

Capstone Project

Hotel booking analysis

Apoorva KR

apoorvargowda1@gmail.com

Introduction

- Hotel industry is a very volatile industry and the bookings depend on variety of factors such as type of hotels, seasonality, days of week and many more. This makes analyzing the patterns available in the past data more important to help the hotels plan better. Using the historical data, hotels can perform various campaigns to boost the business. We can do EDA to predict the future bookings

Agenda

To discuss the analysis of given hotel bookings data set from 2015-2017

Analysis of given data set in following ways:

- Hotel wise analysis
- Distribution channel wise analysis
- Booking cancellation analysis
- Time wise analysis

By doing this trying to find out key factors driving the hotel booking trends

Points to discuss

- Agenda
- Data summary
- Hotel wise analysis
- Distribution channel wise analysis
- Booking cancellation analysis
- Time wise analysis
- Heat map
- Some important questions
- conclusion

Data summary

Given data set has different columns of variables for hotel bookings.some of them are:

Hotel: the category of hotels ,which are two resort hotel and city hotel.

Is_cancelled : the value of column show the cancellation type.if the booking was cancelled or not values[0,1], where 0 indicates not cancelled

Lead_time: the time between reservation and actual arrival

Stayed_in_weekend_nights:the number of weekend nights stay per reservation

Stayed_in_weekdays_nights:the number of weekdays nights stay per reservation

Meal: meal preferences per reservation.[BB,FB,HB,SC,Undefined]

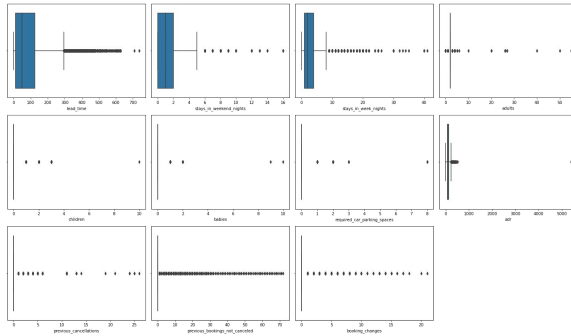
Country: the origin country of guest

Data summary (cont..)

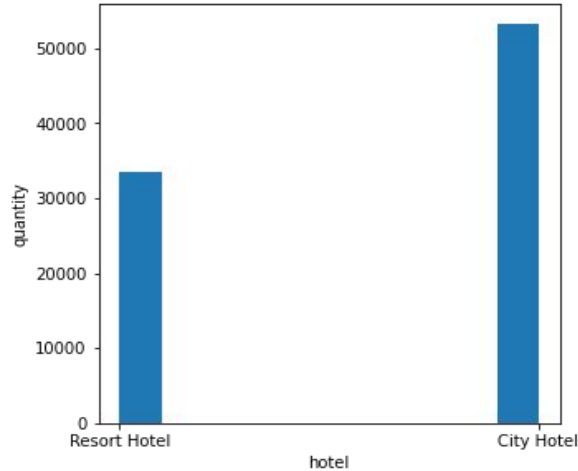
- **Market segment** : (a group of people who share one or more common characteristics, lumped together for marketing purposes)
 1. TA: Travel agents
 2. TO: Tour operator
- **distribution_channel** (A distribution channel is a chain of businesses or intermediaries through which a good or service passes until it reaches the final buyer or the end consumer)
- **day_in_waiting_list**: Number of days the booking was in the waiting list before it was confirmed to the customer
- **is_repeated_guest** (value indicating if the booking name was from repeated guest) 1: Yes, 0: no
- **Reservation_status** 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

Data processing

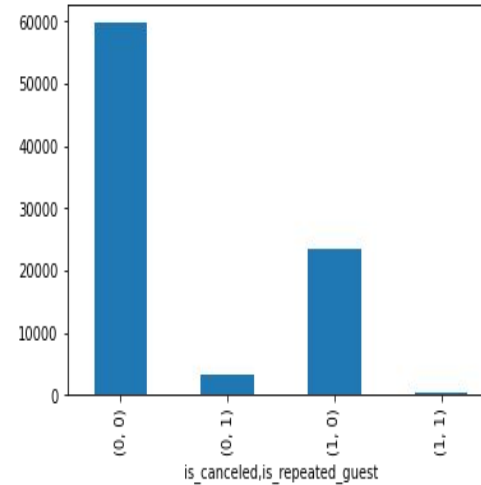
1. Checking whether there is duplicates values or not
2. dropping the duplicates values
3. After dropping the duplicates values country columns has the null values so we drop the null values in order to get a better result.
4. Handling outliers



