AImaBetter

Capston Project

Hotel Booking Analysis

Submitted By :- Harsh Durugkar

Shivangini Gupta
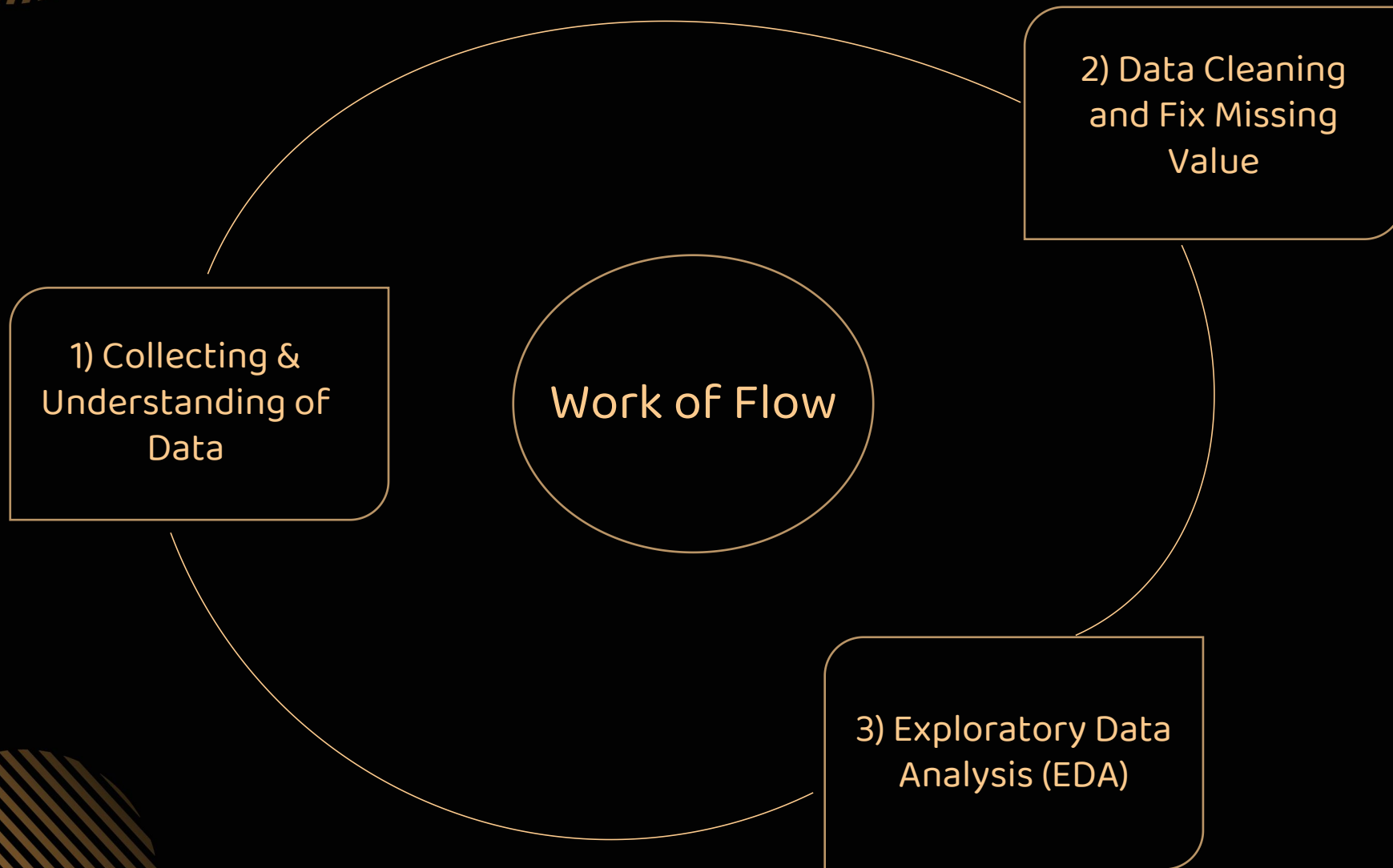
Rajesh Khangar

AI maBetter

Contents

AImaBetter

1) We Have a DataSet of Hotel Booking and we discuss about this dataset as well as cover some querry with analysing.

2) In Analysing the data we work with Hotel Booking DataSet in this dataset containing include the information such as hotel resort vs city , booking cancellation , types of customers , meal of customers , agent information & many more.

3) Hotel Industry is Highly Traffic Industry with Lakh of People with Lakh of Data and with collecting and maintain the information for analysing , give big benifit to Hotel Industry.

4) The main target behind this project is to explore and then analyze the data for discover important factors and find out the insights to solve out problem in Hotel Management , So they easily observe advantages & weakness and perform campaigns to boost there bussiness as well as Performance.

Data Review

In the Given Hotel Booking Dataset there are 119390 number of rows and 32 number of columns.
So let's understand every columns which is contain in dataset :--

1) <u>hotel</u> :- In this column contain two categorical data which is Resort Hotel & City Hotel .

2) <u>is_cancelled</u> :- In this column value show the cancellation type. If the booking was cancelled
Value indicate 1 & 0 indicates not cancelled.

3) <u>lead_time</u> :- Data show the time between reservation and actual arrival .

4) <u>arrival_date_year</u> :- Data show Year of arrival date .

5) <u>arrival_date_month</u> :- Data show Month of arrival date .

6) <u>arrival_date_week_number</u> :- Data of Week number of year for arrival date .

7) <u>arrival_date_day_of_month</u> :- Day of arrival date .

8) <u>stays_in_weekend_nights</u> :- Total Number of weekend nights .

9) <u>stays_in_week_nights</u> :- Total Number of week nights .

10) <u>adults</u> : - Total Number of adults .

11) <u>children</u> :- Total Number of children in Hotel .

12) <u>babies</u> :- Total Number of babies in Hotel .

13) <u>meal</u> :- Type of meal booked by a Customers .

14) <u>country</u> :- Country of origin of Customers .

15) <u>market_segment</u> :-Designation of market segment . (TA/TO) .

16) <u>distribution_channel</u> :- Booking distribution channel.(T/A/TO) .

17) <u>is_repeated_guest</u> :- is a repeated guest or not. (1) means yes & (0) means not.

18) <u>previous_cancellations</u> :- Number of previous bookings that were cancelled by the customer prior to the current booking .

19) <u>previous_bookings_not_canceled</u> :- Number of previous bookings not cancelled by the customer prior to the current booking

20) <u>reserved_room_type</u> :- Code of room type reserved.

21) <u>assigned_room_type</u> : Code for the type of room assigned to the booking.

22) <u>booking_changes</u> : Number of changes made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

23) <u>deposit_type</u> :- Values of No Deposit, Non Refund , Refundable.

24) <u>agent</u> :- Values of ID of the travel agent that made the booking.

25) <u>company</u> :- Contain ID of the company or entity that made the booking .

26) <u>days_in_waiting_list</u> :- Number of days the booking was in the waiting list before it
    was confirmed to the customer.

27) <u>customer_type</u> :- Data Contain types of customer & Contract,Group,transient,Transient party.

28) <u>adr</u> :- Average Daily Rate as defined by dividing the sum of all lodging transactions by the total
    number of staying nights.

29) <u>required_car_parking_spaces</u> :- Data of Number of car parking spaces required by the customer

30) <u>total_of_special_requests</u> :- Data of Number of special requests made by the customer

31) <u>reservation_status</u> :- Reservation status.

32) <u>reservation_status_data</u> :- Date of reservation status.

AImaBetter

04

Types of Data in Dataset

AI maBetter

Types of data

Binary Values

is_canceled
is_repeated_guest

Categorical Data

hotel
arrival_date_month
meal
country
market_segment
distribution_channel
reserved_room_type
assigned_room_type
deposite_type
customer_type
reservation_status

Dataset

Numeric Data

arrival_date_year
children
babies
adults
stays_in_week_nights
stay_in_weekend_nights
arrival_date_day_of_month
arrival_date_week_of_number
booking_changes
total_of_specail_request
required_car_parking_spaces
adr
company
agents
days_in_waiting_list

AImaBetter

05

Insights to find out

AImaBetter

1. How Many Booking Were Cancelled?
2. What is the booking ratio between Resort Hotel and City Hotel?
3. What is the percentage of booking for each year?
4. Which is the most busy month for hotel?
5. From which country most guest come?
6. How Long People Stay in the hotel?
7. Which was the most booked accommodation type (Single, Couple, Family)?
8. How many guests repeated in hotel ?
9. Types of Customers ?
10. Meal Plan Of Customers.
11. Which agent makes highest no. of bookings ?
12. Which is the most preferred room type by the customers ?
13. Which Hotel type has the highest ADR ?

