Capstone Project
Hotel Booking Analysis(EDA)

by Sayan Bandopadhyay

Synopsis

Who does not like to travel and spend some quality vacation in a good & snuggly hotel ? Surely, maximum people out there would love to go for a vacation and enjoy their stay in a good hotel. So, basically hotels play an essential part in the overall vacation experience.

With the growth in travel and tourism, 'Hotel Industry' which comes under 'Hospitality Industry' is also scaling up and of course a large number of factors remain responsible for the success of this industry and one of those are the booking pattern of hotels.

So, here I have a dataset regarding the data of hotel bookings and related information. With all this data, I managed to analyze and extract some meaningful insights from it.

So, let's get straight into it.

AI

Data Briefing

The dataset is quite large and upon using the "shape" method I am able to find the initial number of rows and columns present in the dataset.

- Initially the dataset had 119390 rows and 32 columns (119390,32).

Now, regarding the extraction of the data information from the dataset, I used "info()" method which gave me the information about data types(dtypes), count of non-null values & memory usage by the dataset.

- There were 4 columns with float64 dtypes, 12 columns with object dtypes & 16 columns with int64 dtypes.

I used "column" method to get the list of all the column label names and used "isnull()" method to get the count of null values if any in the dataset.
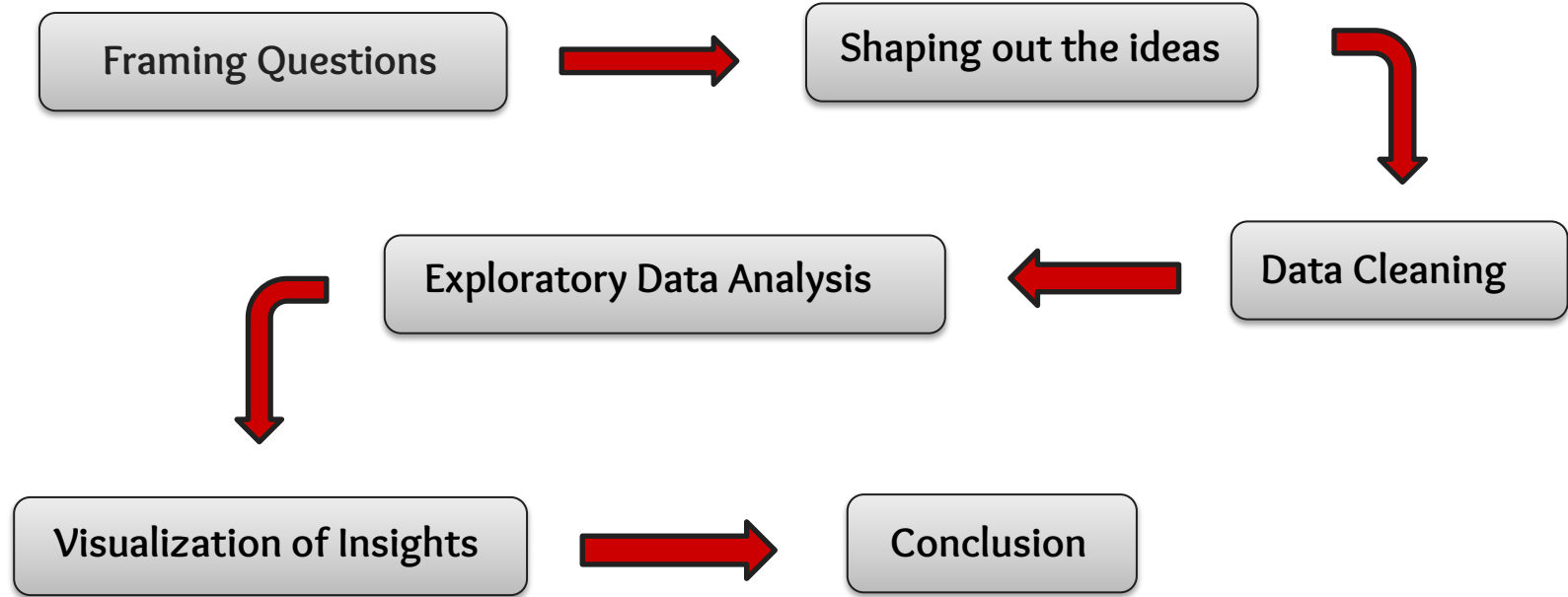
- 4 null values found in 'children' column.
- 488 null values found in 'country' column.
- 16340 null values found in 'agent' column.
- 112593 null values found in 'company' column.

Steps involved in analysis

Framing Questions → Shaping out the ideas → Data Cleaning → Exploratory Data Analysis → Visualization of Insights → Conclusion

ANALYSIS REPORT

Hotel types & preferences

In the dataset, we have two types of hotel that is 'City hotel' & 'Resort hotel'. Upon using the "value_counts()" method, I managed to find out the total counts of each type of hotel.

- Count of 'City hotel' is 79330 which is 66.4% of total count.
- Count of 'Resort hotel' is 40060 which is 33.6% of total count.

Above data can be clearly visualized by this pie chart which shows the percentage share of each hotel.

Here, it seems that out of these two hotels, 'City hotel' is more preferred by the customers as compared to the 'Resort hotel'.

