**ASSIGNMENT – II**

Data Visualization and Pre-processing:

Code:

import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from sklearn.preprocessing import OneHotEncoder  
import numpy as np  
from sklearn.model\_selection import cross\_val\_score, train\_test\_split  
  
df = pd.read\_csv('Churn\_Modelling.csv', header=0)  
# Univarite Analysis  
  
print(df.head())  
print(df.info())  
  
print(df[['Age', 'Tenure', 'Balance', 'CreditScore', 'IsActiveMember']].describe())  
print(sns.histplot(df.Age, kde=True))  
  
#Bivarite Analysis  
  
print(df[['Age', 'Tenure', 'Balance', 'CreditScore', 'IsActiveMember']].corr())  
sns.scatterplot(df.IsActiveMember)  
plt.ylim(0, 1)  
  
#Statiscal Informations  
  
print(df['CreditScore'].mean())  
print(df['CreditScore'].median())  
print(df['CreditScore'].mode())  
  
df = df[df.IsActiveMember != 0]  
df = df.drop('IsActiveMember', 1)  
  
print(df.head())  
  
#categorical columns  
  
df = df[['Gender', 'Age', 'CreditScore']]  
print(df['Gender'].unique())  
  
gender\_encoder = OneHotEncoder()  
gender\_reshaped = np.array(df['Gender']).reshape(-1, 1)  
gender\_values = gender\_encoder.fit\_transform(gender\_reshaped)  
  
print(df['Gender'][:5])  
print()  
print(gender\_values.toarray()[:5])  
print()  
print(gender\_encoder.inverse\_transform(gender\_values)[:5])  
  
labeled = df[['CreditScore', 'Age', 'Gender']]  
Y = labeled.iloc[:, 0].values  
X = labeled.iloc[:, 1:].values  
  
Y = Y.reshape(-1 ,1)  
  
print(Y)  
  
X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.3, random\_state=3)

Output:





