


CAPSTONE PROJECT HOTEL BOOKING ANALYSIS

- SHUBHAM

TABLE OF CONTENTS:

- Problem Statement
 - Process Flow
 - Data Summary
 - Data Cleaning
 - Data Visualization Objectives
 - Hotel-wise yearly bookings
 - Finding how many Bookings were canceled or not canceled?
 - Visualizing the number of hotel bookings that were canceled and not canceled Yearly.
 - From which Country the most guests are coming?
 - Market Segment wise hotel bookings
 - Finding the relationship between cancellation and market segmentation
 - Finding cancellations with respect to customer types
 - Busy months for Hotels
 - The most booked accommodation types as Family, Couple, Single.
 - Deposit Type with respect to cancellation.
 - Reservation Status of visitors.
 - Finding visitors that stay on weekends and weekdays.
 - Visualizing type of visitors with respect to hotel type.
 - Finding meals for visitors.
 - Finding Repeat guests in hotels.
 - Conclusion
 - Challenges
- 

Problem Statement:

The data set includes booking information for a city hotel and a resort hotel, and also 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, etc.

We have explored and analyzed the data to discover important factors that govern the bookings in the hotel, which could help in finding the loopholes to increase the overall productivity of the hospitality industry. This would also help hotels to provide the best offers and facilities to their customers.




EDA Process Flow:



Data Summary:

df_hotel:

- It contains all the information related to City and Resort hotel bookings.
 - It contains following features that are used for exploring and analyzing the data set.
- **arrival_date_year:** Year of arrival date
 - **customer_type:** Type of booking such as Contract (booking has an allotment), Group (booking is associated with a group booking), Transient (booking is not part of a group or contract), Transient-party (booking is transient but is associated with at least other transient bookings).
 - **is_canceled:** Value indicates if the booking was canceled (1) or not (0).
 - **market_segment:** Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators.
 - **meal:** Type of meal booked such as Bed & Breakfast (BB), Half board (breakfast and one other meal – usual dinner) (HB), Full board (breakfast, lunch, and dinner) (FB).
 - **stays_in_weekend_nights:** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
 - **stays_in_week_nights:** Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
 - **adults/children/babies:** Number of adults, children, and babies visiting hotels.
 - **is_repeated_guest:** Value indicates if the booking name was from a repeated guest (1) or not (0).
- 

Data Cleaning:

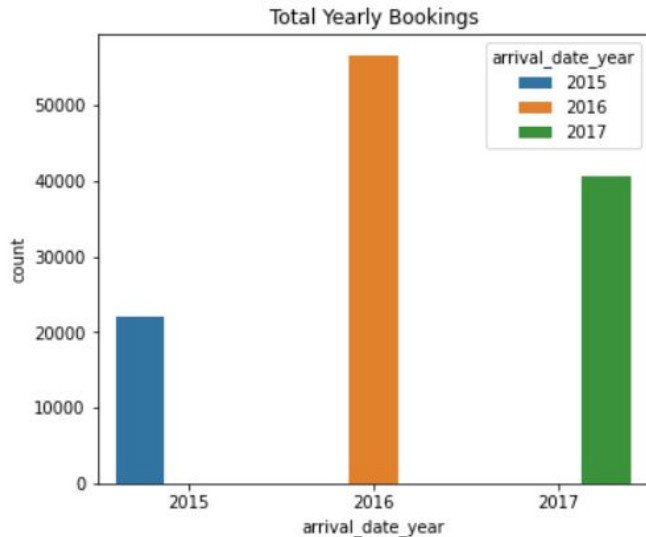
In this step we drop the unnecessary columns to reduce noise in our data set.

- Firstly, we have created a copy of our data set, so that the original data set does not get affected in further steps.
- Afterward, we found the columns with null values and replaced them with 0.
- The missing values in the country column are replaced with mode value that appears mostly and the missing children values are replaced with the rounded mean value for proper evaluation of the data set.
- Lastly, we drop the rows where there is no baby, children, and adult, and convert the datatype of children, company, and agent columns from float to integer for noise-free calculation of the data set.



Hotel-wise yearly bookings

- The highest number of hotel bookings took place in 2016.
- Every year, the number of **City Hotel Booking** is more than **Resort Hotel Booking**.
- More than double bookings were made in 2016 as compared to 2015. Although, the bookings were decreased by almost 15% in 2017.



Finding how many Bookings were canceled or not

- The countplot shows the total number of canceled vs not canceled bookings.
- It shows that the majority of bookings were not canceled which shows that most visitors came to hotel after doing booking.
- Appx. 62% bookings are not canceled.

