# Capstone Project-1

EDA on Hotel Booking Analysis

Team Project:

- ❖ Kishor Kunal
- ❖ Praveen Sikarwar
- ❖ Punam Nagrale
- ❖ Boncheruvu Teja
- ❖ Rishi Chaturvedi

Problem Statement

➤ For this project we will be analysing Hotel Booking data. This data set contains booking information for a city hotel and a resort hotel, and includes information such as when booking was made, length of stay, the number of adults, children or babies and number of available parking spaces.

➤ Hotel industry is a very volatile industry and the bookings depends on above factors and many more.

➤ The main objective behind this project is to explore and analyse data to discover important factor that govern the bookings and give insights to hotel management, which can perform various campaigns to boost the business and performance.
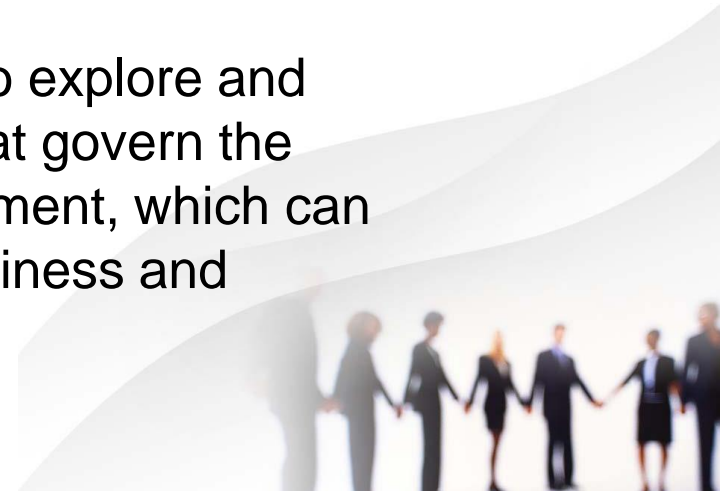
So we will divide our work flow into following 3 steps.

Data Collection and Understanding → Data Cleaning and Manipulation → Exploratory Data Analysis(EDA)

❖<u>Data Collection and Understanding</u>

❑After collecting data it's important to understand your data. So we had hotel booking analysis data which had 119390 rows and 32 columns. So lets understand this 32 columns.

<u>Data Description</u>:-

➢hotel: Name of hotel (City or Resort)
➢ is_canceled: Whether the booking is canceled or not (0 for no canceled and 1 for canceled)
➢lead_time: time (in days) between booking transaction and actual arrival
➢arrival_date_year: Year of arrival
➢arrival_date_month: month of arrival
➢arrival_date_week_number: week number of arrival date.
➢arrival_date_day_of_month: Day of month of arrival date
➢stays_in_weekend_nights: No. of weekend nights spent in a hotel
➢Stays_in_week_nights: No. of weeknights spent in a hotel
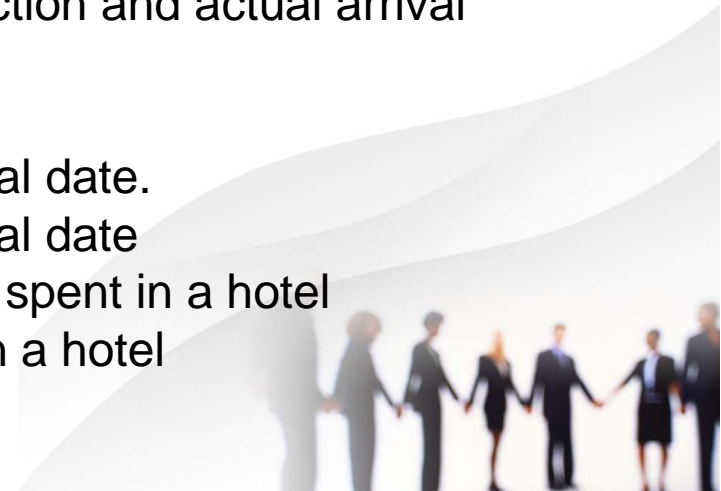
- ➤ adults: No. of adults in single booking record.
- ➤children: No. of children in single booking record.
- ➤ babies: No. of babies in single booking record
- ➤meal: Type of meal chosen
- ➤country: Country of origin of customers (as mentioned by them)
- ➤market_segment : What segment via booking was made and for what purpose.
- ➤distribution_channel: Via which medium booking was made.
- ➤is_repeated_guest: Whether the customer has made any booking before(0 for No and 1 for Yes)
- ➤previous_cancellations: No. of previous canceled bookings.
- ➤previous_bookings_not_canceled: No. of previous non-canceled bookings.
- ➤reserved_room_type: Room type reserved by a customer.
- ➤assigned_room_type: Room type assigned to the customer.
- ➤booking_changes: No. of booking changes done by customers
- ➤deposit_type: Type of deposit at the time of making a booking (No deposit/ Refundable/ No refund)
- ➤agent: Id of agent for booking
- ➤company: Id of the company making a booking
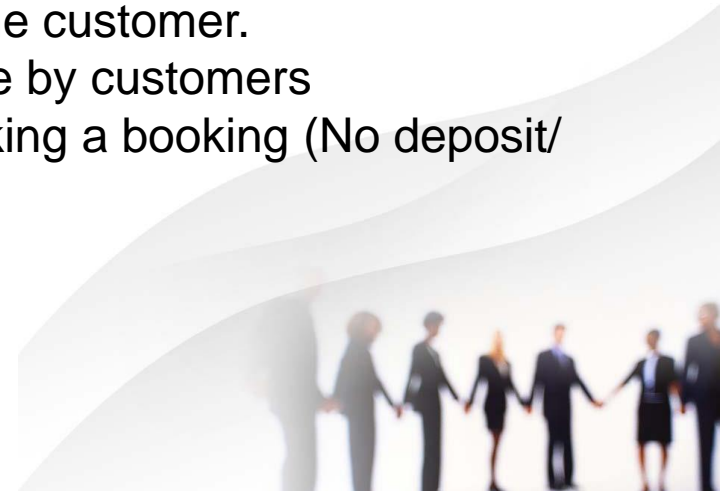
➢days_in_waiting_list: No. of days on waiting list.

