**Hotel Booking Analysis**

1. **Ms.Madhuri Bonela**
2. **Mr.Pratik Ghodke**

**Data science trainees,**

**AlmaBetter, Bangalore**

**Abstract:**

Hotel enterprise is a very unstable enterprise and the bookings rely on range of elements such as kind of hotels, seasonality, days of week and many more. This makes inspecting the patterns accessible in the previous records greater vital to assist the motels layout better. Using the historic data, motels can function a variety of campaigns to enhance the business. We can use the patterns to predict the future bookings the use of time collection or selection trees.We will be tackling this problem statement in three stages:

we can analyse some key metrics for hotel bookings like :

1.The number of cancellations

2.Number of bookings on weekday vs weekends

3.Type of hotels booked by customers like City hotels or resort hotels

4.Type of platform most of the customers used for booking a hotel

5. Checking if the booking was from repeated guest or not

6. Type of booking acquired by most of the customers

7. Checking the arrival count for every month and analysing the patterns

Therefore by analysing every data and applying them in future analysis makes profitable business and satisfies every customer from the service provided .

**1. Problem Statement**

These millions of listings generate a lot of dataProblem Statement

Have you ever wondered when the best time of year to book a hotel room is? Or the optimal length of stay in order to get the best daily rate? What if you wanted to predict whether or not a hotel was likely to receive a disproportionately high number of special requests? This hotel booking dataset can help you explore those questions!

1.The type of hotel preferred by most customers.

2.The number of cancelled bookings.

3.The average daily rate (ADR) and it's calculated by dividing the sum of all lodging transactions by the total number of staying nights.

4. Checking the arrival date by months.

5.Market segment.

6.year of arrival date.

7. Customer Type .

8.Distribution Channel.

9.The total number of repeated guests.

10.which are the months of highest and least occupation.

11.which is the most reserved room type .

We are provided with a hotel bookings dataset.

Our main objective is perform EDA on the given dataset and draw useful conclusions about general trends in hotel bookings and how factors governing hotel bookings interact with each other.

**2. Introduction**

Hotel booking analysis is an tourist agent which offers the facilities for reserving hotels, Events, Travels, houses, vacation apartments and different accommodations for customers.

They operate their enterprise in KSA . Any clients want to e bookhotel, or rental they want to go to their workplace which for checking the availability as properly as negotiation.

Recently they have decided to exchange their reserving sample from guide to web primarily based system. It will assist them to manipulate the clients.

**3. What is Exploratory Data Analysis (EDA)?**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

It is a good practice to understand the data first and try to gather as many insights from it. EDA is all about making sense of data in hand, before getting them dirty with it.

.**4. Exploratory Data Analysis (EDA)**

The following step-by-step example shows how to perform exploratory data analysis in Excel.

Step 1: Create the Dataset.

Step 2: Summarize the Data.

Step 3: Identify Missing Values.

Step 4: Visualize the Data.

Importing Library

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

**4.1.1 import numpy as np**

NumPy can be used **to perform a wide variety of mathematical operations on arrays**. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.

**4.1.2 import pandas as pd**

Import = “Bring this functionality or library to my python script”

* Pandas = The library you want to import, in this case, it's pandas.
* As = The python nomenclature for creating as alias. ...
* pd = The standard short name for referencing pandas.

Pandas stands for “**Python Data Analysis Library** ”. It is used for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

**4.1.3 matplotlib.pyplot**

pyplot is **a plotting library used for 2D graphics in python programming language**. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

**Following steps were followed:**

1. Define the x-axis and corresponding y-axis values as lists.
2. Plot them on canvas using . plot() function.
3. Give a name to x-axis and y-axis using . xlabel() and . ylabel() functions.
4. Give a title to your plot using . title() function.
5. Finally, to view your plot, we use . show() function.

