

Capstone Project - 1

EDA on Hotel Bookings

By

Gaurav Saini

Lini Saxena

Ujjwal Ajit Singh

(Cohort Hudson)

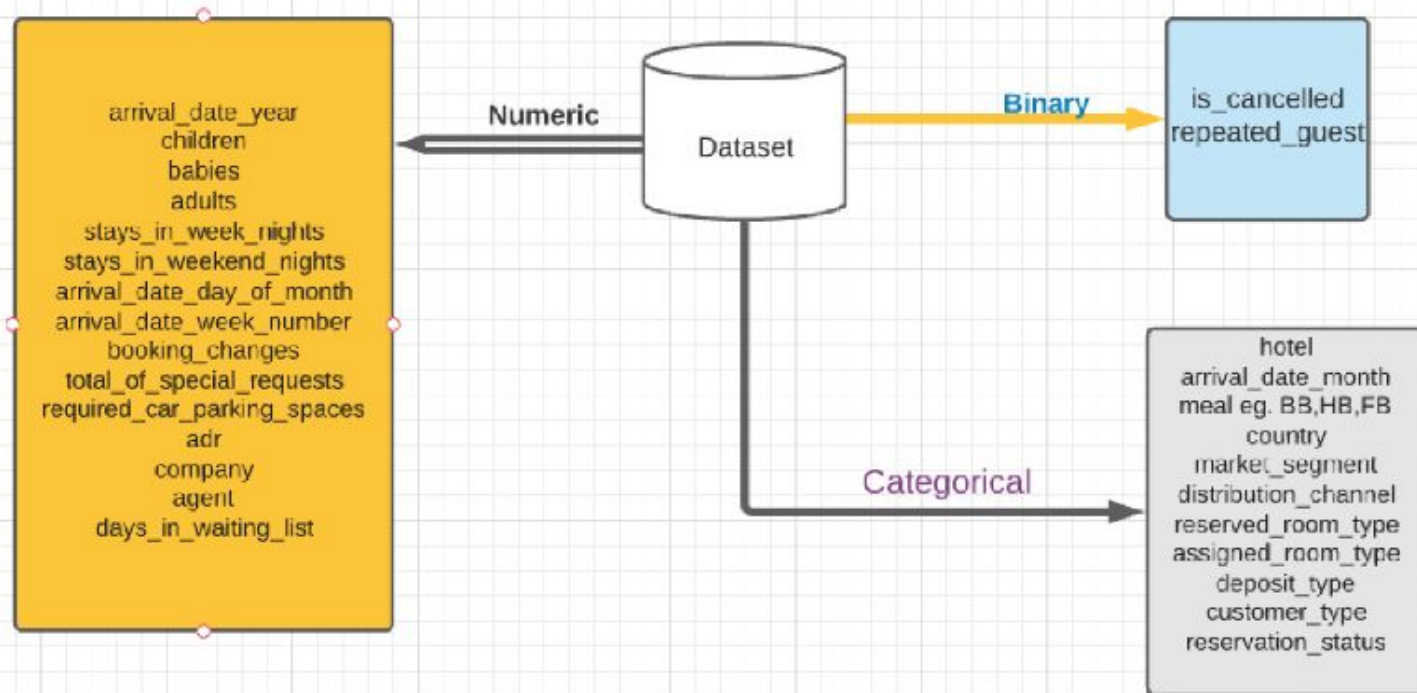
Challenges Faced during data cleaning



- 1. There were Null data in the columns 'children', 'company', 'agent', 'country', which needed to be handled.***
- 2. There were duplicate rows present in the dataset which needed to be handled.***
- 3. The Data type of columns 'reservation_status_date', 'children', 'agent' was not appropriate.***
- 4. We had to add new columns such as 'month', 'total_guests', 'stays_in_total_night'.***
- 5. There was an outlier in the column 'adr' which was needed to be removed.***
- 6. There were some fake bookings/false entries which showed total number of guests is zero.***

Data Exploration

Data Summary



Data Summary

The summary of our findings and our understandings about various columns after the data exploration of dataset is as follows:

This data in the Dataset is of hotel bookings of the year 2015, 2016 and 2017. it contains 32 columns, the description of the data present in the columns are as follows:

- **hotel:** Type of hotel (City or Resort).
- **is_canceled:** Whether the booking is canceled or not (0 for not canceled and 1 for canceled).
- **lead_time:** Time in days between booking transaction and actual arrival.
- **arrival_date_year:** The year of arrival ([2015, 2016, 2017]).
- **arrival_date_month:** The month of arrival.
- **arrival_date_week_number:** The week number of the year of arrival date.
- **arrival_date_day_of_month:** The day of month of arrival date.
- **stays_in_weekend_nights:** Number of nights spent in a hotel on weekends.
- **stays_in_week_nights:** Number of nights spent in a hotel on Weekdays.

Data Summary

- **adults:** Number of adults in single booking record.
- **children:** Number of children in single booking record.
- **babies:** Number of babies in single booking record.
- **meal:** Type of meal chosen [BB-Bed and Breakfast, FB- Full Board(Breakfast Lunch and Dinner), HB-Half Board(Breakfast and Dinner), SC-Self Catering]
- **country:** Country of origin of customers.
- **market_segment:** By which segment was booking made and for what purpose [Direct, Corporate, Online TA(Travel Agents), Offline TA(Travel agents)//TO(Tour Operators), Complementary, Groups, Undefined, Aviation].
- **distribution_channel:** Via which medium booking was made(Medium type=[Direct', Corporate, TA(Travel agents)/TO(Tour Operators), Undefined, GDS]).
- **is_repeated_guest:** Whether the customer has made any booking before(0 for No and 1 for Yes).
- **previous_cancellations:** Number of previous canceled bookings.

Data Summary

- **previous_bookings_not_canceled**: Number of previous non-canceled bookings.
- **reserved_room_type**: Room type reserved by a customer(Room Type=[C, A, D, E, G, F, H, L, P, B]).
- **assigned_room_type**: Room type assigned to the customer (Room Type=[C, A, D, E, G, F, H, L, P, B]).
- **booking_changes**: Number of of booking changes done by customers.
- **deposit_type**: Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund).
- **agent**: Id of agent for booking.
- **company**: Id of the company making a booking.
- **days_in_waiting_list**: Number of days on waiting list.
- **customer_type**: Type of customer(Transient, Contract, Transient-Party, Group).
- **adr**: Average Daily rate=(Rooms revenue/Total rooms sold).

Data Summary

- **total_of_special_requests**: total number of special request.
- **reservation_status**: Whether a customer has checked out or canceled, or not showed.
- **reservation_status_date**: Date of making reservation status. ""

Exploratory Data Analysis

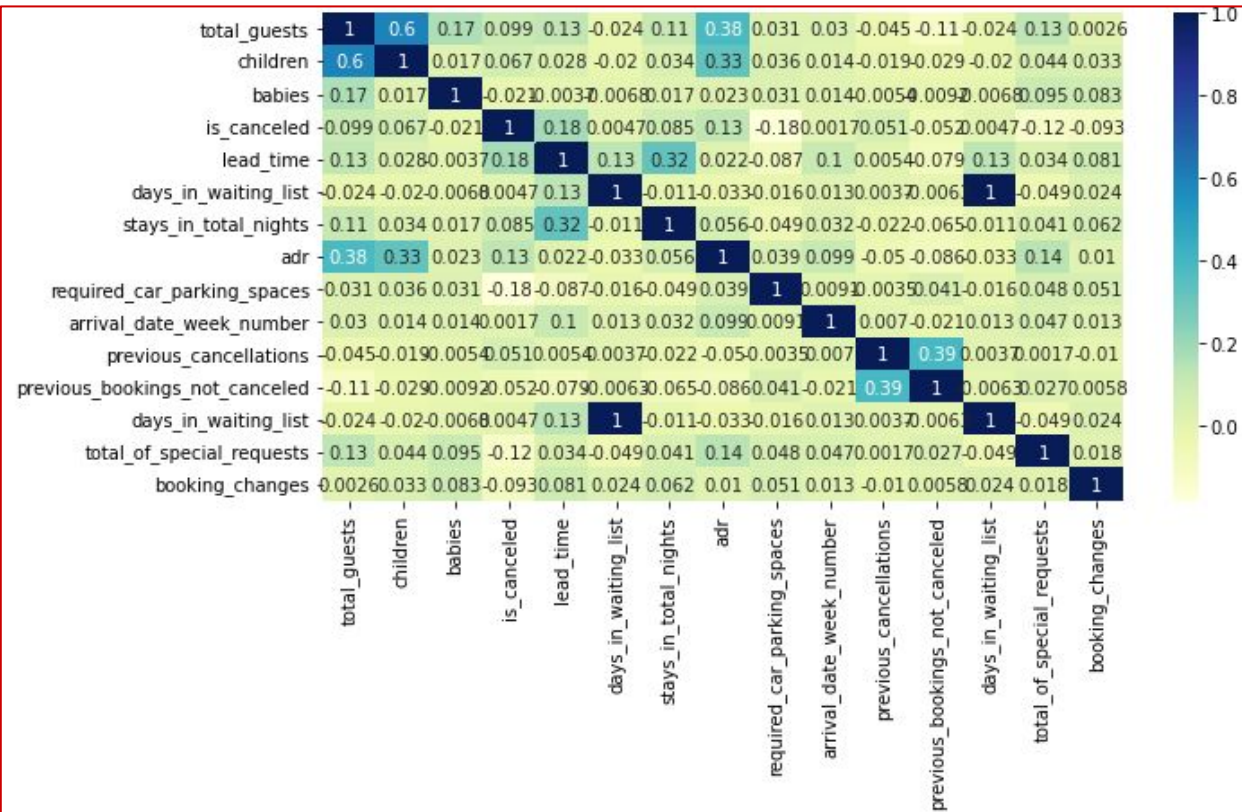
- 1. Correlation Heatmap**
- 2. Univariate Analysis**
- 3. Hotel Type Analysis**
- 4. Distribution Channel Analysis**
- 5. Timewise Analysis**
- 6. Nationality Analysis**
- 7. Optimal Length of stay to get best daily rate**
- 8. Special request scenario**

