

Capstone Project-1

EDA on Hotel Booking Analysis

BY
Nikhil G. Umare

Problem Statement:



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? 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.

The main objective behind this project is to explore and analyze data to discover important factors that govern the bookings and give insights to hotel management ,which can perform various campaigns to boost the business and performance.

Steps Followed:



**Data Collection and
Exploration**

**Data cleaning and
Manipulation**

**Exploratory Data
Analysis(EDA)**

Data Description:

- **hotel** : Hotel(Resort Hotel or City Hotel)
- **is_canceled** : Value indicating if the booking was canceled (1) or not (0)
- **lead_time** : Number of days that elapsed 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 number of year for arrival date
- **arrival_date_day_of_month** : Day of arrival date
- **stays_in_weekend_nights** : Number of weekend nights (Saturday or Sunday) the guest stayed or booked hotel
- **stays_in_week_nights** : Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **adults** : Number of adults
- **children** : Number of children
- **babies** : Number of babies
- **meal** : Type of meal booked. Categories are presented in standard hospitality meal packages:
- **country** : Country of origin.`

Data Description(contd):

market_segment : Market segment designation.(TA/TO)

distribution_channel : Booking distribution channel.(TA/TO)

is_repeated_guest : a repeated guest (1) or not (0)

previous_cancellations : Number of previous bookings that were cancelled by the customer prior to the current booking

previous_bookings_not_canceled : Number of previous bookings not cancelled by the customer prior to the current booking

reserved_room_type : Code of room type reserved.

assigned_room_type : Code for the type of room assigned to the booking.

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

deposit_type :No Deposit,Non Refund,Refundable

agent : ID of the travel agency that made the booking

company : ID of the company/entity that made the booking or responsible for paying the booking.

days_in_waiting_list : Number of days the booking was in the waiting list before it was confirmed to the customer

customer_type : Type of booking, assuming one of four categories

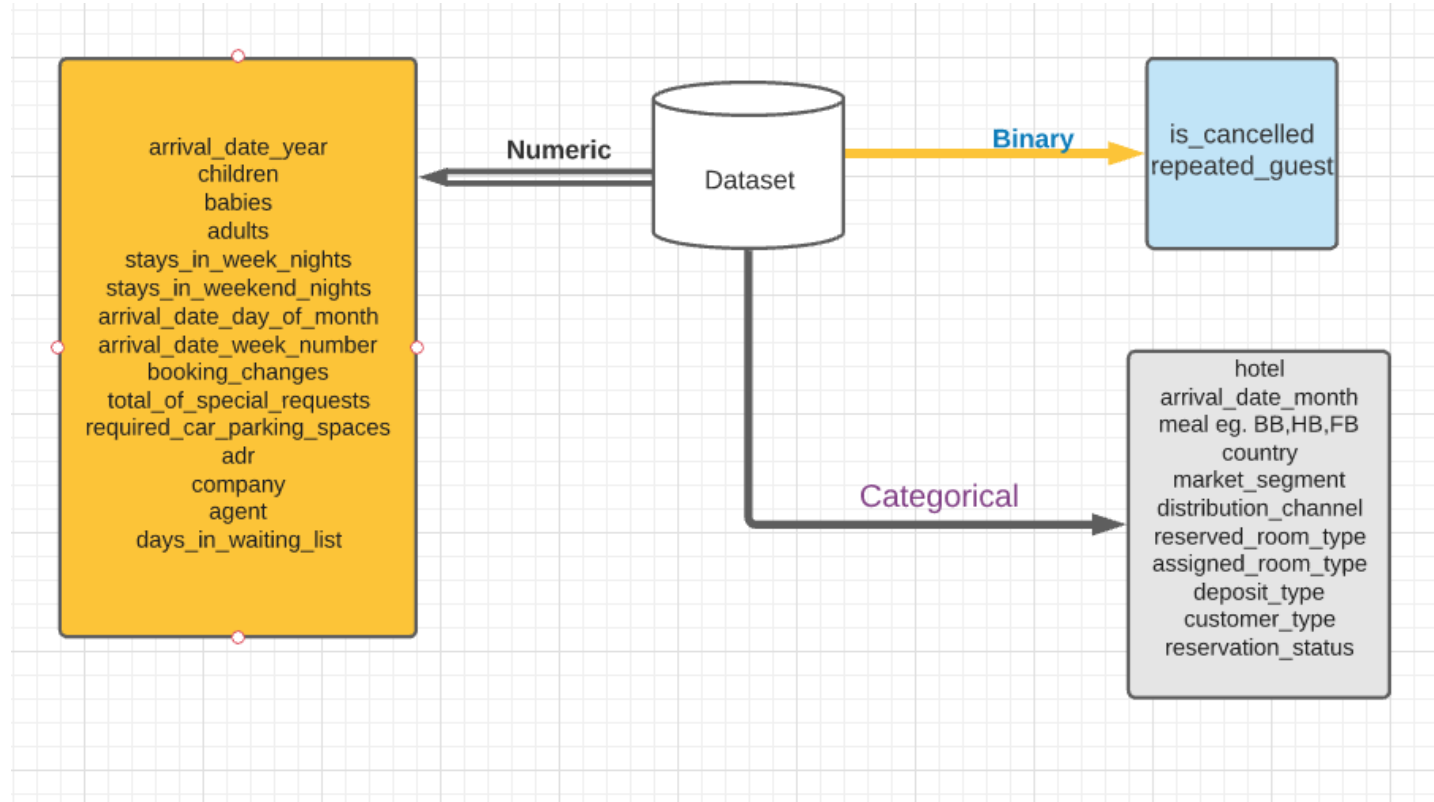
adr : Average Daily Rate

required_car_parking_spaces : Number of car parking spaces required by the customer

total_of_special_requests : Number of special requests made by the customer (e.g. twin bed or high floor)

reservation_status: Reservation last status

Data summary



Data Cleaning:

1)Missing values: There were 4 columns company, agent, country and children with missing values.

