

Capstone Project Hotel Booking Analysis

Team Members
Anamika
Ayush Goyal

Durgesh Mishra



Content

We can divide the Analysis into 2 sections--

- We deal with Data Collection, Data Exploration, Data Cleaning part.
- We extract information from our data and try to answer some really important questions that govern the bookings.

Data collection--

Al

- Importing the required libraries
- Load the file

Data Exploration--

- Shape & Size
- Descriptive statistics

Data Cleaning--

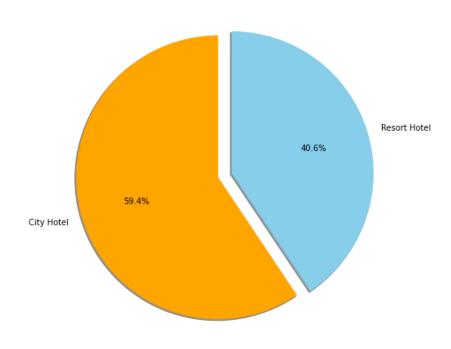
- Duplicate Data & Missing Data
- Feature selection
- Converting datatype & Replacing values
- Datetime operations
- Dealing with Outliers



Types of hotel booking & Booking ratio?

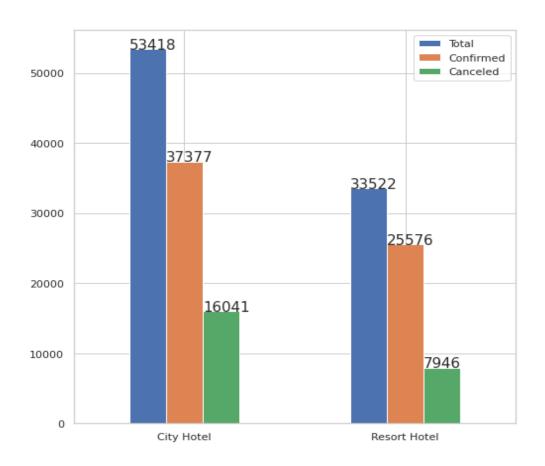






Distribution of Bookings





Bi-Variate Analysis



- 0.8

- 0.6

- 0.4

- 0.2

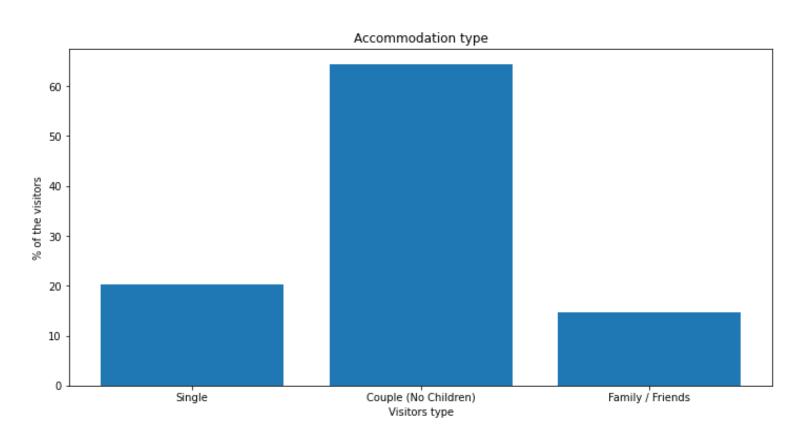
- 0.0

- -0.2

lead_time	1.0	0.1	0.0	0.2	0.3	0.1	0.0	-0.0	0.0	-0.1	0.1	0.1	0.0	-0.1	0.0	0.3	-0.0	-0.1
arrival_date_year	0.1	1.0	-0.0	0.0	0.0	0.0	0.0	-0.0	-0.1	0.0	0.0	-0.0	0.2	-0.0	0.1	0.0	0.0	0.1
arrival_date_day_of_month	0.0	-0.0	1.0	-0.0	-0.0	-0.0	0.0	-0.0	-0.0	0.0	0.0	0.0	0.0	0.0	-0.0	-0.0	-0.0	0.0
stays_in_weekend_nights	0.2	0.0	-0.0	1.0	0.6	0.1	0.0	0.0	-0.0	-0.1	0.0	-0.0	0.0	-0.0	0.0	0.8	-0.0	-0.0
stays_in_week_nights	0.3	0.0	-0.0	0.6	1.0	0.1	0.0	0.0	-0.0	-0.1	0.1	0.0	0.1	-0.0	0.0	0.9	-0.0	-0.0
adults	0.1	0.0	-0.0	0.1	0.1	1.0	0.0	0.0	-0.0	-0.1	-0.0	-0.0	0.2	0.0	0.1	0.1	0.0	-0.3
children	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	-0.0	-0.0	0.0	-0.0	0.3	0.0	0.0	0.0	-0.0	-0.1
