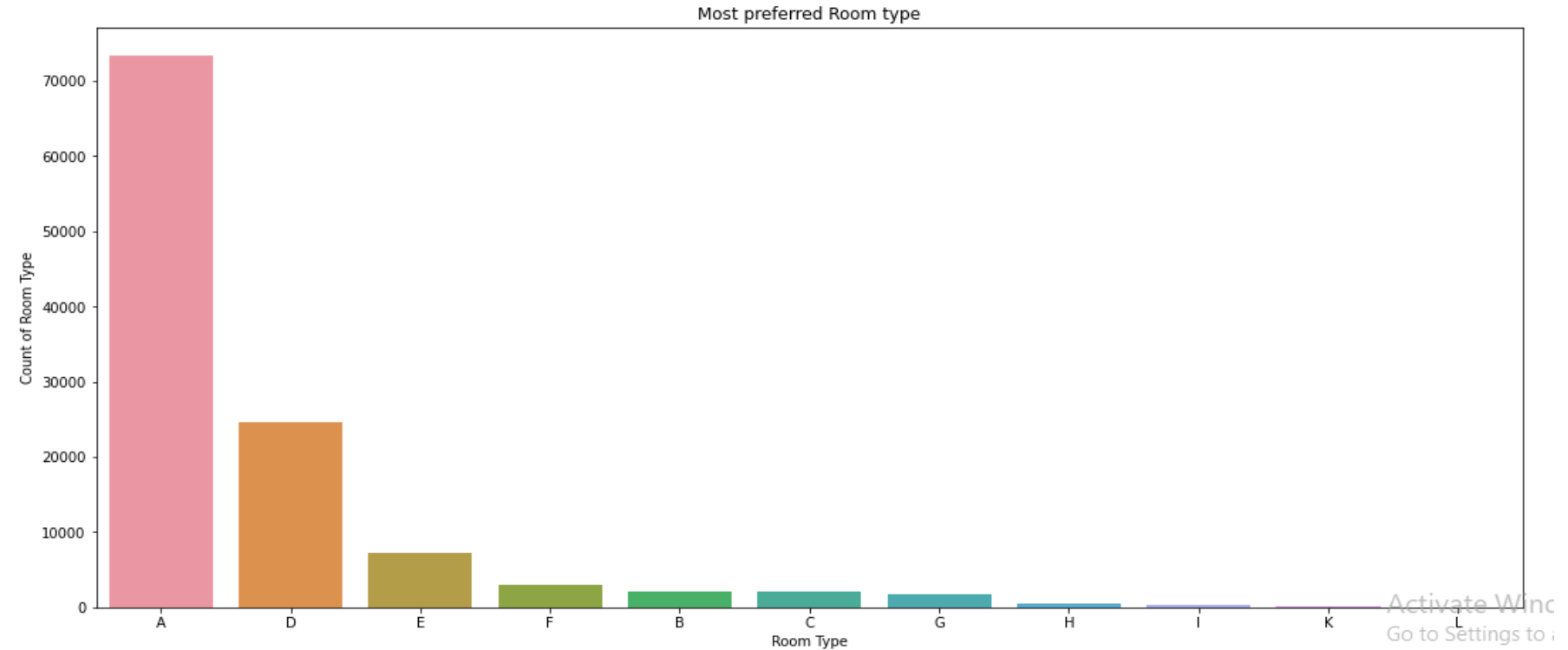
HB: Half Board (Breakfast and Dinner normally)

FB: Full Board (Breakfast, Lunch and Dinner)

