Capstone Project-1

Hotel Booking Analysis

Team Members

Pankaj Rathod

Rahul Chouhan

Content :-

- Data Exploration
  - Observe the Data
  - Find Missing
- Data Cleaning
  - Replace the Null Values
  - Drop unnecessary columns
- Analyse the Data
- Data Visualisation
- Conclusion

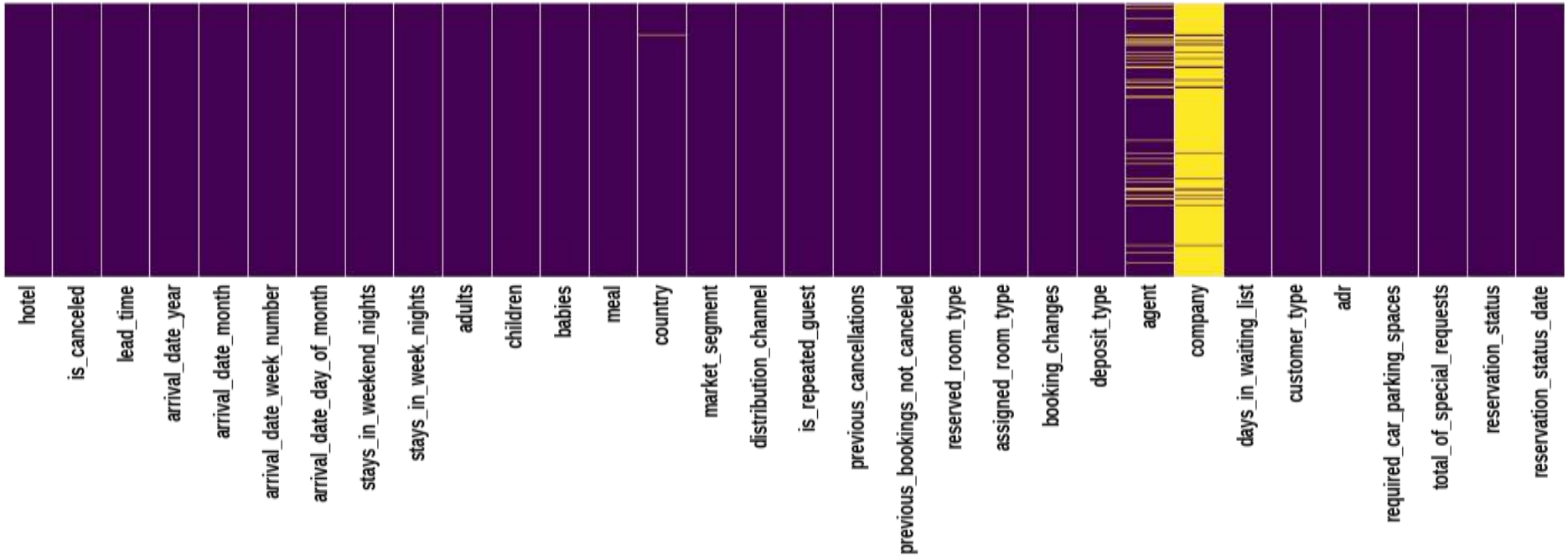
Why Hotel Booking Demand Analysis?

- Hospitality industry is a big contributor to the economic growth of any country.
- With a consistently growing middle class and increasing disposable income, the tourism and hospitality sector is witnessing healthy growth.
- The growth in the hospitality sector and its contributions to the GDP will continue to substantially increase.
- A hotel is an establishment that provides lodging, meals, and other services for travelers and other paying guests.

Exploratory Data Analysis with Python

Preparing Our Dataset

- Libraries used: **Numpy, Pandas, Seaborn, Matplotlib, and Plotly**.
- Shape: **119390, 32**
- Most of the **null values were present in columns - company and agent**



First Five Rows Data

```
# Checking first 5 rows in dataset
df.head(5)
```

	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_w
0	Resort Hotel	0	342	2015	July	27	1	0	
1	Resort Hotel	0	737	2015	July	27	1	0	
2	Resort Hotel	0	7	2015	July	27	1	0	
3	Resort Hotel	0	13	2015	July	27	1	0	
4	Resort Hotel	0	14	2015	July	27	1	0	

5 rows × 34 columns

Last Five Rows Data

```
# Checking the last 5 rows in dataset
df.tail(5)
```

	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
119385	City Hotel	0	23	2017	August	35	30	2	
119386	City Hotel	0	102	2017	August	35	31	2	
119387	City Hotel	0	34	2017	August	35	31	2	
119388	City Hotel	0	109	2017	August	35	31	2	
119389	City Hotel	0	205	2017	August	35	29	2	