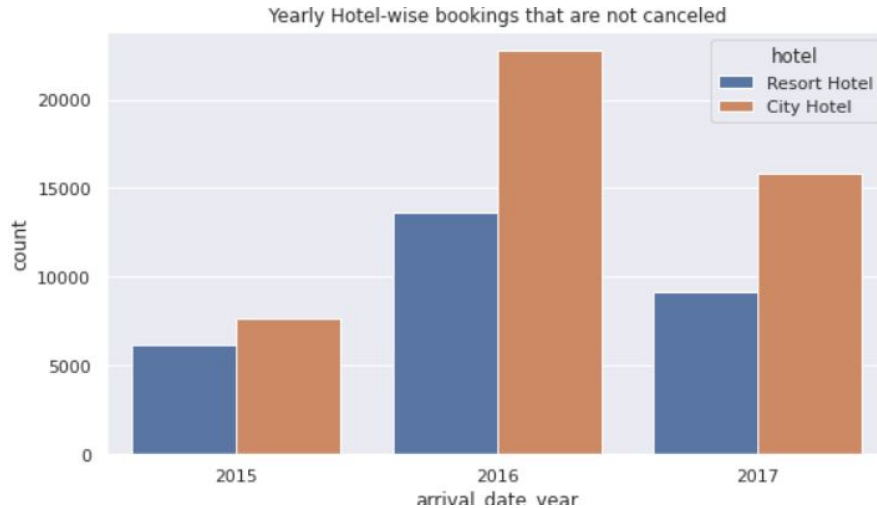


Yearly visualizing the number of hotel bookings that were canceled and not canceled

- The majority of bookings took place in City Hotel.
- The plot shows around 30% bookings were cancelled in resort hotel and 40% city hotel.
- There is a high rise in cancellation of booking in City hotel as compared to Resort hotel.

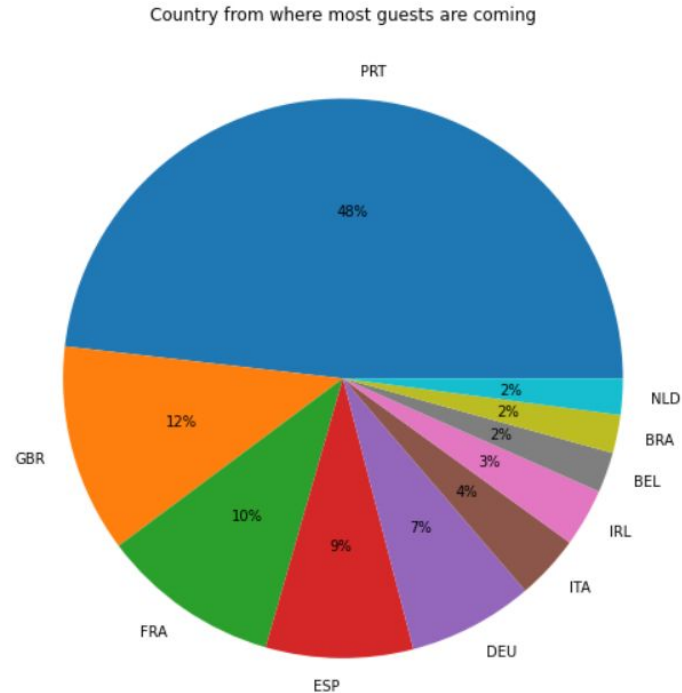


- In every year number of canceled and not canceled bookings are more in City hotel bookings than compared to Resort hotel.
- Further, it should be noted that in 2016 maximum number of cancelation took place and the ratio of bookings that not canceled between City and Resort hotel is least in 2015.



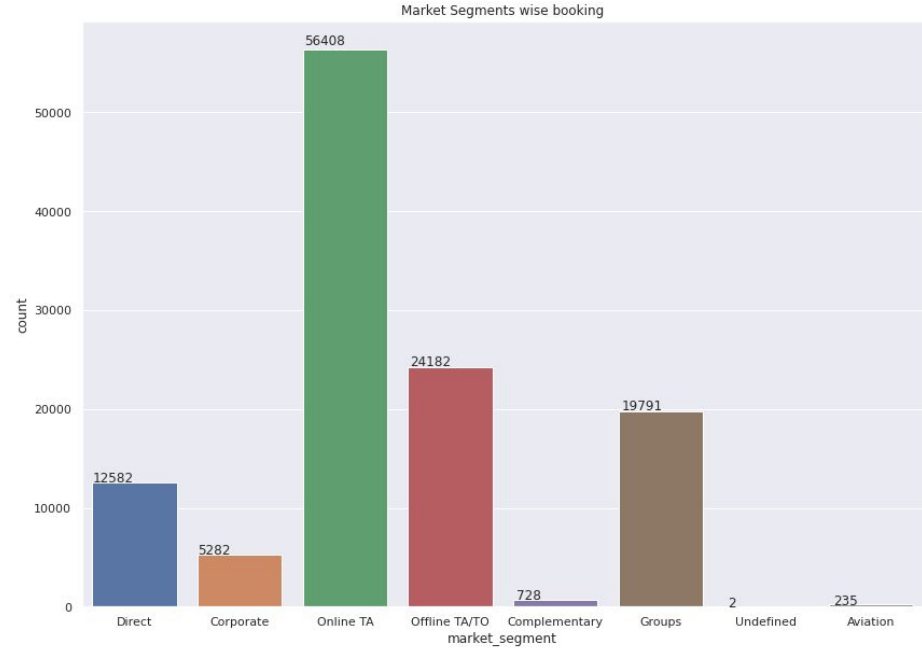
From which Country the most guests are coming