2)Handling Duplicates: Data had 31994 duplicates values.so we dropped it from data.

3) Removing Outliers:Removing an outlier from Average_daily_rate column

4) Feature Engineering:

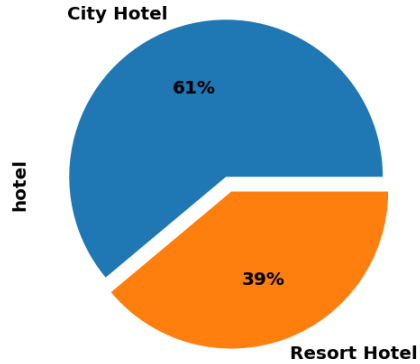
We created 2 new columns

a) 'Total _People'=from the childrens, adults, babies.

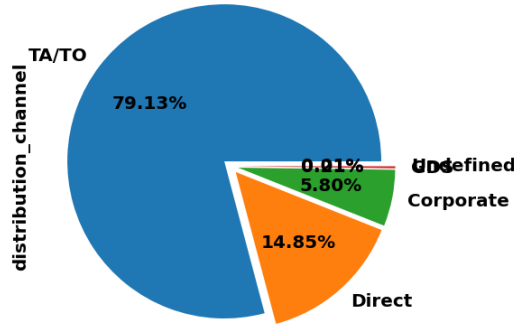
b) 'Total_stay'=From weekend nights and weekdays nights.

Exploratory Data Analysis:

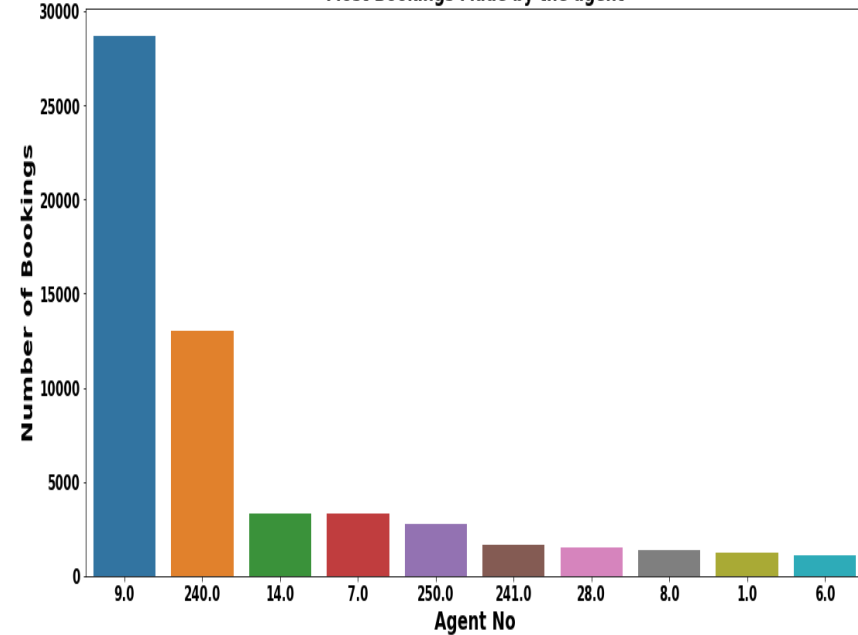
Pie Chart for most preferred hotel



Booking % by distribution channel



Most Bookings Made by the agent

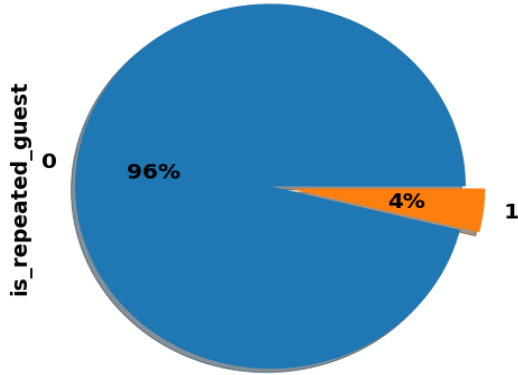


CONCLUSIONS:

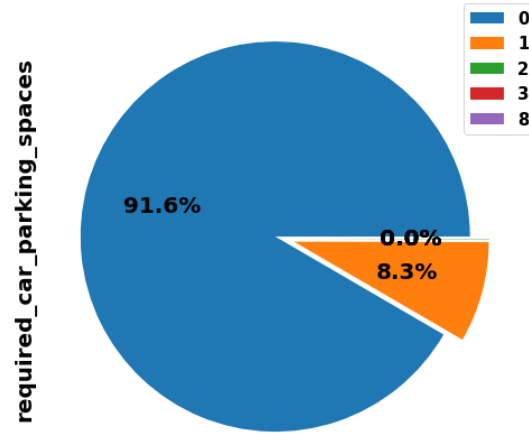
1. City Hotel is most preferred hotel by guests. Thus city hotels has maximum bookings.
2. Most 79% people prefer 'TA/TO' for booking.
3. Agent ID no: 9 made most of the bookings

Exploratory Data Analysis:

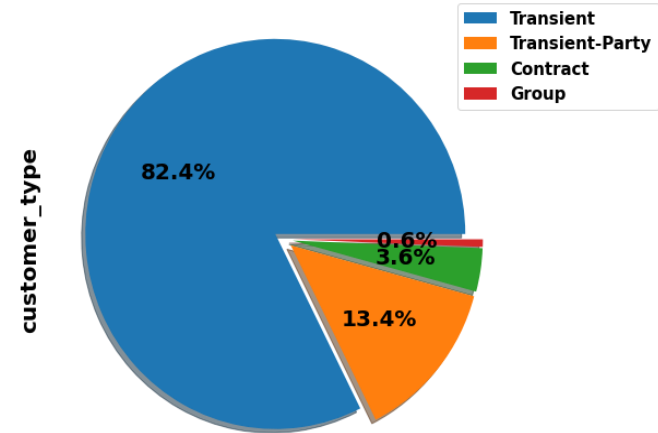
Percentage (%) of repeated guests



% Distribution of required car parking spaces



% Distribution of Customer Type



CONCLUSIONS:

1. Only 4 % people were revisited the hotels. Rest 96 % were new guests. Thus retention rate is low.
2. 91.6 % guests did not required the parking space. only 8.3 % guests required only 1 parking space.
3. Transient customer type is more which is 82.4 %. percentage of Booking associated by the Group is very low.

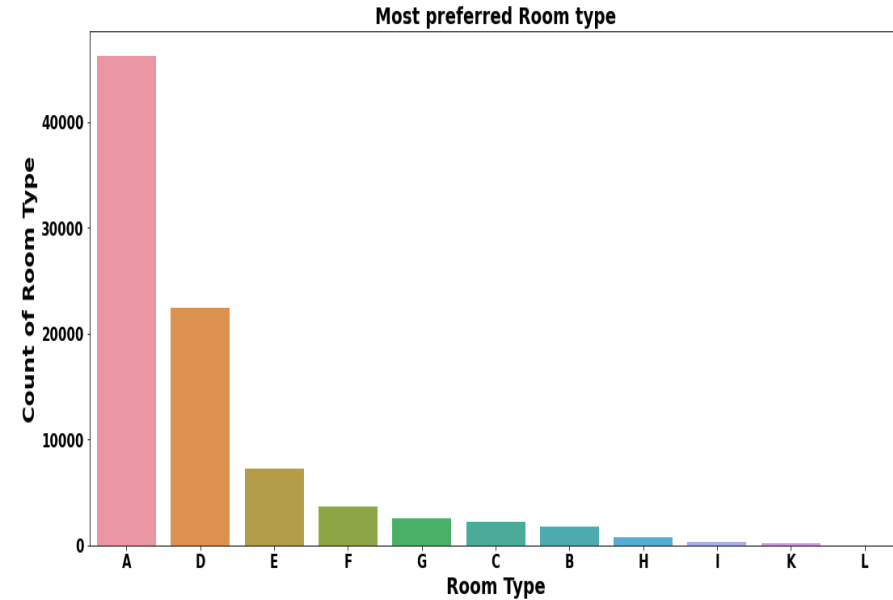
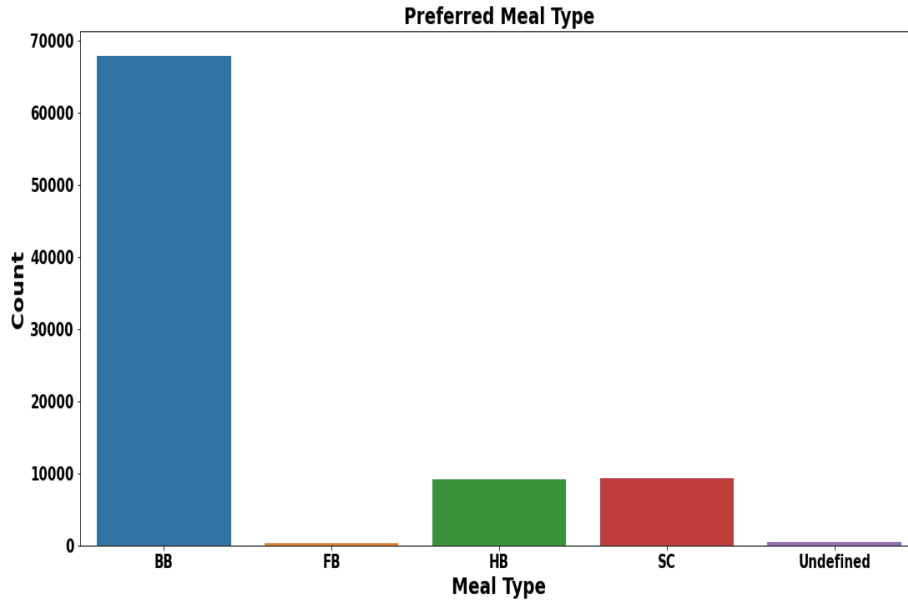
Contract-when the booking has an allotment or other type of contract associated to it

Group -when the booking is associated to a group

Transient-when the booking is not part of a group or contract, and is not associated to other transient booking

Transient-party-when the booking is transient, but is associated to at least other transient booking

Exploratory Data Analysis:

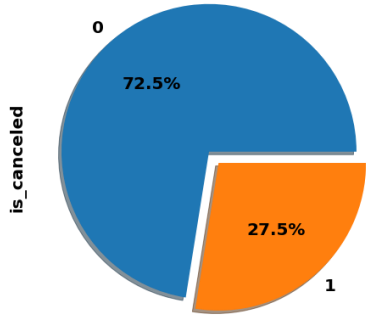


CONCLUSION:

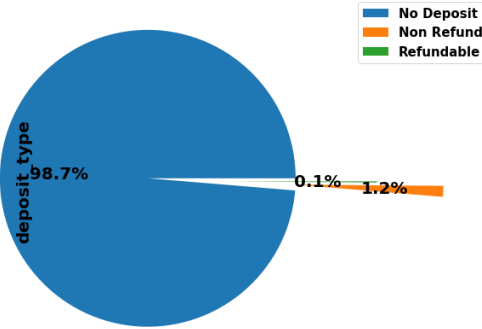
1. So the most preferred meal type by the guests is BB(Bed and Breakfast)
- HB- (Half Board) and SC- (Self Catering) are equally preferred.
2. So the most preferred Room type is "A".

Exploratory Data Analysis:

Cancellation and non Cancellation



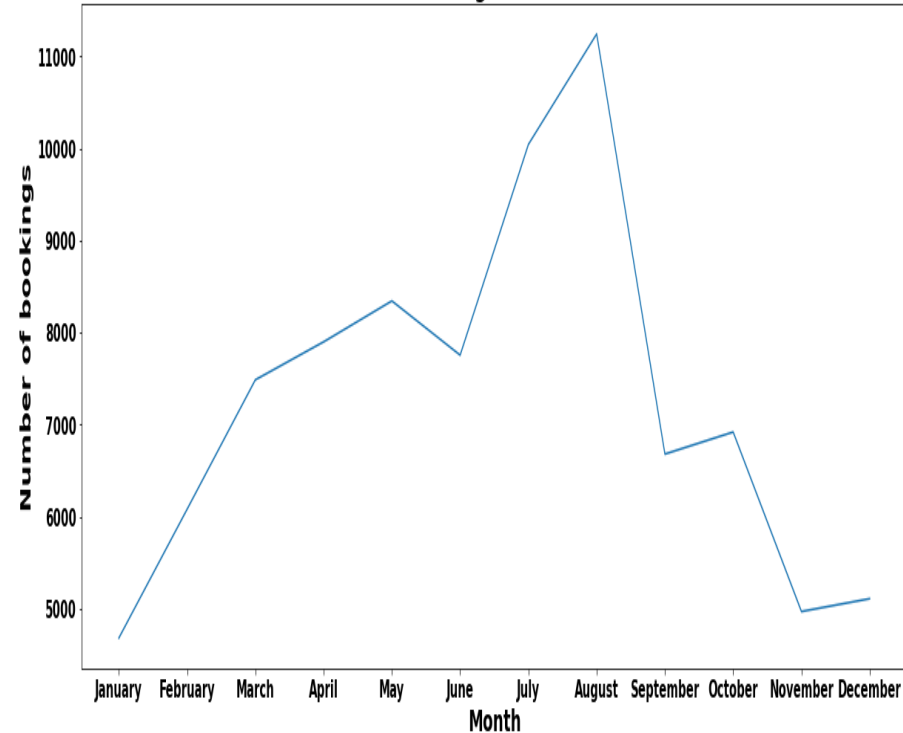
% Distribution of deposit type



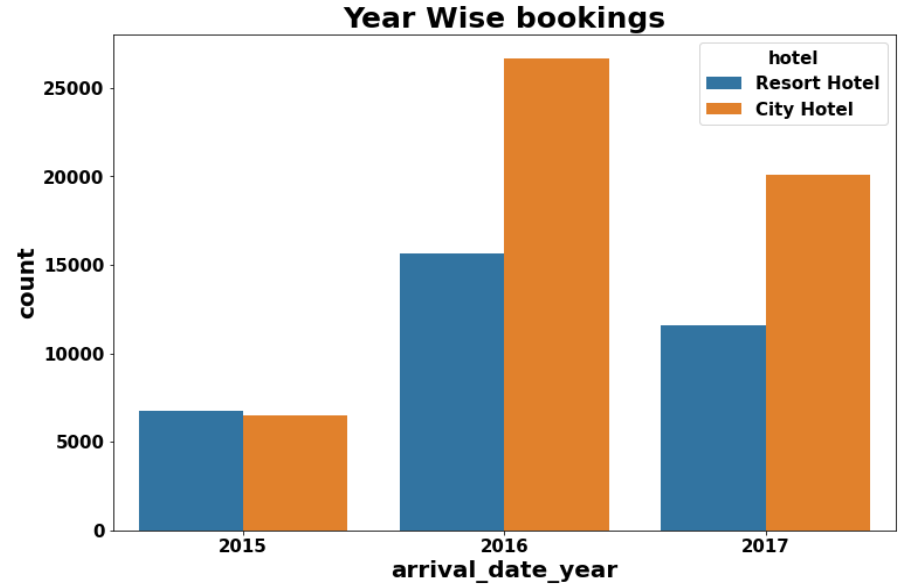
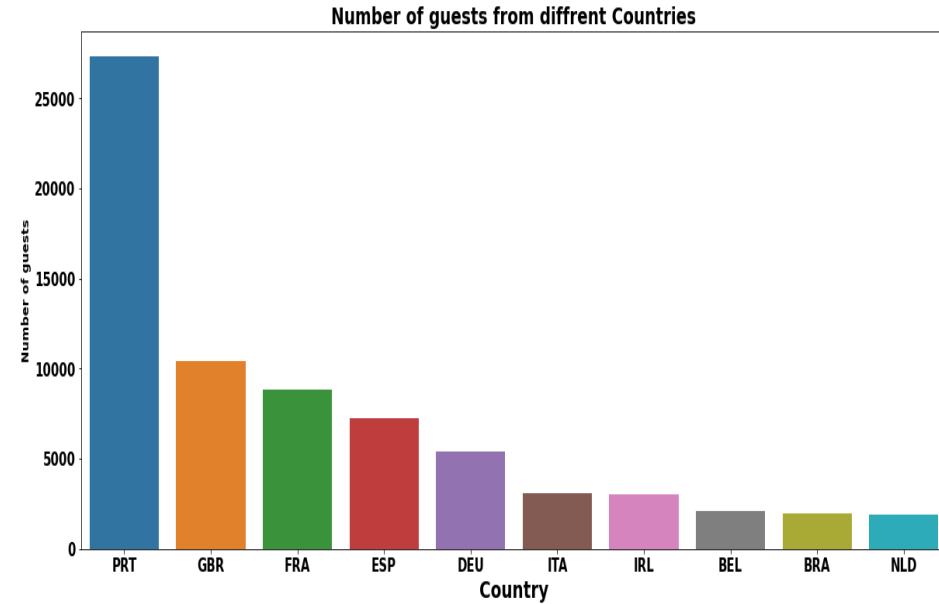
CONCLUSIONS:

1. 27.5 % of the bookings were cancelled.
2. 98.7 % of the guests prefer "No deposit" type of deposit.
3. July and August months had the most Bookings.
Summer vacation can be the reason for the bookings.