5 rows × 34 columns

Explore The Dataset

Go Through The Dataset

hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number
City Hotel	0	44	2017	August	35
City Hotel	0	188	2017	August	35
City Hotel	0	135	2017	August	35
City Hotel	0	164	2017	August	35
City Hotel	0	21	2017	August	35
City Hotel	0	23	2017	August	35
City Hotel	0	102	2017	August	35
City Hotel	0	34	2017	August	35
City Hotel	0	109	2017	August	35

```
country                          488
market_segment                     0
distribution_channel               0
is_repeated_guest                  0
previous_cancellations             0
previous_bookings_not_canceled     0
reserved_room_type                 0
assigned_room_type                 0
booking_changes                    0
deposit_type                       0
agent                          16340
company                       112593
days_in_waiting_list               0
customer_type                      0
```

Checking Null Values In Dataset

Replacing The Null values with There mean

```
country                          0
market_segment                   0
distribution_channel             0
is_repeated_guest                0
previous_cancellations           0
previous_bookings_not_canceled   0
reserved_room_type               0
assigned_room_type               0
booking_changes                  0
deposit_type                     0
agent                            0
company                          0
days_in_waiting_list             0
customer_type                    0
```

Description of the Data

```
[8] # Exploring descriptive statistical parameter

    df.describe()
```

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	stays_in_week_nights	
count	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.000000	119390.0
mean	0.370416	104.011416	2016.156554	27.165173	15.798241	0.927599	2.500302	1.8
std	0.482918	106.863097	0.707476	13.605138	8.780829	0.998613	1.908286	0.5
min	0.000000	0.000000	2015.000000	1.000000	1.000000	0.000000	0.000000	0.0
25%	0.000000	18.000000	2016.000000	16.000000	8.000000	0.000000	1.000000	2.0
50%	0.000000	69.000000	2016.000000	28.000000	16.000000	1.000000	2.000000	2.0
75%	1.000000	160.000000	2017.000000	38.000000	23.000000	2.000000	3.000000	2.0
max	1.000000	737.000000	2017.000000	53.000000	31.000000	19.000000	50.000000	55.0

Visualization

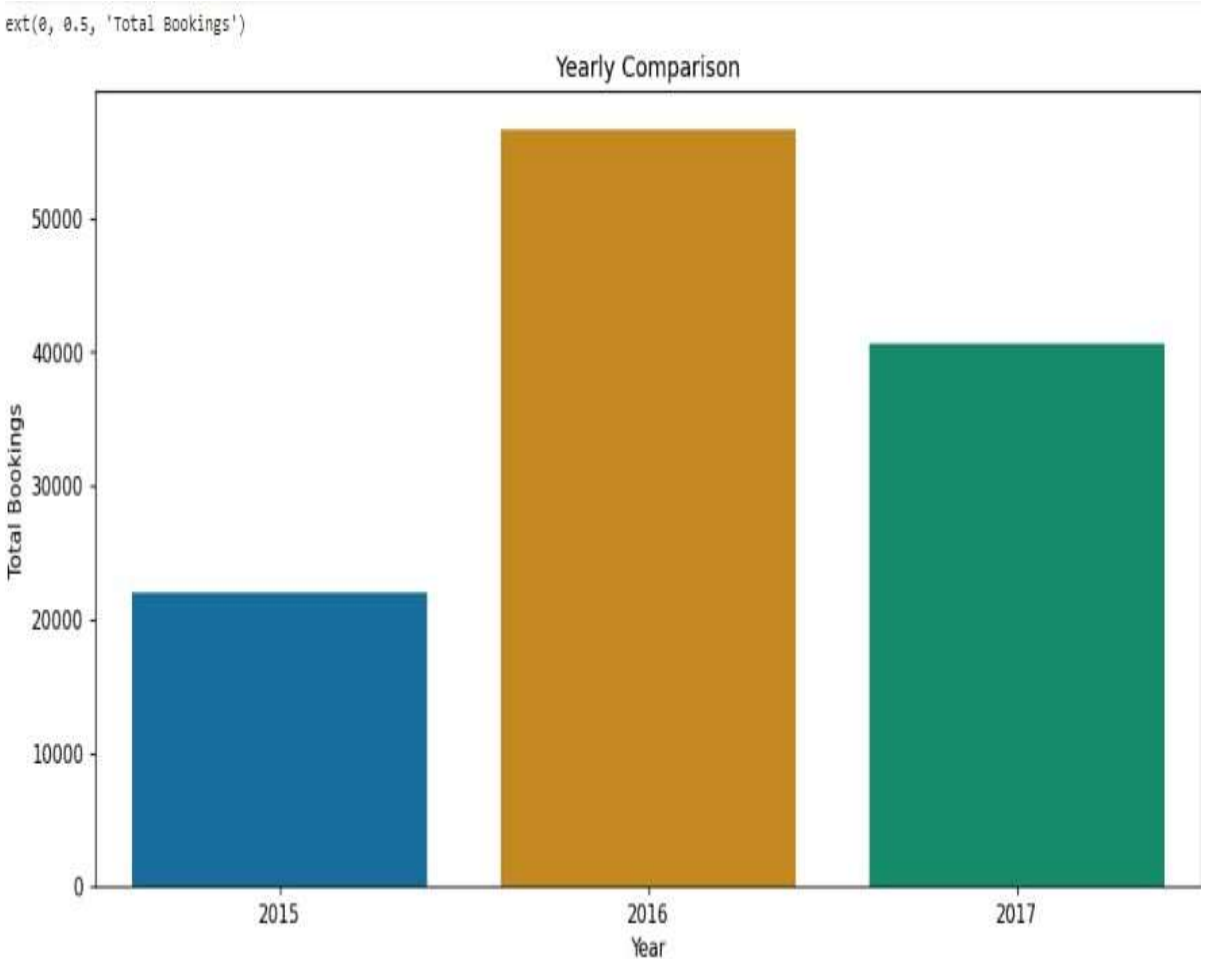
Total Number Of Bookings Across Different Years