Hotel wise analysis



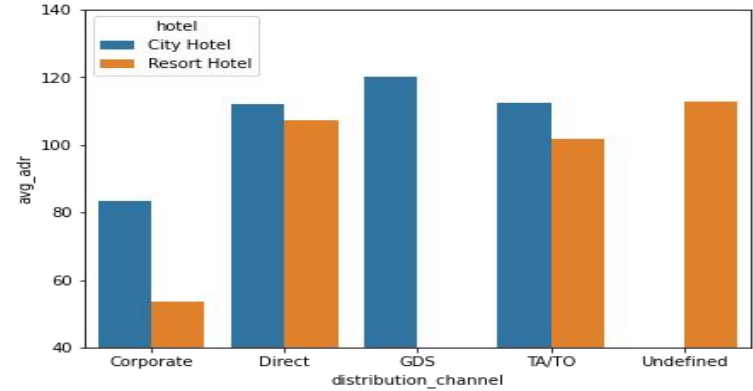
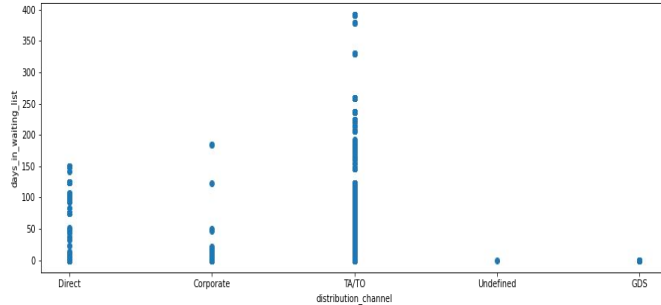
- so from the graph we can see that resort has 33,000 bookings and city hotel has around 53,000 bookings



- when the hotel booking is canceled and the customer is repeated guest the entry is almost zero this means the repeated guest is less likely

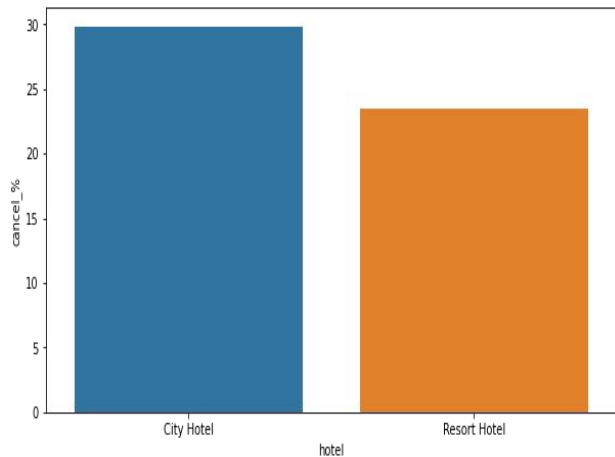
Distribution channel wise analysis

H

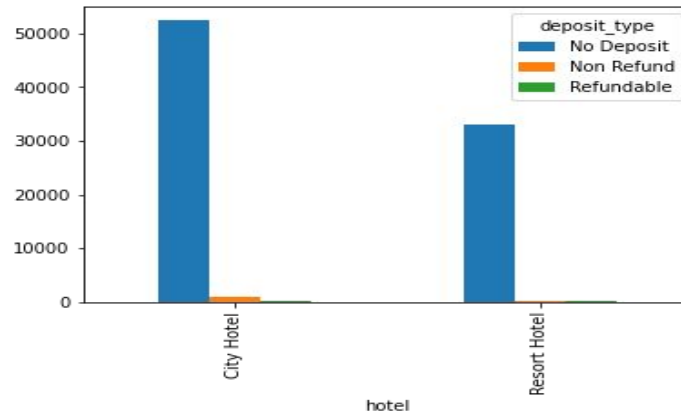


- Here we can see that the most of the guest are making reservation through TA/TO which is travel agency and tour operator
- Second most channel is direct
- TA/TO has the highest number of days on waiting list
- GDS channel brings higher revenue generating deals for City hotel, in contrast to that most bookings come via TA/TO. City Hotel can work to increase outreach on GDS channels to get more higher revenue generating deals.
- Resort hotel has more revenue generating deals by direct and TA/TO channel. Resort Hotel need to increase outreach on GDS channel to increase revenue

