**Name : Dion D Rodrigues Assignment.no: 02**

**Roll.no : 53 Class : FYMCA B Subject : ML Lab Batch : B3**

**# IMPORTING LIBRARIES**

***import pandas as pd***

***import numpy as np***

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

***from sklearn.preprocessing import OneHotEncoder***

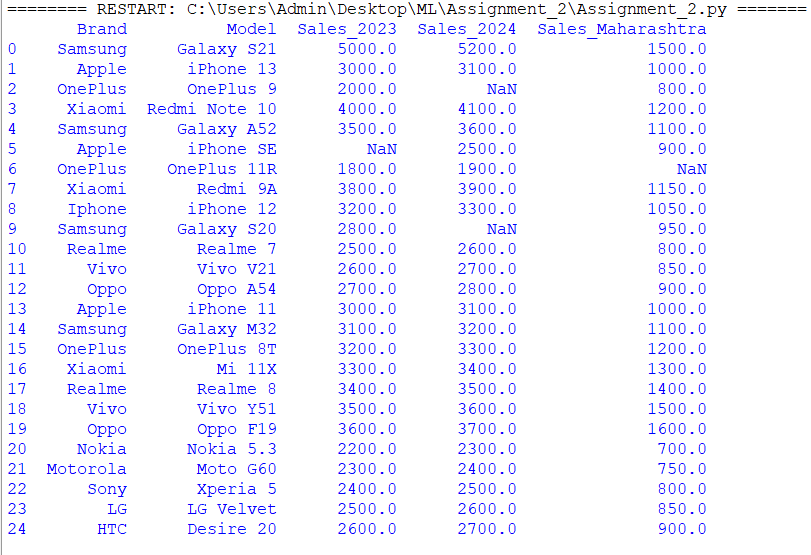
***from sklearn.preprocessing import StandardScaler***

***from sklearn.decomposition import PCA***

**#IMPORTING DATASET**

***data = pd.read\_csv('C:/Users/Admin/Desktop/ML/Assignment\_2/mobile\_sales.csv')***

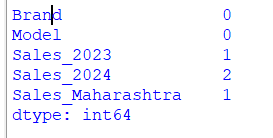
***print(data)***

****

**#CHECKING FOR MISSING DATA**

***print()***

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

****

**#HANDLING MISSING VALUES IF ANY**

***numerical\_col= data.select\_dtypes(include=[np.number]).columns***

***imputer = SimpleImputer(strategy='median')***

***data[numerical\_col] = imputer.fit\_transform(data[numerical\_col])***

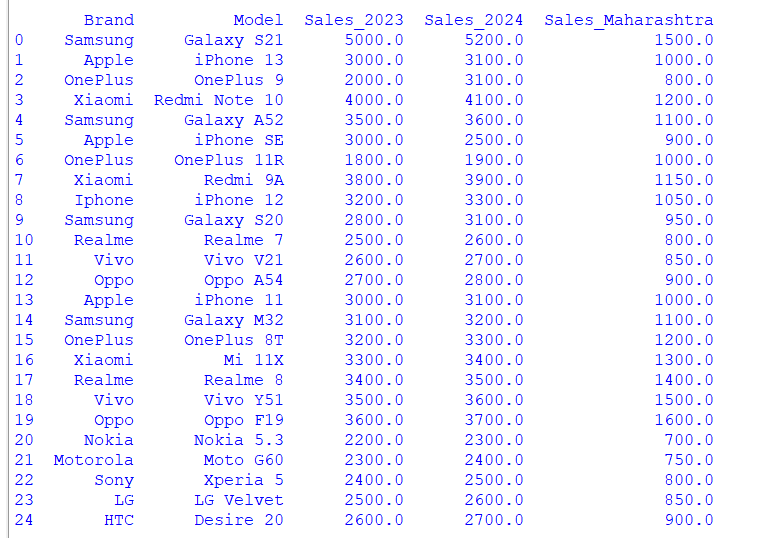
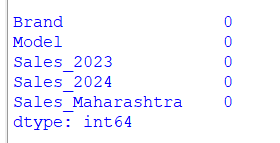
**#VERIFY IF MISSING VALUES ARE HANDLED**

***print()***

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

***print()***

***print(data)***

****

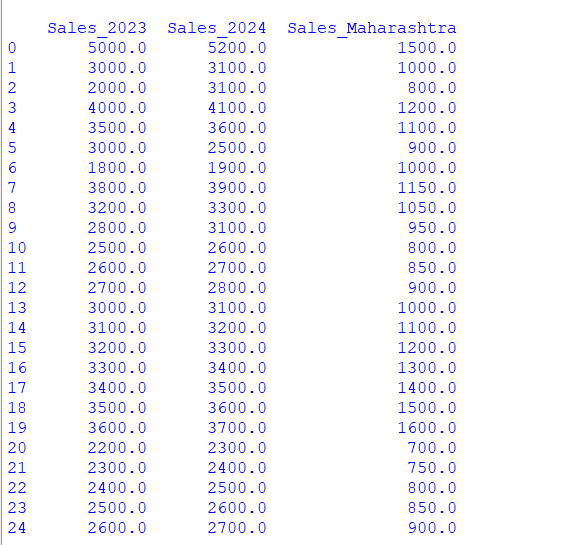
**#SELECT NUMERICAL COLUMNS (REPLACE WITH ACTUAL COLUMN NAMES)**

***numerical\_features = ['Sales\_2023','Sales\_2024','Sales\_Maharashtra']***

***x= data[numerical\_features]***

***print()***

***print(x)***

******

**#STANDARDIZE THE DATA**

**x= StandardScaler().fit\_transform(x)**

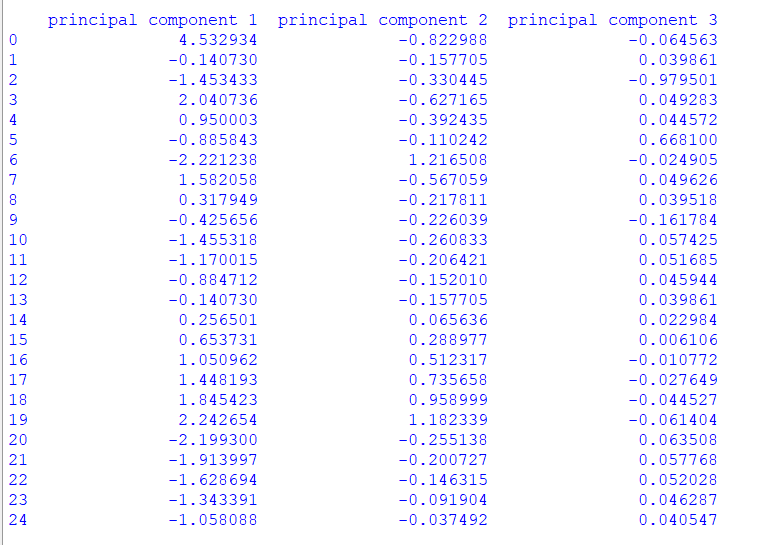
**pca= PCA(n\_components=3) #choose the number of components you want to keep**

**principalComponents = pca.fit\_transform(x)**

**principalDf = pd.DataFrame(data = principalComponents , columns =['principal component 1','principal component 2', 'principal component 3])**

**print()**

**print(principalDf)**

****