Name : Dion D Rodrigues Assignment.no: 07

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

import pandas as pd

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

# Load the dataset

file\_path = 'customer\_spending\_data.csv'

data = pd.read\_csv(file\_path)

# Extract the relevant features (ignoring the 'Cluster' column)

X = data[['Spending on Food', 'Spending on Clothing']]

# Initialize K-Means with a reasonable number of clusters (let's start with 3)

kmeans = KMeans(n\_clusters=3, random\_state=42)

# Fit the model

kmeans.fit(X)

# Predict the cluster labels

clusters = kmeans.predict(X)

# Add the new cluster labels to the dataset

data['New Cluster'] = clusters

# Plotting the results to visualize the clusters

plt.figure(figsize=(8, 6))

plt.scatter(X['Spending on Food'], X['Spending on Clothing'], c=clusters, cmap='viridis', s=50)

plt.scatter(kmeans.cluster\_centers\_[:, 0], kmeans.cluster\_centers\_[:, 1], s=200, c='red', label='Centroids')

plt.title("K-Means Clustering of Customer Spending")

plt.xlabel("Spending on Food")

plt.ylabel("Spending on Clothing")

plt.legend()

plt.grid(True)

plt.show()

# Step 1: Display the cluster centers (centroids)

centroids = kmeans.cluster\_centers\_

print("Cluster Centers (Centroids):")

print(pd.DataFrame(centroids, columns=['Spending on Food', 'Spending on Clothing']))

# Step 2: Explore the distribution of customers in each cluster

cluster\_distribution = data['New Cluster'].value\_counts()

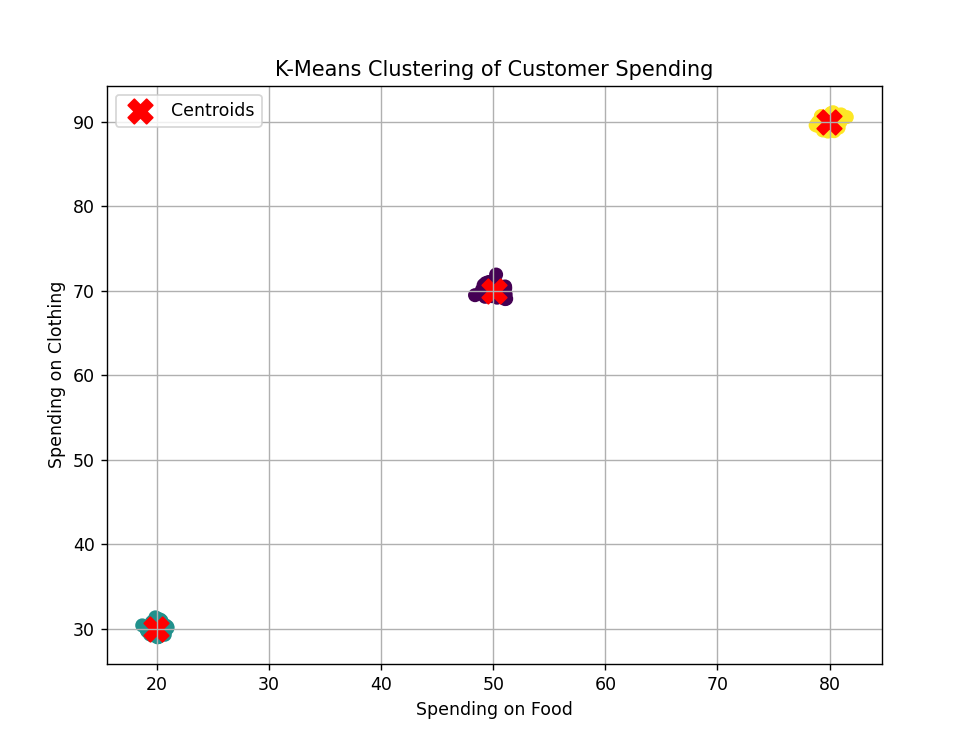
print("\nCustomer Distribution Across Clusters:")

print(cluster\_distribution)

# Step 3: Display the first few rows of the updated dataset

print("\nUpdated Dataset with Cluster Labels:")

print(data.head())



print(data.head()) 