Clustering Results Report

The goal was to perform customer segmentation using any clustering algorithm on customer and transactions data and identify actionable insights. I have used Kmeans algorithm and metrics like the Davies-Bouldin Index and Silhouette Score to determine the optimal number of clusters.

Key Results

1. **Optimal Number of Clusters (k)**:
    - **Best k**: 2
    - **Davies-Bouldin Index**: 0.63
    - **Silhouette Score**: 0.55
2. **Cluster Characteristics**:
    - Clear segmentation was achieved based on standardized Quantity and TotalValue.
    - Unique customer segments were identified, ranging from high-value, high-frequency customers to low-value, low-frequency ones.
3. **Visual Insights**:
    - **Scatter Plot**: Customers were segmented into clusters, with centroids marked.
    - **Cluster Sizes**: A bar chart showed the distribution of customers across clusters.
4. **Trends**:
    - Lower Davies-Bouldin Index values and higher Silhouette Scores confirmed the validity of the clusters.

Actions Taken:

- I used StandardScaler() to standardized the features to ensure equal scaling.
- Then I applied KMeans clustering for k ranging from 2 to 10.
- The best k was determined based on the lowest Davies-Bouldin Index.
- Then I created the visualizations to highlight customer segmentation and cluster sizes.

Recommendations:

- **High-Value Customers**: Company should focus on loyalty programs and premium services for the high-value customers.
- **Low-Value Customers**: Company should develop the strategies to improve engagement and retention for low-value customers.
- **Cluster-Based Offers**: Company should target the specific clusters with personalized marketing campaigns.