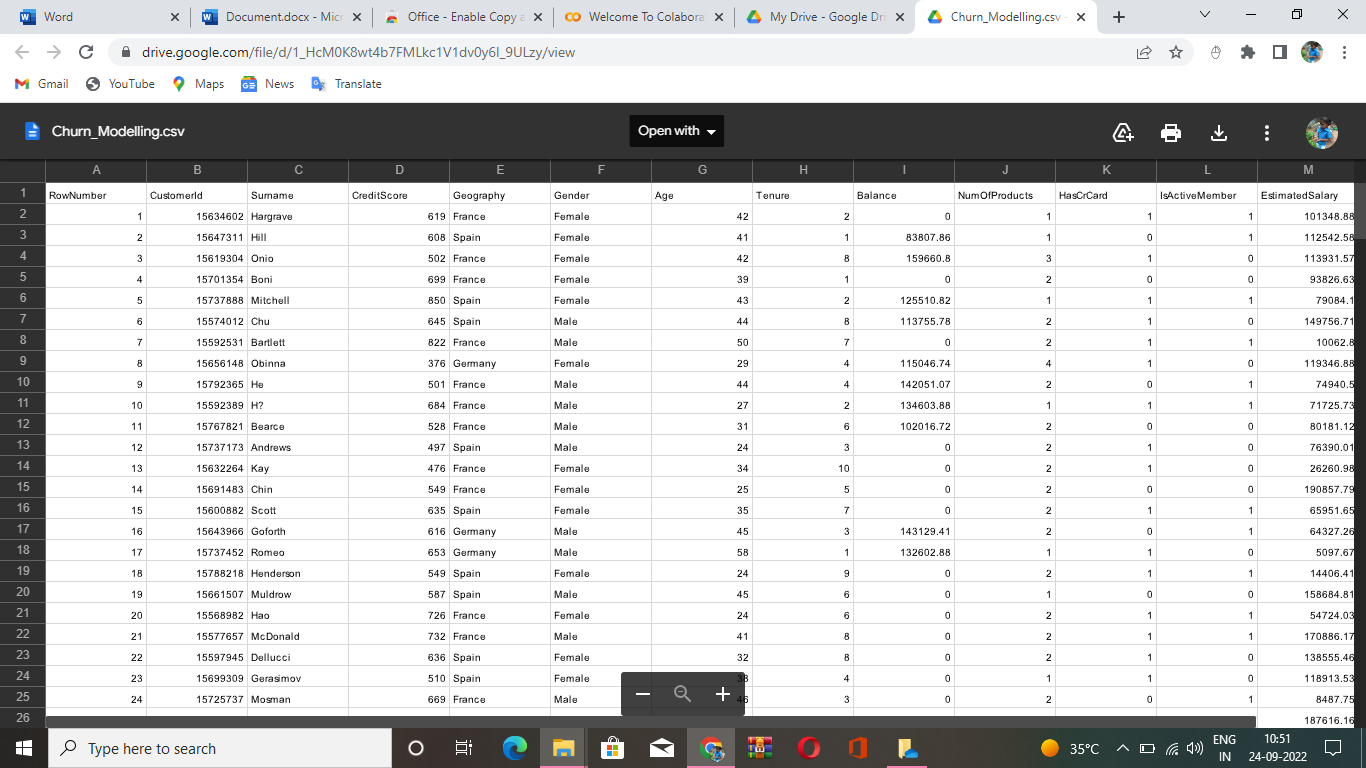
| Assignment Date | 19 September 2022 |
| --- | --- |
| Student Name | K.Yoka Muthu Nivethitha |
| Student Roll Number | 962719106040 |
| Maximum Mark |  |

Question –1

1. Download the dataset:



Question-2

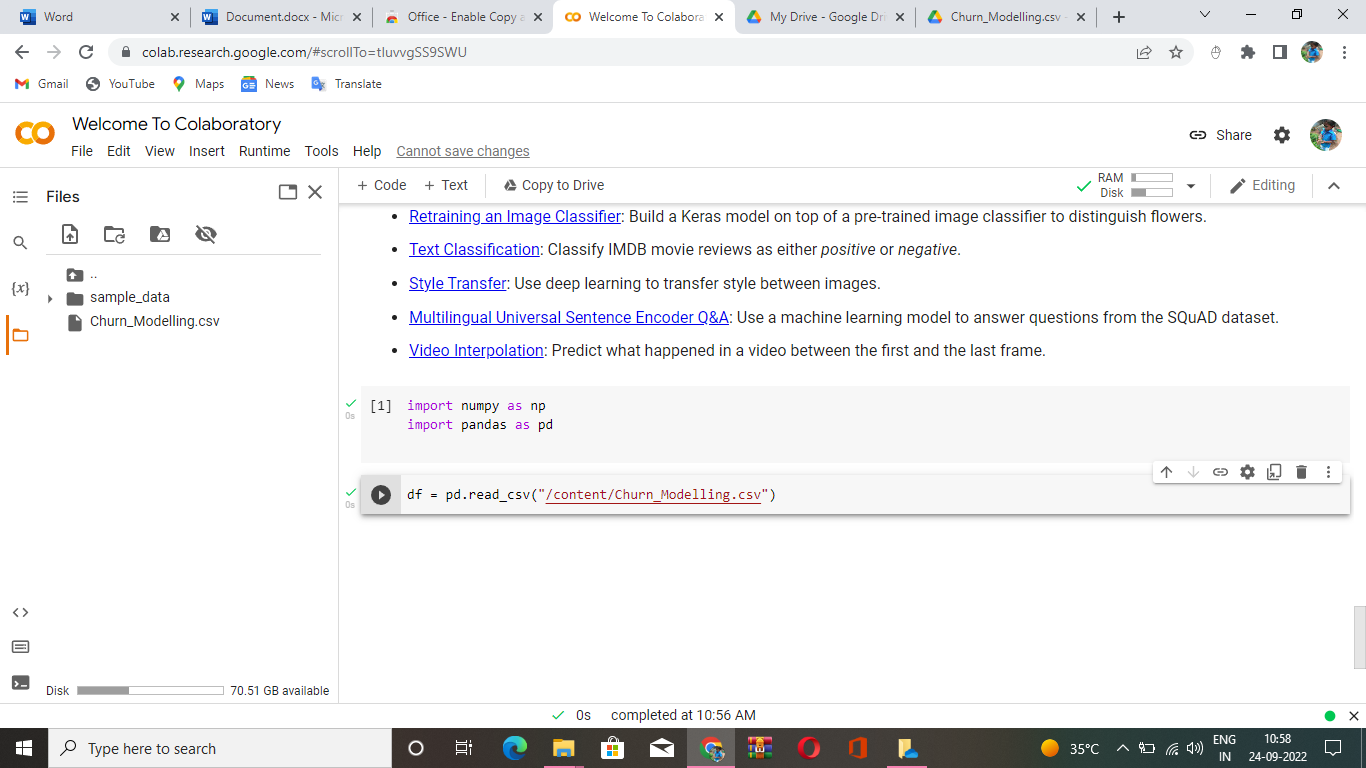
2. Load the dataset.

