**NAME : MANGESH A. GHADWAJE**

**ROLL NO:24**

**BATCH : B2**

**COURSE: ML PRACTICAL**

**Assginment No. 9**

**Problem Statement :**

**Implement Hierarchical clustering on the shopping trends data set.**

**Code :**

***import numpy as np***

***import pandas as pd***

***import matplotlib.pyplot as plt***

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

***from scipy.cluster.hierarchy import dendrogram, linkage***

***from sklearn.cluster import AgglomerativeClustering***

***from sklearn.metrics import silhouette\_score***

***# Load the dataset (replace 'shopping\_trends.csv' with your actual file)***

***df = pd.read\_csv('shopping\_trends.csv')***

***# Data preprocessing***

***df = pd.get\_dummies(df, drop\_first=True) # One-hot encoding***

***df.fillna(df.mean(), inplace=True) # Fill missing values with column mean***

***# Standardize the data***

***X\_scaled = StandardScaler().fit\_transform(df)***

***# Generate the linkage matrix and plot dendrogram***

***linked = linkage(X\_scaled, method='ward')***

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

***dendrogram(linked)***

***plt.title('Dendrogram for Hierarchical Clustering')***

***plt.xlabel('Samples')***

***plt.ylabel('Euclidean Distance')***

***plt.show()***

***# Apply Agglomerative Clustering***

***n\_clusters = 3***

***hc = AgglomerativeClustering(n\_clusters=n\_clusters, linkage='ward')***

***df['Cluster'] = hc.fit\_predict(X\_scaled)***

***# Calculate silhouette score***

***silhouette\_avg = silhouette\_score(X\_scaled, df['Cluster'])***

***print(f'Silhouette Score: {silhouette\_avg:.2f}')***

***# Scatter plot for clusters based on original features***

***plt.figure(figsize=(8, 5))***

***plt.scatter(df.loc[df['Cluster'] == 0].iloc[:, 0], df.loc[df['Cluster'] == 0].iloc[:, 1], s=50, c='red', label='Cluster 0')***

***plt.scatter(df.loc[df['Cluster'] == 1].iloc[:, 0], df.loc[df['Cluster'] == 1].iloc[:, 1], s=50, c='blue', label='Cluster 1')***

***plt.scatter(df.loc[df['Cluster'] == 2].iloc[:, 0], df.loc[df['Cluster'] == 2].iloc[:, 1], s=50, c='green', label='Cluster 2')***