- Booking cancellation analysis

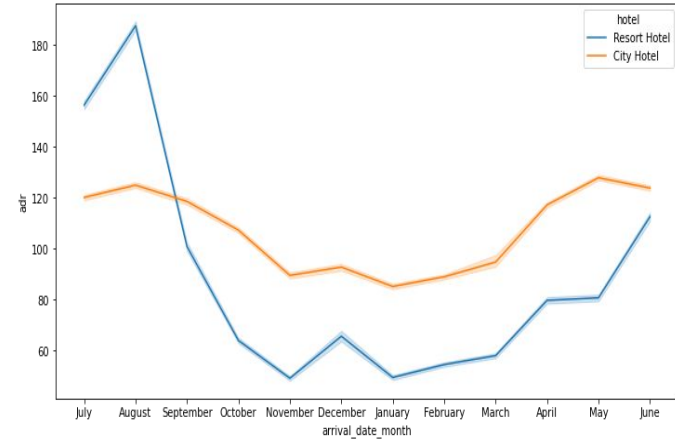
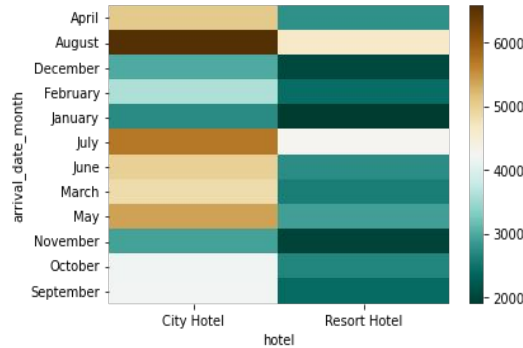


- There is 30% cancellation for city hotels and 23% cancellation for resort hotels



- Bookers are not required to send in a deposit in most bookings which could explain the high rate of cancellations

Time wise analysis

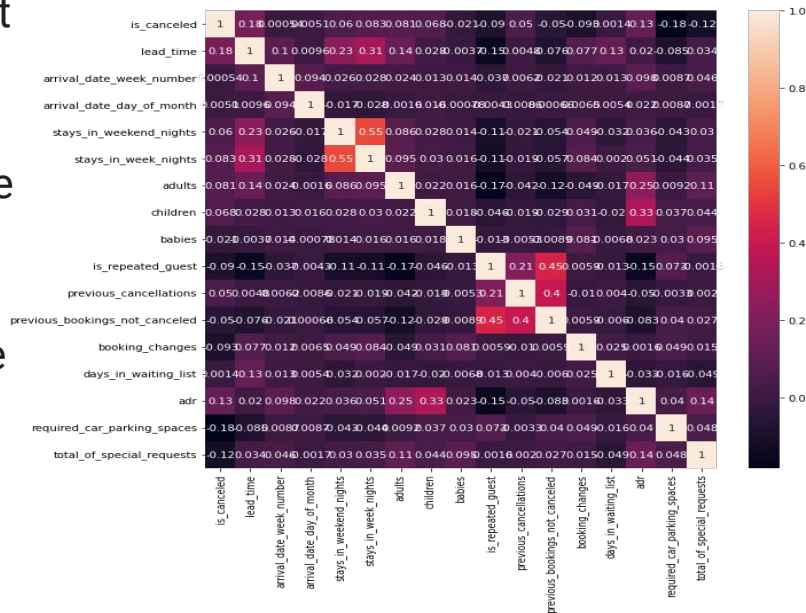


- this map gives the insight of which are the busiest month of the hotel ,, in city hotel the month with the highest bookings seems to be august and followed by july in resort hotel only busiest month seems to be august
- For resort hotels, the average daily rate is more expensive during august, july and september. For city hotels, the average daily rate is more expensive during august, july, june and may

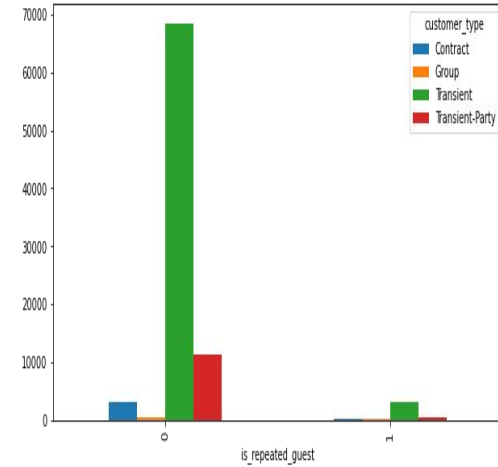
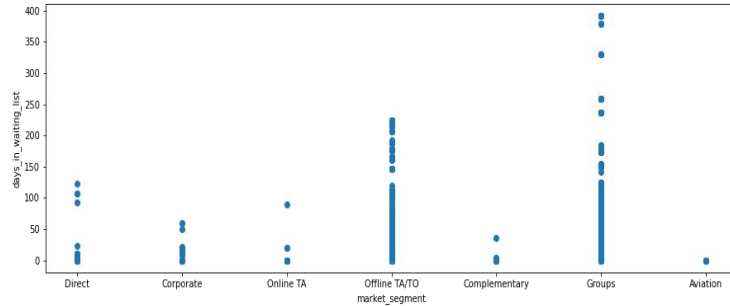
correlation

