

Task 3 : Customer Segmentation / Clustering

Objective

The goal of this analysis was to perform customer segmentation using clustering techniques. The analysis aimed to group customers based on transaction and customer data to uncover patterns and insights useful for targeted marketing and decision-making.

Methodology

Data Preprocessing:

- Data from customers.csv, products.csv, and transaction.csv was loaded and cleaned.
- Relevant features were extracted and scaled to ensure uniformity for clustering algorithms.

Clustering Technique:

- The K-Means clustering algorithm was utilized for this analysis.
- The optimal number of clusters was determined using the Elbow Method, evaluating inertia for various cluster counts.

Evaluation Metrics:

- Davies-Bouldin (DB) Index: Measures cluster compactness and separation.
- Silhouette Score: Indicates how well each point fits within its assigned cluster.

Key Findings

- **Number of Clusters (Optimal k): 4**
This was determined using the Silhouette score, which peaked around k=4
- **Davies-Bouldin Index: 1.384**
A lower value of this metric indicates better-defined clusters. This score suggests a moderate clustering quality.
- **Silhouette Scores:**
Scores ranged from 0.18 to 0.23 across k=2 to k=10 with the highest score at k=5.
However, k=4 is chosen as it balanced Silhouette score and interpretability.

- **Inertia values (Elbow method):**

The elbow point was observed around $k=4$, indicating diminishing returns in minimizing inertia beyond this point.

Conclusion

This clustering analysis provides actionable insights into customer behavior. The segmentation can serve as a foundation for personalized marketing, improved customer experience, and strategic decision-making. Further refinement of the model can be achieved by experimenting with additional clustering techniques or incorporating domain-specific features.