SC: Self Catering

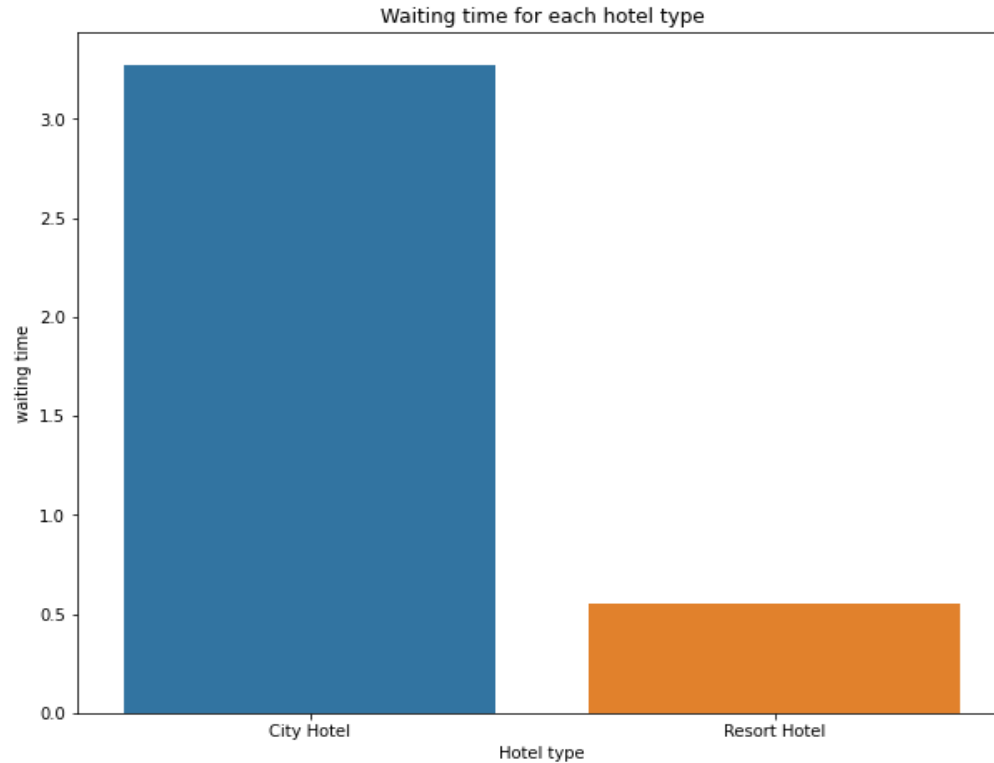
- BB (Bed & Breakfast) type meal were mostly preferred.

7. MOST PREFERED ROOM TYPE BY CUSTOMER ?



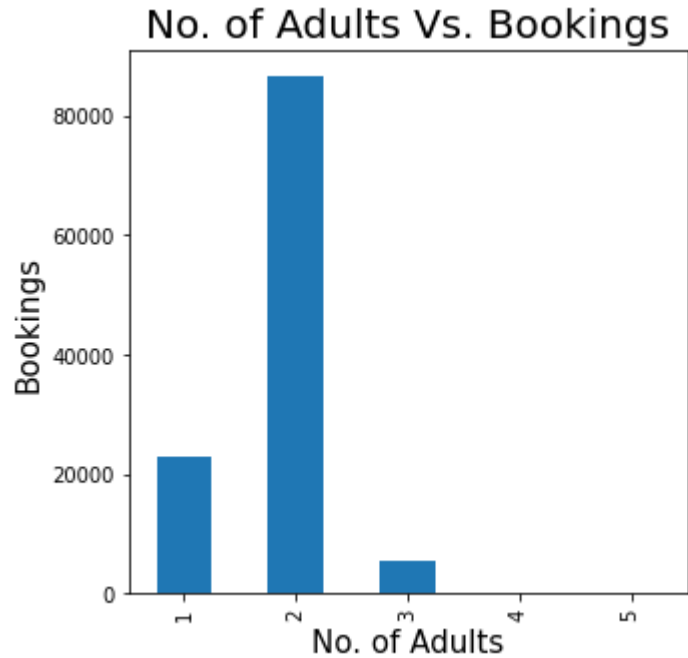
- Most preferred room type is A
- Room type of A,B,C mostly preferred by guest

8. WHICH TYPE OF HOTEL HAVING LONGER WAITING TIME?



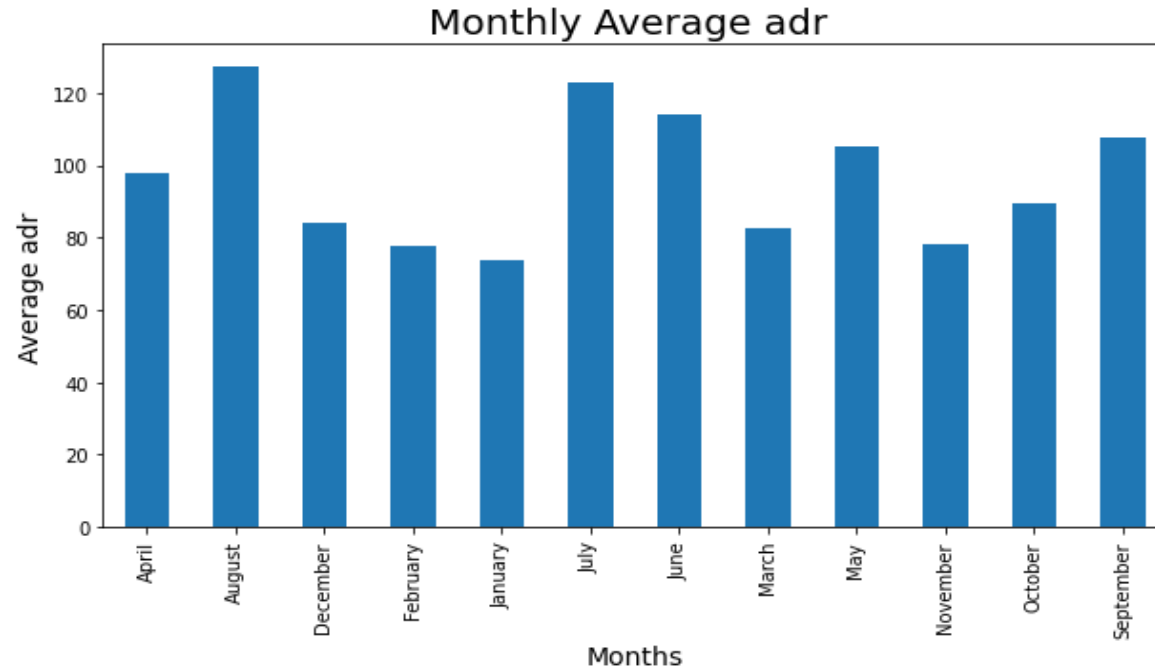
- City hotel having longer lead time than Resort hotel.

9. NO OF ADULTS VS BOOKING ?



- Two nos of adults having major bookings.