How many customers actually checked in to the hotel across different years?

Let us find out with a simple bar chart, in 2015 there are 18.5% of customers checked in.

Whereas in 2016 we can see that there is an increase in bookings up to 47.4%.

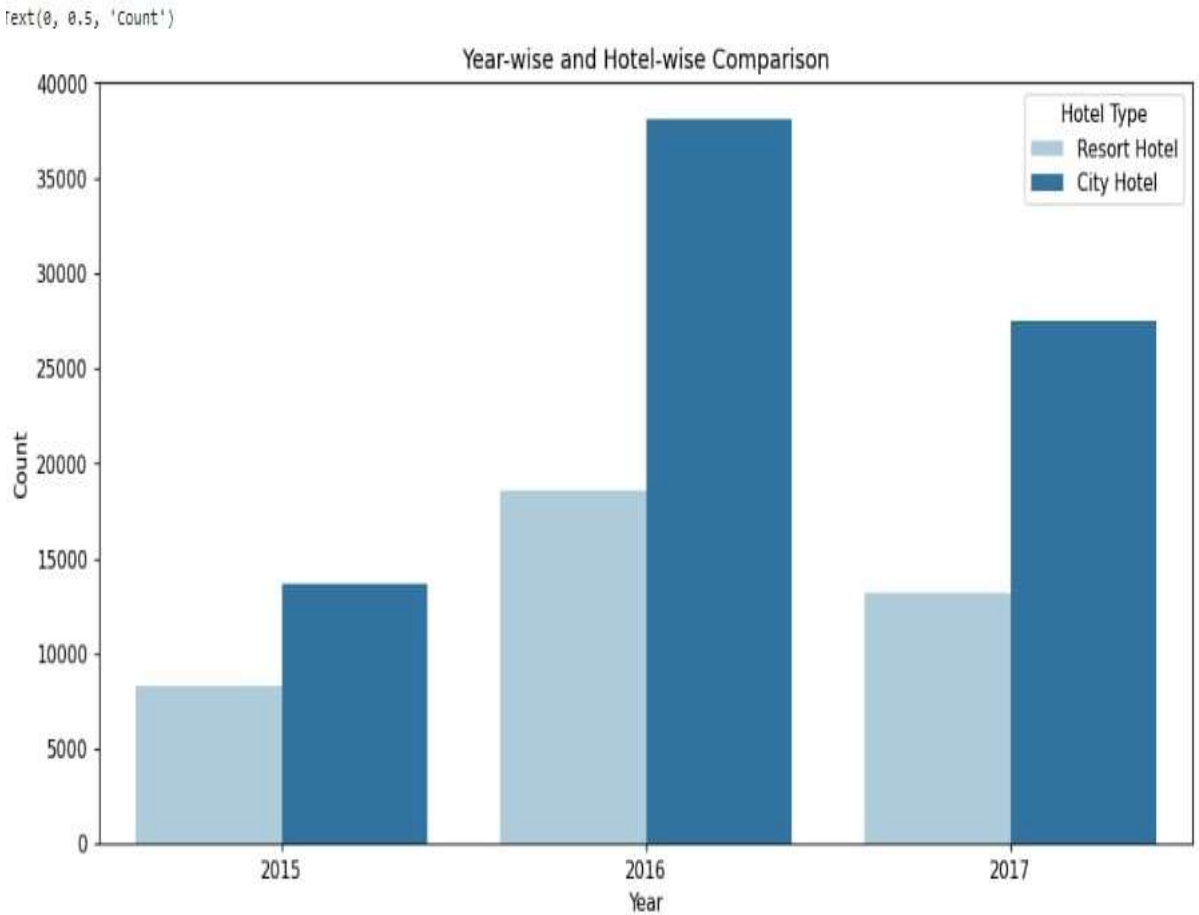
This increase in trend did not sustain for more time, going downward in 2017 with only 34.1% bookings.



