Customer Segmentation Using Clustering Algorithms
DS605 - Lab 6 Report

Devam Gandhi - 202518008

26 Sep 2025

Abstract

In this report, three clustering algorithms—K-Means, Agglomerative Hierarchical Clustering, and DBSCAN—are all utilized for analyzing the Mall Customers dataset. We compare the clusters produced through each method, evaluate how well it performs, and utilize dendrogram visualizations, the Elbow Method, and silhouette scores to back up our choices.

1 Introduction

Targeted marketing campaigns and tailored services can be made possible by customer segmentation, making it an essential role in retail analytics. We prioritize on two features: Spending Score (1–100) and Annual Income (k$). We can find important groups of consumers by combining these parameters.

2 Methodology

2.1 Clustering Methods

- **K-Means:** The ideal amount of k was identified using the Elbow technique and silhouette scores.

- **Hierarchical Clustering:** Ward's linkage method was used. The dendrogram guided the choice of cluster number.

- **DBSCAN:** Parameter tuning was done for `eps` and `min_samples`, with evaluation using silhouette score.

3 Results and Discussion

3.1 1. Optimal Clusters

The ideal amount of clusters for K-Means was determined to be **5** based on silhouette scores and the Elbow Method (Figure 1). The same choice for Hierarchical Clustering

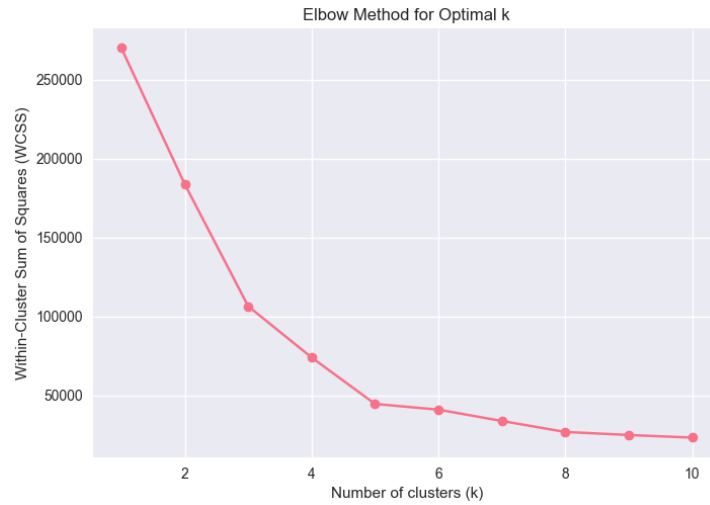was supported by the dendrogram (Figure 2), which demonstrated a notable jump at 5 clusters.

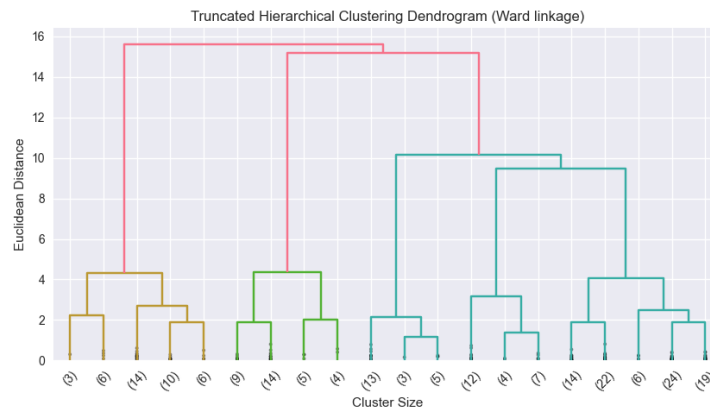Figure 1: Elbow Method for K-Means showing the optimal $k = 5$.

Figure 2: Dendrogram (Ward linkage) suggesting 5 clusters.

