Capstone Project

Hotel Booking EDA

Team Member
Mohd. Navaid Ansari
Mohd Atif Ansari
Mohammad Anas Ansari

Content

Introduction to Hotel Management

Hotel Booking

The hotel reservation system software, guests can choose how long they will stay, the type of room they want, get add-ons, and pay securely online through a payment platform. In this data set we'll go in depth about what a hotel reservation system is, how it works, and the benefits that it offers to the hospitality sector. So let's get started.

Hotel Booking Data Set Insight

➢ The data sets of resort hotels and city hotels will be used which have the same structure, that describe with 119,390 rows and 32 columns observations and each observation represents hotel reservations.

➢ The data sets were taken in the period 2015 to 2017, including that bookings arrived effectively and orders that were cancelled.

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies
0	Resort Hotel	0	342	2015	July	27	1	0	0	2	0.0	0
1	Resort Hotel	0	737	2015	July	27	1	0	0	2	0.0	0
2	Resort Hotel	0	7	2015	July	27	1	0	1	1	0.0	0
3	Resort Hotel	0	13	2015	July	27	1	0	1	1	0.0	0
4	Resort Hotel	0	14	2015	July	27	1	0	2	2	0.0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
119385	City Hotel	0	23	2017	August	35	30	2	5	2	0.0	0
119386	City Hotel	0	102	2017	August	35	31	2	5	3	0.0	0
119387	City Hotel	0	34	2017	August	35	31	2	5	2	0.0	0
119388	City Hotel	0	109	2017	August	35	31	2	5	2	0.0	0
119389	City Hotel	0	205	2017	August	35	29	2	7	2	0.0	0

119390 rows × 32 columns

meal	country	market_segment	distribution_channel	is_repeated_guest	previous_cancellations	previous_bookings_not_canceled	reserved_room_type	assigned_room_type	booking_changes	deposit_type	agent
BB	PRT	Direct	Direct	0	0	0	C	C	3	No Deposit	NaN
BB	PRT	Direct	Direct	0	0	0	C	C	4	No Deposit	NaN
BB	GBR	Direct	Direct	0	0	0	A	C	0	No Deposit	NaN
BB	GBR	Corporate	Corporate	0	0	0	A	A	0	No Deposit	304.0
BB	GBR	Online TA	TA/TO	0	0	0	A	A	0	No Deposit	240.0
...	...	...	...	...	...	...	...	...	...	...	...
BB	BEL	Offline TA/TO	TA/TO	0	0	0	A	A	0	No Deposit	394.0
BB	FRA	Online TA	TA/TO	0	0	0	E	E	0	No Deposit	9.0
BB	DEU	Online TA	TA/TO	0	0	0	D	D	0	No Deposit	9.0

company	days_in_waiting_list	customer_type	adr	required_car_parking_spaces	total_of_special_requests	reservation_status	reservation_status_date
NaN	0	Transient	0.00	0	0	Check-Out	2015-07-01
NaN	0	Transient	0.00	0	0	Check-Out	2015-07-01
NaN	0	Transient	75.00	0	0	Check-Out	2015-07-02
NaN	0	Transient	75.00	0	0	Check-Out	2015-07-02
NaN	0	Transient	98.00	0	1	Check-Out	2015-07-03
...	...	...	...	...	...	...	...
NaN	0	Transient	96.14	0	0	Check-Out	2017-09-06
NaN	0	Transient	225.43	0	2	Check-Out	2017-09-07
NaN	0	Transient	157.71	0	4	Check-Out	2017-09-07
NaN	0	Transient	104.40	0	0	Check-Out	2017-09-07
NaN	0	Transient	151.20	0	2	Check-Out	2017-09-07

Column Names

```
['hotel',
 'is_canceled',
 'lead_time',
 'arrival_date_year',
 'arrival_date_month',
 'arrival_date_week_number',
 'arrival_date_day_of_month',
 'stays_in_weekend_nights',
 'stays_in_week_nights',
 'adults',
 'children',
 'babies',
 'meal',
 'country',
 'market_segment',
 'distribution_channel',
 'is_repeated_guest',
 'previous_cancellations',
 'previous_bookings_not_canceled',
 'reserved_room_type',
 'assigned_room_type',
 'booking_changes',
 'deposit_type',
 'agent',
 'company',
 'days_in_waiting_list',
 'customer_type',
 'adr',
 'required_car_parking_spaces',
 'total_of_special_requests',
 'reservation_status',
 'reservation_status_date']
```

Problem Statement

1) What type of hotel does a customer book in general?
2) Which year most of customers arrive?
3) Which are the busiest months in the year of hotel booking?
4) How many booking were cancelled.
5) From which country most guests arrive.
6) Which month is the most occupied and which is the least occupied.
7) How customers booked their hotel from market segment.
8) How many customers were stays in week nights and booking changes?
9) How many repeated customers?
10) Deposit type payment of customer?

Methodology: Data Cleaning and Processing

- **Data processing-1:** In first part we have to remove unnecessary features. Since there were many column with all null values.

- **Data processing-2:** we have manually go through each features select from part 1, and encoded the categorical features.

- **EDA:** In this part we do some exploratory data analysis (EDA) on the features selected in part-1 and part-2 to see the trend.

- **Create the model:** Finally, in the last part but not the last part we creates models and function, and import some libraries it's not the easy task. Its also an iterative process. We show how to start with simple models and then add complexity for better performance.

Data processing and visualization

➢ This Data will take you through the major considerations to keep in mind regarding hotel management and throw some tips and ideas along the way, to help you run a better hotel business.

