**APPLIED DATA**

**SCIENCE**

DATA VISUALIZATION AND PRE-PROCESSING

ASSSIGNMENT-2

DONE BY.,

K.BALAJI

REG:961819106009

BATCH:B12-6A2E

from google.colab import files

uploaded=files.upload()

**Churn\_Modelling.csv**(text/csv) - 684858 bytes, last modified: 9/24/2022 - 100% done

Saving Churn\_Modelling.csv to Churn\_Modelling.csv

#Importing packages

import numpy as np

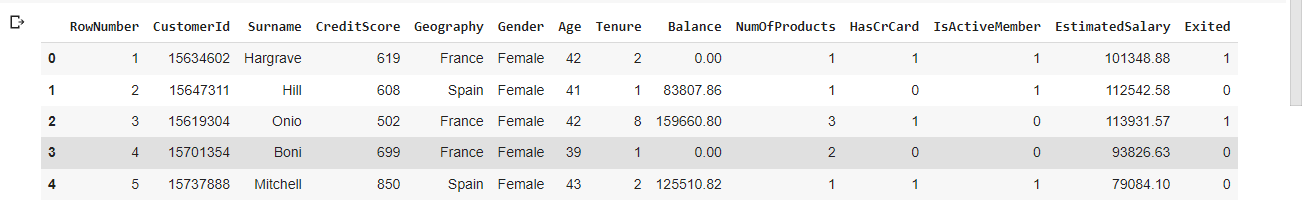
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data=pd.read\_csv(' Churn\_Modelling.csv')

data.head()



#Importing packages

import numpy as np

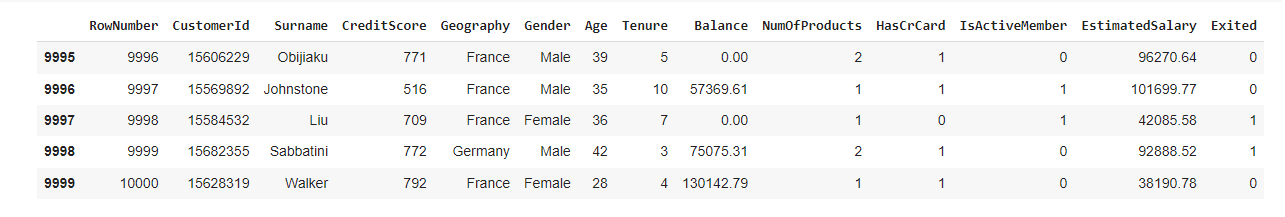
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data=pd.read\_csv('Churn\_Modelling.csv')

data.tail()



data.shape

(10000, 14)

data.info

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 10000 entries, 0 to 9999

Data columns (total 14 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 RowNumber 10000 non-null int64

1 CustomerId 10000 non-null int64

2 Surname 10000 non-null object

3 CreditScore 10000 non-null int64

4 Geography 10000 non-null object

5 Gender 10000 non-null object

6 Age 10000 non-null int64

7 Tenure 10000 non-null int64

8 Balance 10000 non-null float64

9 NumOfProducts 10000 non-null int64

10 HasCrCard 10000 non-null int64

11 IsActiveMember 10000 non-null int64

12 EstimatedSalary 10000 non-null float64

13 Exited 10000 non-null int64

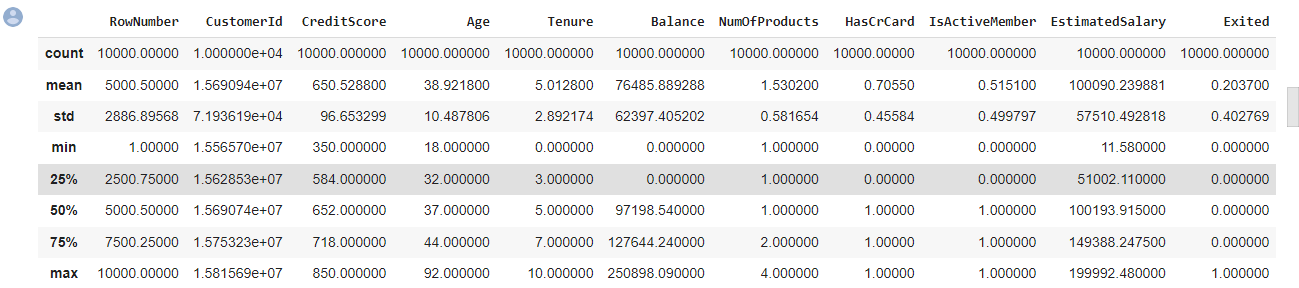
dtypes: float64(2), int64(9), object(3)

memory usage: 1.1+ MB

data.isnull().sum()



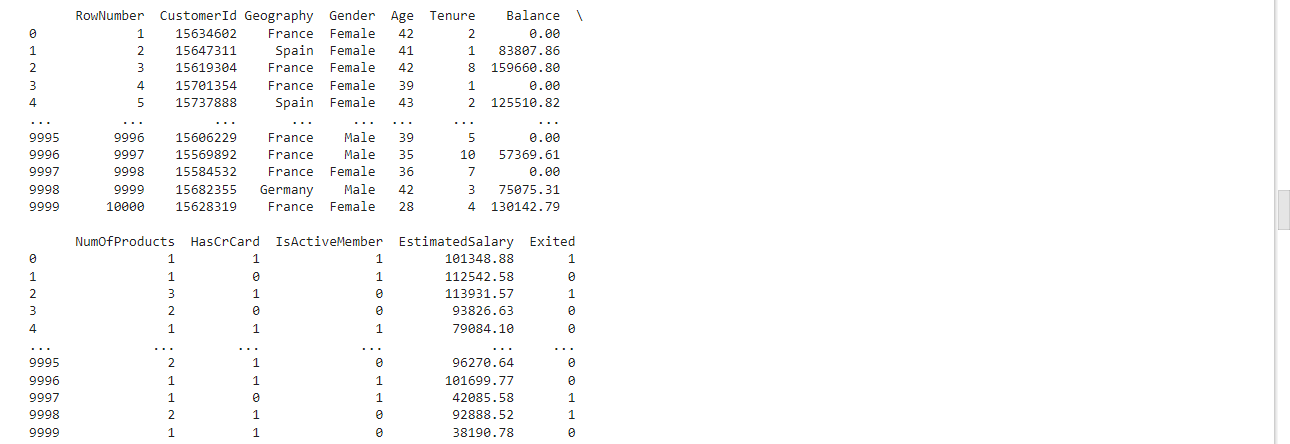
data.describe()



x=data.drop(['Surname','CreditScore'],axis=1)

y=data['CreditScore']

print(x)



print(y)

0 619

1 608

2 502

3 699

4 850

...

9995 771

9996 516

9997 709

9998 772

9999 792

Name: CreditScore, Length: 10000, dtype: int64

#splitting Training and Test data

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

from sklearn.linear\_model import LinearRegression

x\_train,x\_test,Y\_train,Y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=2)

plt.xlabel("Balance")

plt.ylabel("EstimatedSalary")

plt.title("Balance Vs EstimatedSalary ")

plt.show

<function matplotlib.pyplot.show(\*args, \*\*kw)>



data=pd.get\_dummies(data,columns=['Geography','Gender',])

print(data)

