DS605 Lab Assignment 6 Report

202518036 Kashyap S. Patel

September 26, 2025

1 Optimal Clusters

- **K-Means:** From the Elbow Method plot, the "bend" occurs at k = 5, so the optimal number of clusters = 5.

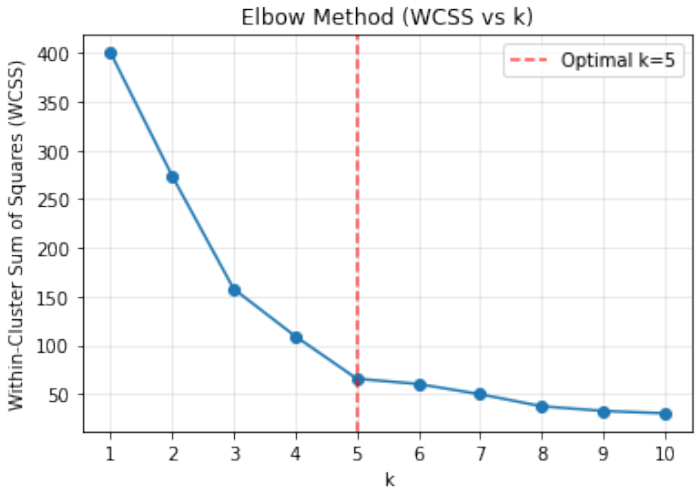

Figure 1: K-Means Elbow Method

- **Hierarchical:** From the dendrogram, cutting around distance 150 also gives 3 clusters as a clean split.

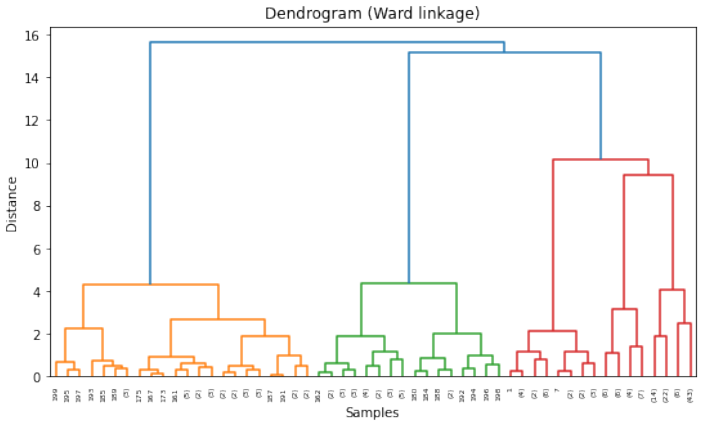

Figure 2: Dendrogram

2 Cluster Comparison

- **K-Means (k=5):** Created 5 spherical clusters. Works well for compact groups.

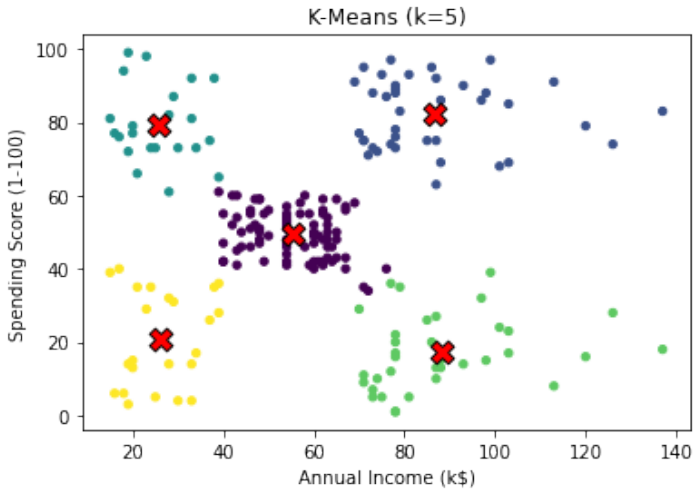

Figure 3: K-Means Clusters

- **Hierarchical (3 clusters):** Produced 3 larger clusters, merging some smaller groups that K-Means separated.

