Clustering Report

Megh Nanavati
202518018

26th September 2025

1 Optimal Clusters

1.1 K-Means Clustering

The Elbow Method was used to determine the optimal number of clusters. The plot of WCSS versus k shows a sharp decrease until k=5, after which the decrease becomes gradual. Therefore, the optimal cluster number for K-Means is 5.
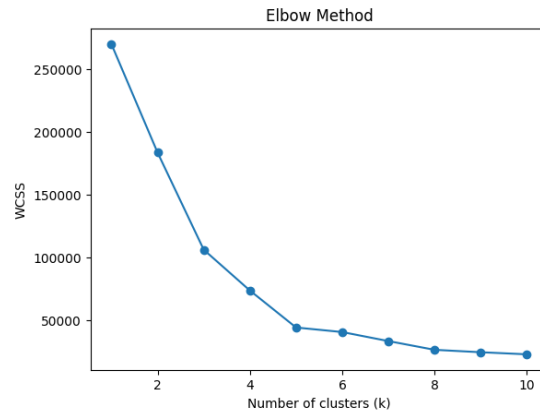

Figure 1: Elbow Method plot indicating optimal K-Means clusters at k=5.

1.2 Hierarchical Clustering

The dendrogram generated using Ward linkage shows 5 clear clusters before the first major merge. Therefore, 5 clusters are also chosen for Hierarchical Clustering.
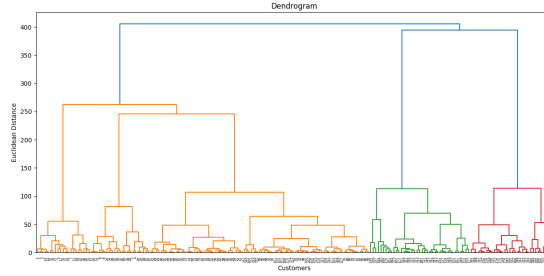
Figure 2: Dendrogram indicating 5 clusters for Hierarchical Clustering.

2 Cluster Comparison

The following figure shows the clusters from K-Means, Hierarchical, and DB-SCAN side by side for comparison. Here we can see that the performance of k-means and hierarchical clustering is similar; however, DBSCAN classifies some points as noise. Due to this the clusters are not as good as compared to k-means and hierarchical clustering.
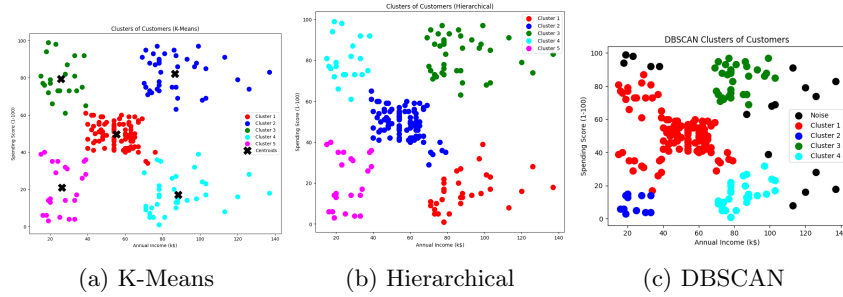

(a) K-Means   (b) Hierarchical   (c) DBSCAN

Figure 3: Comparison of clusters from the three algorithms.

3 DBSCAN Performance

DBSCAN identified 4 clusters and labeled some points as noise. This differs from K-Means and Hierarchical because DBSCAN is density-based, ignoring sparse points.
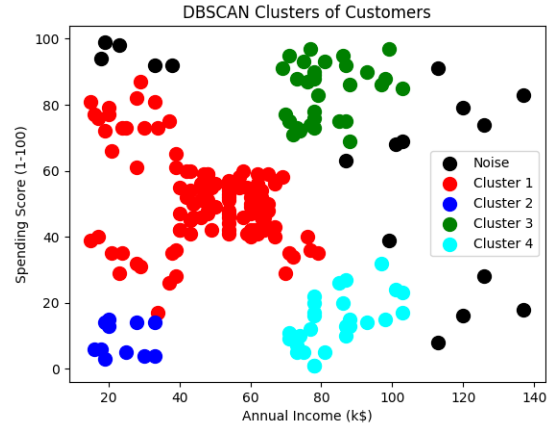
Figure 4: DBSCAN clustering showing 4 clusters with noise points.

4 Algorithm Suitability

K-Means and Hierarchical are more suitable for this dataset because clusters are compact and roughly spherical. DBSCAN is less suitable due to uniform density and minimal noise.

5 Real-world Application

The customer segments can be used for targeted marketing:

- High income, high spending: luxury promotions.

- High income, low spending: incentives or loyalty programs.

- Low income, high spending: budget-friendly premium offers.

- Low income, low spending: general campaigns or discounts.