Name : Dion D Rodrigues Assignment.no: 09

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

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

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import AgglomerativeClustering

from scipy.cluster import hierarchy

from sklearn.metrics import silhouette\_score

# Reading the Data

data = pd.read\_csv('IRIS.csv')

# Describing the Data

print(data.head())

print(data.info())

print(data.describe())

print(data[['species']].describe())

# Checking for null values in the data

print("\nMissing values:\n", data.isnull().sum())

# Splitting the data (excluding the species column)

X = data.iloc[:, :-1].values

# Constructing a Dendrogram

plt.figure(figsize=(15, 10))

plt.title("Dendrogram for Iris Dataset")

dendrogram = hierarchy.dendrogram(hierarchy.linkage(X, method='ward'))

plt.xlabel('Sample Index')

plt.ylabel('Euclidean Distance')

plt.show() # Display the dendrogram

# Training the model using Agglomerative Clustering

hc = AgglomerativeClustering(n\_clusters=3, metric='euclidean', linkage='ward')

# Predicting clusters

pred = hc.fit\_predict(X)

# Visualizing the Clustering Result

plt.figure(figsize=(10, 7))

plt.scatter(X[pred == 0, 0], X[pred == 0, 1], c='green', label='Cluster 1')

plt.scatter(X[pred == 1, 0], X[pred == 1, 1], c='blue', label='Cluster 2')

plt.scatter(X[pred == 2, 0], X[pred == 2, 1], c='red', label='Cluster 3')

plt.xlabel('Sepal Length')

plt.ylabel('Sepal Width')

plt.title('Sepal Length vs Sepal Width (Clustered)')

plt.legend()

plt.show() # Display the clustering visualization

# Evaluating with silhouette score

score = silhouette\_score(X, pred)

print(f"Silhouette Score: {score:.2f}")