Graphs used for analysis

1. **Line Chart**
2. **Bar/Column graph**
3. **Scatter Plot**
4. **Pie Chart**
5. **Histogram**
6. **Box Plot**
7. **Correlation Heatmap**

Correlation Heatmap

Correlation between columns

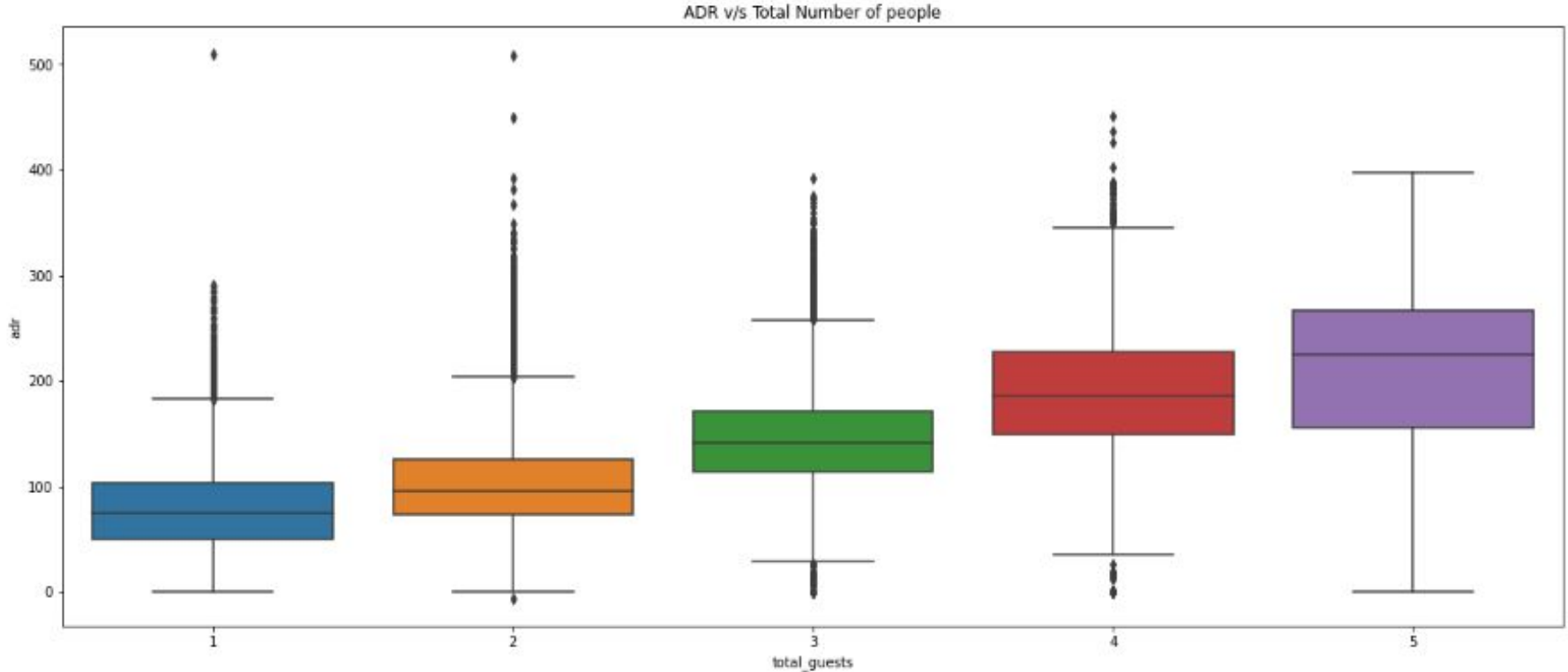


Conclusions:

- **ADR and total people are highly correlated**
- **stays_in_total_nights and lead_time have slight correlation. This shows that long stays or vacations are planned way before the arrival.**
- **is_canceled and same_room_alloted_or_not are negatively correlated. That means customer is unlikely to cancel his bookings if he don't get the same room as per reserved room. We have visualized it above.**

Univariate Analysis

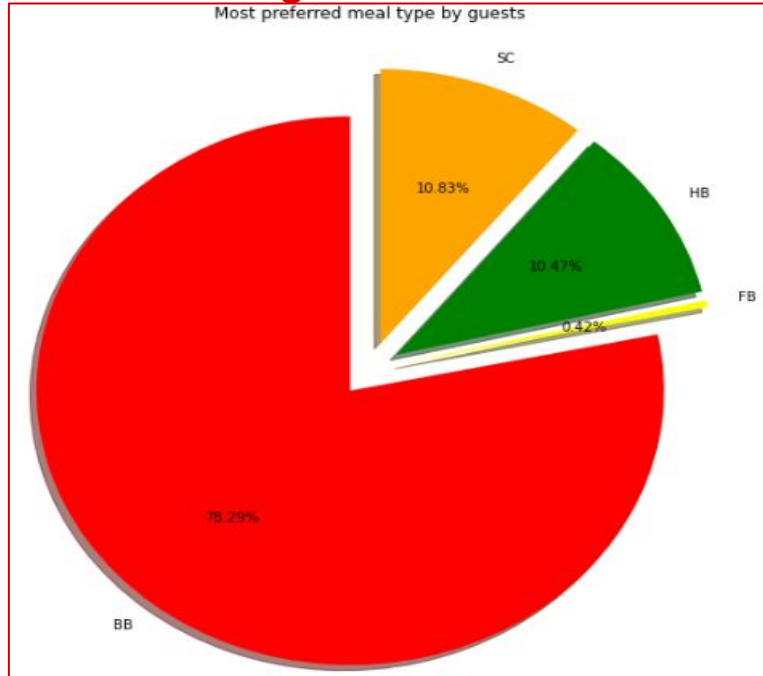
Relationship between ADR & Total No. of Guests



Conclusions:

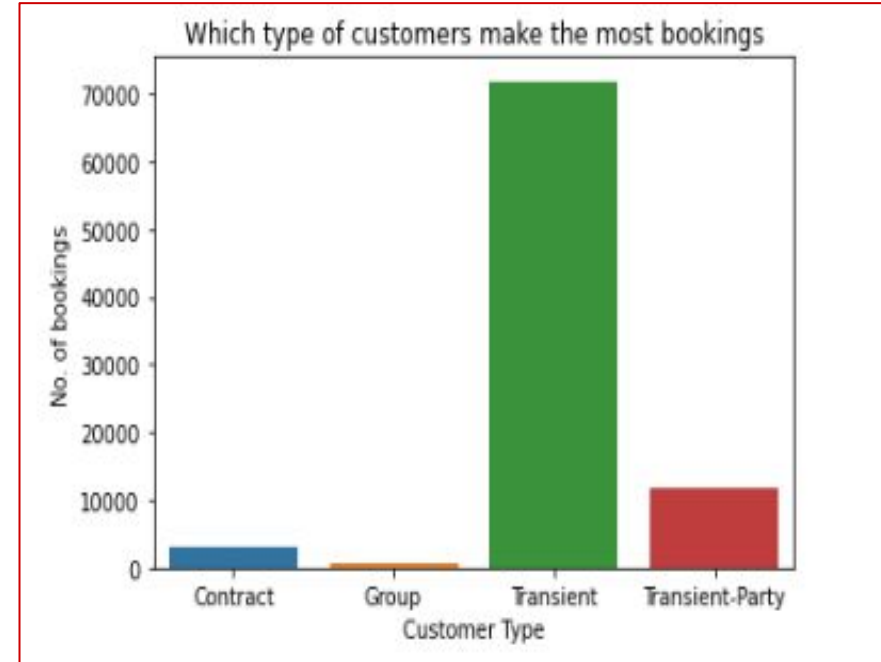
ADR and total_people are positively correlated which is true as more number of people means more booking hence more revenue therefore more profit.