➢customer_type: Type of customer(Transient, Group, etc.)

➢ adr: Average Daily rate

➢required_car_parking_spaces: No. of car parking asked in booking

➢ total_of_special_requests: total no. of special request.

➢ reservation_status: Whether a customer has checked out or canceled,or not showed

➢ reservation_status_date: Date of making reservation status

❖ Data Cleaning & Manipulation:

1.We make a copy of Data copy as new DataFrame, So that the changes made in the copied data will not reflect in original data

2.And then remove duplicate out of this

➤ After collection of Data, we got all the record about Hotel Type, Meal type ..etc.

```
[7] Hotel_df.shape

    (119390, 32)
```

```
[8] Hotel_df[Hotel_df.duplicated()].shape

    (31994, 32)
```

```
[10] Hotel_df['hotel'].unique()

    array(['Resort Hotel', 'City Hotel'], dtype=object)
```

```
[13] Hotel_df['arrival_date_year'].unique()

    array([2015, 2016, 2017])
```

```
[14] Hotel_df['meal'].unique()

    array(['BB', 'FB', 'HB', 'SC', 'Undefined'], dtype=object)
```

Data cleaning

➤ Data cleaning helps to remove the redundant and useless data from the data set.

Data Cleaning

Dollowing steps are perfomed as a part of data cleaning:

1. Remove duplicate rows
2. Handling missing values

```
[ ]  # How many Duplicate Rows are there
     Hotel_df[Hotel_df.duplicated()].shape

     (31994, 32)
```

```
[ ]  # Drop Duplicate Rows
     Hotel_df.drop_duplicates(inplace = True)
```

```
[ ]  # DataFrame after Drop Duplicate Rows
     Hotel_df.shape

     (87396, 32)
```

Find out Null Values in Data

➢ After creating a copy, we performed data cleaning by deleting the duplicate and null values from the copy data set.

```
df.isnull().sum().sort_values(ascending = False)[:6]
```

```
company              82137
agent                12193
country                452
children                 4
reserved_room_type       0
assigned_room_type       0
dtype: int64
```

```
# Columns having missing values.
Hotel_df1.isnull().sum().sort_values(ascending = False)[:6]
```

```
hotel                        0
is_canceled                  0
reservation_status           0
total_of_special_requests    0
required_car_parking_spaces  0
adr                          0
dtype: int64
```

❖ Exploratory Data Analysis (EDA) :

➢ <u>EDA</u> will be divided into following 2 analysis.

1.Univariate analysis:->Univariate analysis explores each variable in a data set, separately. Like which hotel is mostly preferred by visitors.

2.Bivariate analysis:->It **involves the analysis of two variables**. Like Which type of room generating more ADR(**average daily rate).**
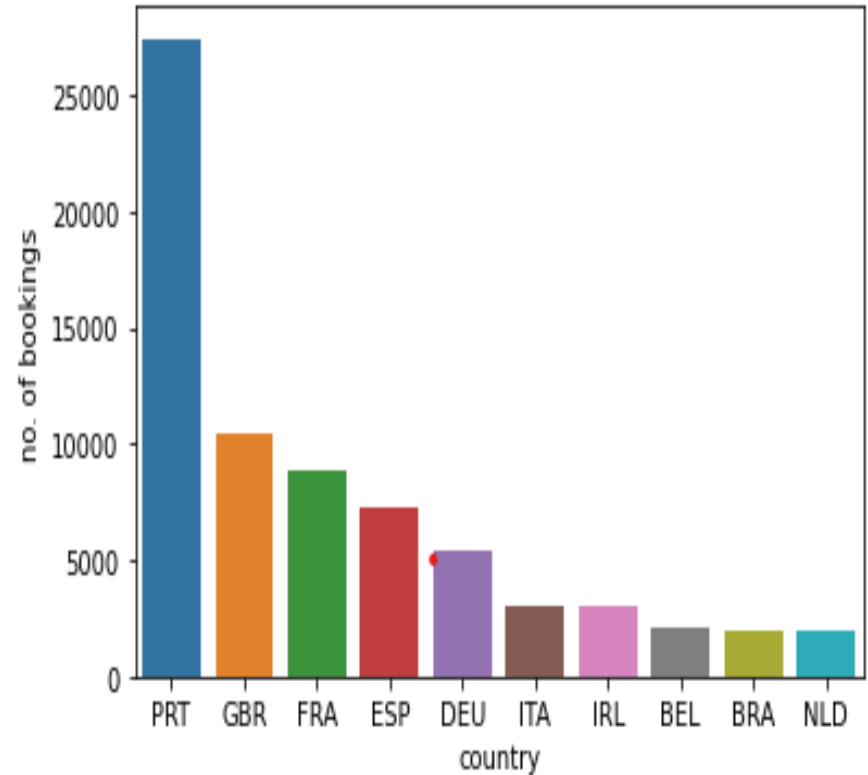
3.Multivariate analysis:->It **involves the analysis of two or variables**.

