Name: Aarya Tiwari

RollNo.: 16010421119

EXP 4

K-means clustering algorithm

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn.metrics import silhouette\_score

data = pd.read\_csv('Wholesale customers data.csv')

# Elbow Method

wcss = []

for i in range(1, 11):

    kmeans = KMeans(n\_clusters=i, init='k-means++', max\_iter=300, n\_init=10, random\_state=0)

    kmeans.fit(data)

    wcss.append(kmeans.inertia\_)

# Plot the Elbow Method graph

plt.figure(figsize=(8, 4))

plt.plot(range(1, 11), wcss)

plt.title('Elbow Method')

plt.xlabel('Number of clusters (K)')

plt.ylabel('WCSS')

plt.show()

# Silhouette Score

silhouette\_scores = []

for k in range(2, 11):

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

    kmeans.fit(data)

    score = silhouette\_score(data, kmeans.labels\_)

    silhouette\_scores.append(score)

plt.figure(figsize=(8, 4))

plt.plot(range(2, 11), silhouette\_scores)

plt.title('Silhouette Score')

plt.xlabel('Number of clusters (K)')

plt.ylabel('Silhouette Score')

plt.show()

chosen\_k\_elbow = 4  chosen\_k\_silhouette = 4  kmeans = KMeans(n\_clusters=chosen\_k\_elbow, random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

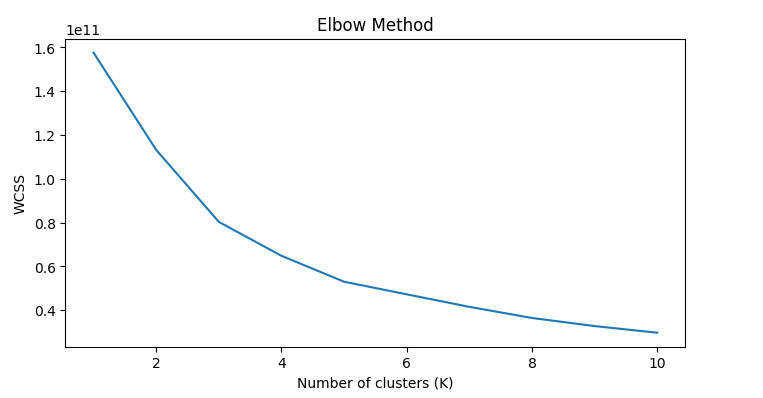
plt.scatter(data['Frozen'], data['Grocery'], c=labels)

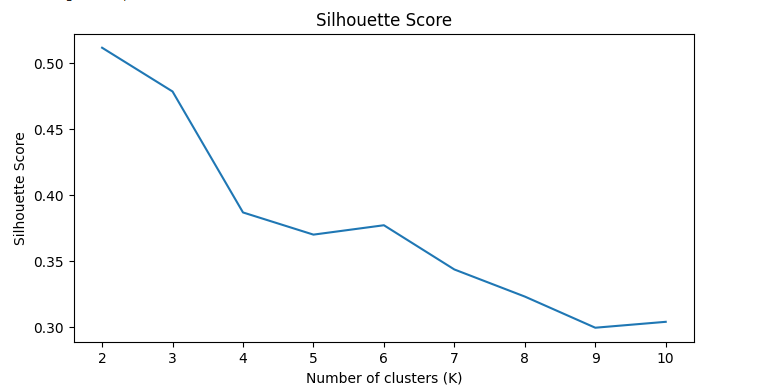
plt.xlabel('Frozen')

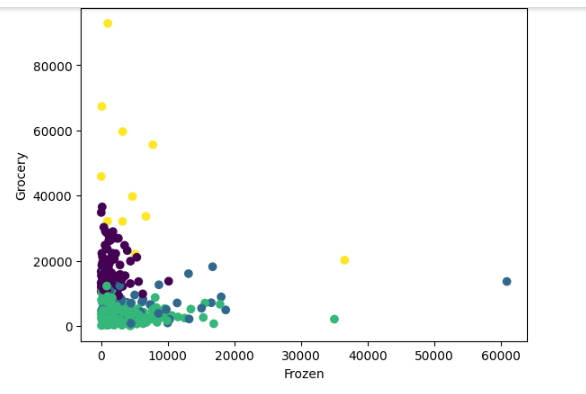
plt.ylabel('Grocery')

plt.show()

Output







1. What is the main difference between K-means and K-nearest neighbours?

K-means and K-nearest neighbors (KNN) are two popular machine learning algorithms with distinct purposes and operational approaches.

K-means is an unsupervised clustering algorithm that groups unlabeled data points into K clusters based on their similarity. It starts by selecting K centroids randomly, assigns data points to the nearest centroid, and updates centroids iteratively until convergence.

K-nearest neighbors (KNN), on the other hand, is a supervised algorithm used for classification and regression. It determines the class or value of a new data point by considering the classes of its K nearest neighbors in the training dataset, using distance metrics like Euclidean or Manhattan distance.

Despite their differences, both K-means and KNN share commonalities. They are iterative algorithms and rely on distance metrics to analyze data, such as Euclidean or Manhattan distance.

K-means is for unsupervised clustering of unlabeled data, while KNN is for supervised classification or regression of labeled data. Their commonalities include iterative processes and the use of distance metrics.

1. What are some stopping criteria for k-means clustering?

Here are some common stopping criteria for K-means clustering:

1. Centroid Convergence: Stop when cluster centroids no longer change significantly.

2. Maximum Iterations: Set a limit on the number of iterations.

3. Change in Cluster Assignments: Stop when data point assignments stabilize.

4. Inertia or Distortion: Monitor the sum of squared distances within clusters and stop when it no longer decreases significantly.

Outcome: CO Comprehend radial-basis-function (RBF) networks and Kernel learning method

Conclusion : we learnt about, K-means , an unsupervised clustering algorithm that groups unlabeled data points into K clusters based on their similarity.