06
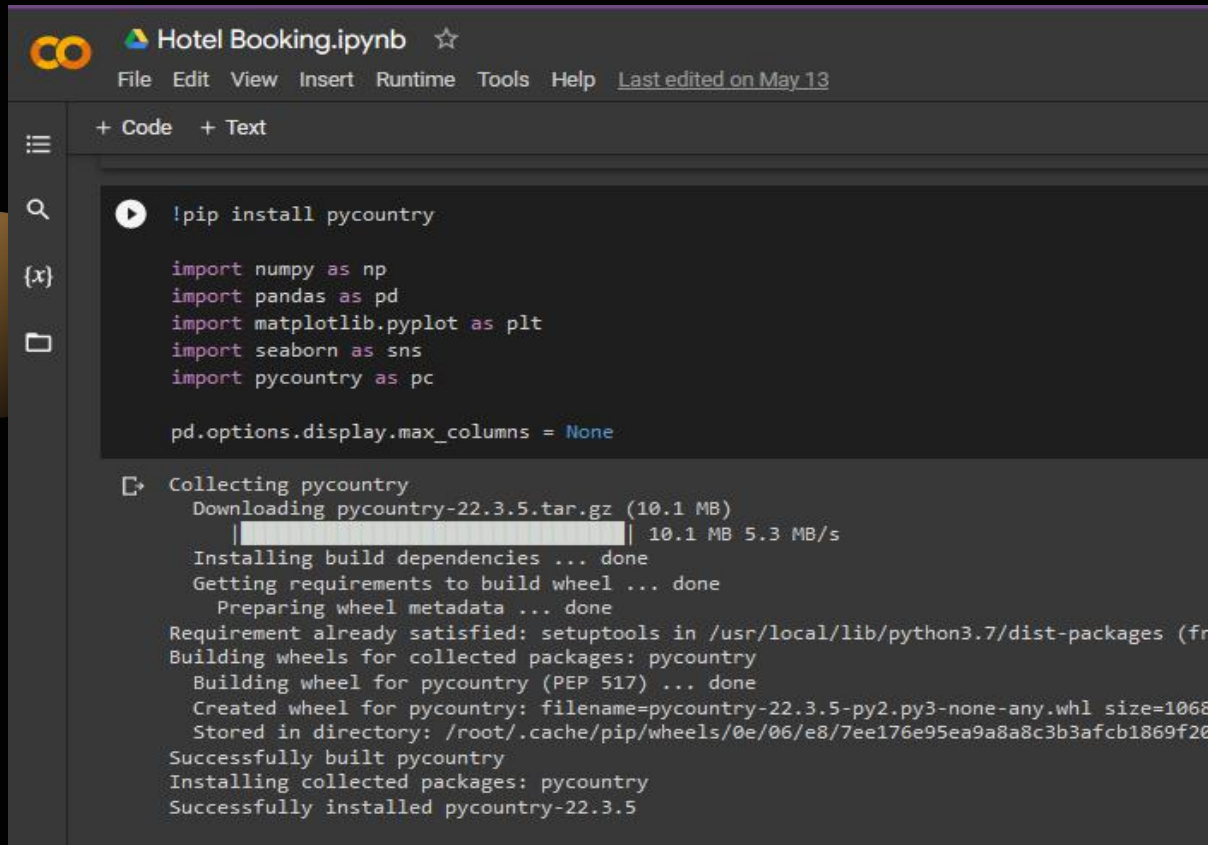
Exploratory Data Analysis (EDA)

AImaBetter

EDA

1) We Import all required library in code so we take advantages of library to solve out our problem . If in future we need more library so we import in this colab. Currently we add numpy , pandas , matplotlib , seaborn, pycountry etc.
2)Then we add our Data-Set file i.e excel file in it . Our Data-Set file is in google drive so we import google drive to link with that file & we import google drive then we give location of our file then call file with pandas library with the function of pd.read_csv() .This function read file excel file.
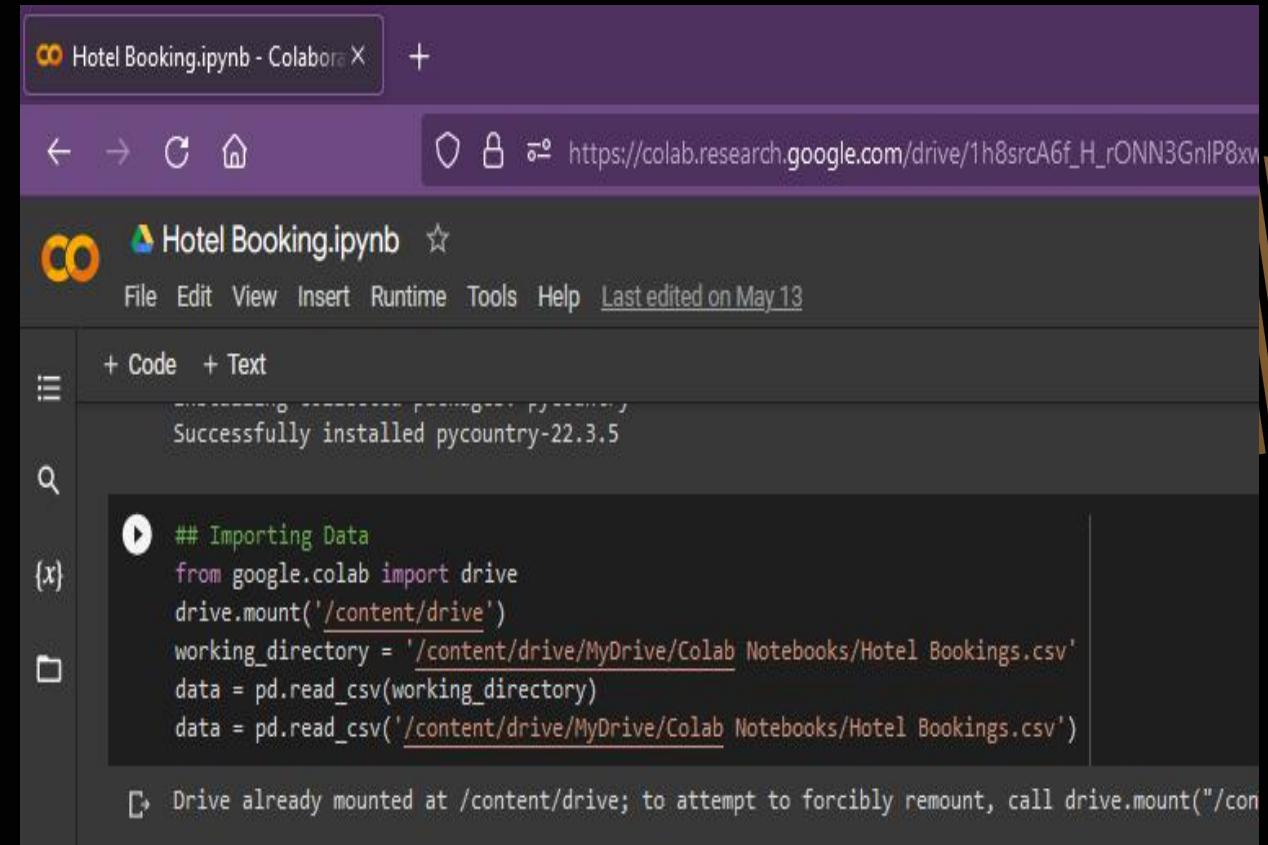
```
!pip install pycountry

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import pycountry as pc

pd.options.display.max_columns = None

Collecting pycountry
    Downloading pycountry-22.3.5.tar.gz (10.1 MB)
    |                              | 10.1 MB 5.3 MB/s
    Installing build dependencies ... done
    Getting requirements to build wheel ... done
      Preparing wheel metadata ... done
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (fr
Building wheels for collected packages: pycountry
    Building wheel for pycountry (PEP 517) ... done
    Created wheel for pycountry: filename=pycountry-22.3.5-py2.py3-none-any.whl size=1068
    Stored in directory: /root/.cache/pip/wheels/0e/06/e8/7ee176e95ea9a8a8c3b3afcb1869f20
Successfully built pycountry
Installing collected packages: pycountry
Successfully installed pycountry-22.3.5
```

```
Successfully installed pycountry-22.3.5

## Importing Data
from google.colab import drive
drive.mount('/content/drive')

working_directory = '/content/drive/MyDrive/Colab Notebooks/Hotel Bookings.csv'
data = pd.read_csv(working_directory)
data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Hotel Bookings.csv')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/con
```

3) Lets Check Missing Value in Dataset & Then Target Every missing value to fill & make data complete.