➢ The process of analyzing data to better understand the contents of the data (preprocessing and visualization).

➢ Clustering the process of finding a group structure in data in such a way that each data in a group is similar to each other and not similar to data in other groups .

Discussion on problem statement
What type of hotel does a customer book in general?

- As we can see on graph most of the people prefer around 67% of the customers visit city hotel and around 33% customers visit resort hotel.

- The reason behind why most of people like city hotel. A city hotel is what you probably know best. It provides accommodation and meals to travelers. Often times, people come from all over the world to stay at a hotel so that they can tour around the place that they are staying.

- A resort is the destination. Oftentimes, resorts are located in scenic areas but not near a city or anything really. People visit a resort to get away from other people and cities. They stay in the resort and probably don't travel around as they may do when staying at a city hotel.
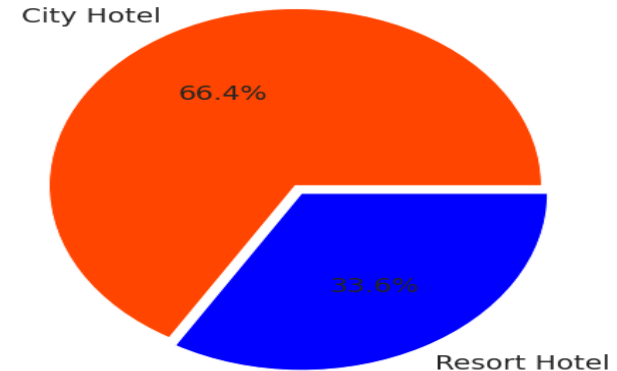
Which year most of customers arrive?

- As you can see on graph most of guest arrive in year 2016. And the most of customer arrive in city hotel as compare to resort hotel. Then the least of number customer was arrive in 2015 and this graph increase in 2016. And again this graph will be down in 2017 as compare 2016.
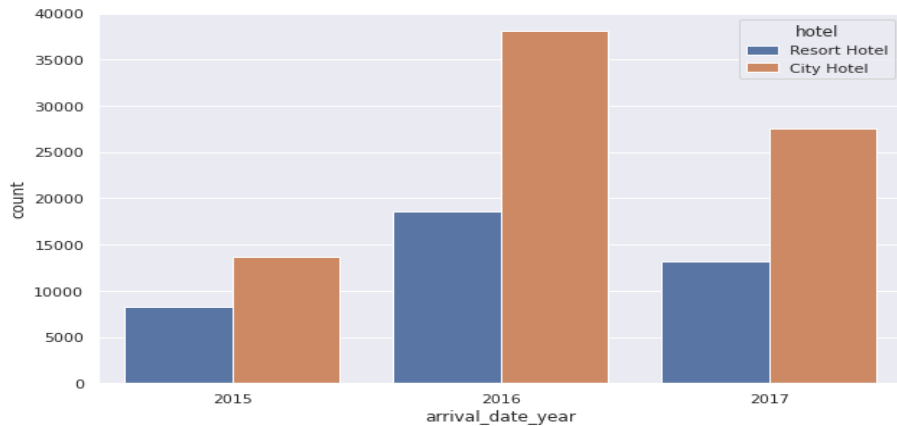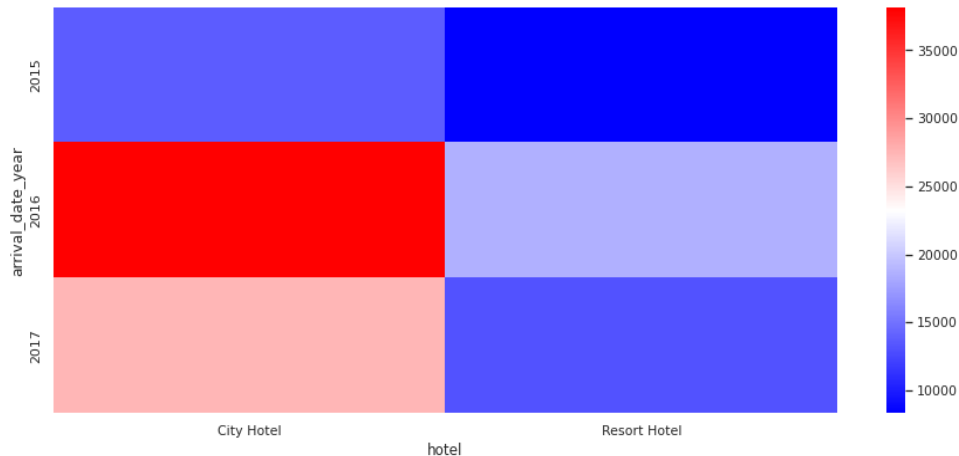
Which are the busiest months in the year of hotel booking?

- That illustrate us the busiest month of the year as you can see in the heatmap graph city hotel makes on the top with highest busiest month in maroon colour that define most busiest month is august.

