

EDA CAPSTONE PROJECT

FOR HOTEL BOOKING ANALYSIS

PREPARED BY:

FAIZAN .N. KHAN

TABLE OF CONTENT

AI

- **Aim**
- **Data Summary**
- **Variable Description**
- **Data Visualization**
 1. Most Preferred Hotel
 2. Hotel type with highest ADR(Average Daily Rate)
 3. Relationship between ADR and Total Stay
 4. Percentage of Repeated Guests.
 5. Percentage Distribution Of required car parking spaces.

6. Meal Type Distribution.
 7. Mostly used for Distribution Channels and Relationship of Distribution channel and ADR.
 8. Booking By month and optimal stay length in Hotels.
 9. Plotting Histogram.
 10. Year and Hotel wise confirmed booking and cancellation Distribution.
 11. ADR across different Month
 12. Weekly Stay Distribution and Calculation and Non-Calculation.
 13. Room type preference and Customer Types.
 14. Correlation Heatmap.
- **Suggestion for the client to achieve Business Objective?**
 - **Conclusion.**

This project contains the real world data record of hotel bookings of a city and a resort hotel containing details like bookings, cancellations, guest details etc. from 2015 to 2017. Main aim of the project is to understand and visualize dataset from hotel and customer point of view.

DATA SUMMARY

The logo consists of the letters 'AI' in a white, bold, sans-serif font, centered within a solid red square.

- Exploratory Data Analysis(EDA) is an approach to analyze the Data Using Visual Techniques .It is used to discover trends, patterns or check assumptions with the help of Graphical Representations.
- In This Project, I am Going to Begin to Explore the Dataset in Which Hotel Booking Comprises between Two Types of Hotels for Example, City Hotels and Resort Hotels.
- In This Dataset We Have 119390 Rows, 32 Columns.
- This Datasets Have Duplicate And Null Values
- In This Datasets have 31994 duplicate values and four Columns have missing Values
- In This Dataset Have Float64(4), Int64(16), Object(12) dtype
- In this Project I have Divided Data manipulation workflow into three different categories (1)Data Collection, (2)Data Cleaning, (3)Data Manipulation and EDA(EXPLORATORY DATA ANALYSIS).

VARIABLE DESCRIPTION

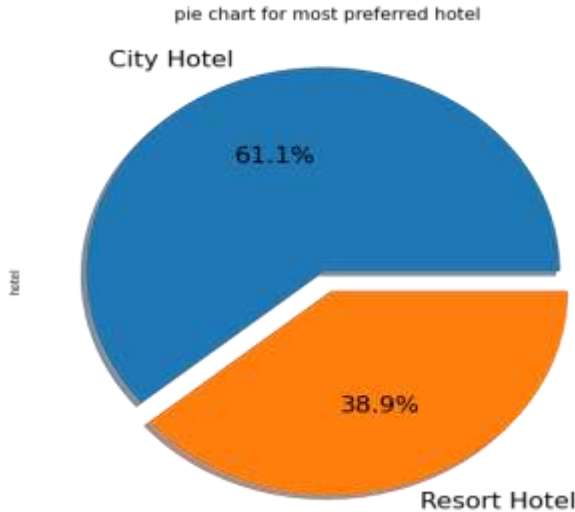
AI

- **hotel** : Name of the hotel (Resort Hotel or City Hotel)
- **is_Canceled** : if the booking was Cancelled(1) or not (0)
- **lead_time** : Numbers of days has elapsed between entering data of the booking into 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
- **arrival_date_day** : day of arrival date
- **stays_in_weekend_nights** : number of weekend nights(Saturday or sunday)the guest stayed or booked to stay at the hotel
- **stays_in_week_nights** : Number in week nights(Monday to Friday)the guest stayed or booked to stay at the hotel
- **adults** : Numbers of adults
- **children** : Number of children Among Guest
- **babies** : Numbers of babies Among Guest
- **meal** : Kind of meal opted By Customer

- **country** : Code of respective Country
- **make_segment** : Which segment the customer belongs to
- **Distribution_channel** : How the customer accessed the stay-corporate booking/direct/TA.TO.(TA : travel agent, TO : Tour Operator)
- **is_repeated_guest** : guest coming for the first time or not
- **previous_cancellation** : was there a cancellation before
- **previous_bookings** : count of previous booking
- **resevered_room_type** : code of room type resevered
- **assigned_room_type** : code of room type assigned
- **booking_changes** : count of changes made to booking
- **deposit_type** : Type of deposit made by the guest
- **agent** : booked through agent
- **days_in_waiting_list** : number of days in waiting list
- **customer_type** : Type of customer Arrived.
- **required_car_parking** : if car parking required.
- **total_of_special_req** : Number of special requirement by customers
- **reservation_status** : Reservation Status(cancelled, check-out, no show)
- **reservation_status_date** : Date of last reservation status

DATA VISUALIZATION

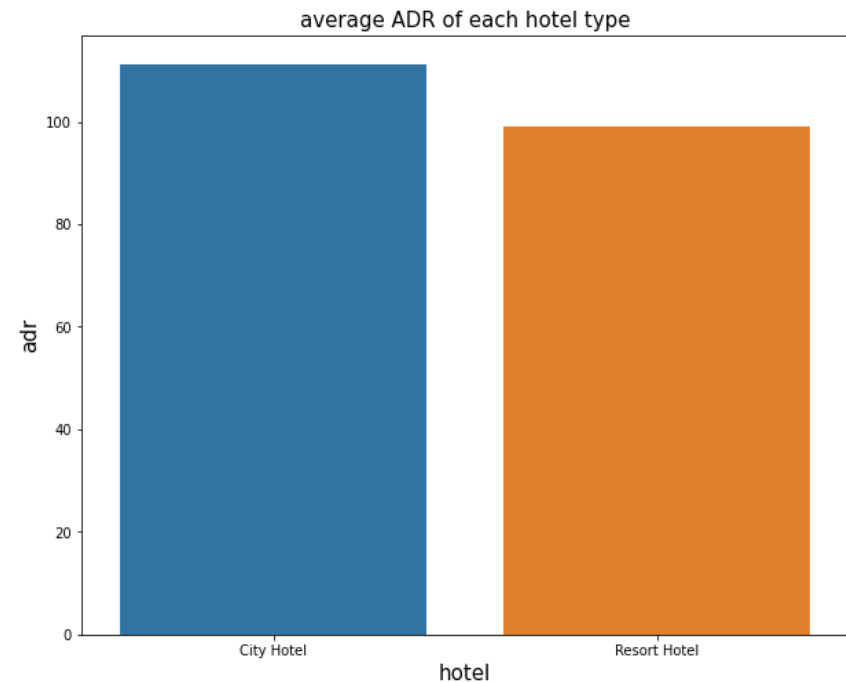
AI



- Most preferred Hotel
- From the above Chart, we get to know that City Hotel is most preferred hotel by guests. thus city hotel has maximum bookings 61.1% guests preferred city hotel while 39.9% guests preferred Resort Hotel.
- City Hotel Doing very well so they are providing more services for attracting more guest to increases his revenue, so they are doing really well but in case of resort hotel guest have shown less interest than city hotel so resort hotel need to find our better solution for attracting more guest and find what city hotel has done to attract guests, so there is very high scope of growth of resort hotel if they upgrade their service and adopt there better way