1. Dealing with Missing Values

```python
## Find the missing value, show the total null values for each column and sort it in descending order
df.isnull().sum().sort_values(ascending=False)[:10]

company                    112593
agent                       16340
country                       488
children                        4
reserved_room_type              0
assigned_room_type              0
booking_changes                 0
deposit_type                    0
hotel                           0
previous_cancellations          0
dtype: int64
```

```python
## Drop Rows where there is no adult, baby and child
df = df.drop(df[(df.adults+df.babies+df.children)==0].index)


## If no id of agent or company is null, just replace it with 0
df[['agent','company']] = df[['agent','company']].fillna(0.0)


## For the missing values in the country column, replace it with mode (value that appears most often)
df['country'].fillna(data.country.mode().to_string(), inplace=True)


## for missing children value, replace it with rounded mean value
df['children'].fillna(round(data.children.mean()), inplace=True)
```
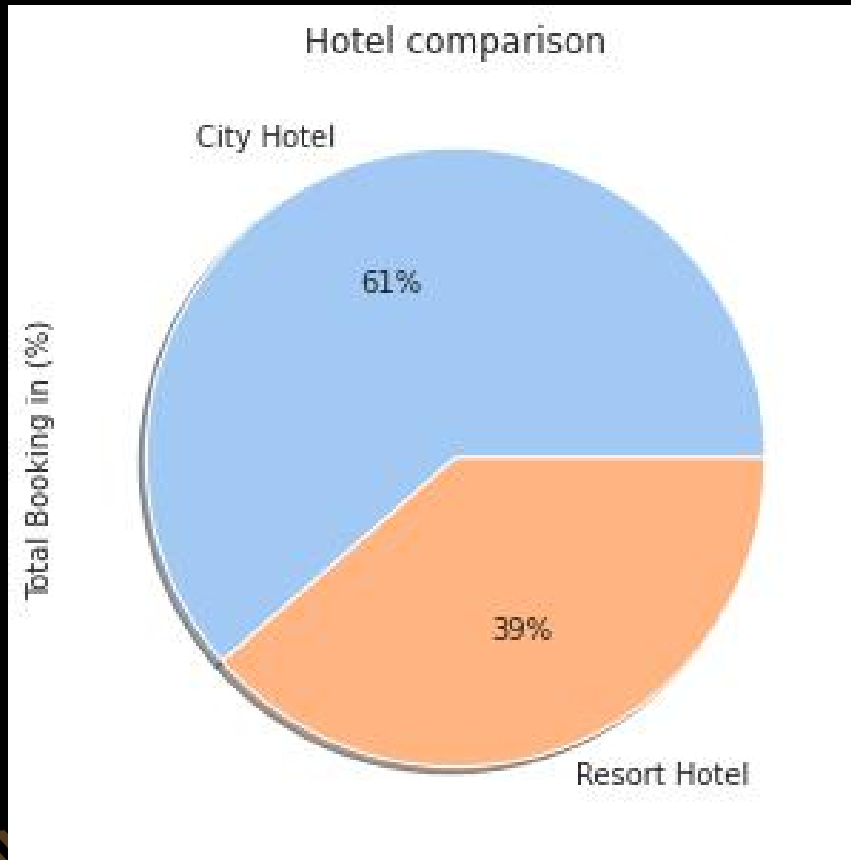
![AImaBetter logo]

1) How Many Booking Were Cancelled ?

Fig Shows, How many percentage of Booking Cancel in Hotel

Booking (%)

Booking Cancelled (No = 0, Yes = 1)
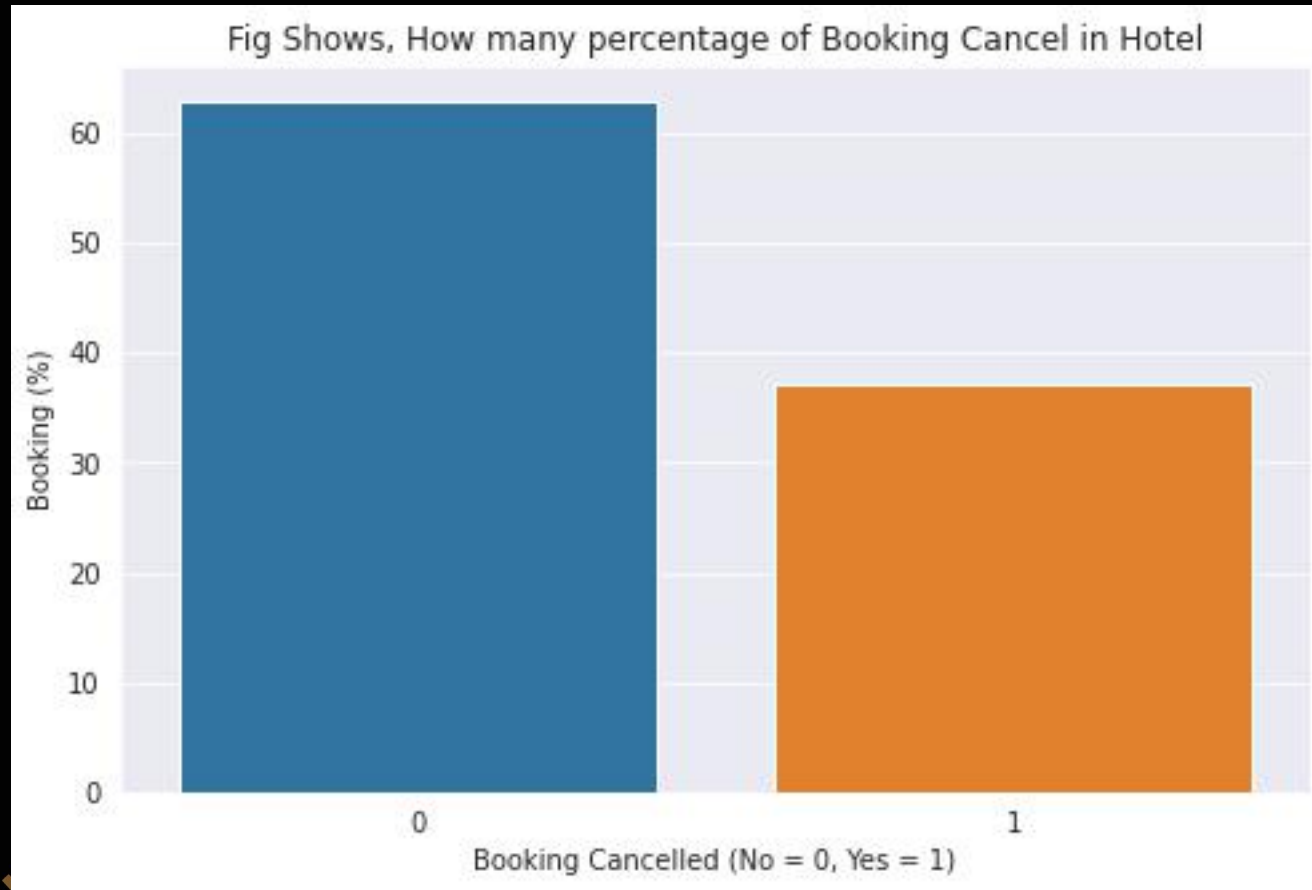
KEY INSIGHTS

According to visualization ,
There are near 37% Booking
Cancelled By Customers &
remaining 63% of Booking is
Safe.

