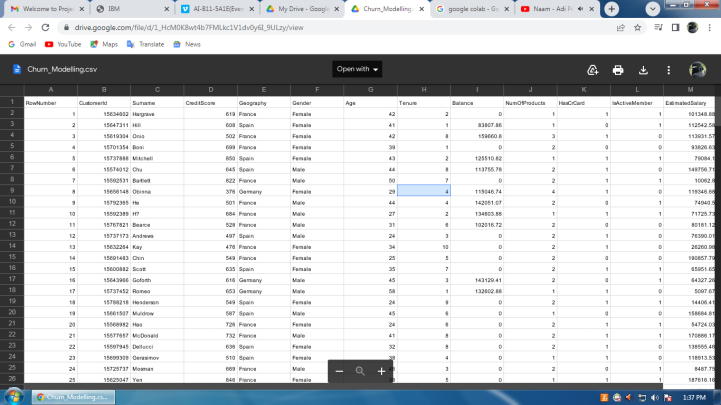
|  |  |
| --- | --- |
| Assignment Date | 19.09.2022 |
| Student Name | M.Hema |
| Student Roll Number | 962719106009 |
| Maximum Marks |  |

**Question 1:** **Download the dataset:**

**solution:**



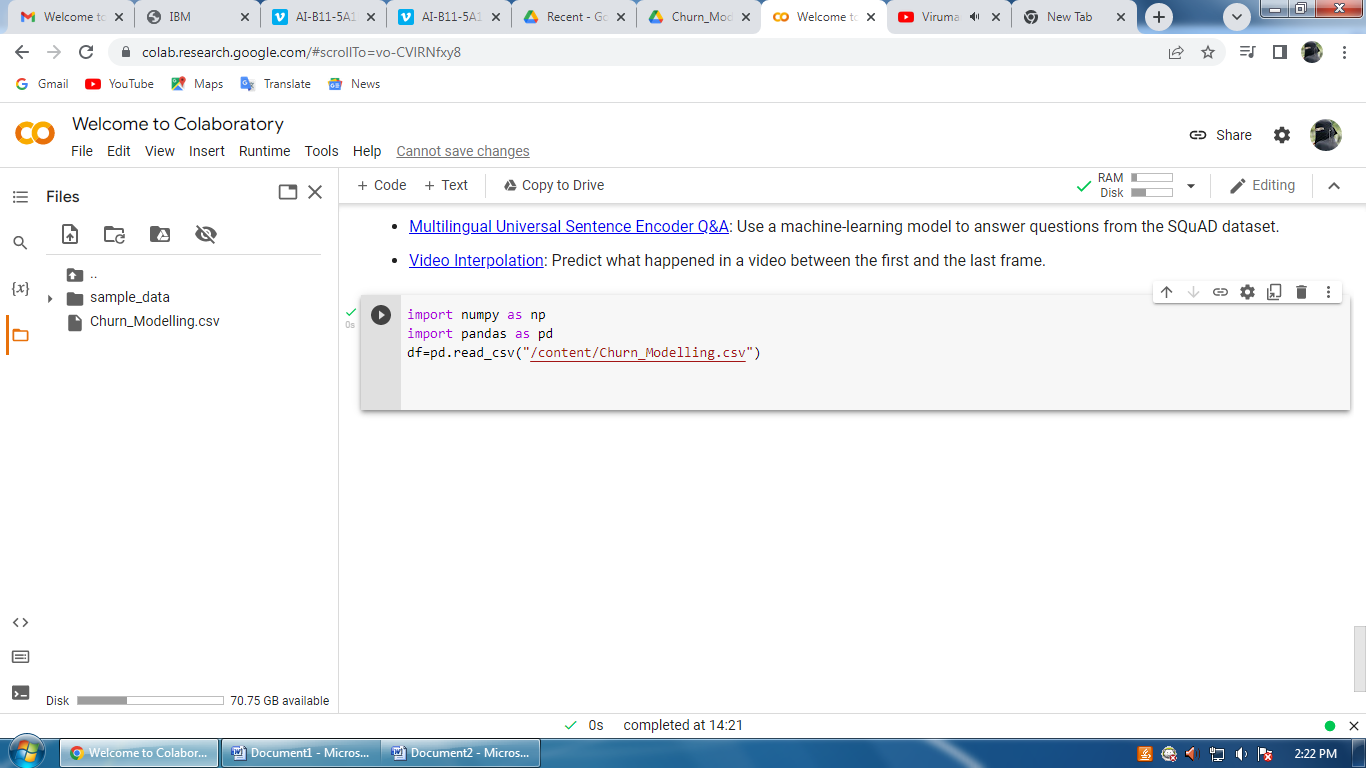
**Question 2: Load the dataset**.