Which meal type is preferred more by guests? & which type of customers make most number of bookings?



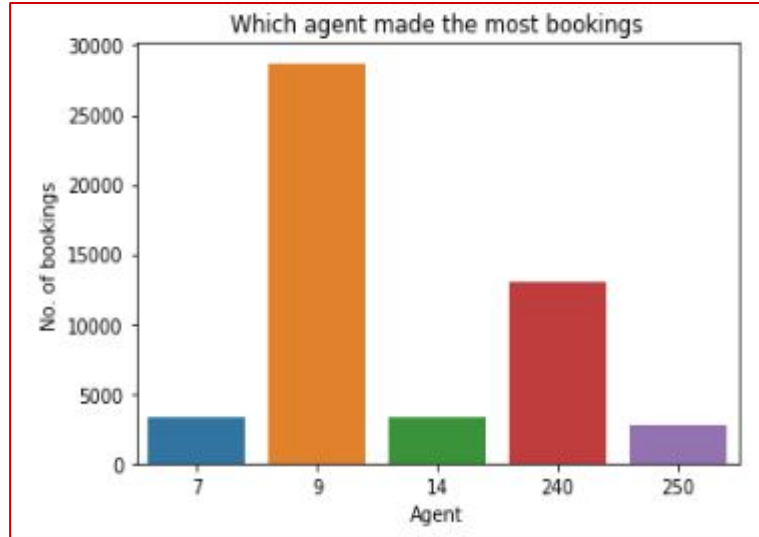
Conclusions:

- **BB-Bed and Breakfast, FB- Full Board(Breakfast Lunch and Dinner), HB-Half Board(Breakfast and Dinner), SC-Self Catering**
- **The meal type 'Bed and Breakfast' is preferred the most by the guests.**



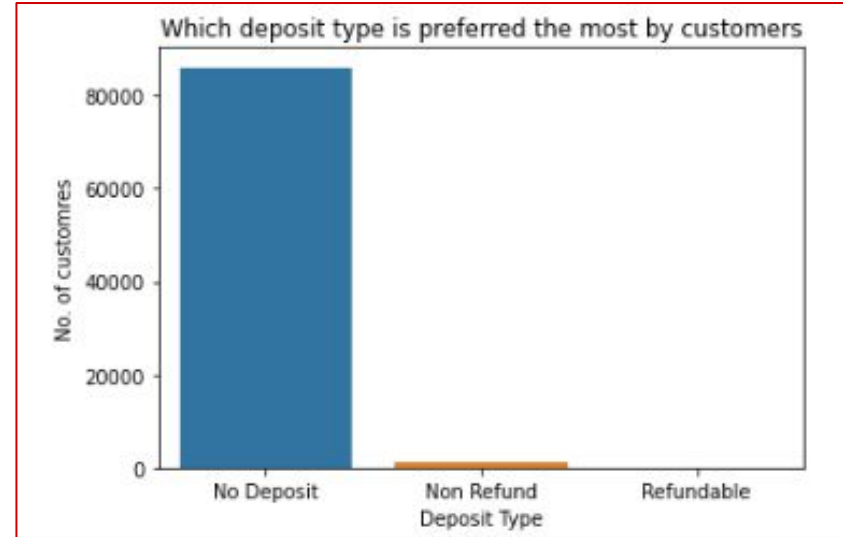
We can see that 'Transient'(Lasting or continuing for a short period of time) customer type made the most number of bookings.

Which agent made the most number of bookings? & Which deposit type is preferred more by the customers?



Conclusion:

Agent ID 9 made the most number of bookings followed by Agent ID 240.

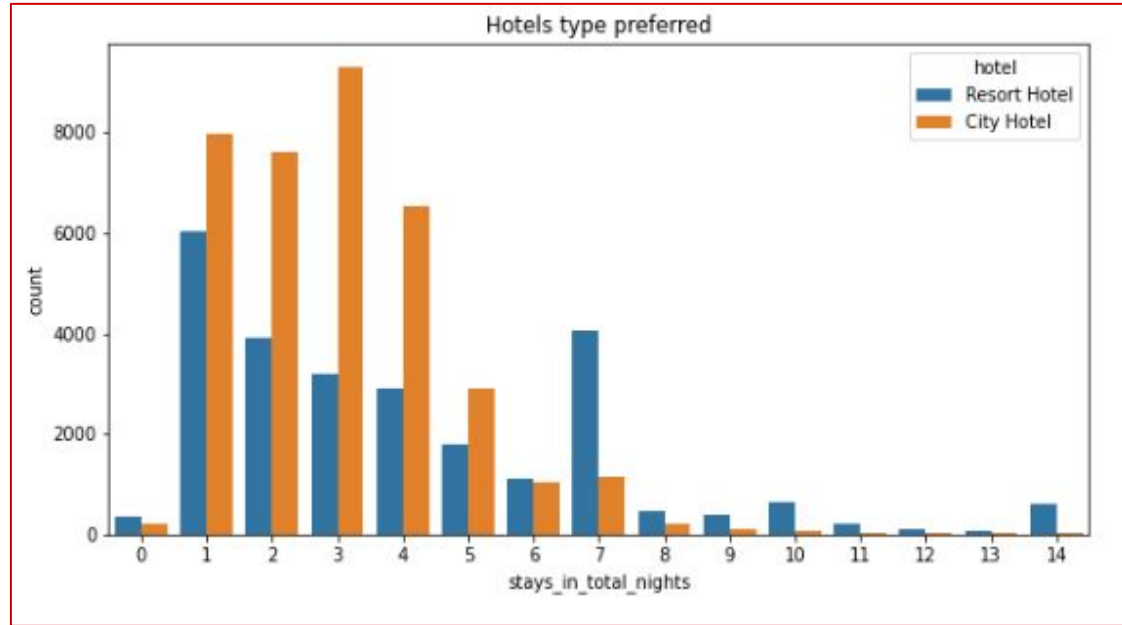
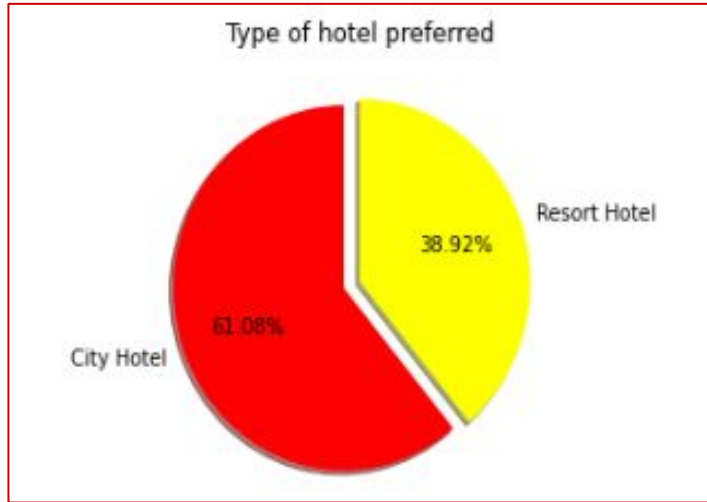


Conclusion:

Almost all customers preferred 'No Deposit' type booking.