PREFFERENCE OF HOTEL TYPES

City Hotel

66.4%

hotel

33.6%

Resort Hotel
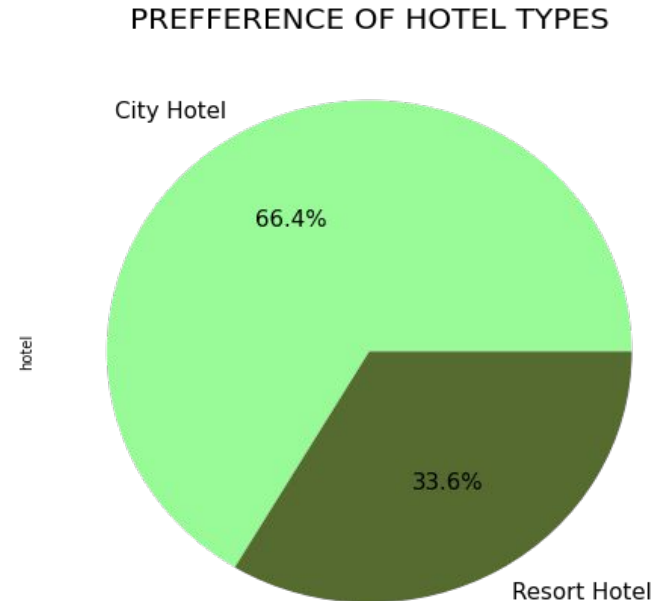
Year-wise bookings

Similarly, I found out the different year data which is present in the dataset. Dataset have booking data of 3 years and those are 2015, 2016 & 2017.
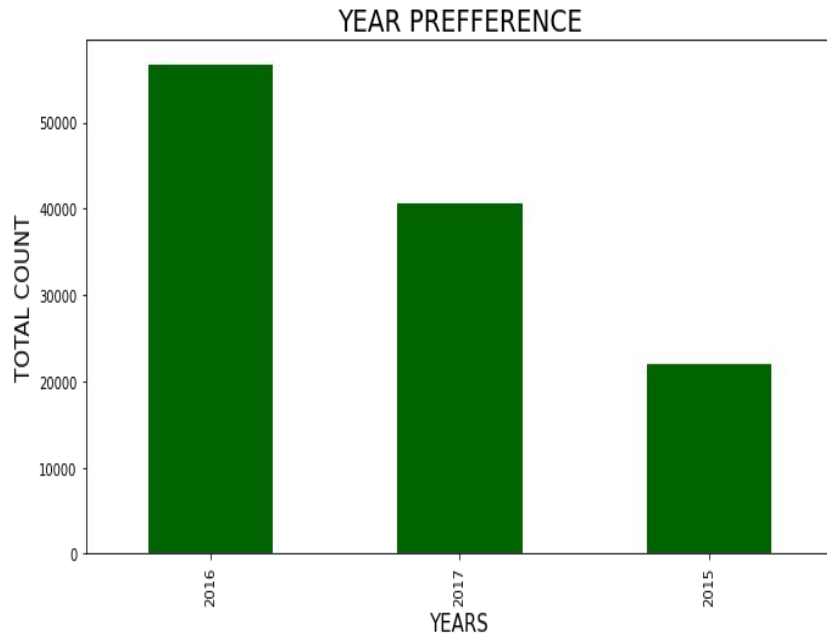
- Total booking that took place in 2015 was 21996.
- Total booking that took place in 2016 was 56707.
- Total booking that took place in 2017 was 40687.

From this bar chart, it can be said that most hotel booking took place in the year 2016 whereas 2015 has witnessed least bookings among the 3 years.

Month-wise bookings

MONTH PREFFERENCE
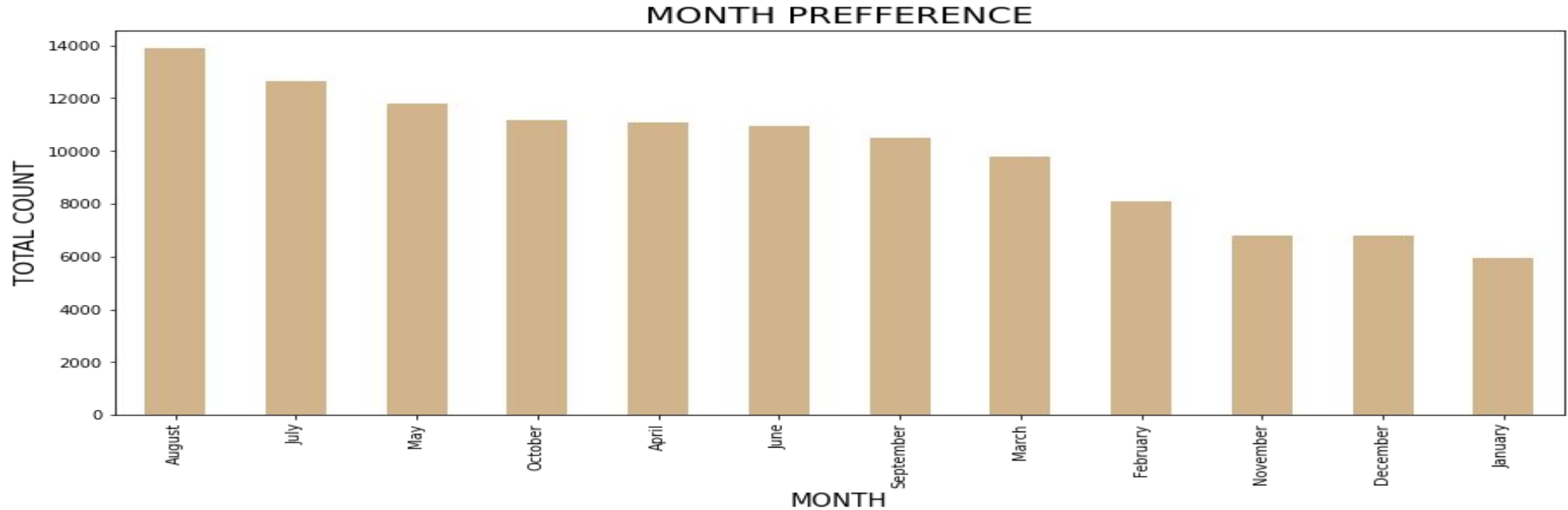
Here, the data of 12 months are given and its prominent enough that 'August' has got the highest number of hotel bookings whereas 'January' has got the least number of hotel bookings.
Therefore, it can be said that customers prefer 'August' very often for tourism purpose.

Country preference

While analyzing on the different types of countries and total count of bookings in each of them, I observed that there were a lot of countries so, I decided to analyze only on the basis of top 10 countries with maximum bookings.
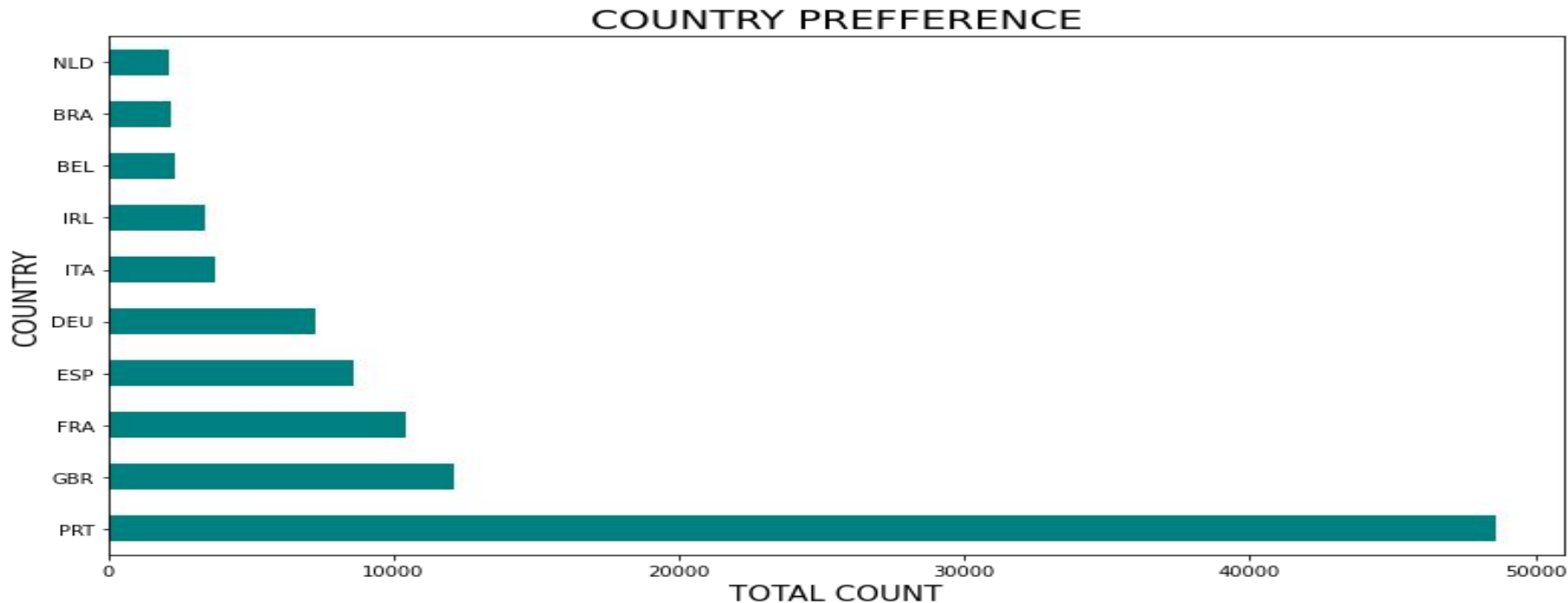
The top 10 countries with maximum bookings are as follows :-
1. PRT (Portugal)
2. GBR (Great Britain)
3. FRA (France)
4. ESP (Spain)
5. DEU (Germany)
6. ITA (Italy)
7. IRL (Ireland)
8. BEL (Belgium)
9. BRA (Brazil)
10. NLD (Netherlands)

Out of these 10 countries, PRT(Portugal) has got the highest number of bookings that is 48590. Because, of its beauty and landscape diversity people loves to travel to Portugal more, hence leading to more and more hotel bookings.
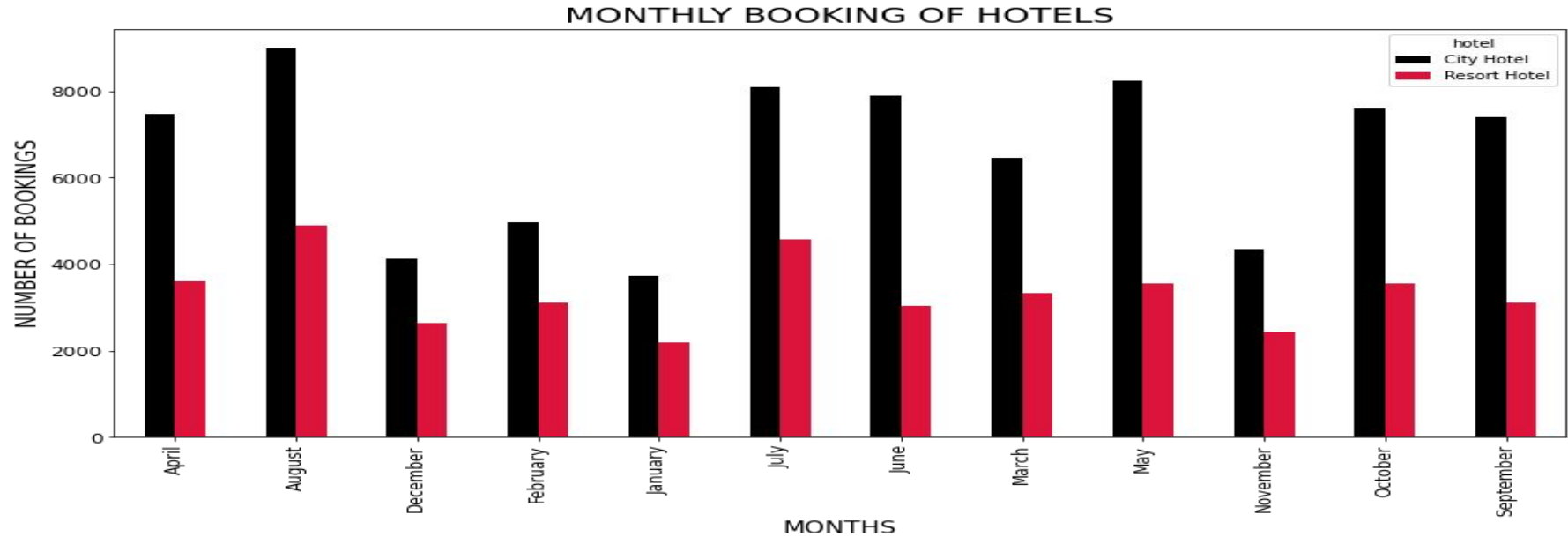
Hotel preference (Monthly basis)

Grouping the hotel types along with monthly booking data by using the "groupby()" method, I got the count of bookings in different months of different types of hotels.

Hotel	City Hotel	Resort Hotel
Months		
April	7480	3609
August	8983	4894
December	4132	2648
February	4965	3103
January	3736	2193
July	8088	4573
June	7894	3045
March	6458	3336
May	8232	3559
November	4357	2437
October	7605	3555
September	7400	3108

These are the counts of total bookings of 'City hotel' & 'Resort hotel' in various months.

Visualization of this data will make it more trouble-free to understand.

MONTHLY BOOKING OF HOTELS

- In each & every month, bookings of 'City hotel' is more than that of bookings of 'Resort hotel'.
- Both 'City hotel' & 'Resort hotel' has got the maximum bookings in August & minimum bookings in January.

Hotel preference (Yearly basis)

Hotel		City Hotel	Resort hotel
Years			
2015		13682	8314
2016		38140	18567
2017		27508	13179

- ‘City hotel’ bookings > ‘Resort hotel bookings in each year.
- Both the hotels have witnessed maximum booking in 2016.
- Fall in the number of bookings can be spotted for both the hotels after 2016.

Just as the previous one this time, I grouped along with yearly booking data.