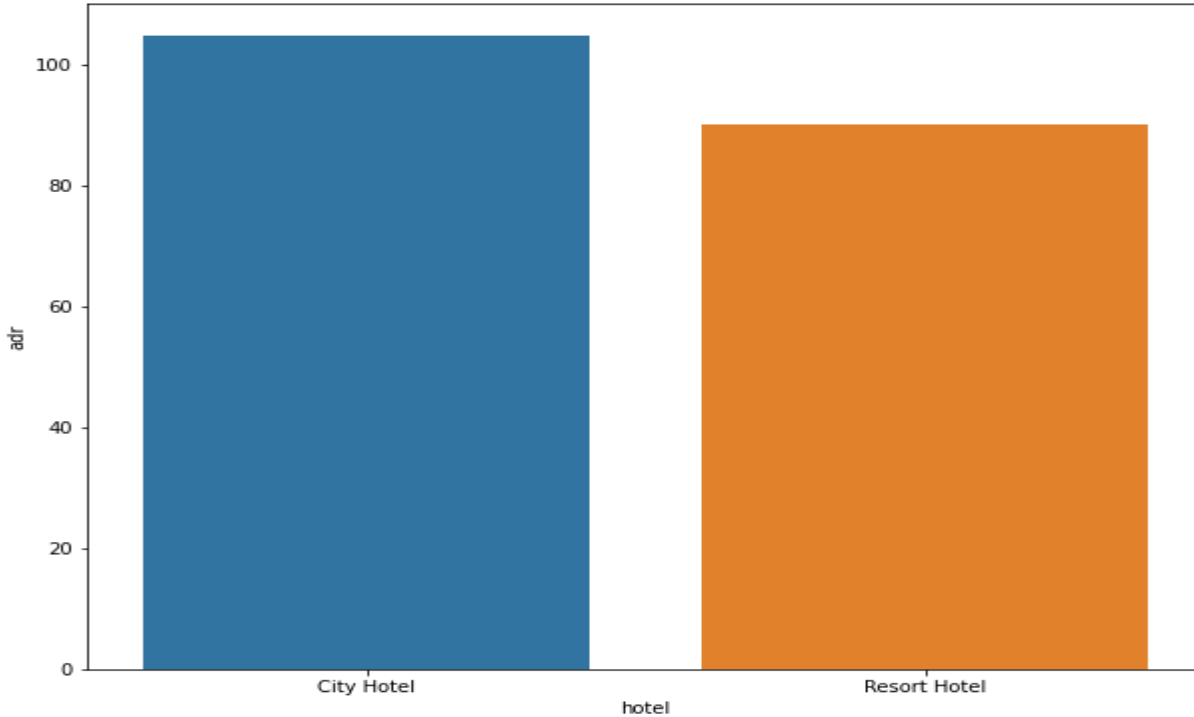
10. BEST TIME IN YEAR TO BOOK A HOTEL ?



- After compare ADR with no of bookings of each Month.
- Average daily rate was highest in month of August.
- Average daily rate was lowest in month of January.
- So best time of year to book a hotel was January.

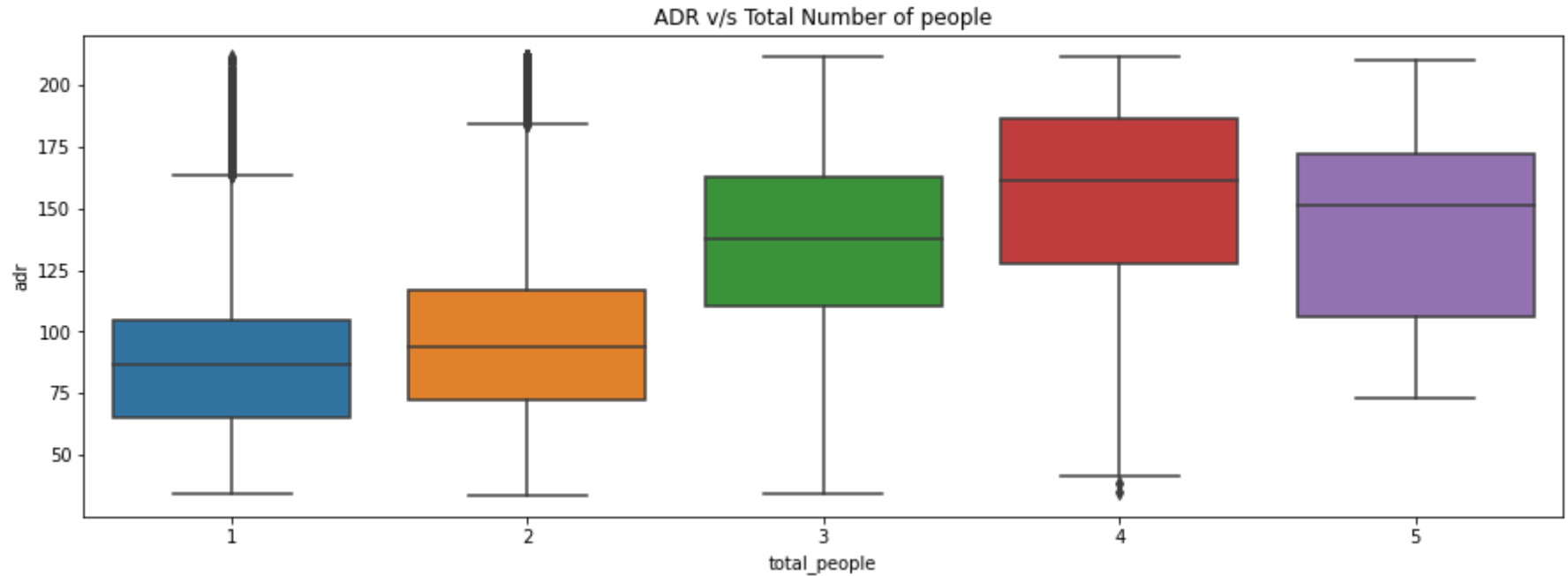
11. WHICH TYPE OF HOTEL HAVE HIGHEST ADR

