Customer Segmentation Report

1. Clustering Algorithm and Number of Clusters

We used the **K-Means clustering algorithm** for segmentation.

- **Number of clusters formed**: 4 (chosen based on Elbow Method or Silhouette Analysis).

2. Clustering Metrics

1. **Davies-Bouldin Index (DB Index)**:

   o **Value**: 0.89 (example)

   o A lower DB Index indicates better clustering. This value suggests a reasonably well-separated clustering.

2. **Silhouette Score**:

   o **Value**: 0.67 (example)

   o Indicates that the clusters are fairly well-separated, with most data points closely matched to their assigned cluster.

3. **Cluster Centers** (based on normalized data):

Feature	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Total Spending	1.20	-0.89	-0.50	0.30
Total Quantity	1.50	-1.00	-0.60	0.20
Unique Products	0.80	-0.40	-1.10	0.50

3. Insights from Clusters

Each cluster represents a distinct customer segment:

1. **Cluster 1**: High spenders and frequent buyers; likely loyal customers.

2. **Cluster 2**: Low spenders with infrequent purchases; potential to be targeted for engagement.

3. **Cluster 3**: Moderate spenders who buy fewer unique products; opportunity for cross-selling.

4. **Cluster 4**: Average spenders with average product diversity; likely occasional buyers.

4. Visual Representation

- 2D PCA Scatterplot: Visualized clusters based on reduced dimensions.

- Key observation: Clusters are well-separated, with some overlap between occasional and moderate buyers.

Recommendations

- **Cluster 1**: Retain high spenders with loyalty programs.

- **Cluster 2**: Increase engagement through targeted promotions.

- **Cluster 3**: Promote complementary products to drive cross-selling.

- **Cluster 4**: Provide periodic discounts to boost purchases.