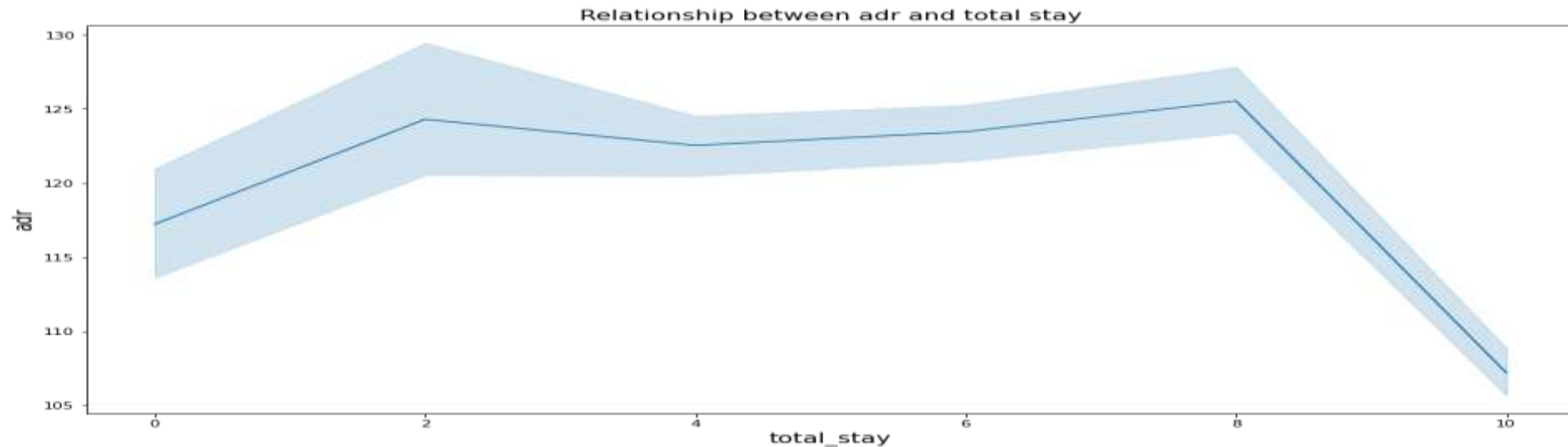
Hotel type with highest adr(Bivariate with Categorical – numerical)

- City Hotel has highest ADR. That Means City hotel is generating more revenue than resort hotel.
- so the city hotel has can do more advertising to get more revenue. Thus the city hotel are already enjoying high adr but bit more of positive efforts towards more definitely add alot to their growth and overall revenue.



Relationship between ADR and Total Stay (Bivariate with numerical-numerical)

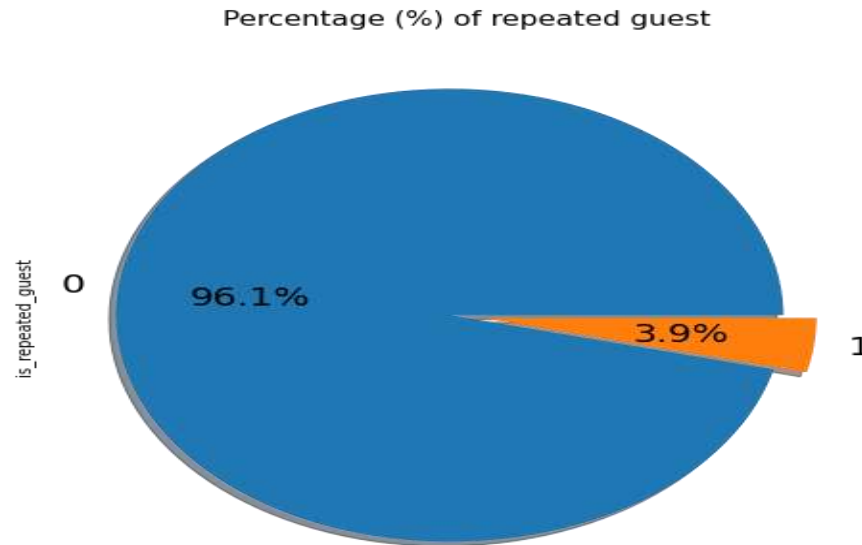
AI



- From this chart we have found that as the total stay increases the adr also increases so, adr is directly proportional to the total_stay.
- The hotel should focus on increasing thier ADR and more advertisement and better facilities and good offers will let the guests to stay more that will directly result in increasing adr. so, Hotels should offer more attractive offers and facilities so that total stay can be increased that will directly multiply their edr and ultimately will increases.

PERCENTAGE OF REPEATED GUESTS

AI



- Repeated Guest are very less which is only 3.9% while 96.1% guests are not returning to the same hotel So, it is amttter of deep thinking and taking proper step to increases a repeated guest numbers. In Order to retained the guests management should take feedback from the guests and try to impress with his Sevices.

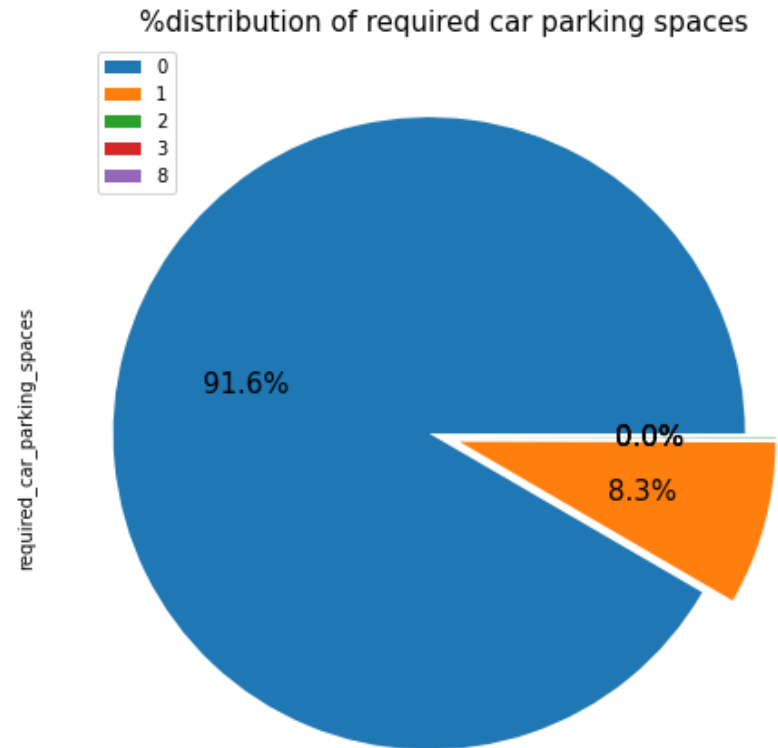
Percentage Distribution Of required car parking spaces (Univariate)

- This chart shows that 91.6% guests did not required the parking spaces. only 8.3% guests required only parking spaces
- yes that insight definately help the hotel to provide better services.it can be said that hotels need to work less on car parking spaces as only 1 car parking spaces were required by 8.3% guest. so it is better to have a limited car parking spaces and use that spaces for other important purpose rather than just leeting go unused. so it is better to focus on other areas to increase quality of hotel rather than focusing mainly on car parking area only. the demand of car parking area is less . this might be the reason as too many guests prefer to use public vehicles to travel.

