Customer Segmentation - Clustering Report

Introduction

This report presents the results of customer segmentation using KMeans clustering. The objective was to group customers into distinct segments based on their transaction history and profile information. The data used for clustering includes customer spending behavior, product preferences, and the quantity of products purchased. The clustering approach helps in understanding customer behavior, which can be leveraged for targeted marketing, product recommendations, and customer retention strategies.

Methodology

For customer segmentation, KMeans clustering was used, a widely adopted technique for unsupervised learning. The features considered for clustering include:

- **Total Value**: Total money spent by each customer.

- **Quantity**: Total quantity of products purchased.

- **Product ID**: The number of unique products purchased by each customer.

These features were normalized to ensure each contributed equally to the clustering process. The optimal number of clusters was determined using the **Davies-Bouldin Index**, which measures the separation and cohesion of the clusters. A lower value indicates better clustering quality.

Clustering Results

The KMeans algorithm identified **3** distinct customer segments based on the features provided. The optimal number of clusters was selected by evaluating the **Davies-Bouldin Index**, which yielded a value of **0.7963**. This low value indicates well-separated and compact clusters, making the segmentation meaningful for business purposes.

- **Davies-Bouldin Index**: 0.7963

- **Inertia (Sum of squared distances from samples to their closest cluster center)**: 157.0369

The clustering results were validated based on the lowest Davies-Bouldin Index, which confirmed that 3 clusters represent the best segmentation of the customer base.

Clusters Visualization

A scatter plot of the clusters was created to visualize the segmentation based on the **Total Spend** and **Quantity Purchased**. Each customer was assigned to one of the three clusters, and the plot demonstrates how customers with similar spending behavior and product preferences are grouped together.

(Here, you can include the scatter plot image of the clusters you generated with KMeans.)

Conclusion

The customer segmentation using KMeans clustering effectively identified **3 distinct customer groups**. This segmentation can be used for targeted marketing campaigns, tailored product recommendations, and improving customer retention strategies. By analyzing the features driving these clusters, businesses can focus on high-value segments, develop personalized offers, and optimize the overall customer experience.

The **Davies-Bouldin Index** and **inertia** metrics validate the clustering quality, suggesting that the segments identified are meaningful and well-separated.