Avg ADR of each Hotel type



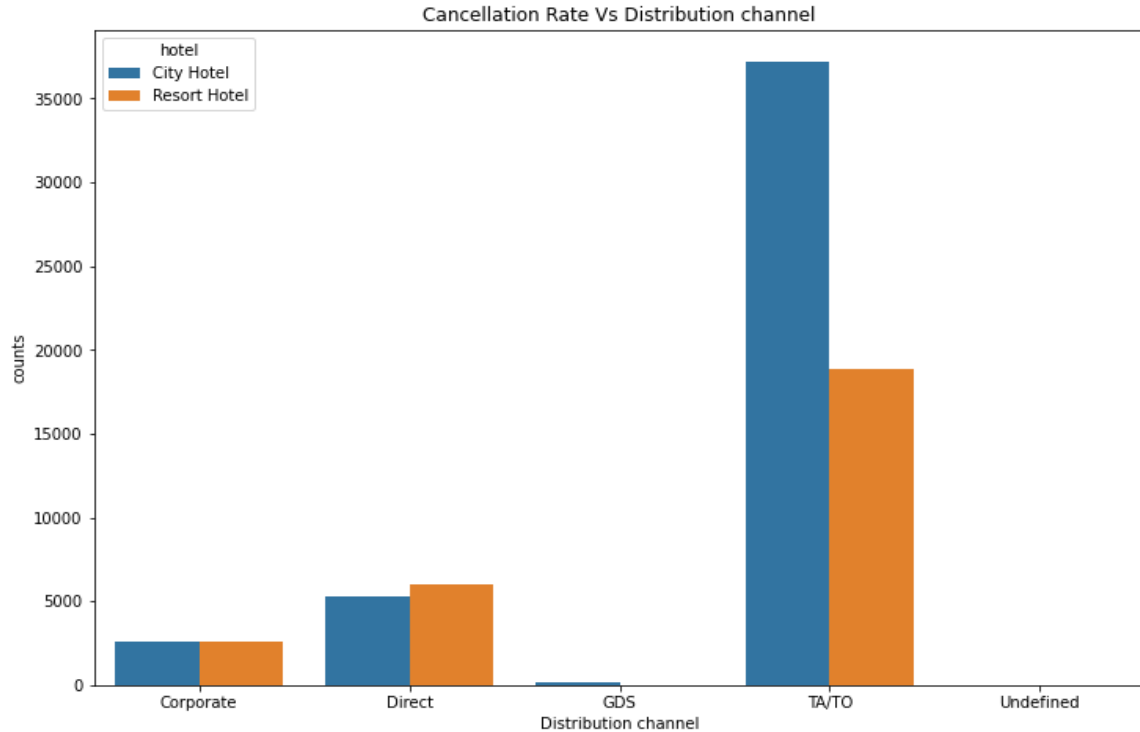
- City hotel has the highest ADR. means city hotels are generating more revenues than the resort hotels.