Hotel type analysis

Which type of hotel is mostly booked by guests & relationship between Hotel Type and Total Stay

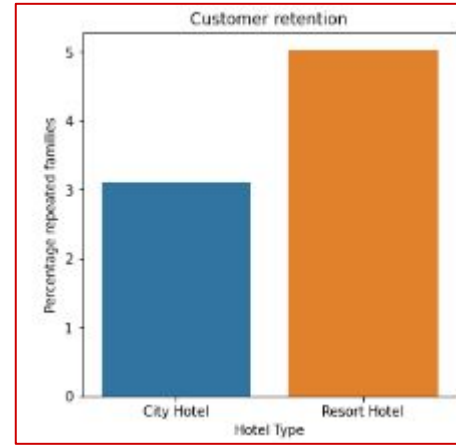
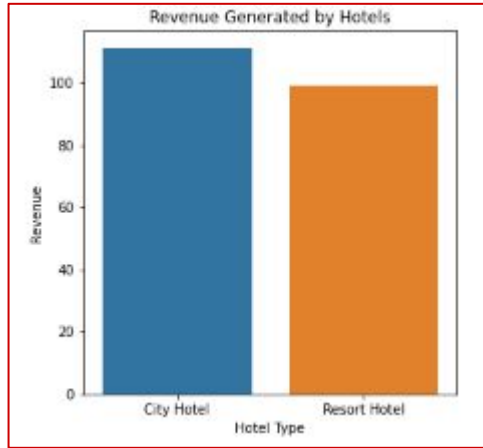


Conclusions:

As we can see that City Hotel has more bookings(almost 1.5 times) as compared to Resort hotel.

1. Customers are preferring City Hotel for shorter stays(less than 6 days) & Resort Hotels are preferred for longer stays(more than or equal to 6 days).
2. For the stay of exactly 7, 10 or 14 days (especially 7 days) customers prefer the Resort hotels (May be for mini vacation or to celebrate functions like engagements or marriages, office success parties or other close intimate events.).

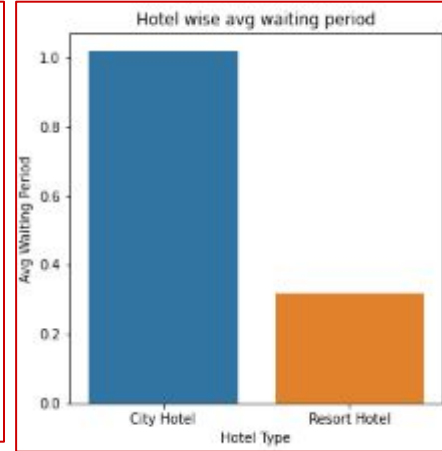
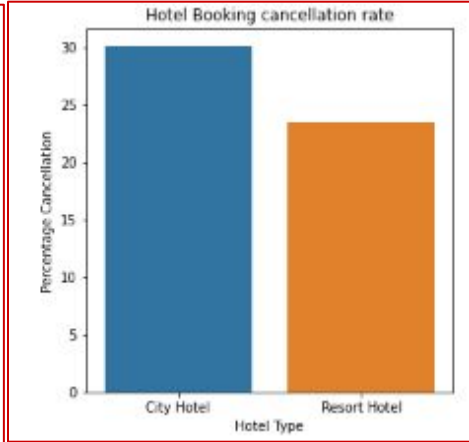
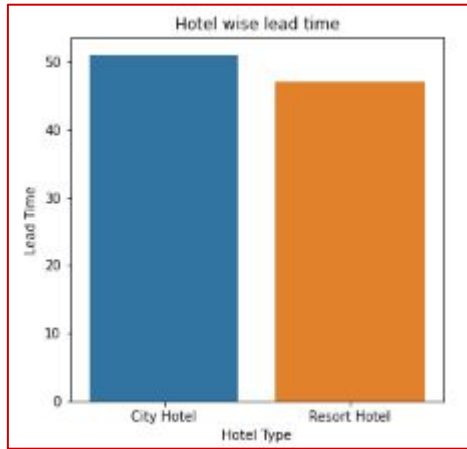
Which hotel generates more revenue? & Which hotel type has higher number of repeated guests?



Conclusions:

- *The City Hotels are generating more revenue in comparison to Resort Hotels.*
- *Although City Hotels are booked almost 1.5 times more than Resort Hotels, the difference in revenue of City Hotels and Resort Hotels is very less as people prefer resort hotels for longer stays and as we have assumed that Resort hotels are preferred for close intimate events which increase the revenue of the Resort Hotels.*
- *As we can see customer retention of Resort Hotel is slightly more than City Hotel one of the reason can be as we have assumed that City hotels are used more for Casual stays but the Resort hotels are more used for intimate events.*

Which Hotel has high lead time? Which type of hotel has higher waiting period? Which type to hotel has high booking cancellation rate? & is booking cancellation rate correlated with Lead time and waiting period?

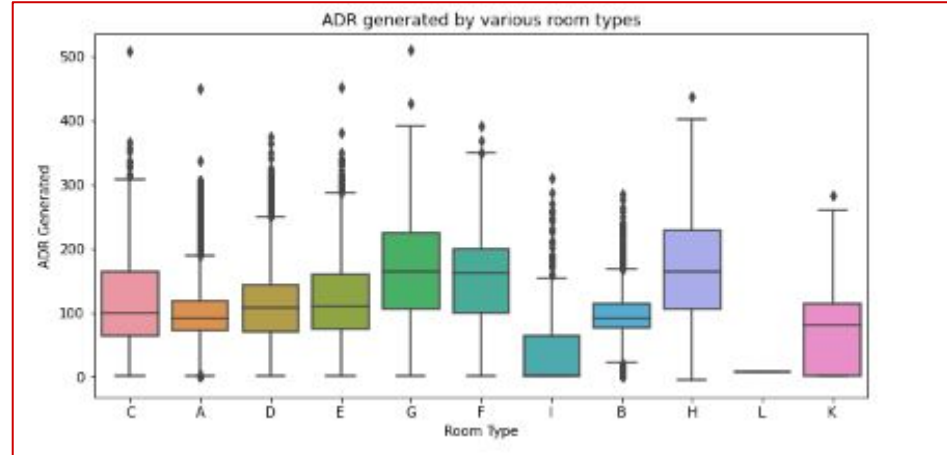
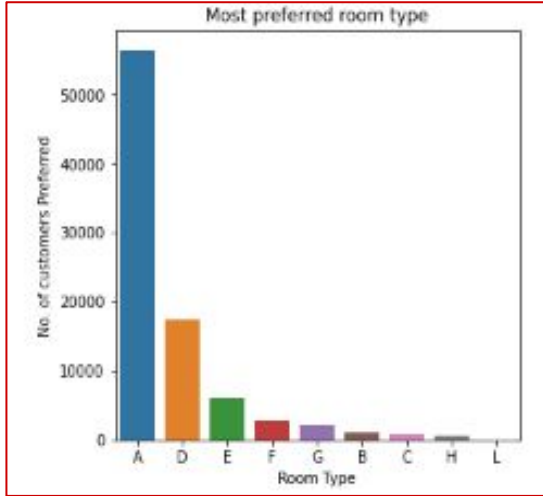


Hotel cancellation is not related with waiting period or lead time. Thus our Correlation heatmap is relevant.

Conclusions:

- **Lead time of both type of hotels are more or less similar, so from the above graph we can conclude that customers prefer to book hotel approximately 45 days before the check in.**
- **City Hotels have more waiting period hence City hotels are much more busy in comparison to Resort Hotels.**
- **The booking cancellation rate of City hotel is approximately 30% & for Resort hotels it is approximately 24%**
- **Hotel cancellation is not related with waiting period or lead time. Thus our Correlation heatmap is relevant.**
- **City Hotels are preferred more for casual stays and Resort Hotels are preferred more for events hence cancellation is more in City Hotels.**

Which room type is more in demand? Which room type generates the higher ADR?