***plt.title('Clusters of Shopping Trends Data')***

***plt.xlabel(df.columns[0])***

***plt.ylabel(df.columns[1])***

***plt.legend()***

***plt.grid()***

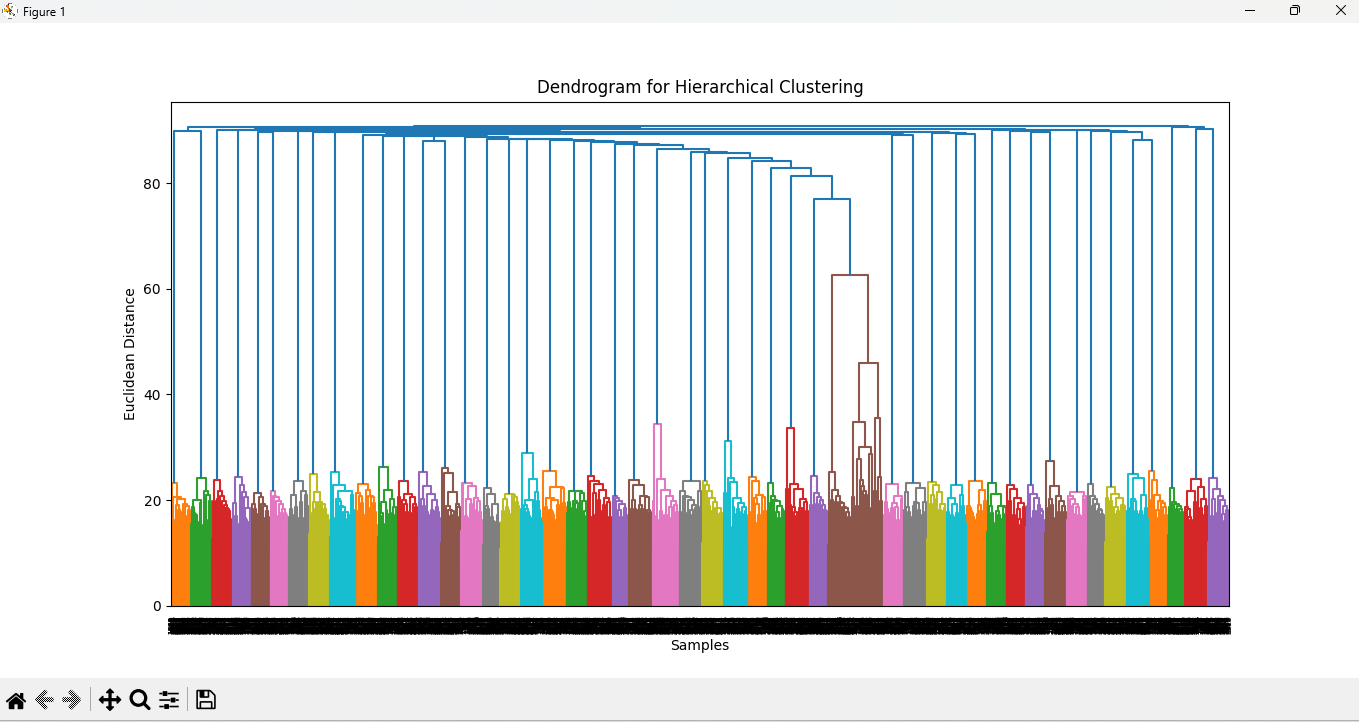