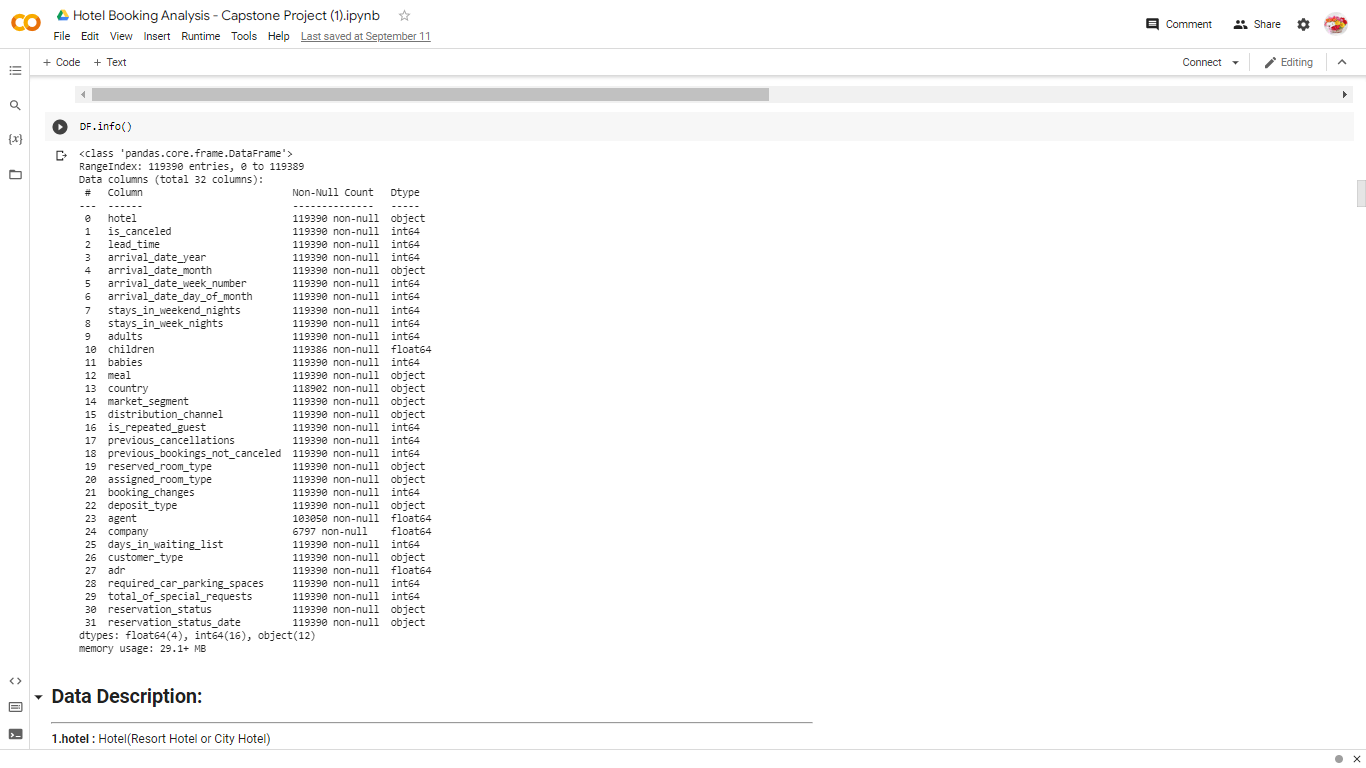
#### **4.2 Mounting Google Drive**

from google.colab import drive

drive.mount('/content/drive/')

#### **4.3 Importing Dataset From Drive**

df = pd.read\_csv(path + 'Airbnb NYC 2019.csv')



We can observed the, there is non null values in data set and

4.4 **Feature descriptions**

**1.hotel :** Hotel(Resort Hotel or City Hotel)

**2.is\_canceled :**Value indicating if the booking was canceled (1) or not (0)

**3.lead\_time :**Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

**4.arrival\_date\_year :** Year of arrival date

**5.arrival\_date\_month :** Month of arrival date

**6.arrival\_date\_week\_number :**Week number of year for arrival date

**7.arrival\_date\_day\_of\_month :** Day of arrival date

**8.stays\_in\_weekend\_nights :** Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

**9.stays\_in\_week\_nights :** Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

**10.adults :** Number of adults

**11.children :**Number of children

**12.babies :** Number of babies

**13.meal :**Type of meal booked. Categories are presented in standard hospitality meal packages:

**14.country :** Country of origin.`

**15.market\_segment :** Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”

**16.Distribution\_channel :** Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”

**17.is\_repeated\_guest :** Value indicating if the booking name was from a repeated guest (1) or not (0)

**18.previous\_cancellations :** Number of previous bookings that were cancelled by the customer prior to the current booking

**19.previous\_bookings\_not\_canceled :** Number of previous bookings not cancelled by the customer prior to the current booking

**20.reserved\_room\_type :** Code of room type reserved. Code is presented instead of designation for anonymity reasons.

**21.assigned\_room\_type :** Code for the type of room assigned to the booking.

**22.booking\_changes :** Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

**23.deposit\_type :** Indication on if the customer made a deposit to guarantee the booking.

**24.agent :** ID of the travel agency that made the booking

**25.company :** ID of the company/entity that made the booking or responsible for paying the booking.

**26.days\_in\_waiting\_list :** Number of days the booking was in the waiting list before it was confirmed to the customer

**27.customer\_type :** Type of booking, assuming one of four categories

**28.adr :** Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights

**29.required\_car\_parking\_spaces :** Number of car parking spaces required by the customer

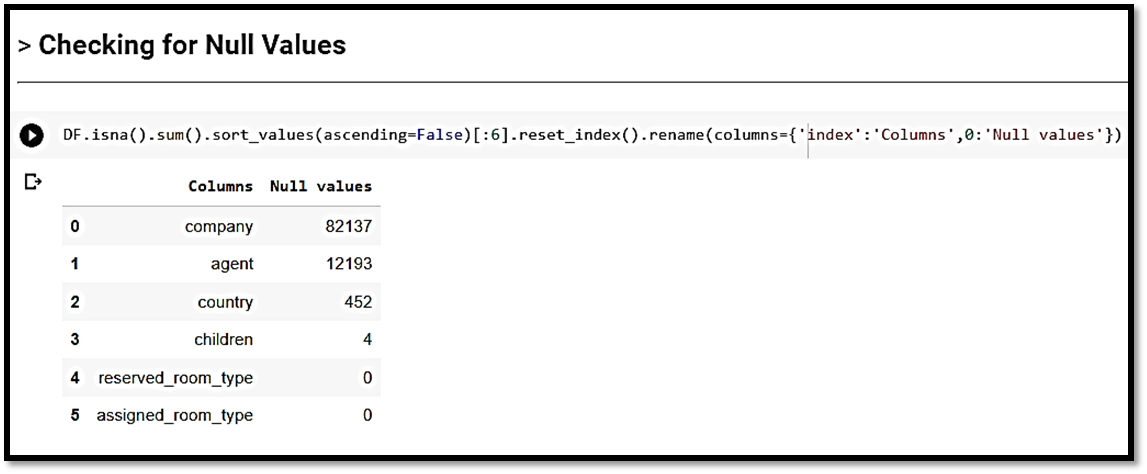
**30.total\_of\_special\_requests :** Number of special requests made by the customer (e.g. twin bed or high floor)