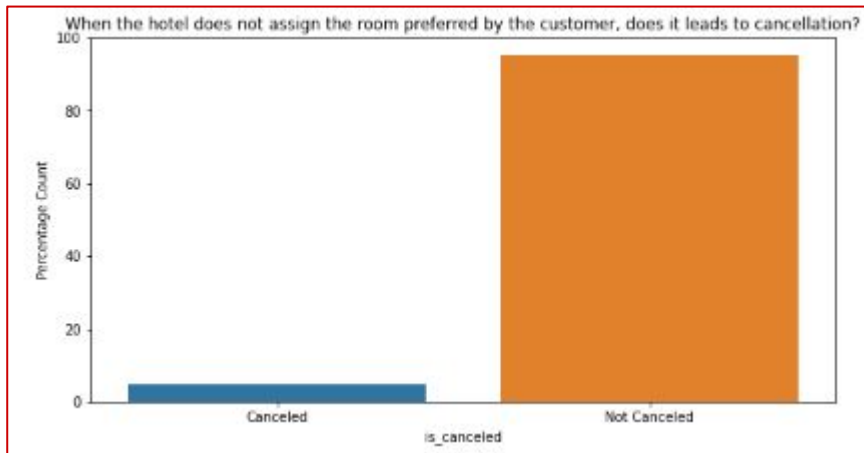


Conclusions:

- Customers prefer Room type 'A' the most followed by room type 'D' and room type 'E'.
- The Room type 'H' generates the max adr followed by room type 'G', 'F'.
- Room type 'A' is most preferred in economy class, followed by room type 'D'.
- Room type 'H', 'G', 'F' are preferred very less by customers and yet create high adr so we can assume that they fall in luxury class and very less customers can afford it.

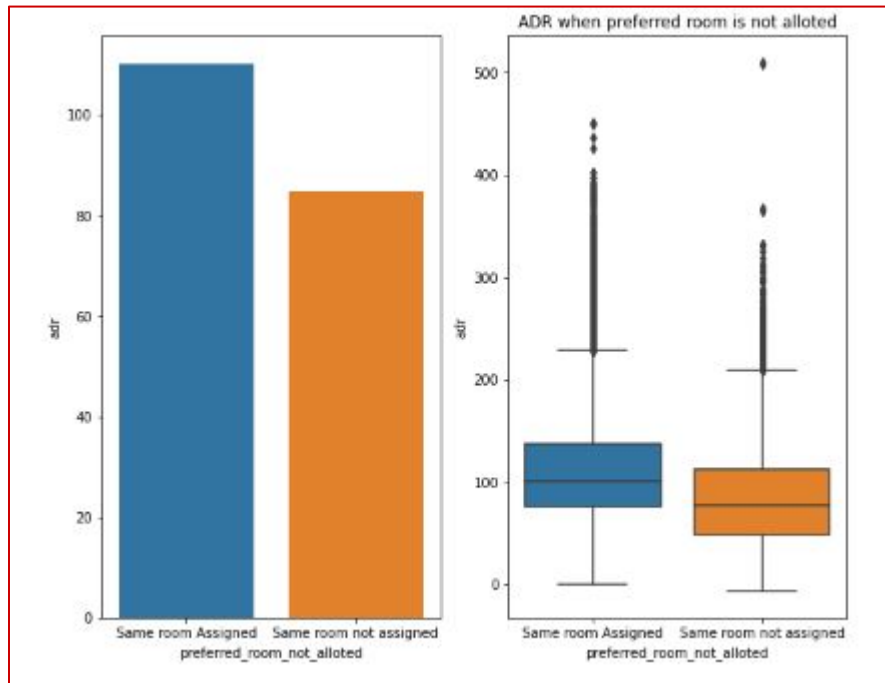
When the hotel does not assign the room preferred by the customer, does it leads to cancellation?

&



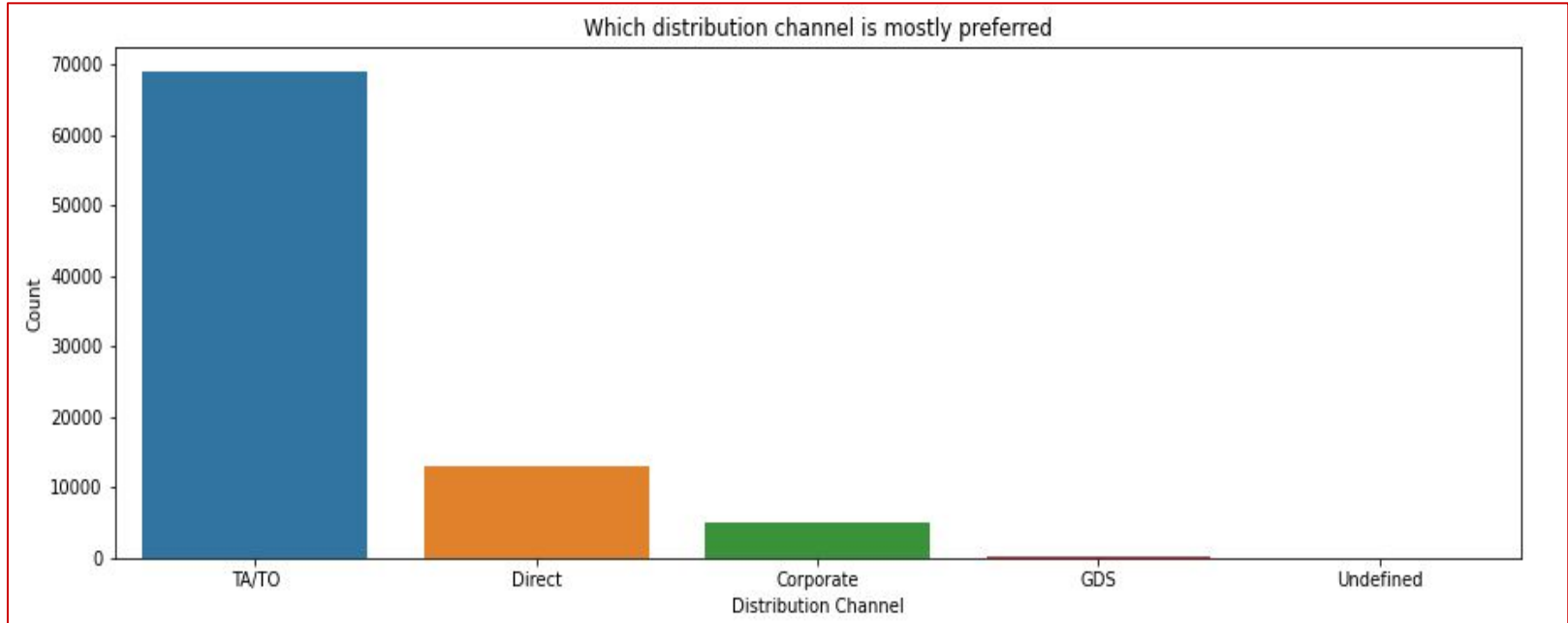
Conclusions:

- ***Customer generally doesn't cancel the booking when the hotel doesn't assign the same room as preferred by the customers.***
- ***Although the hotel not assigning the preferred room type by customer generally doesn't leads to cancellation. But when hotels don't assign the preferred room type to customer their ADR decreases hence profit decreases.***



Distribution channel wise analysis

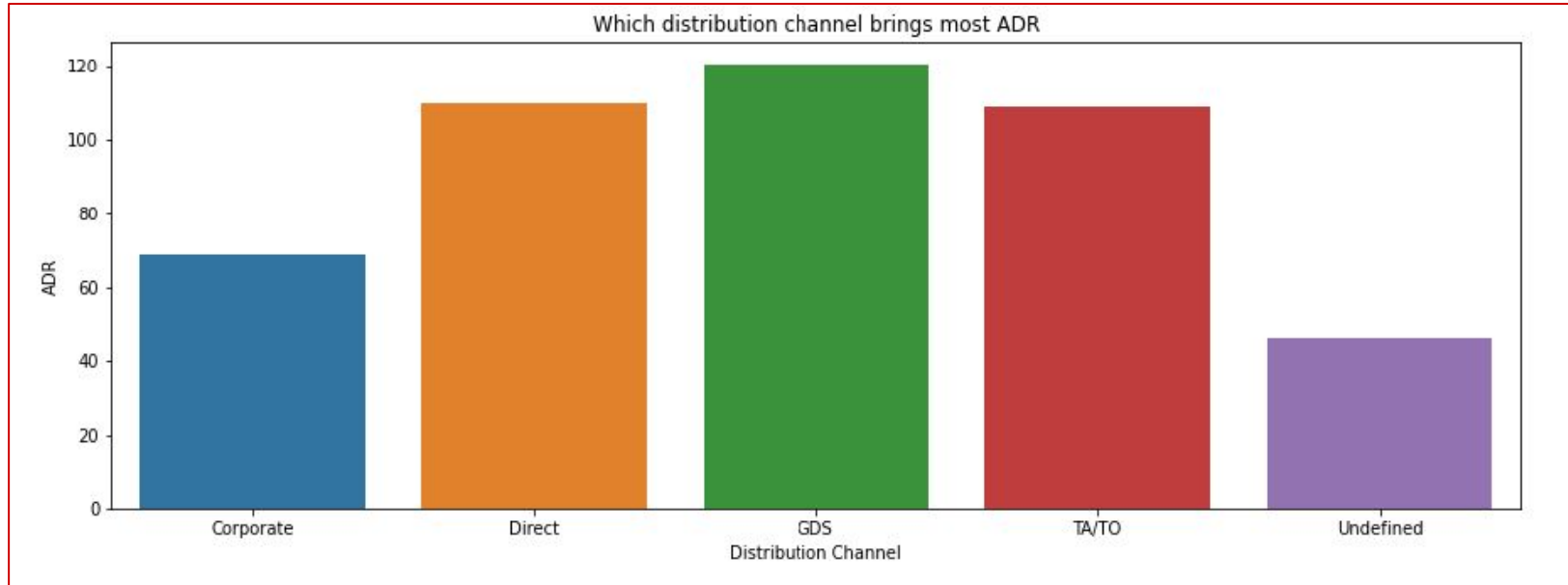
Which distribution channel is mostly preferred by the customers?



