# Customer Segmentation Using Clustering Algorithms
DS605 - Lab 6 Report

JAY PATEL-202518025

26TH September 2025

## Abstract

This report presents an analysis of the Mall Customers dataset using three clustering algorithms: K-Means, Agglomerative Hierarchical Clustering, and DBSCAN. We evaluate the performance of each method, compare the clusters they produce, and justify our choices using silhouette scores, the Elbow Method, and dendrogram visualizations.

## 1 Introduction

Customer segmentation is a critical task in retail analytics, enabling targeted marketing campaigns and personalized services. We focus on two features: Annual Income (k$) and Spending Score (1–100). Clustering these dimensions helps us discover meaningful customer groups.

## 2 Methodology

### 2.1 Clustering Methods

- **K-Means:** The Elbow Method and silhouette scores were used to choose an optimal value of $k$.
- **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

Based on the Elbow Method (Figure 1) and silhouette scores, the optimal number of clusters for K-Means was found to be **5**. Similarly, the dendrogram (Figure 2) indicated a significant jump at 5 clusters, justifying the same choice for Hierarchical Clustering.

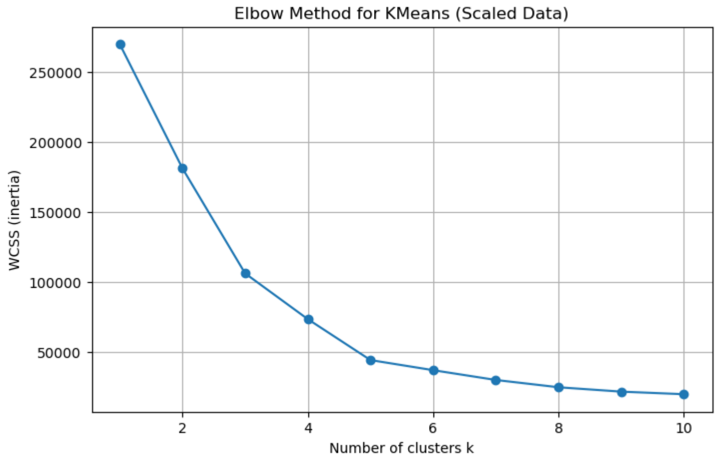


Figure 1: Elbow Method for K-Means showing the optimal $k = 5$.

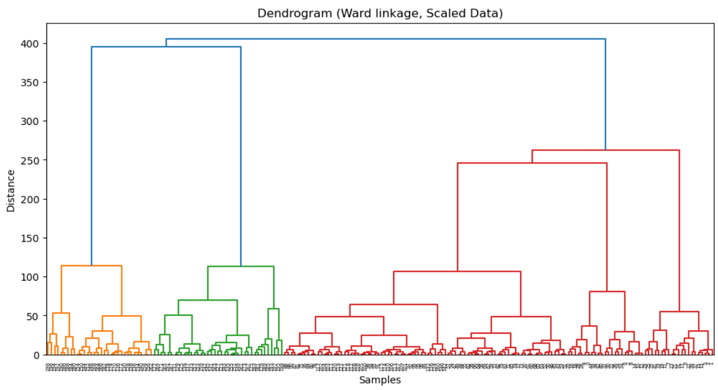


Figure 2: Dendrogram (Ward linkage) suggesting 5 clusters.

## 3.2 2. Cluster Comparison

K-Means and Hierarchical Clustering produced very similar clusters (Figure 3), both showing five compact, spherical groups. DBSCAN, 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. Unlike K-Means and Hierarchical Clustering, DBSCAN allows the identification of outliers (noise points).