- The highest numbers of visitors are from countries like Portugal, United Kingdom, France, Spain and Germany.
- More than 80% visitors come from these five countries.



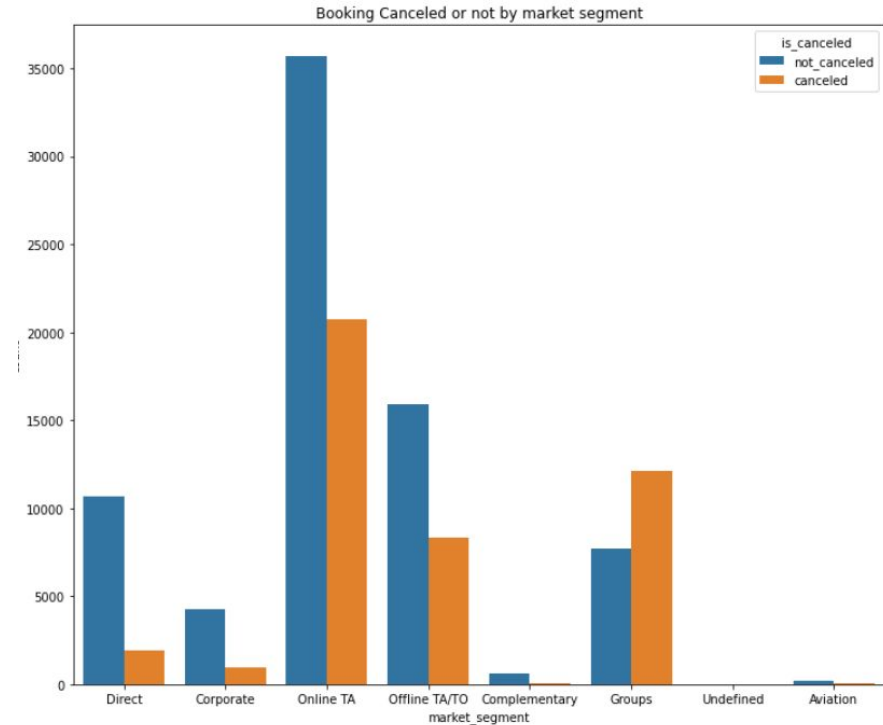
Market Segment-wise hotel bookings

- Bookings through online and offline travel agents are higher as compared to group bookings and direct bookings.
- The majority of people go for travel agents for hotel bookings.



Finding relationship between cancelation and market segmentation

- Plot shows that Groups segment has cancellation rate of around 60%.
- Offline TA/TO (Travel Agents/Tour Operators) and Online TA has cancellation rate of more than 33%.
- It is surprising that the cancellation rate in these segments is high despite of a deposit.
- It is fact that cancellations that are made collectively like group reservations has high cancellation rate as different visitors may not free at same to visit hotel.
- It is interesting to note that direct segment has cancellation rate of less than 20%.



Finding cancellations with respect to customer types

- The plot shows that majority of bookings are transient.
- Transient booking is a booking that is not a part of a contract or group.
- This shows that Booking online is becoming increasingly consumer friendly.