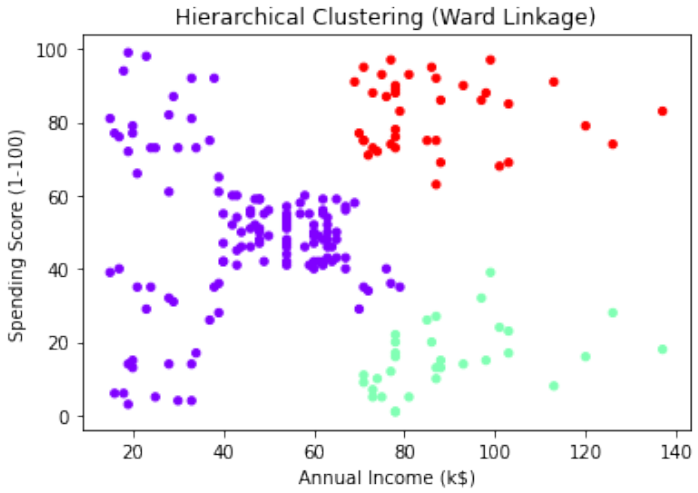

Figure 4: Hierarchical Clustering Clusters

- **DBSCAN:**
  - Detected arbitrary shaped clusters.
  - Identified noise points (black) that K-Means and Hierarchical forced into clusters.

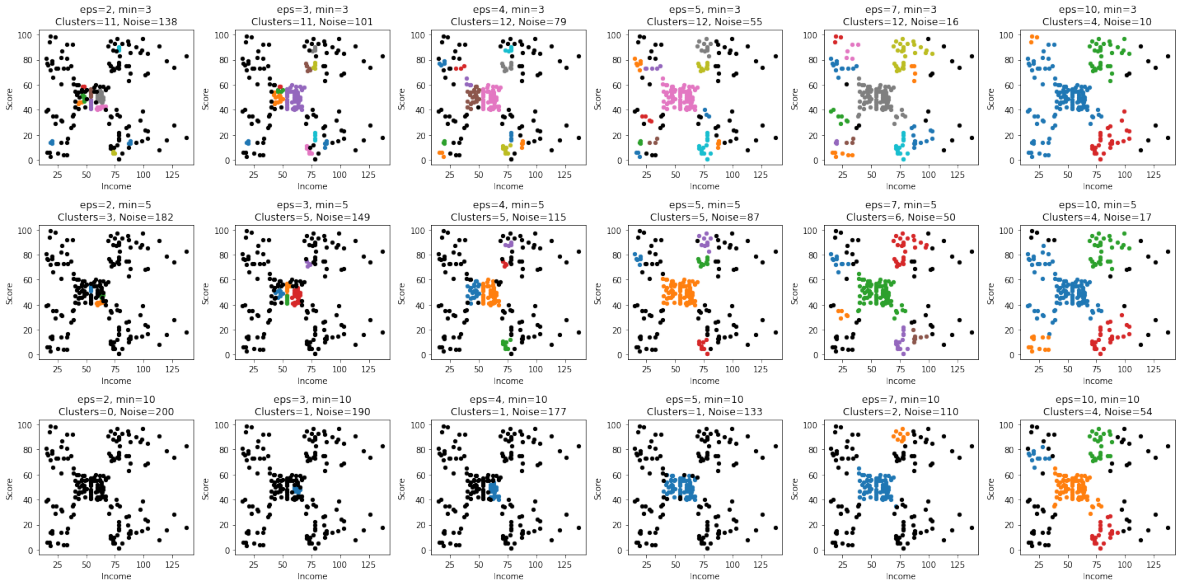

Figure 5: DBSCAN Clusters

- **Notable difference:**
  - K-Means & Hierarchical assign all points to a cluster.
  - DBSCAN isolates outliers instead of forcing them into clusters.

3 DBSCAN Performance

- DBSCAN clearly detected noise/outliers (label -1).
- Found fewer clusters depending on (eps, min_samples).
- Performance varies with parameters: sometimes 2–3 clear clusters, sometimes everything labeled as noise.
- **Conclusion:** it is better at handling irregular shapes + noise. But it is parameter sensitive; poor eps/min_samples → bad clustering.

4 Algorithm Suitability

- Since clusters are fairly well-separated, K-Means (k=5) works best.
- DBSCAN adds value by highlighting noise but is sensitive to parameters.
- Hierarchical gave broader groupings, less detailed than K-Means.

5 Real-World Application

- **Cluster with high income, low spending score:** These are wealthy but cautious customers. Marketing team could target them with exclusive luxury promotions, premium memberships, or personalized offers to encourage higher spending.
- **Cluster with low income, high spending score:** Likely young or impulsive shoppers. They could be targeted with budget-friendly deals, discounts, loyalty rewards.

- **Cluster with medium income & medium score:** Average customers. Can be engaged with general seasonal campaigns.
- **Noise (DBSCAN -1):** Outliers (unusual customers). Marketing might ignore them or study them separately for niche campaigns.