**solution:**

import numpy as np

import pandas as pd

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

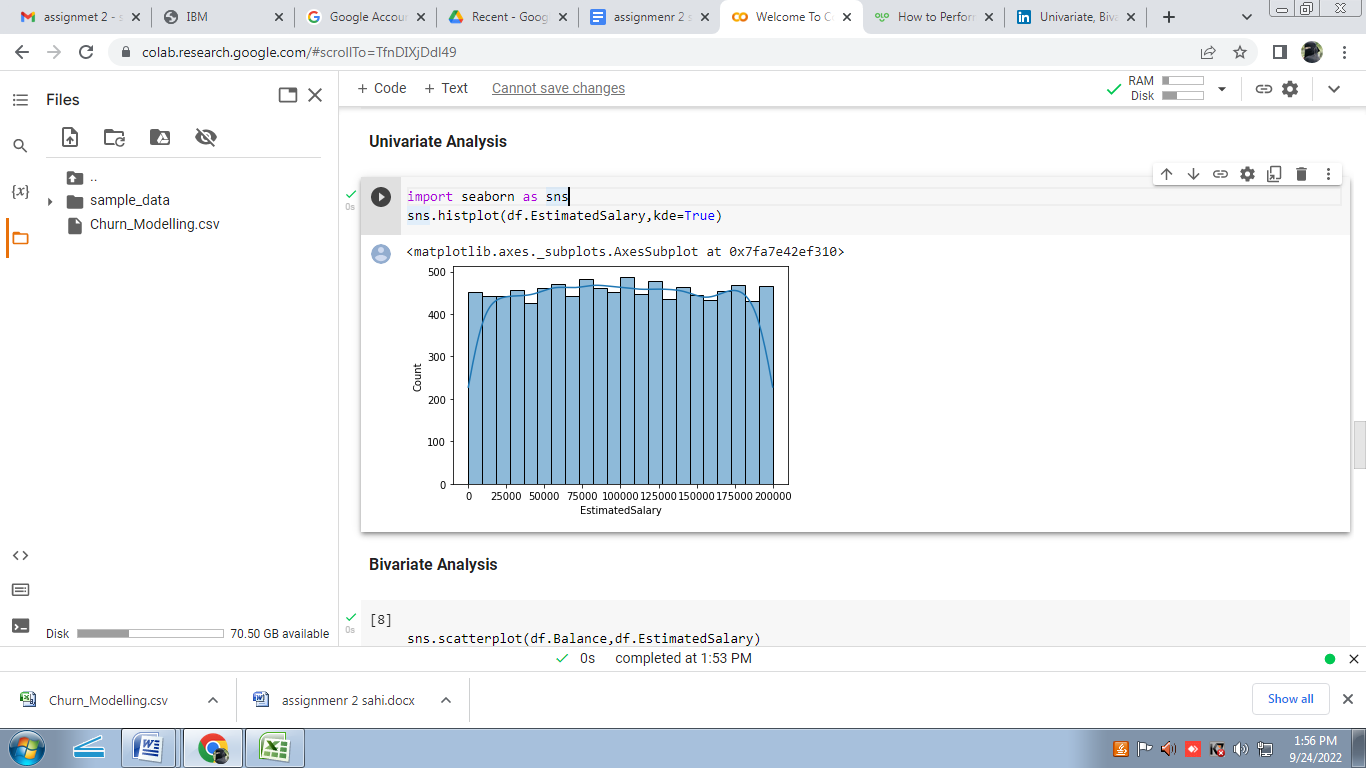


**Question 3. Perform Below Visualizations.**

**● Univariate Analysis**

**solution:**

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

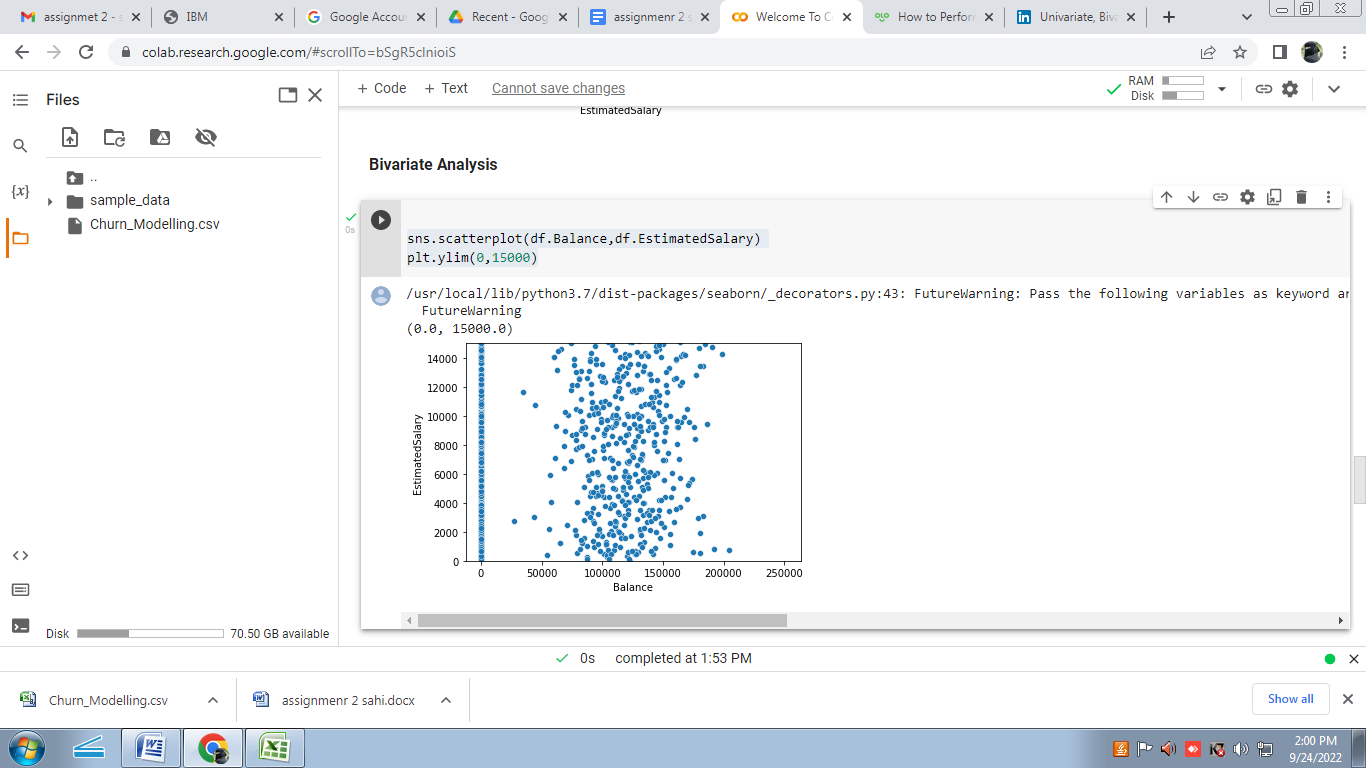


**● Bi - Variate Analysis**

**solution:**

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

plt.ylim(0,15000)