➢ Demand Trend Of Hotels Year wise

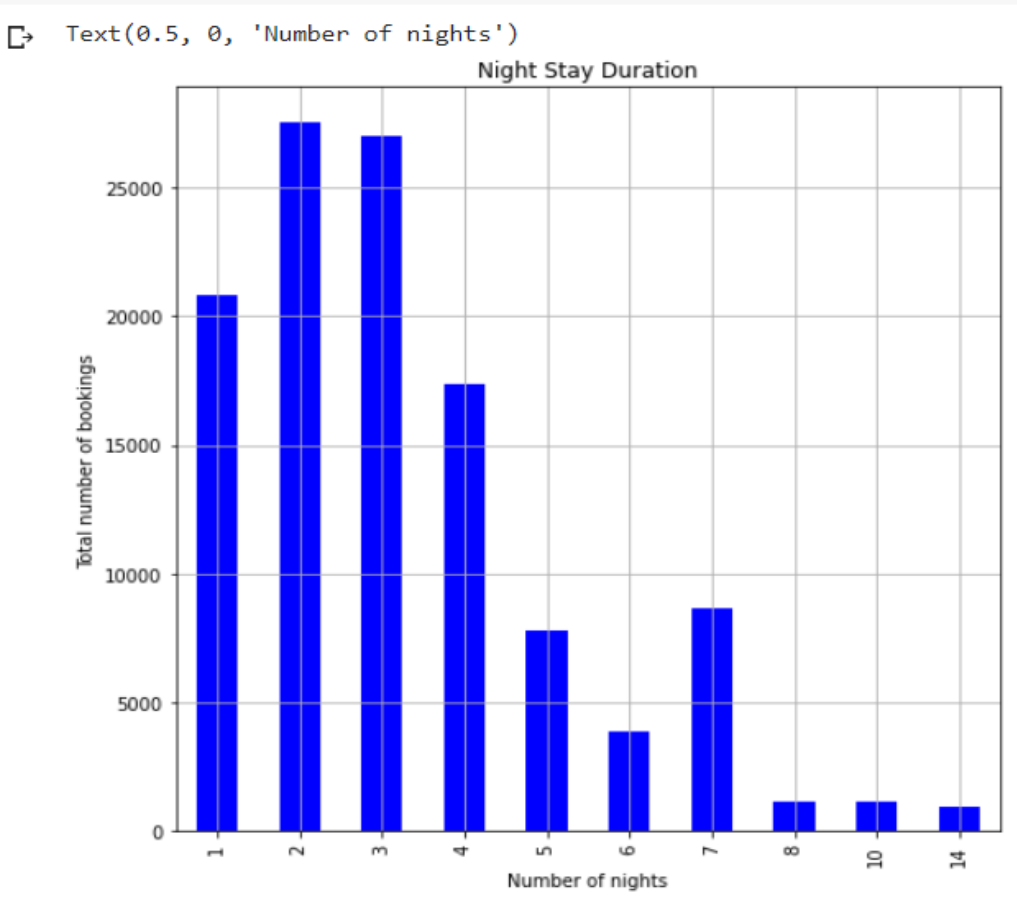
Which type of hotels customers preferred to stay in different years?

Here, we plotted a subplot for Resort hotel and City hotel. From these columns we can conclude that there is always demand of City hotels as compared to Resort hotels across three different years 2015, 2016 and 2017. As we discussed early, after increasing the booking trend it got decreased again. This happened in both cases – Resort as well as for City hotels.