Country with more no. of guests

Conclusions:

➢ Maximum number of guests were from Portugal. i.e. more than 25000 guests.

➢ After Portugal, GBR(Great Brittan),France and Spain are the countries from where most of the guests came.

Conclusion:

➢ City hotel has the highest ADR. That means city hotels are generating more revenues than the resort hotels. More the ADR more is the revenue.
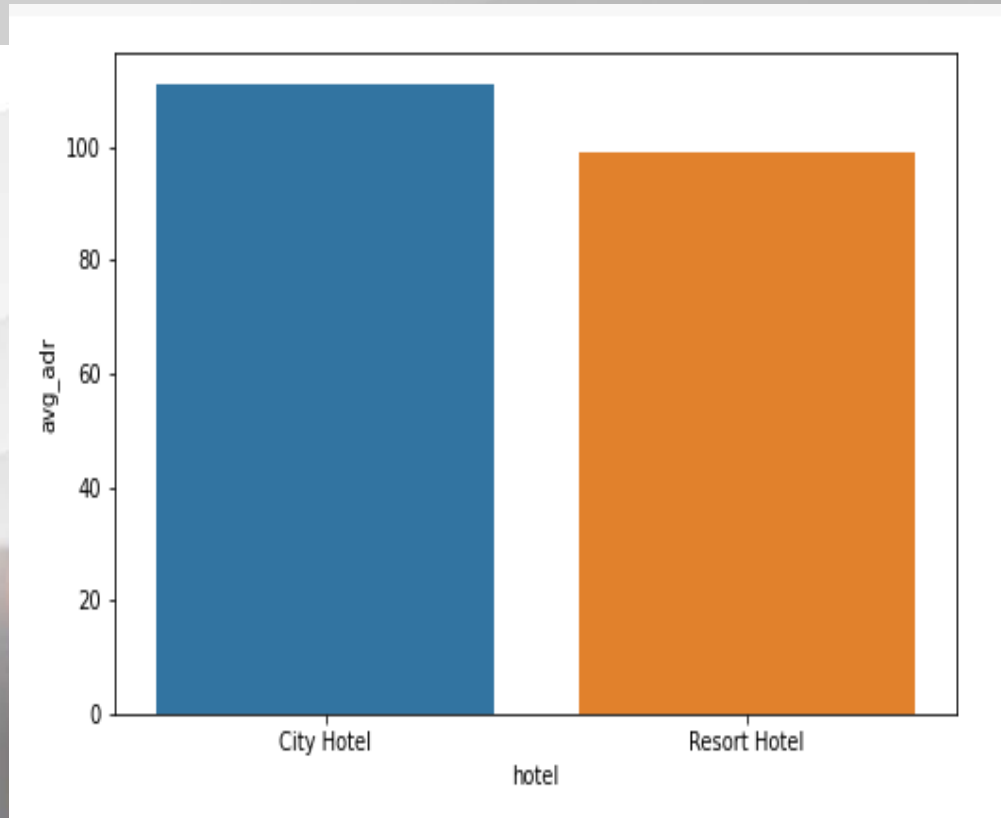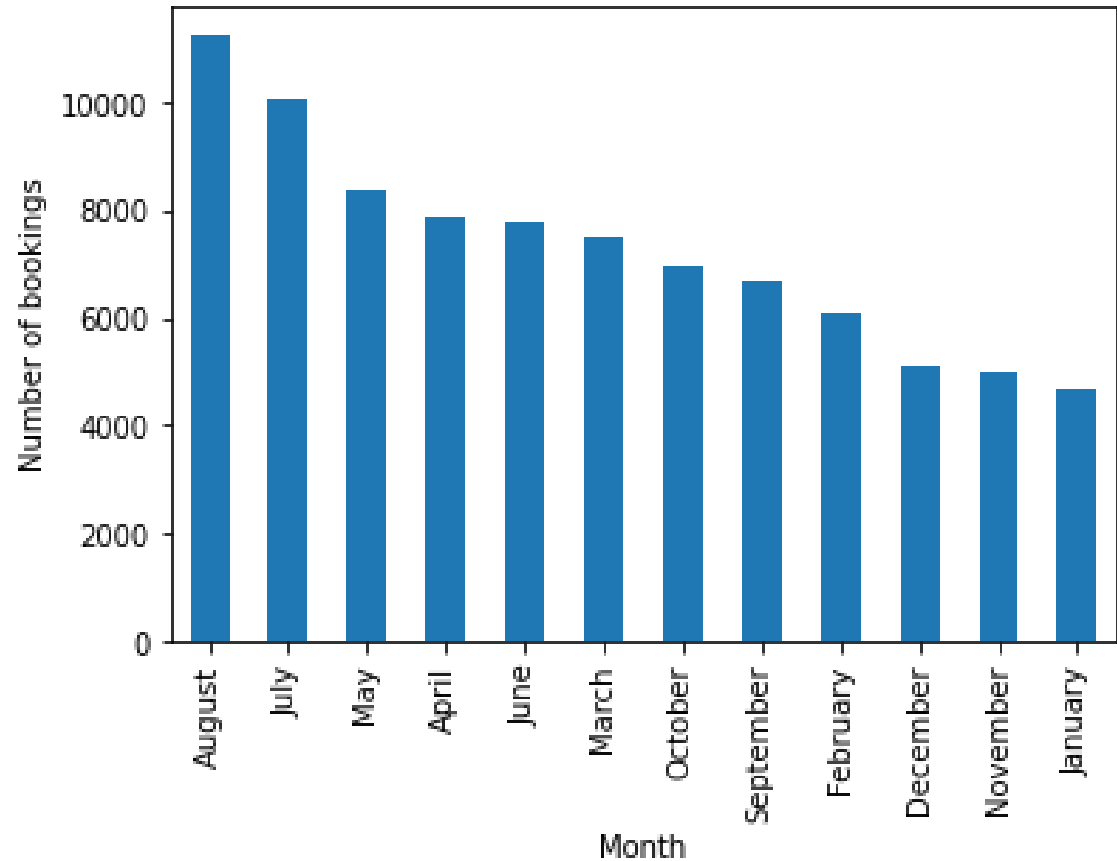
Month wise bookings of each hotel & busiest month

