EXP NO: 7	CLUSTERING USING K-MEANS AND ENSEMBLE METHODS
DATE: 18/9/25	

Aim:

To perform clustering on customer and wine datasets using K-Means and to combine multiple clustering results using CSPA ensemble clustering for improved partitioning.

Program:

Step 1: Import Libraries

```
import kagglehub import pandas as pd import numpy as np
import matplotlib.pyplot as plt import seaborn as sns
from sklearn.cluster import KMeans, SpectralClustering
from sklearn.preprocessing import StandardScaler from
sklearn.decomposition import PCA from sklearn.datasets
import load_wine
from sklearn.metrics import silhouette_score
```

Step 2: Load Dataset

```
path = kagglehub.dataset_download("shwetabh123/mall-customers") df
= pd.read_csv("/kaggle/input/mall-customers/Mall_Customers.csv")
df.head()
```

Step 3: Apply K-Means Clustering

```
kmeans = KMeans(n_clusters=5, random_state=42)
df["Cluster"] = kmeans.fit_predict(df[["Annual Income (k$)", "Spending Score (1-
100)"]])
df.head()
```

Step 4: Elbow Method for Optimal Clusters

```
distortions = [] for i in   range(1, 11):   km
 = KMeans(n_clusters=i, random_state=42)
    km.fit(df[["Annual Income (k$)", "Spending Score (1-100)"]])
distortions.append(km.inertia_)

plt.plot(range(1, 11), distortions, marker="o")
plt.title("Elbow Method") plt.xlabel("Number of
Clusters") plt.ylabel("Inertia") plt.show()
```

Step 5: Visualize Clusters

```
plt.figure(figsize=(8, 6))
sns.scatterplot(
data=df,
(k$)x="Annual Income
y="Spending Score (1-100)",
hue="Cluster",
palette="Set2",     s=80 )
```

```
plt.title("Customer Segments (KMeans)") plt.show()
```

Step 6: Load Wine Dataset

```
wine = load_wine()
X = pd.DataFrame(wine.data, columns=wine.feature_names)
X_scaled = StandardScaler().fit_transform(X)
```

Step 7: Create Base Clusterings

```
base_clusterings = [] for k in [3, 4, 5]:
km = KMeans(n_clusters=k, random_state=42)
    base_clusterings.append(km.fit_predict(X_scaled))
```

Step 8: Define CSPA Ensemble Function

```
def   cspa_ensemble(clusterings):
n_samples = len(clusterings[0])
    similarity_matrix = np.zeros((n_samples, n_samples))
     for clustering in clusterings:         for i
in           range(n_samples):    for j in
range(n_samples):           ==     if clustering[i]
clustering[j]:
similarity_matrix[i][j] += 1

    similarity_matrix /= len(clusterings)

    ensemble_labels = SpectralClustering(
        n_clusters=3, affinity="precomputed", random_state=42
    ).fit_predict(similarity_matrix)

    return ensemble_labels
```

Step 9: Apply Ensemble Clustering and Evaluate

```
ensemble_labels = cspa_ensemble(base_clusterings)
print("Silhouette Score:", silhouette_score(X_scaled, ensemble_labels))
```

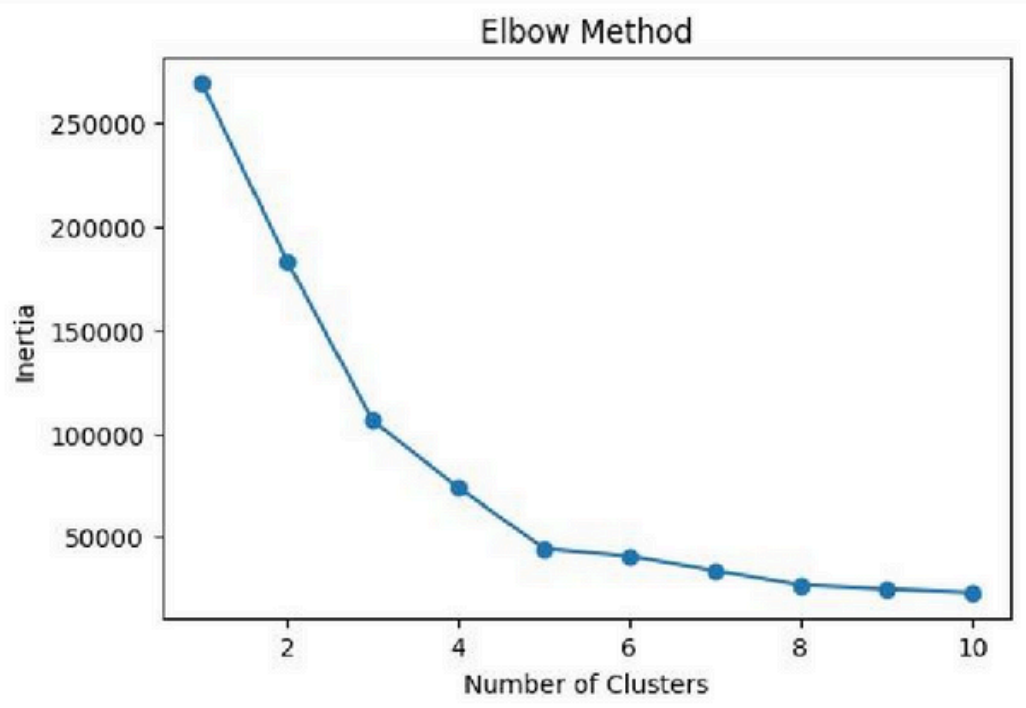
Step 10: Visualize Ensemble Clusters

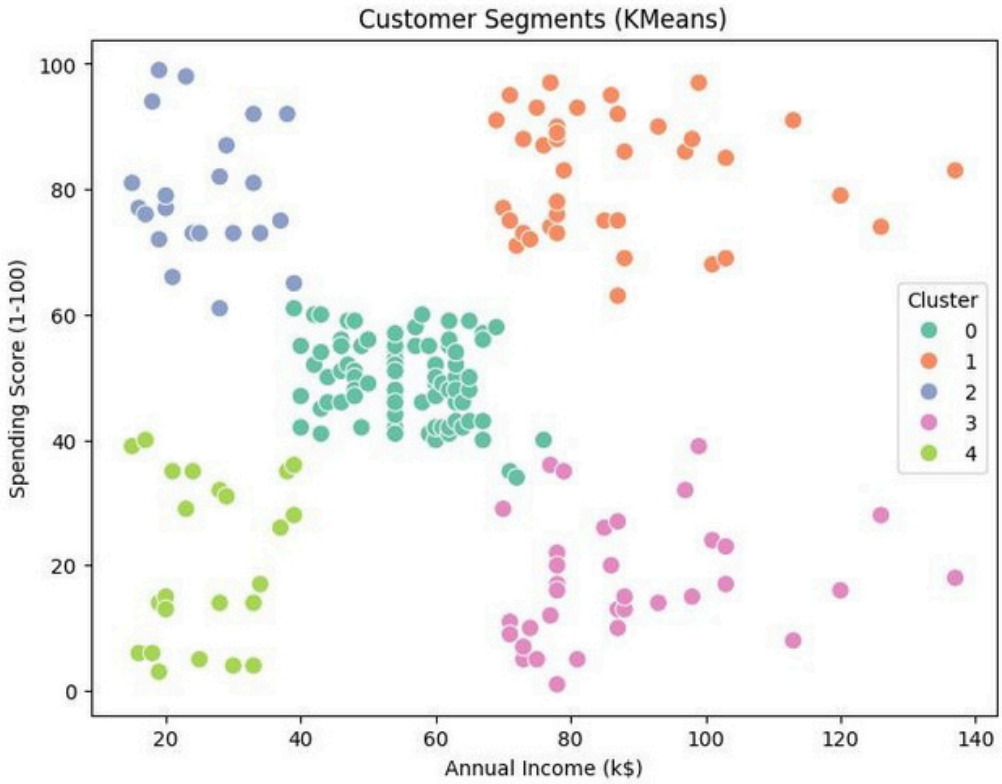
```
pca = PCA(n_components=2) X_pca =
pca.fit_transform(X_scaled)

plt.figure(figsize=(10, 6)) plt.scatter(
    X_pca[:, 0], X_pca[:, 1],
    c=ensemble_labels, cmap="viridis", s=50, edgecolor="k"
)
plt.title("CSPA Ensemble Clustering on Wine Dataset (PCA-reduced)")
plt.xlabel("PCA Component 1") plt.ylabel("PCA Component 2")
plt.colorbar(label="Cluster Label") plt.grid(True)
plt.show()
```

