Name: siddhesh shende

Class: SY-A Roll No: 59 Batch: A3

Sub: ML

**Assignment No: 2 (MLL)**

**code:**

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

# Load the data

data = pd.read\_csv('C:/Users/MCA3/Desktop/SY-B(MLL)/iris.csv')

print(data)

# Check for missing values

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

# Handling missing data by replacing with the mean (for numerical columns)

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

imputer = SimpleImputer(strategy='mean')

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

# Verify if missing data is handled

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

# Select numerical columns (replace with actual column names)

numerical\_features = ['Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width']

x = data[numerical\_features]

print(x)

# Standardize the data

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

pca = PCA(n\_components=2) # 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'])

print(principalDf)

OUTPUT:



