Clustering Analysis Report

1 Optimal Clusters

K-Means: Using the Elbow Method, the distortion (inertia) decreased rapidly until around $k = 5$, after which the curve began to flatten. This indicated that 5 clusters was the optimal choice.

Hierarchical Clustering: The dendrogram showed clear separation when cutting at around 5 clusters as well, since below that the linkage distance started to increase sharply.

Conclusion: Both methods converged on 5 clusters as optimal.

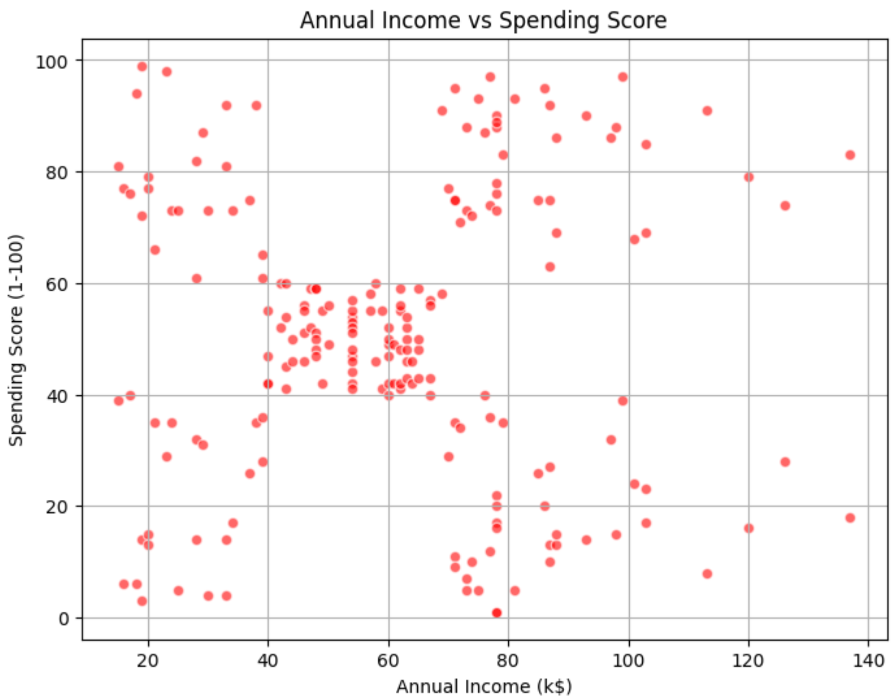

Figure 1: Annual Income vs Spending Score

2 Cluster Comparison

- **K-Means:** Produced spherical, well-separated clusters.
- **Agglomerative (Hierarchical):** Produced clusters very similar to K-Means, with minor boundary differences (since it groups bottom-up).
- **DBSCAN:**
  - Detected arbitrary-shaped clusters.

– Some customers were labeled as noise (-1), unlike K-Means and Hierarchical which force every point into a cluster.

Overall: K-Means and Agglomerative gave very similar results, while DBSCAN gave fewer but denser clusters with outliers excluded.

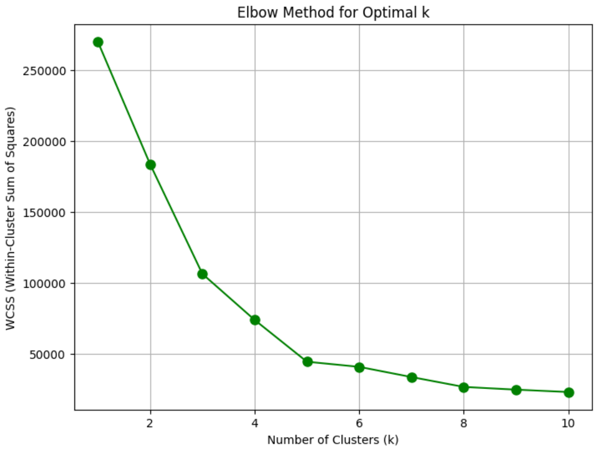

Figure 2: Cluster Curve

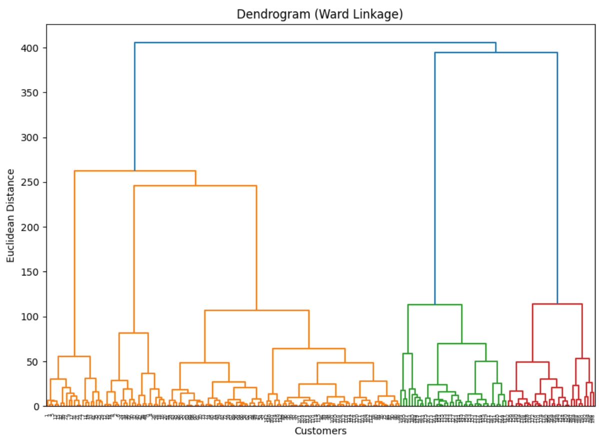

Figure 3: Dendogram

3 DBSCAN Performance

- DBSCAN identified some noise points (customers that did not belong to any cluster).
- Unlike K-Means and Hierarchical, which divided all customers into groups, DBSCAN only grouped customers that were close in density.
- On this dataset, which tends to form well-separated circular clusters, DBSCAN was less effective than K-Means/Hierarchical, but it did help highlight outliers (e.g., very high-income customers with unusual spending habits).