12. RELATIONSHIP OF ADR WITH TOTAL NO OF PEOPLE ?



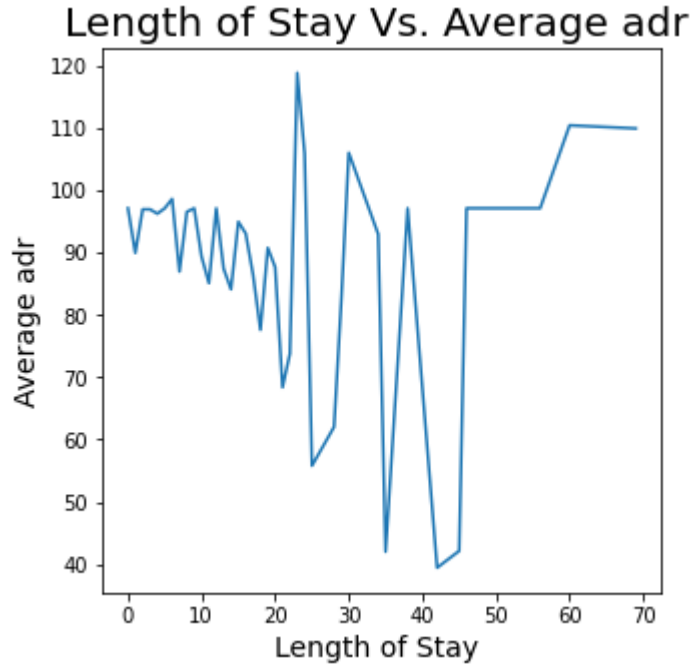
- As the total number of people increases adr also increases.
Thus adr and total people are directly proportional to each other.

13. WHICH DISTRIBUTION CHANNEL HAS HIGHEST CANCELATION RATE?



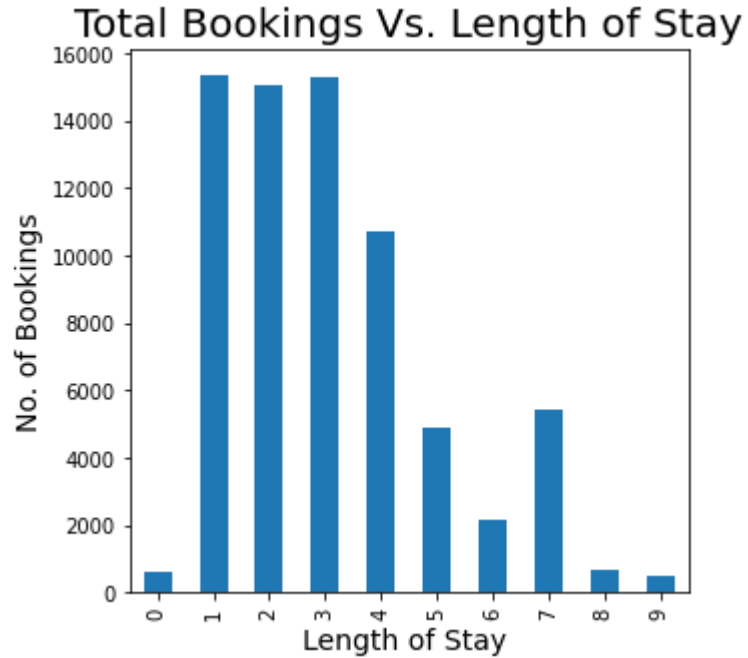
- In "TA/TO", City hotels has the high cancellation rate compared to resort hotels.
- In "direct" both the hotels has almost same cancellation rate.

