

Capstone Project

Hotel Booking EDA

Team Member

Mohd. Navaid Ansari

Mohd Atif Ansari

Mohammad Anas Ansari

Capstone Project

Hotel Booking EDA

By – Mohd. Navaid Ansari

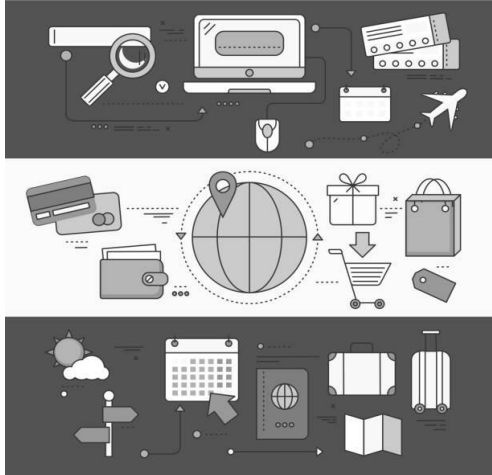
Content

- Introduction to Exploratory Data Analysis
- Hotel Booking Data Set Insight
- Problem Statement
- Methodology
- Discussion on problem statement
- Correlation of all data
- Conclusion
- References

Exploratory Data Analysis

- EDA is a technique for comprehending, summarizing, and analyzing the contents of a dataset.
- Usually EDA is preferred to answer a specific question or prepare for more sophisticated modelling.
- Plotting in EDA consist of Box plot, Histograms, Scatter plot and many more.

Why EDA on Hotel Booking Data



Data analytics in the hotel industry is key to marketing strategy, building customer loyalty, and enhancing productivity. It enables hotels to personalize experiences for their guests, introduce better hotel pricing strategies, and expand their customer base.

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 x 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) How customers booked their hotel from market segment.
- 7) How many customers were stays in week nights and booking changes?
- 8) How many repeated customers?
- 9) Deposit type payment of customer?
- 10) Correlation of Data Set

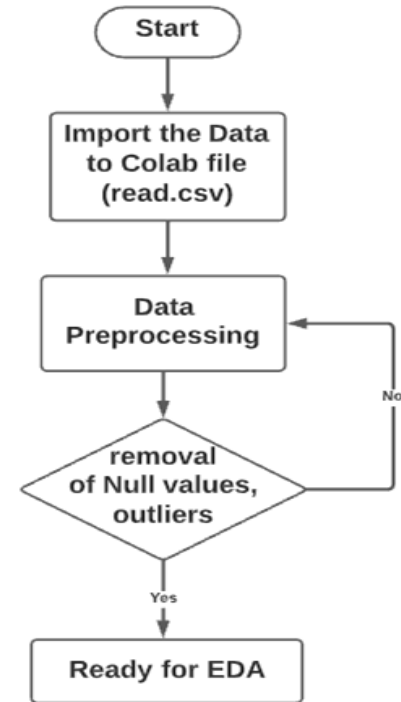
Methodology: Data Cleaning and Processing

Start : Import the data set to google colaboratory file

Data preprocessing : In first part we have to remove unnecessary features. Since there were many column with all null values and outliers.

we have to manually go through each column, and encode the categorical features.

EDA : In this part we do some exploratory data analysis (EDA) on the features selected.



