Mall Customer Segmentation Using Clustering Algorithms

DS605 - Lab 6 Report

Saachi Garg - 202518007

26th September 2025

1 Introduction

Targeted marketing efforts and customised services are made possible by customer segmentation, which is a crucial task in retail analytics. We concentrate on two features: Spending Score (1–100) and Annual Income (k$). We can find significant client groups by grouping these dimensions.

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

Using silhouette scores and the Elbow Method, (Figure 1) the ideal number of clusters for K-Means was determined to be **5**. The similar decision for Hierarchical Clustering was supported by the dendrogram (Figure 2), which also showed a notable leap at cluster number five.

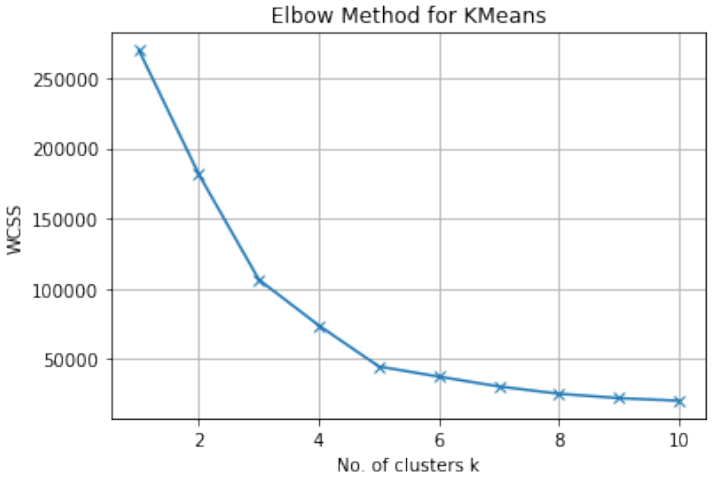

Figure 1: Elbow Method for K-Means showing the optimal $k = 5$.

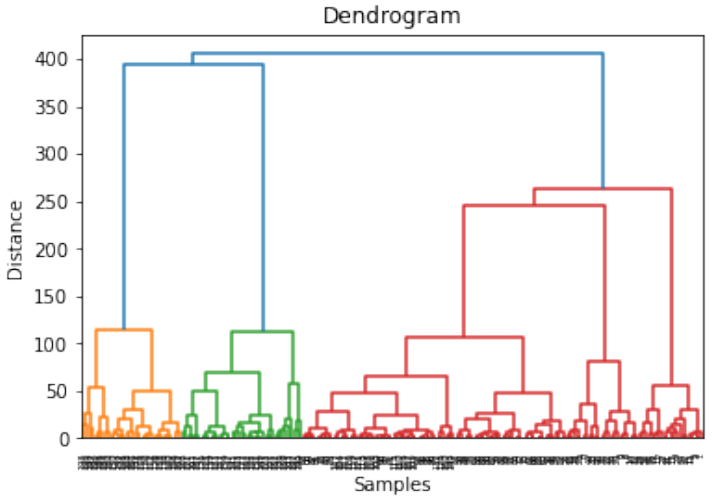

Figure 2: Dendrogram (Ward linkage) suggesting 5 clusters.

3.2 2. Cluster Comparison

The five tight, spherical groups that were created by K-Means and Hierarchical Clustering were very similar(Figure 3). On the other hand, DBSCAN was more density-sensitive. It either found three to four clusters with some noise or gathered the majority of points into a single cluster, depending on the parameter parameters. In contrast to K-Means and Hierarchical Clustering, DBSCAN makes it possible to identify outliers, or noise spots.