**31.reservation\_status :** Reservation last status, assuming one of three categories

**4.5 Data Exploration**

Data exploration is **the first step in data analysis involving the use of data** visualization tools and statistical techniques

**4.6 Identify and remove null values**

****

* The table shows us a null values are present in data set
* 1.for company and agent I will fill the Missing values with Zero (0)  
    
  2.for country we will fill Missing values with object 'others'.  
    
  3.AS the count of missing values in Children Column is only 4, so we can replace with 0 considering no children's.
* **Null values Treatment**

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we dropped them at the beginning of our project inorder to get a better result.

* **Encoding of categorical columns**

We can used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

**5.0 Exploration Data and Visualizing**

Data exploration is the first step of data analysis used to explore and visualize data to uncover insights from the start or identify areas or patterns to dig into more. Using interactive dashboards and point-and-click data exploration, users can better understand the bigger picture and get to insights faster.

## **5.1 1)Hotel Type**

* Type of hotel preferred by most of the customers is very important factor in hotel booking analysis.

#we can see the hotel type count in graph

plt.figure(figsize=(6 ,6))

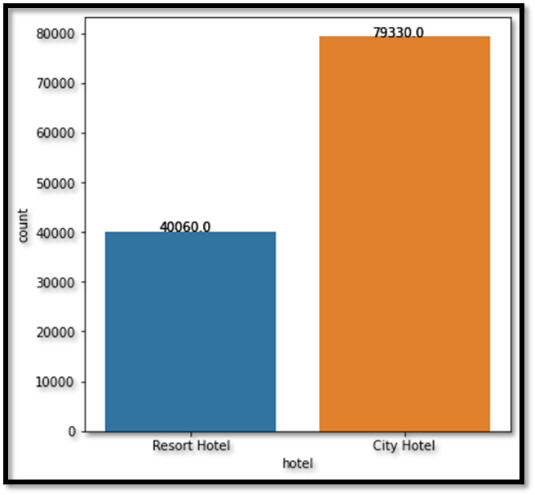
sns.countplot(DF['hotel'], palette='husl')

ax = sns.countplot(x="hotel", data=DF)

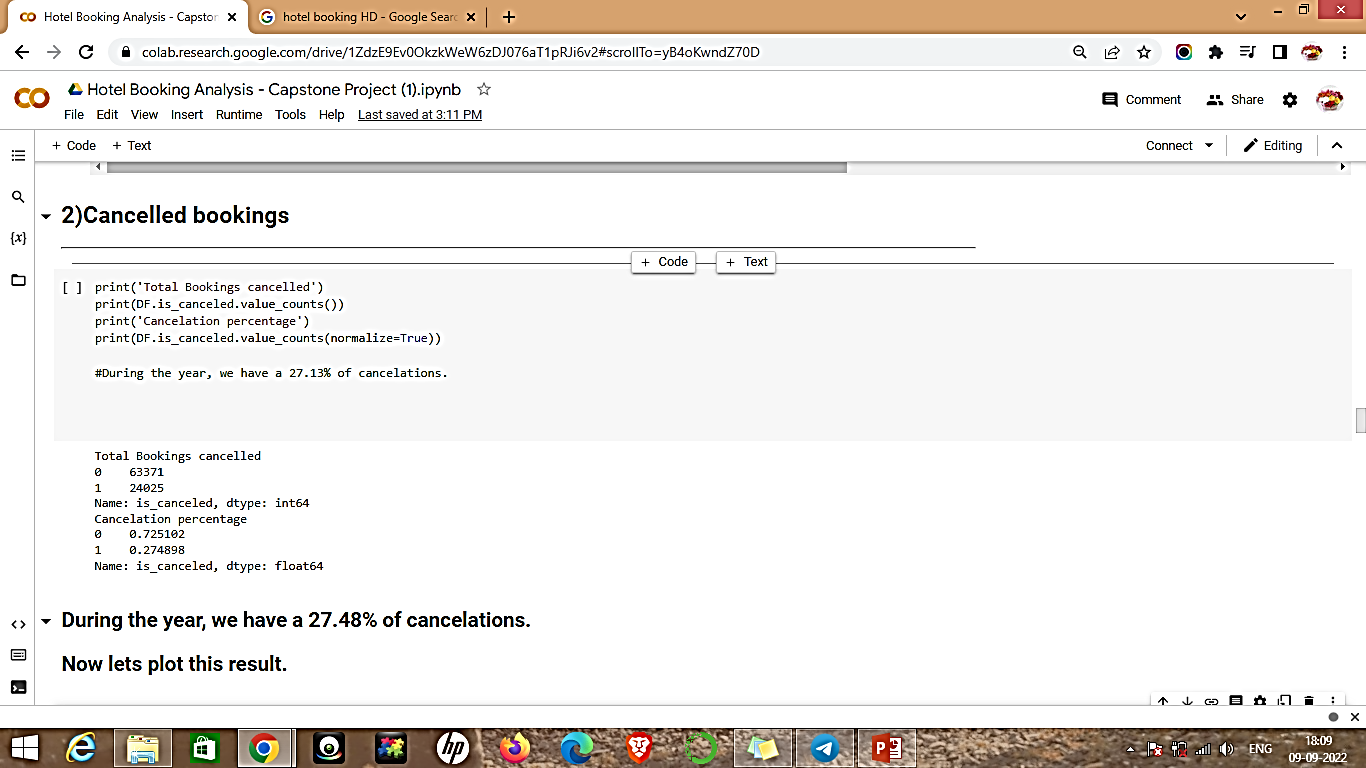
for p in ax.patches:

   ax.annotate('{:.1f}'.format(p.get\_height()), (p.get\_x()+0.25, p.get\_height()+0.90))

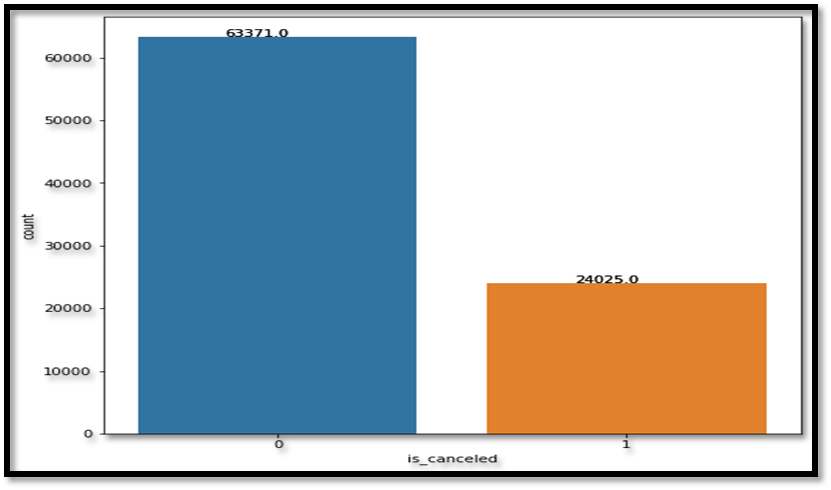
plt.show()



**5.2 Cancelled bookings**

****

* We can clearly seen that amongst the total booking count of cancelled booking is 44224 which is 27.48 % of total booking



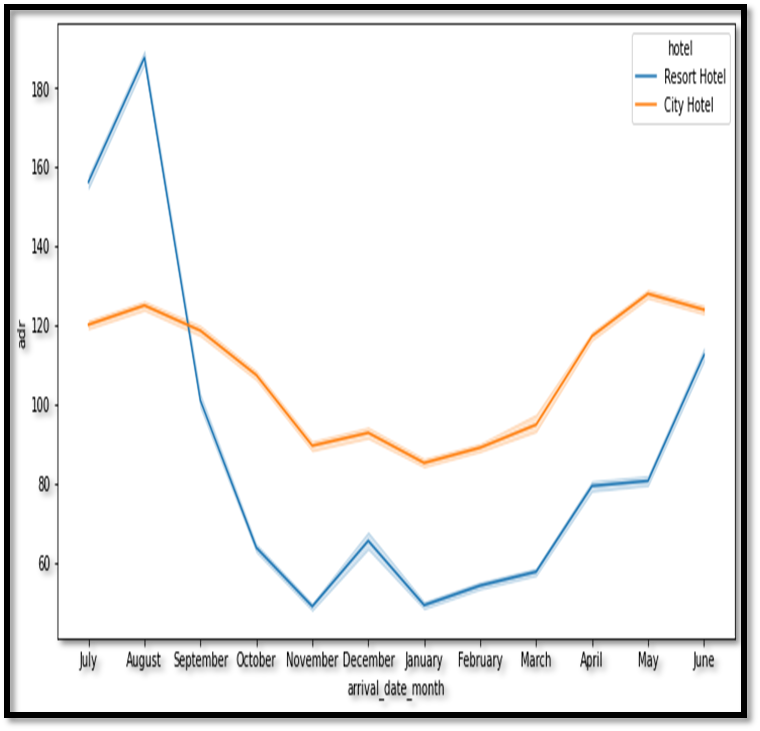
**5.3 Average Daily Rate (ADR) :- dividing the sum of all lodging transactions by the total number of staying nights**

* The average daily rate (ADR) shows how much revenue is made per room on average. The higher the ADR, the better. A rising ADR suggests that a hotel is increasing the money it's making from renting out rooms.
* #Average Daily Rate (ADR)

plt.figure(figsize=(18,6))

sns.lineplot(x='arrival\_date\_month', y='adr', hue='hotel', data= DF)

* plt.show()