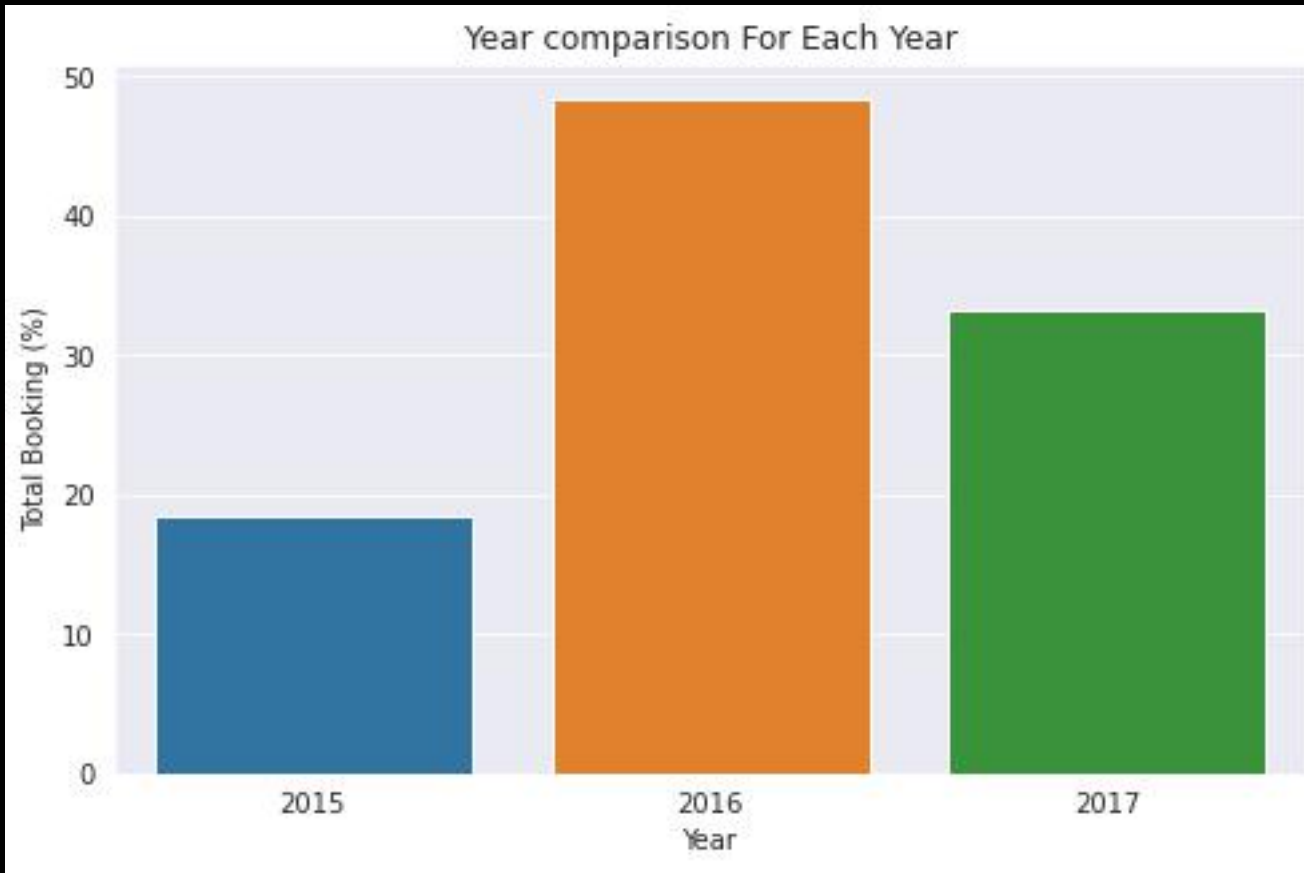## 2. What is the booking ratio between Resort Hotel and City Hotel ?

KEY INSIGHTS

According to visualization ,
There are 61% of Booking done
in City Hotel & remaining 39% of
Booking done in Resort Hotel.

So, City Hotel capture more
customers than the Resort Hotel.

EDA

3. What is the percentage of booking for each year ?

KEY INSIGHTS

Bar Chart Show the Total % of
Booking done in Each Year.
In Visualiztion we see that in
2015 Booking done by nearly 19%
adn in 2016 48% Booking & in
2017 Booking done 33% .
So 2016 is Successful Year For
Hotel Booking.