YEARLY BOOKING OF HOTELS
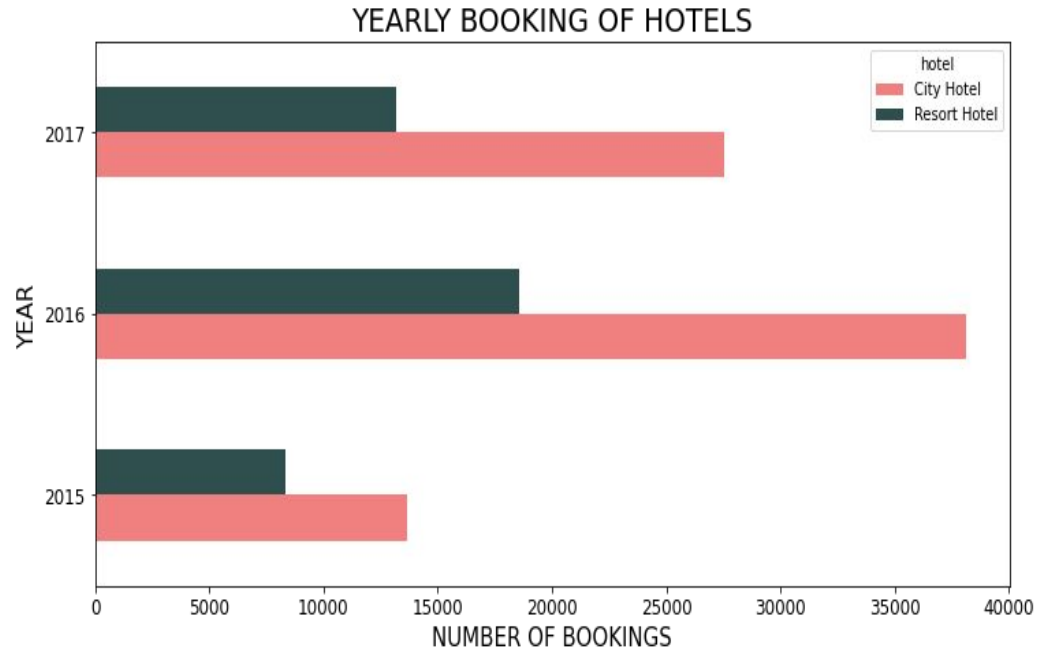
Cancellation of bookings

This particular analysis is based upon the number cancellation of bookings in 'City hotel' & 'Resort hotel'. Upon grouping the cancellation data along with hotel data, I got the total counts of booking being cancelled and total counts of bookings not cancelled.
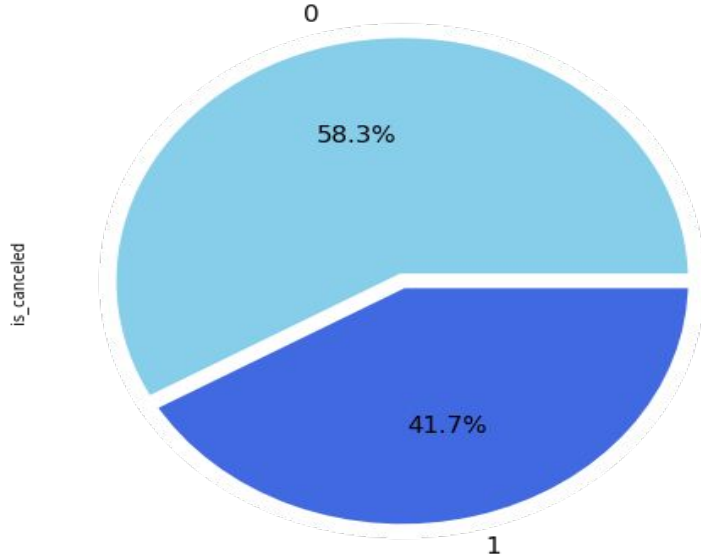

CANCELLATION STATUS
(0 = Not cancelled & 1 = Cancelled)

Here,
0 = Booking is not cancelled.
1 = Booking is cancelled.

Bookings as well as cancellation rate both is higher in 'City hotels' than 'Resort hotels'.
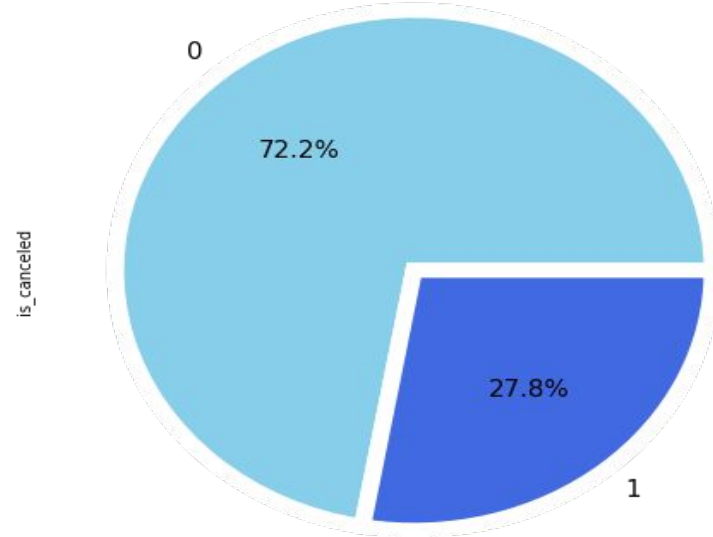
AI

CITY HOTEL CANCELLATION STATUS

(0 = Not cancelled & 1 = Cancelled)

## RESORT HOTEL CANCELLATION STATUS

(0 = Not cancelled & 1 = Cancelled)

- In 'City hotel' 41.7% of total bookings is cancelled whereas 58.3% of total bookings is not cancelled.
- In 'Resort hotel' 27.8% of total bookings is cancelled whereas 72.2% of total bookings is not cancelled.

