Analysis and Questions

Meet Gandhi-202518023

26-September-2025

1 Analysis and Questions

1. Optimal Clusters: What was the optimal number of clusters you identified for K-Means and Hierarchical Clustering? Justify your choices using the Elbow Method plot and the Dendrogram.

5 was the optimal number of clusters for K-Means and Hierarchical Clustering. According to Elbow, a sharp bend (in the plot of cluster vs inertia) was observed around 5. As the inertia decreases with an increase in the cluster. In the Dendrogram, a significant jump was observed after the 4 divisions, so 5 clusters was the optimal solution.

2. Cluster Comparison: Visually compare the results of the three algorithms. Did they produce similar clusters? Describe any notable differences.

K-Means and Hierarchical Clustering produced similar clusters while DBSCAN produced different ones. These can be due to its density-based nature, which is not the case for the K-Means and Hierarchical methods.

3. DBSCAN Performance: How did DBSCAN perform on this dataset? Did it identify any noise points? How did its results compare to the other methods which force every point into a cluster?

DBSCAN can identify some customers as noise. The other 2 clustering methods even consider the outlier as the point in the cluster means, while DBSCAN is best for identifying the outliers, so it produces fewer clusters.

4. Algorithm Suitability: Based on your results, which algorithm do you think was most suitable for this specific dataset and why? Consider the shape and density of the clusters.

K-Means is the best algorithm for this dataset as the shape of the data is best bit for this algorithm. On the other hand, Hierarchical Clustering performed similarly but is heavier for computation. And DBSCAN cannot be used in this case, the data is scattered and not densely populated over one region

5. Real-World Application: Describe a hypothetical real-world business scenario where the customer segments you identified could be used by the mall's marketing team. For example, how would you target the group with high income but low spending score?

The identified clusters can be directly applied in the target marketing. For Example,

- **High income, high spending:** VIP customers to be rewarded with loyalty programs and exclusive offers.
- **High income, low spending:** Potential customers to be targeted with premium product promotions and personalized discounts to encourage spending.
- **Low income, high spending:** Price-sensitive but active customers who could benefit from affordable bundles and value deals.
- **Moderate income, moderate spending:** General customers to be engaged through seasonal campaigns and regular promotions.