Conclusions:

- *Travel Agent and Tour operator are mostly preferred by the customers.*
- *Hotels can partner with these agents and operator or advertise using them as medium in order to increase their business.*

Which distribution channel brings the most 'adr' for the hotels?

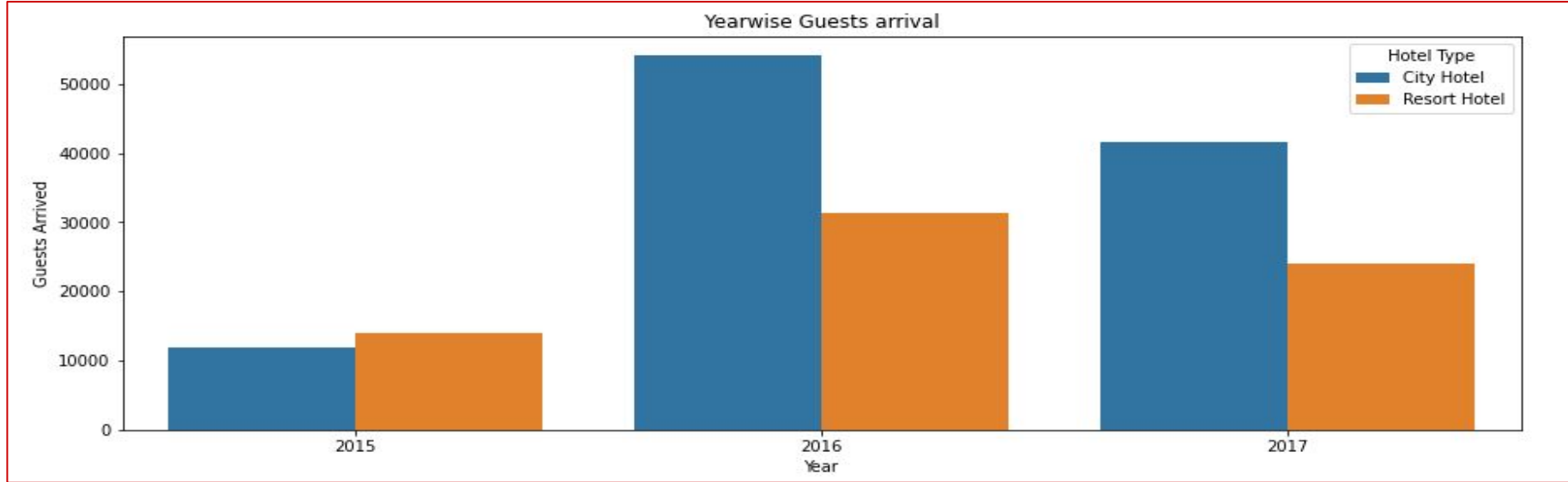


Conclusions:

- *TA(travel agent)/TO(tour operator) brings the most number of customers but the revenue generated per room is most through Global Distribution System(GDS).*
- *Hotels can think about a way to bring more customers through GDS in order to increase overall profitability.*

Time wise Analysis

Yearly arrival of the guests



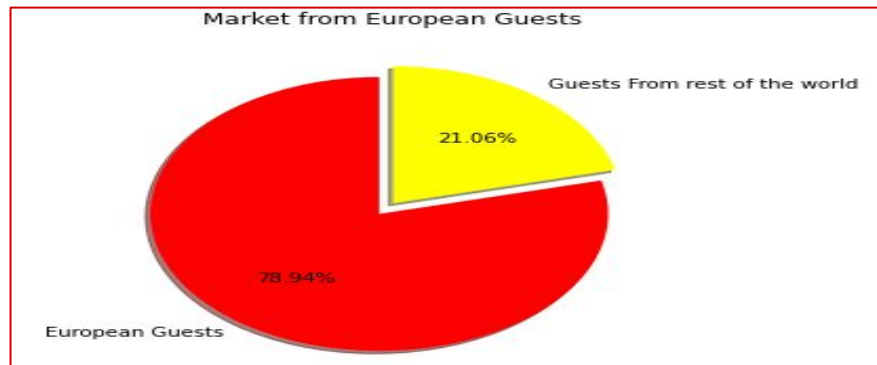
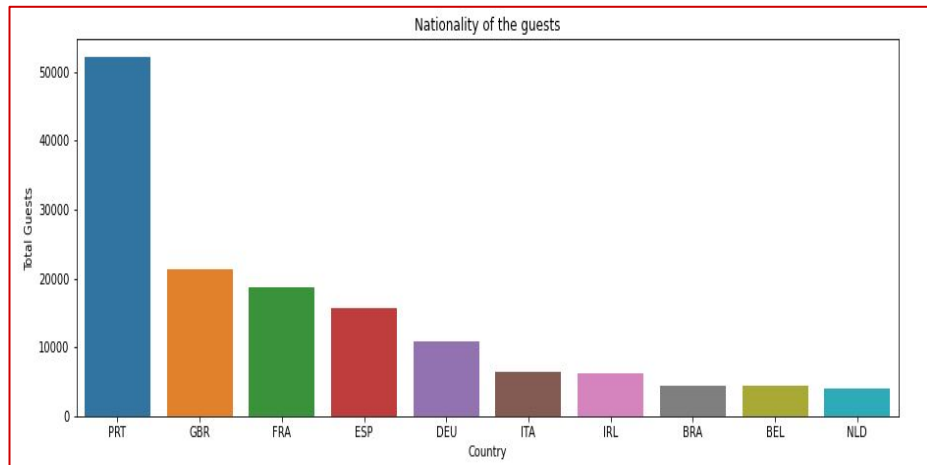
Conclusions:

- ➔ The year 2016 witnessed maximum number of guests followed by the year 2017.
- ➔ Although the year 2015 witnessed the minimum number of guests, in the year 2015 Resort hotel type booking is more than City hotel.

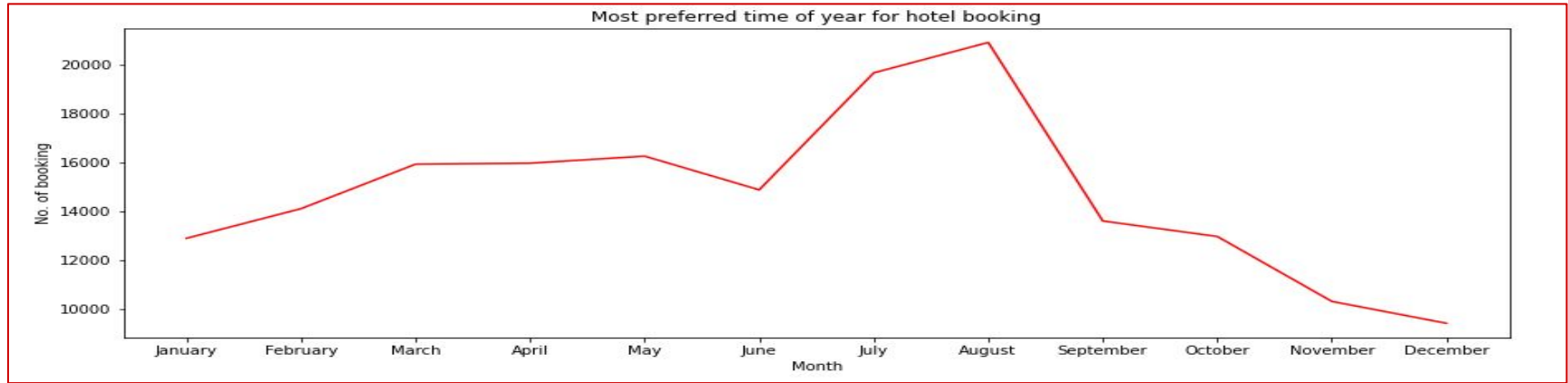
What is the nationality of the guest?