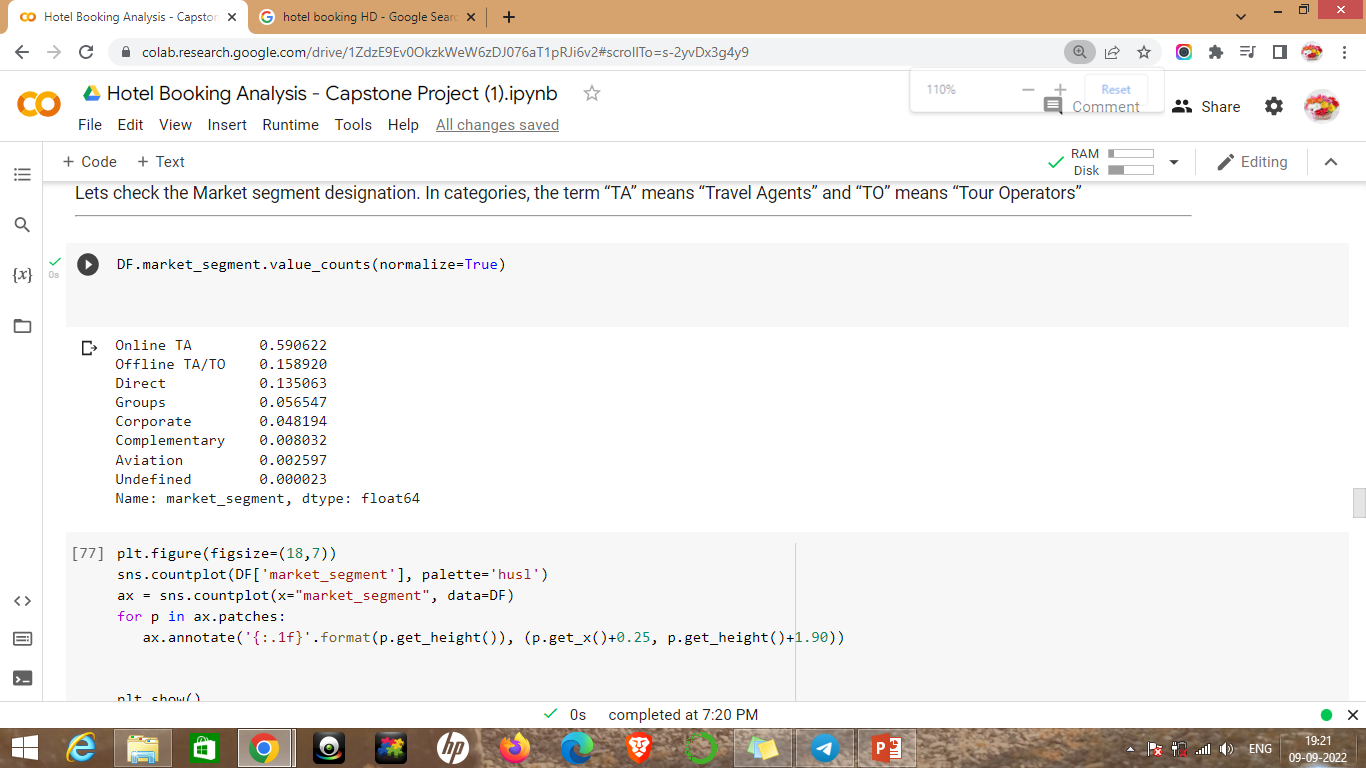


**5.4 Now lets check the arrival date by months**

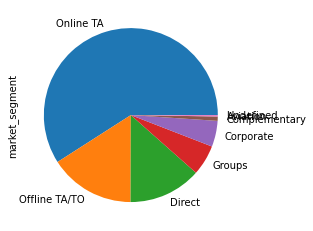
* The arrivals of customers is more in the month of August and next month with more numbers of
* arrivals was in the month of July comparatively less arrivals in other months .
* The number of arrivals were even good in number in the month of September, October, march,

