Clustering Report

Number of Clusters

4 clusters were formed using the KMeans clustering algorithm.

Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) value for this clustering is 1.046.

The DB Index is a measure of clustering quality where lower values indicate better-defined clusters. A value of 1.046 suggests that the clusters are reasonably well-separated and compact.

Clustering Logic

The clustering was performed using a dataset aggregated with the following features:

TotalValue: Sum of transaction values for each customer.

Quantity: Total number of items purchased by the customer.

ProductID: Number of unique products purchased by the customer.

Category: Number of unique product categories purchased by the customer.

Numerical features were normalized using StandardScaler to ensure all features had a mean of 0 and a standard deviation of 1, which prevents bias caused by different feature scales.

Principal Component Analysis (PCA) was used to reduce the dimensionality of the data to 2 components for visualization purposes.

Visualization

A scatter plot of the clusters was generated using PCA components:

Each point represents a customer.

Clusters are visualized using distinct colors.

The plot shows that the clusters are fairly well-separated, indicating effective segmentation.