Conclusions:

- From the above graph we can see that most number of guests are from PORTUGAL followed by GREAT BRITAIN and FRANCE.
- 9 out of 10 max guests's nationality belongs to the European countries.
- Out of 176990 guests 139708 guest's nationality belongs to EUROPE, which is roughly 79% of total guests. So it is safe to conclude that data in this dataset is of European hotels.



Which is the the best month of the year to stay in hotel?



Conclusions:

- ➔ As we can see that instead of the month of DECEMBER the total number of guests in every month are more than 10k. But from the graph there is a very clear visible spike is in the month of JULY and AUGUST, so we can conclude that the best time of the year to book a hotel is in JULY and AUGUST.
- ➔ The above graph further supports our assumption that the data in the dataset is of European Hotels, as during the month of JULY and AUGUST it is summers in Europe and the weather, natural sceneries and landscapes are breathtaking therefore the tourism industries boom during this period of time thus there is a clear visible spike in total number of guests during this period.
- ➔ The in the months of November, December, January and February there is extreme winters in Europe, therefore there is a clear visible decline in total guests during this period of time.

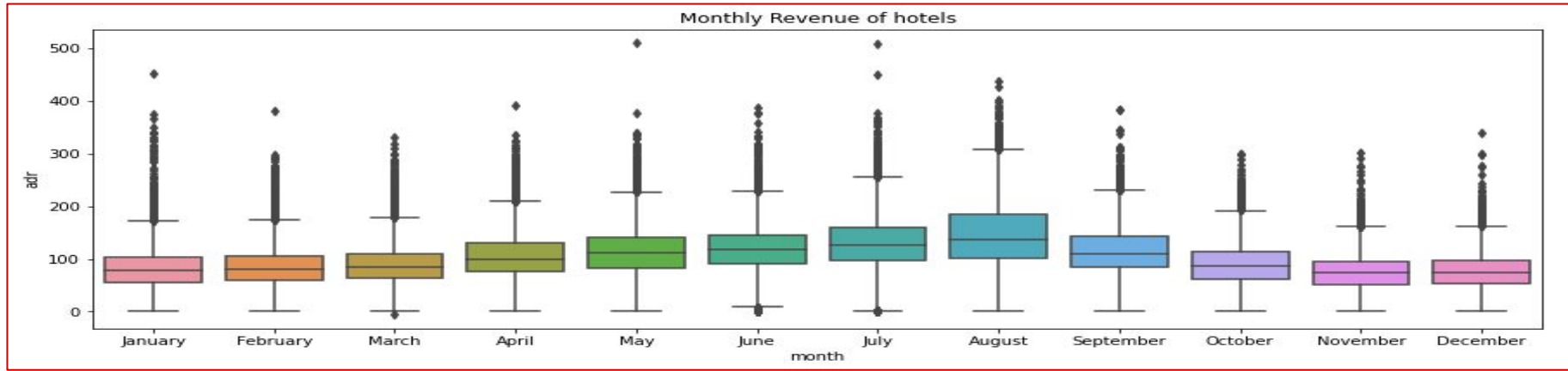
Resort And City Hotel Monthly Comparison



Conclusions:

- ➔ After looking at the above graph we can say that even after analyzing the data of CITY and RESORT hotel separately the fact that the "Best time to book an hotel is month of JULY and AUGUST" remains the same.
- ➔ From the above table we can also conclude that throughout the year the people prefer CITY hotels more than the RESORT hotel.

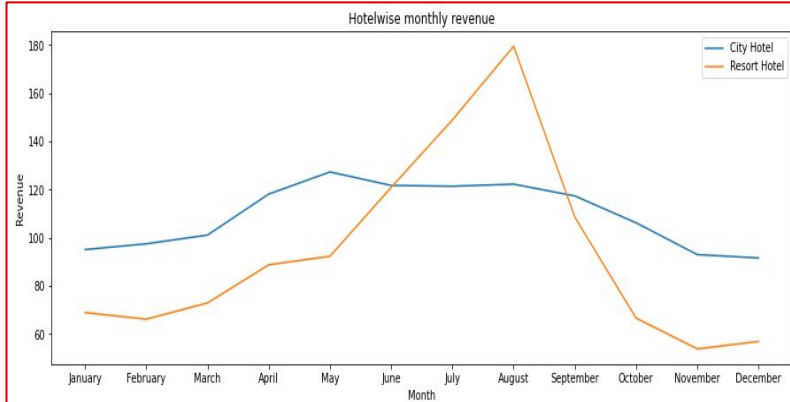
In which month of the year Revenue is the highest?



Conclusions:

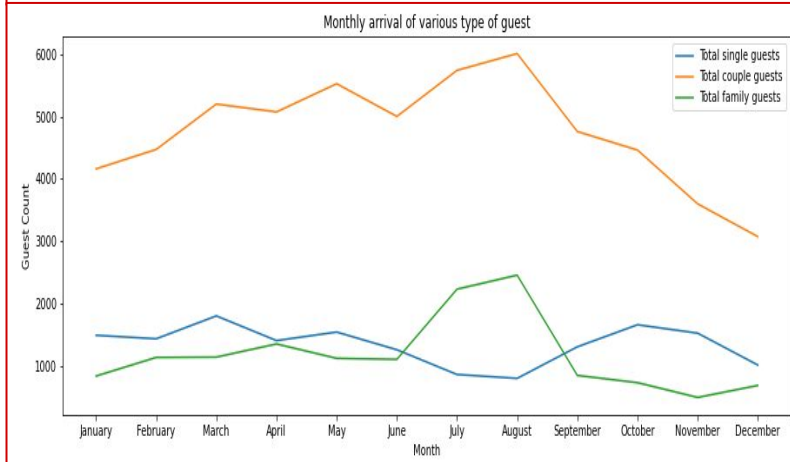
- ➔ The above analysis further supports our assumption that the data in the dataset is of european hotels. And as the month of June, July, August, September witness summers in the Europe, the number of bookings are more and Hotels make most profit during these months only. So we can also assume that hotels increase prices of the room during these months.
- ➔ And as the month of November, December, January, February witness Winters in europe The number of booking are low thus hotels reduce the prices of room during these months and their revenue also reduce.

Monthly ADR of Resort and City Hotel

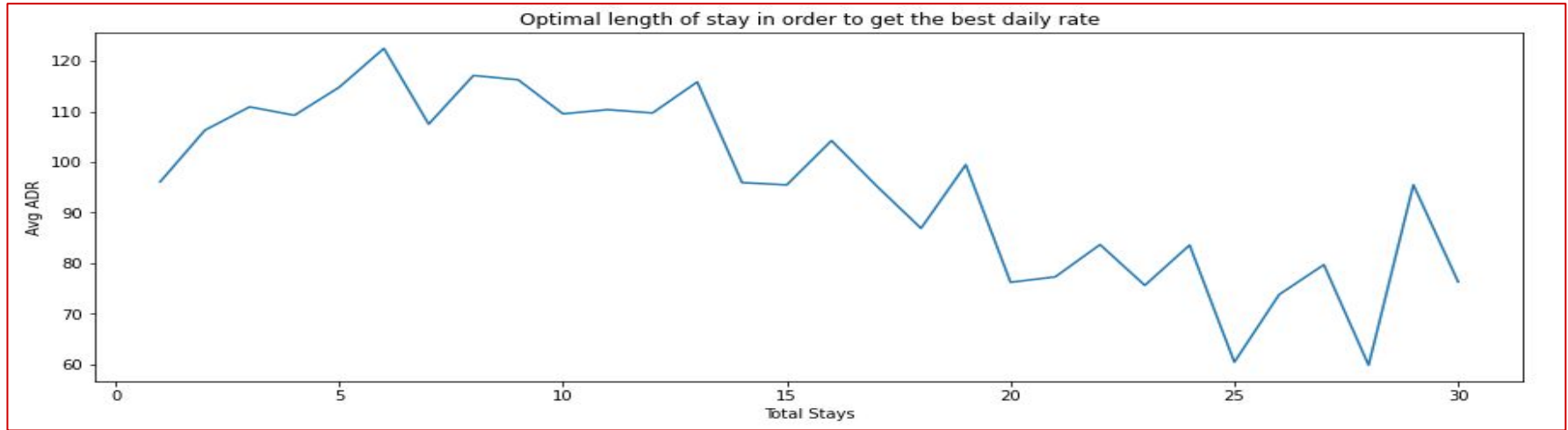


Conclusions:

- We can observe from the above graph that during the peak months (June to September) the Resort hotel's earning of revenue is way more than City hotels (which earn more or less same throughout the year), which further proves our assumption that City hotels are booked more for Casual stays thus their price increase very less during summers, while Resort hotels are booked more for Close intimate events which happen during peak of tourism season, so the Resort hotel owner earn more during this period of time.
- As we can observe that most of the bookings throughout the year is made by guests of type couple (or two individual) so the hotels can devise their plans and offers keeping this in mind.
- The hotel booking trend of singles is opposite to the booking trend of couples and families. The number of bookings of singles reduces during the peak of the tourism season which is strange, one of the reason may be, they are travelling in a group.