babies	-0.0	-0.0	-0.0	0.0	0.0	0.0	0.0	1.0	-0.0	-0.0	0.1	-0.0	0.0	0.0	0.1	0.0	-0.0	-0.0
previous_cancellations	0.0	-0.1	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	1.0		-0.0	0.0	-0.1	-0.0	0.0	-0.0	-0.0	-0.0
previous_bookings_not_canceled	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.0	-0.0		1.0	0.0	-0.0	-0.1	0.0	0.0	-0.1	-0.0	0.0
booking_changes	0.1	0.0	0.0	0.0	0.1	-0.0	0.0	0.1	-0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
days_in_waiting_list	0.1	-0.0	0.0	-0.0	0.0	-0.0	-0.0	-0.0	0.0	-0.0	0.0	1.0	-0.0	-0.0	-0.0	-0.0	0.0	-0.0
adr	0.0	0.2	0.0	0.0	0.1	0.2	0.3	0.0	-0.1	-0.1	0.0	-0.0	1.0	0.0	0.1	0.1	0.0	0.7
required_car_parking_spaces	-0.1	-0.0	0.0	-0.0	-0.0	0.0	0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	1.0	0.0	-0.0	-0.0	-0.0
total_of_special_requests	0.0	0.1	-0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.0	0.1	0.0	1.0	0.0	-0.0	0.0
total_nights	0.3	0.0	-0.0	0.8	0.9	0.1	0.0	0.0	-0.0	-0.1	0.1	-0.0	0.1	-0.0	0.0	1.0	-0.0	-0.0
new_days_in_waiting_list	-0.0	0.0	-0.0	-0.0	-0.0	0.0	-0.0	-0.0	-0.0	-0.0	0.0	0.0	0.0	-0.0	-0.0	-0.0	1.0	0.0
adr_pp	-0.1	0.1	0.0	-0.0	-0.0	-0.3	-0.1	-0.0	-0.0	0.0	0.0	-0.0	0.7	-0.0	0.0	-0.0	0.0	1.0
	lead_time	arrival_date_year	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	adults	children	babies	previous_cancellations	evious_bookings_not_canceled	booking_changes	days_in_waiting_list	adr	required_car_parking_spaces	total_of_special_requests	total_nights	new_days_in_waiting_list	ad_pp

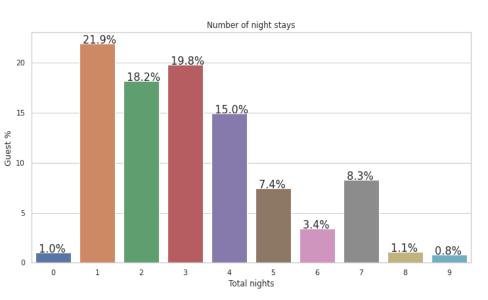


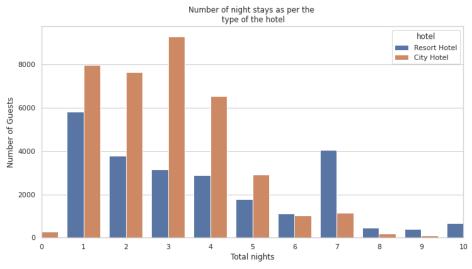
Which was the most booked accommodation type?