Let's separte it by hotel
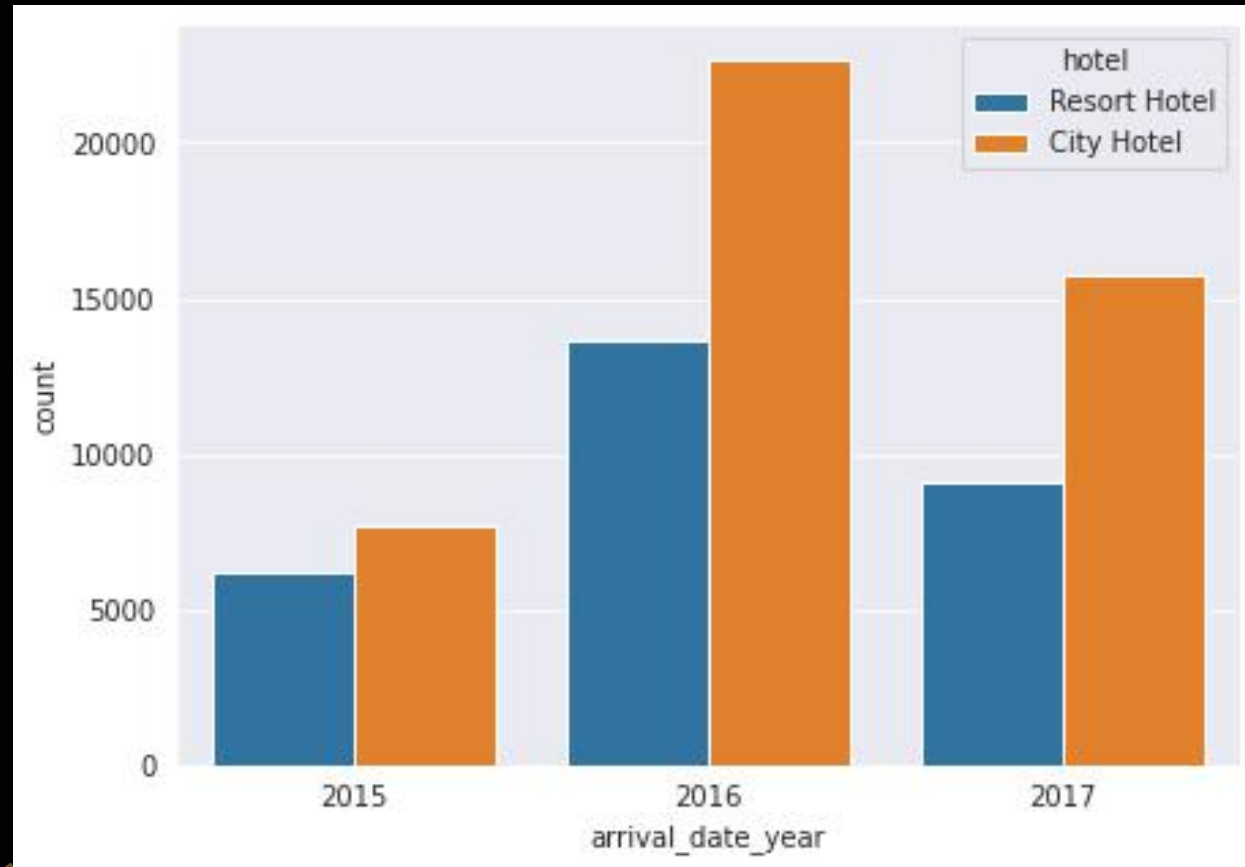
KEY INSIGHTS

After Comparing the Booking with City vs Resort Hotel. We see City Hotel is most successful for booking in 2015 , 2016 & 2017 with the comparision of Resort Hotel.
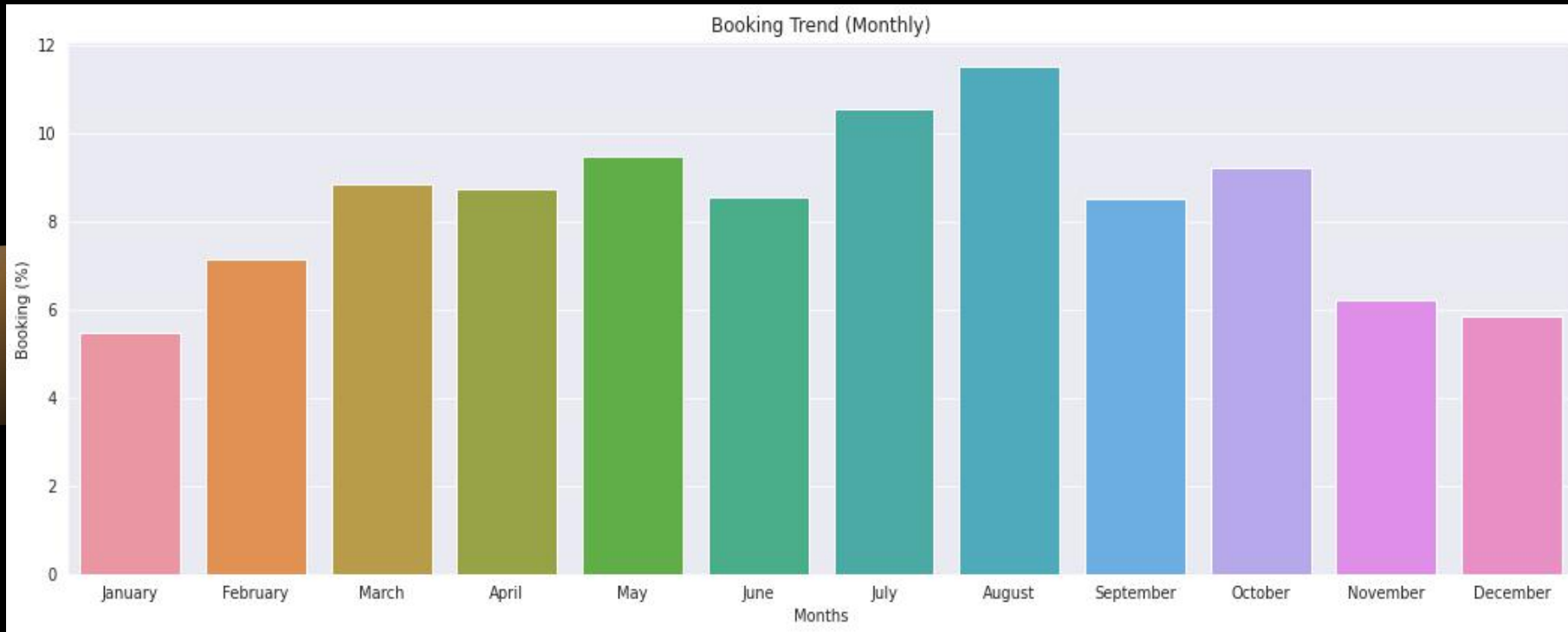
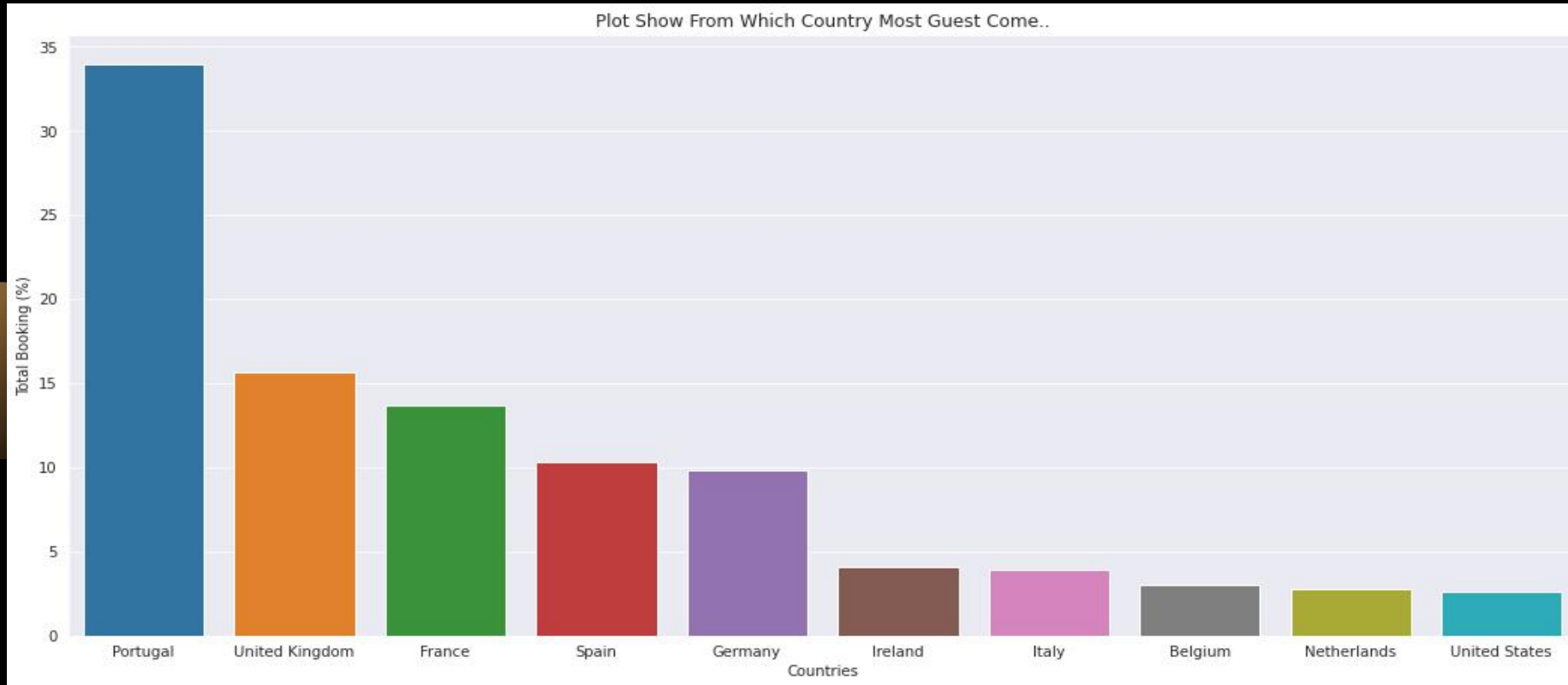# Compare Busy Month for Hotel Between Resort vs City

Booking Trend (Monthly)

KEY INSIGHTS

Here city Hotel is busy month than Resort in May, June, Aug, Sep, oct month and in Jan, Feb, Mar, april, july, nov & dec month busy for resort hotel more than city hotel.
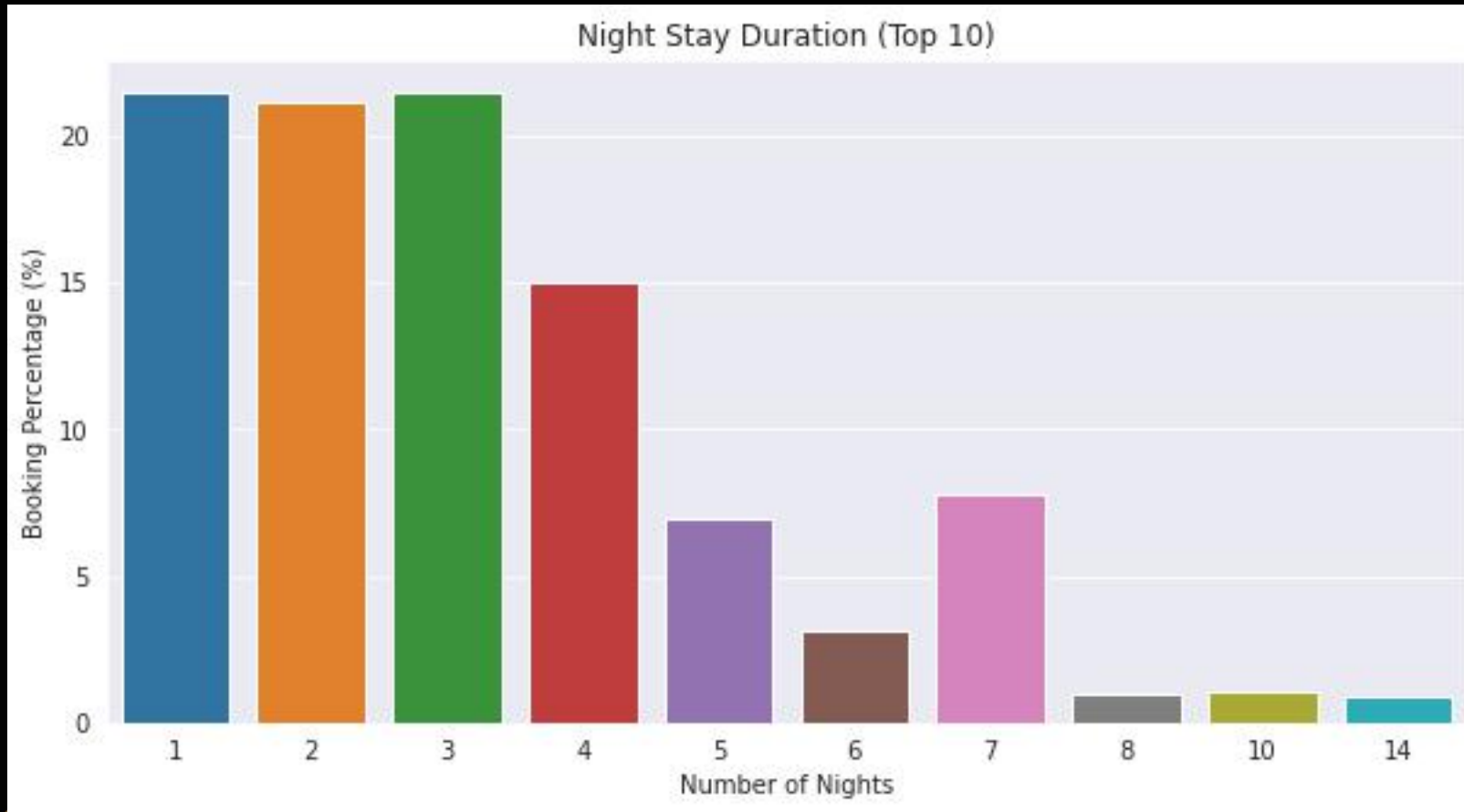
5) From which country most guest come ?

Plot Show From Which Country Most Guest Come..

KEY INSIGHTS

Most Guest come from Portugal. Nearly 34% Guest come from portugal and united kingdom 2nd highest country with near 16% & France is in 3rd position with near 14% of guest.

6) How Long People Stay in the hotel ?

Night Stay Duration (Top 10)

KEY INSIGHTS

In visualization we see most of people choose 3 Night in hotel. Near 22.5% of people choose 3 night & 22% of people choose 1 night & 21% people choose 2 nights.
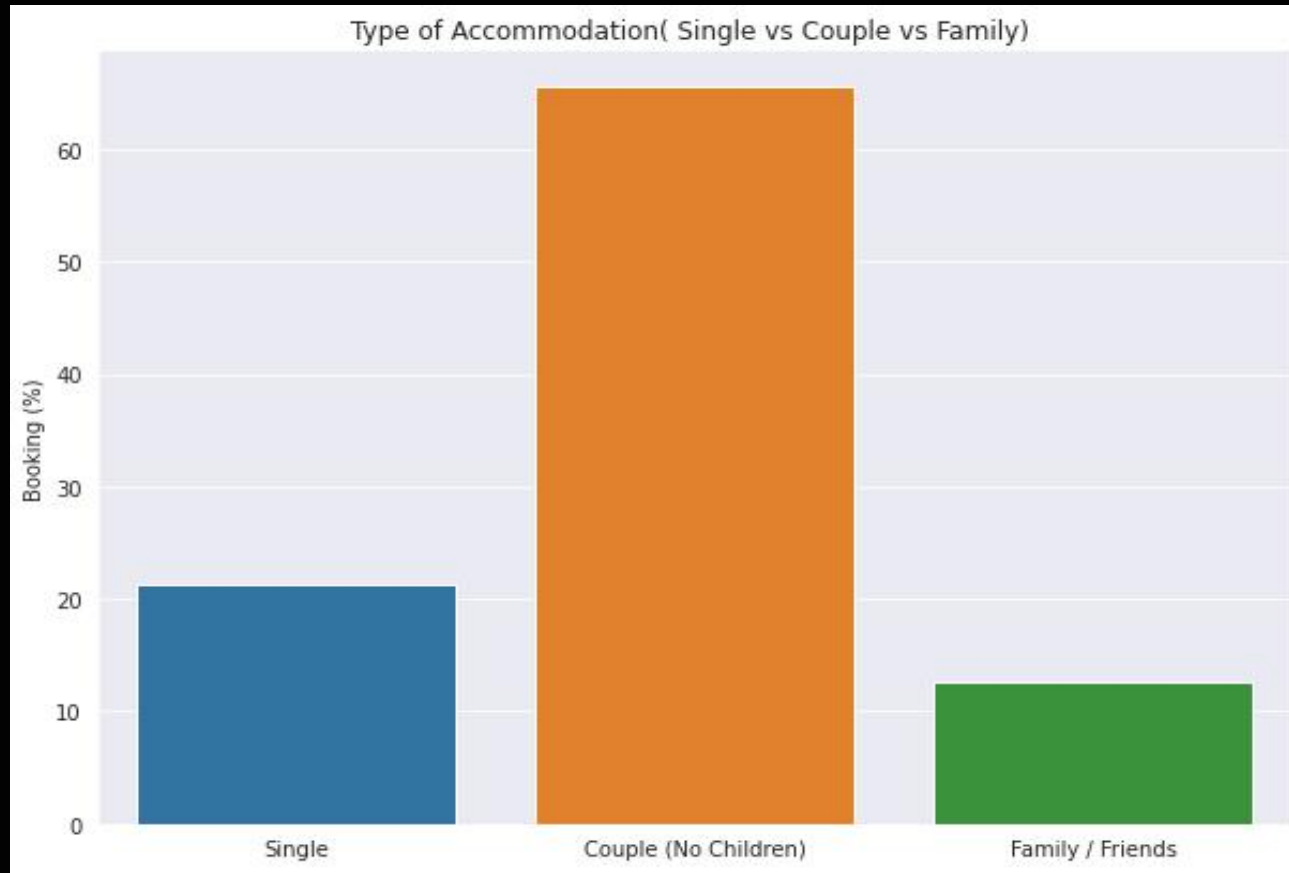
8) How many guests repeated in hotel ?

Pie Chart Show Only 3% Guest is Repeated in Hotel. It means 3755 guest out of 115455 is repeated.

Guest Not Repeated 97%
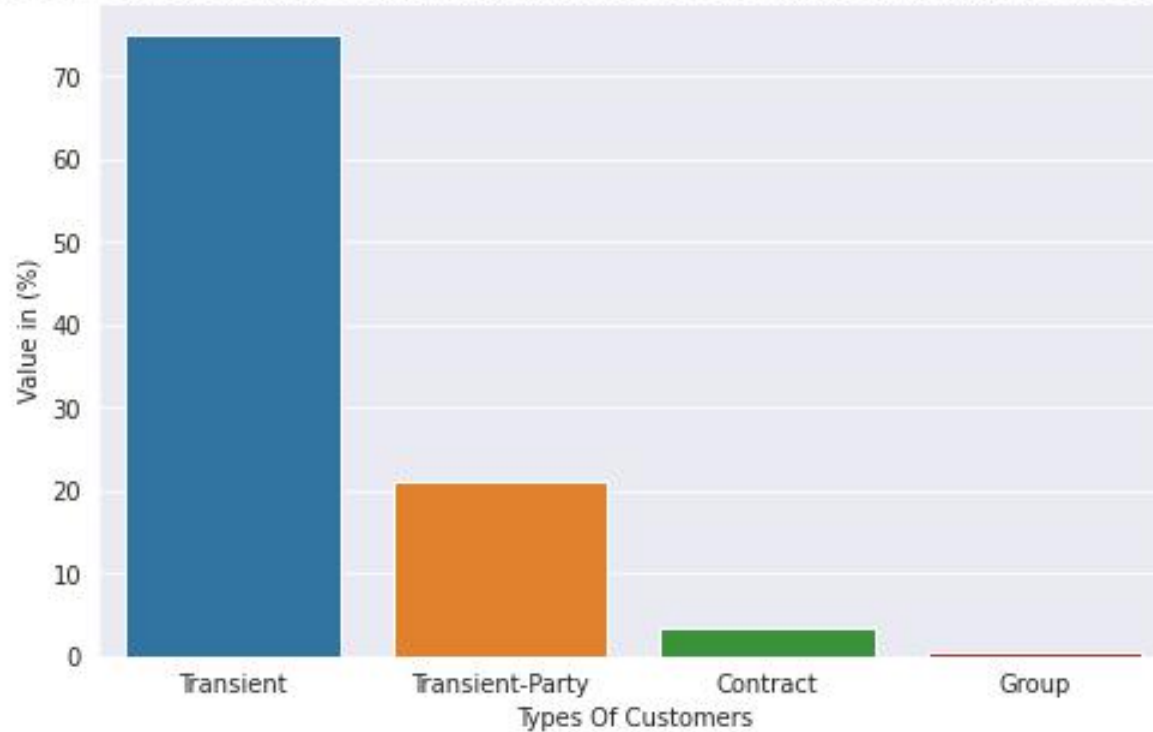
3% Guest is Repeated

