EXPERIMENT 10

Clustering

Aim:

To understand the importance of clustering.

Algorithm:

1. Import the required libraries — numpy, pandas, matplotlib, seaborn, and KMeans from sklearn.cluster.

2. Load the Mall Customers dataset using pd.read\_csv() and display basic information.

3. Visualize relationships between features using pair plots.

4. Select relevant features — Annual Income and Spending Score — for clustering.

5. Apply the K-Means algorithm with 5 clusters and assign cluster labels to the dataset.

6. Plot a scatter plot showing customer segments based on income and spending score.

7. Calculate WCSS (Within-Cluster Sum of Squares) for cluster counts from 1 to 10.

8. Plot the Elbow Method graph to determine the optimal number of clusters.

Program:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.cluster import KMeans

df = pd.read\_csv("C:\Mall\_Customers.csv")

print(df.info())

print(df.head())

sns.pairplot(df)

plt.show()

X = df[['Annual Income (k$)', 'Spending Score (1-100)']].values

kmeans = KMeans(n\_clusters=5, n\_init='auto', random\_state=42)

labels = kmeans.fit\_predict(X)

df['Cluster'] = labels

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

sns.scatterplot(x='Annual Income (k$)',y='Spending Score (1-100)',hue='Cluster',data=df,palette='tab10',s=100)

plt.title("Customer Segments based on Income & Spending")

plt.legend(title='Label', bbox\_to\_anchor=(1.05, 1), loc='upper left') # moves legend outside

plt.tight\_layout()

plt.show()

wcss = []

for i in range(1, 11):

kmeans = KMeans(n\_clusters=i, n\_init='auto', random\_state=42)

kmeans.fit(X)

wcss.append(kmeans.inertia\_)

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

plt.plot(range(1, 11), wcss, marker='o', linestyle='--')

plt.title("Elbow Method for Optimal K")

plt.xlabel("Number of Clusters")

plt.ylabel("WCSS")

plt.show()

Output:

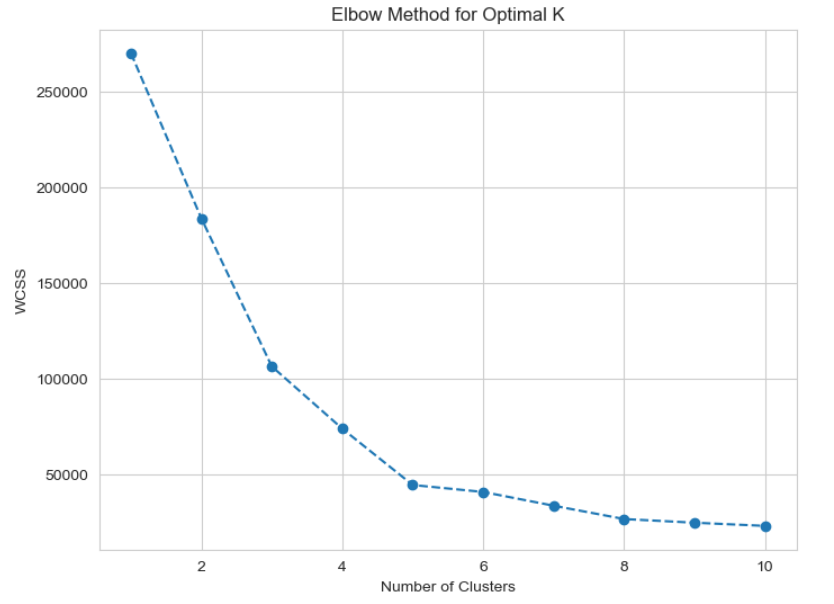
A screenshot of a computer

AI-generated content may be incorrect.

A group of blue dots

AI-generated content may be incorrect.





Result:

Hence a python program for clustering is written and executed successfully.