Customer Segmentation / Clustering

Clustering Algorithm Used: K-Means clustering

Number of Clusters: 4 (chosen based on the elbow method and evaluation metrics)

Features Used: Total spending per customer (Total Value), Number of transactions per customer (TransactionID), Total quantity purchased per customer (Quantity)

1. The customers were grouped into **4 distinct clusters** based on their spending and transaction behaviour.

2. **DB Index Value**

   - **Davies-Bouldin Index (DB Index)**: 0.865

   - The DB Index measures the quality of clustering. A lower value indicates better clustering. A value of **0.865** suggests that the clusters are **reasonably well-separated**, but there is some overlap between clusters.

3. **Other Relevant Clustering Metrics:**
   - **Silhouette Score**: 0.374
        The silhouette score ranges from -1 to 1. A score of **0.374** indicates that the clusters are **somewhat distinct**, but there is still significant overlap between clusters.
   - **Inertia**: 111.99
        Inertia measures the sum of squared distances of samples to their closest cluster center. A lower inertia value indicates tighter clusters. The value of **111.99** suggests that the clusters are **moderately tight**.

1. **Cluster 0**: High spenders with frequent purchases.
2. **Cluster 1**: Occasional buyers with moderate spending.
3. **Cluster 2**: Budget-conscious customers buying low-priced products in bulk.
4. **Cluster 3**: New or inactive customers with minimal activity.

Customer Clusters