## Assignment -2

# Data Visualization and Pre-processing

| | |
|---|---|
| Assignment Date | 5 September 2022 |
| Student name | SHANMUGAM M |
| Student roll number | **112519106004** |
| Maximum marks | 2 marks |

```
import pandas as pd
```

```
import numpy as np import sklearn
data=pd.read_csv("/content/drive/MyDrive/New
folder/Churn_Modelling.csv") import seaborn as sn import
matplotlib.pyplot as plt data.head() data.tail()
plt.boxplot(data['Age']) sn.boxplot(data['Tenure'])
sn.displot(data['Age']) sn.lineplot(x="Age",y="Balance",data=data)
sn.barplot(x='Age',y='Balance',data=data)
sn.countplot(x='Age',data=data) corr_matrix=data.corr()
sn.heatmap(corr_matrix) data.describe() data.describe()


data.isnull().sum() import
seaborn as sn import seaborn as
sn
qnt=data.quantile(q=[0.25,0.75])
qnt
IQR =qnt.loc[0.75]-qnt.loc[0.25] IQR
upper_extreme=qnt.loc[0.75]+1.5*IQR
upper_extreme
lower_extreme=qnt.loc[0.25]-1.5*IQR
lower_extreme









from sklearn.impute import SimpleImputer imp
=SimpleImputer(missing_values=np.nan,strategy='main')
data[data['Age']>88] data[data['Age']>92]
data['Age']=np.where(data['Age']>88,data['Age'].mean(),data['Age'])
data[data['Age']>88]
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data['Surname']=le.fit_transform(data['Surname']) data.head()


y=data['Exited'] x=data.drop(columns=['Exited'],axis=1)
names=x.columns names from sklearn.preprocessing import scale x
x=pd.DataFrame(x,columns=names) x.head() 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)
x_train.shape x_train.shape
```