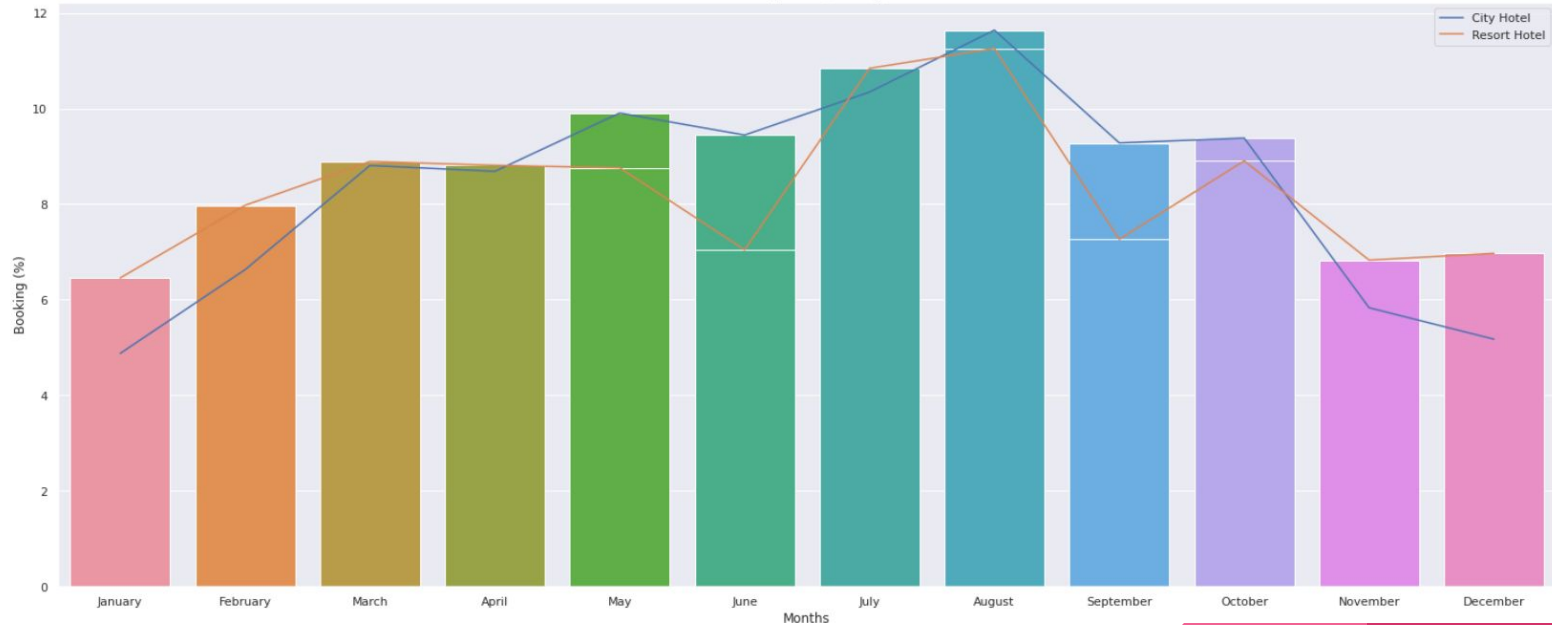
Conclusion:

- Booking online is becoming first choice of visitors so, hotels can advertise and provide offers through this channel to increase their sales.

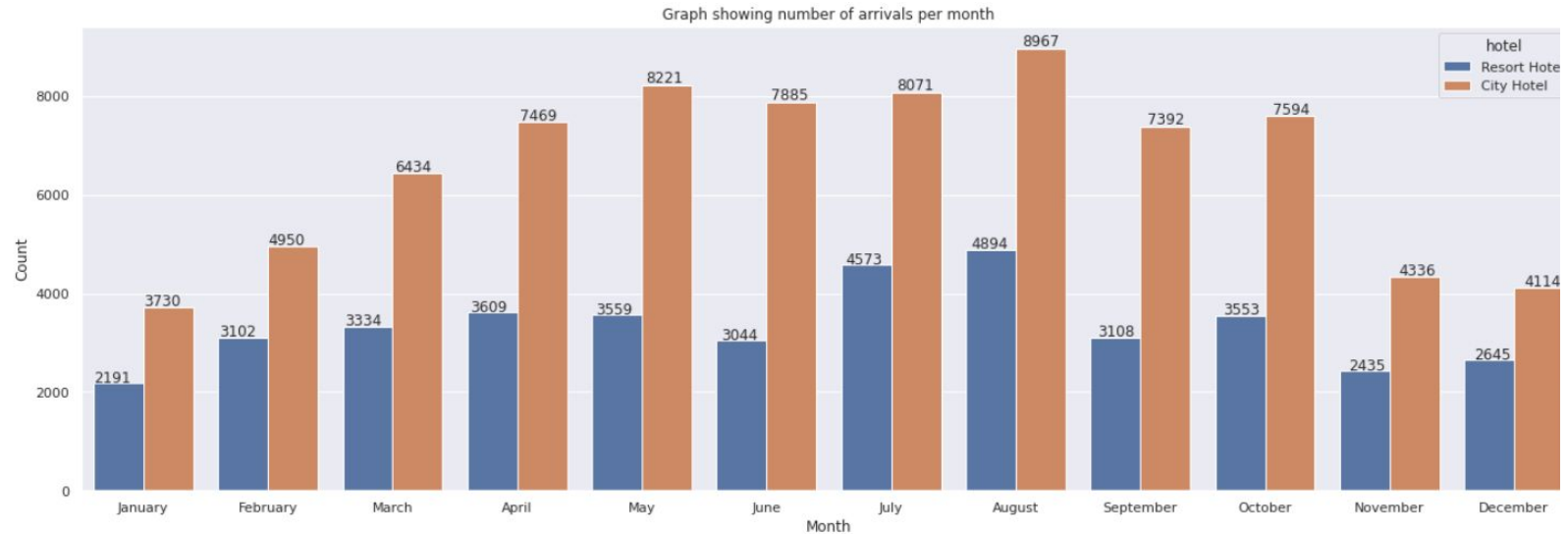


Busy months for Hotels

The plot shows most of the bookings were made from July to August, and the least bookings were placed at the start and end of the year.



We can see that bookings for both City and Resort Hotel are high in July and August, and lower in start and end of the year. Further, period from May to August is busiest for both the hotels.

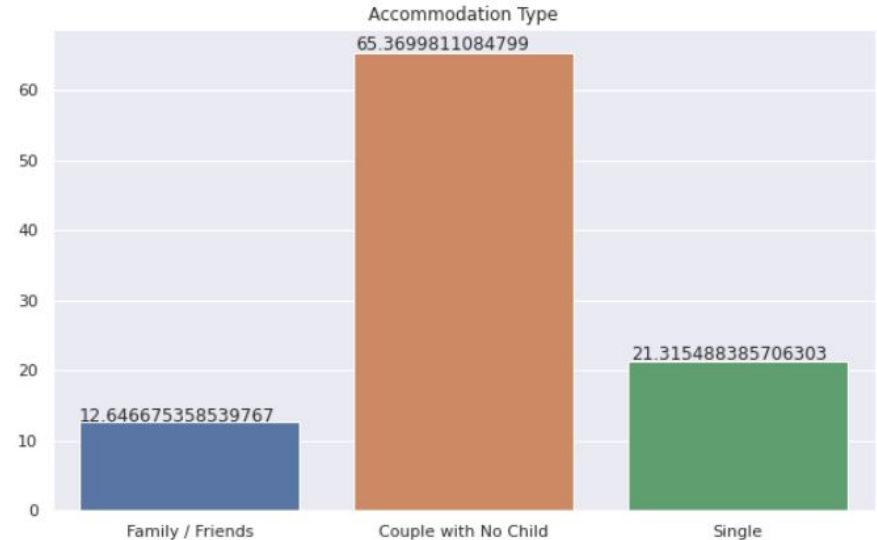


The most booked accommodation type as Family, Couple, Single

- Couple with no children is the most popular accommodation type.

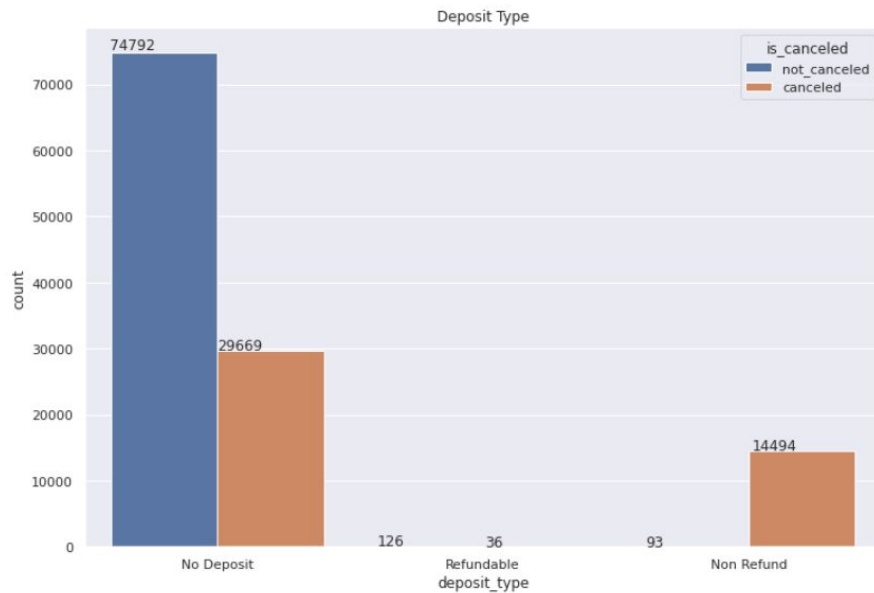
Conclusion:

- Hotels can provide special plans and services for couples and singles.