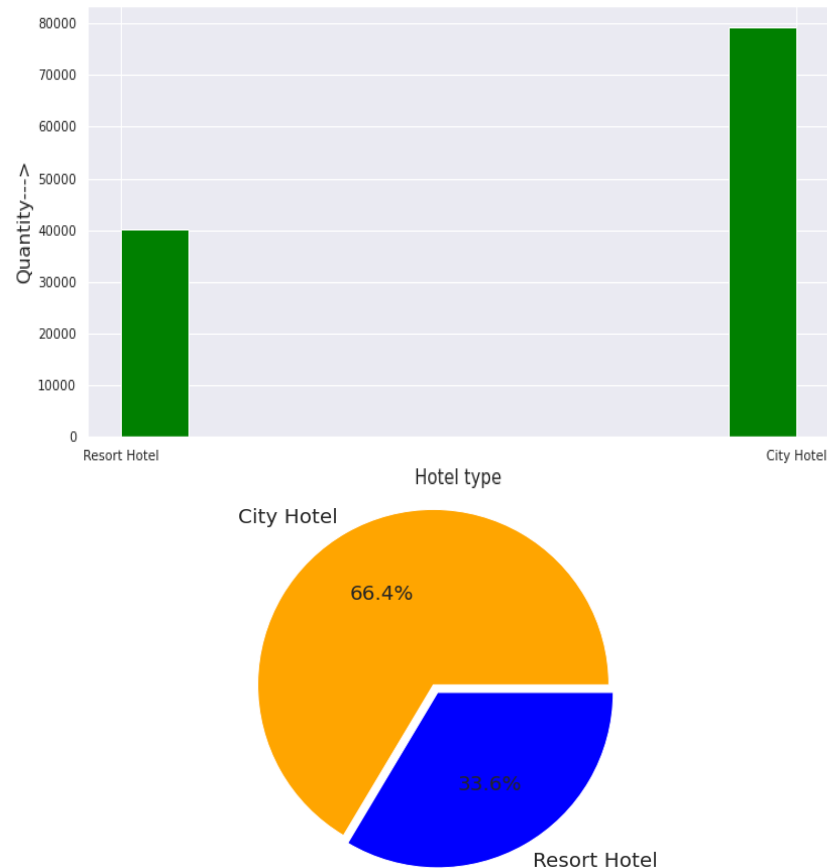
Data processing and visualization

- 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?

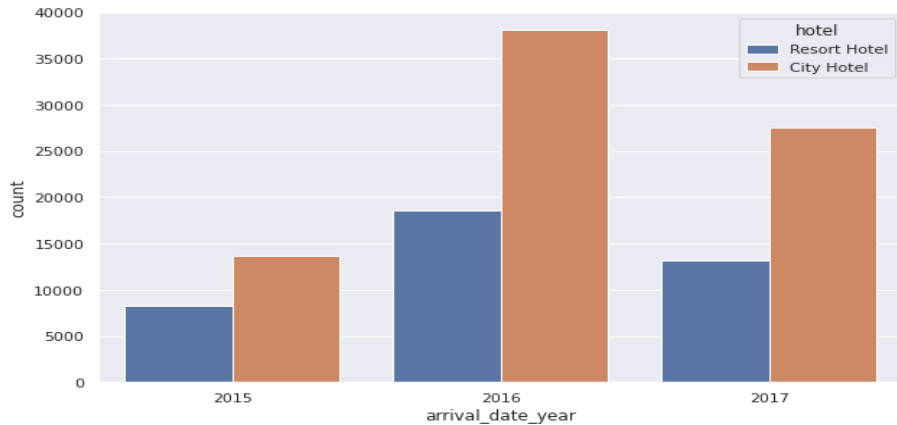
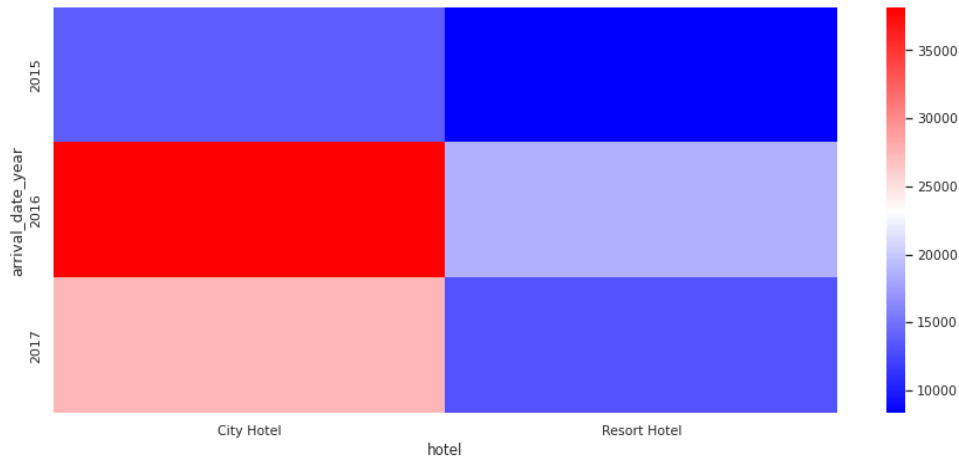
- As we can see on graph most of the people prefer around 67% of the customers visit city hotel and around 33% customers visit resort hotel.
- The reason behind why most of people like city hotel. A city hotel is what you probably know best. It provides accommodation and meals to travelers. Often times, 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.



