Unsupervised Machine Learning Demo using Hierarchical Clustering (Python)

OBJECTIVE:
Demonstrate a simple Unsupervised Machine Learning example using Hierarchical (Agglomerative) Clustering in Python.

PROBLEM STATEMENT:
We have student data with:
1. Hours Studied
2. Exam Score

We do NOT have labels.
We want to automatically group students using clustering.

STEP 1: Install Required Libraries

```
pip install numpy pandas matplotlib scikit-learn scipy
```

STEP 2: Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster.hierarchy import dendrogram, linkage
```

STEP 3: Create Dataset

```
data = {
   "Hours": [1, 2, 3, 4, 5, 8, 9, 10, 11, 12],
   "Score": [35, 40, 45, 48, 50, 75, 78, 80, 82, 85]
}

df = pd.DataFrame(data)
print(df)
```

STEP 4: Visualize Data

```
plt.scatter(df["Hours"], df["Score"])
plt.xlabel("Hours Studied")
plt.ylabel("Score")
plt.title("Student Data")
plt.show()
```

STEP 5: Create Dendrogram

```
linked = linkage(df, method='ward')

plt.figure(figsize=(8,5))
dendrogram(linked)
plt.title("Dendrogram")
plt.xlabel("Students")
plt.ylabel("Distance")
plt.show()
```

STEP 6: Apply Hierarchical Clustering

```
hc = AgglomerativeClustering(n_clusters=2)
labels = hc.fit_predict(df)
print(labels)
```

STEP 7: Add Cluster Column

```
df["Cluster"] = labels
print(df)
```

STEP 8: Visualize Clusters

```
plt.scatter(df["Hours"], df["Score"], c=df["Cluster"])
plt.xlabel("Hours Studied")
plt.ylabel("Score")
plt.title("Student Clusters")
plt.show()
```

INTERPRETATION:

Cluster 0 → Low study hours, Low scores
Cluster 1 → High study hours, High scores

WHY THIS IS UNSUPERVISED LEARNING:

- No target column
- No labels provided
- Algorithm finds patterns automatically

BONUS: Change Number of Clusters

```
hc = AgglomerativeClustering(n_clusters=3)
labels = hc.fit_predict(df)
df["Cluster"] = labels
```