KEY INSIGHTS

Here is only 3% of guest repeated in hotel and remaining 97% of guest not repeated.
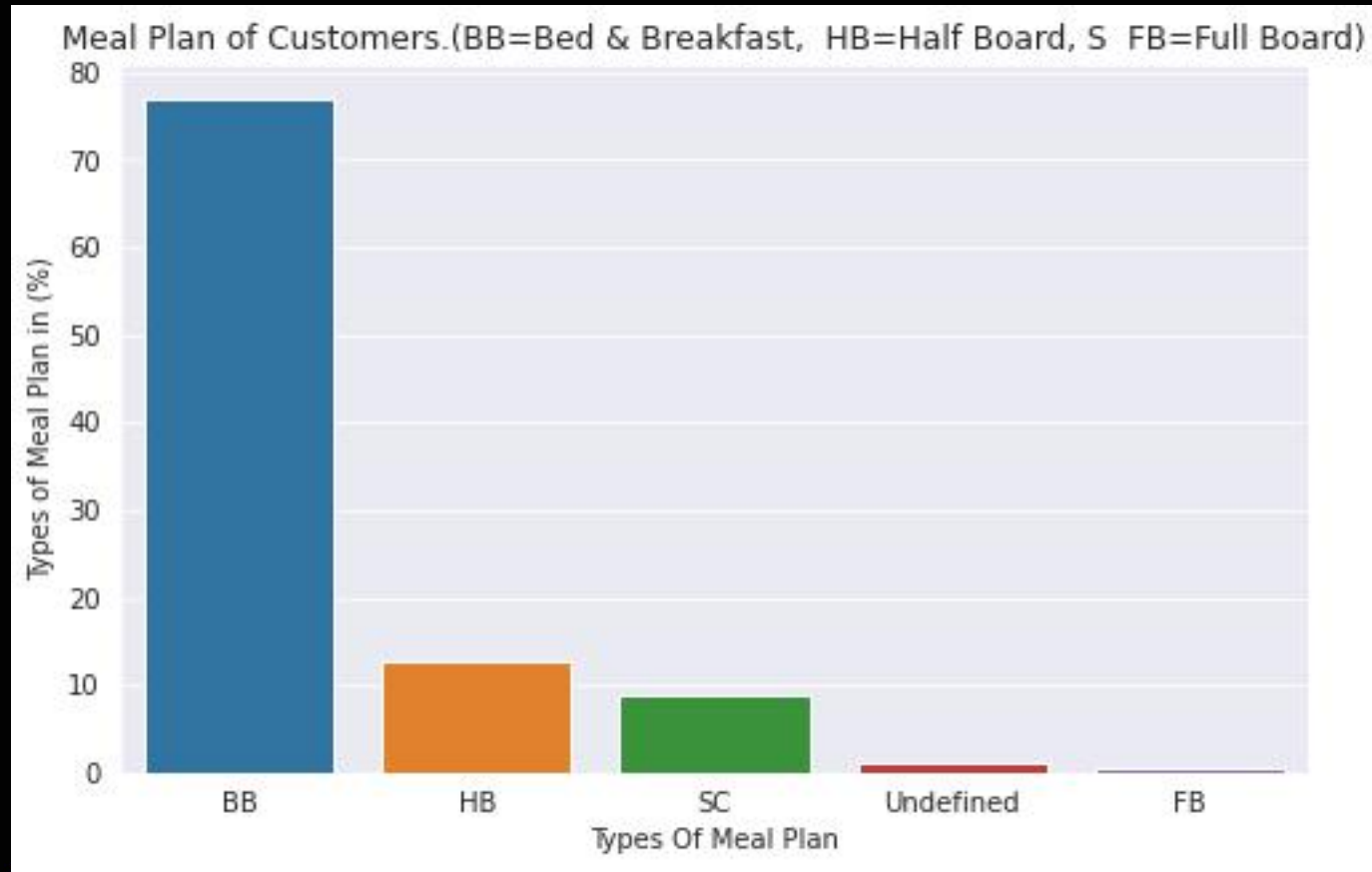
9) Types of Customers ?

Bar Chart Show Types of customers in Hotel.In Which Transient is 75%, Transient-Party is 21%, Contract is 3%, Group is 0%

KEY INSIGHTS

Transient Customer is mostly visited customer more than Transient-party, contract & Group.

10) Meal Plan of Customers ?

Meal Plan of Customers.(BB=Bed & Breakfast, HB=Half Board, S FB=Full Board)
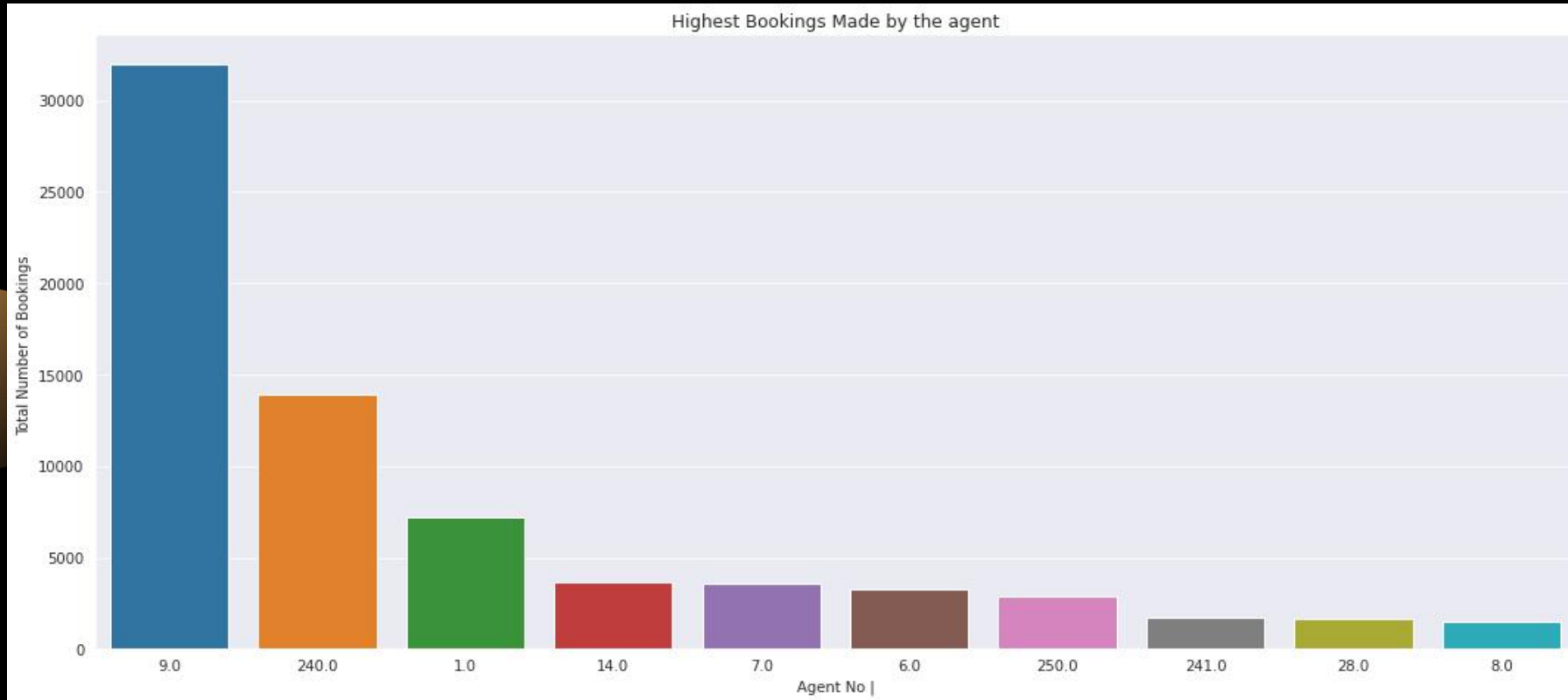
KEY INSIGHTS