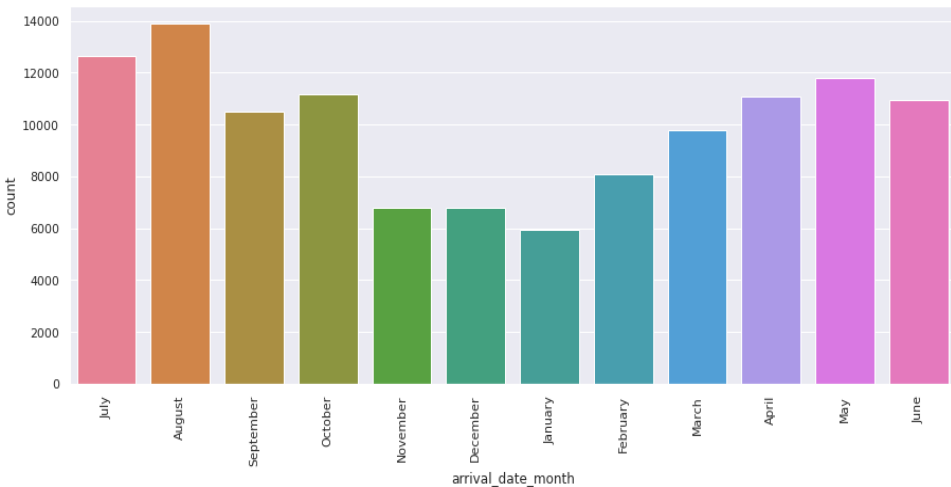
Which year most of customers arrive?



- As you can see on graph most guests arrive in year 2016. And the most of customer arrive in city hotel as compare to resort hotel. Least number of customer was arrive in 2015 and this graph boost in 2016. And again this graph will be down in 2017 as compare 2016.



Which are the busiest month in the year of hotel booking

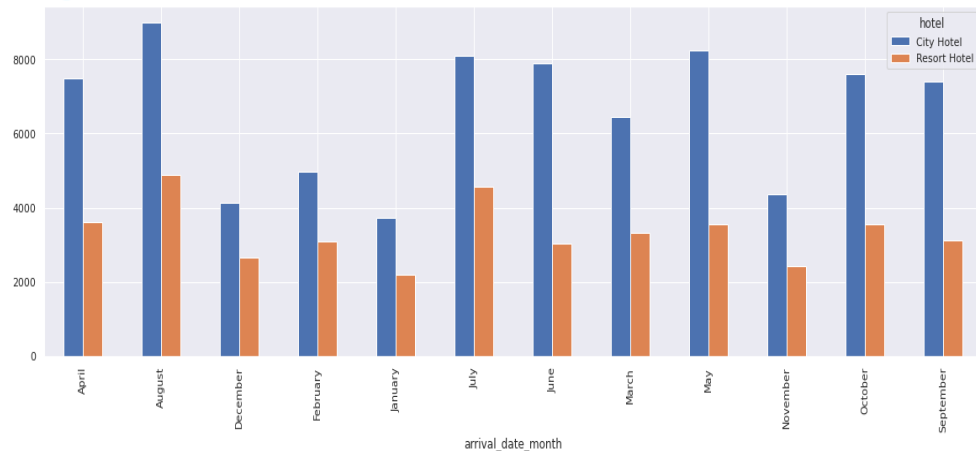


As you can see in the left graph that show how many customers arrive in months of year time span.

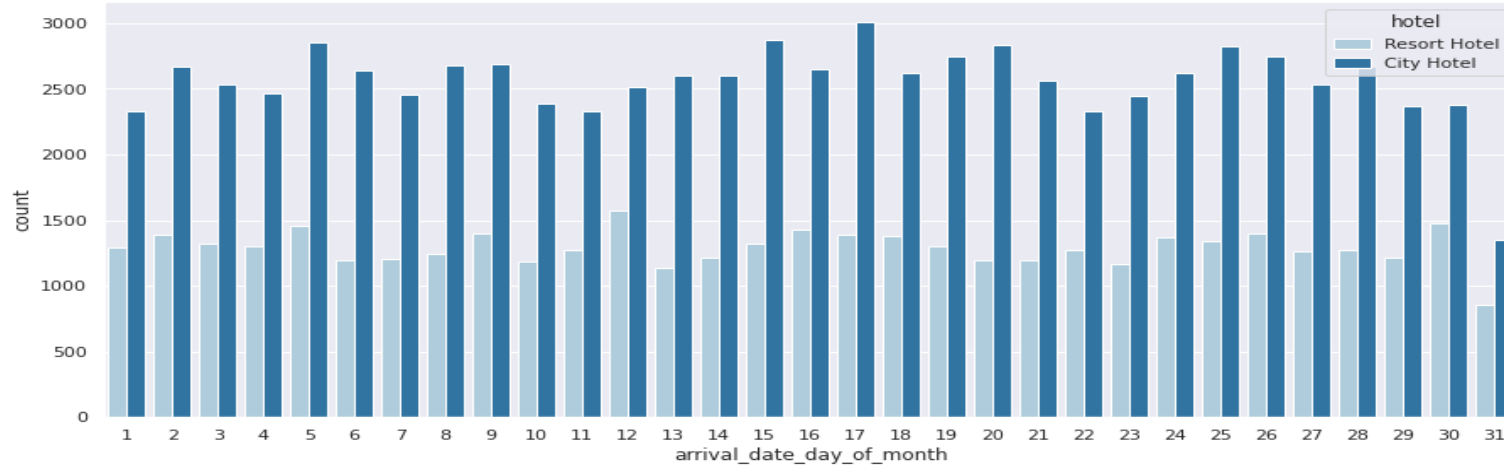
The highest number of customers came from almost greater than 13000 in August month the orange colour show the line. And least number of customers came in start of the month.