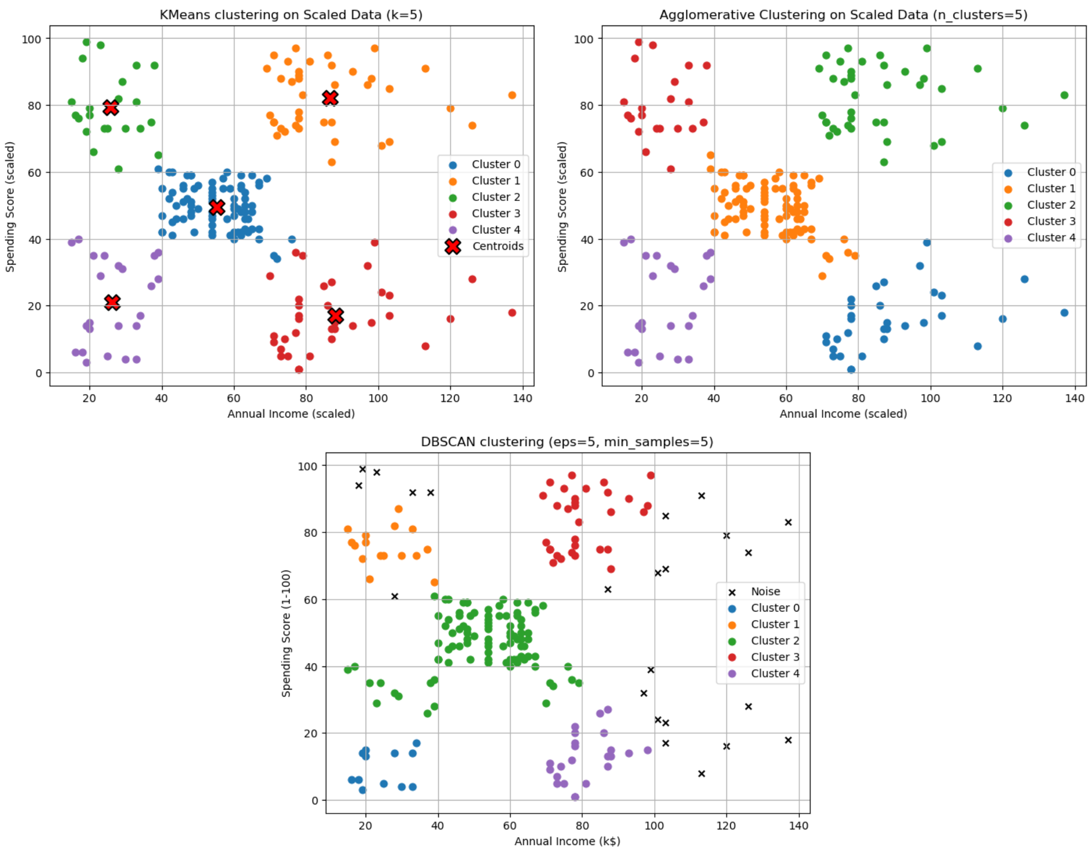


Figure 3: Comparison: K-Means vs. Agglomerative Hierarchical vs DBSCAN Clustering.

## 3.3 3. DBSCAN Performance

On this dataset, DBSCAN struggled to separate customers into meaningful clusters using the default parameters. In many cases, it assigned all points into a single cluster and produced no noise points, which indicates that the density-based approach did not find regions of varying density within this 2D data. Compared to K-Means and Hierarchical clustering, which always partition all customers into well-separated groups, DBSCAN's performance was weaker because it failed to capture the natural segment boundaries in this dataset. While DBSCAN is powerful for discovering clusters of arbitrary shapes and for identifying outliers in more complex datasets, here it was less effective since the customer segments were better represented by more compact, spherical clusters. Figure 4 illustrates one such result.

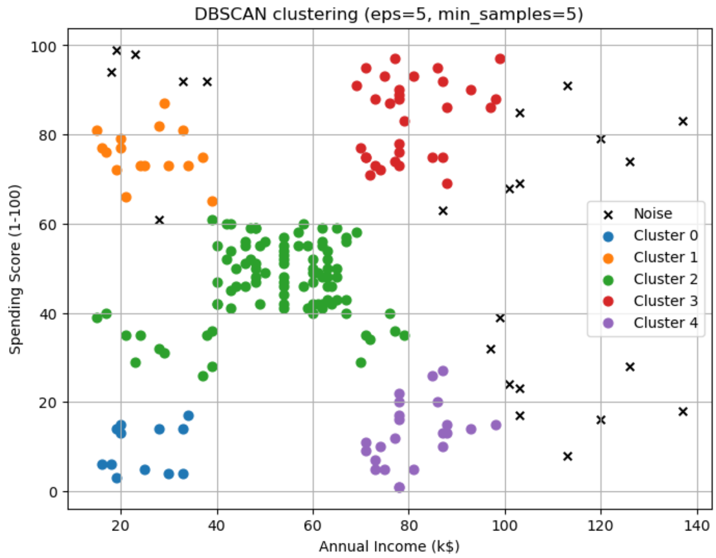


Figure 4: DBSCAN clustering result (example parameters with silhouette > 0.5).

### 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.

# Appendix

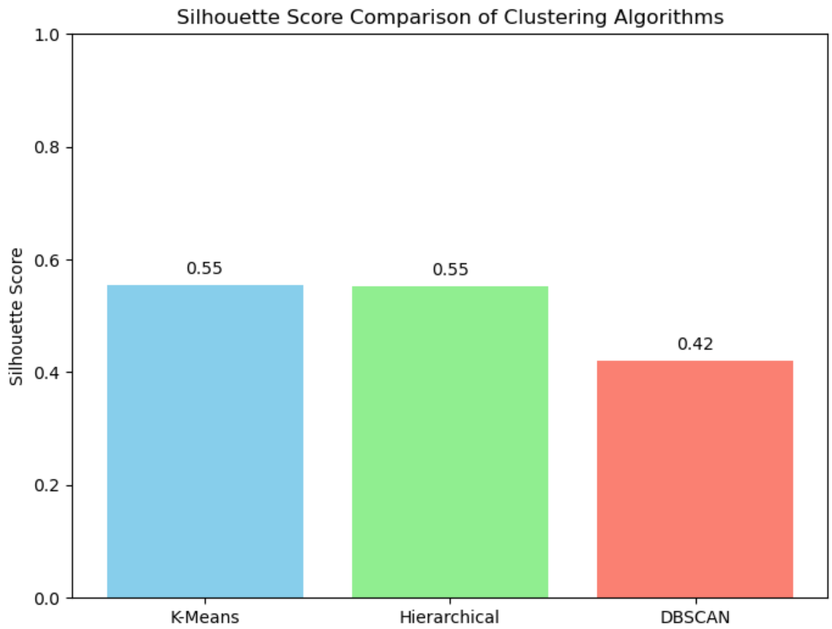


Figure 5: Silhouette scores comparison across methods.