- The value is more than eight thousand in numbers.
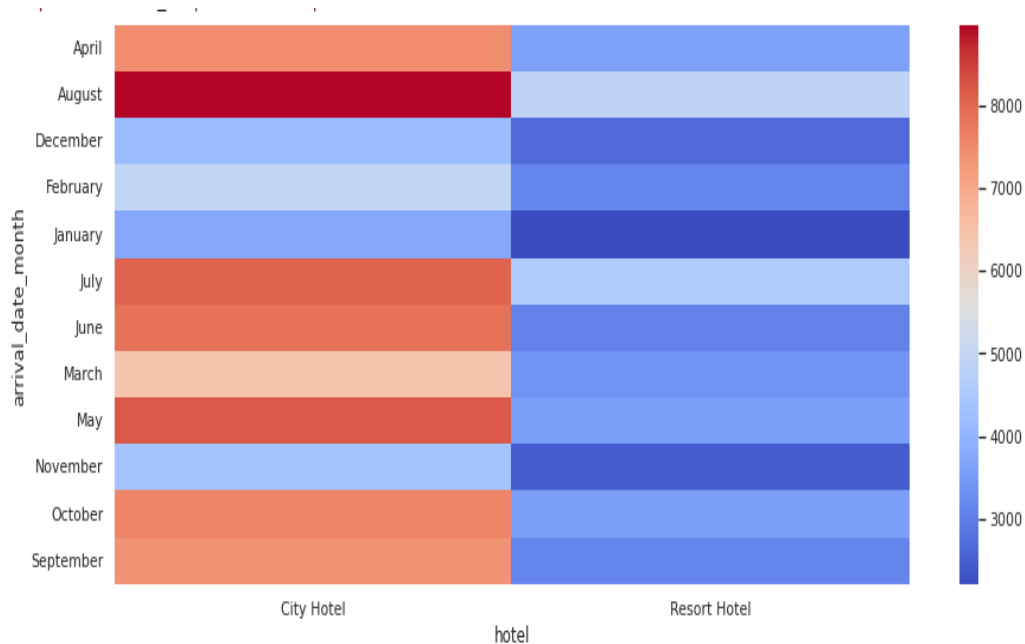
How many booking were cancelled

Total no. of Cancelation = 44224
Resort Hotel Cancelation = 11122
City Hotel cancelation = 33102

	hotel	arrival_date_year	is_canceled
0	City Hotel	2015	6004
1	City Hotel	2016	15407
2	City Hotel	2017	11691
3	Resort Hotel	2015	2138
4	Resort Hotel	2016	4930
5	Resort Hotel	2017	4054

Total Bookings Canceled = 37.041628277075134
Resort Hotel Bookings Canceled = 27.763354967548675
City Hotel Bookings Canceled = 41.72696331778646

- As we can see in data set the number of customers were canceled the hotel.

- The maximum number of customers canceled the city hotel in 2016, which around (15400) and the canceled the resort hotel in 2016, which around (4900).

- The proportion of guests who canceled the hotel, which is around 37%, is still quite large. So that further review is needed regarding the cause of the guest canceling the booking.

Graph of total cancelation by hotel type

- In this graph 0 & 1 represents:
  0 : not canceled the hotel
  1 : is canceled the hotel

- In this stats bar chart illustrates city hotel that makes on top with cancelled booking.

- The value of cancellation booked at city hotel and resort hotel ,the value of city hotel is more than 30000 and the value of resort hotel is more than 10000.
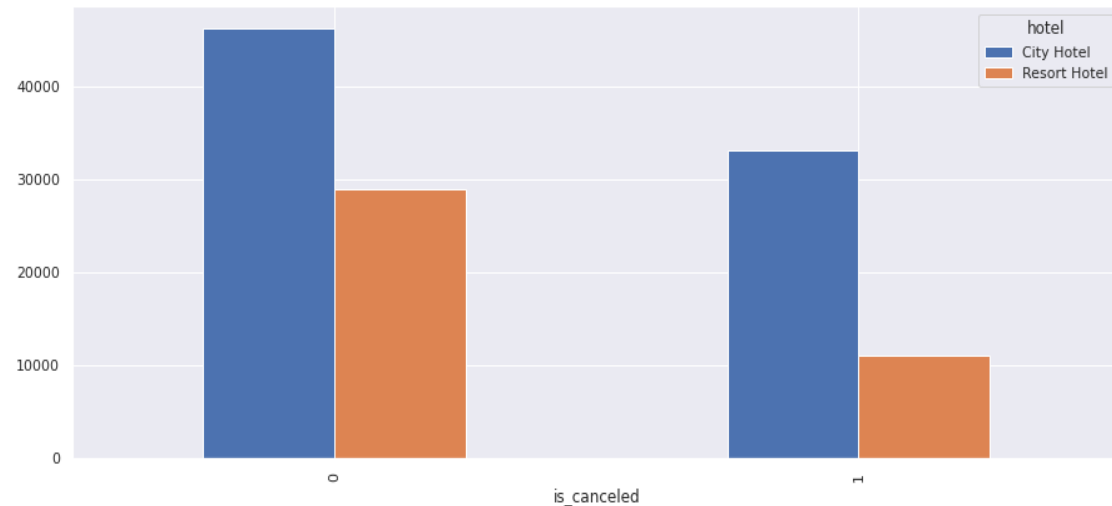
➢ The chart values states that cancelled ratio of resort hotel is less than the city hotel.

From which country most guests arrive

- Guests who came from countries are dominated by Portugal (PRT), Great Britain (GBR), France (FRA), Spain (ESP), and Germany (DEU) etc.