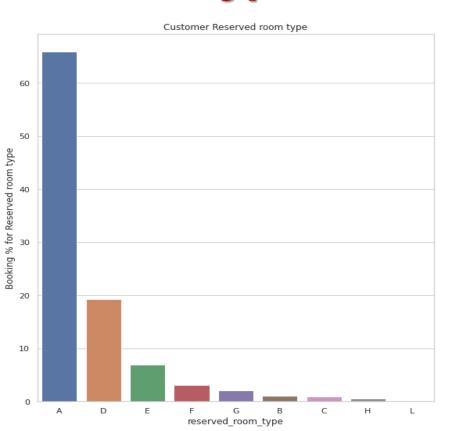
How many nights does visitors prefer to stay?

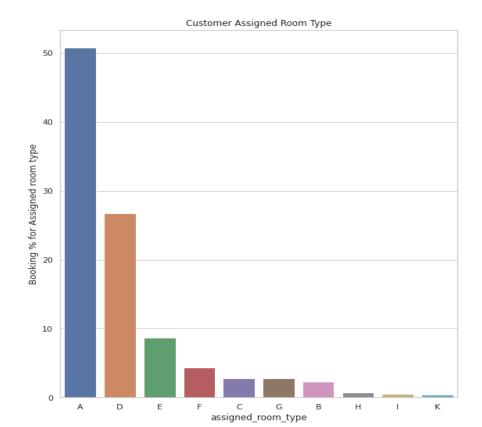






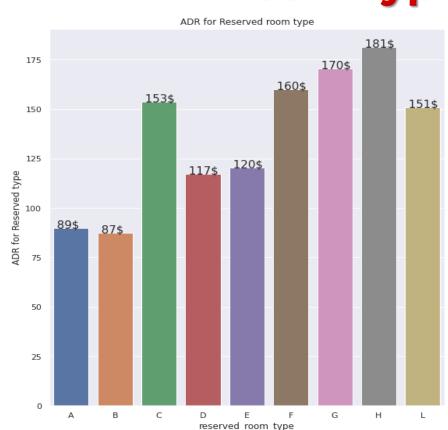
Reserved Room & Assigned Room type Bookings

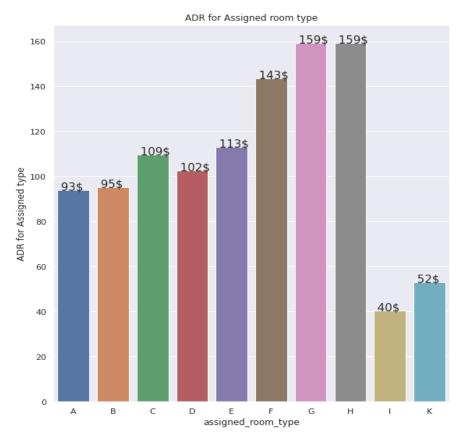




A

ADR for Reserved & Assigned Room type





From which Country most Visitors come from?



