Mall Customers Clustering Analysis

Your Name

September 26, 2025

1 Clustering Results

K-Means Clustering

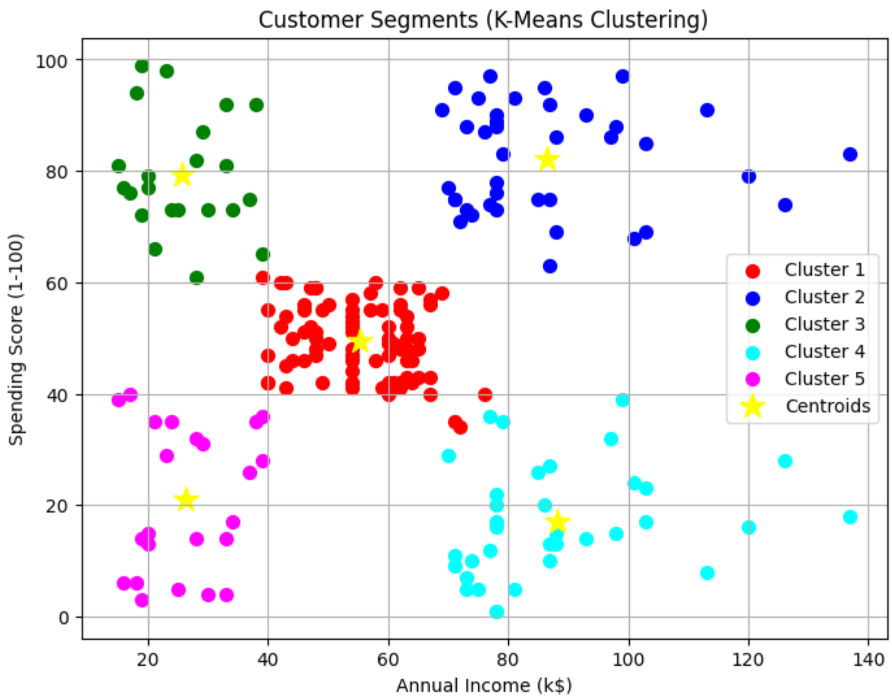

Figure 1: K-Means Clustering Results. Each color represents a distinct customer cluster, and the centroids are shown as larger markers.

Hierarchical Clustering

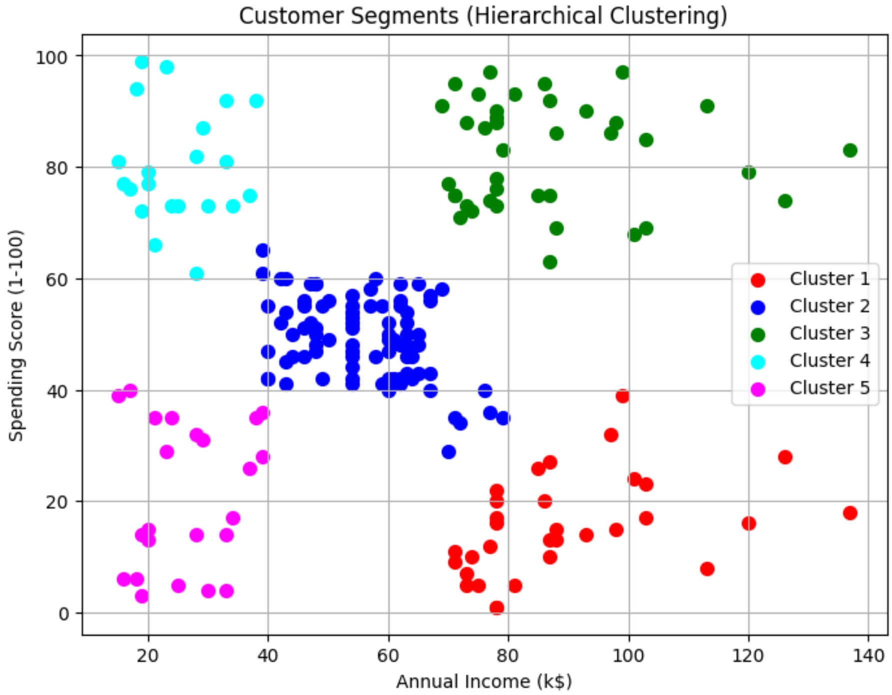

Figure 2: Agglomerative Hierarchical Clustering Results. The Ward linkage method was used to form compact clusters.