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 assess 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 many booking were cancelled

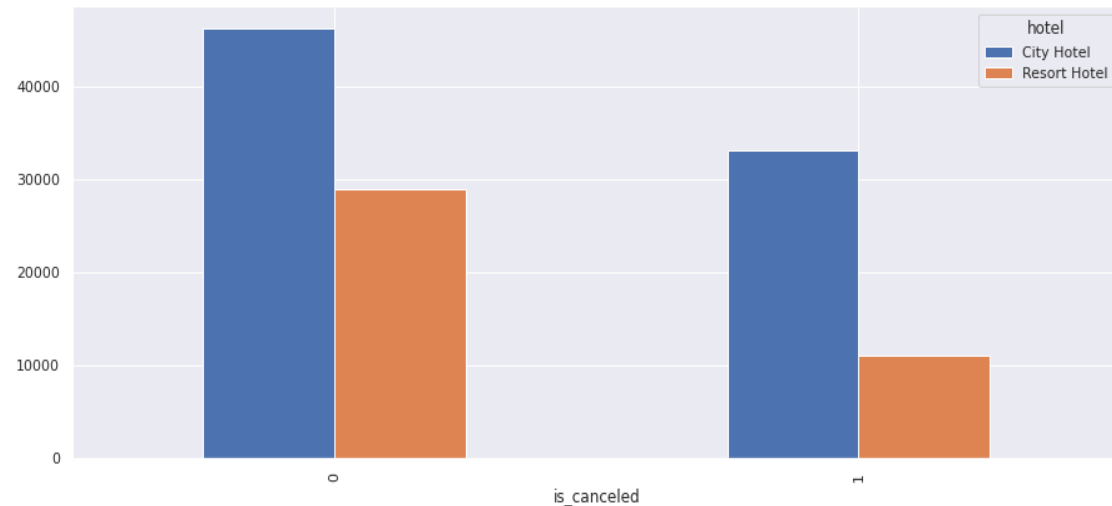
Total no. of Cancellation = 44224
Resort Hotel Cancellation = 11122
City Hotel cancellation = 33102

	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

- 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.

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

Graph of total cancellation by hotel type

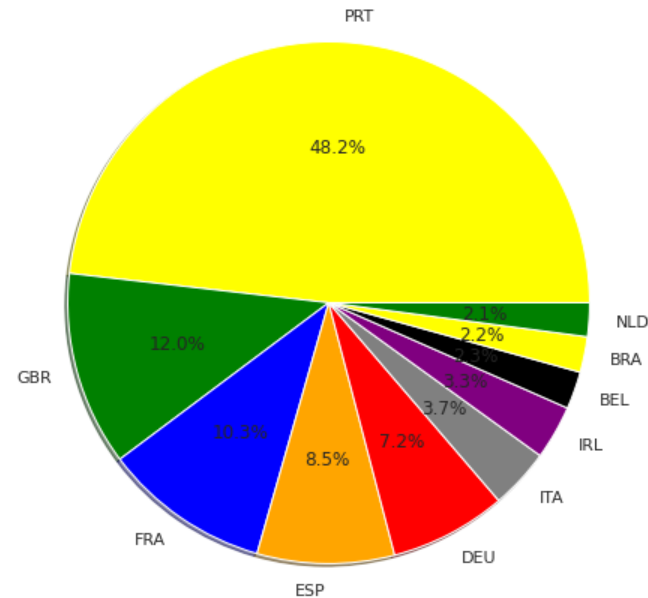


➤ The chart values states that cancelled ratio of resort hotel is less than the city hotel.

- 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.