Percentage Distribution Of required car parking spaces (Univariate)

AI

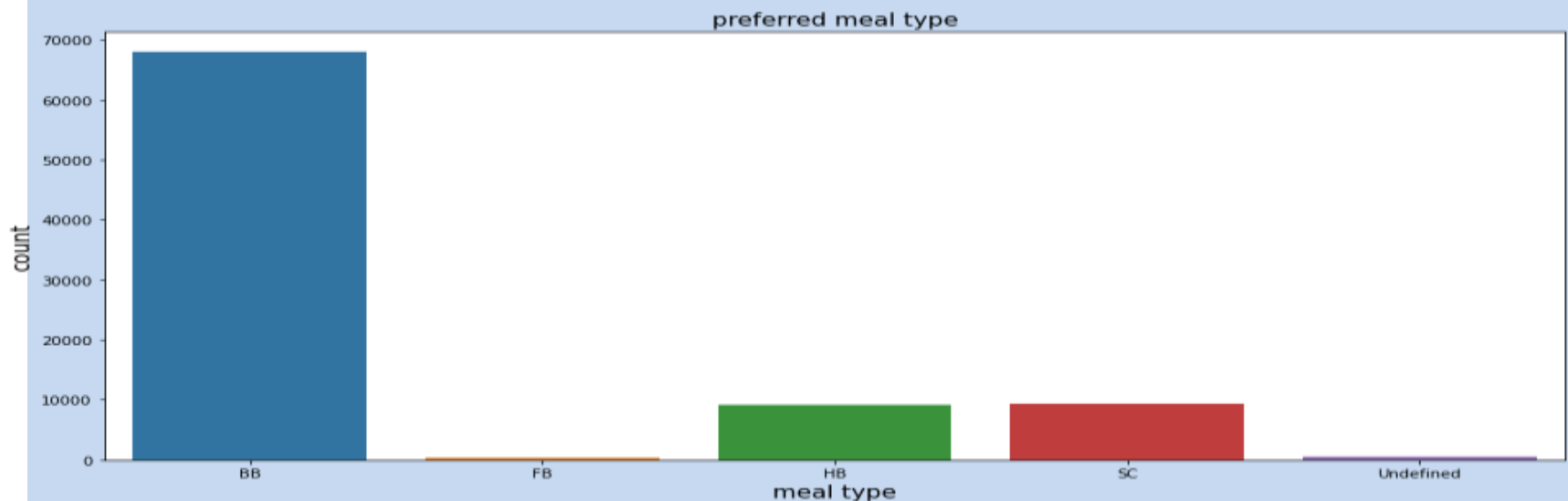
- This chart shows that 91.6% guests did not required the parking spaces. only 8.3% guests required only parking spaces
- yes that insight definately help the hotel to provide better services.it can be said that hotels need to work less on car parking spaces as only 1 car parking spaces were required by 8.3% guest. so it is better to have a limited car parking spaces and use that spaces for other important purpose rather than just leeting go unused. so it is better to focus on other areas to increase quality of hotel rather than focusing mainly on car parking area only. the demand of car parking area is less . this might be the reason as too many guests prefer to use public vehicles to travel.



Meal Type Distribution(Univariate)

AI

- Inside the chart we found the meals which is most preferred by the guests is BB(Bed and Breakfast) while HB(Half board) and SC (self catering) are equally preferred. types of meals are as follows:
- BB-(bed and breakfast) HB-(Half Board) FB-(Full -Board) SC -(Self- Catering)
- So, the insight has positive impact of hotels needs to focus more on the BB meal type so that the majority of customers are satisfied while others types of meals should be given equal importance with proper management of food services so as the guests we need to give best offer or services to customers.



Mostly used for Distribution Channels and relationship of Distribution channel and ADR

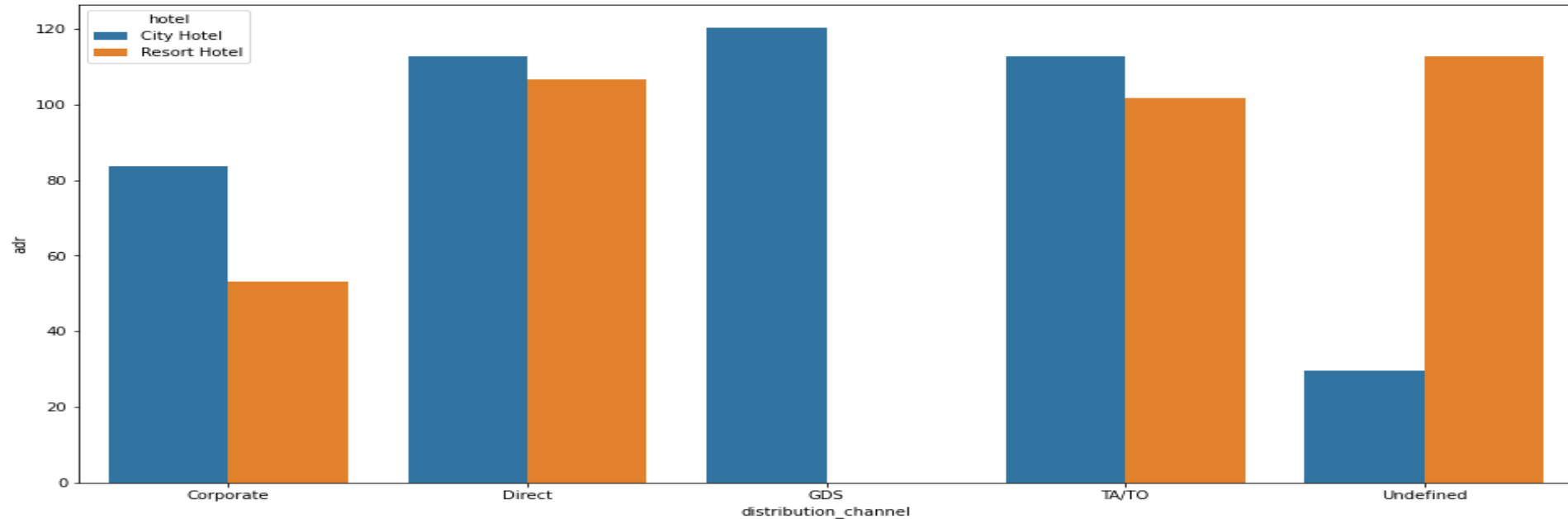
AI

- From the above 1st Chart, we have found that 'TA/TO' has been mostly (79.1%) used for booking hotels. direct constitutes of 14.9%. Corporates Constitutes of 5.8% GDS Constitutes of only 0.2% and rest unidentified are 0.
- From the above 2nd Chart it is clear that 'direct' and 'TA/TO' have almost equally contribution in ADR on both type 'hotels' While GDS has highly Contributed in ADR in both type of 'city hotels' type. GDS needs to increase Resort Hotel Bookings. Corporate these are Corporate hotel Booking Companies which makes booking possible. the definition of abbreviation used in this graph are as follows:
- GDS-A GDS is a worldwide contact between travel bookers and suppliers, Such as hotels and other accommodation Providers. It communicates live product, price and availability Data to Travel Agents and online Booking Engines, And Allows for automated
- transaction.
- Direct Means that Bookings are Directly Made with the respective Hotels.
- *TA/TO means that Bookings Are through travel agents and travel operator.
- Undefined Bookings are undefined it may be the reason that customers made their bookings on arrival.