- Recommended hotel can adjust the culture of the hotel. In accordance with the customs of the guest country, such as making food that tastes acceptable to guests from other countries. It is recommended that marketing from 5 dominant countries be optimized to increase the number of visitors.
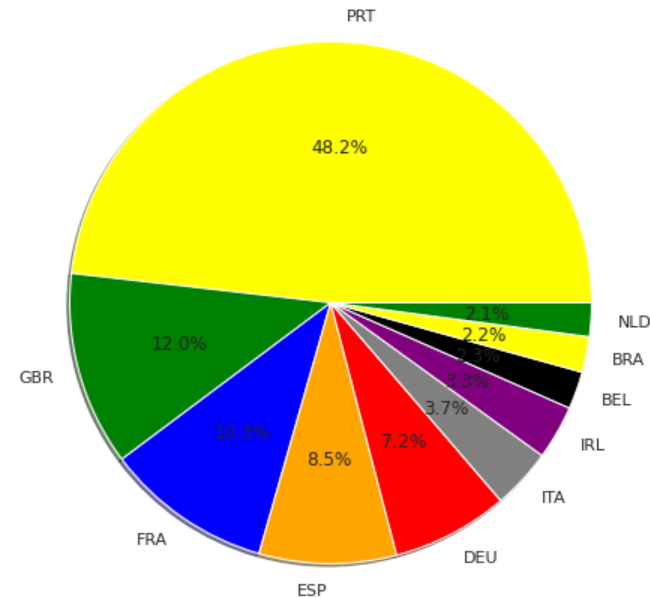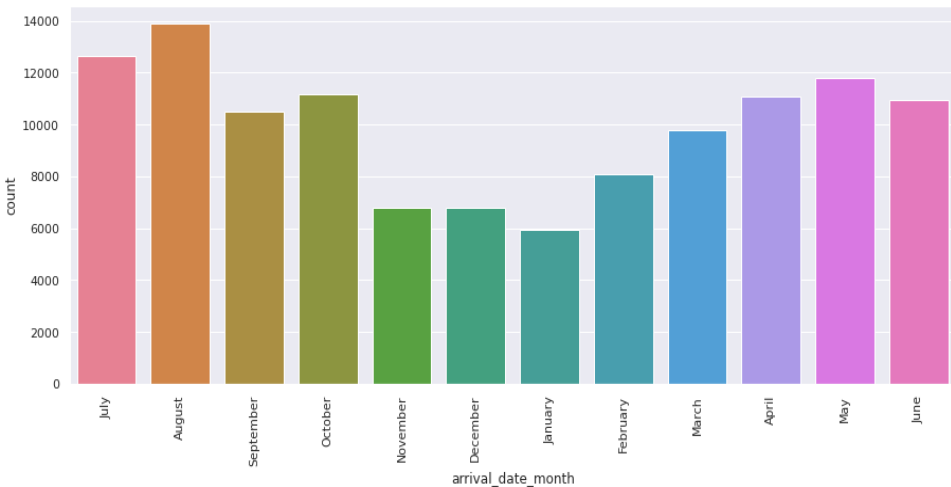


```
PRT     48590
GBR     12129
FRA     10415
ESP      8568
DEU      7287
         ...
PYF         1
BWA         1
NCL         1
HND         1
NPL         1
Name: country, Length: 177, dtype: int64
```

# Which month is the most occupied and which is the least occupied

As you can see in the left graph that show how many customers arrive in months of year time span.

The highest number of customers came from almost greater than13000 in august month the orange colour show the line. And least number of customers came in start of the month.

The bar chart right explicit information about customer visit in city hotel and resort hotel. And above graph show the total number of customers arrival in hotel. from that how many in days of month.

The blue color bar line define the city hotel while orange color bar line illustrates resort hotel.
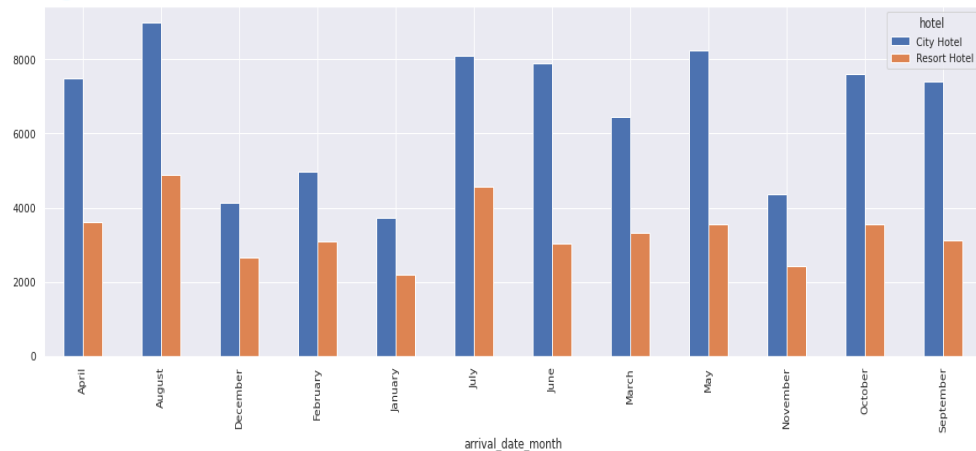
Customers arrive per day in a month
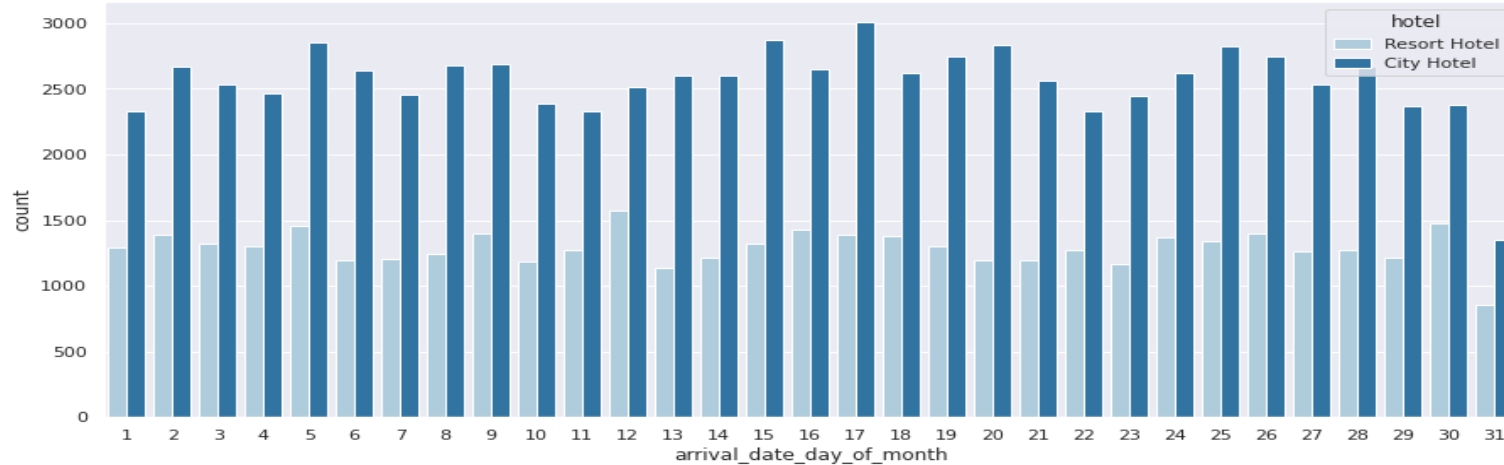
- As you asses that explicit contrast from other bar chart cause it illustrates day's stats when in most of customers arrive in a month.

