

Summary of Task 3: Customer Segmentation (Clustering)

Objective:

Perform customer segmentation using clustering techniques to group customers based on their profile and transaction data. Evaluate clustering performance using relevant metrics and visualize the results.

Key Findings from Clustering

1. Clustering Algorithm:

- **Agglomerative Clustering** with the **Ward linkage method** was used to segment customers into 3 clusters.

2. Clustering Evaluation Metrics:

- **Silhouette Score:** 0.0606
 - Indicates low cohesion within clusters and poor separation between clusters.
- **Calinski-Harabasz Index:** 60.34
 - Lower values indicate that the clusters are less compact and less distinct.
- **Davies-Bouldin Index:** 2.652
 - A higher value indicates less distinct clusters (ideal DBI is close to 0).

3. Dendrogram Analysis:

- The dendrogram provides a hierarchical visualization of how clusters are formed.
- It shows the distances at which clusters merge, with most data points merging at lower distances.

4. Cluster Labels:

- Each customer was assigned to one of three clusters (0, 1, 2).
 - Cluster distribution:
 - Cluster 0: Majority of customers.
 - Cluster 1: Small segment of unique customer behavior.
 - Cluster 2: A mid-sized cluster of customers with similar profiles.
-

Visualizations

1. Dendrogram:

- Hierarchical clustering results were visualized with a dendrogram.

- The Ward linkage method grouped customers based on transaction data, showing the hierarchy of customer similarities.
-

Actionable Insights

1. Cluster 0:

- Represents the largest group of customers, potentially general consumers with standard purchasing behavior.
- Focus marketing strategies on appealing to this group to maximize revenue.

2. Cluster 1:

- A small, distinct cluster that may represent premium or infrequent customers.
- Consider offering personalized loyalty programs or exclusive offers to retain this segment.

3. Cluster 2:

- A mid-sized group, possibly consisting of customers with moderate purchasing frequency.
- Design targeted campaigns or upselling opportunities for this cluster.

4. Improvement Opportunity:

- Low silhouette and Davies-Bouldin scores suggest overlapping clusters. Consider exploring more features or advanced algorithms (e.g., DBSCAN or K-Means with PCA) to improve segmentation.
-

Recommendations for Future Work

1. Feature Engineering:

- Include more features like average spend, frequency of transactions, and product preferences for better clustering.

2. Algorithm Selection:

- Try different clustering algorithms (e.g., K-Means, DBSCAN) to compare results.

3. Cluster-Specific Strategies:

- Develop personalized engagement strategies based on cluster characteristics.