



Task 8:

Clustering with K-Means — Full Explanation + Output



Objective

Perform unsupervised learning using **K-Means clustering** and evaluate cluster quality.



Tools

scikit-learn

pandas

matplotlib

PCA (optional)



Step-by-Step Explanation

1 Load & Visualize Data

We will use the **Iris dataset** (3 natural clusters).

2 Fit K-Means

We train the model and assign cluster labels.

3 Use Elbow Method

Find the best number of clusters (K) by checking the “bend” in the WCSS graph.

4 Visualize Clusters

Use PCA to convert 4D → 2D for plotting.

5 Evaluate with Silhouette

Score

Higher score = better clustering

Range = **-1 to 1**

0.5+ is good.



Full Working Code (Copy-Paste Ready)

```
import pandas as pd

import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_score

# -----
# 1. Load the dataset

# -----
iris = load_iris()
X = pd.DataFrame(iris.data,
columns=iris.feature_names)

# -----
# 2. Elbow Method to find
optimal K

# -----
wcss = []
K_values = range(1, 10)

for k in K_values:
```

```

    kmeans =
KMeans(n_clusters=k,
random_state=42)
    kmeans.fit(X)

wcss.append(kmeans.inertia_)

plt.plot(K_values, wcss,
marker='o')
plt.title("Elbow Method")
plt.xlabel("Number of Clusters
(K) ")
plt.ylabel("WCSS")
plt.show()

# -----
# 3. Apply K-Means with 3
clusters

# -----

kmeans = KMeans(n_clusters=3,
random_state=42)
cluster_labels =
kmeans.fit_predict(X)

# -----

# 4. PCA for 2D visualization

# -----
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

plt.scatter(X_pca[:, 0],
X_pca[:, 1], c=cluster_labels)
plt.title("K-Means Clustering
(Iris Dataset)")
plt.xlabel("PCA 1")
plt.ylabel("PCA 2")
plt.show()

# -----

```

```
# 5. Evaluate with Silhouette  
Score
```

```
# -----  
  
score = silhouette_score(X,  
cluster_labels)  
print("Silhouette Score:",  
score)
```



Sample Output

1

Elbow Method Plot

You will see a graph where the WCSS drops sharply at **K = 3**, meaning **3 clusters** is optimal.



Your graph will look like this (example):

K = 1 → WCSS = very high

K = 2 → WCSS drops

K = 3 → Bend (best K)

K = 4-9 → gradual decrease

2

Cluster Visualization

Plot

You will see a 2D scatter plot (3 colors) showing the Iris clusters:

Cluster 0 → Setosa

Cluster 1 → Versicolor

Cluster 2 → Virginica

(Colors may vary)

3

Silhouette Score

Output

Example output when running:

Silhouette Score: 0.55

A score around **0.5+** means K-Means worked well.



Final Summary

Step

Optimal K (Elbow Method)

Result

PCA visualization

Silhouette Score

If you want, I can also provide:

3 clear clusters

~0.55 (good)