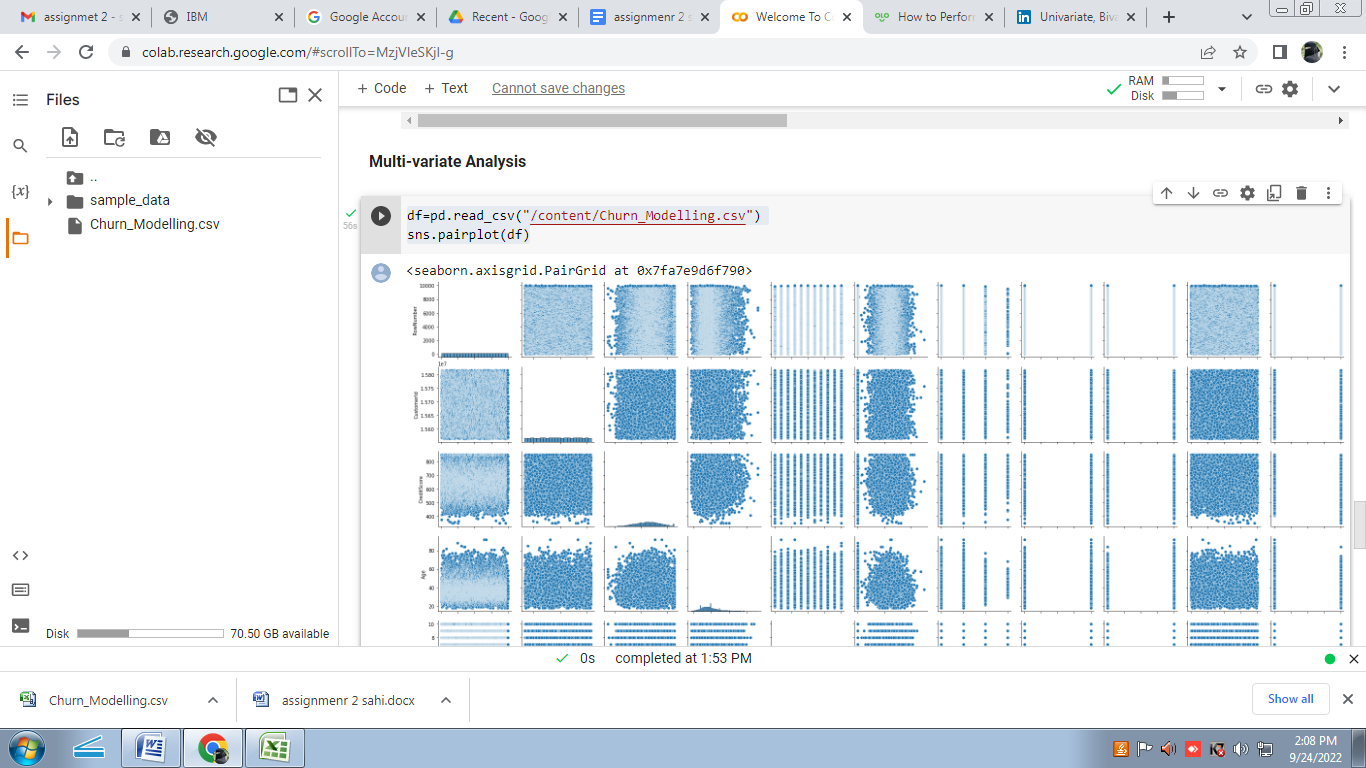
****

**● Multi - Variate Analysis**

**solution:**

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

sns.pairplot(df)

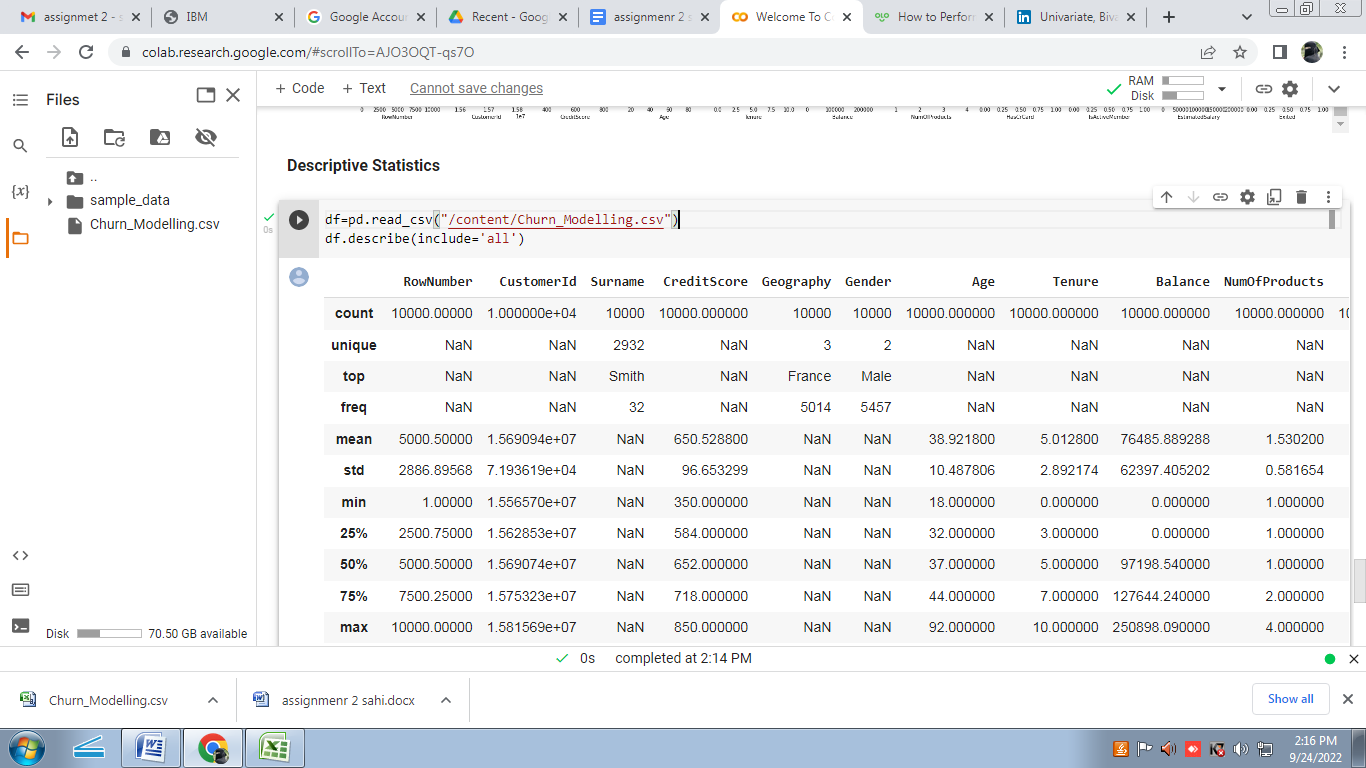
****

**Question 4. Perform descriptive statistics on the dataset.**

**solution:**

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

df.describe(include='all')