mostly used distribution channel for hotel bookings



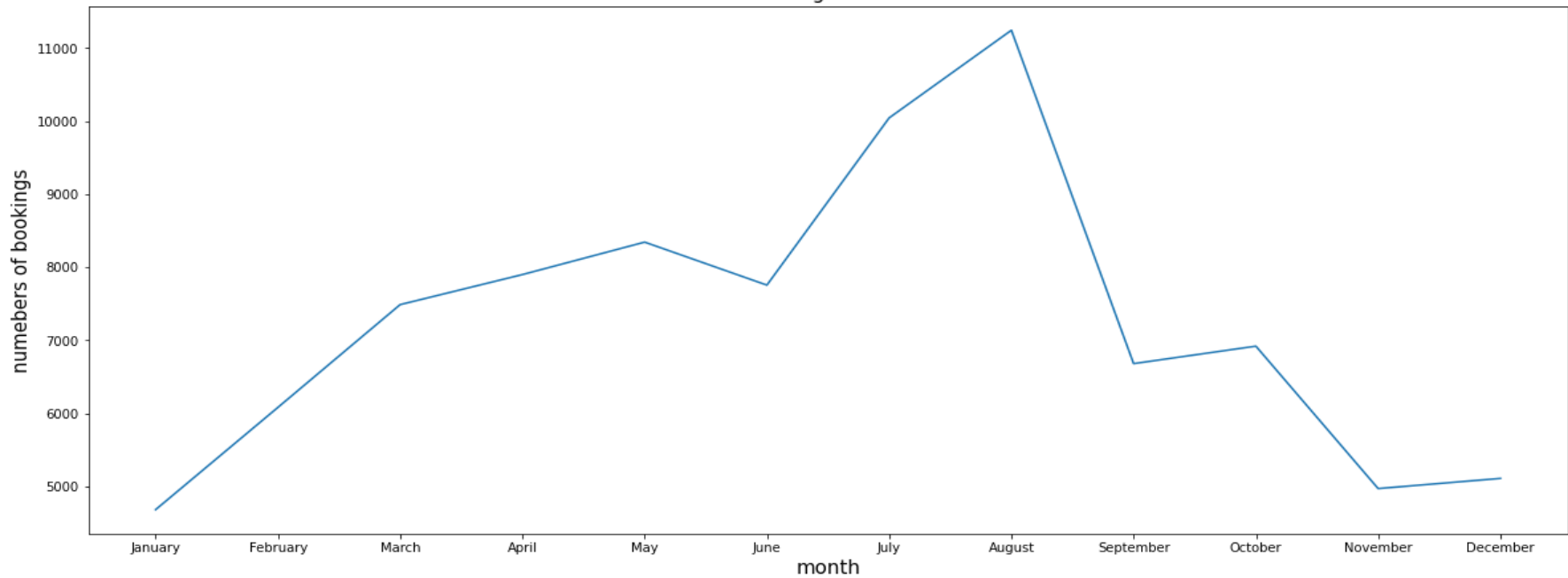
ADR across Distribution channel



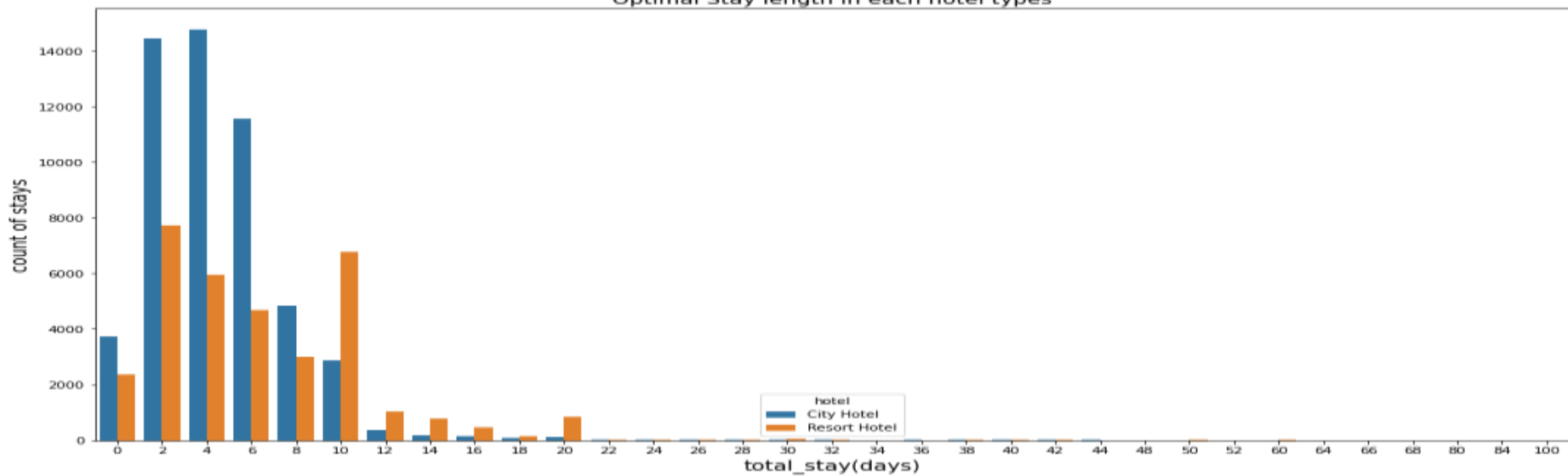
Booking By month and optimal stay length in hotels

- From the first chart found most numbers of booking in the month of july and august. as august and july is come near to the summer vacation, so the summer vacation the reason for the most numbers of bookings.
- While From the Second Chart Givess diffrent insight, so from the above observation we have found the optimal stay in both type hotel is less than 7 days so after staying number is drastically Declined.