Number of bookings across each month

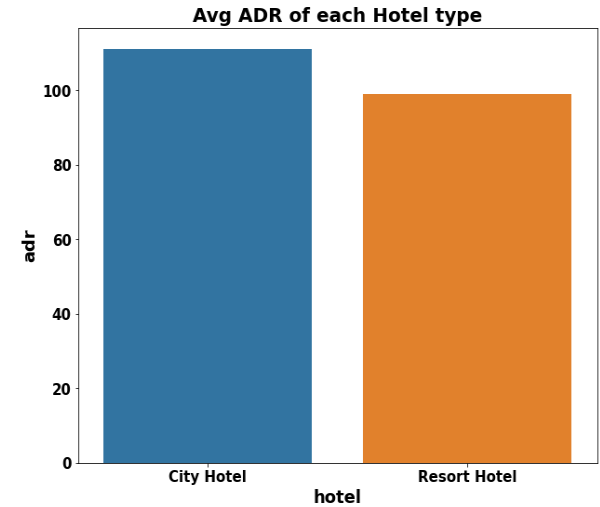
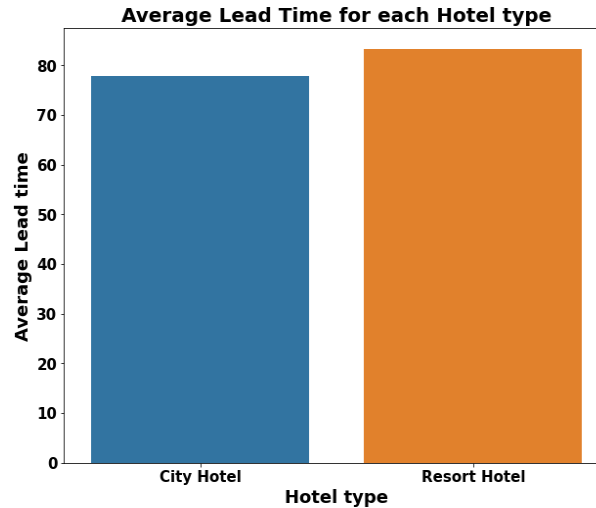


Exploratory Data Analysis:



- CONCLUSIONS:
- 1. Most of the guests are coming from Portugal i.e. more than 25,000 guests are from Portugal.
- 2. 2016 had the highest bookings. 2015 had less than 7,000 bookings. Overall, City hotels had the most of the bookings.

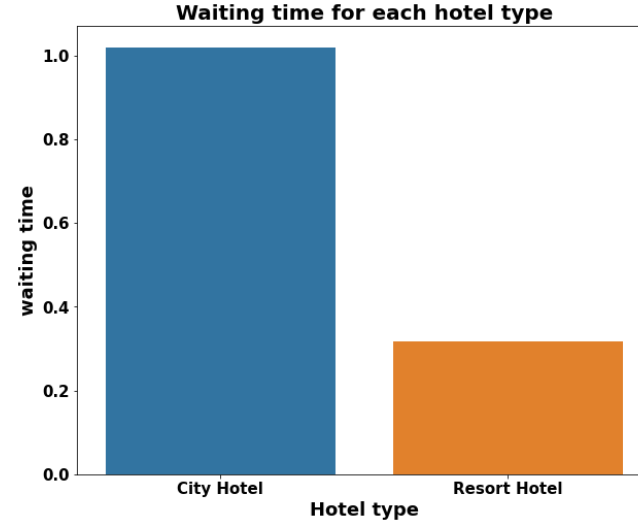
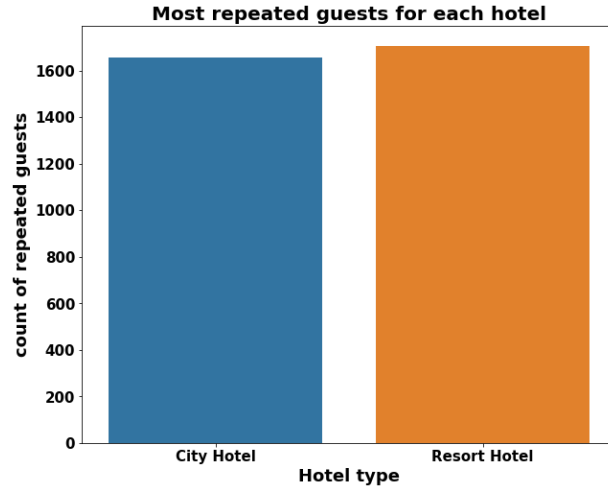
Exploratory Data Analysis:



CONCLUSION:

1. Booking cancellation rate is high for City hotels which almost 30 %.
2. Average lead time for resort hotel is high. It means people plan their trip too early. Usually people prefer resort hotels for longer stays. That's why people plan early.
3. Average ADR for city hotel is high as compared to resort hotels. These City hotels are generating more revenue than the resort hotels.

