ML-Lab Assignment 6

Name: Divya Mashruwala
ID: 202518042

September 26, 2025

1 Analysis and Discussion

1.1 Optimal Clusters

For K-Means clustering, the Elbow Method indicated an optimal number of clusters as **5**, where the WCSS curve showed a clear bend, suggesting a good balance between compactness and simplicity.

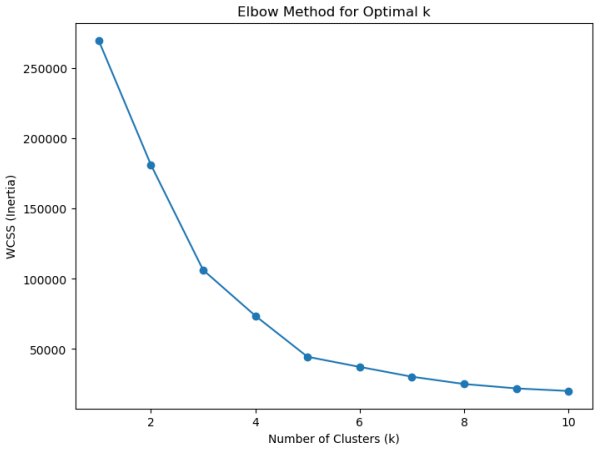

Figure 1: Elbow Method for optimal k in K-Means

For Hierarchical Clustering, the dendrogram suggested **3 clusters** as the best choice, identified by cutting the tree at the longest vertical distance without crossing any horizontal lines.

1.2 Cluster Comparison

Visually comparing the results of the three algorithms: - K-Means produced five distinct clusters with balanced sizes, capturing finer granularity. Cluster sizes were not listed here, but all points were assigned. - Hierarchical Clustering produced three clusters with the following sizes: Cluster 1: 129, Cluster 2: 32, Cluster 3: 39, with a silhouette score of 0.4618. - DBSCAN produced three clusters: Cluster 1: 138, Cluster 2: 12, Cluster 3: 33, and identified 17 noise points, with a silhouette score of 0.3747.

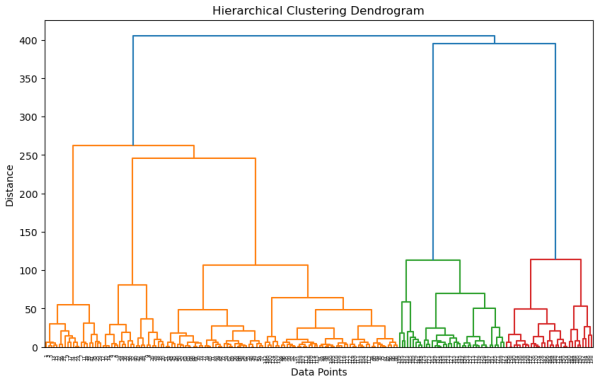

Figure 2: Dendrogram for Hierarchical Clustering

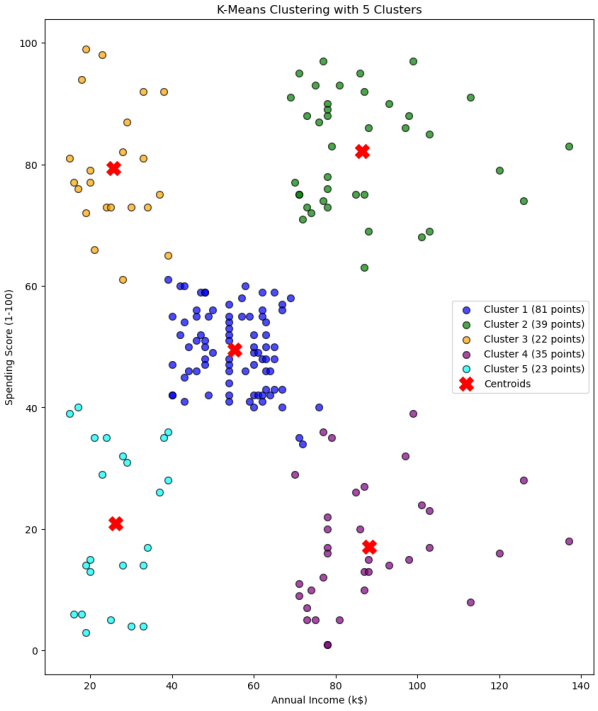

Figure 3: Enter Caption

1.3 DBSCAN Performance

DBSCAN effectively identified dense clusters, producing three clusters with sizes: 138, 12, and 33 points, while labeling 17 points as noise. Unlike K-Means and Hierarchical Clustering, DBSCAN does not force every point into a cluster, which helps detect outliers and unusual customer behavior. However, it is sensitive to parameter selection ($\epsilon = 0.5$, min_samples $= 14$).

1.4 Algorithm Suitability

Considering cluster shapes, density, and practical application, K-Means is the most suitable for this dataset. Although DBSCAN isolates noise, K-Means provides a fixed number of clusters (5), assigns all customers to a cluster, and produces interpretable, balanced clusters. Hierarchical Clustering is useful for visualization but produces fewer clusters

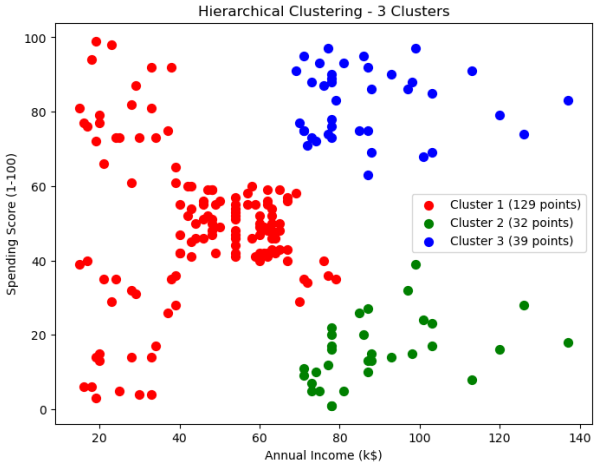

Figure 4: Hierarchical Clustering of Customers

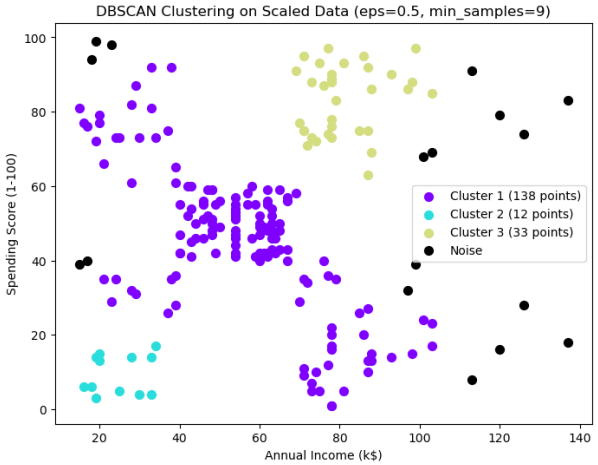

Figure 5: DBSCAN Clustering of Customers

(3) and less granularity. Silhouette scores:

- K-Means: 0.554
- Hierarchical Clustering: 0.4618
- DBSCAN: 0.3747

1.5 Real-World Application

The clusters can guide targeted marketing:

- **High-income, high-spending customers:** Premium products, loyalty programs, exclusive promotions.
- **High-income, low-spending customers:** Personalized discounts or bundled offers to encourage spending.
- **Low-income, high-spending customers:** Budget-friendly deals or installment options.

- **Outliers (DBSCAN noise points):** Investigate unusual buying patterns for special campaigns.

This segmentation allows the mall to optimize marketing strategies, improve engagement, and increase revenue.