***plt.show()***

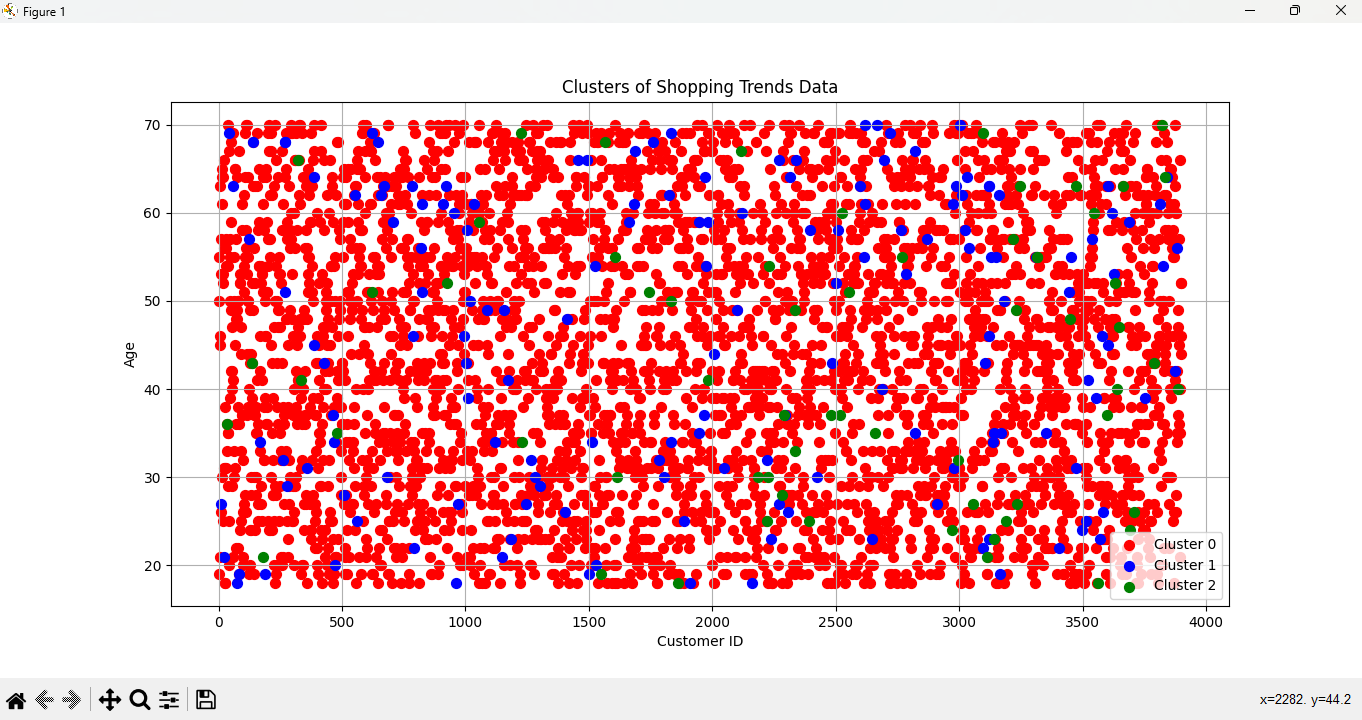
***# Print cluster counts***

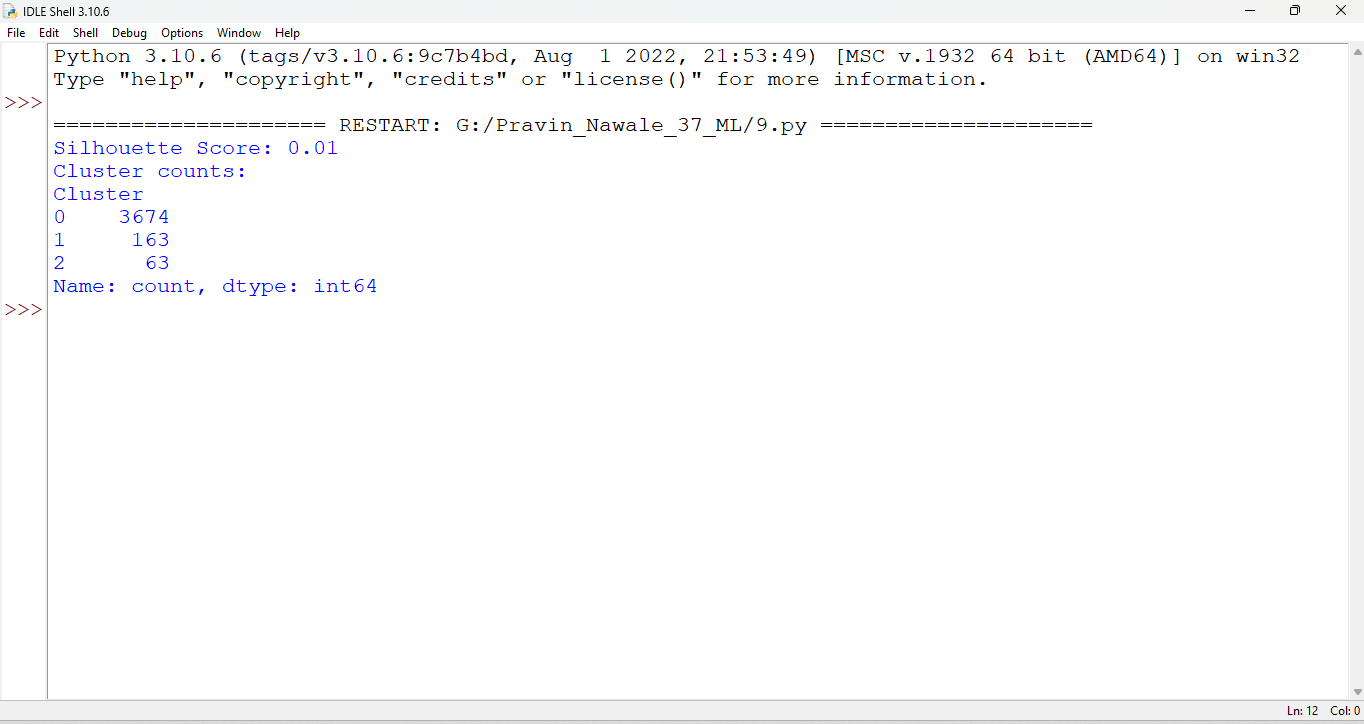
***print("Cluster counts:")***

***print(df['Cluster'].value\_counts())***

**Output :**

****

****

****