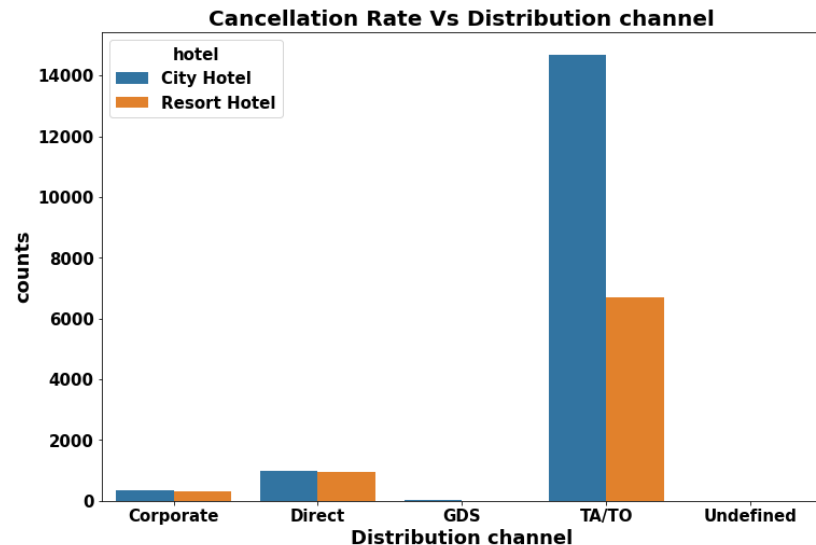
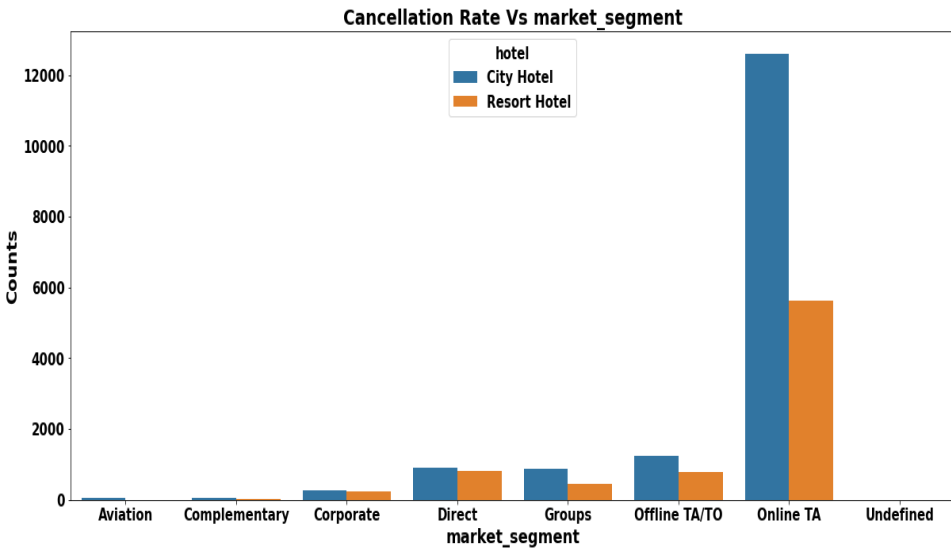
Exploratory Data Analysis:



CONCLUSION:

1. Resort hotels has the most repeated guests. In order to get increase the count of repeated guests hotel management need to take the valuable feedbacks from the guests and try to give good service.
2. Waiting time period for City hotel is high as compared to resort hotels. That means city hotels are much busier than Resort hotels.

Exploratory Data Analysis:



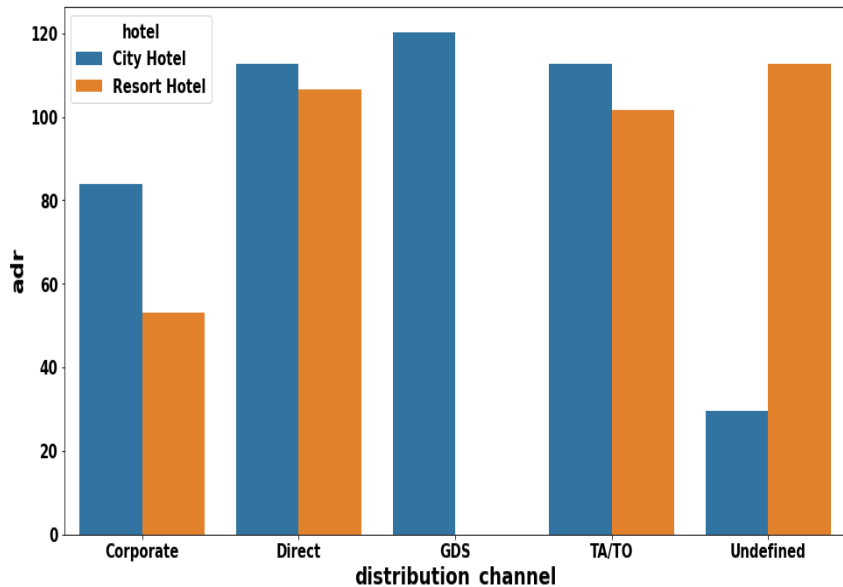
CONCLUSIONS: TA/TO' market segment has highest cancellations for city hotels.