Solution:

import numpy as np

import pandas as pd

df = pd.read\_csv("/content/Churn\_Modelling.csv")



Question\_3

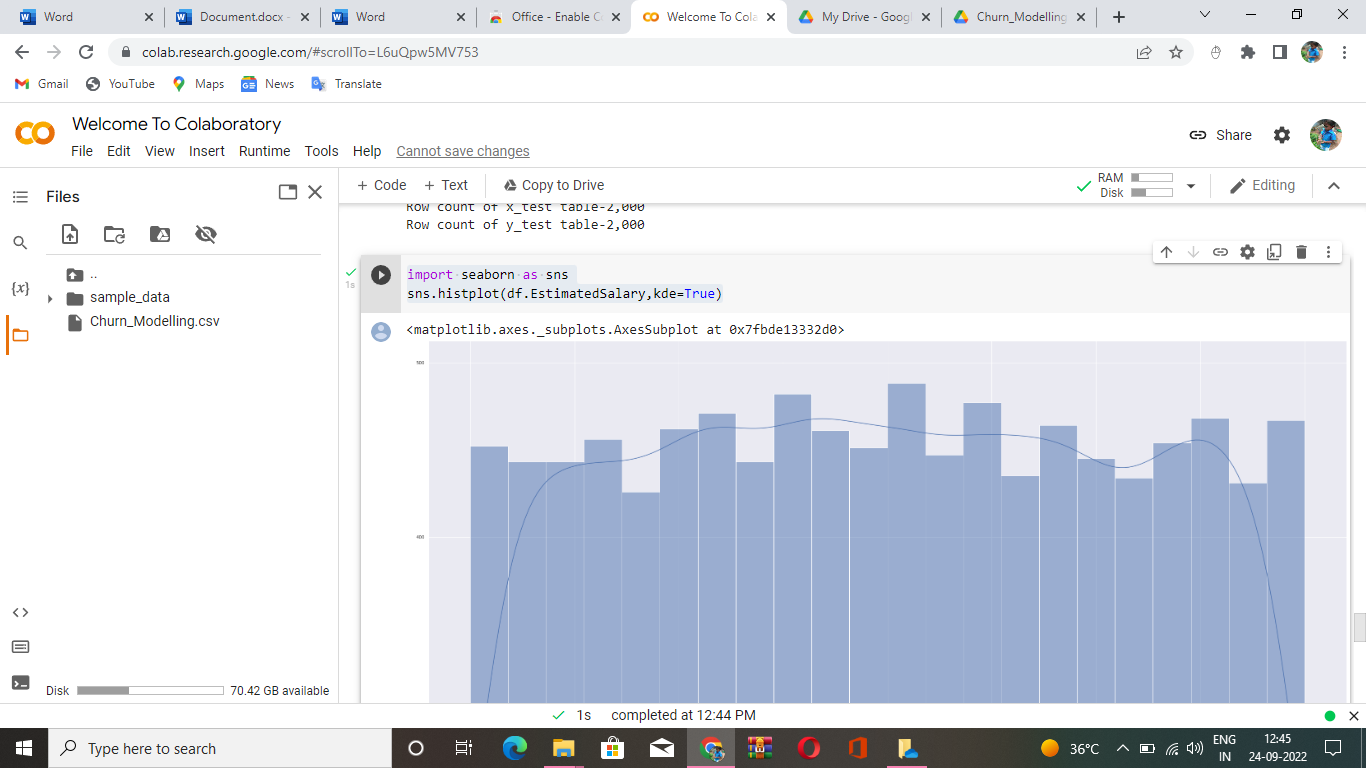
3. Perform Below Visualizations.