Analysis on the basis of market segment

Market segments	Counts
Online TA	56477
Offline TA/TO	24219
Groups	19811
Direct	12606
Corporate	5295
Complementary	743
Aviation	237
Undefined	2
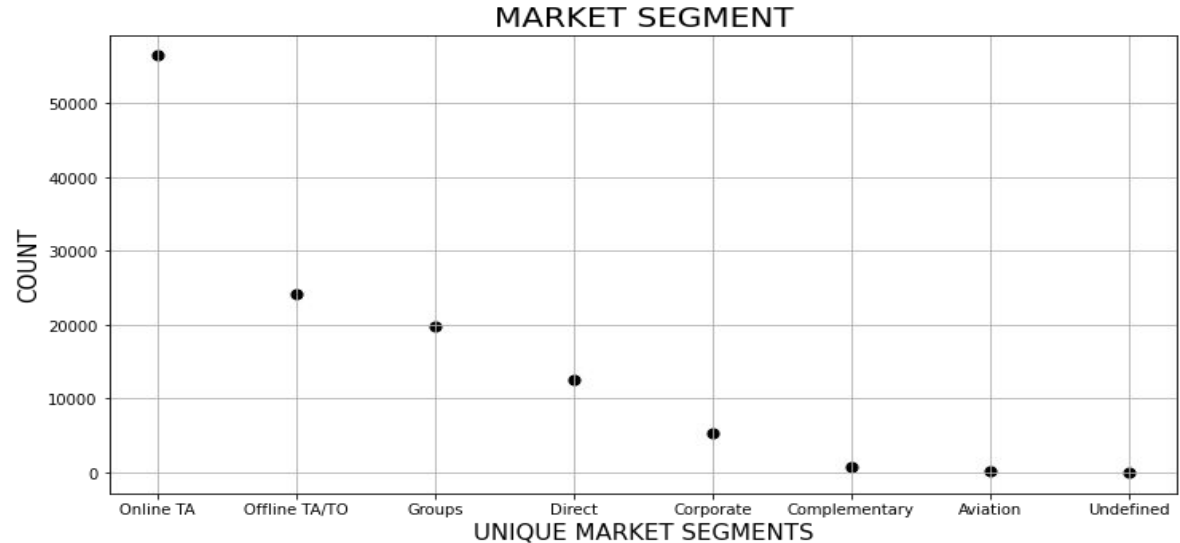
MARKET SEGMENT

From the counts table, it's clear that in market segment 'Online TA' has the highest counts out of all so, it can be stated that 'Online TA' brings the maximum number of bookings.

ADR (Average Daily Rate)

'Average Daily Rate' (ADR) is one of the key performance indicator (KPI) of the industry. Its used to measure the average price customers are paying per room per night on a given period of time.


The higher the ADR, the better.

A rising ADR suggests that a hotel is renting out there rooms more often and scaling up there revenue. To increase the ADR, hotels should look into ways to boost price per room.

Hotel operators seek to increase ADR by focusing on pricing strategies. This includes upselling, promotions, and complimentary offers.

This histogram shows the distribution of ADR values of the hotel data in this dataset.

- The dotted line marks the average ADR value which is approximately 100.

- Most of the hotels have the ADR values in between 90 to 120.

- Very few hotels have ADR above 150.

AVERAGE DAILY COUNT RATE
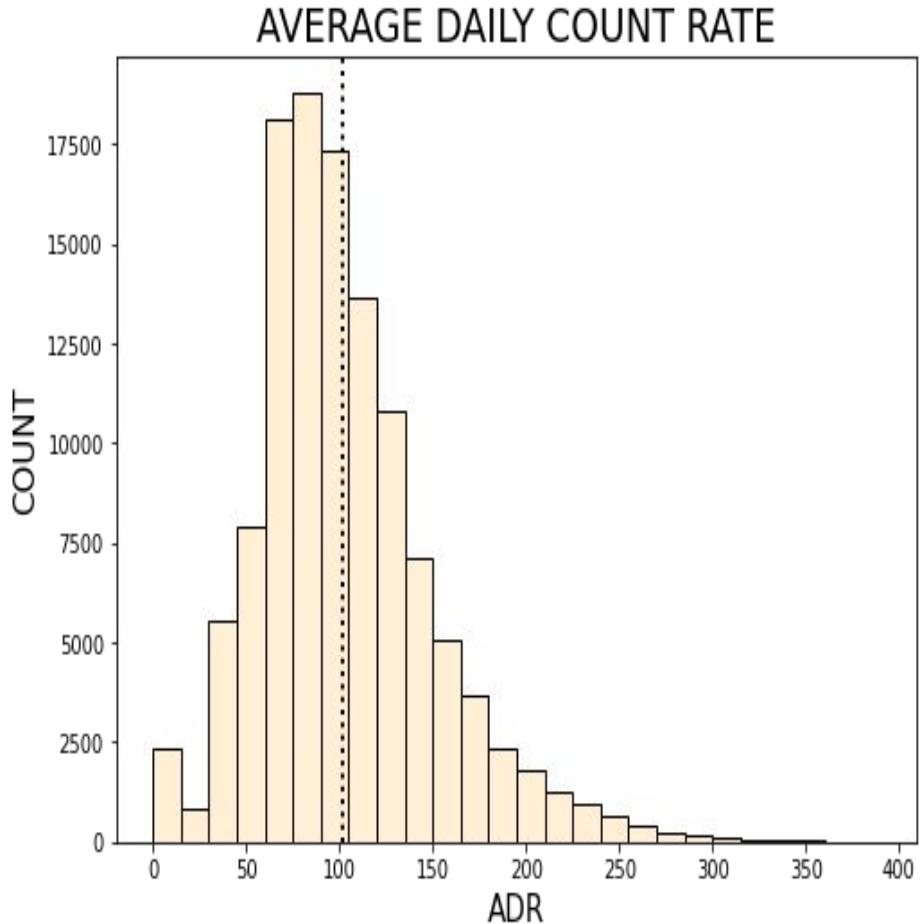
Let's have a look at the average ADR of two different types of hotels present in the dataset which are 'City hotel' & 'Resort hotel'.