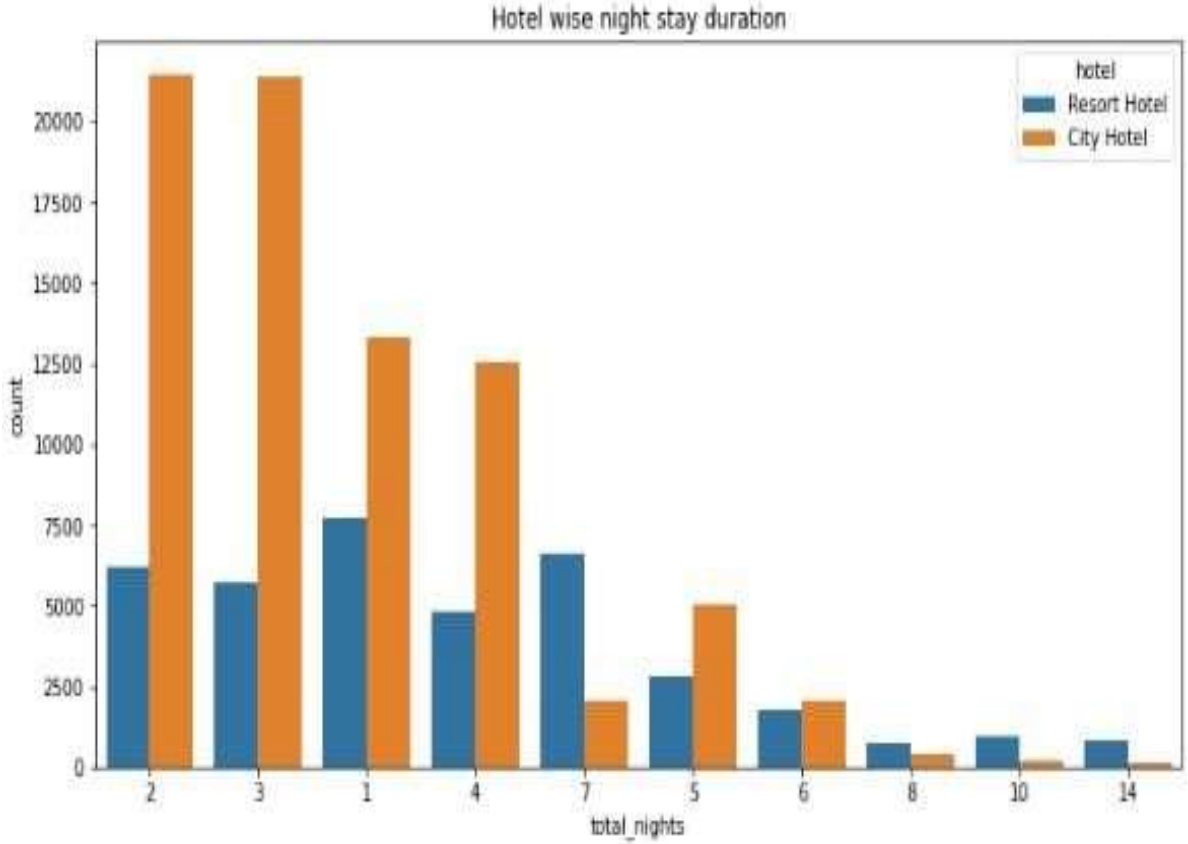
➢ Night Stay Duration

By combiningthe two columns of and stays_in weekend_nights we got a total number of nights. Hence, we can say that more customers like to spend 2 – 3 nights while some customers prefer to stay for 1 – 4 nights. Very few customers are there who are interested to stay for more than 5 days.