4 Algorithm Suitability

Most suitable: K-Means (or Agglomerative).

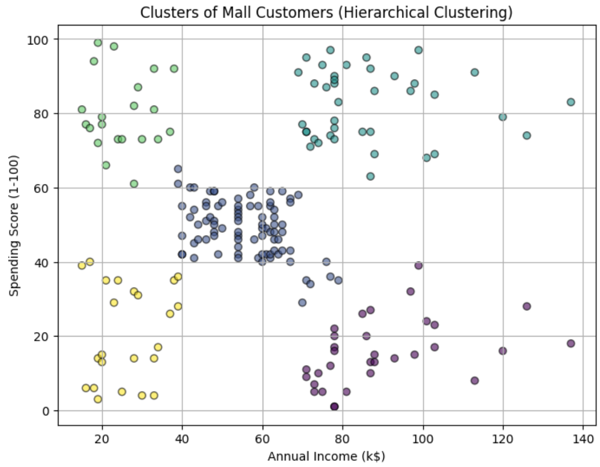

Figure 4: Clustering

Reason:

- The clusters in this dataset (Annual Income vs Spending Score) are fairly compact and spherical.
- K-Means and Hierarchical both separated them clearly into 5 groups.
- DBSCAN is more suitable when clusters are non-spherical or overlapping in density, which was not the case here.

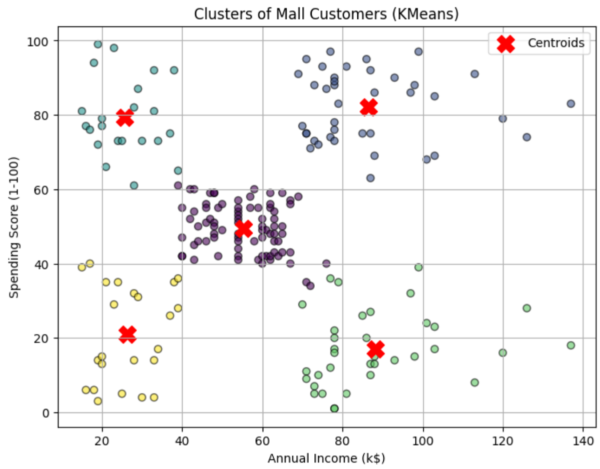

Figure 5: centroid

5 Real-World Application

Example: Mall Customer Segmentation

- **Cluster 1: High Income – High Spending Score** Premium customers → target with luxury promotions, VIP memberships, exclusive events.
- **Cluster 2: High Income – Low Spending Score** Wealthy but cautious → target with personalized offers, discounts, premium loyalty programs to encourage spending.
- **Cluster 3: Medium Income – Medium Spending Score** Average spenders → target with seasonal sales, bundle offers.

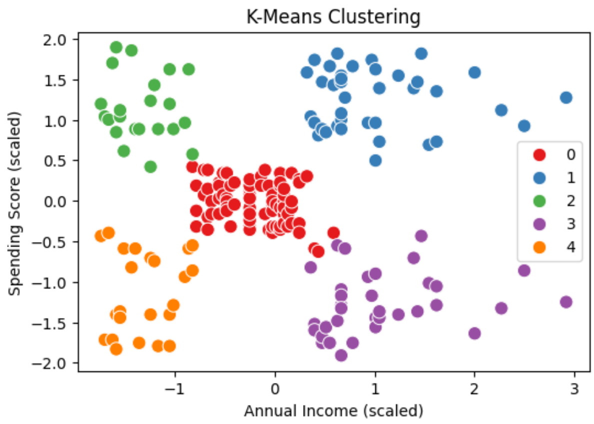

Figure 6: K-mean

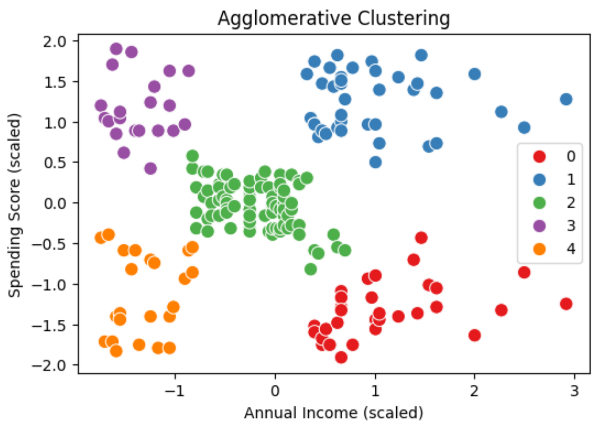

Figure 7: Agglomerative

- **Cluster 4: Low Income – High Spending Score** Price-sensitive but frequent shoppers → target with budget-friendly promotions, loyalty rewards.
- **Cluster 5: Low Income – Low Spending Score** Least valuable segment → minimal marketing effort, maybe general discounts.

Example (Cluster 2): For the high-income but low-spending score group, the marketing team might:

- Send personalized luxury product recommendations.
- Offer exclusive store experiences (e.g., personal shopping assistants).
- Provide targeted promotions to shift them toward higher spending.