From which country most guests arrive

- Guests who came from countries are dominated by Portugal (PRT), Great Britain (GBR), France (FRA), Spain (SPN), 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

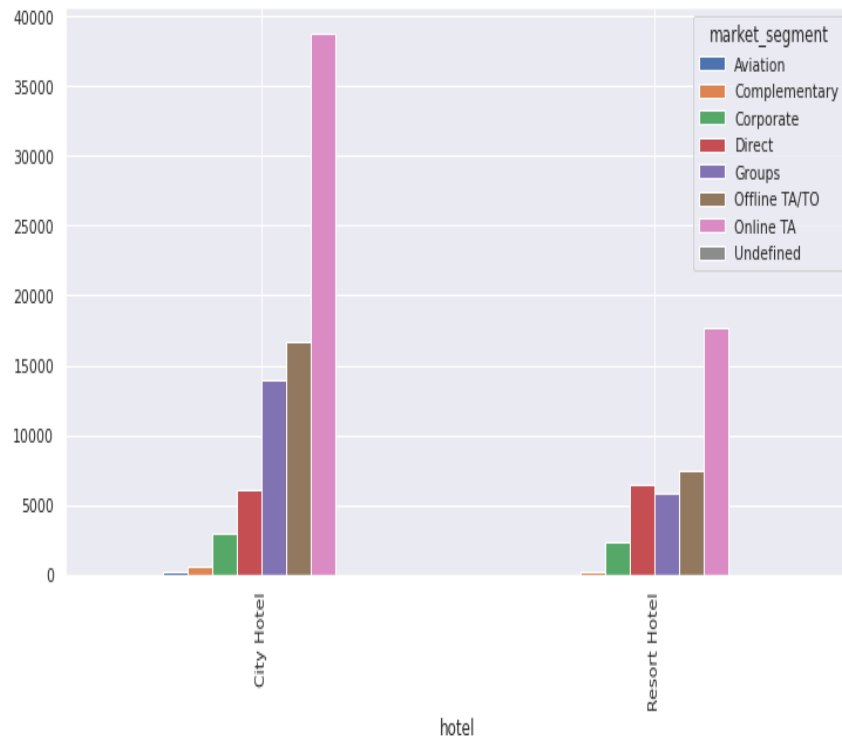
...

PYF	1
BWA	1
NCL	1
HND	1
NPL	1

Name: country, Length: 177, dtype: int64

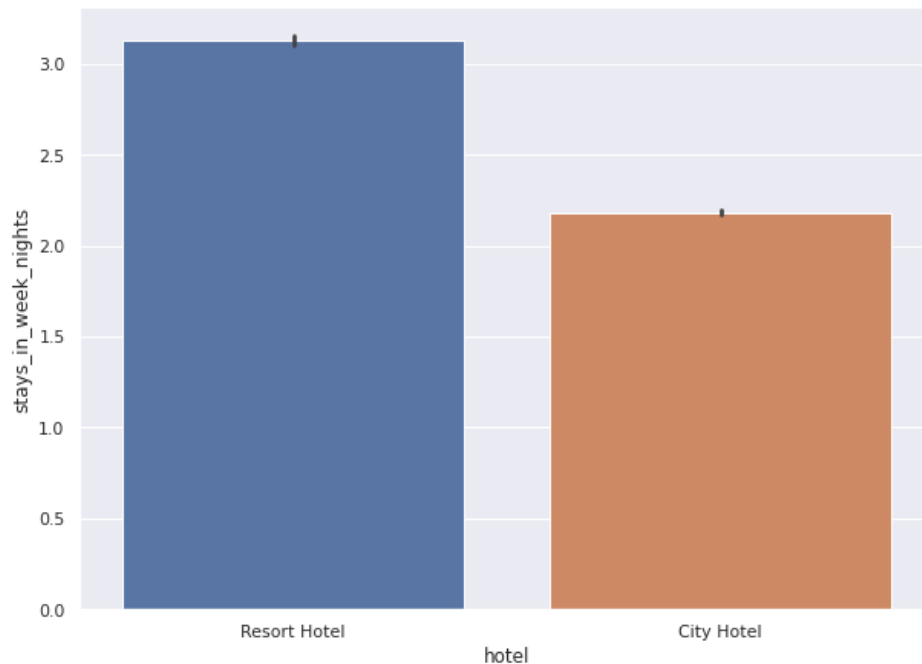
How customers booked their hotel from market segment

- 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.