Conclusion:

➢ The month wise booking are as below:

<mark>August</mark>	<mark>11257</mark>
July	10057
May	8355
April	7908
June	7765
March	7513
October	6934
September	6690
February	6098
December	5131
November	4995
January	4693

➢ The Busiest Month is August, which have highest Bookings.

Figure 3.1: Monthly bookings

Customers asking for car parking space

Conclusion:

➢ More customers from resort hotels are asking for carparking space.
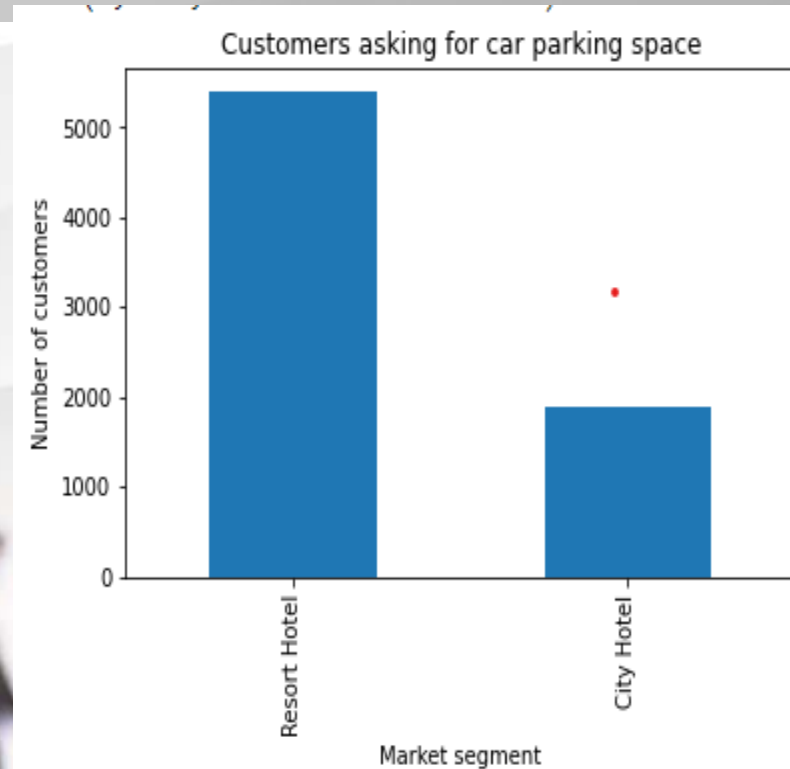
Hotel with maximum booking in weekend

Conclusion:

➤ City hotels have more bookings in weekend nights.

Weekend bookings in each hotel

Most preferred meal type

Conclusion:

➤ BB(Bed & Breakfast) is the most preferred type of meal by the guests.

➤ Full Board i.e. FB is least preferred.
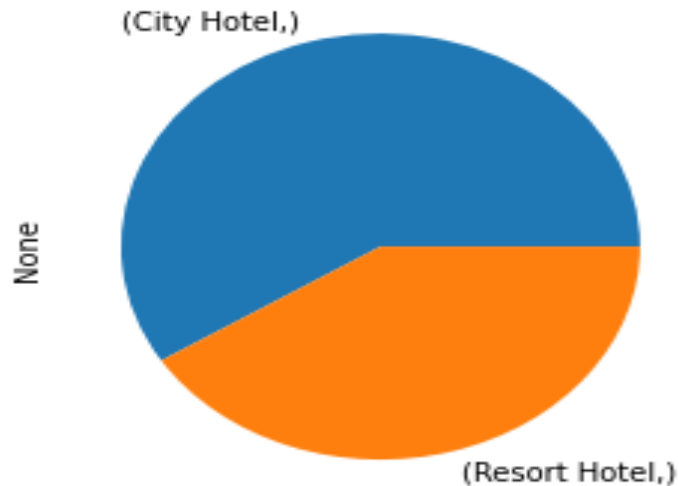
➤ HB (Half Board) and SC(Self Catering) are equally preferred almost.