The result shows that most of customer booked hotel from city hotel on 17th day of the month and resort hotel on 12th day of the month. The least number of customers came at the end of the month.

How customers booked their hotel from market segment

- In hotel marketing, the marketing budget should be prioritized for marketing in the online field. This is because more customers choose to book online.

- You will ultimately increase your bookings. At the end of the day, the point of every feature within your hotel management business solution is to boost the bookings that you get at your hotel.

- Whether you want to increase your off-season bookings or you want to expand your offerings to new market segments, you will be successful if you select the right hotel management software for your property.
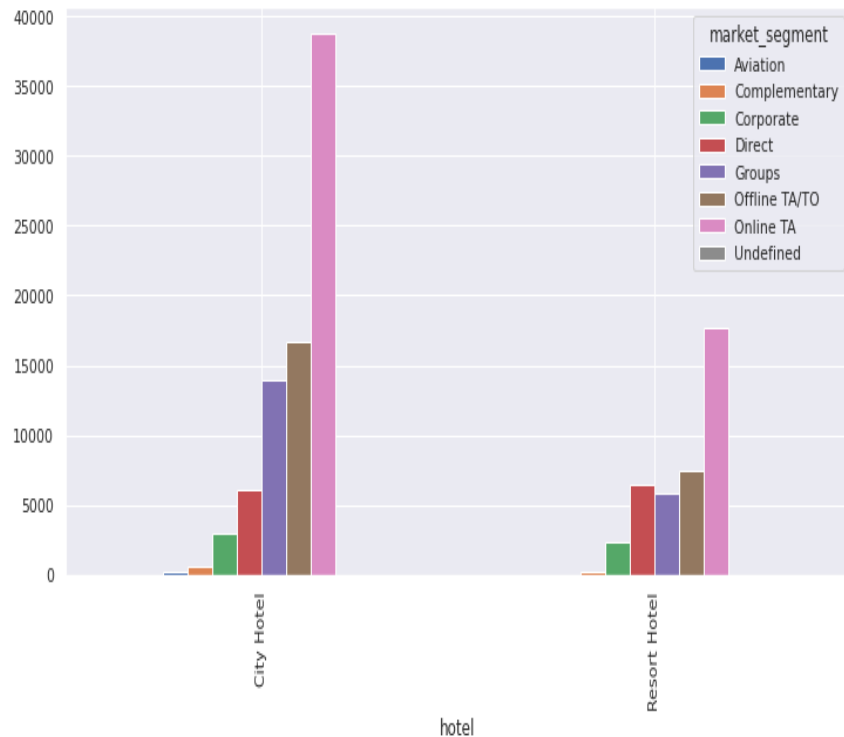
How many customers were stays in week nights

- Resort hotel makes on the top on the stats that explicit that most of the customers stay week night.
- Almost customers stay 3.5 weeks in resort type hotel.
- While on other hand city hotel is bit less than resort hotel stats.
- Customers almost stay 2.0 weeks in city hotel.
- The most of the customers stay in night which prefer the resort hotel.
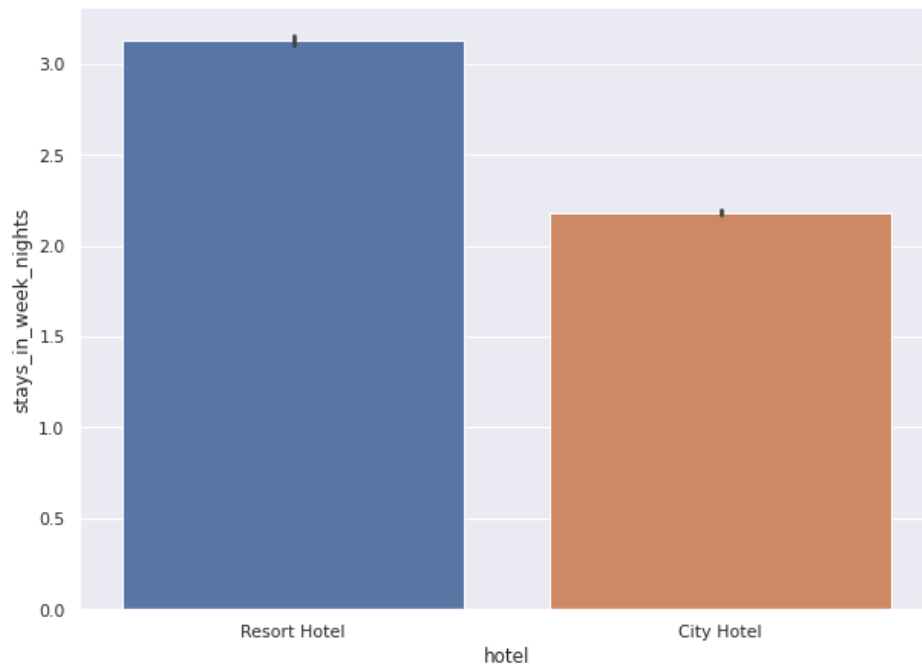
How many customers changed the booking

- That indicates the comparison between which customers change the booking most resort hotel or city hotel.