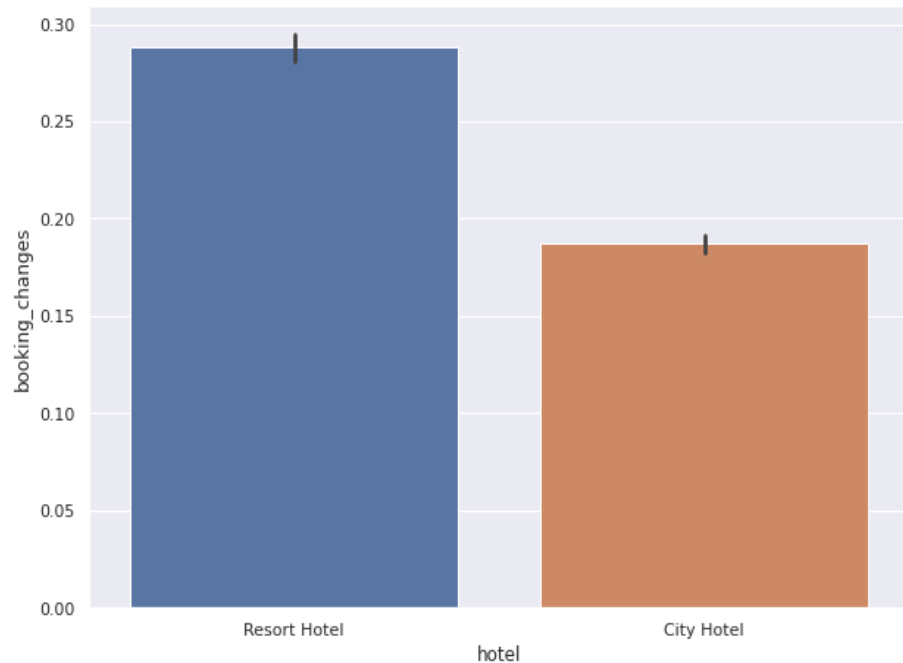
How many customers were stays in week nights

- 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.



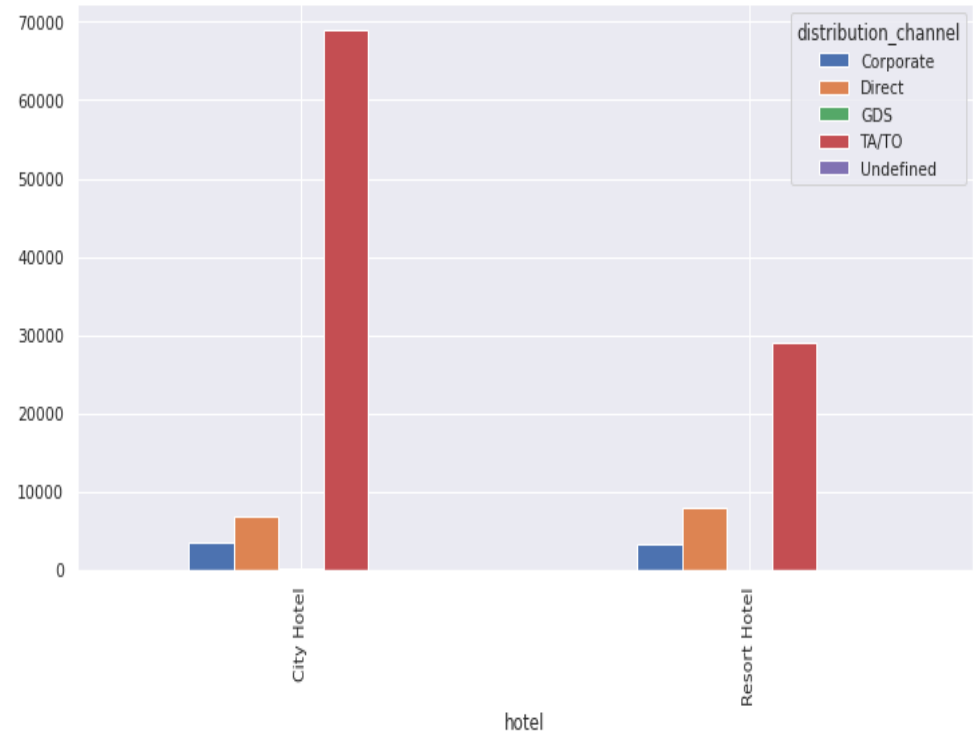
How many customers changed the booking

- 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.