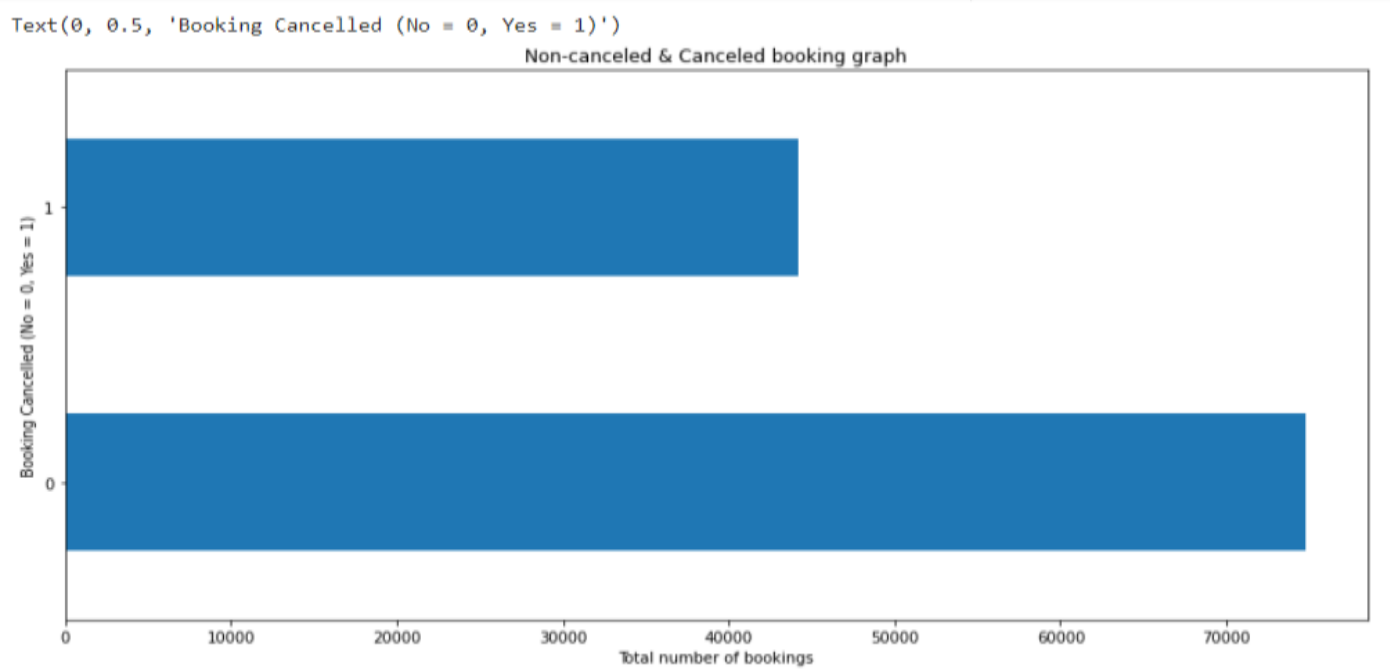
Hotel-wise Night Stay Duration

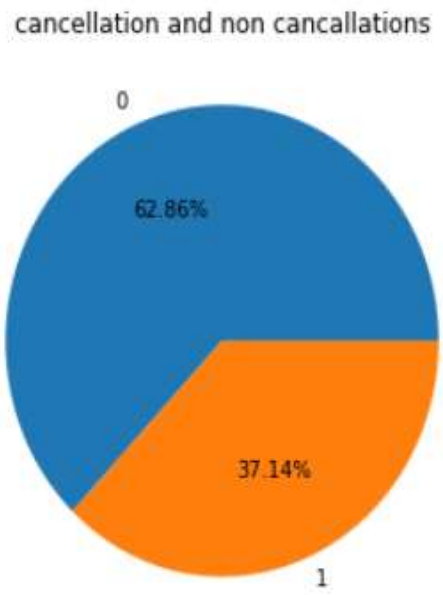
According to this visual representation, people are interested in city hotel to stay for two, three, one and four days respectively whereas for resort hotel people are interested to stay one, seven, two and three days respectively.



Non-Canceled vs Canceled Booking Percentage

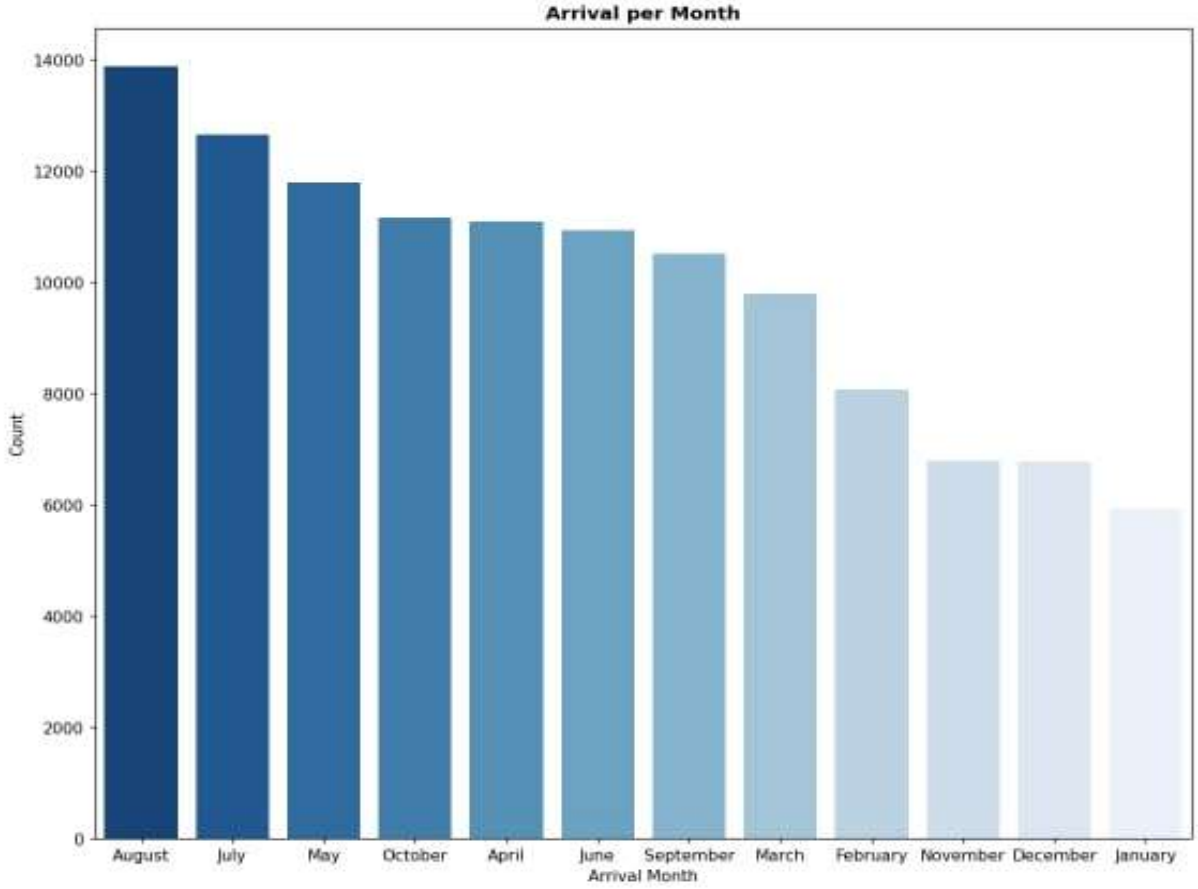
As we saw the total number of booking from our last slide, here we are going to see the same but in terms of percentage. This bar graph Representing that 63% of customers are check-in at hotels where 37% of customers canceled their booking