1. Market Segment: Online TA/TO' market segment has highest cancellations for city hotels.

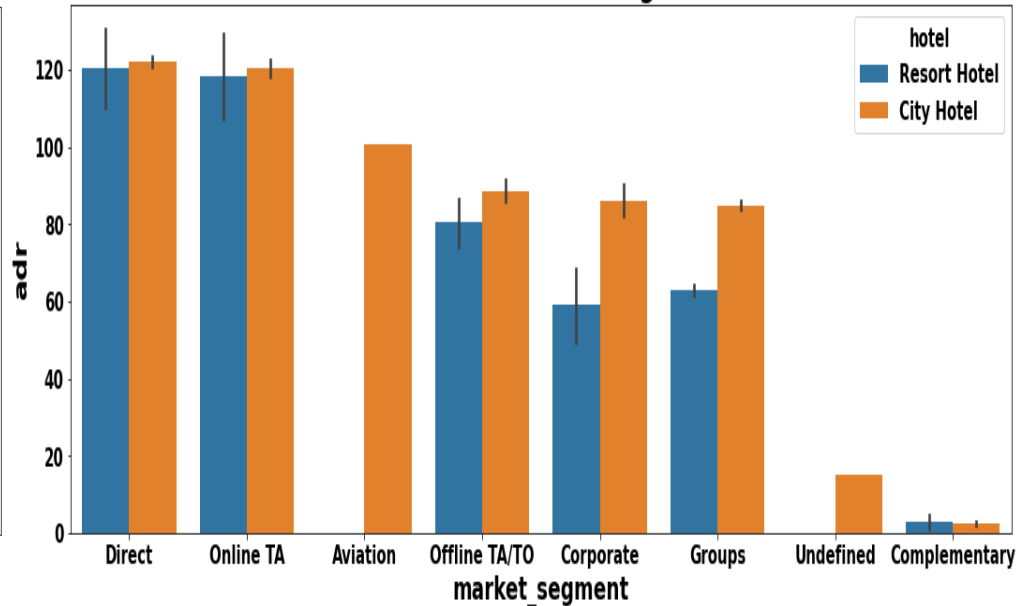
2. Distribution channel: TA/TO' distribution channel has highest cancellations for city hotels and more than 6000 cancellations for resort hotels. In order to reduce the cancellations they should improve their cancellation policies and deposit policies.

Exploratory Data Analysis:

ADR across Distribution channel



Adr across market segment



CONCLUSION:

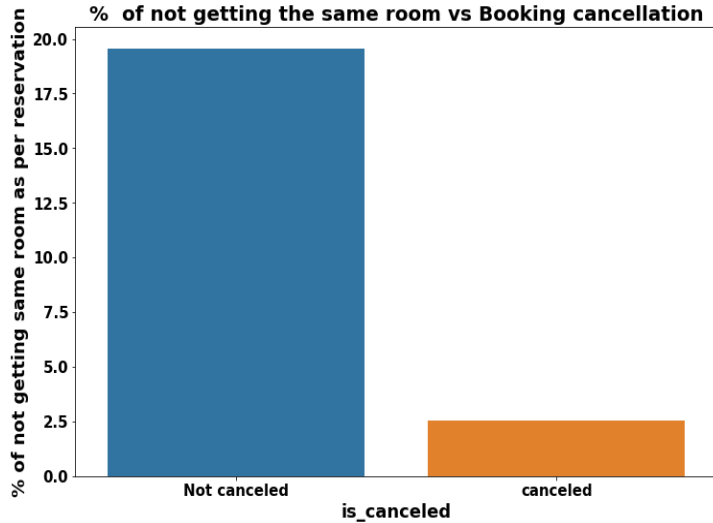
Distribution channel:

1. 'Direct' and 'TA/TO' has almost equal adr in both type of hotels which is high among other channels.
2. GDS has high adr in 'City Hotel' type. GDS needs to increase Resort Hotel bookings. From this we can say that "Direct" and 'TA/TO' are generating more revenue than the other channels.

Market Segment:

1. Here "Direct" and 'Online Travel Agency' has high adr for both hotel types. Aviation segment needs to increase Resort hotel bookings.

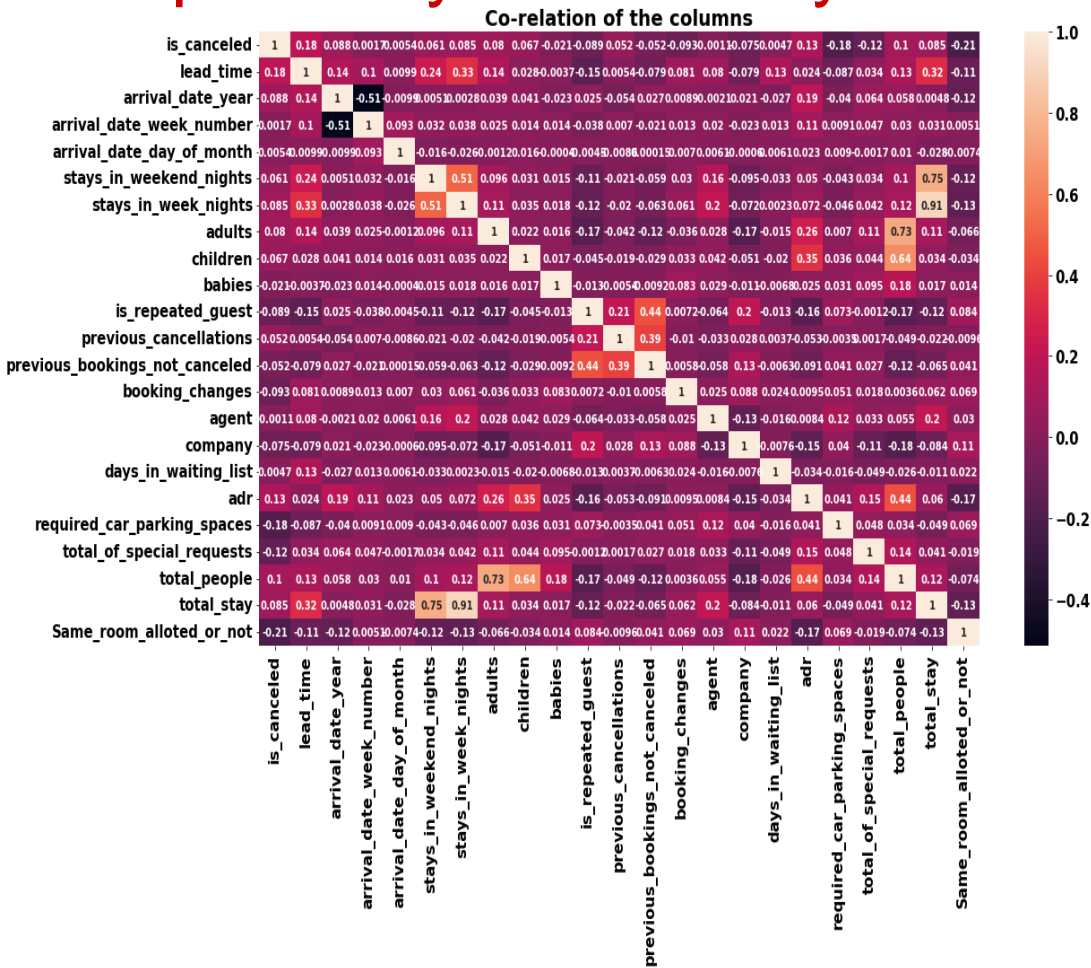
Exploratory Data Analysis:



CONCLUSION:

1. Almost 19 % people did not canceled their bookings even after not getting the same room which they reserved while booking hotel. Only 2.5 % people cancelled the booking.
2. Thus not getting the same room as per reserved room is not the reason for booking cancellations

Exploratory Data Analysis:

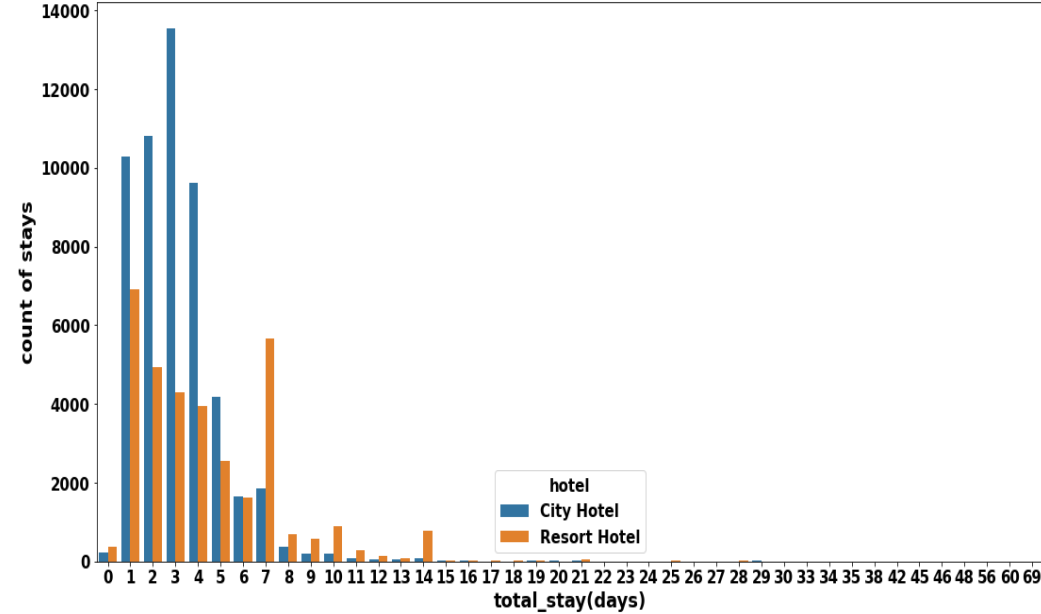


CONCLUSION:

- 1.is canceled and same_room_allotted_or_not are negatively correlated. Not getting the same room as per reserved room is not the reason for booking cancellations.
- 2.lead-time and total stay is positively correlated means more is the stay of customer more will be the lead time.
- 3.ADR and total people are highly correlated. That means more the people more will be adr. High adr means high revenue
- 4.is_repeated_guest and previous_bookings Not_canceled has strong correlation. May be repeated guests are not more likely to cancel their bookings.

Exploratory Data Analysis:

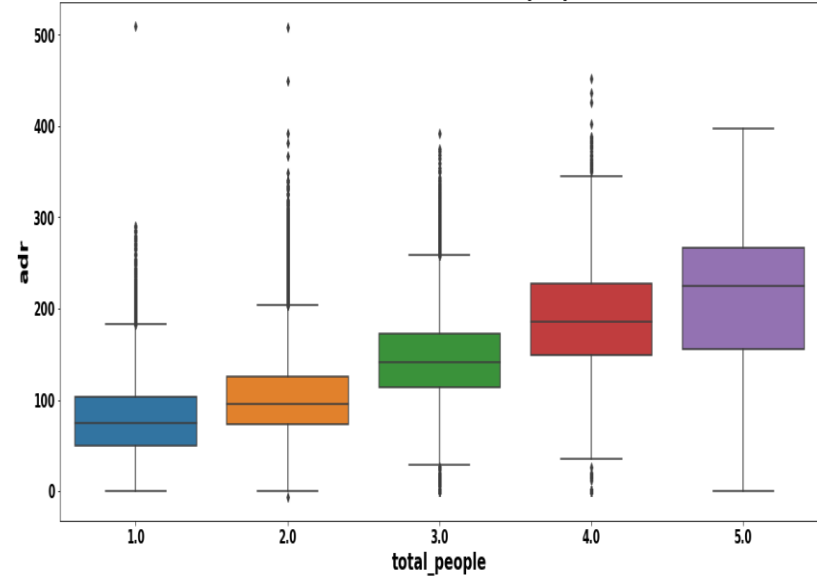
Optimal Stay Length in Both hotel types



CONCLUSION:

1. Optimal stay in both the type hotel is less than 7 days. Usually people stays for a week.
2. For stay more than 7 days people likes to stay in Resort hotels. As we can see after 7 days City Hotel Bookings are very less as compared to Resort hotels.
3. As we saw in Correlation heatmap, total people and adr are positively correlated. Thus for 2 people ,adr is almost 100 and for 5 people its more than 200.
4. Thus more the people more will revenue of the hotel

ADR v/s Total Number of people



THANK YOU