1) Total stay length and lead time have slight correlation. This may mean that for longer hotel stays people generally plan little before the actual arrival.

2) adr is slightly correlated with total_people which makes sense as more no. of people means more revenue, therefore more adr.

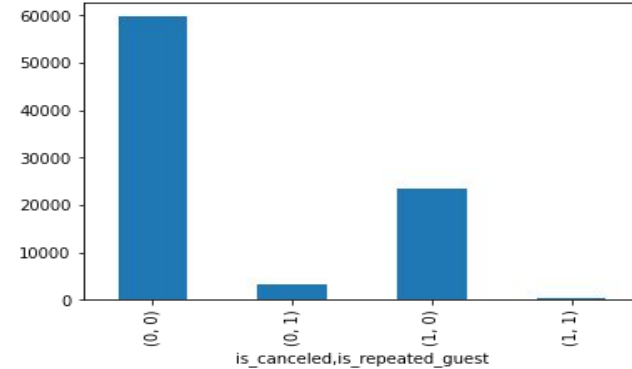
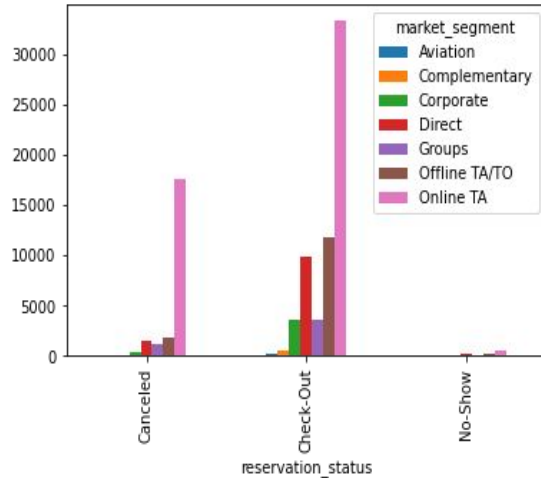


Some important questions



- aviation industry has the minimum number of days on the waiting list
- transient has the highest bookings

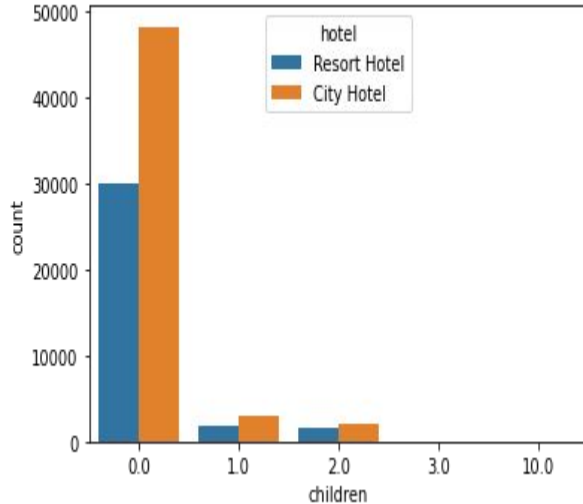
Some important questions(cont..)



- most of the bookings either in the canceled or checkout done by online TA
- when the hotel booking is canceled and the customer is repeated guest the entry is almost zero this means the repeated guest is less likely