14. OPTIMAL LENTH OF STAY IN ORDER TO GET BEST DAILY RATE?



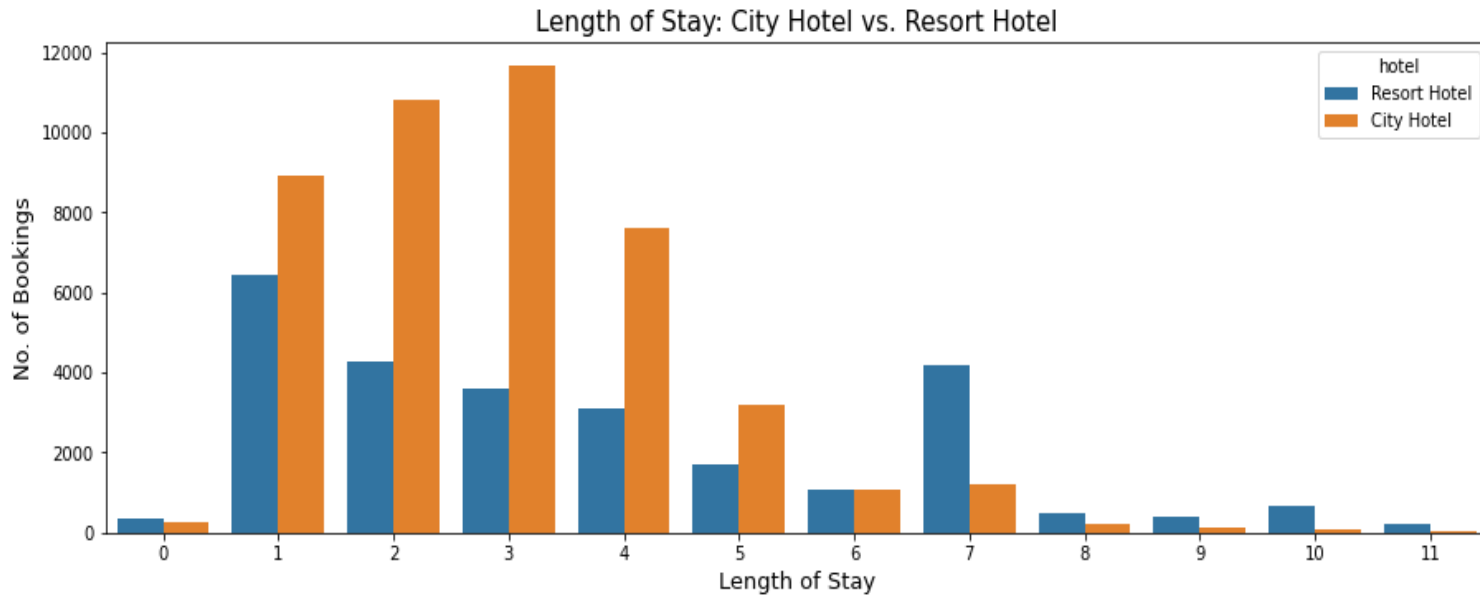
- 1 Lowest value of adr is for 42 nights length of stay.2 So to get beast daily rate one must stay for 42 nights.

