Customer Segmentation Clustering

Overview

This report summarizes the results of the customer segmentation analysis conducted using clustering techniques on the provided datasets: customer.csv, transaction.csv and product.csv .The analysis aimed to group customers based on their purchasing behavior and profile information.

Number of Clusters Formed

- Total Clusters: 4

Evaluation Metrics

1. Davies-Bouldin Index (DB Index):
    - Value: 1.0604240399923033
    - Interpretation: A lower DB Index indicates better clustering quality, as it reflects the average similarity ratio of each cluster with its most similar cluster.

2. Silhouette Score:
    - Value: 0.3135106549790539
    - Interpretation: The silhouette score ranges from -1 to 1. A value closer to 1 indicates that the samples are well clustered, while a value near or below 0 indicates overlapping clusters.

Additional Relevant Metrics

- Average Transaction Value per Cluster:

    - Cluster 0: 734.517317
    - Cluster 1: 459.001818
    - Cluster 2: 1042.066379
    - Cluster 3: 708.906416

- Number of Transactions per Cluster:
    - Cluster 0: 436
    - Cluster 1: 275
    - Cluster 2: 116
    - Cluster 3: 173

Visualizations

- Clusters were visualized using PCA (Principal Component Analysis) to reduce the dimensionality of the data. The scatter plot illustrates how customers are grouped into different clusters based on their transaction behavior.

Conclusion

The analysis successfully segmented customers into distinct groups based on their purchasing behavior and profile information. The clustering metrics indicate a reasonable quality of clustering, with potential areas for further exploration or refinement in future analyses.