- The graph illustrates that the stats dominant by resort hotel's customers in changing the booking.
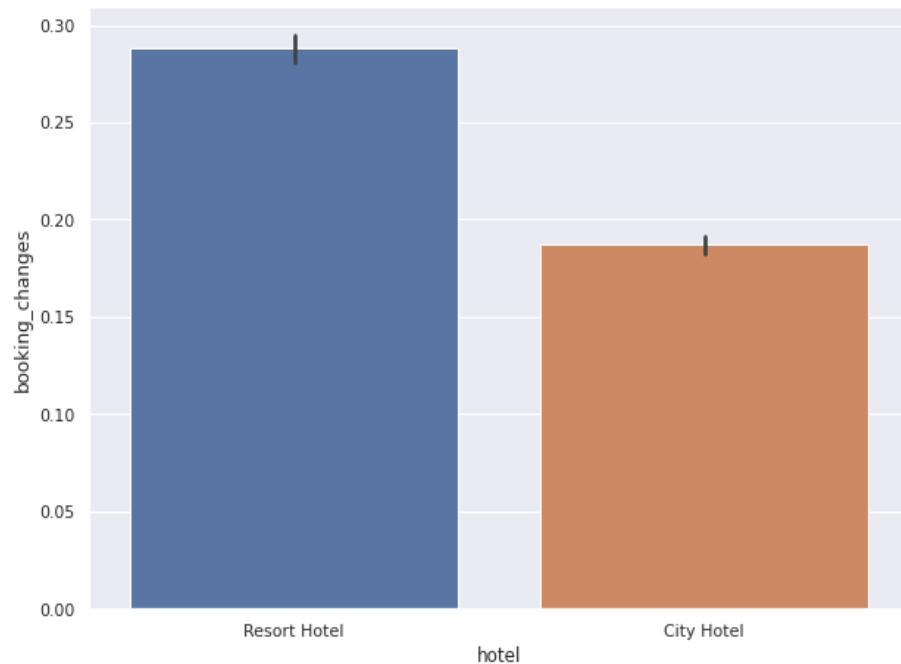
Distribution channel of customers arrival

- Channel management system is a single interface to control and distribute inventories across different channels such as Corporate, Direct, GDS, TA/TO etc.

- As we can see on bar chart the most of customer comes from TA/TO (Travel Agencies / Tour operators) in hotel.

- A channel manager connects directly to a central reservation system that holds information about the availability and cost of hotel rooms, sharing this information via the distribution channels. Add more attractive facilities and as per discount facilities so that the demand for them increases.
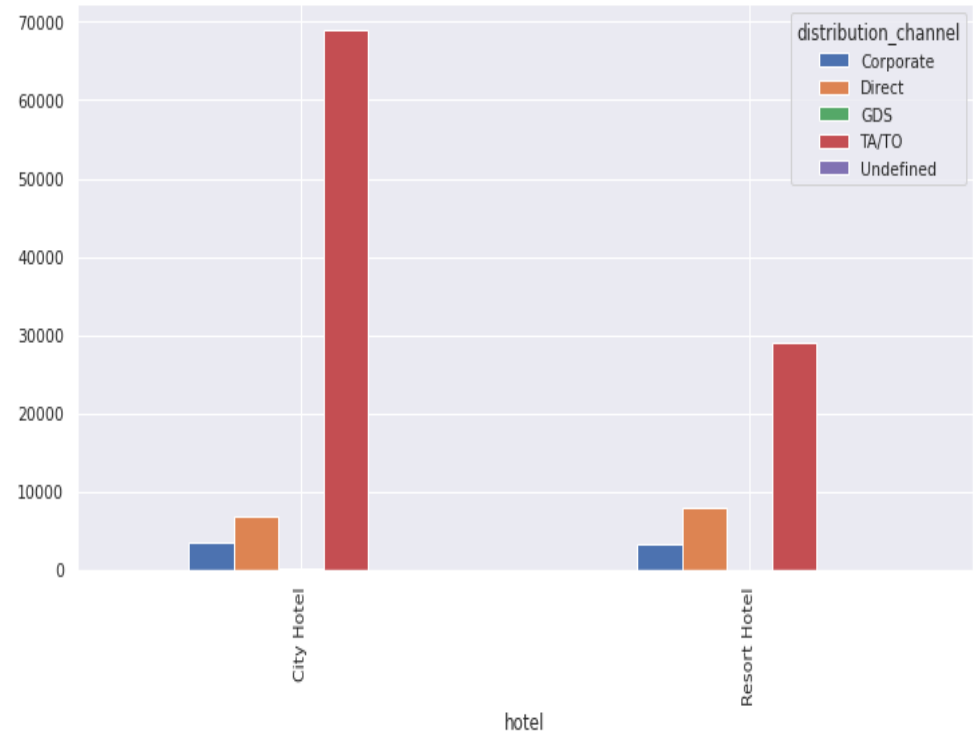
How many repeated customers

- As we can see in the graph that the number of guests who have come, the guest does not come again in this hotel their number is very less.