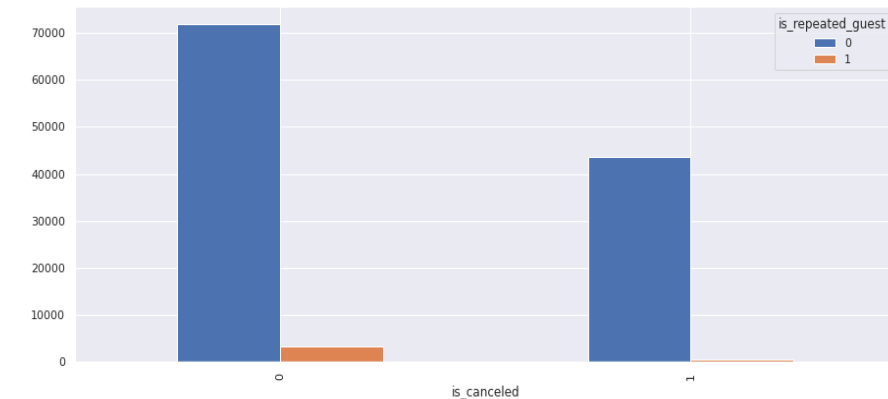
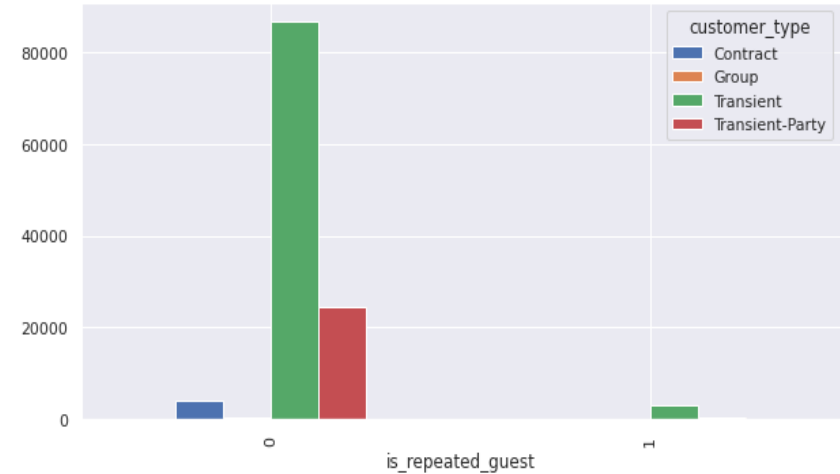
Distribution channel of customers arrival

- 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.



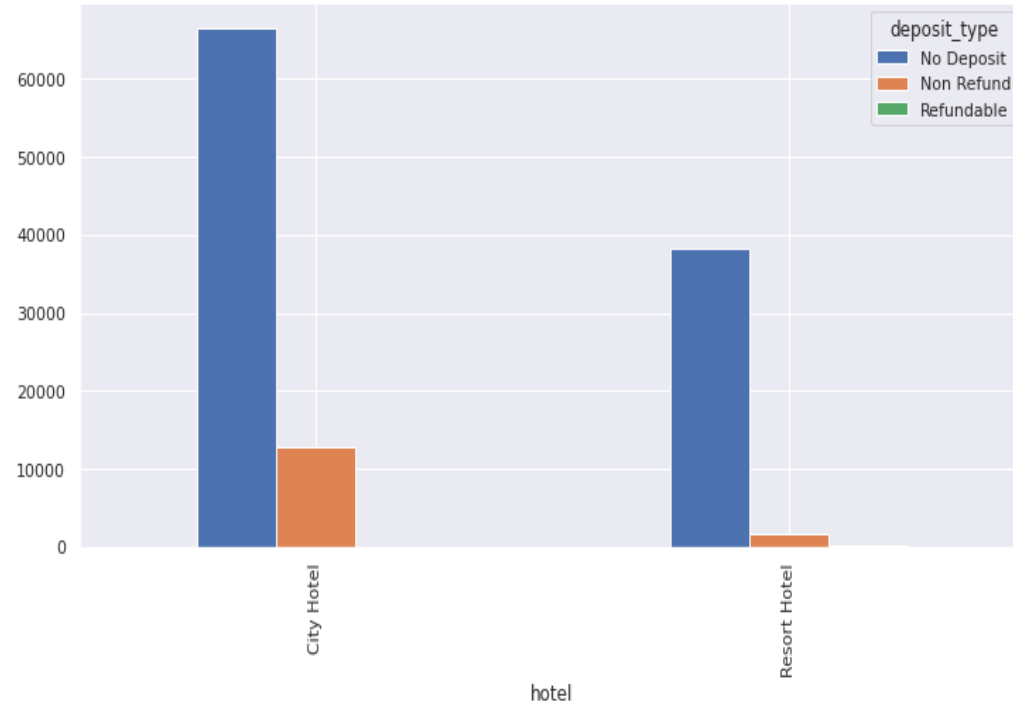
How many repeated customers

- 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 null.
- 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.