3.2   2. Cluster Comparison

hierarchical clusteringDBSCAN, however, was more sensitive to density. Depending on parameter settings, it either grouped most points into a single cluster or identified 3–4 clusters with some noise. hierarchical clustering,
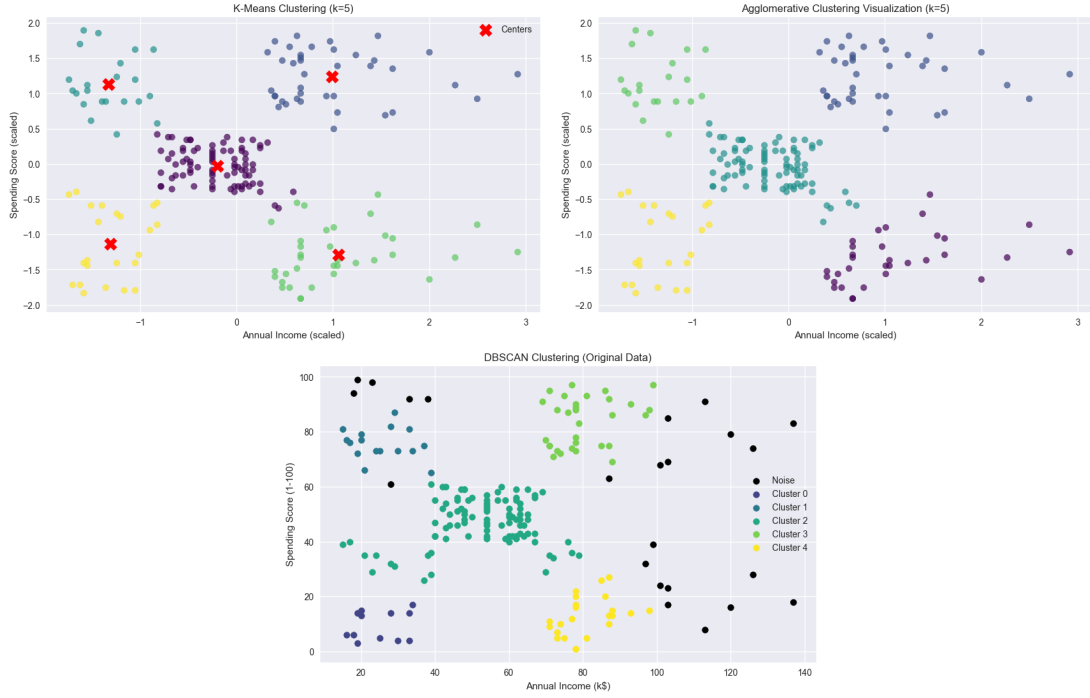
Figure 3: Comparison: K-Means vs. Agglomerative Hierarchical vs DBSCAN Clustering.

3.3 3. DBSCAN Performance

Using the default parameters, DBSCAN was struggle classifying consumers to significant clusters on this dataset. It frequently generated no noise points and grouped all the points to a single cluster, indicating that the density-based technique failed to recognize zones with various density in this 2D data. given its failure to detect a dataset's natural segment boundaries, DBSCAN did more poorly than K-Means and Hierarchical clustering, which always classify all customers into different categories. Since the customer segments were better represented by more compact, spherical clusters, DBSCAN was not as effective in this case, notwithstanding its strength of detecting clusters of arbitrary shapes and identifying outliers in more complex datasets. Figure 4 illustrates one such result.
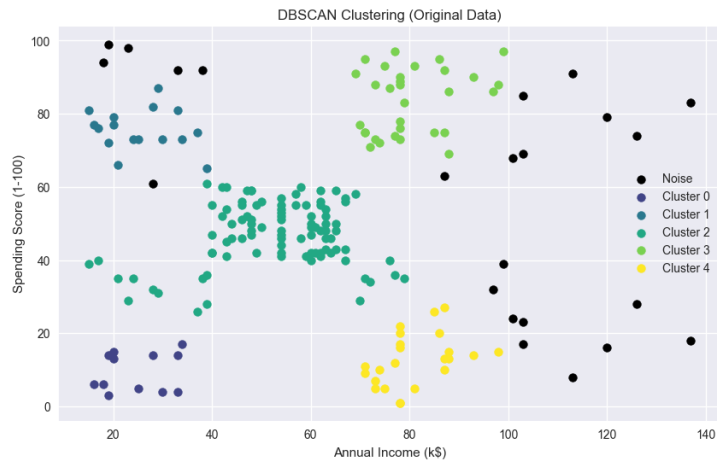

Figure 4: DBSCAN clustering result (example parameters with silhouette > 0.5).

3

3.4   4. Algorithm Suitability

For this dataset, K-Means and Hierarchical Clustering are more suitable since the clusters are compact and well-separated. DBSCAN is better for irregularly shaped clusters or when noise detection is essential. Here, however, DBSCAN struggled due to uniform density and relatively clear separation of groups.

3.5   5. Real-World Application

The identified clusters can guide marketing strategies. For example:

- A cluster of customers with **high income but low spending score** can be targeted with personalized promotions, luxury product launches, or loyalty programs to increase engagement.

- Customers with **low income but high spending score** may be targeted with discounts and bundled offers to maximize retention.

- Middle-income moderate spenders can be offered seasonal campaigns or group discounts.

4   Conclusion

This project demonstrated how clustering techniques can segment mall customers into actionable groups. K-Means and Hierarchical Clustering identified 5 clusters with good silhouette scores, while DBSCAN required careful parameter tuning. In practice, segmentation enables targeted marketing, boosting both customer satisfaction and business revenue.