BB is mostly demanded type of meal according to guest data . Here is BB means Bed & Breakfast. The ratio of BB is nearly 77% & HB is 12.5% & SC is 9.5%. Remaining is Undefined type of meal.

EDA

11) Which agent makes highest no. of bookings ?

KEY INSIGHTS

Agent no 9 is make most number of booking with above 32000. Agent no 240 is in the 2nd position who has make booking more than 14000.

AImaBetter

EDA

12) Which is the most preferred room type by the customers ?

Chart Shows Most Preferred Room Type By The Customers
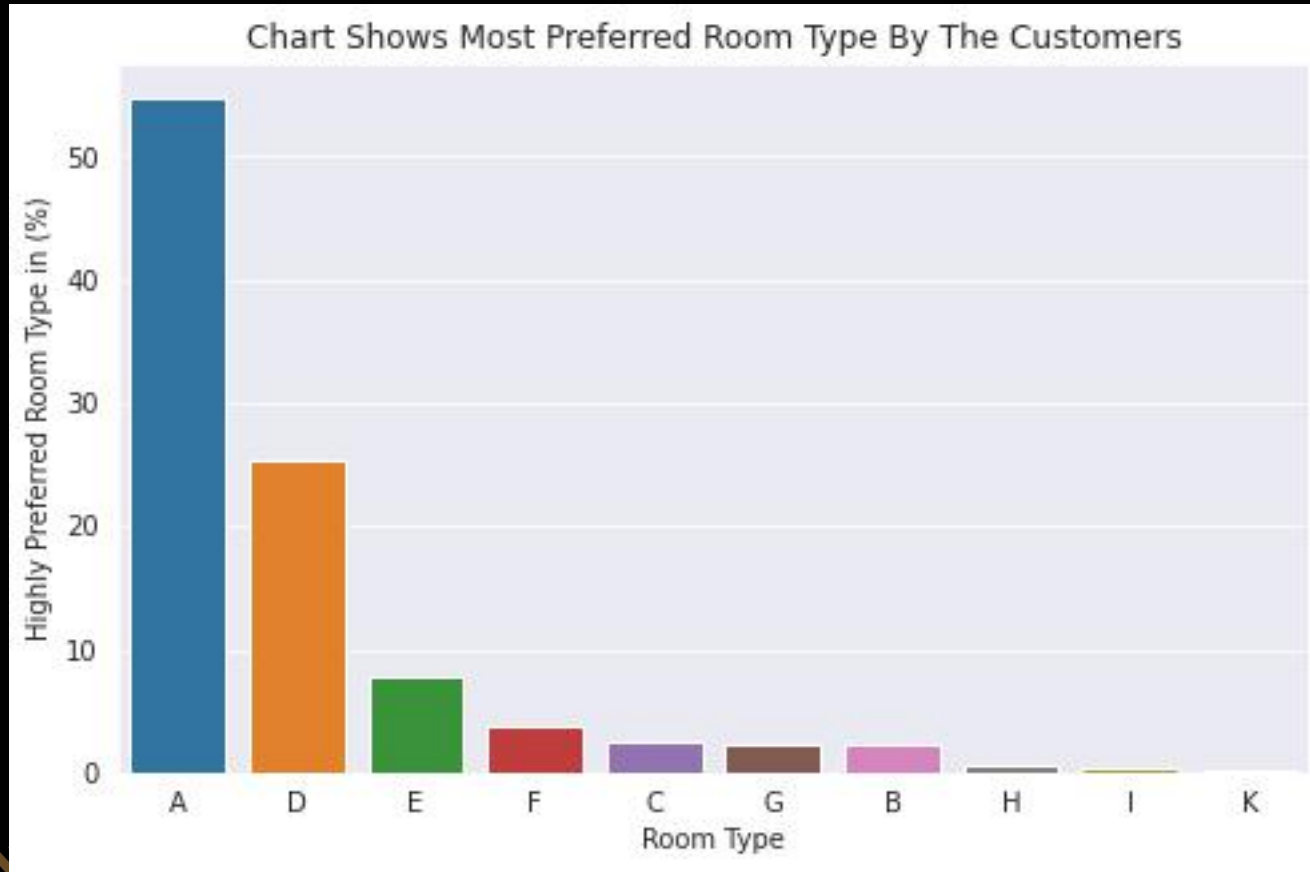
KEY INSIGHTS

'A' type of room most preferred by customers compare of others room types & 'D' type of room 2nd most choice by customer & 'E' is in 3rd position.

EDA

13) Which Hotel type has the highest ADR ?

Avg ADR of Both Hotel type

KEY INSIGHTS

Here is we see in visualization City Hotel is more ADR Than Resort Hotel. City hotel is with nearly 105 of adr & Resort is 85 of adr.
More adr(average daily rate) means more revenue.
so it means city hotels are generating more revenues than the resort hotels.

AImaBetter

07

PART SEVEN

Insights to be found

Insights to be found

1) There are near 37% Booking Cancelled By Customers & remaining 63% of Booking is Safe.

2) There are 61% of Booking done in City Hotel & remaining 39% of Booking done in Resort Hotel.

3) There are in 2015 Booking done by nearly 19% adn in 2016 48% Booking & in 2017 Booking done 33% .
So 2016 is Successful Year For Hotel Booking.
We see City Hotel is most successful for booking in 2015 , 2016 & 2017 with the comparision of Resort Hotel.

4) There are August Month is most busy month for Hotel & July is 2nd most busy month for hotel .
If we compare , city Hotel is busy month than Resort in May, June, Aug, Sep, oct month and in Jan, Feb, Mar
april, july, nov & dec month busy for resort hotel more than city hotel.

5) Most Guest come from Portugal. Nearly 34% Guest come from portugal and united kingdom 2nd highest
country with near 16% & France is in 3rd position with near 14% of guest.

6) In visualization we see most of people choose 3 Night in hotel. Near 22.5% of people choose 3 night & 22% of people choose 1 night & 21% people choose 2 nights. With comparision City hotel is more guest for 1, 2, 3, 4 & 5 nights than the resort hotel. In resort hotel people love to stay with 6, 7, 8, 10 & 14 nights.

7) Couple book hotel more than single & Family/Friends members.
Ratio of couple for booking hotel is near 66% and for single is 22% and Family/Friends with 12% .

8) only 3% of guest repeated in hotel and remaining 97% of guest not repeated.

9) Transient Customer is mostly visited customer more than Transient-party , contract & Group.

10) BB is mostly demanded type of meal according to guest data . Here is BB means Bed & Breakfast.
The ratio of BB is nearly 77% & HB is 12.5% & SC is 9.5%. Remaining is Undefined type of meal.

11) Agent no 9 is make most number of booking with above 32000. Agent no 240 is in the 2nd position who has make booking more than 14000.

12) 'A' type of room most preferred by customers compare of others room types & 'D' type of room 2nd most choice by customer & ' E ' is in 3rd position.

13) In visualization City Hotel is more ADR Than Resort Hotel. City hotel is with nearly 105 of adr & Resort is 85 of adr. More adr(average daily rate) means more revenue. so it means city hotels are generating more revenues than the resort hotels.