numbers of bookings across each month

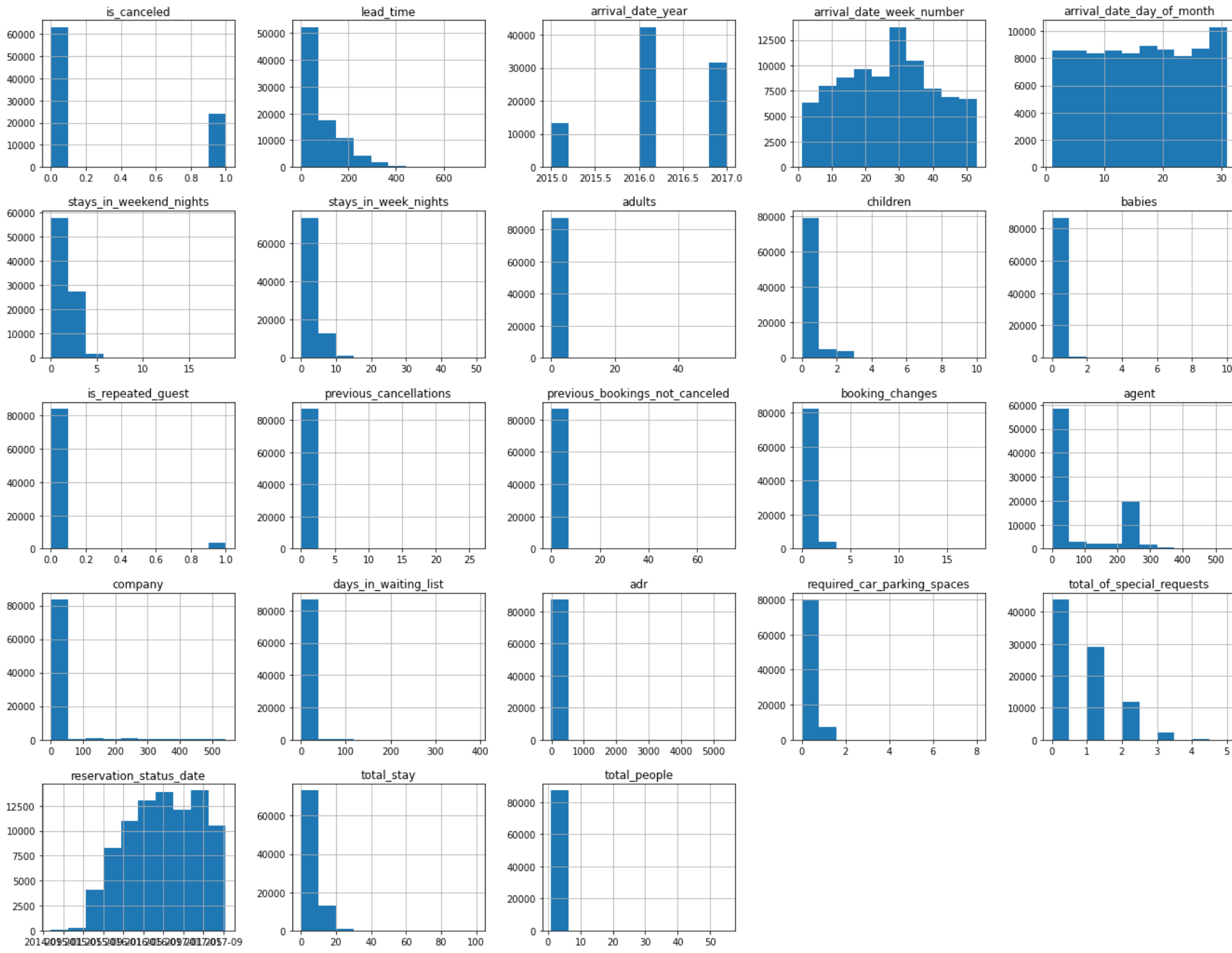


Optimal Stay length in each hotel types



Plotting Histogram

- some insights are found from the chart are as follows:
- we can see maximum guest come in the year of 2016.
- maximum week number is 30.
- maximum arrival happen in the last of the month.
- maximum guest come with no children.
- there is ver less requirement of car parking.

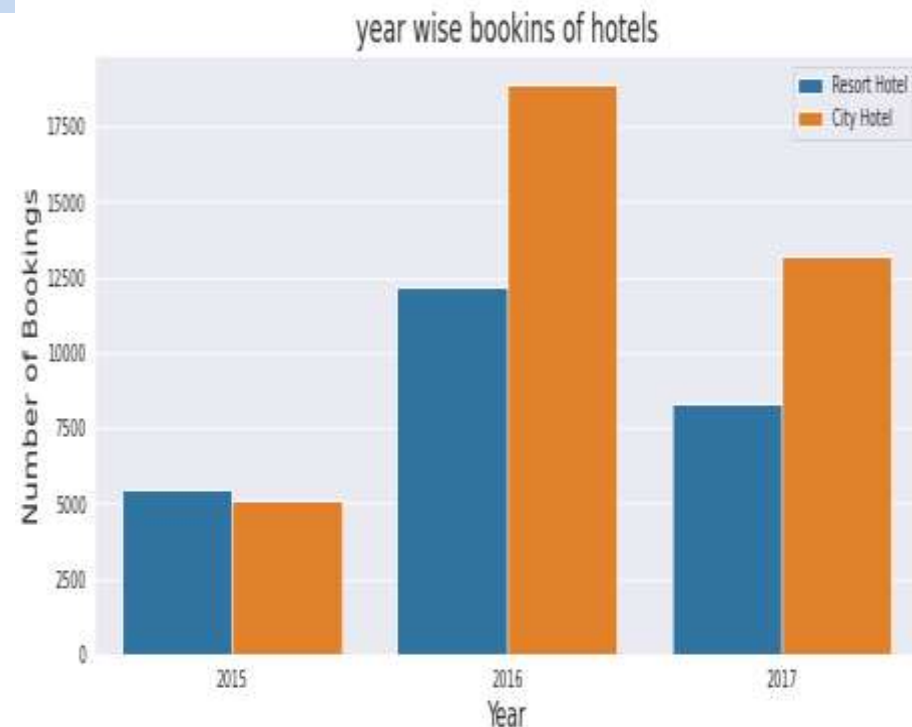
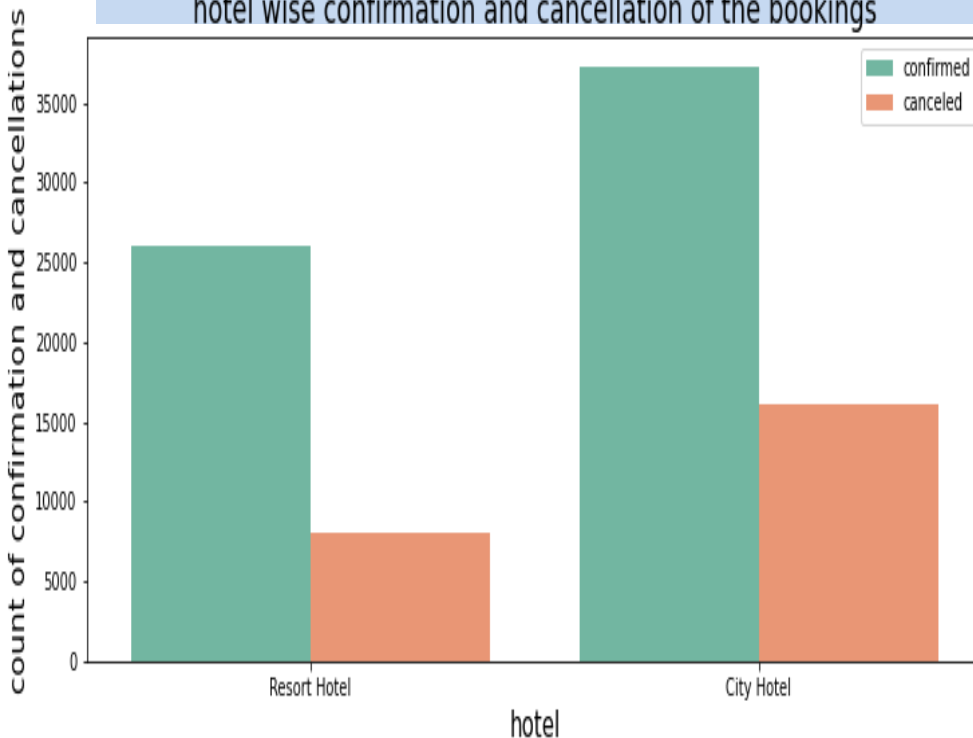


Year and Hotel wise confirmed booking and cancellation Distribution

AI

- we can clearly see above the graph that having city hotel is having greater number of bookings as compared to Resort Hotel.
- From the above graphs, it can be summarized that the year 2016 both the hotel saw massive increases their Booking and by far the year 2016 is the year of highest booking of both hotel. In each Year that is 2015, 2016 and 2017 the city hotel is having the highest number of bookings.