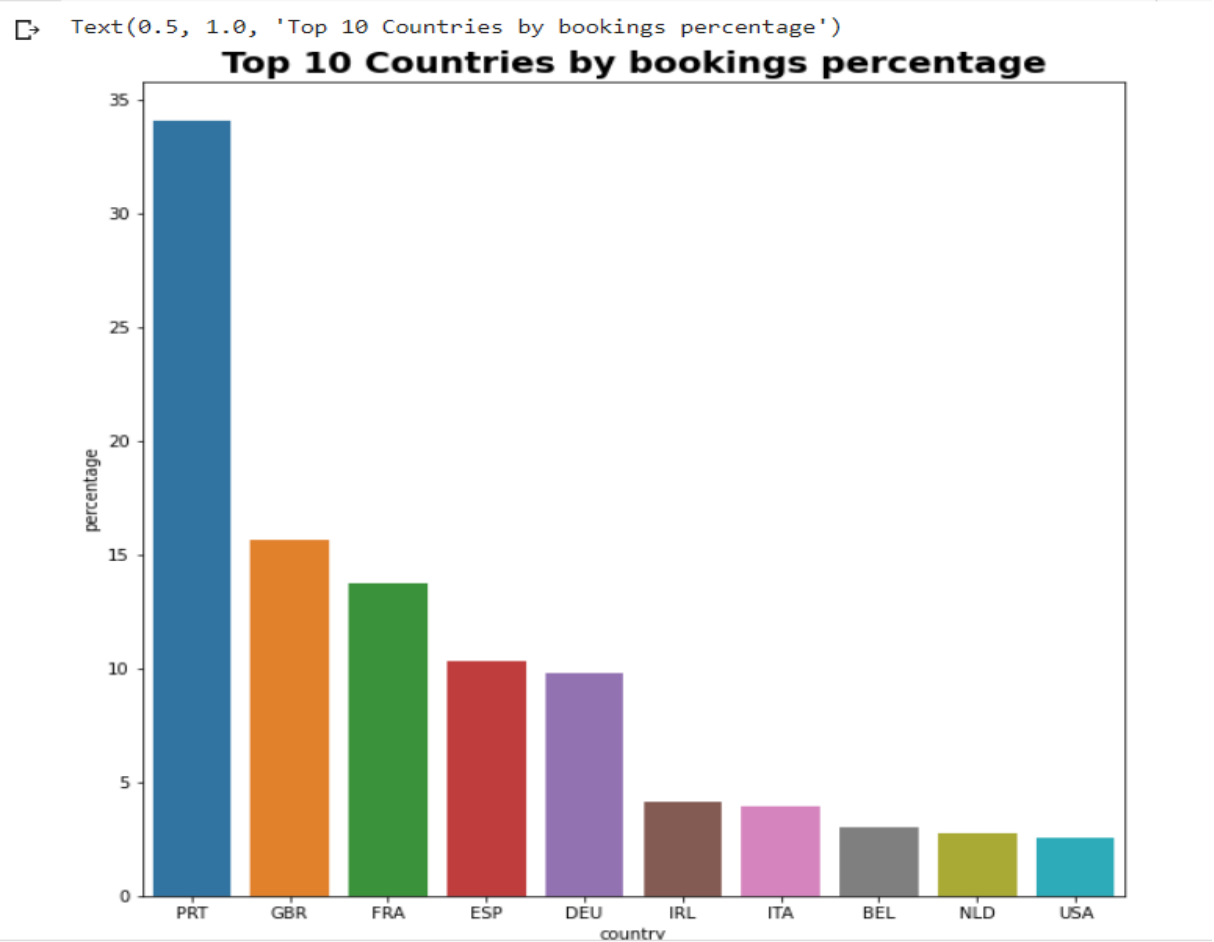
Booking Trend Throught The Year

According to this graph of arrival month and count, August has the highest number of customers followed by July, May, October, and so on and so forth, while January has the lowest rate.