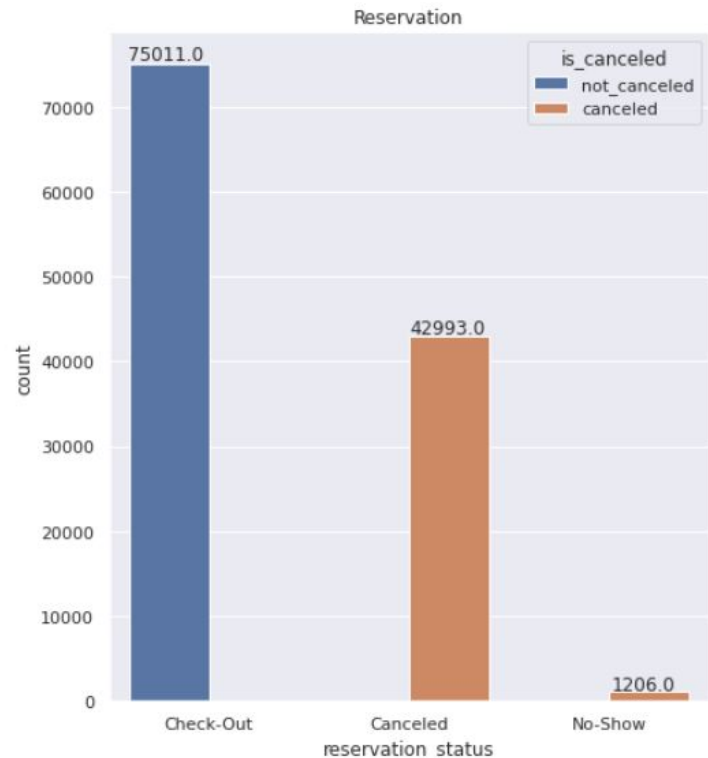
Deposit Type with respect to booking cancelation

- The majority of bookings that are canceled are from no deposit type that does not require any deposit, due to this the high cancelation rate is observed.
- Also it is interesting to note that refundable deposits had less cancellation than non-refundable deposits. Logically one would have assumed that non-refundable deposits have less cancellation as hotel rates are usually higher.



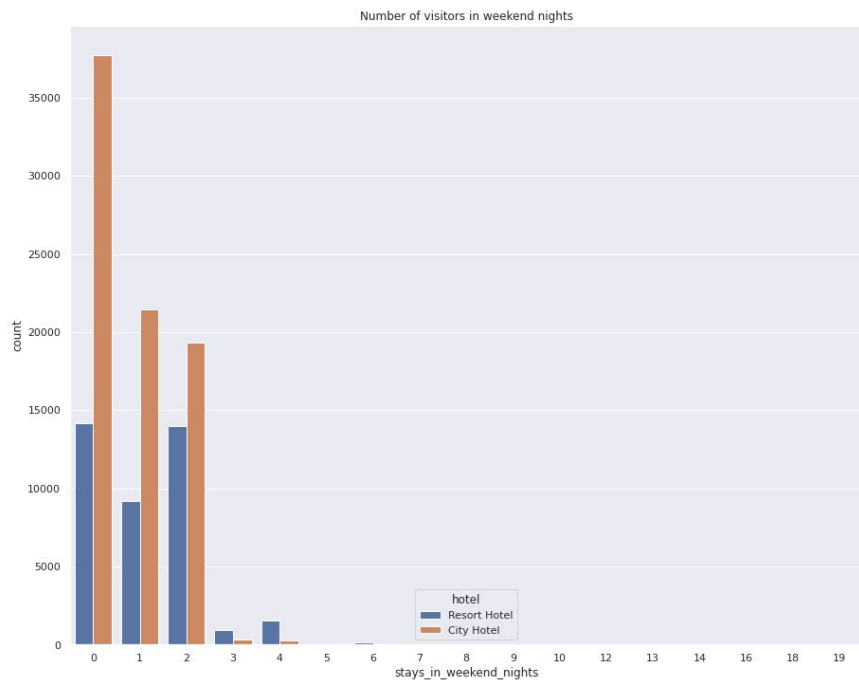
Relationship between Reservation Status and cancelation

- The graph shows most of the people how booked the hotel done checkout.
- This implies most people who do the reservation came to the hotel.

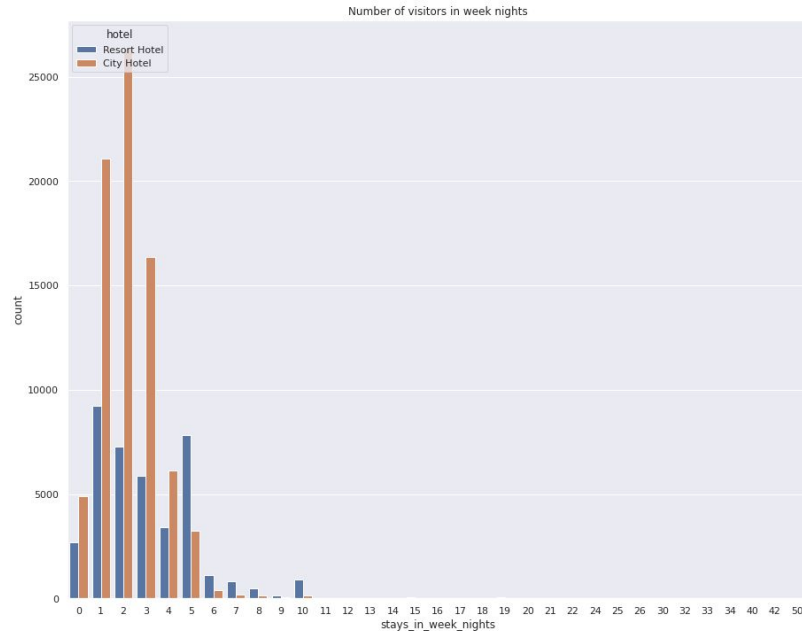


Visitors that stay on weekends and week days

The plots shows that the majority of visitors stay for one or less than one day in weekend nights for both type of hotels. Further, most people stay in the month of July and August.

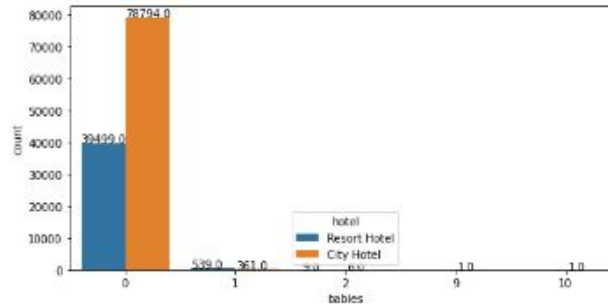
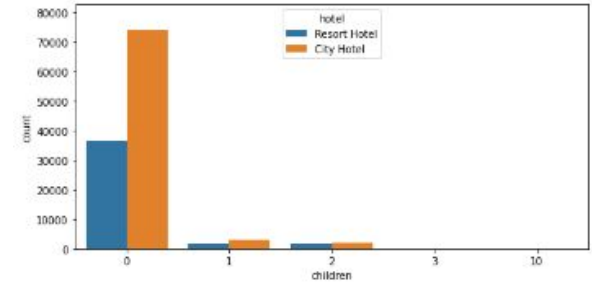
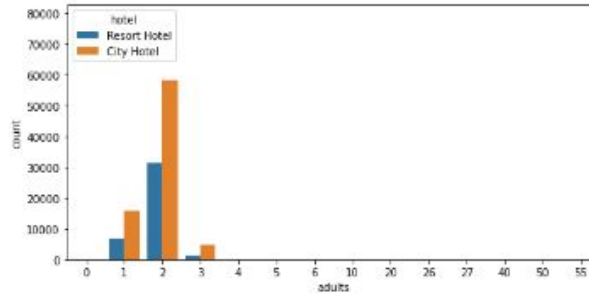


The above plots shows that 75% people stay for three or less than three days in week nights for both type of hotels. It is to be noted that as no of days increases people stay more in resort hotel than city hotel. Further, most people stay in the month of July and August.



Visualizing type of visitors with respect to hotel type

- The plot shows that most of the visitors of both the hotels are adults either couple or single with no children and babies.
- Further, couple with babies prefer Resort Hotels.

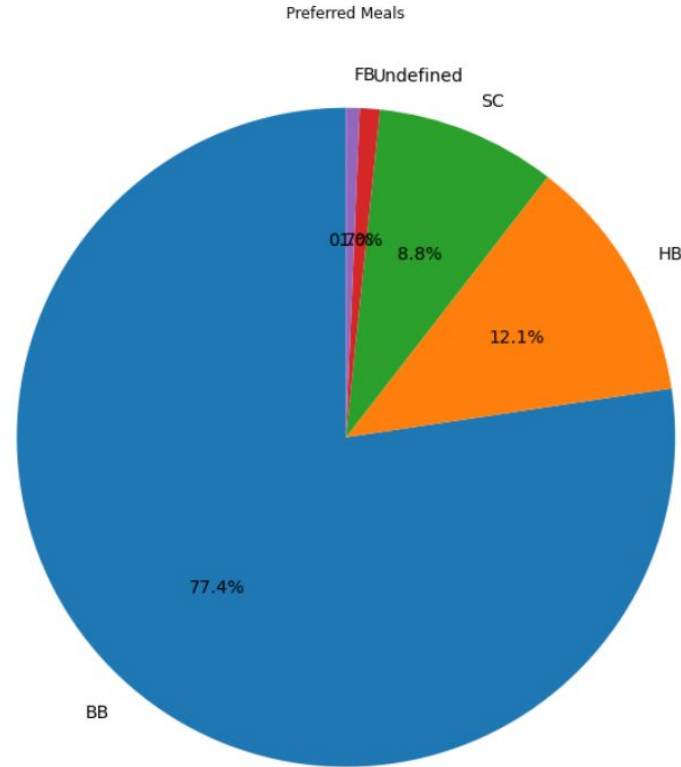


Preferred meals by visitors

Most of the visitors prefer for Bead and Breakfast meal followed by Half board, and then SC (no meal package).