hotel wise confirmation and cancellation of the bookings

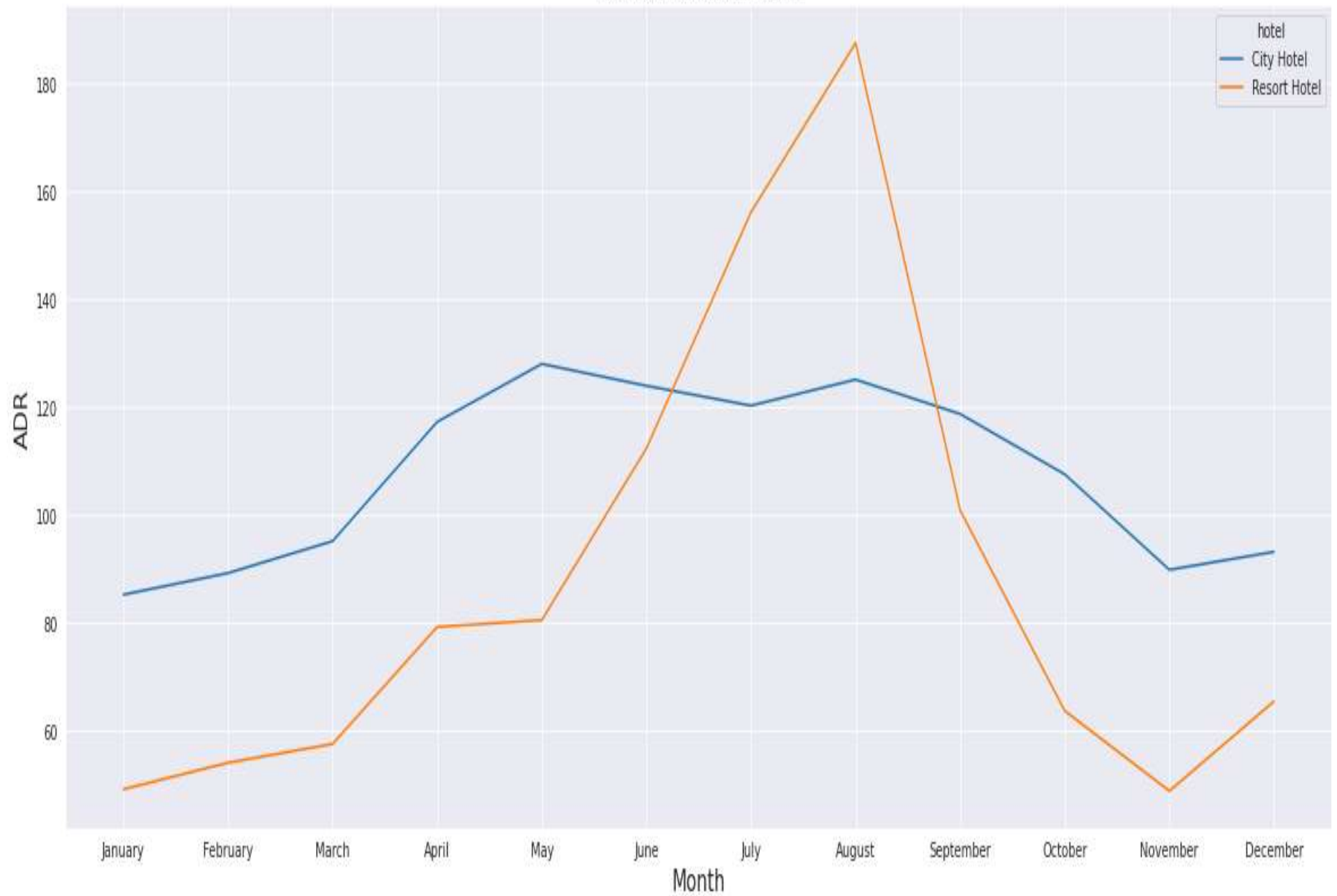


ADR Across Different Month

AI

- for Resort Hotel, ADR is high in the months of June, July, August as compared to city Hotels. The Reason may be that Customers/people want to spend their Summer Vacation In Resorts Hotels.
- Best Time for guest to visit Resort or City Hotel is January, February, March, April, October, November, and December as the average daily rate of these months is very low. so it would be feasible and sustainable.
- So, the Higher ADR, The Higher will be revenue so it's a good sign. Hotels should work more to enhance their ADR by offering good schemes to attract customer in winter vacation also and on other holidays.

