**Name:** Muzammil Rauf Patel

**Class:** SYMCA **Div:** B

**Roll No:** 45 **Batch:** B2

**Course:** ML Lab

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

**Assignment No: 1**

**Problem Statement:**

Perform the following operations using Python on  
the given data sets a) Importing the libraries b) Importing the Dataset c)  
Handling of Missing Data d) Handling of Categorical Data e) Splitting the  
dataset into training and testing datasets f) Feature Scaling

**------------------------------------------------------------------------------------------------**

**Code:**

import pandas

import pandas as pd

from sklearn.preprocessing import StandardScaler

from sklearn.preprocessing import LabelEncoder

from sklearn.linear\_model import LinearRegression

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

from sklearn.tree import DecisionTreeRegressor

data=pd.read\_csv('D:\Muzammil\ML\Assignmnet 1\Country.csv',na\_values=["?"])

print("Data: \n",data)

print("---------------------------------------")

print("\nMissing Data:")

print(data.isnull().sum())

print("---------------------------------------")

print("\nFilling Nan values with 0:")

df=data.fillna(value=0)

print(df)

data.describe()

data['Age'].fillna(data['Age'].mean(),inplace=True)

data.isnull().sum()

data['Income'].fillna(data['Income'].median(),inplace=True)

data.isnull().sum()

print("---------------------------------------")

print("Scaling")

scale = StandardScaler()

df = pd.read\_csv('D:\Muzammil\ML\Assignmnet 1\Country.csv',na\_values=["?"])

X = df[['Age','Income']]

scaledX = scale.fit\_transform(X)

print(scaledX)

print("---------------------------------------")

print("Categorial Data")

label = LabelEncoder()

df = label.fit\_transform(df['Region'])

print(df)

dataset=pd.read\_csv('D:\Muzammil\ML\Assignmnet 1\Country.csv')

print("---------------------------------------")

print("\nDataset Preview")

print(dataset)

X = dataset[['Region','Age','Income']]

y = dataset['Online Shopper']

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

print("---------------------------------------")