****

**Question 5 : Handle the Missing values**.

**solution:**

from ast import increment\_lineno

import pandas as pd

import numpy as np

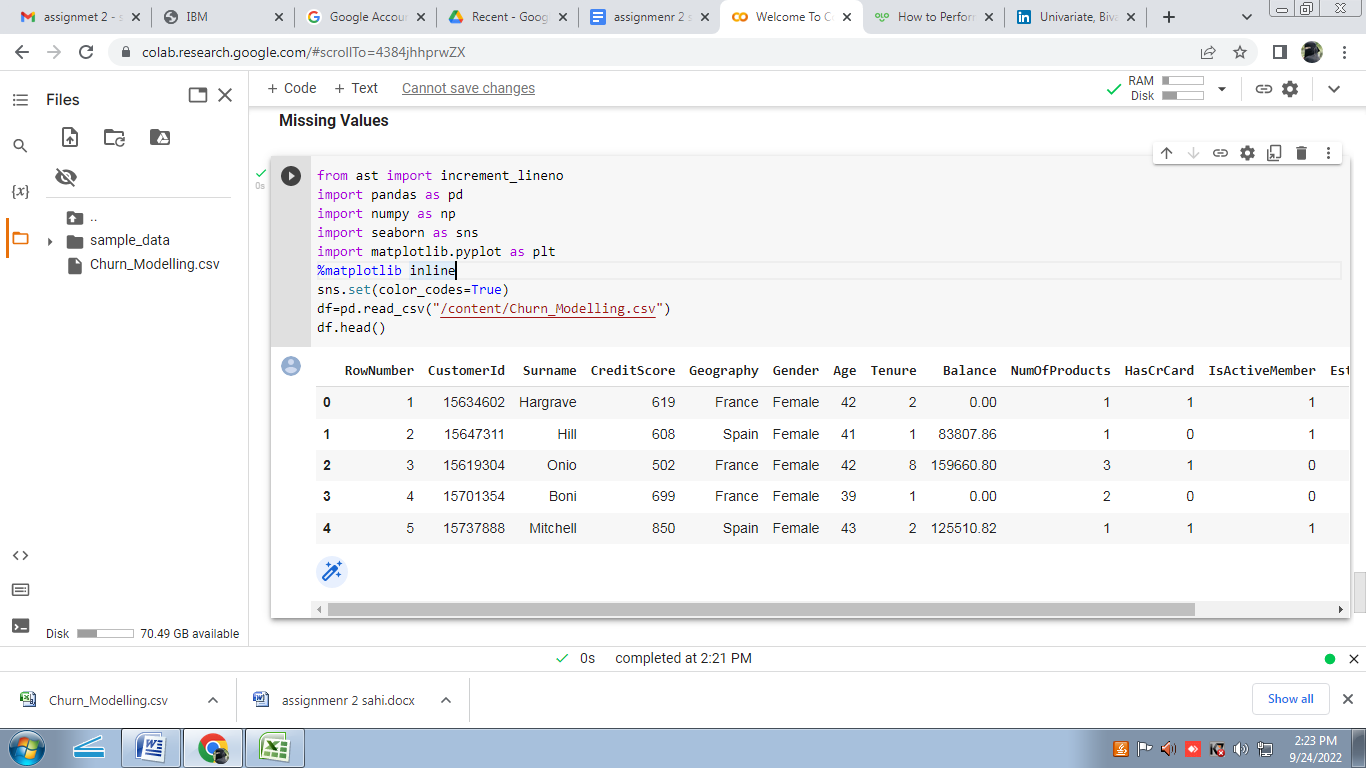
import seaborn as sns

import matplotlib.pyplot as plt

%matplotlib inline

sns.set(color\_codes=True)

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

 df.head()