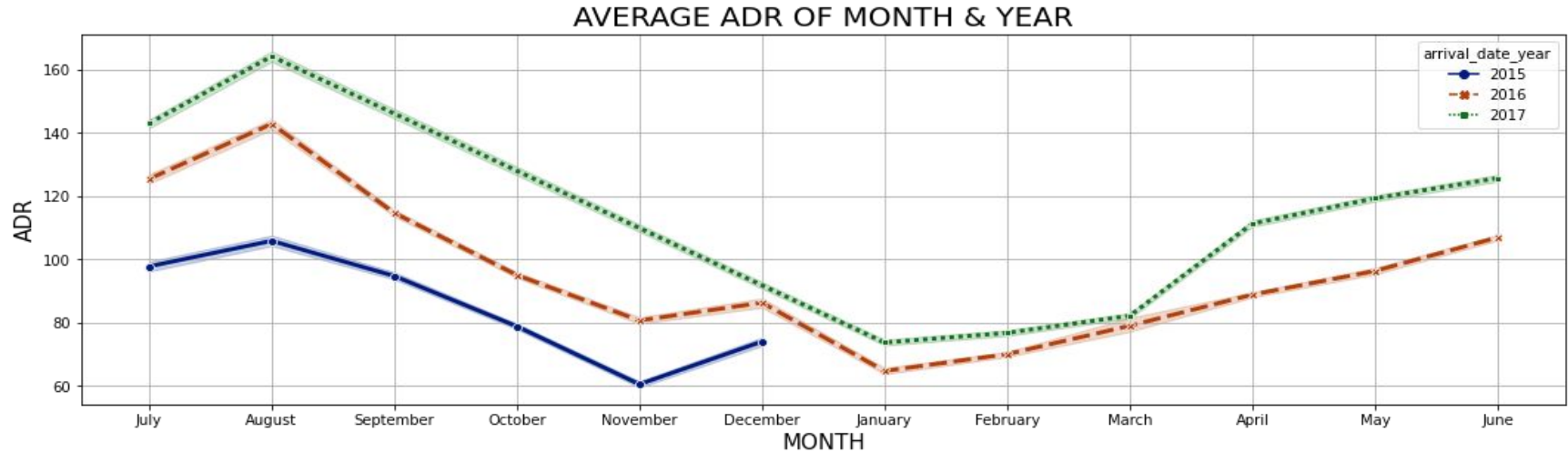| Hotel	Average ADR
City Hotel	105.304465
Resort Hotel	94.952930

So, from the table above & the horizontal bar chart on the right it's clear that average ADR of 'City hotel' is more than 'Resort hotel' though by a minute margin of only 10.35 approximately.

AVERAGE ADR OF HOTEL TYPES
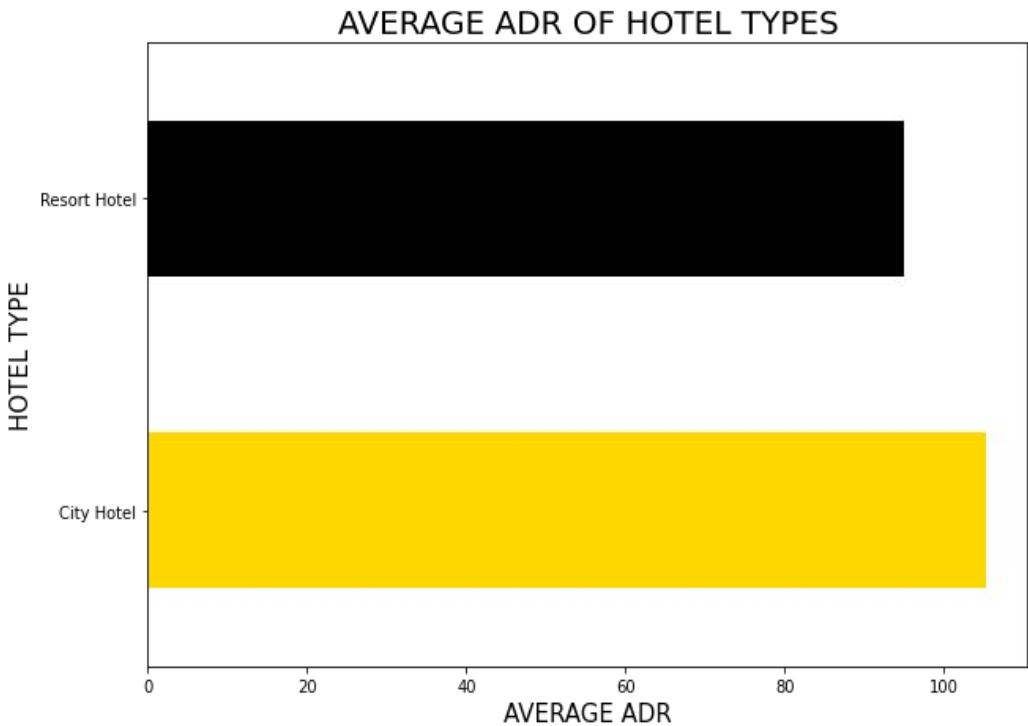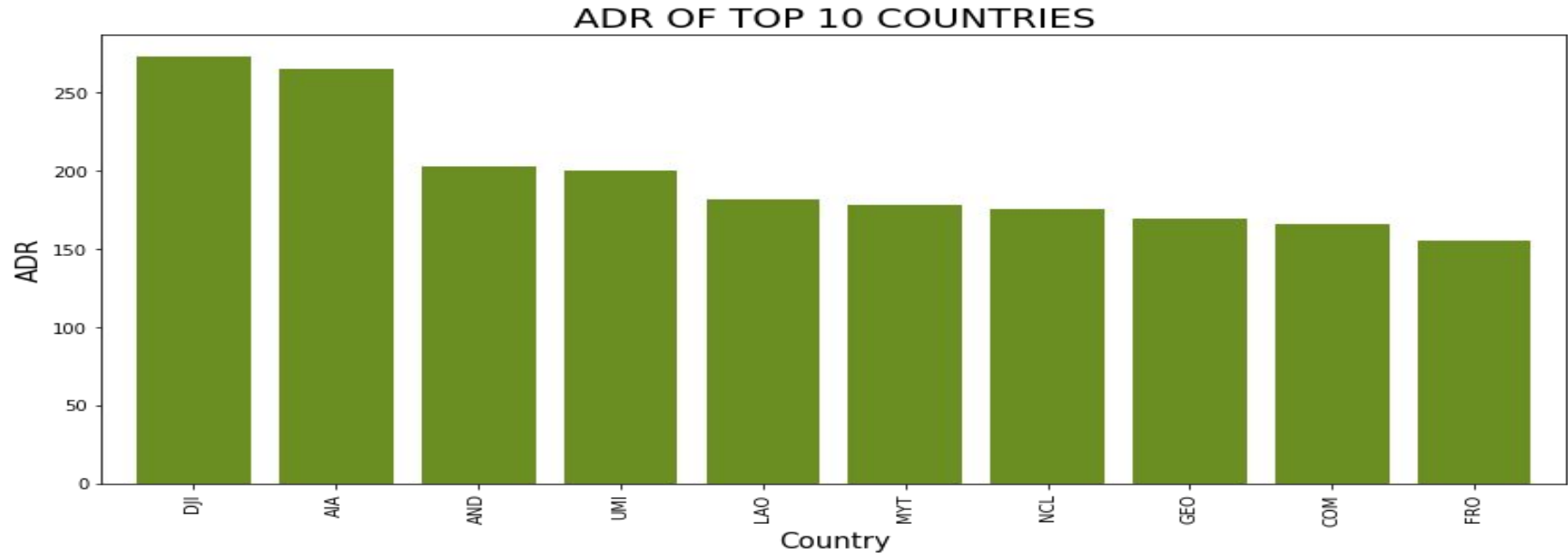
This bar chart shows the country having the best ADR. I have only considered the top 10 countries with high ADR values.