DBSCAN Clustering

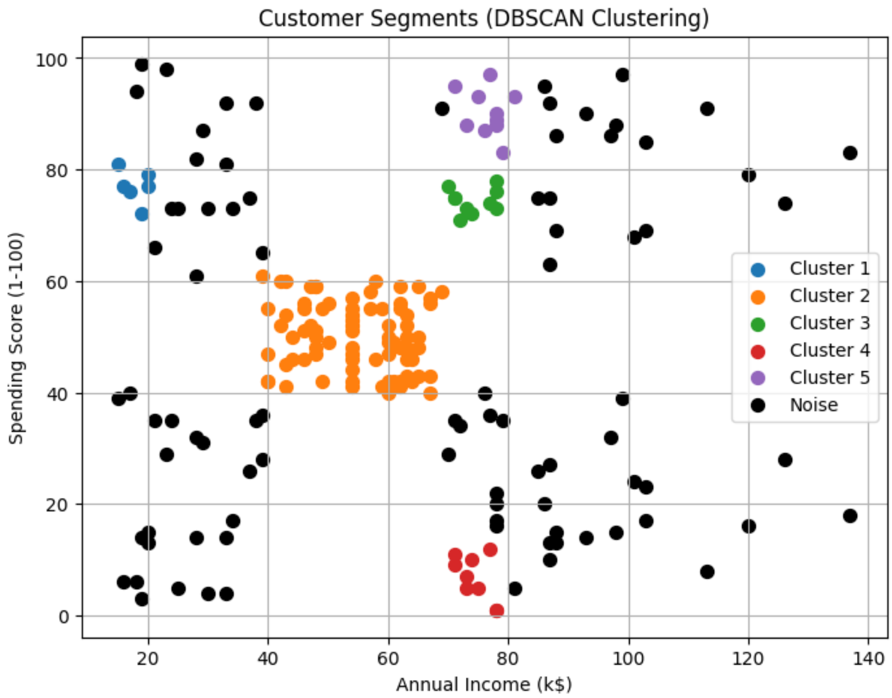

Figure 3: DBSCAN Clustering Results. Noise points are shown in black, while valid clusters are represented by different colors.

2 Analysis and Questions

1. Optimal Clusters

For K-Means, the Elbow Method indicated that the optimal number of clusters was **5**. This was the point where the WCSS (Within-Cluster Sum of Squares) curve began to flatten, suggesting diminishing returns for adding more clusters. For Hierarchical Clustering, the dendrogram cut also suggested **5 clusters**, as the longest vertical distance without crossing horizontal lines intersected five groups. This consistency strengthens confidence in the chosen cluster number.

2. Cluster Comparison

Visually, both K-Means and Hierarchical Clustering produced similar segmentations of customers, dividing them into groups such as:

- Low income, low spending customers,
- High income, high spending customers,
- High income, low spending customers,
- Moderate income, moderate spending customers,
- Low income, high spending customers.

The main difference is that K-Means creates spherical clusters around centroids, while Hierarchical Clustering allows slightly more flexible boundaries. DBSCAN produced fewer clusters and marked some points as noise, which shows it does not force every point into a group.

3. DBSCAN Performance

DBSCAN was able to detect non-linear structures and identify outliers. Several customers were labeled as noise (`-1`), meaning they did not belong to any dense region. Unlike K-Means and Hierarchical Clustering, which force every data point into a cluster, DBSCAN's approach is advantageous in real-world datasets where not all customers fit neatly into groups. However, DBSCAN was sensitive to the choice of parameters `eps` and `min_samples`, and with this dataset it produced fewer clusters compared to K-Means and Hierarchical.

4. Algorithm Suitability

For this dataset, K-Means proved most suitable due to its clear, interpretable clusters and balanced distribution of customers. Hierarchical clustering was also consistent and useful for visualization, especially in understanding cluster merging. DBSCAN, although powerful in detecting arbitrary shapes and noise, did not perform as effectively here because the data forms relatively well-separated circular groups, which align with K-Means' strengths.

5. Real-World Application

The customer segments identified can help the mall's marketing team design targeted campaigns:

- **High income, high spending group:** These customers are premium buyers. They can be targeted with exclusive offers, VIP programs, and luxury product promotions.
- **High income, low spending group:** Although wealthy, these customers spend cautiously. The mall could offer personalized loyalty programs, discounts, or incentives to encourage higher spending.
- **Low income, high spending group:** These are aspirational buyers who spend more relative to their income. They may respond well to installment plans, budget-friendly luxury products, and bundled offers.
- **Moderate income, moderate spending group:** This group represents average buyers. Seasonal sales, family packages, and general marketing strategies would work best.
- **Low income, low spending group:** These customers are the least profitable. However, they can be engaged with affordable product promotions to increase footfall in the mall.

By leveraging these insights, the mall can optimize promotions, improve customer engagement, and maximize revenue.