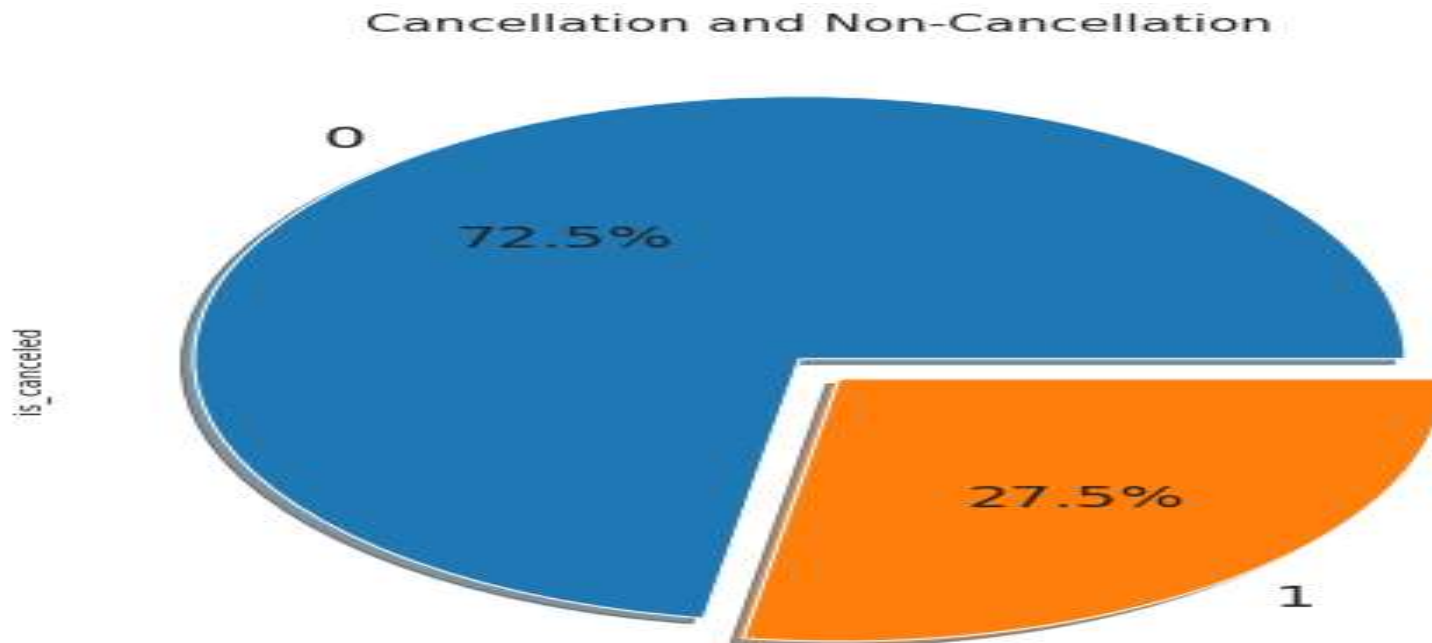
ADR across each month



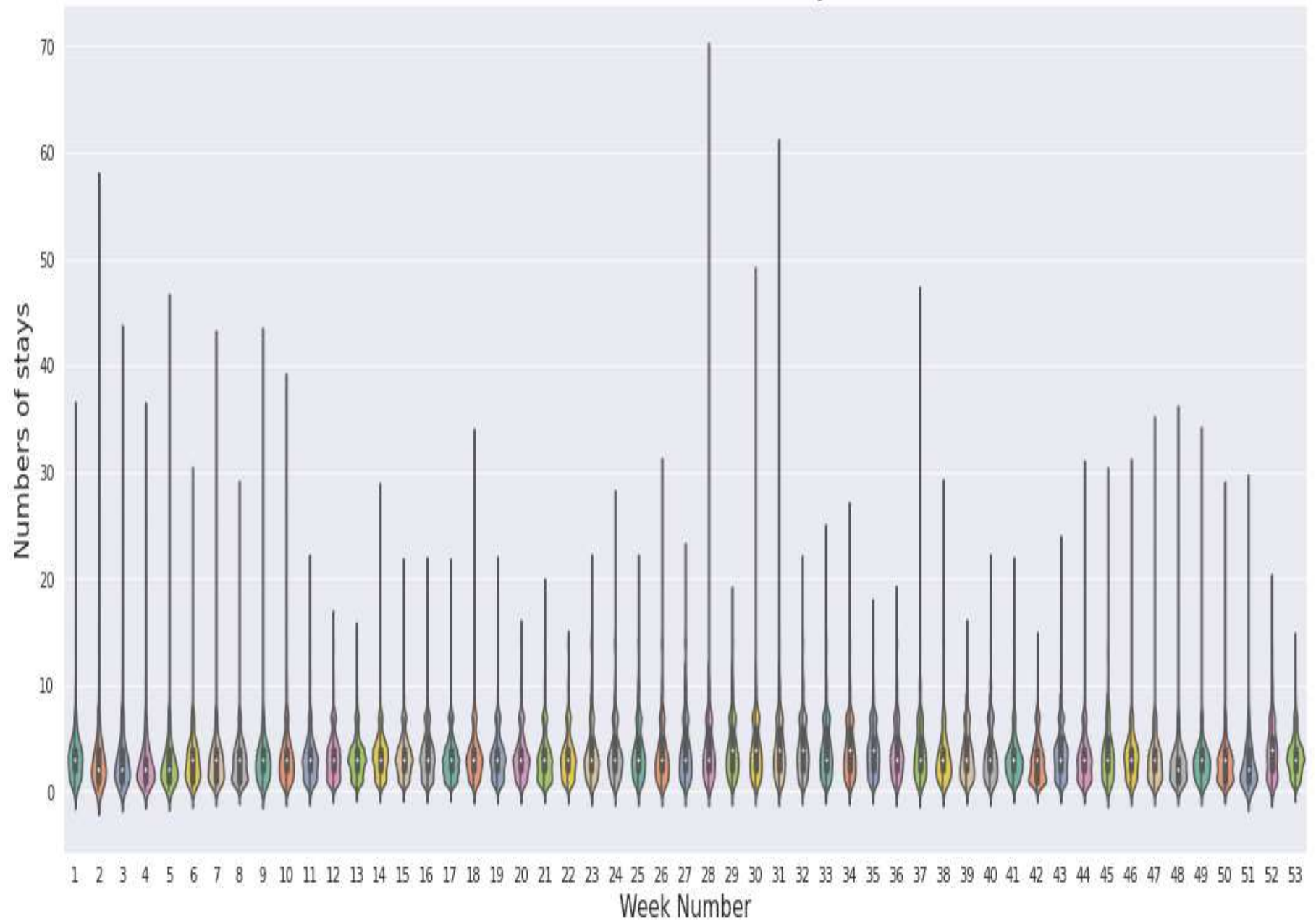
Weekly Stay Distribution and Calculation and Non-Calculation

AI

- From the above Violin plot, we have found that from the week 28 and week 31, it is shown the highest days of stay whereas from the week 1 to 11 has shown a very steady trend in the number of stays and also the week 18 to 22 has the least number of stays by the visitors in aggregate of all 3 years 2015, 2016 and 2017.
- From the graph, we have found insight the more than 1/4th of the overall bookings i.e. approx 27.5% of the tickets was got cancelled.



Week wise number os stays



Room type preference and Customer Types

AI

- From the first Count Plot We got data most prefer Type A so majority of guest have shown interest to this Room.
- From the above graph it can be measured that the transients type of customers visits the most comes in the category of least visitors.
- yes From the graph we got positive impact because 'A','D','E' Most Preferred due to better service offer by in room type. So Overall Booking in a hotel Manner.
- off course the better understanding regarding the different type guests will help to take proper right steps towards services facilities, requirement and offers which will directly results in the growth in business.

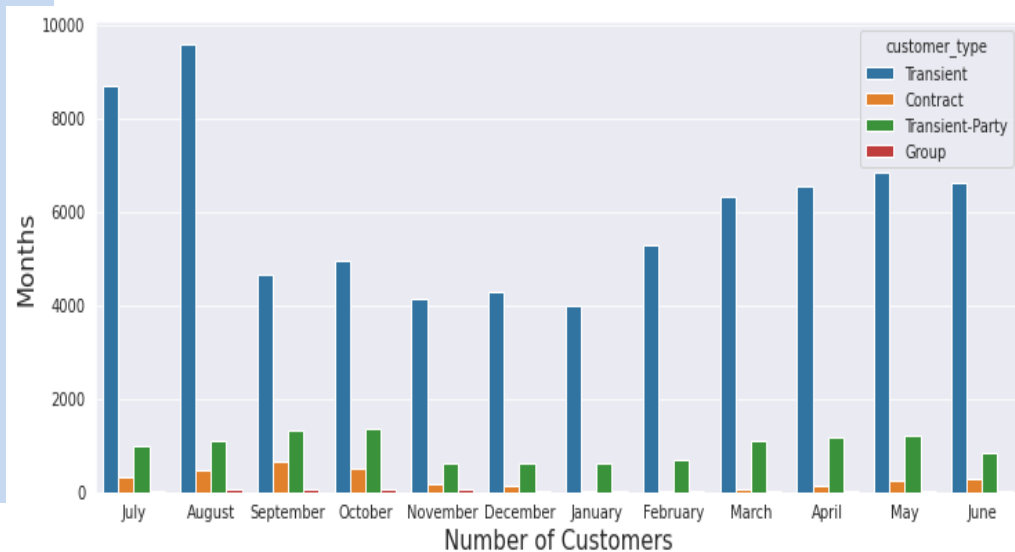
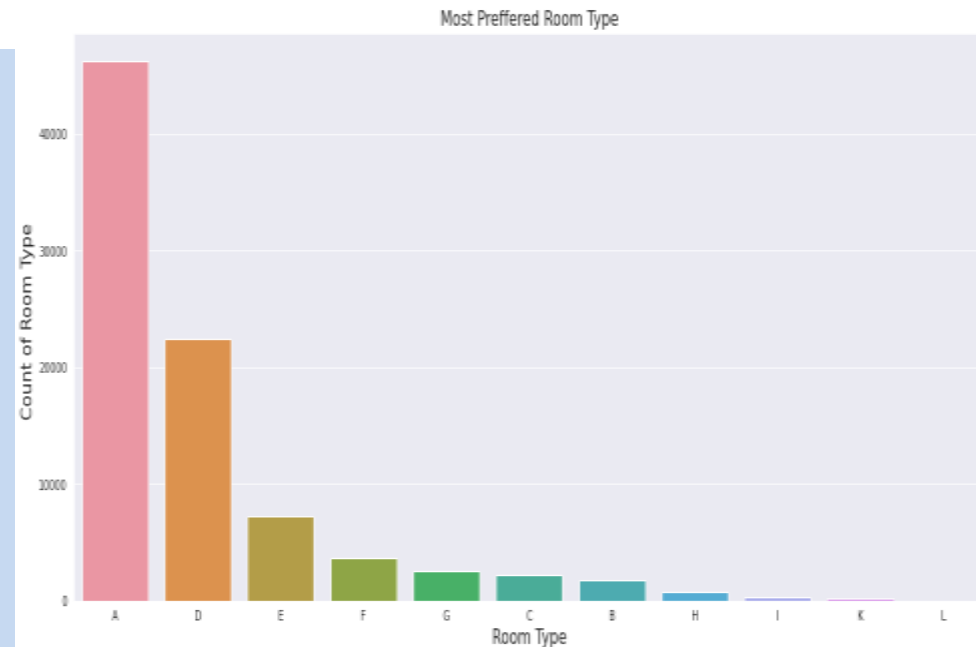
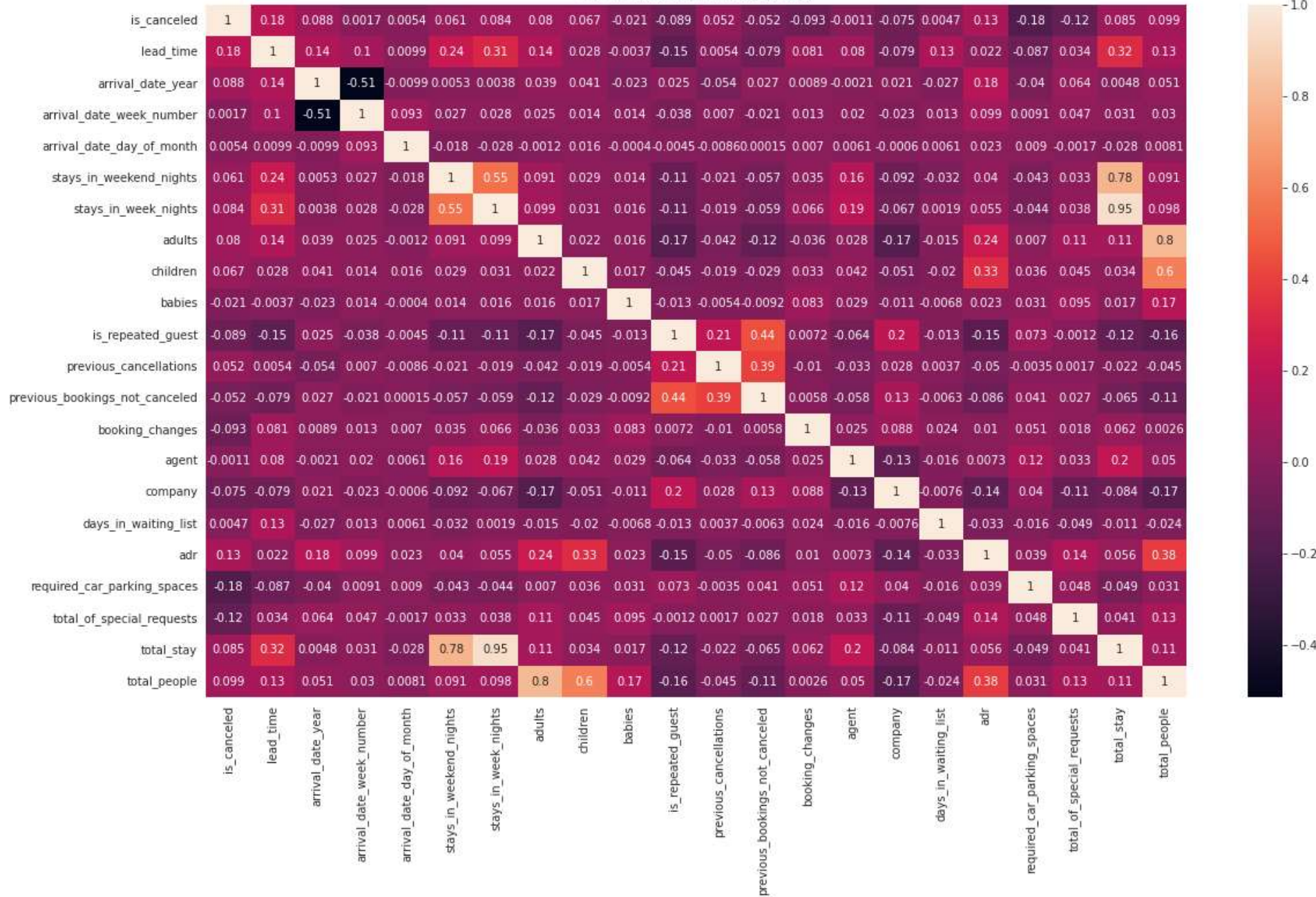


Chart Correlation Heatmap

AI

- The Insight found from the charts are as follows:
- - is_canceled and total stays are negative correlated. This customer is unlikely to cancel their booking. If they don't get the same as they reserved room - lead_time and total_stay is positively correlated. This means more the customer more will be lead time - adult, childrens and babies are correlated with each other. This indicates more the people more will be adr - is_repeated guests and previous booking not_cancel have a strong correlation. This may be due to the reason that repeated guests not more interested to cancel their Bookings. - so these are some more powerful insight found from the chart of correlation heatmap.

Co-relation of the columns



SUGGESTION FOR THE CLIENT TO ACHIEVE BUSINESS OBJECTIVE

AI

- Business Objective attained as follows:
- To Attain high growth and more success, hotel business need to flourish and for that few things which we need to consider high revenue generation, customers satisfaction and employee retention.
- We are able to achieve the same by showing the client which are the month is high revenue generation by using various charts and graphs distribution.
- Enhancing the revenue adopted by bar chart distribution of which type room are most preferred and reserved and which are the months suitable for visitors
- we also have founded the various preference in different categories like meal type, optimal stay length, facilities require by customers like car parking spaces,etc. so this all insight need to better planning for growth and revenue.
- So preparing well by understanding these all the outcomes, the client can be well prepared in advance so that minimum grievances would be faced by clients in long run and would help in further enhancement of their hospitality and service.
- Ask feedback often from the guests visiting the hotels so the quality can be upgraded to the next level to increase more guests.
- Periodically throw offers to attract the old customers so as to increase the no of repeated guests.

CONCLUSION

AI

- City Hotel are most proffered hotel by the guest, So we can say that City Hotel is busiest hotel in comparison of the resort hotel.
- The average ADR of city hotels is higher as compared to Resort hotel, so it can easily said that city hotel is generating more revenue than resort hotels.
- the total number of stay is directly proportional to adr.so the number of stay is higher then higher will be adr and revenue as well.
- The percentage of repeated guest is very low. only 3.9% people revisited the hotels and rest 96.1% were new guests. So the retention rate is much low
- Among the different types of meals,BB(bed and breakfast) is the most preferred meals by the guests. So the guests is loved to opt. this type meals.
- Direct and TA/TO have equally contribution in adr in both types of hotel , City Hotels and Resorts Hotels.
- Optimal stay length in both hotels (City hotels and Resort Hotels) is less than 7 days. Usually people stays for a week after 7 days optimal stay length is declined drastically.
- most numbers of bookings have taken place in the month of July and august.
- The Mostly Used distribution channel for bookings is TA/TO is 79.10% Booking were made through TA&TO(travel agents and Tour Operators)
- 1/4th of the total booking is cancelled.Approx 27.5% have got cancelled out of all bookings.
- Majority of guests have shown interest in the room type 'A'. Room type 'A' is the most prefferd Room type.