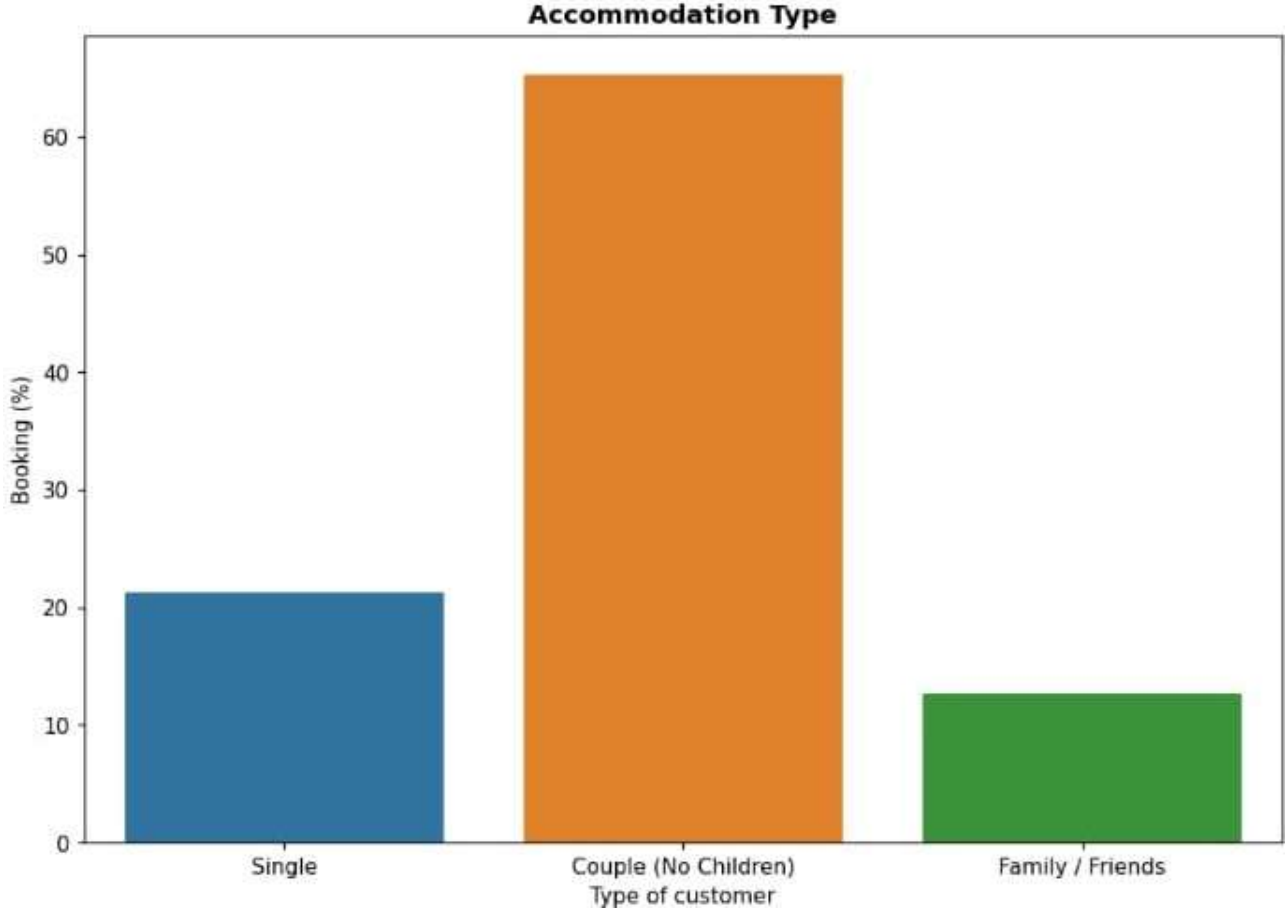
Top 10 Countries With Maximum Customers

We found that Portugal ranks first with 48590 customers followed by the UK, France, Spain, and Germany, with 12129, 10415, 8568, and 7287 customers. The Netherlands has the lowest number of customers in the back seats



Accommodation Type – Single, Couple & Family

From the graph, it is seen that most numbers of customers are containing 2 adults customer and those who have more than 2 either containing adults, children & babies have the lowest number of customers.



Conclusion

The majority of the hotels booked are city hotels. Definitely need to spend the most targeting fund on those hotels.

We should also target months between May to Aug because these are peak months.

The majority of the guests are from Western Europe. We should spend a significant amount of our budget on that area

The most number of customer are containing 2 adults customer and those who have more than 2 either containing adults, children & babies have the lowest number of customer

Challenges

Huge chunk of data was to be handled by keeping in mind not to miss anything which is even of little relevance.

Handling with too many null values and replacing it.

References:-

1) Geeks for Geeks

2) Analytics Vidhya

3) Stack oveíflow.

4) Kaggle.

Thank You!