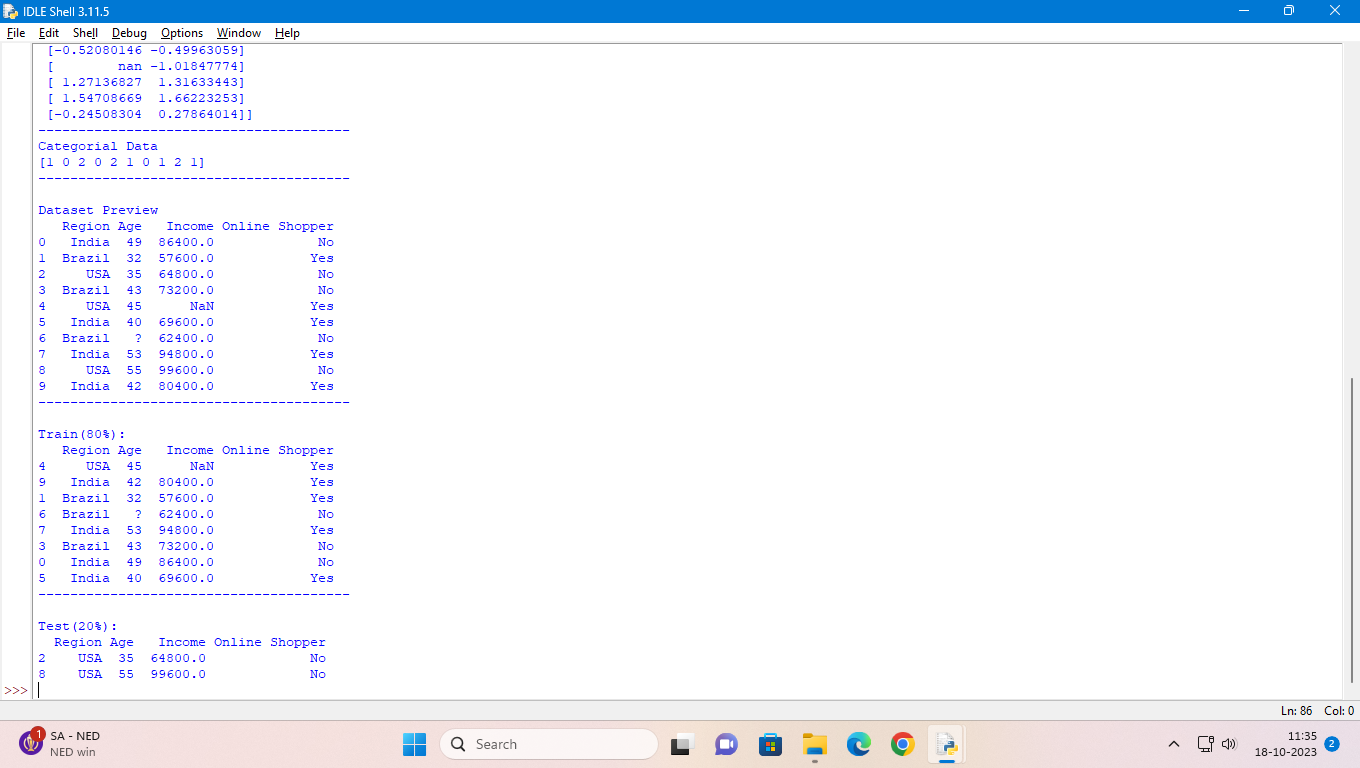
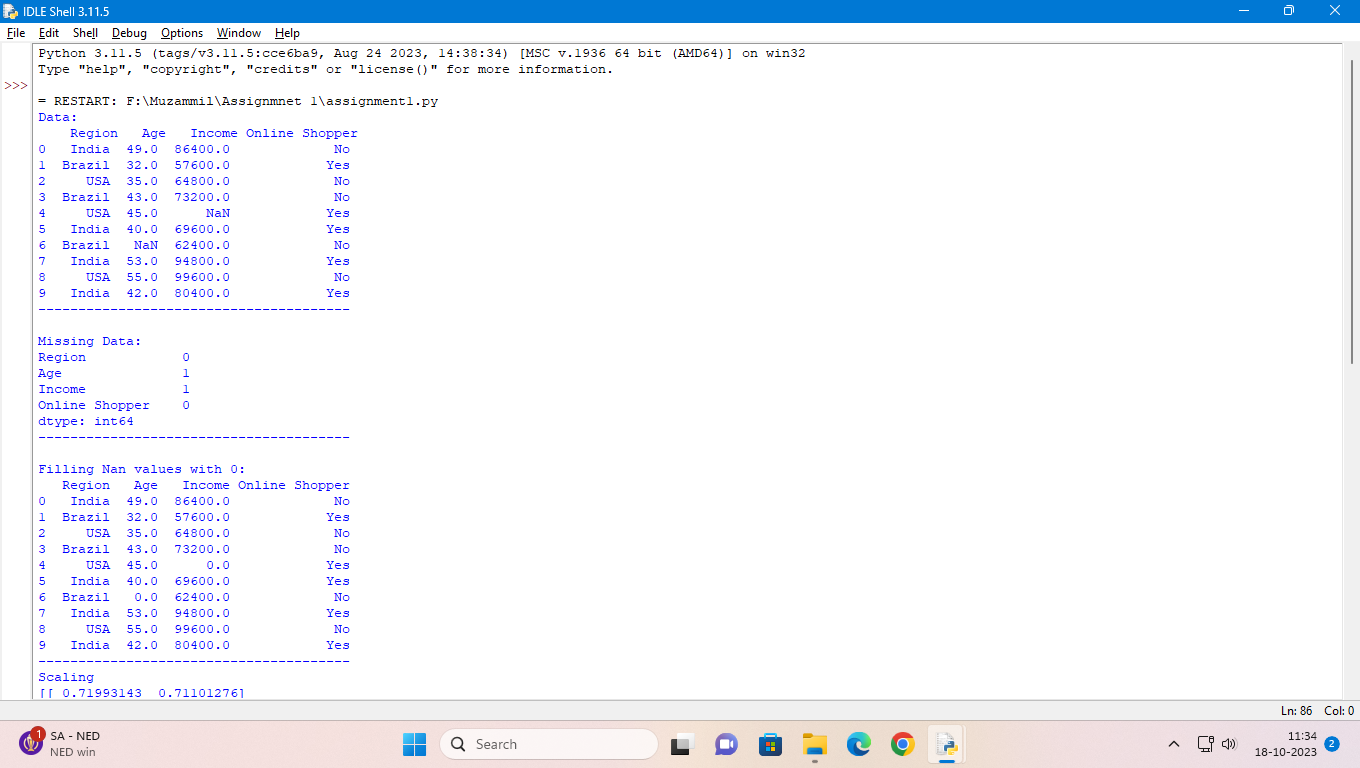
print("\nTrain(80%):")

print(pd.concat([X\_train,y\_train],axis=1))

print("---------------------------------------")

print("\nTest(20%):")

print(pd.concat([X\_test,y\_test],axis=1))

****