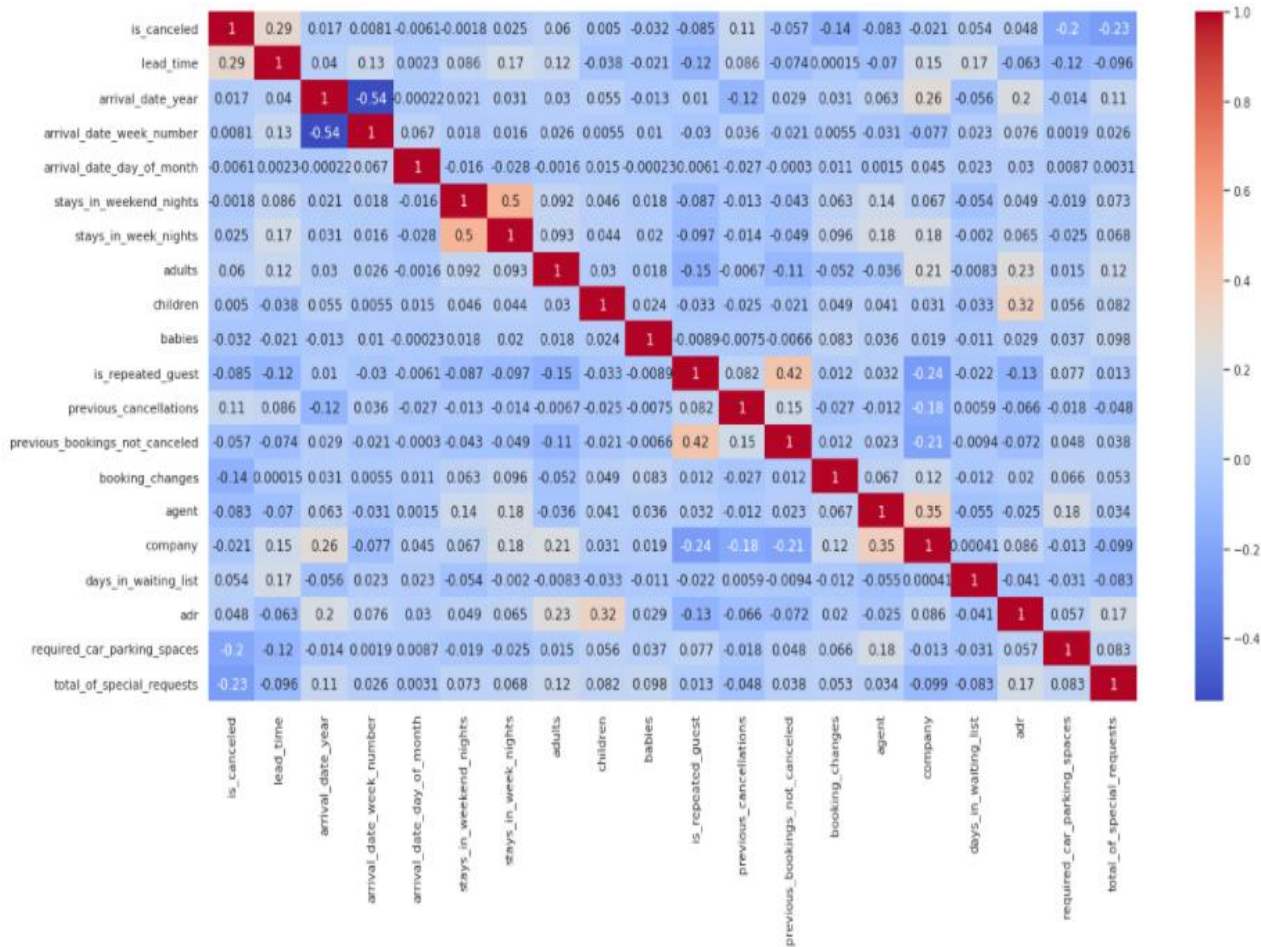
Deposit type payment of customer

- 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 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 decreased 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, cancellation of booking was more that, As repeated guest is less, so we should target our advertisement on guests to increase returning guests.

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.

Key points

- Analyzing customer data allows hoteliers to understand the preferences and purposes of a customer.
- Customers could then be categorized into market segments.
- This segmentation helps target the right audience for marketing and sales.
- Data analytics in hotels enables you to share relevant promotional information with the right market segment.
- Hotels receive bookings from various channels such as online travel agencies (OTA), direct bookings, and website bookings.
- Data analytics in hotels looks at the different mediums to determine the volume of booking received from each channel.
- Hoteliers can use this information to understand which channel yields the most bookings and which channel needs to be worked on.
- This data is important to formulate the right marketing strategy that gives your hotel bookings the right push on all channels.

References

- 1) <https://www.almabetter.com/>
- 2) <https://www.wikipedia.org>
- 3) <https://www.kaggle.com/>
- 4) <https://github.com/>

• Thank You