DJI(Djibouti) tops the list with a value of 273 approximately.

This line chart shows ADR values of each month for 3 different years, which are 2015, 2016, & 2017.

- In each year 'August' has witnessed highest ADR values.
- 'June' to 'August' timeframe has seen rise in hotel business compared to other months.
- Average ADR is rising every year which affirms that hotel business is scaling up significantly each year.

Meal analysis

PERCENTAGE OF MEAL PREFERENCE
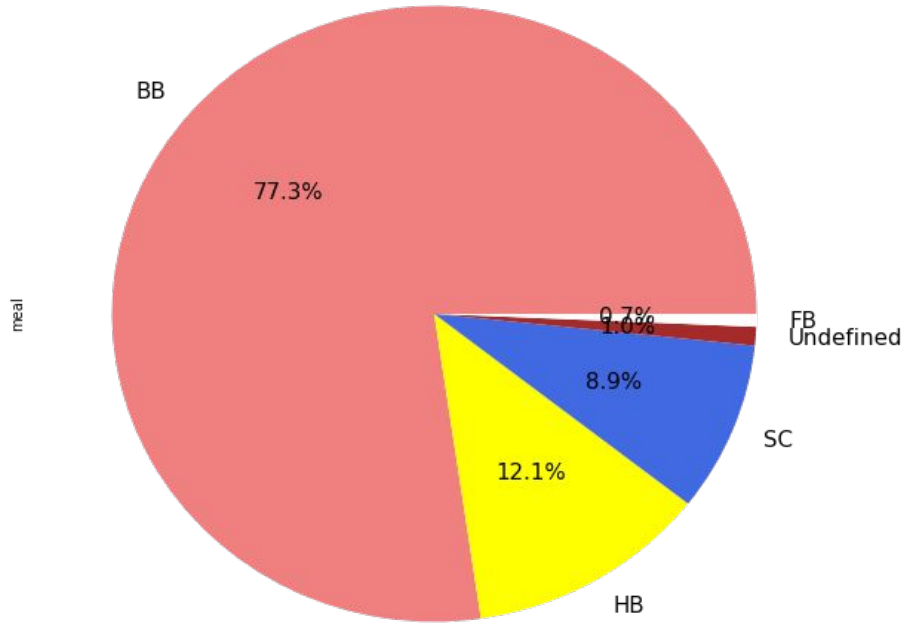
Almost 77.3% of customers prefer BB (Bed & Breakfast).

A Bed & Breakfast or BB is an accommodation type which provides an overnight stay as well as the first meal of the day included in the price.

So, it can be said that customers mainly preferred to take breakfast at their respective hotel & rest meals outside their hotel. So, people like to explore more in the city at different places.

Car parking analysis
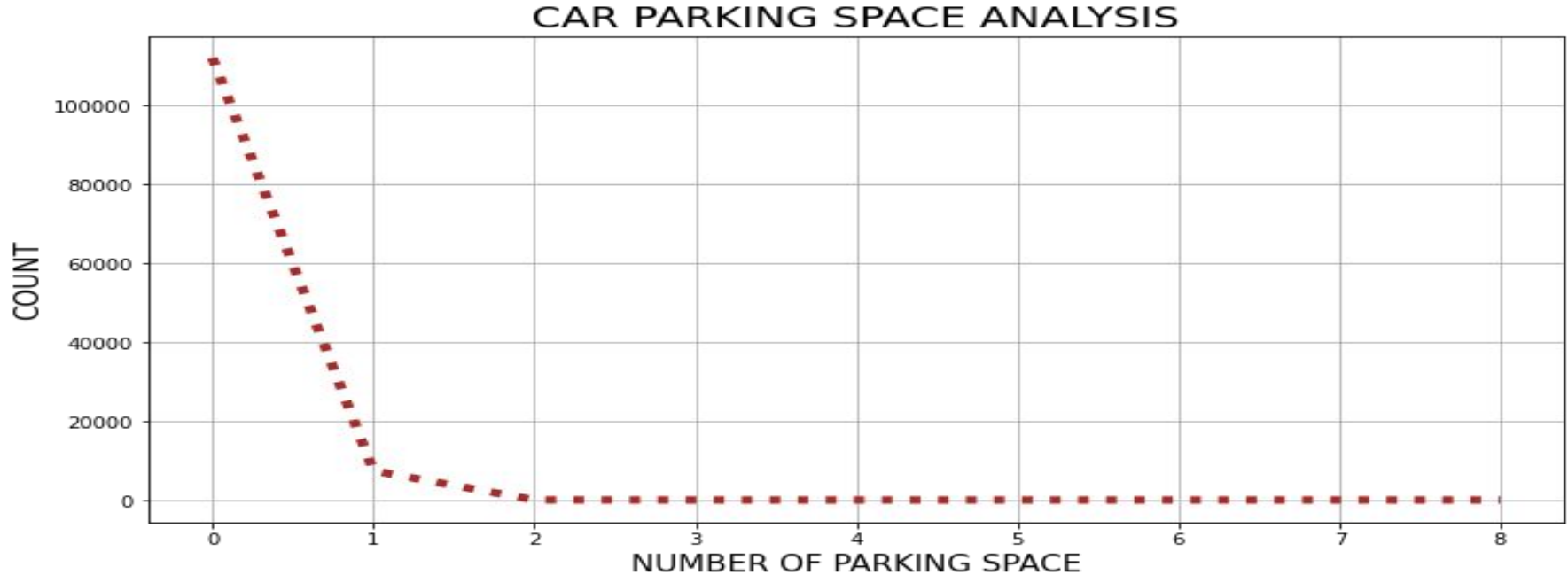
From the line chart above, its intelligible that most of the hotels have 0 to 1 car parking space available, which indicates that most of the people like to use public transport facilities.

Repetition of guests

Here, 0 = Guests did not repeat their bookings
    1 = Guests repeated their bookings.

So, it can be said that most of the guests did not repeat their bookings & maximum are first time guests.

So, hotel management may come with a marketing tactic to increase the bookings of their repeated guests also.