● Univariate Analysis

Solution:

import seaborn as sns

sns.histplot(df.EstimatedSalary,kde=True)



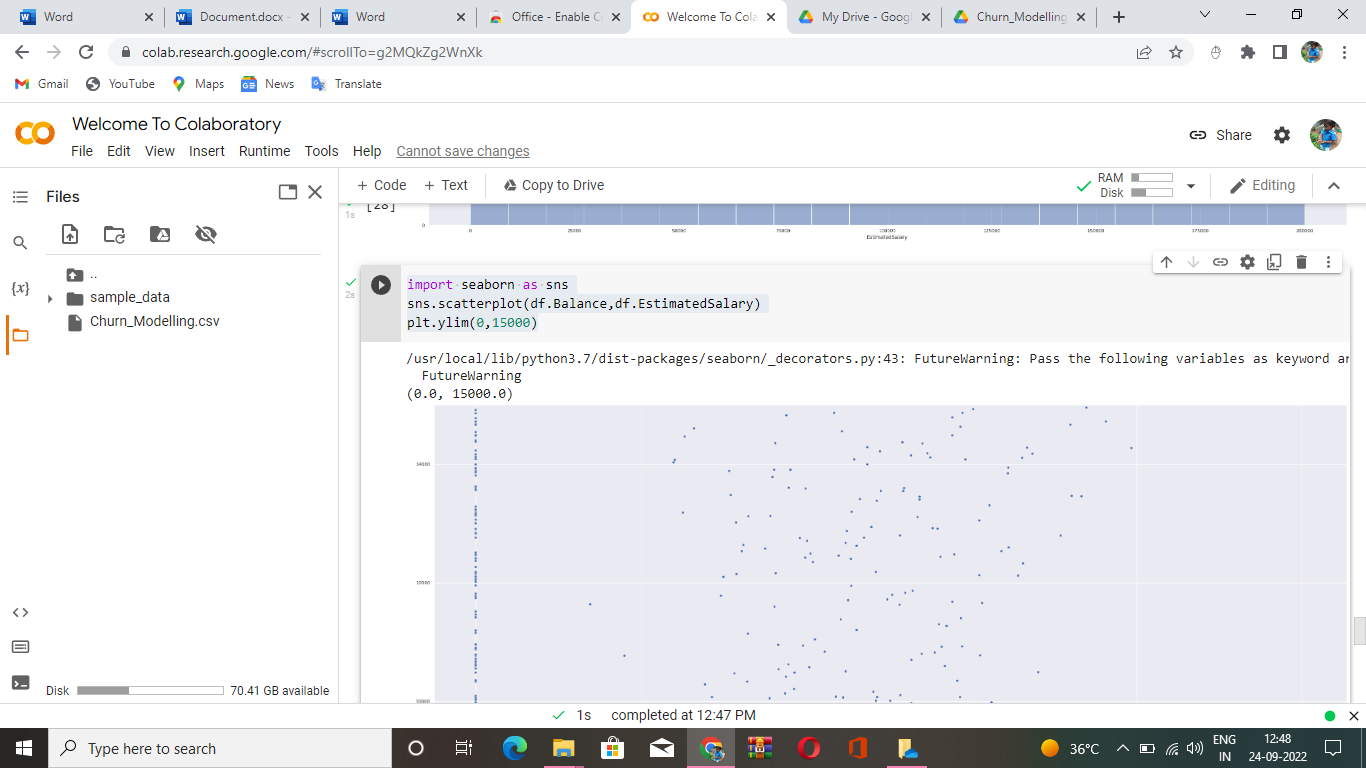
● Bi - Variate Analysis

Solution:

import seaborn as sns

sns.scatterplot(df.Balance,df.EstimatedSalary)

plt.ylim(0,15000)



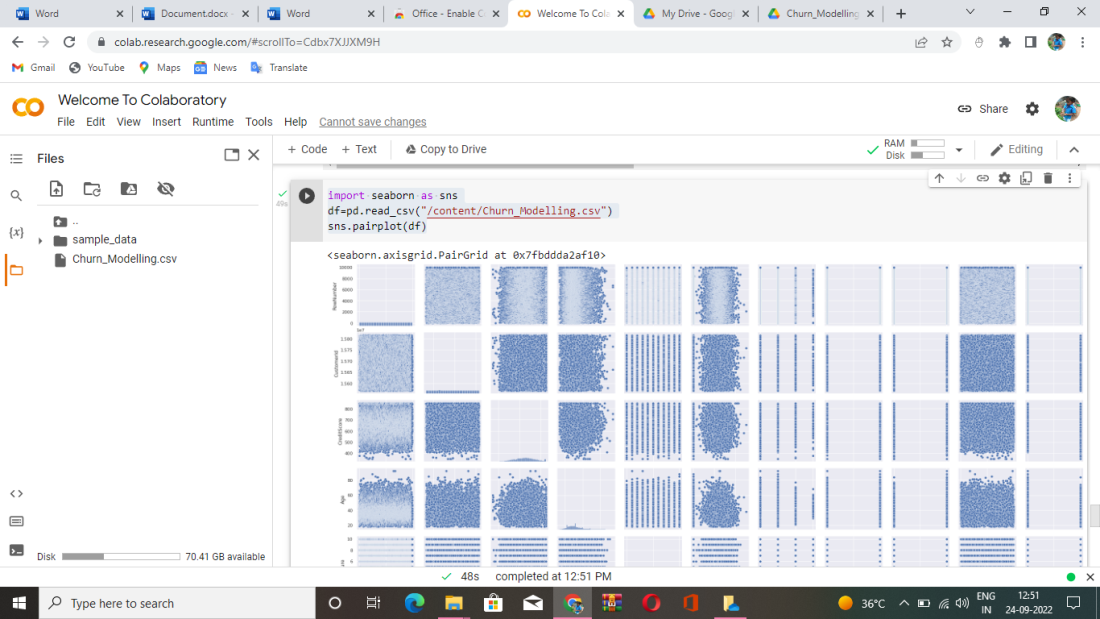
● Multi - Variate Analysis

Solution:

import seaborn as sns

df=pd.read\_csv("/content/Churn\_Modelling.csv")

sns.pairplot(df)



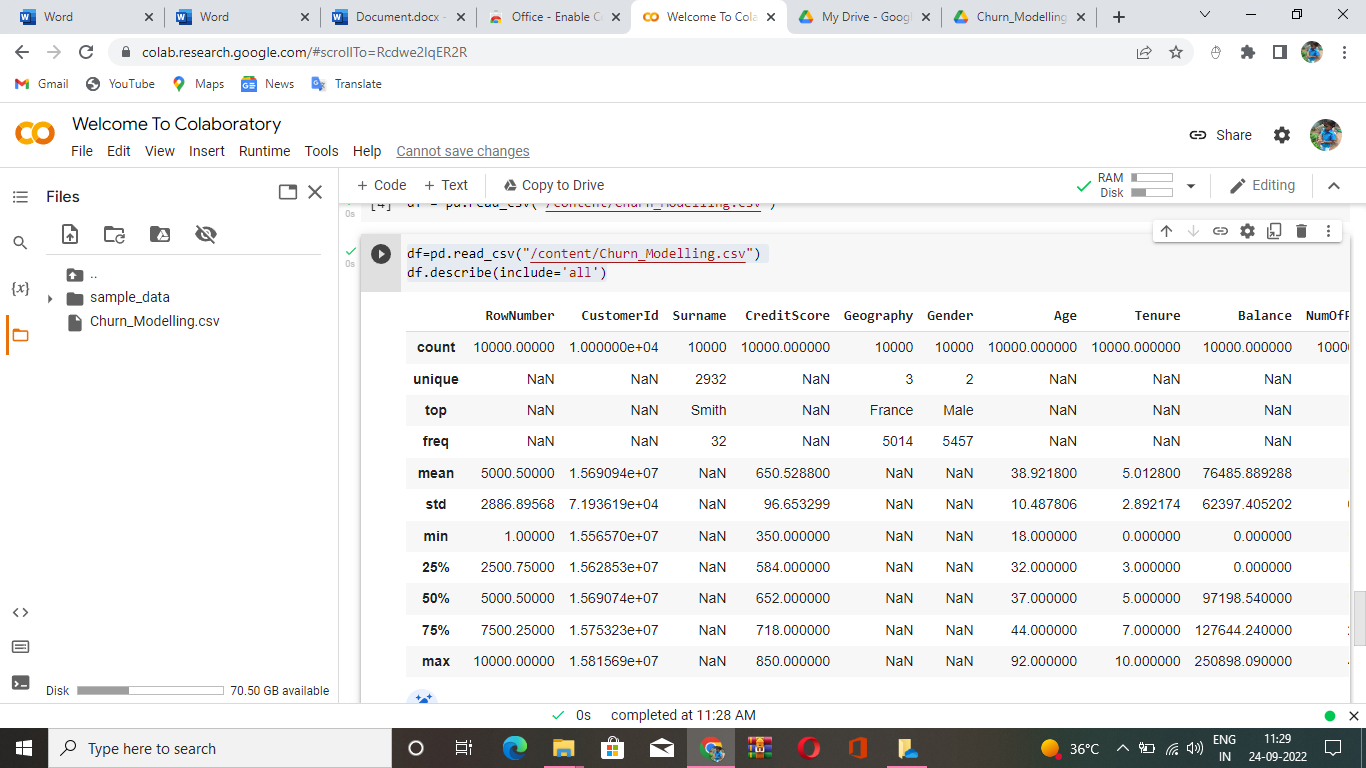
Question\_4

4. Perform descriptive statistics on the data set

Solution:

df=pd.read\_csv("/content/Churn\_Modelling.csv")

df.describe(include='all')



Question\_5

5. Handle the Missing values.

Solution:

from ast import increment\_lineno

import pandas as pd

import numpy as np

import seaborn as sns

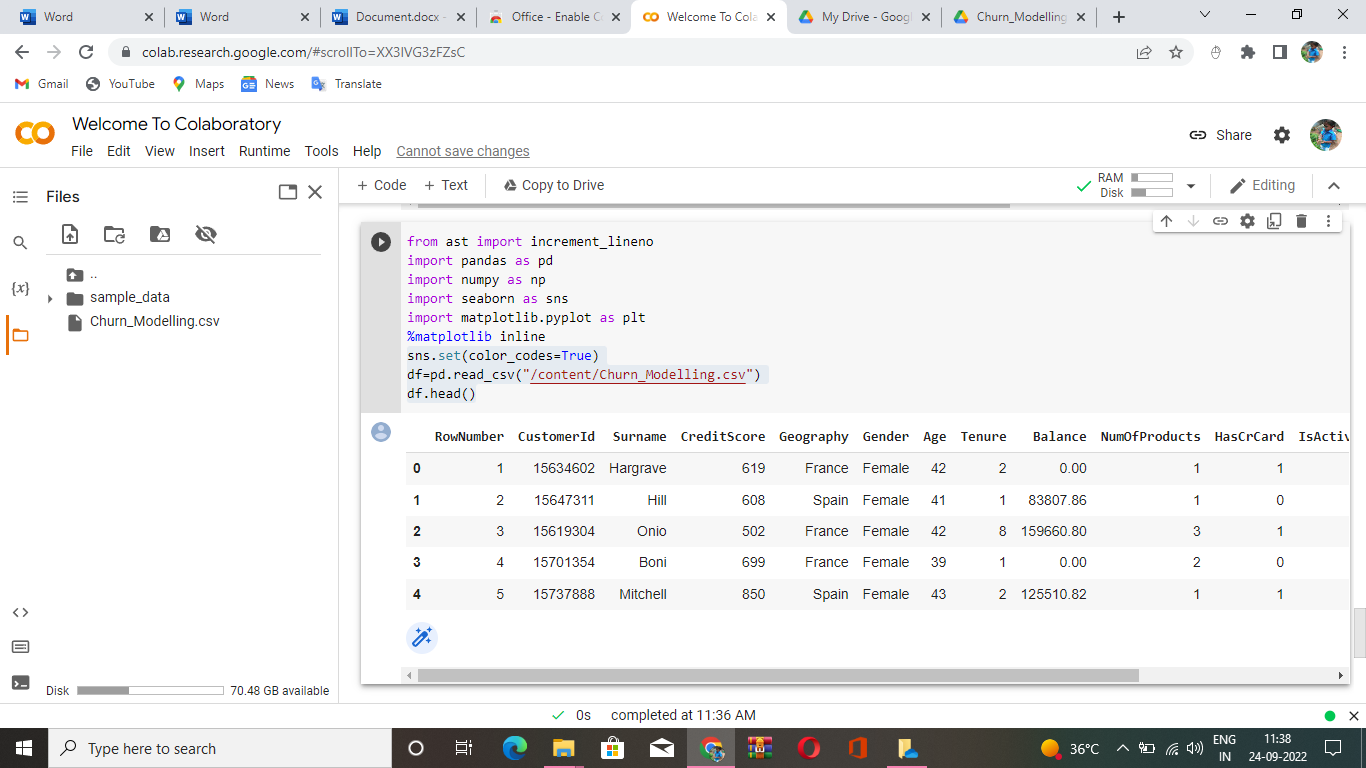
import matplotlib.pyplot as plt

%matplotlib inline

sns.set(color\_codes=True)

df=pd.read\_csv("/content/Churn\_Modelling.csv")

df.head()



Question\_6

6. Find the outliers and replace the outliers

Solution:

import pandas as pd

import matplotlib

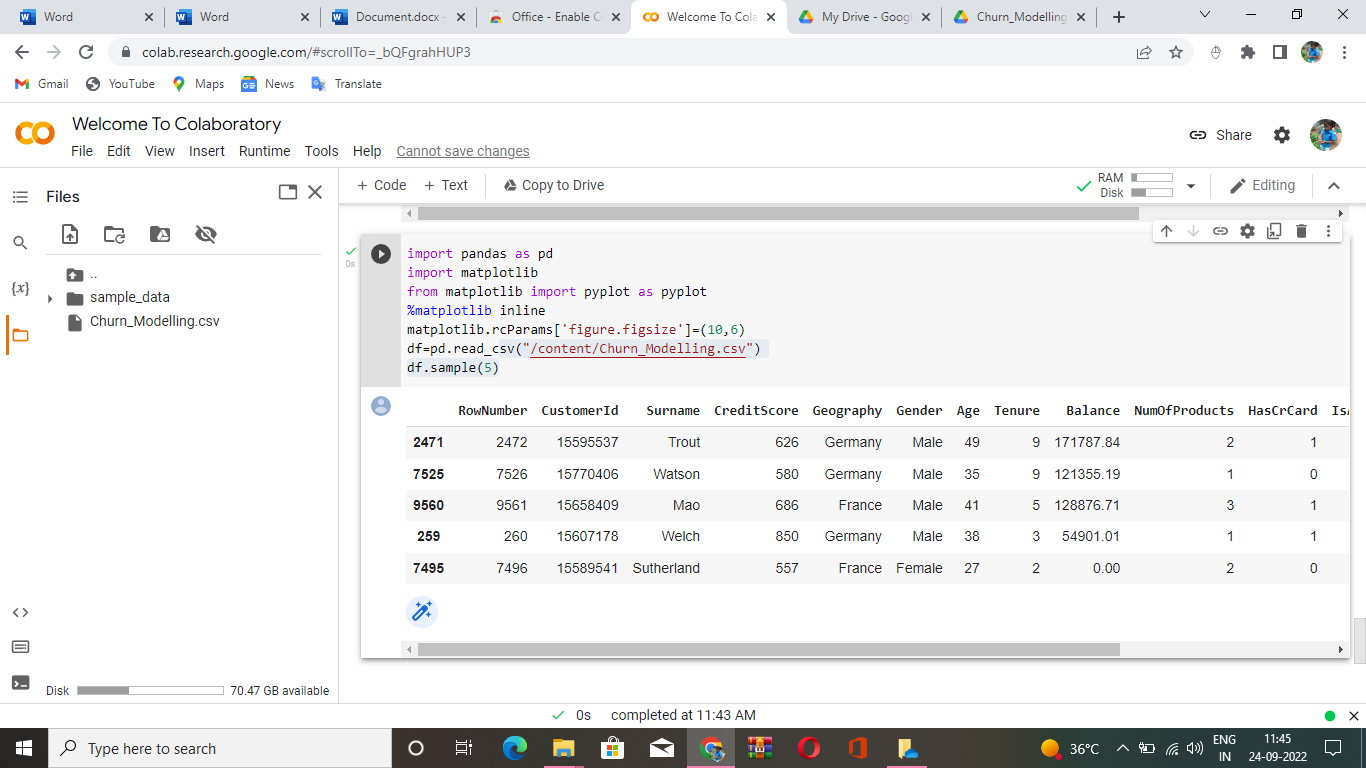
from matplotlib import pyplot as pyplot

%matplotlib inline

matplotlib.rcParams['figure.figsize']=(10,6)

df=pd.read\_csv("/content/Churn\_Modelling.csv")

df.sample(5)



Question\_7

7. Check for Categorical columns and perform encoding.

Solution:

df=pd.read\_csv("/content/Churn\_Modelling.csv")

df.columns

import pandas as pd

import numpy as np

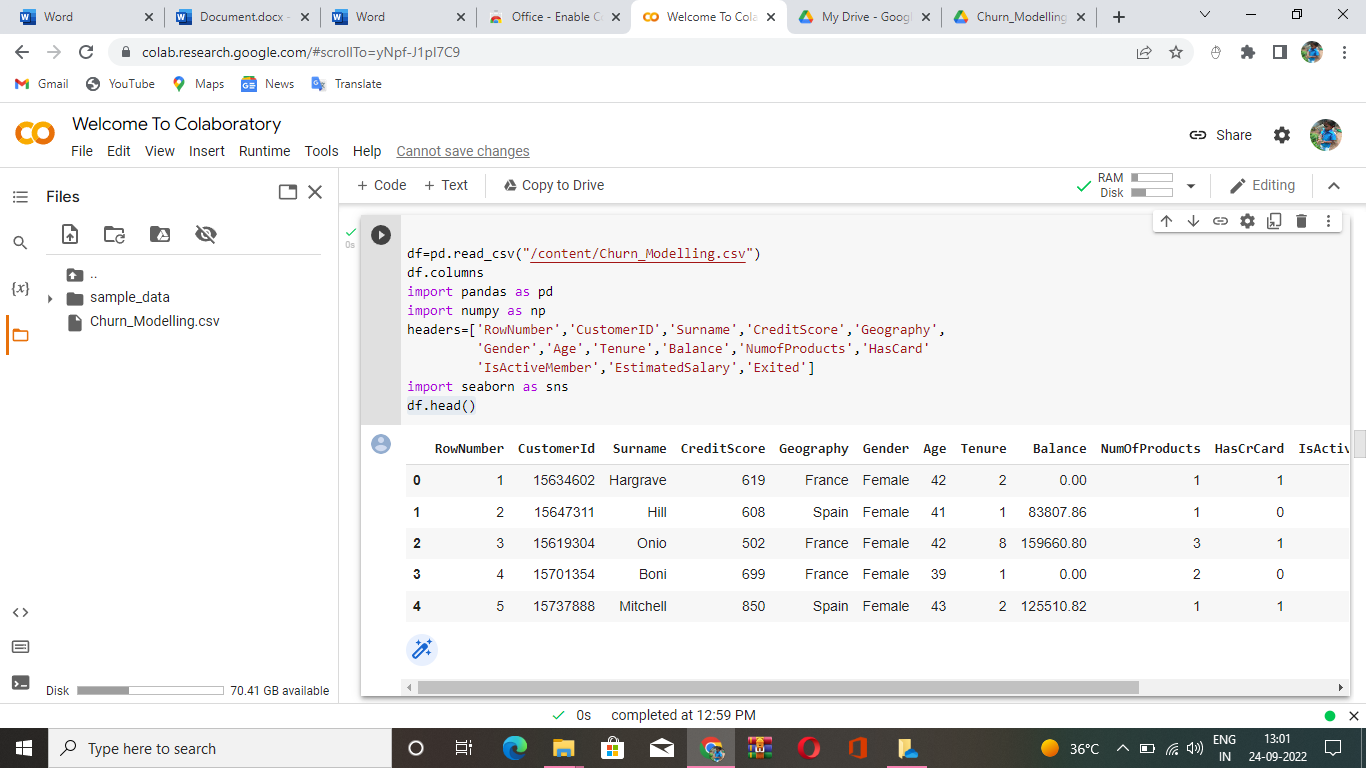
headers=['RowNumber','CustomerID','Surname','CreditScore','Geography',

'Gender','Age','Tenure','Balance','NumofProducts','HasCard'

'IsActiveMember','EstimatedSalary','Exited']

import seaborn as sns

df.head()