Optimal length of stay in order to get the best daily rate

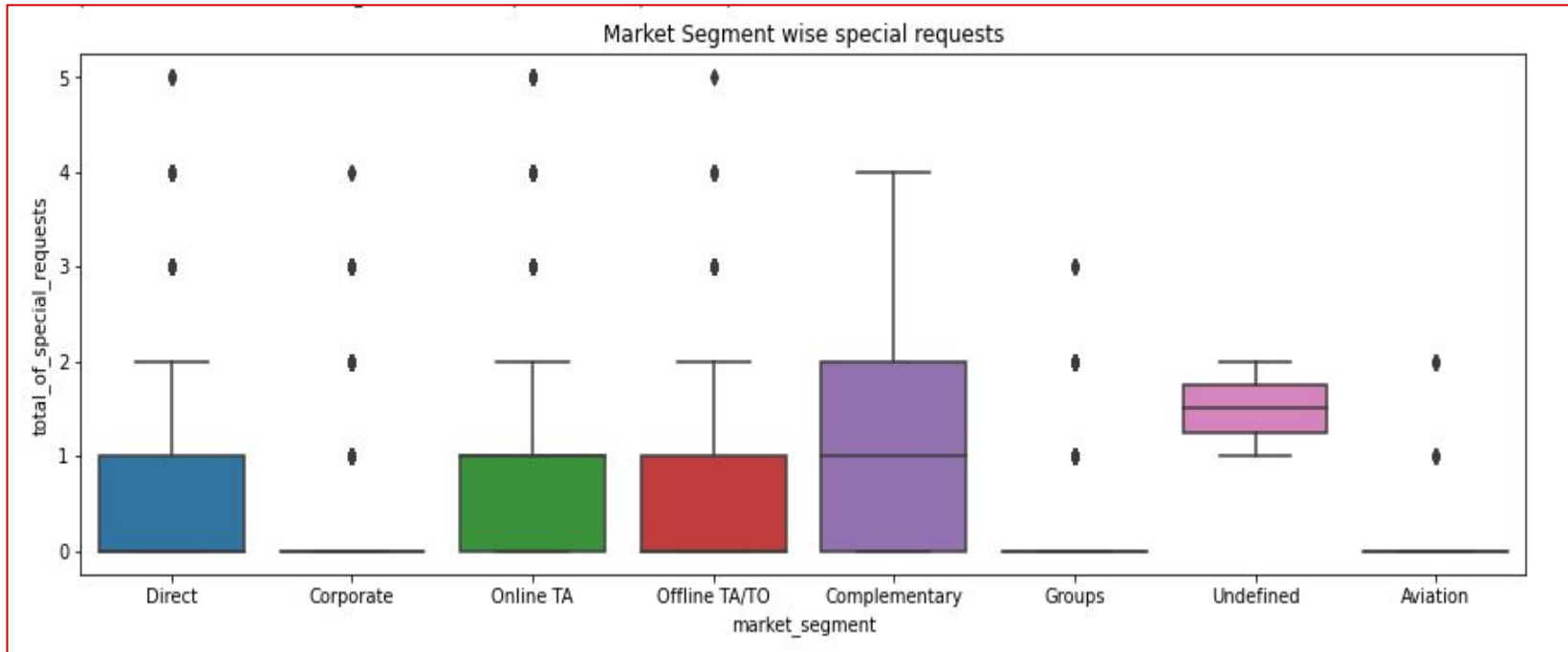


Conclusions:

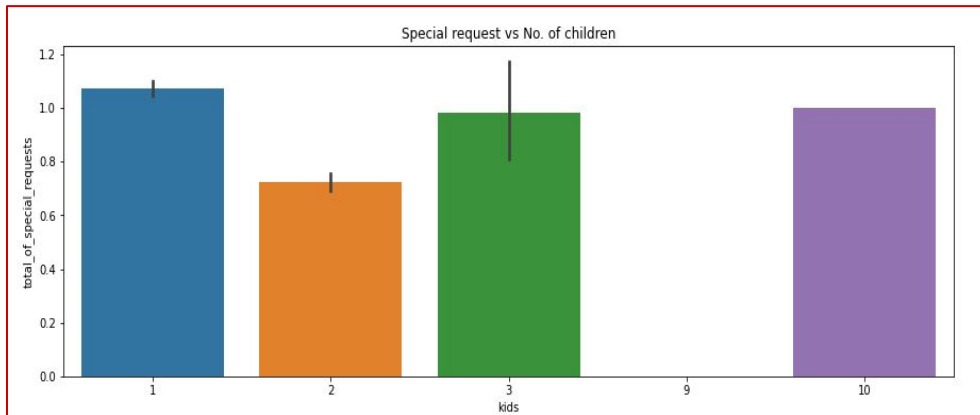
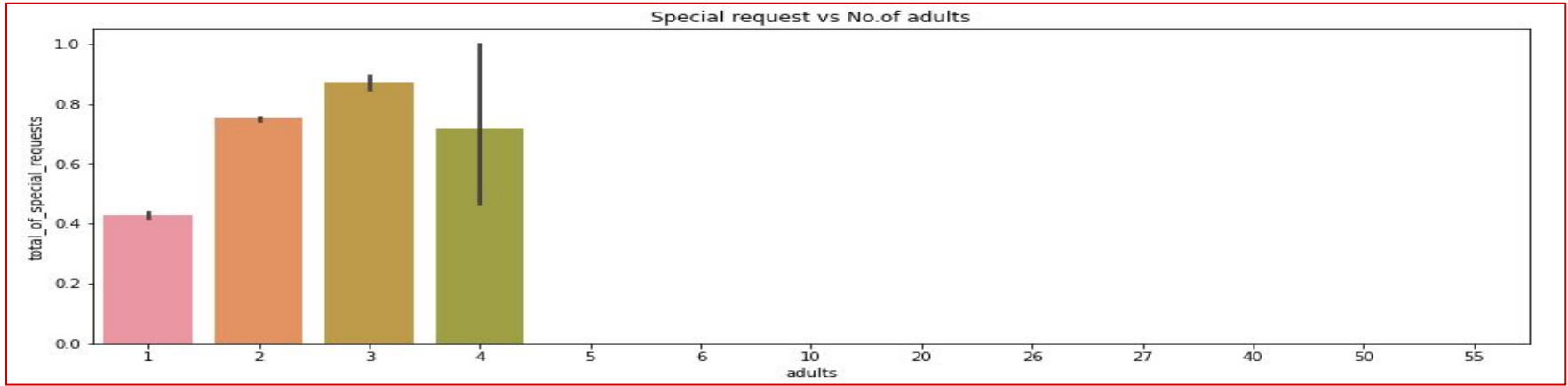
- **Hotel's Point of view:** The optimal length of stay to get the best daily rate is less than or equal to 2 weeks, because it is the most frequently booked length of stay by the customer.
- **Customer's point of view:** The optimal length of stay to get the best daily rate is more than 2 weeks.

Special Request Scenario

Market Segment (Special Request Analysis)



Children And Adult (Special Request Analysis)



Conclusion:

- Hotels will likely receive a disproportionately high number of special requests in the case of children among the guests and also in case of 2 to 4 adults in the group.
- Hotels can expect disproportionately high number of special requests from every market segments except 'corporate', 'groups', 'aviation'.

Thank You