Output:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

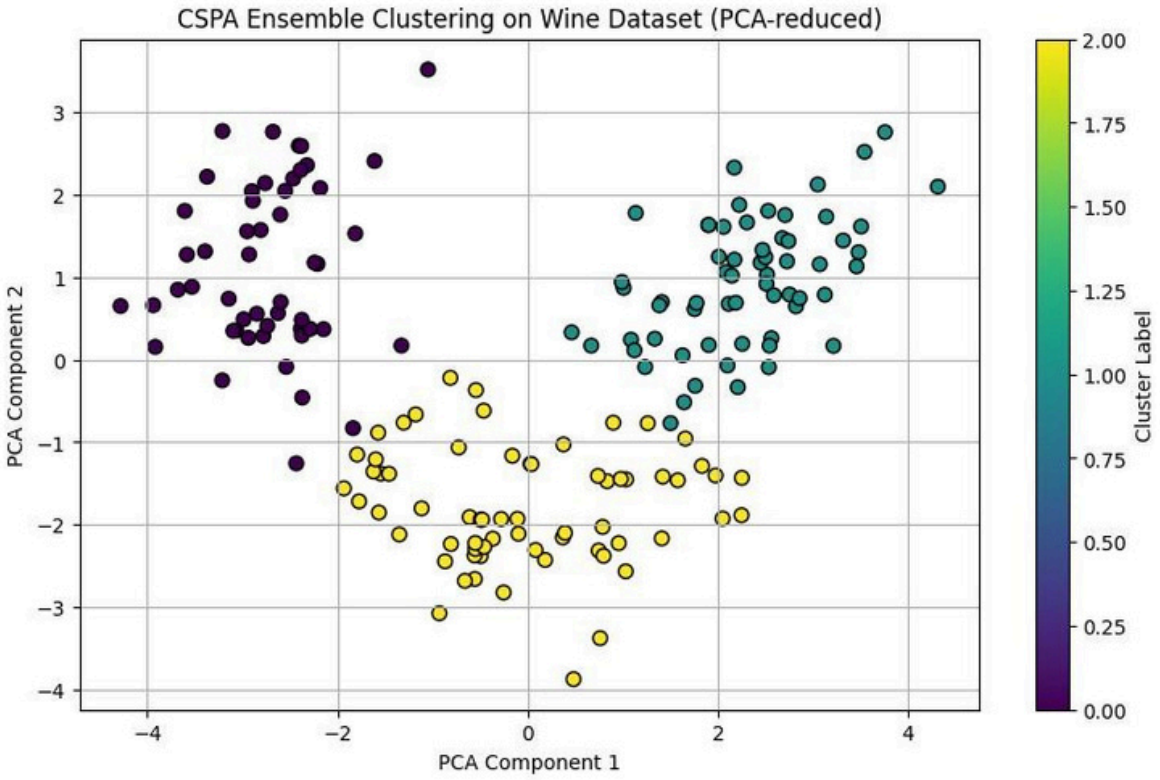
	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)	Cluster
0	1	Male	19	15	39	4
1	2	Male	21	15	81	2
2	3	Female	20	16	6	4
3	4	Female	23	16	77	2
4	5	Female	31	17	40	4





	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavanoid_phenols	proanthocyanins	color_intensity	hue	od280/od315_of_diluted_wines	proline
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065.0
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050.0
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185.0
3	14.37	1.95	2.50	16.8	113.0	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480.0
4	13.24	2.59	2.87	21.0	118.0	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735.0

```
Silhouette Score: 0.2848589191898987
```


CSPA Ensemble Clustering on Wine Dataset (PCA-reduced)
PCA Component 2
PCA Component 1
Cluster Label
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00

Result:

Customer groups and wine samples were successfully clustered using K-Means and ensemble methods, with visualization and silhouette analysis confirming the quality of clustering.