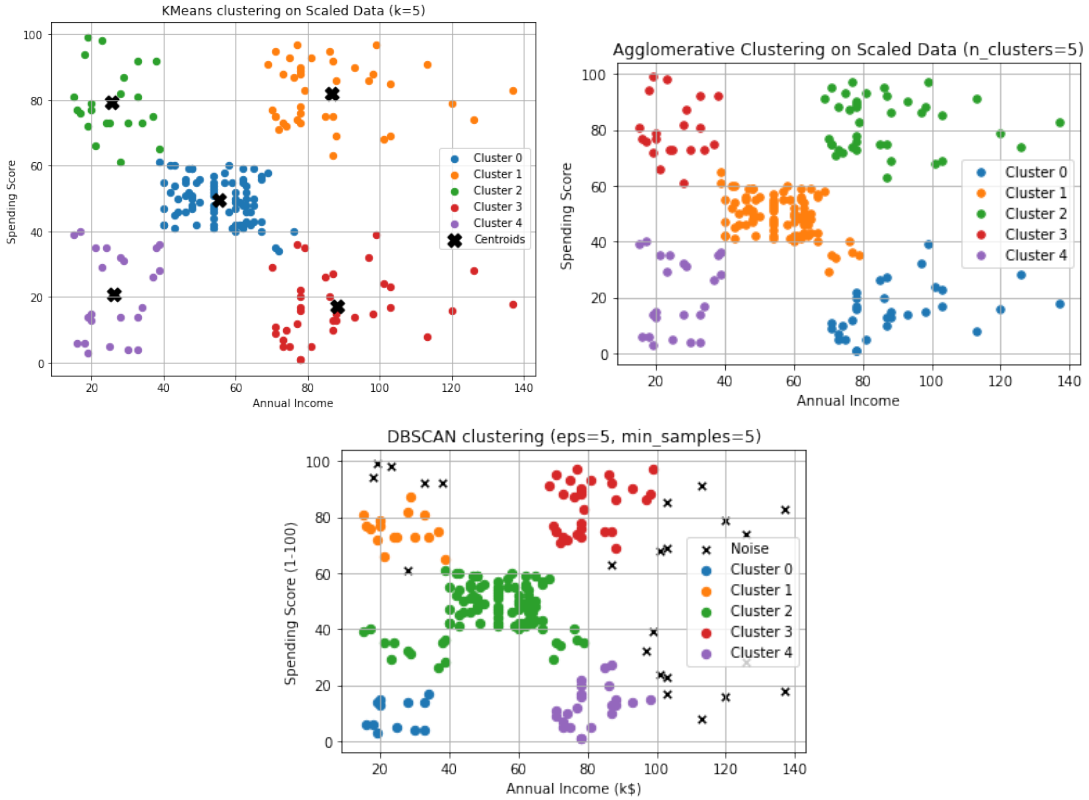

Figure 3: Comparison: K-Means vs. Agglomerative Hierarchical vs DBSCAN Clustering.

3.3 3. DBSCAN Performance

Using the default parameters, DBSCAN has trouble separating consumers into meaningful clusters on this dataset. It frequently produced no noise points and grouped all the dots into a single cluster, suggesting that the density-based method was unable to identify areas of different densities in this 2D data. Due to its inability to recognise the dataset's natural segment borders, DBSCAN performed worse than K-Means and Hierarchical clustering, which always divide all consumers into distinct groups. Although DBSCAN is a great tool for finding clusters of any shape and for detecting outliers in more complicated datasets, it was less successful in this case because more compact, spherical clusters better represented the customer segments. Figure 4 illustrates one such result.

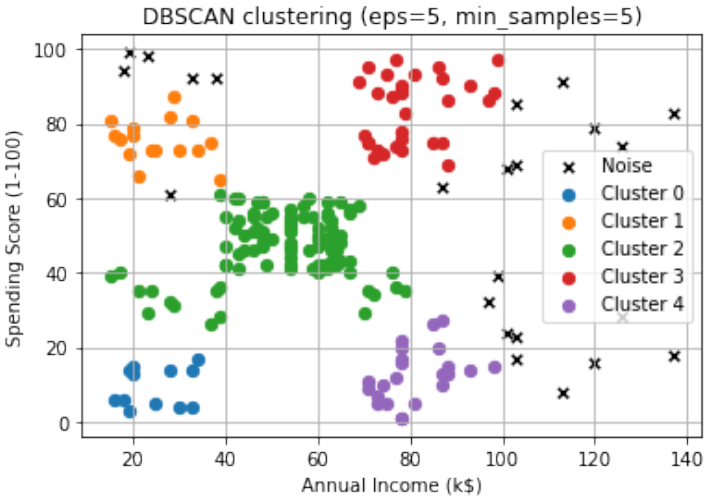

Figure 4: DBSCAN clustering result (example parameters with silhouette > 0.5).

3.4 4. Algorithm Suitability

For this dataset, K-Means and Hierarchical Clustering are better suitable since the clusters are compact and well-separated. When noise detection is crucial or the clusters are irregularly formed, DBSCAN performs better. However, because of the uniform density and somewhat obvious group distinction, DBSCAN had trouble in this case.

3.5 5. Real-World Application

Clustering can be applied to patient data (e.g., age, BMI, blood pressure, lifestyle habits, and medical history) to group individuals with similar health profiles. For instance

- A cluster of patients with high BMI and high blood pressure can be flagged for preventive programs focusing on diet, exercise, and hypertension management.
- Patients with low physical activity but high cholesterol can be targeted with lifestyle counseling and regular check-ups.
- A group of young, healthy, low-risk individuals may only require annual screenings and digital health reminders, reducing unnecessary medical costs.

4 Conclusion

This report showed how mall patrons can be divided into actionable groups using clustering algorithms. While DBSCAN needed careful parameter adjustment, K-Means and Hierarchical Clustering found five clusters with good silhouette scores. In actuality, segmentation makes customised marketing possible, increasing client satisfaction and company profits.