Figure 4.1 Types of meal and count
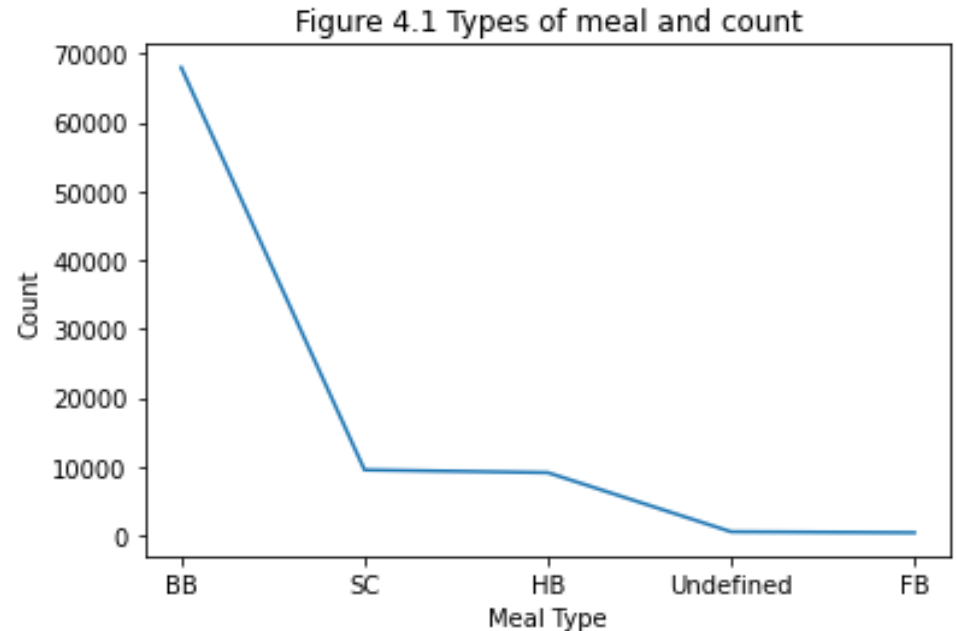
Total bookings in each hotel type

Conclusion:

➢ City Hotels have more number of confirmed and are receiving more number of guests than Resort Hotels.

Confirmed bookings in each hotel

Comparing cancellations in each hotel type

Conclusion:

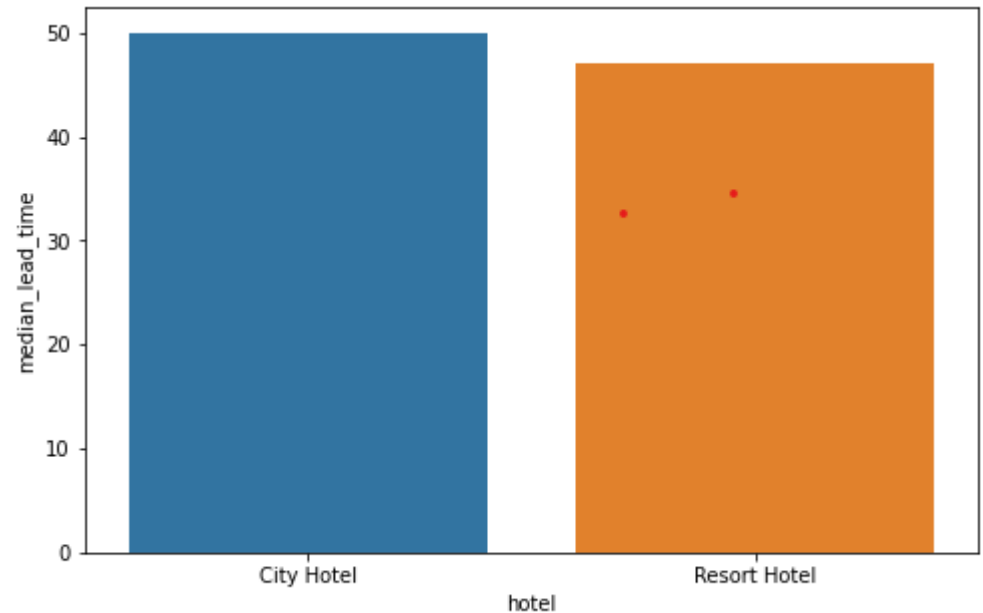➢ More customers are cancelling their bookings in the city hotels as compared to resort hotels.

Cancelled bookings in each hotel

(City Hotel,)

None

(Resort Hotel,)

Hotel with higher lead time

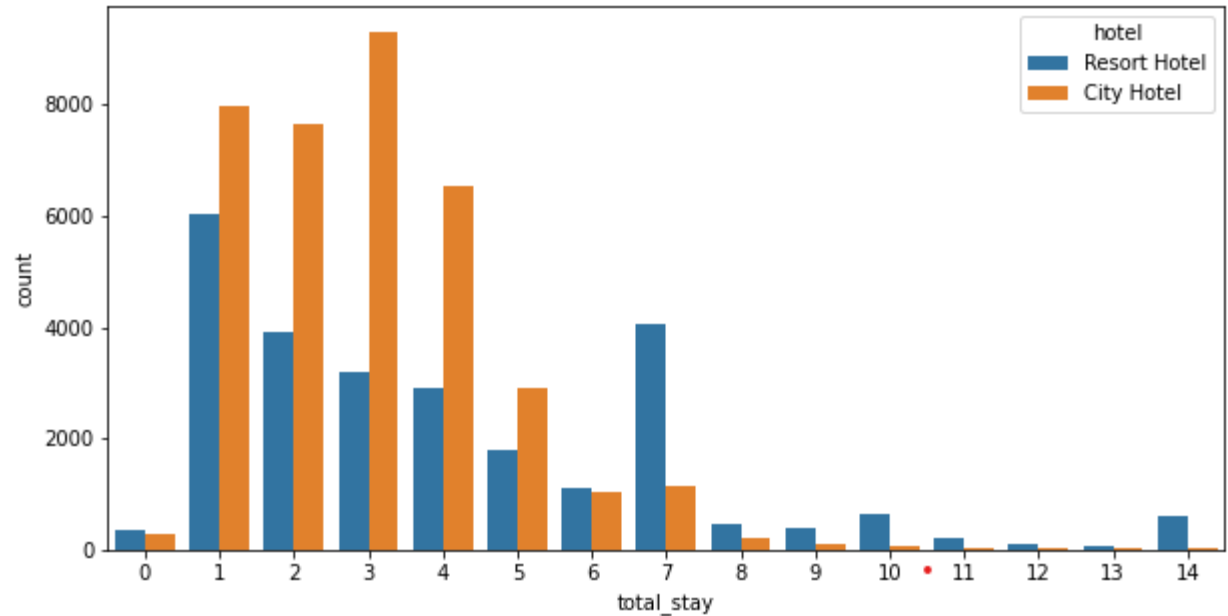Lead time:->The time gap between the date of booking and date of arrival.

Conclusion:

➢ City Hotels have higher lead time than resort hotels.
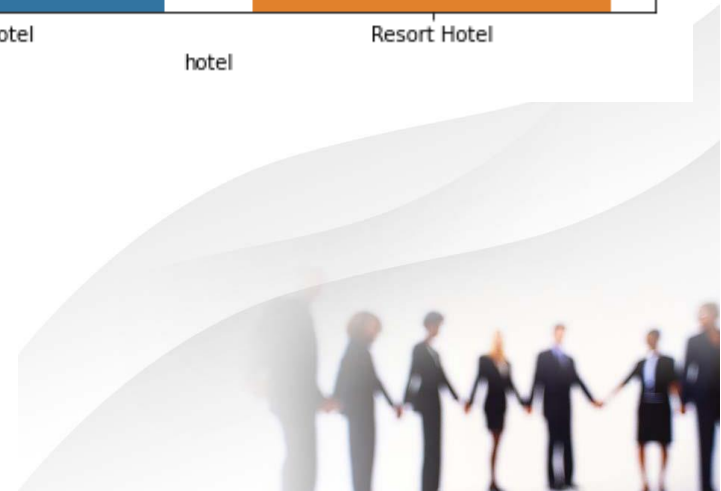
How many days people generally preferred stay in each hotel

Conclusion:

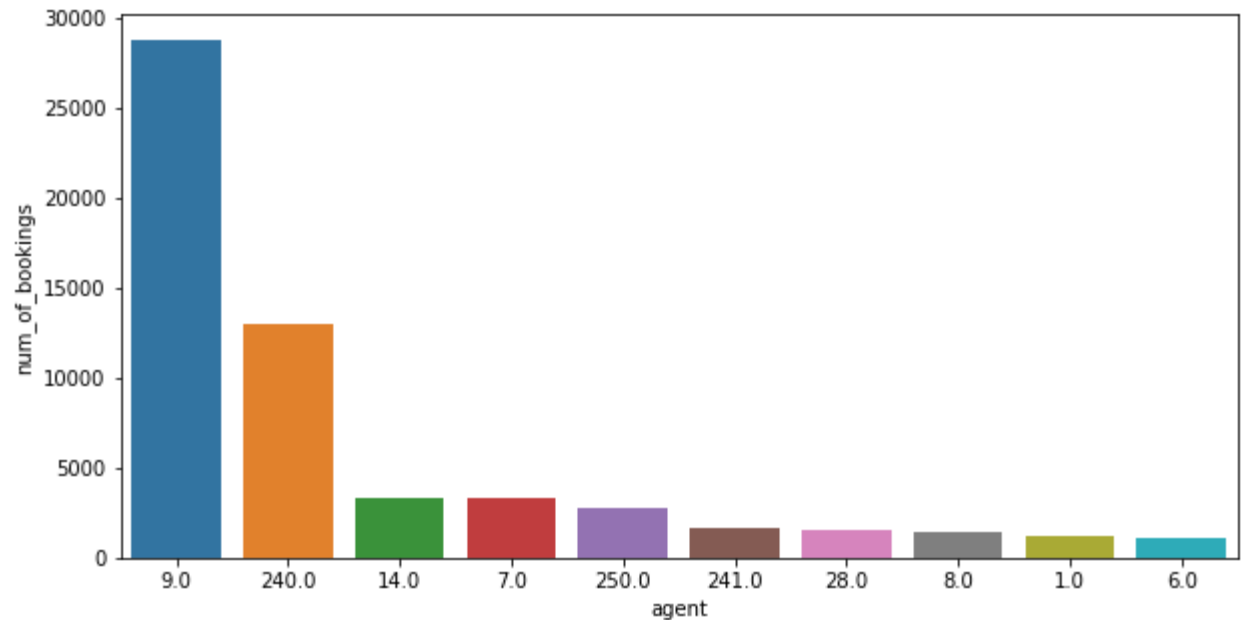➤ Preferred stay length is less than '7' days. In both Hotel
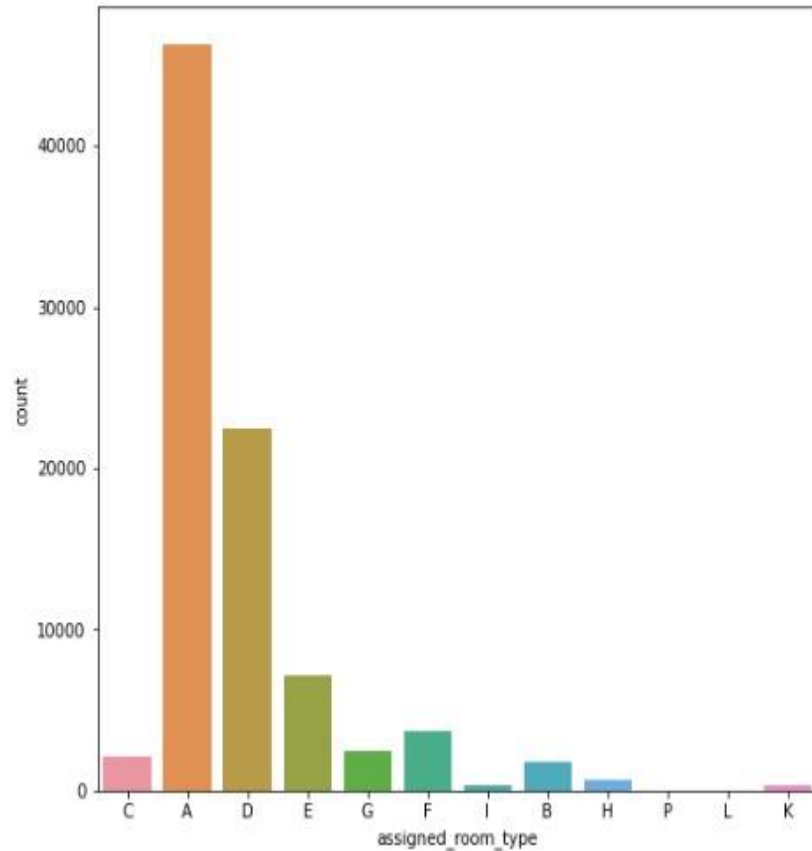
Agent wise bookings
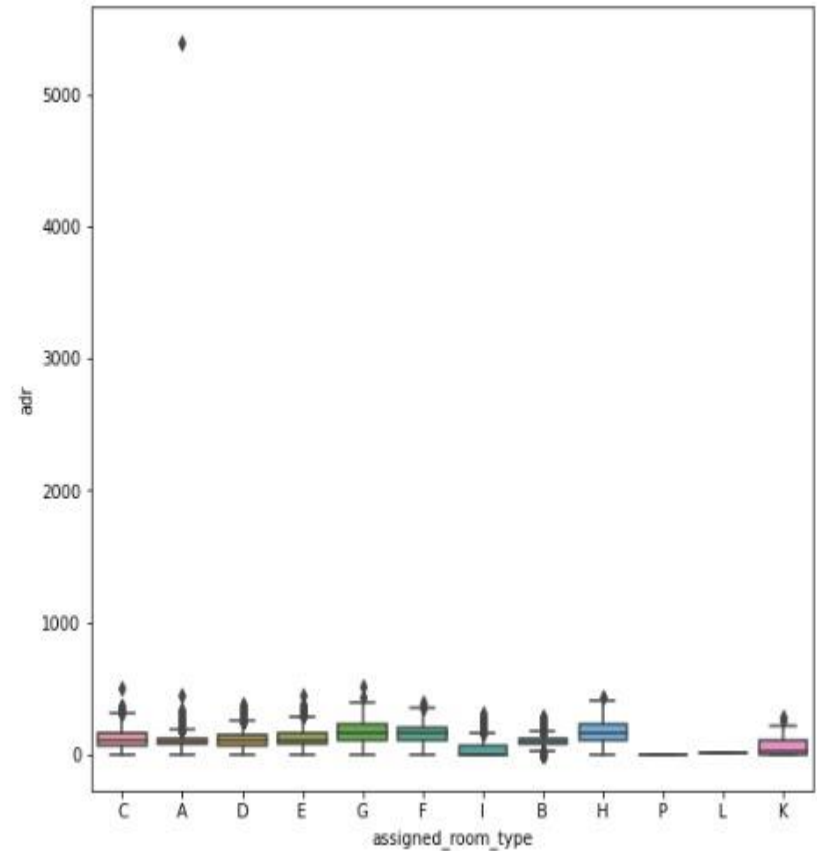
Conclusion:

➢ Agent with ID'9.0' is making more bookings.

