Question31:

31. Scenario: You work as a data scientist for an e-commerce company that sells a wide range of

products online. The company collects vast amounts of data about its customers, including their

purchase history, browsing behavior, demographics, and more. The marketing team wants to

understand their customer base better and improve their targeted marketing strategies. They have

asked you to perform customer segmentation using clustering techniques to identify distinct groups

of customers with similar characteristics.

Question: Your task is to use Python and clustering algorithms to segment the customers into

different groups based on their behavior and characteristics. The marketing team will use these

segments to tailor their marketing campaigns and promotions effectively.

Answer:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.cluster import KMeans

from sklearn.decomposition import PCA

file\_path =r"C:\Users\jampa\Downloads\customer\_segmentation\_data.csv"

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

features = ['Age', 'Annual\_Income', 'Spending\_Score', 'Online\_Sessions\_per\_Month', 'Average\_Order\_Value', 'Total\_Purchases', 'Customer\_Lifetime\_Value']

X = df[features]

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

inertia = []

cluster\_range = range(2, 11)

for k in cluster\_range:

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

kmeans.fit(X\_scaled)

inertia.append(kmeans.inertia\_)

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

plt.plot(cluster\_range, inertia, marker='o', linestyle='-', color='teal')

plt.title('Elbow Method for Optimal k')

plt.xlabel('Number of Clusters (k)')

plt.ylabel('Inertia')

plt.xticks(cluster\_range)

plt.show()

optimal\_k = 5

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

kmeans.fit(X\_scaled)

df['Cluster'] = kmeans.labels\_

pca = PCA(n\_components=2)

X\_pca = pca.fit\_transform(X\_scaled)

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

sns.scatterplot(x=X\_pca[:, 0], y=X\_pca[:, 1], hue=df['Cluster'], palette='viridis', s=60, alpha=0.8)

plt.title('Customer Segmentation Clusters (PCA)')

plt.xlabel('PCA Component 1')

plt.ylabel('PCA Component 2')

plt.legend(title='Cluster')

plt.show()

df.to\_csv('segmented\_customers.csv', index=False)

print('Customer segmentation complete. Results saved to segmented\_customers.csv.')

Output:



