**NAME : MANGESH A. GHADWAJE**

**ROLL NO:24**

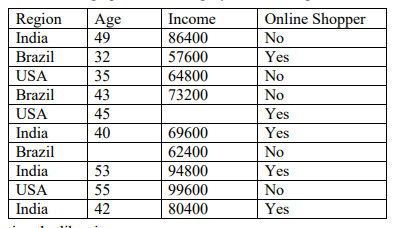
**BATCH : B2**

**COURSE: ML PRACTICAL**

**Assginment 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 :**

***#!pip install scikit-learn pandas numpy dtale notebook***

***import pandas as pd***

***import numpy as np***

***#import dtale***

***from sklearn.impute import SimpleImputer***

***from sklearn.compose import ColumnTransformer***

***from sklearn.preprocessing import LabelEncoder,OneHotEncoder***

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

***from sklearn.preprocessing import StandardScaler***

***data={' Region' : ['India','Brazil','USA','Brazil','USA','India','Brazil','India','nan','India'],***

***' Age':['49','32','35','43','45','40','nan','53','55','42'],***

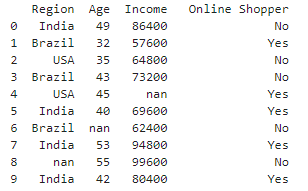
***' Income':['86400','57600','64800','73200','nan','69600','62400','94800','99600','80400'],***

***' Online Shopper':['No','Yes','No','No','Yes','Yes','No','Yes','No','Yes']***

***}***

***dataset=pd.DataFrame(data)***

***print(dataset)***

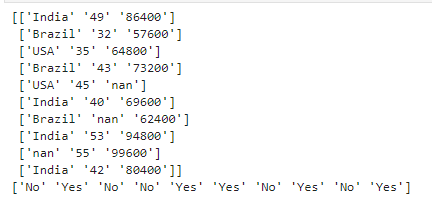


***X=dataset.iloc[:,:-1].values***

***Y=dataset.iloc[:, -1].values***

***print(X)***

***print(Y)***

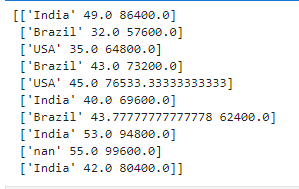


***imputer = SimpleImputer(missing\_values=np.nan,strategy="mean")***

***imputer = imputer.fit(X[:,1:])***

***X[:,1:] = imputer.transform(X[:,1:])***

***print(X)***



***labelencoder\_X = LabelEncoder()***

***X[:,0]=labelencoder\_X.fit\_transform(X[:,0])***

***ct=ColumnTransformer([("Region",OneHotEncoder(),[0])],remainder='passthrough')***

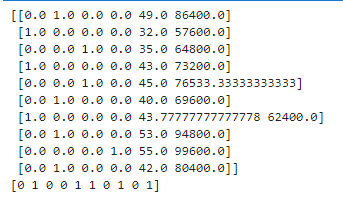
***X=ct.fit\_transform(X)***

***labelencoder\_Y=LabelEncoder()***

***Y=labelencoder\_Y.fit\_transform(Y)***

***print(X)***

***print(Y)***



***labelencoder\_X = LabelEncoder()***

***X[:,0]=labelencoder\_X.fit\_transform(X[:,0])***

***ct=ColumnTransformer([("Region",OneHotEncoder(),[0])],remainder='passthrough')***

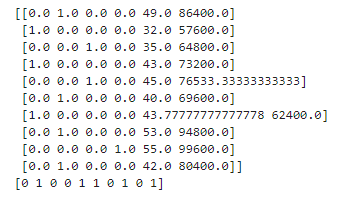
***X=ct.fit\_transform(X)***

***labelencoder\_Y=LabelEncoder()***

***Y=labelencoder\_Y.fit\_transform(Y)***

***print(X)***

***print(Y)***



***labelencoder\_X = LabelEncoder()***

***X[:,0]=labelencoder\_X.fit\_transform(X[:,0])***

***ct=ColumnTransformer([("Region",OneHotEncoder(),[0])],remainder='passthrough')***

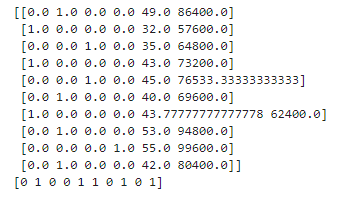
***X=ct.fit\_transform(X)***

***labelencoder\_Y=LabelEncoder()***

***Y=labelencoder\_Y.fit\_transform(Y)***

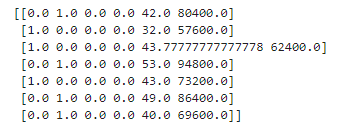
***print(X)***

***print(Y)***



***X\_train,X\_test,Y\_train,Y\_test = train\_test\_split(X,Y,test\_size=0.3,random\_state=0)***

***print(X\_train)***



***sc\_X = StandardScaler()***

***X\_train = sc\_X.fit\_transform(X\_train)***

***X\_test = sc\_X.transform(X\_test)***

***print(X\_train)***