april, may, June

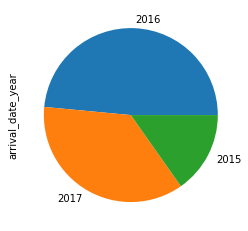
**5.5 Market Segment**

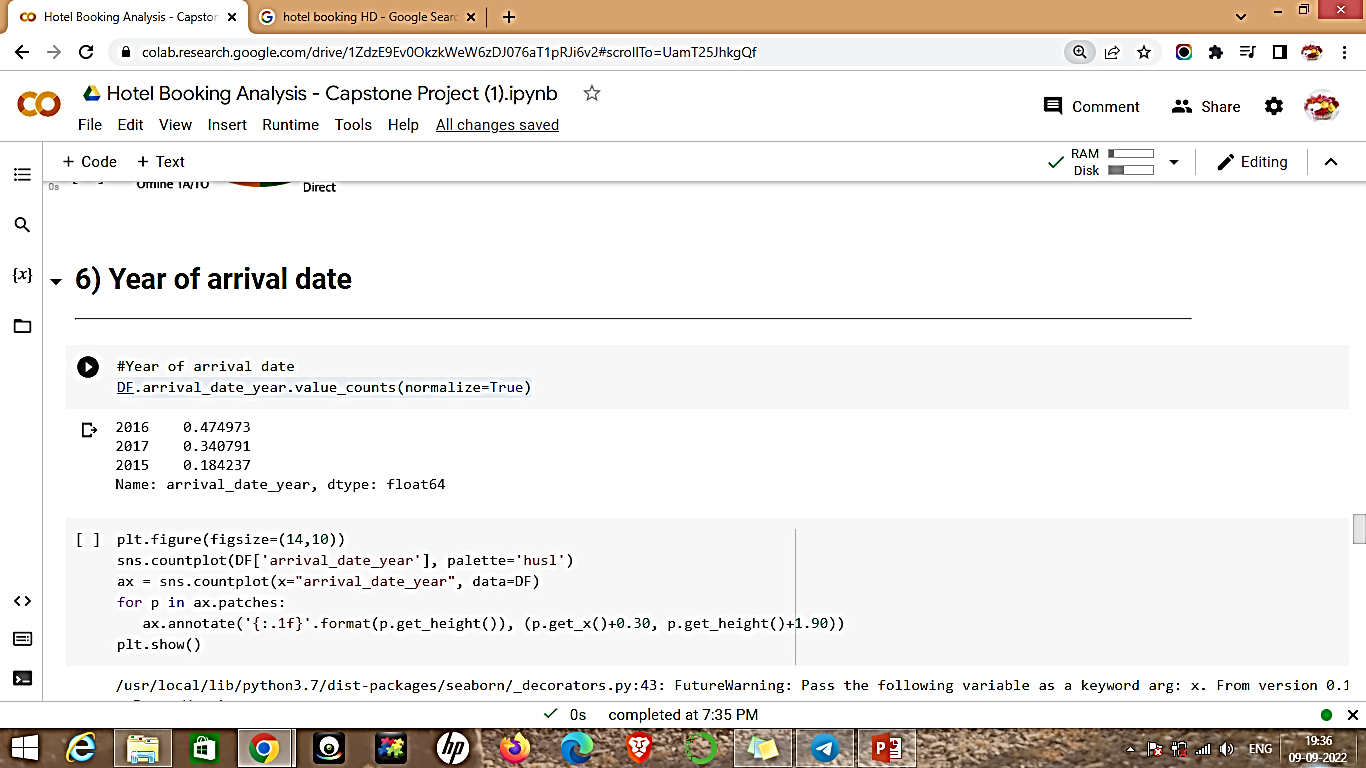
****

* Lets check the Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”.
* When hotels collaborate effectively with online TA they reduce their inventories of uncooked rooms and boost revenue .

****

**5.6 Year of arrival date**

****

****

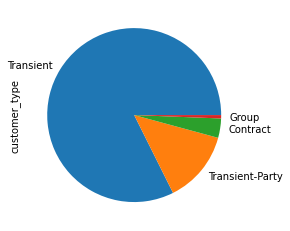
**5.7 Customer Type**

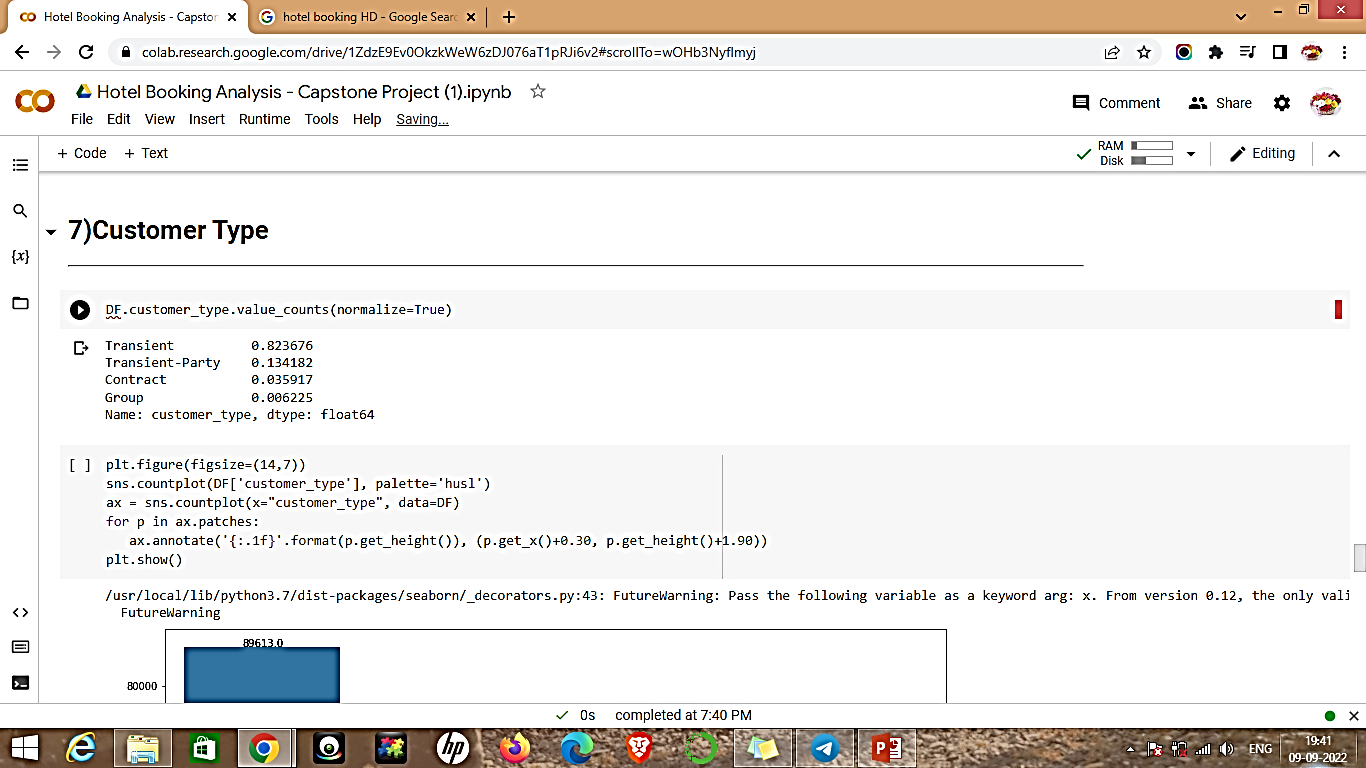
a)Contract - when the booking has an allotment or other type of contract associated to it;

b)Group – when the booking is associated to a group;

c)Transient – when the booking is not part of a group or contract, and is not associated to other transient booking;

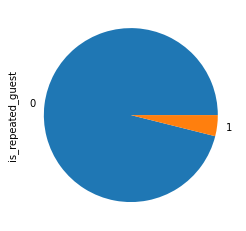
d)Transient-party – when the booking is transient, but is associated to at least other transient booking

****

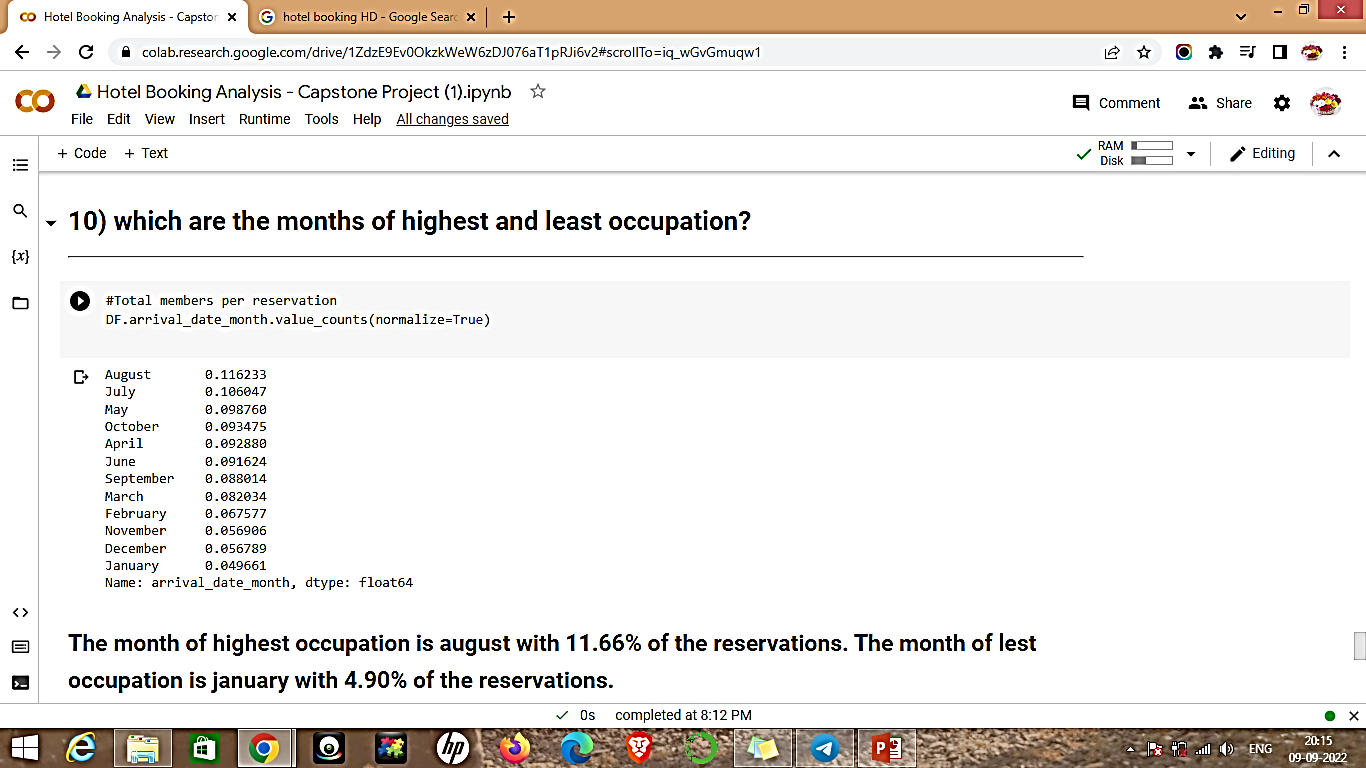
****

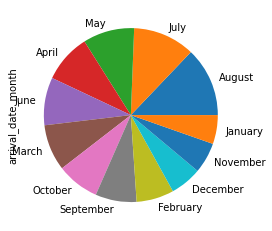
**5.8 Repeated Guest**

Check if the booking name was from a repeated guest (1) or not (0)