REPEATED & NON REPEATED GUEST
(0 = Not repeated & 1 = Repeated)
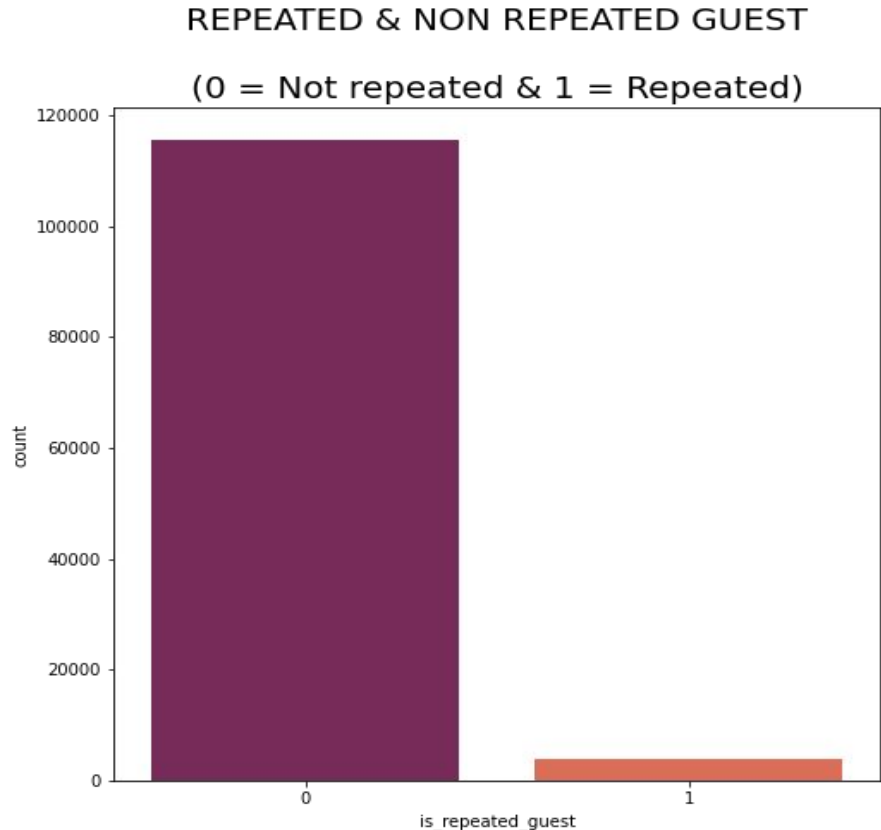
Analysis on night stays
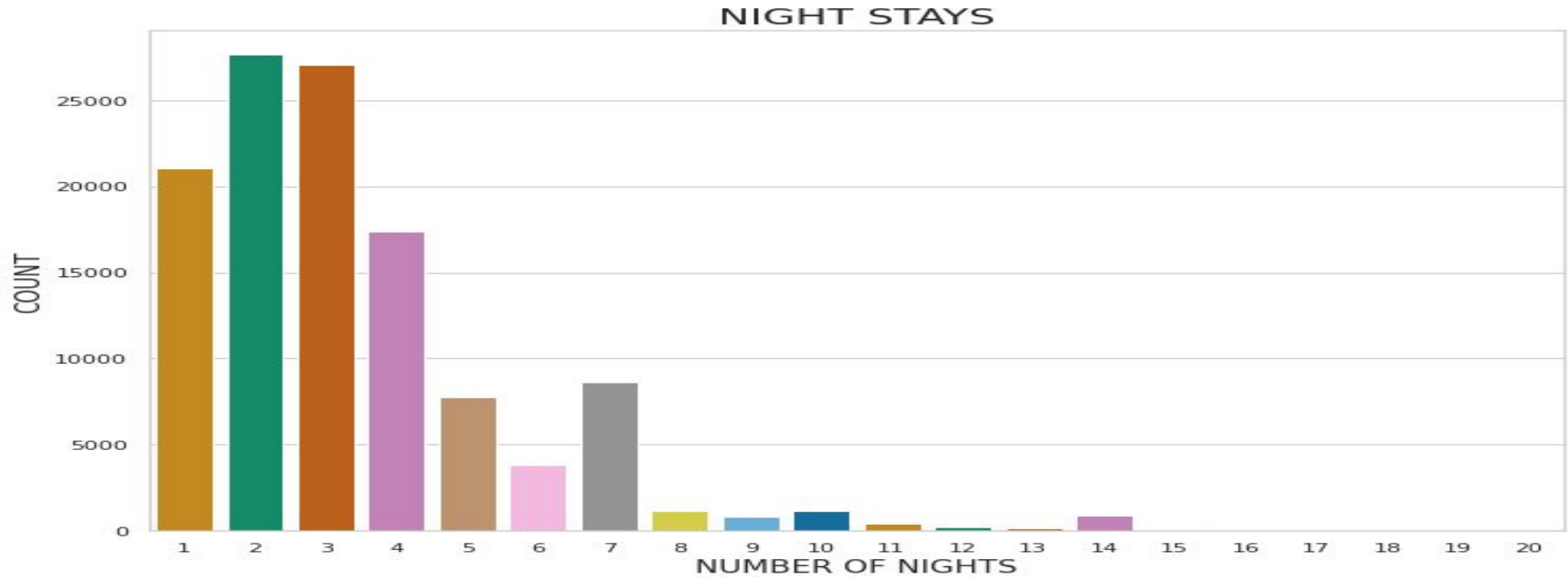
Here, it's clearly visible that most of the guest stayed in the hotel for 1, 2, 3, 4, 5 & 7 nights

Conclusion

- Most preferred hotel by customer is 'City hotel'.
- Maximum bookings took place in 'August' & minimum in 'January'.
- 'Portugal' is the most preferred place for tourism with highest number of bookings.
- 2016 witnessed highest number of bookings.
- Along with the highest bookings in 'City hotel' it also has highest cancellation rate.
- 'Online TA' brings maximum bookings.
- Average ADR of hotel is approximately 100.
- 'City hotel' also leads in the case of average ADR in comparison with 'Resort hotel'.
- Again, 'August' is more favourable for hotel business with highest ADR in all 3 years.
- 'Djibouti' has got the highest ADR value out of all countries.
- Each year hotel business is scaling up with the increase in average ADR every year.
- Most of the guest prefers 'Bed & Breakfast'(almost 77.3%).
- Most of the hotels have 0 to 1 car parking space.
- Hotels are more booked by new guests rather than repeated guests.
- Most of the guest stays for 1, 2, 3, 4, 5 & 7 nights in the hotel.

THANK YOU