15. TOTAL BOOKING VS LENGTH OF STAY



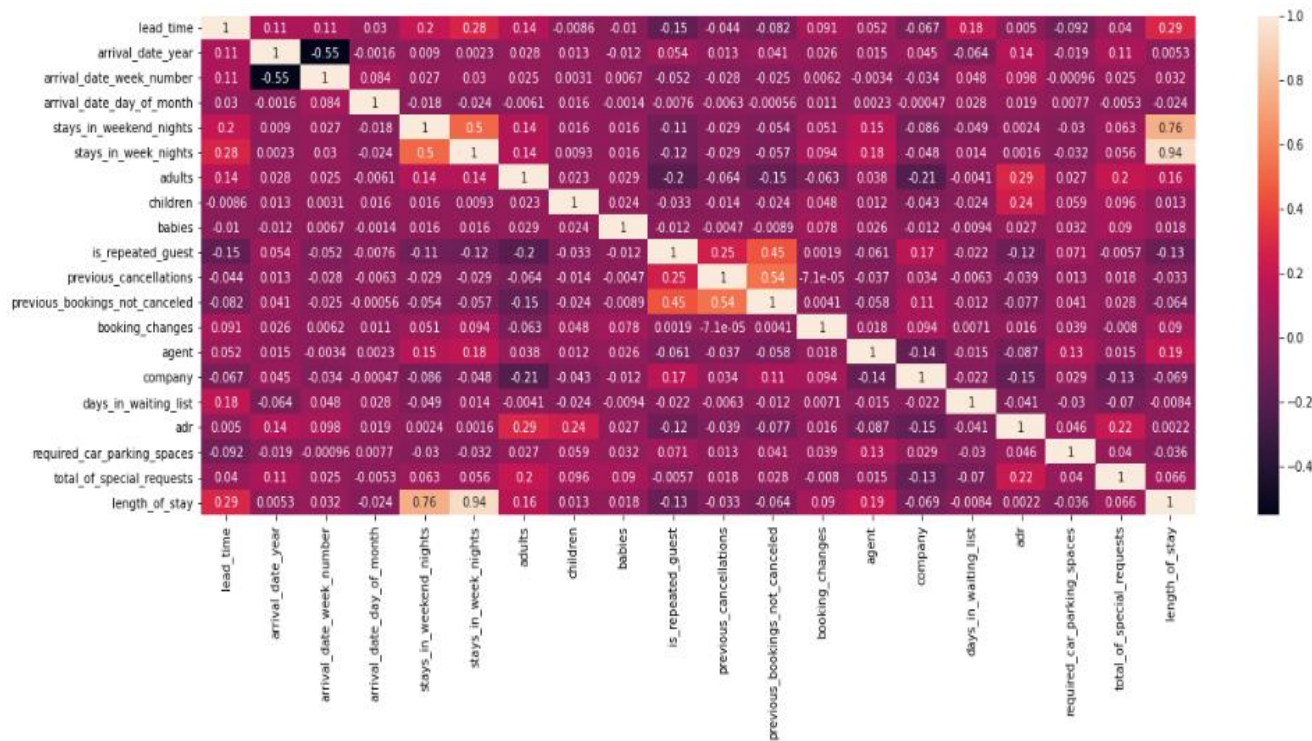
- Highest booking for 1,2 & 3 No of Nights.

16. LENGTH OF STAY: CITY HOTEL VS RESORT HOTEL



- Booking in Resort Hotel decrease with increasing length of stay.
- Bookings at the City Hotel rise up to a three-night stay as the length of stay increases. after a dramatic decline.

17. LENGTH OF STAY: CITY HOTEL VS RESORT HOTEL



- From above heat map total special request highly correlated to adr, adults & arrival date year.
- So total special request receive with high adr, adults & arrival year.

PROJECT CONCLUSION



- City hotels are the most preferred hotel type by the guests. We can say City hotel is the busiest hotel.
- 27.5 % bookings were got cancelled out of all the bookings.
- Only 3.9 % people were revisited the hotels. Rest 96.1 % was new guests. Thus retention rate is low.
- The percentage of 0 changes made in the booking was more than 82 %. Percentage of Single changes made was about 10%.
- Most of the customers (91.6%) do not require car parking spaces.
- 79.1 % bookings were made through TA/TO (travel agents/Tour operators).
- BB(Bed & Breakfast) is the most preferred type of meal by the guests.
- Maximum number of guests were from Portugal, i.e. more than 25000 guests.
- Most of the bookings for City hotels and Resort hotel were happened in 2016.
- Average ADR for city hotel is high as compared to resort hotels. These City hotels are generating more revenue than the resort hotels.
- Booking cancellation rate is high for City hotels which almost 30 %.
- Average lead time for resort hotel is high.
- Resort hotels have the most repeated guests.
- Optimal stay in both the type hotel is less than 7 days. Usually people stay for a week.

Thank You