

In [1]:

```

import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
# Load data from CSV file
data = pd.read_csv('C:/Users/it3119/Desktop/data.csv')
X = data.values
# Determine the optimal number of clusters using silhouette score
scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, random_state=0).fit(X)
    scores.append(silhouette_score(X, kmeans.labels_))
best_k = np.argmax(scores) + 2 # Add 2 because we started from k=2
# Apply k-Means with the optimal number of clusters
kmeans = KMeans(n_clusters=best_k, random_state=0).fit(X)
# Print the clustering solution
print("Clustering solution:")
for i in range(best_k):
    cluster_i = np.where(kmeans.labels_ == i)[0]
    print(f"Cluster {i+1}: {cluster_i}")
# Calculate the quality of the clustering solution
score = silhouette_score(X, kmeans.labels_)
print(f"Silhouette score: {score}")

```

Clustering solution:

```

Cluster 1: [ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 2
1 22 23
24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
72 73 74]
Cluster 2: [ 75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
91 92
93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110
111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128
129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146
147 148 149]
Silhouette score: 0.6204656046551029

```

In []: