Lab 6: Clustering Methods – Analysis and Questions

Dhruv Parmar
Student ID: 202518030

September 26, 2025

1. Optimal Clusters

Using the Elbow Method (Figure 1), the within–cluster sum of squares (WCSS) begins to flatten at $k = 5$, indicating that five clusters is optimal for K-Means. For hierarchical clustering, the dendrogram (Figure 2) shows a clear separation at six vertical lines, so we selected 6 clusters for Agglomerative Hierarchical Clustering.
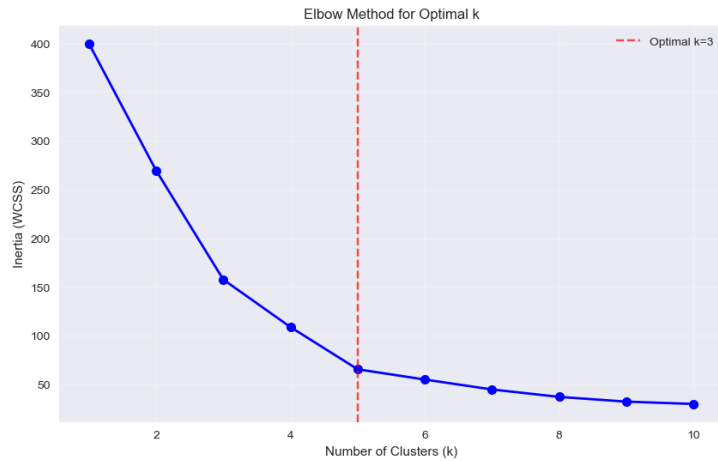

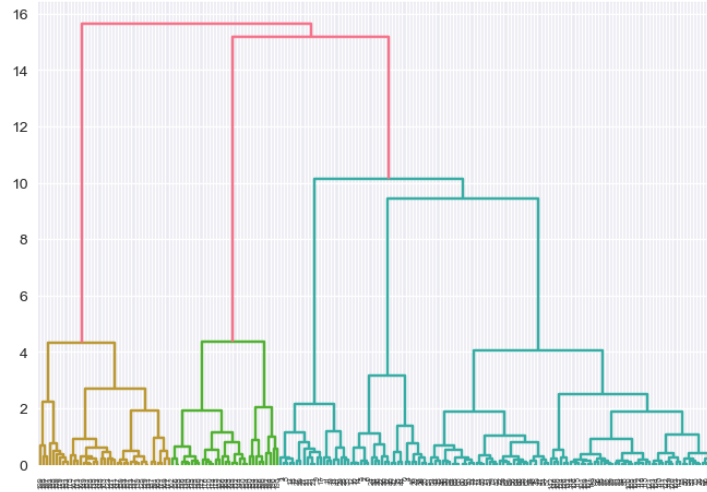Figure 1: Elbow Method showing optimal $k = 5$.

Figure 2: Dendrogram indicating optimal clusters for hierarchical clustering.

2. Cluster Comparison

K-Means produced five compact, roughly spherical clusters. Agglomerative Hierarchical Clustering with six clusters produced a very similar grouping but with slightly different boundaries due to its bottom–up merging process. DBSCAN produced four dense clusters and marked several points as noise, which the other two algorithms forcibly assigned to clusters.

3. DBSCAN Performance

DBSCAN does not require a predefined number of clusters. With parameters $\varepsilon = 5$ and `min_samples`=5, it identified four main clusters and labelled about 12 points as noise (-1). Compared to K-Means and Hierarchical, DBSCAN separated dense regions well but treated sparse regions as noise rather than forcing them into a cluster.

4. Algorithm Suitability

For this dataset, the customer groups appear compact and well-separated in two dimensions (Annual Income vs. Spending Score). Therefore, K-Means was the most suitable algorithm. Hierarchical clustering produced comparable results but is less efficient for large datasets. DBSCAN is more suitable for irregularly shaped clusters or when detecting outliers is important.

5. Real-World Application

The identified customer segments can guide the mall's marketing strategy. For example, the cluster with high annual income but low spending score could be targeted with personalised premium offers or loyalty incentives to increase spending. High-spending groups could be offered exclusive rewards to maintain loyalty, while low-income, high-spending clusters could receive discounts or bundle offers.