**Question 6. Find the outliers and replace the outliers**

**solution:**

import pandas as pd

import matplotlib

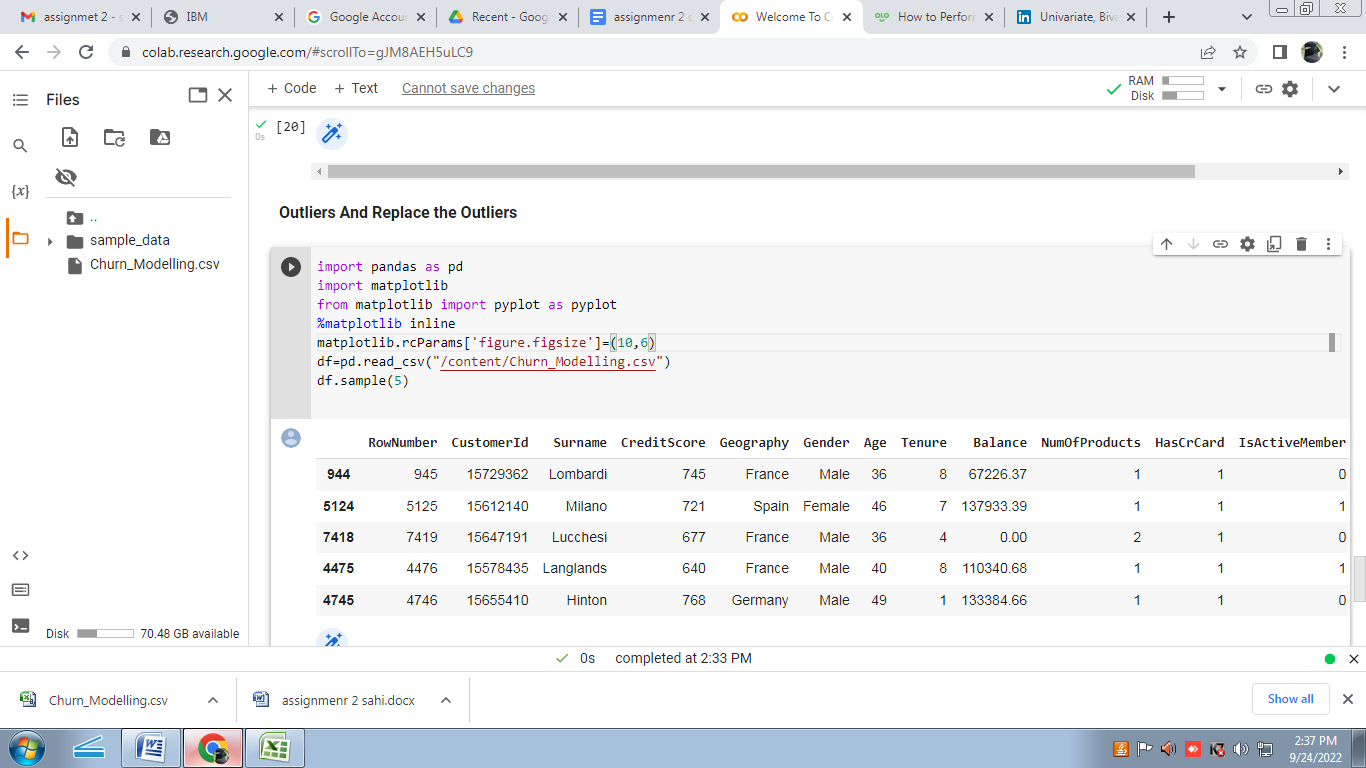
from matplotlib import pyplot as pyplot

%matplotlib inline

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

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

df.sample(5)