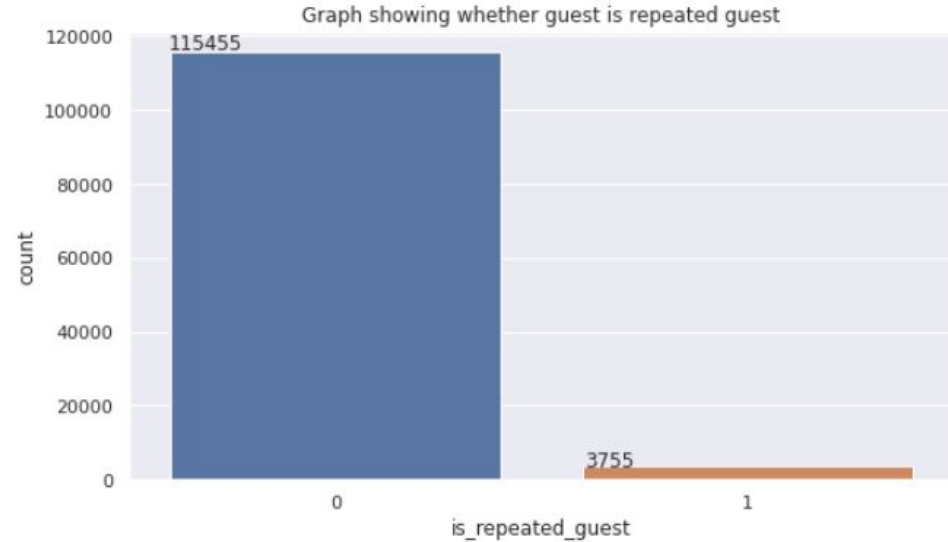
Conclusion:

More than 70% of people prefer BB meals, so menu prices can be increased. As most people prefer to do breakfast in hotels. At the same time, additional offers for lunch and dinners should be added to attract visitors.



Finding Repeated Guest

- The plot shows very low number of repeated guests.
- Further, marketing or business should be drafted to improve repeated guests.



Conclusion

Hotel's Perspective:

- Overall, People prefer City hotels over Resort Hotels. So, the hospitality industry can create new City hotels and can provide offers to visitors in Resort hotels to engage more visitors.
- People for a shorter duration of stay prefer City hotels over Resort hotels and for longer stays go for Resort hotels.
- From May to August the hotels seem busy. During this period hotels can increase their price for increasing their profits. Although, at the end and start of the year very few visitors arrive, in this period hotels can provide promotional offers to attract visitors.
- Couples with no children and babies prefer to stay in both types of hotels. But couples with babies prefer to stay in Resort hotels. So, additional couple services can be provided to attract visitors towards hotels. Also, services like baby care can be provided in Resort hotels.
- Visitors mostly prefer online/offline travel agents and tour operators for their hotel bookings. So, hotel advertising can be processed from this channel to increase the range of hotels.
- More than 70% of people prefer BB meals, so menu prices can be increased. As most people prefer to do breakfast in hotels. At the same time, additional offers for lunch and dinners should be added to attract visitors.
- The majority of bookings that are canceled are from a no deposit type that does not require any amount to be deposited, due to this a high cancellation rate is observed. Also, it is interesting to note that refundable deposits had fewer cancellations than non-refundable deposits. Logically one would have assumed that non-refundable deposits have fewer cancellations as hotel rates are usually higher.
- The majority of bookings are transient (booking that is not a part of a contract or group). This shows that Booking online is becoming increasingly consumer friendly. Hotels can advertise and provide offers through this channel to increase their sales.
- Countries like Germany, Italy, Ireland, Belgium, Brazil, and the Netherlands have a very less number of visitors. Since they are in the top 10 countries from where visitors are coming so these country visitors have the potential to increase further if a proper advertising channel is established to engage them.
- The most painful analysis in this data set is very less number of people are retain by the hotels, this shows people are not satisfied with the hotel services. Proper feedback should be taken at checkouts. Plus proper social connectivity channels should be established with the visitors and hotels should send promotional offers and reminders to the visitors regularly to retain them.

Visitor's Perspective

- People planning big parties in resort and city hotels can plan at the start or end of the year as there is no rush and they can get heavy discounts in this period.
- People can use online/offline travel agents and tour operators for booking their tickets since they seem to be the most trustworthy while booking hotels and people can also get better deals in the future through this channel.
- People can go for BB meals in hotels since 77% of visitors preferred it.



Challenges:

- The major challenge was to identify the columns that are taken into consideration for analyzing the data set and the columns that can be dropped to reduce noise from the data set.
- Framing objectives that provides fruitful conclusion on the data set.
- Computation time.



Thank You !!!!!