-16000

-12000

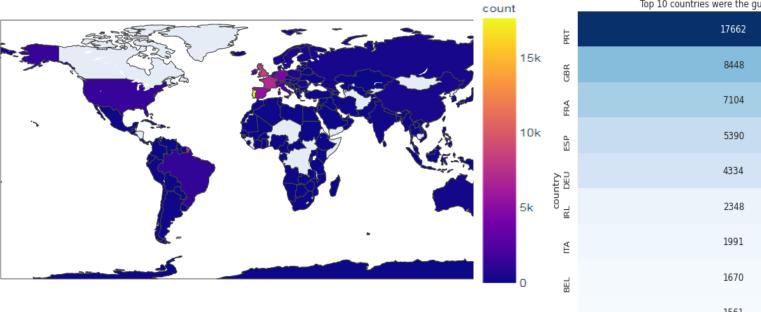
-10000

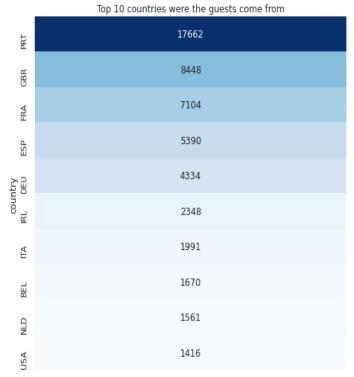
- 8000

- 6000

- 4000

- 2000

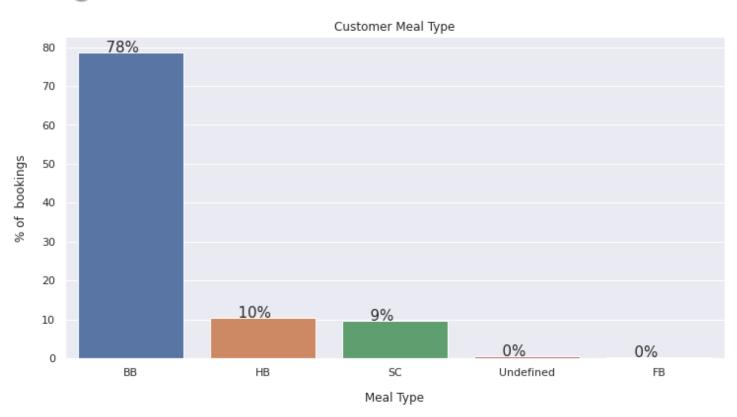




count

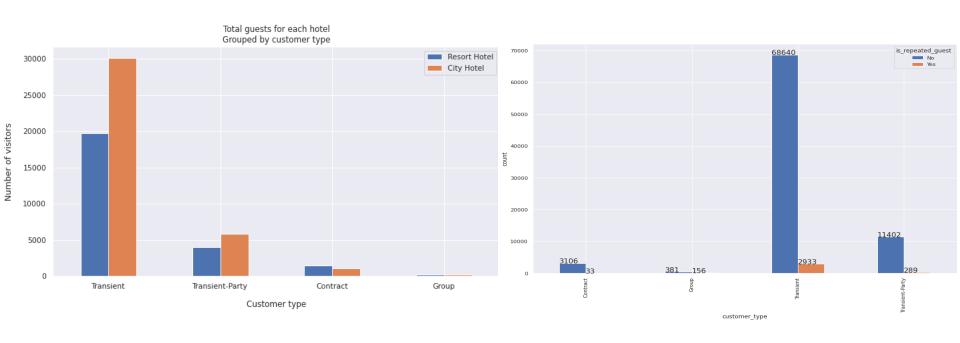
Al

Types of meal booked by the customers?



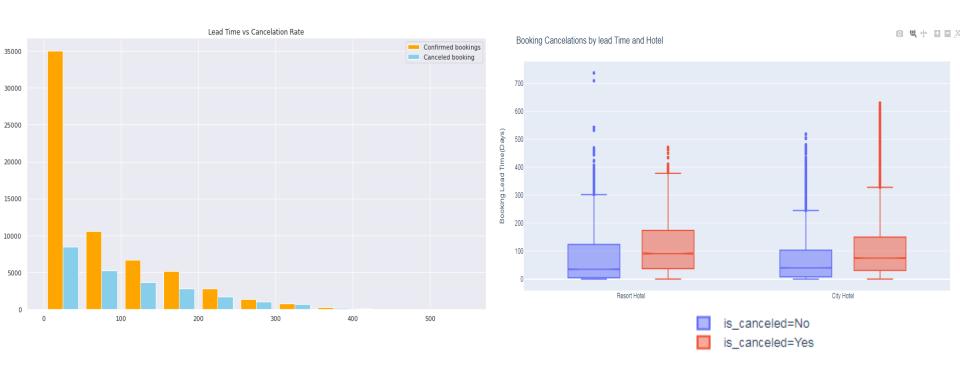
Customer Type & Repeated Guest





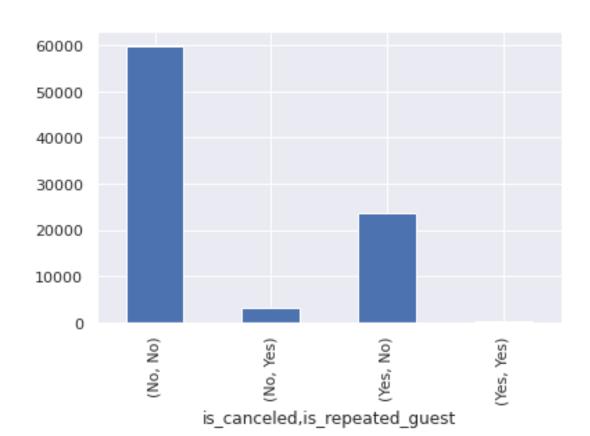
Lead Time & Cancelation Rate?