Questioh\_8

8. Split the data into dependent and independent variables.

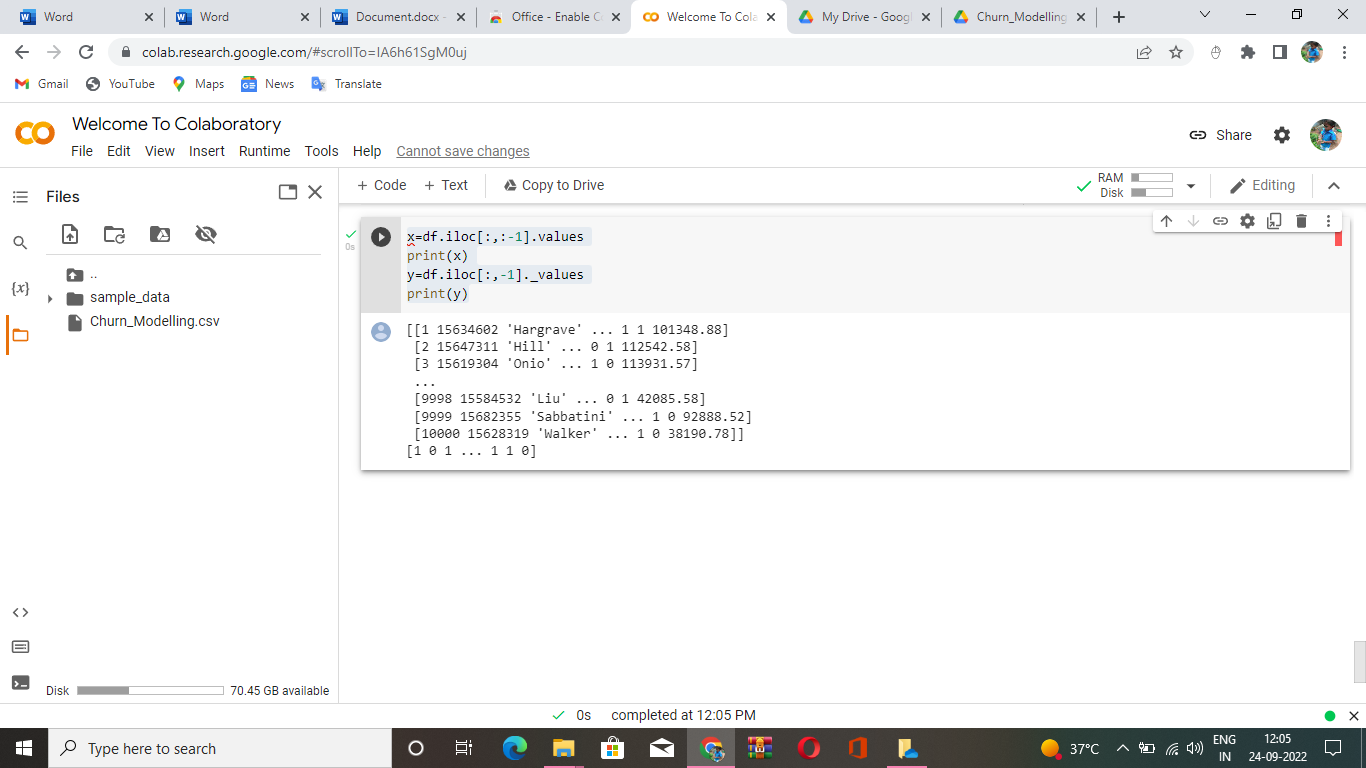
Solution:

x=df.iloc[:,:-1].values

print(x)

y=df.iloc[:,-1].\_values

print(y)



Question\_9

9. Scale the independent variables

Solution:

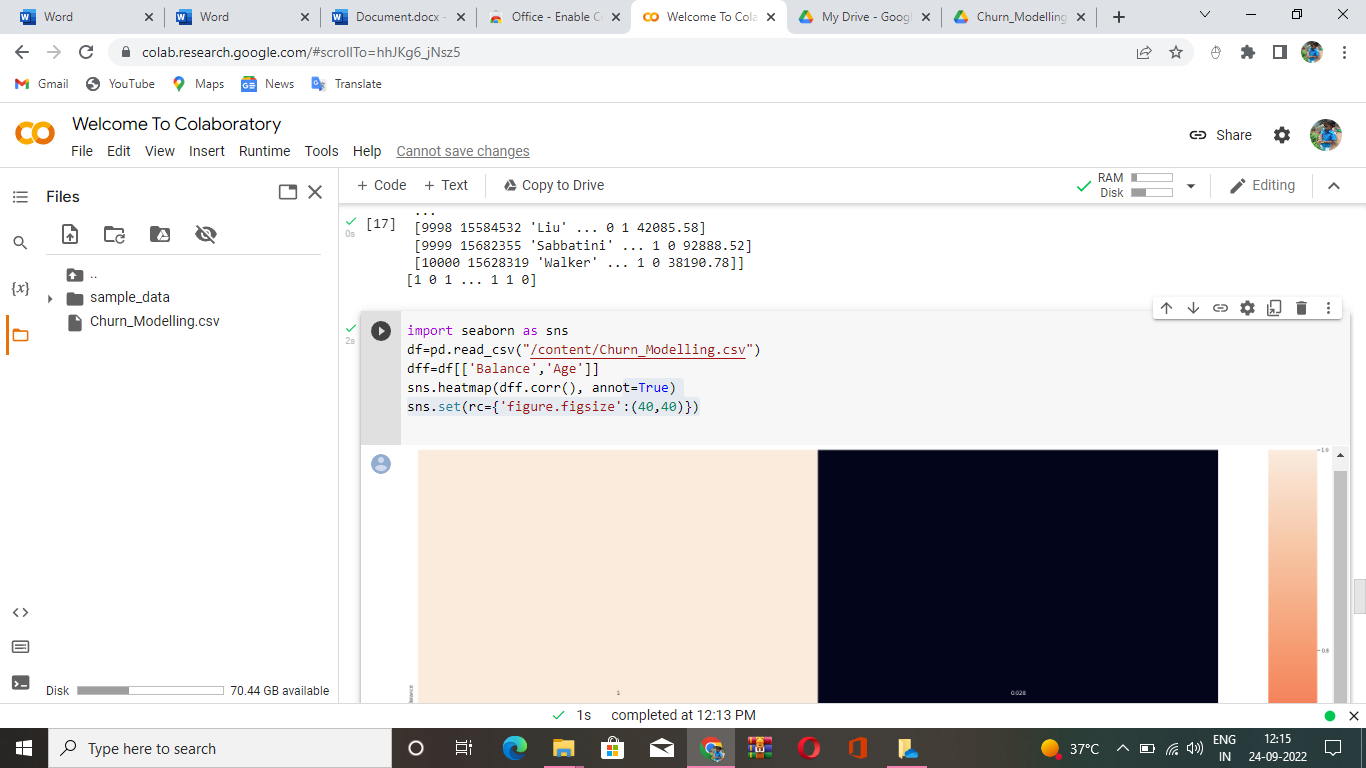
import seaborn as sns

df=pd.read\_csv("/content/Churn\_Modelling.csv")

dff=df[['Balance','Age']]

sns.heatmap(dff.corr(), annot=True)

sns.set(rc={'figure.figsize':(40,40)})



Question\_10

10. Split the data into training and testing

Solution:

from scipy.sparse.construct import random

x=df.iloc[:, 1:2].values

y=df.iloc[:,2].values

from sklearn.model\_selection import train\_test\_split

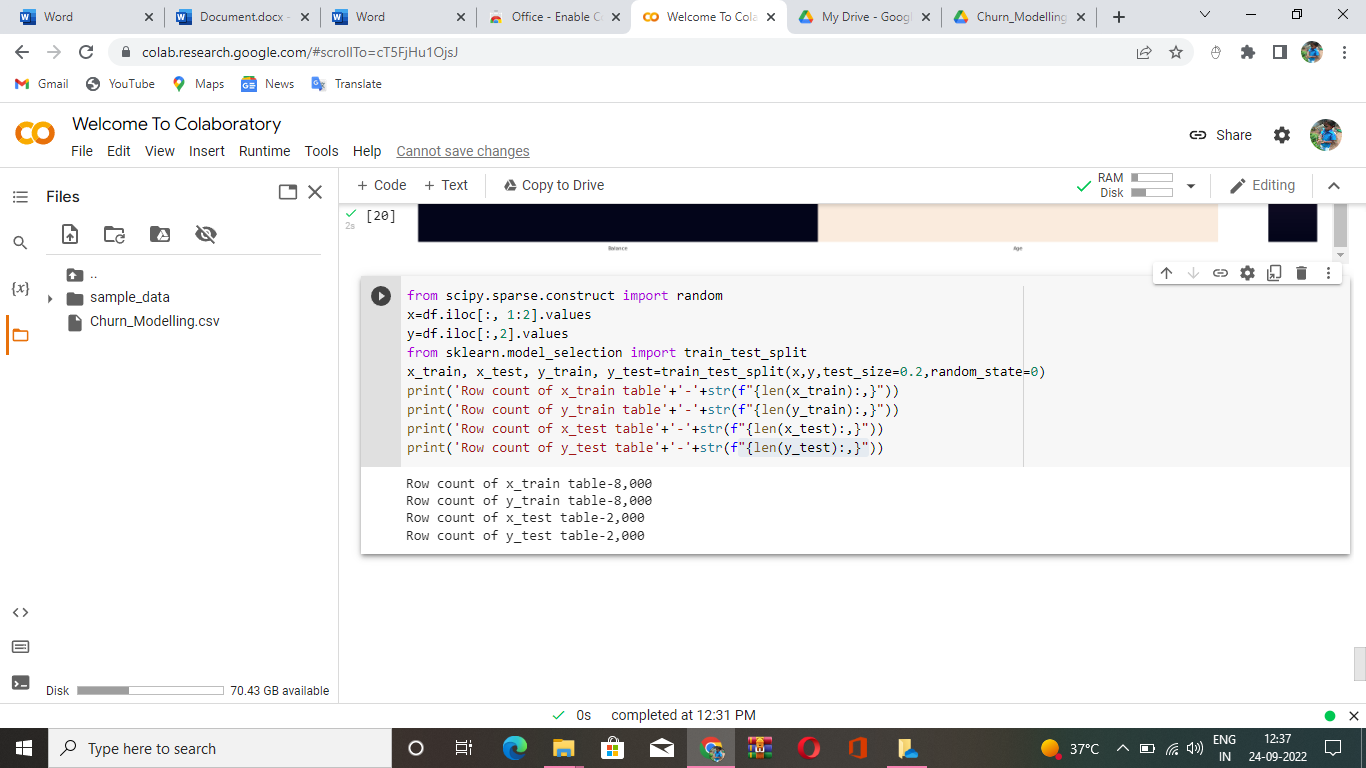
x\_train, x\_test, y\_train, y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=0)

print('Row count of x\_train table'+'-'+str(f"{len(x\_train):,}"))

print('Row count of y\_train table'+'-'+str(f"{len(y\_train):,}"))

print('Row count of x\_test table'+'-'+str(f"{len(x\_test):,}"))

print('Row count of y\_test table'+'-'+str(f"{len(y\_test):,}"))

s