Task 3: Customer Segmentation / Clustering

Clustering Analysis Report

1. Number of Clusters Formed: The clustering analysis identified **4 clusters** as the optimal number of groups. This was determined using the Elbow Method, which indicated a distinct inflection point at 4 clusters.

2. Davies-Bouldin Index (DBI): The Davies-Bouldin Index for the clustering results is **1.1218**. The DBI measures the average similarity ratio of each cluster with its most similar cluster. A lower DBI value represents better clustering (compact within clusters and well-separated clusters). While a DBI of 1.1218 is acceptable, there might be room for slight improvement in separation or compactness of clusters.

3. Other Relevant Clustering Metrics:

- **Silhouette Score:** The silhouette score achieved is **0.3709**, which provides a moderate indication of cluster quality. A higher score closer to 1 indicates that the clusters are well-separated and cohesive.
- **Cluster Visualizations:** The clusters were visualized in two-dimensional space using the standardized features `TotalValue` and `AveragePrice`. This visualization highlights the degree of separation between the clusters, though some overlapping points suggest possible improvements with further feature selection or alternative clustering methods.

4. Recommendations:

- The clusters can be further refined by:
    - Exploring additional features or engineering new variables to better capture distinctions between data points.
    - Experimenting with alternative clustering algorithms (e.g., DBSCAN, Gaussian Mixture Models) to compare results.
    - Tuning the hyperparameters of the clustering algorithm for improved results.
- A further investigation into the cluster properties (e.g., distribution of `TotalValue` and `AveragePrice` within clusters) may provide actionable insights or validate the relevance of the clusters for business or research objectives.

5. Conclusions: The clustering analysis successfully identified 4 clusters with moderate compactness and separation. The Davies-Bouldin Index and Silhouette Score indicate reasonable clustering quality, but there is scope for further refinement. These results provide a meaningful basis for downstream applications such as customer segmentation, anomaly detection, or targeted analysis.