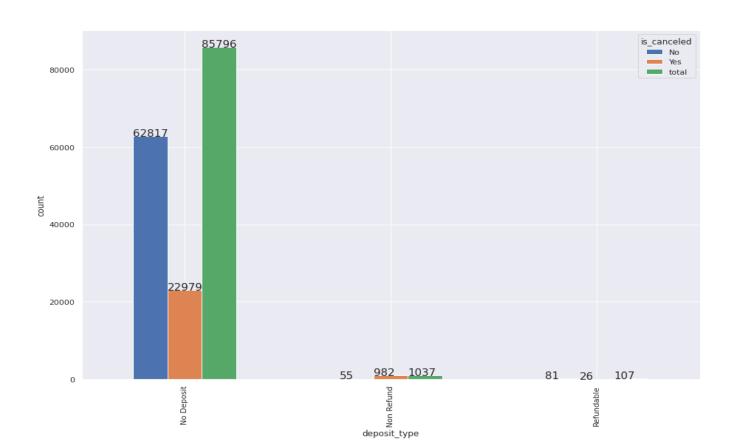


Repeated Customer & Cancelation



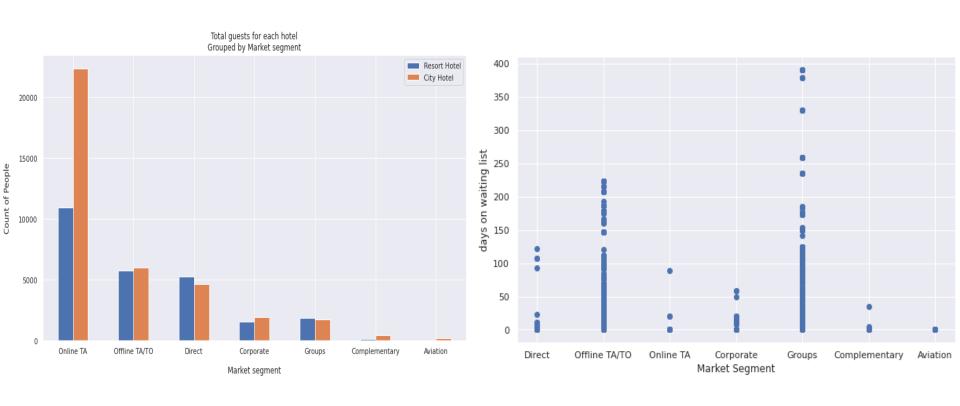


Deposit & Cancellation



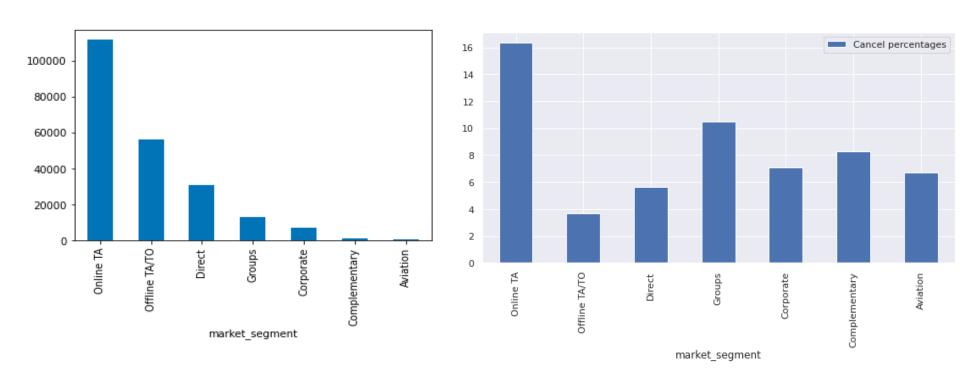


Market Segments & Waiting list days



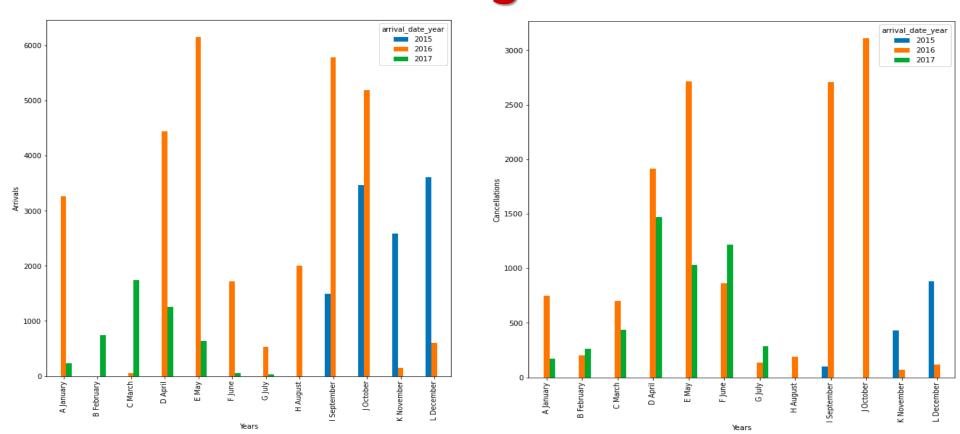


Accommodated Nights & Cancellation % for Market segment



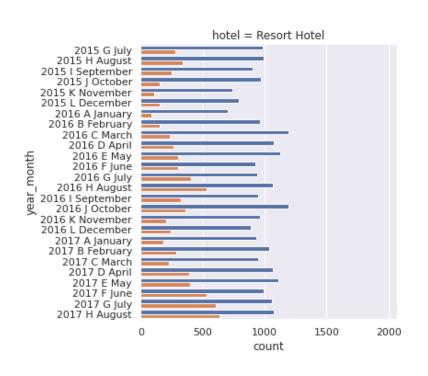
Waiting days of Cancelled & Confirmed bookings

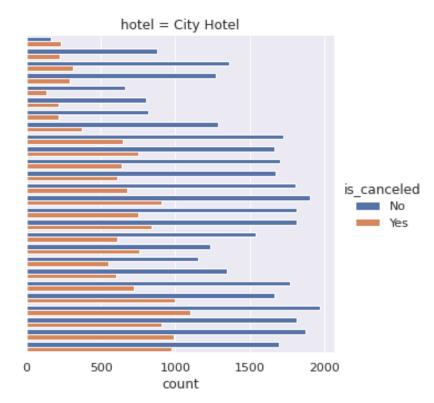






Cancellation & Arrival status







Conclusion:

- Easy and hassle free bookings are mostly preferred like through online agents or offline agents
 therefore hotels can also introduce their own travel agent facility within their premises, to attract more
 bookings.
- The cancellation rate of groups arriving is a great concern as we have second highest cancellation rate despite of a very low arrival rate, therefore we can introduce non refundable bookings for groups in exchange of discounted prices.
- Direct bookings should be promoted as well through various perks and discounts.
- In resort hotels we are having a higher cancellation rate in 2nd and third quarter which needs to be focused upon.
- First and third quarter needs to have some attention due to having higher cancellation rate and, apart from this 4th quarter needs even more attention due to its constant cancellation and very few arrivals which is a major concern.



- Around 60% of booking is done in City Hotels as compared to Resort Hotels. Resort hotels could reduce prices to increases booking percentages & also promote the hotels more.
- Couple (or 2 adults) is the most popular accommodation type. So hotels can make arrangement plans accordingly.
- For most cases there are no deposits for both City Hotels & Resort Hotels.
- We can see that there is some seasonality during the years, which is normal in hospitality, which gets more bookings in the Spring. This could be because of weather conditions as people prefer going on vacation during more comfortable seasons such as Spring/Summer and not during Winters.



Challenges

- Non uniform data due to lack of enough information of all months of different years.
- Different variable names by each member due to lack of coordination
- There are many irrelevant features, duplicate & missing data.
 Moreover, null values and duplicate values identification was a great challenge due to lack of proper unique Identification number.
- The Dataset contains many outliers such as 55 adults in a single room, there is a booking with a lead time of 737 days, \$5,400 for ADR, and so on.