- A management approach focused on customer satisfaction can improve customer loyalty, thus increasing the positive image of the hotel booking.

- Research on the topic of guest satisfaction, which translates into the consideration of whether or not customers will return to a hotel or advise it to other customers, is pivotal to the success of the hotel business.
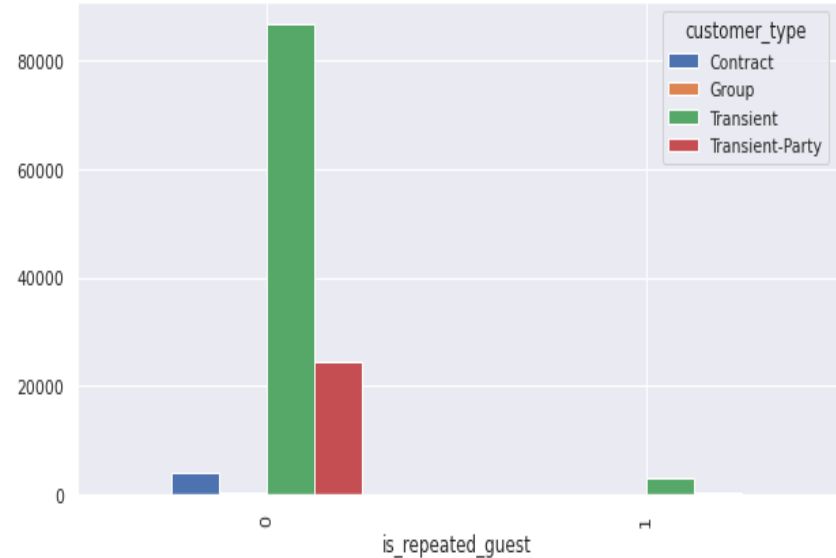
Deposit type payment of customer

- The term reservation is defined as blocking or booking a particular room type for a guest, for a definite period of time, for a particular guest.

- Revenue management is another huge part of managing your hotel. Because there are very few non-refundable deposit types, it is necessary types, it is necessary to add more attractive facilities and health insurance so that the demand for them increases.

- The role of reservation department is not limited to making reservation. It also maintains records of the hotel occupancy, which helps in planning sales and marketing strategies. Deposit type is one of most rule play in reservation system.
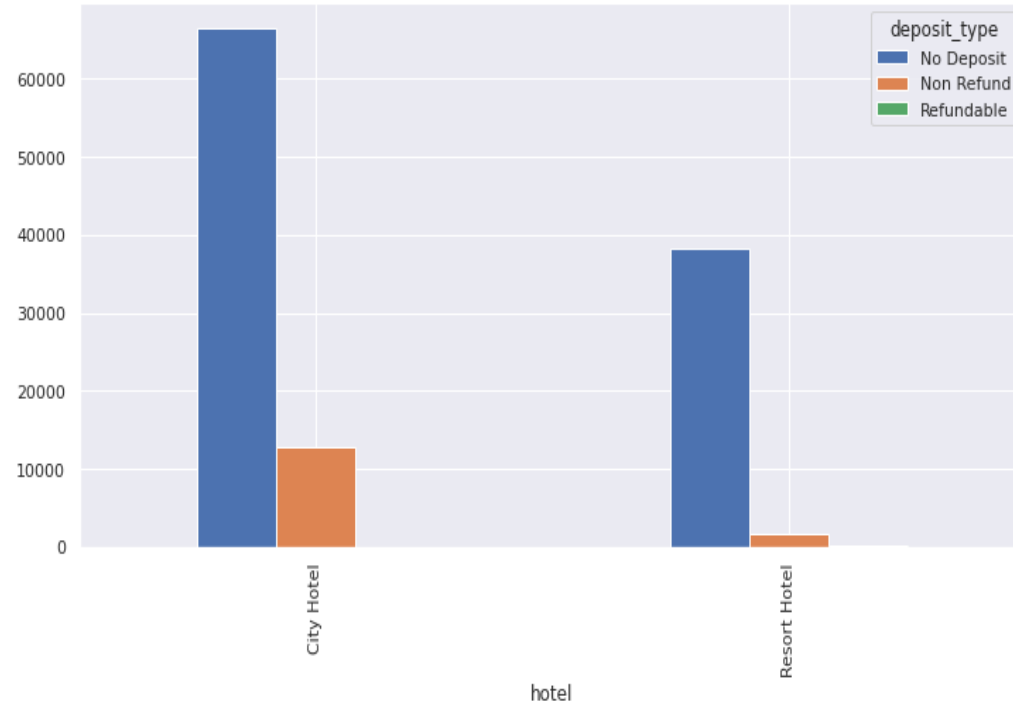
Correlation of all data

- Heat maps are a great tool for visualizing complex statistical data.

- There are one red colored squares that get my attention. This is refers to the is cancelled.