**5.9 which are the months of highest and least occupation?**

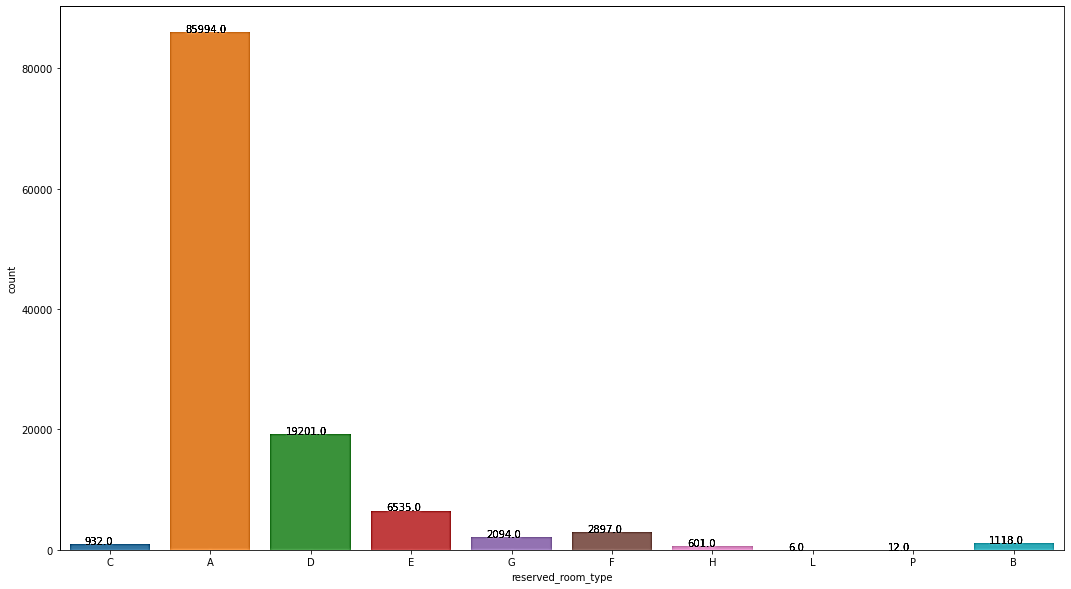
****

****

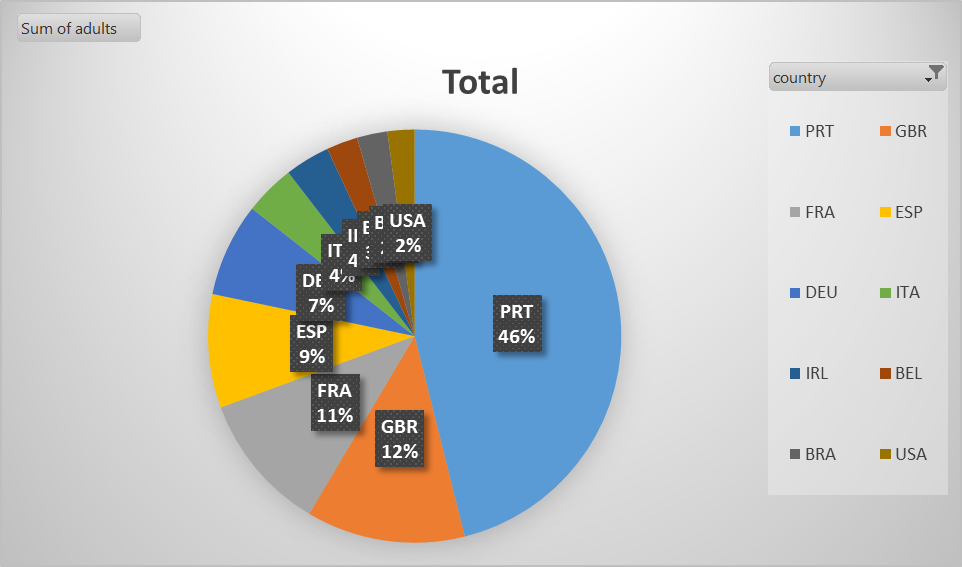
* The month of highest occupation is august with 11.62 % of the reservations.
* The month of lest occupation is january with 4.90 % of the

reservations.

**5.10 Now lets check the arrival date by months**

****

**5.11 Which is the most visited customer from which country?**



**Conclusion**

* City hotel got most of booking in last three years of time span.
* During the year, we have a 27.48% of cancelations.in that city hotels boking cancellation is more than resort hotel
* Resort hotels, the average daily rate is more expensive during august, July and September.
* For city hotels, the average daily rate is more expensive during august,jully, june and may.
* The arrivals of customers is more in the month of August and next month with more numbers of arrivals was in the month of July comparatively less arrivals in other months .
* Compared to other market segment designation online TA is high in range. With 56477.
* The number of arrivals in the year of 2016 is high with 56707 and the next year in 2017 is 40667
* The frequency of customer repetition was low hotel need to have opportunity in that area where they can make improvements
* The month of highest occupation is august with 11.62 % of the reservations. The month of lest occupation is January with 4.90 % of the reservations.
* The most of reserved room by customer is type A category which is 72 %