In-Class Exercise: Reading Clustering Output

Scenario #1: Three Clusters

1. How many distinct customer groups (segments) are there?

    3

2. Explain how the customers in cluster 1 are different from cluster 2?

    Younger, lower credit score, home value, and income.

3. What aspect of the customer data most differentiates cluster 1 from cluster 3?|

    The amount they withdraw from the ATM.

4. Which cluster has the highest cohesion? In practical terms, what does that mean?

    Cluster 1. This means that the customers in this group are more similar than the customers in the other two clusters.

Scenario #2: Ten Clusters

5. Is the root mean squared standard deviation of these clusters higher or lower than they were in the three cluster scenario? Why?

    Lower. Because there are more centroids, placing each data point nearer to a cluster center.

6. Is the distance to the nearest cluster higher or lower than in the three cluster scenario? Why?

    Lower. Because there are more centroids in the same space; there is less space between centroids.

7. Which scenario (#1 or #2) has higher cohesion among its clusters?

    Scenario #2 has higher cohesion (lower root mean squared standard deviation).

8. Which scenario (#1 or #2) has higher separation between its clusters?

    Scenario #1 has the higher separation (higher distance between clusters).