`<matplotlib.axes._subplots.AxesSubplot at 0x7fd6fc8323a0>`

Most Preferred room type

More revenue generating room type

Conclusion:
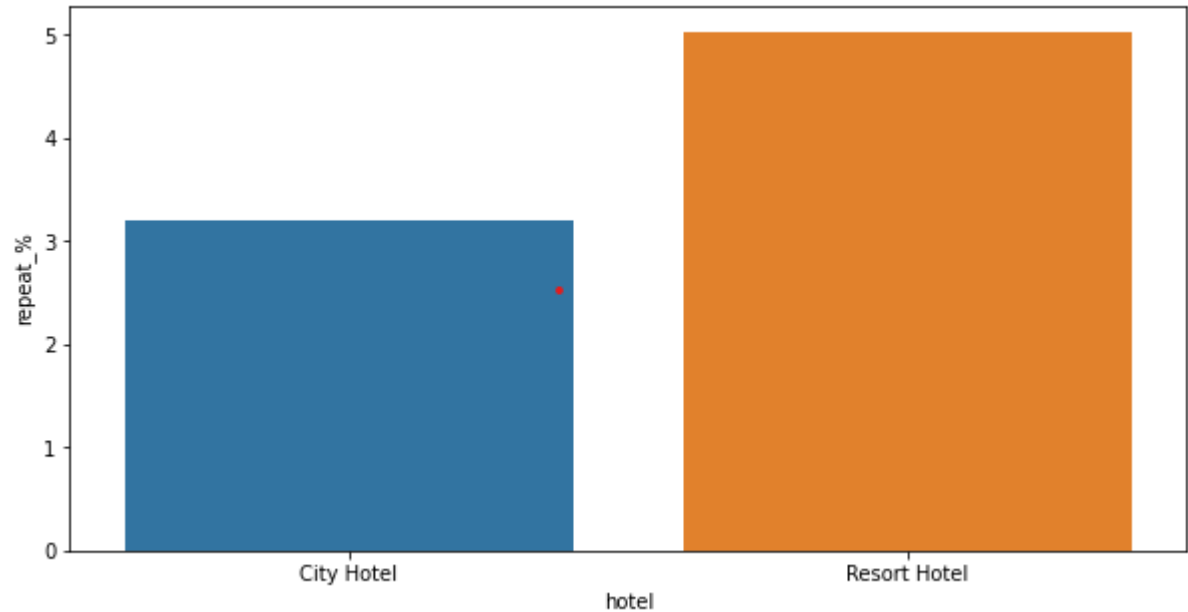
➢ Most preferred room type is 'A'

Conclusion:

➢ Most revenue generating room type is 'G'

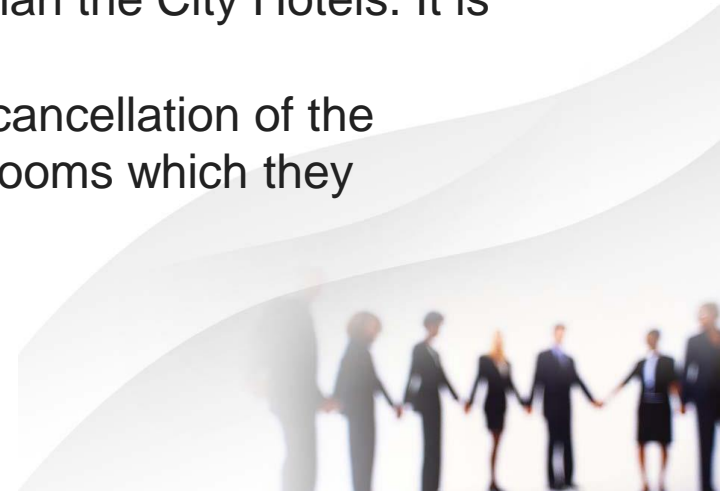# Hotel having more repeated customer

Conclusion:

➤ Resort hotels have more repeated customers.

Conclusion:

1. We observe that the City Hotel is mostly preferred by the visitors i.e. about 61.1% in comparison to Resort Hotel(38.9%).Thus City Hotel has maximum booking.
2. The most prefer room type is A
3. We observe that the Agent having ID 9.0 made the most booking
4. we clearly se that Meal Type BB is most preferred
5. **August** is the Busiest month and July and August months had the most Bookings. Summer vacation can be the reason for the bookings.
6. TA/TO' is mostly(79.1%) used for booking hotels.
7. The City Hotel faces maximum hotel cancelation
8. Resort Hotel has slightly more repeated guests than the City Hotels. It is almost similar for both hotels.
9. Its is clear that there is no much(2.5%) effect on cancellation of the bookings even if the guests are not assigned with rooms which they reserved during booking

Thank You!