- We can see our new features, Room and net cancelled have a higher correlation with is cancelled than most of the other columns.
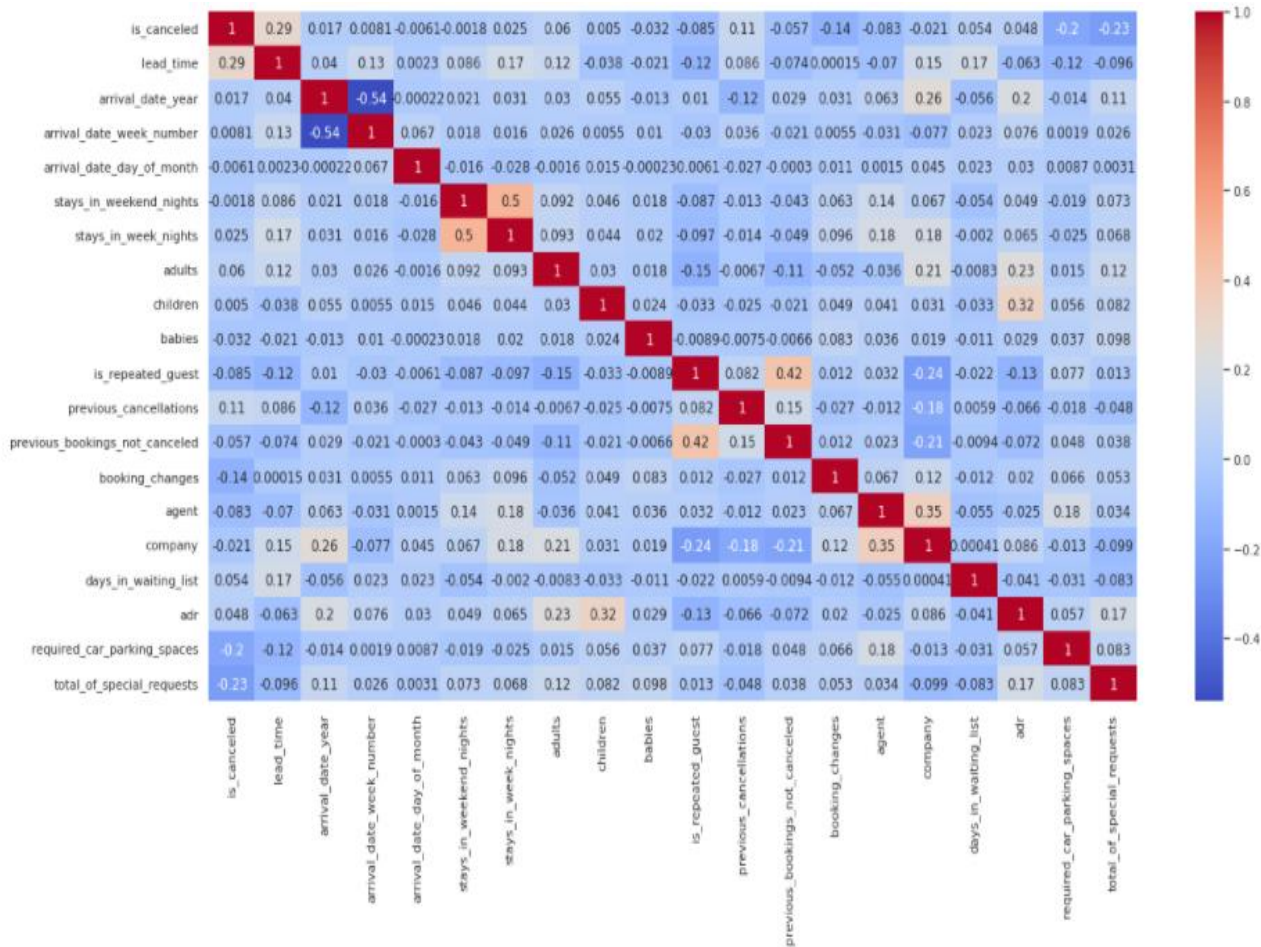
Conclusion:

- City hotel was booked most by the customer so we should spend most targeting fund in these hotel. The population booked the City hotel more than 60% .

- We should target the month between July and August as this was peak month due to summer period.   And the least bookings were made at the start and end of the year. In the year 2016 bookings more than double ,compared to the previous year. But the bookings decre ased by almost 15% in that year 2017.

- Total number of bookings were canceled around 37%.

- Majority of the guests are from Western Europe. We should spend a significant amount of our budget on those area. Portugal(PRT), the UK, and France(FRA), Spain and Germany are the top countries from most guests come, more than 80% come from these 5 countries.

- We also saw that due to no deposit policy, cancelation of booking was more that,
  As repeated guest is less, so we should target our advertisement on guests to increase returning guests.

- We then performed feature selection and feature engineering, and then made the predictive model using the Decision Tree to predict whether our customer/guest will cancel the booking or not.
  And we achieved 99% accuracy.

From the discussion above, it can be concluded that the recommendations that can be made are:

1. Further review is welcome regarding the cause of the guest who canceled the booking.

2. Have a discount/promo in November-January

3. To adjust the number of hotel guests, you can hire contract employees during peak months.

4. It is recommended that the hotel can adjust its hotel culture according to the habits of the guest country, such as making food that tastes acceptable to guests from other countries.

5. It is recommended that marketing from 5 dominant countries (Portugal, Great Britain, Germany, Spain, and France) be optimized to increase the number of visitors.

AI

- We've looked at what a hotel reservation system is, how it works, and the benefits to the hotel booking. The benefits of an Hotel booking(EDA) system range from reducing human error to improved efficiency, from making data collection and rewards management easier to giving you a competitive edge.

- With the advent of technology, having a hotel reservation system software is the one of the easiest changes a hotel can implement to reap the benefits mentioned in this data set. It allows you to keep occupancy up and ensure everything is running smoothly with little intervention from hotel staff. The more time that is freed up for staff to take care of administrative duties, the more time they can dedicate to offering the best service possible.

References

1) https://www.almabetter.com/

2) https://www.wikipedia.org

3) https://www.kaggle.com/

4) https://github.com/