**Question 7. 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

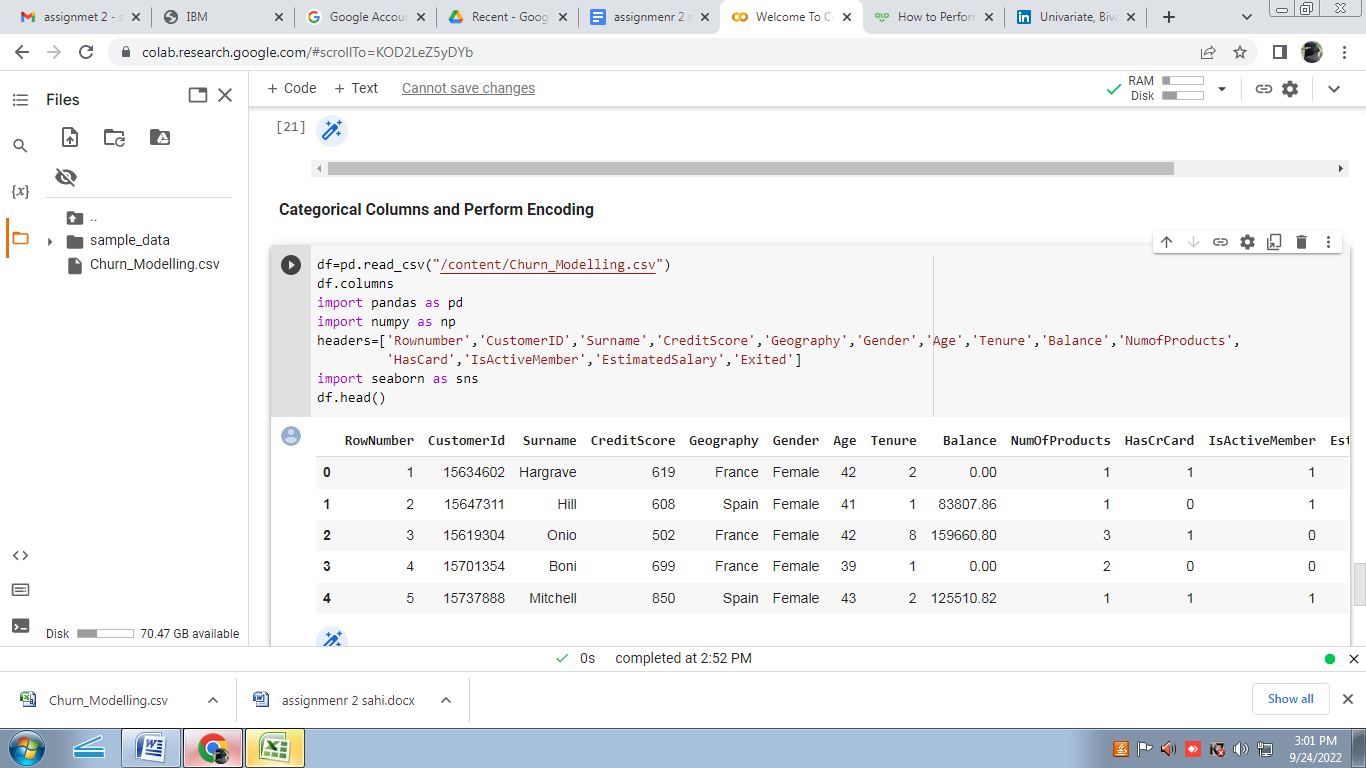
headers=['Rownumber','CustomerID','Surname','CreditScore',

'Geography','Gender','Age','Tenure','Balance','NumofProducts',

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

import seaborn as sns

df.head()

****

**Question 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. 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. 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):,}"))