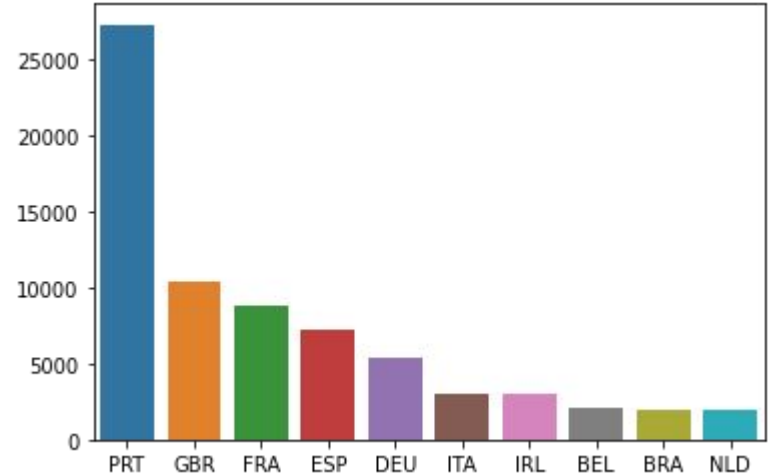
Children

Number of children



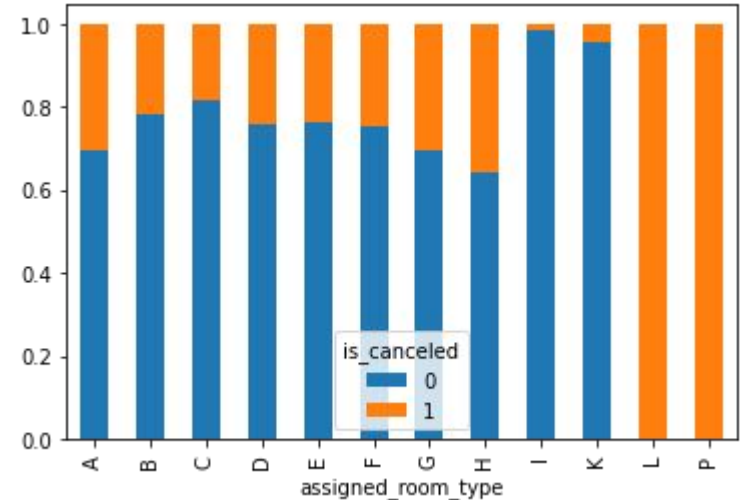
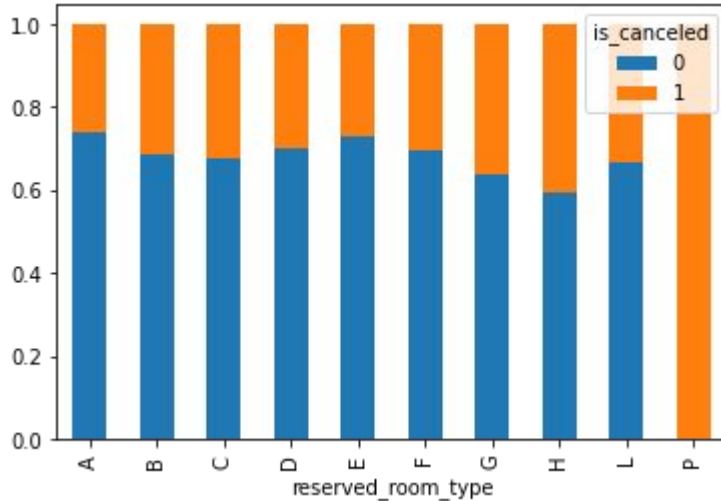
- if a customer is accompanied by a baby or children will not make request to hotels
- Families with children have no particular preference for the hotel type

country



- More visitors are from western europe, namely Portugal, France, Great Britain, and Spain being the highest.

Room type



1. **reserved_room_type vs is_canceled**- Booking for Room Type "P" will definitely get cancelled.

2. **assigned_room_type vs is_canceled** - A traveller who is assigned Room Type "L" or "P", will definitely cancel his booking. Travellers assigned Rooms "I" and "K" are least likely to cancel.

conclusion

- Around 60% bookings are for city hotels and around 40% bookings are for resort hotels therefore city hotel is busier than resort hotel ,also overall adr of city hotel is slightly higher than resort hotel
- the majority of reservations are for city hotels
- The number of repeated guests is too low
- most of the bookings either in the canceled or checkout done by online TA
- city hotels and resort hotels maximum number of bookings by online TA
- that aviation industry has the minimum number of days on the waiting list
- More visitors are from western europe, namely Portugal,France, Great Britain, and Spain being the highest.
- Families with children have no particular preference for the hotel type

- August and July are the most profitable and busiest months for both the hotels
- The confirmed bookings go from their lower value in January to their highest value in August
- Transients are the most common customer type, they represent 75% of the total customers.
- There is a disproportionate amount of cancellations on hotel bookings. Bookers are not required to send in a deposit in most bookings which could explain the high rate of cancellations
- Data suggests that hotel business could be improved by targeting working travelers or improving daily rates for weekdays