Customer Results Report

Overview

In this clustering analysis, we segmented customers based on their total spending (`TotalSpend`) and total quantity purchased (`TotalQuantity`). Other features were evaluated but did not significantly contribute to meaningful segmentation.

Methodology

- **Feature Selection**: `TotalSpend` and `TotalQuantity` were chosen as the primary clustering parameters.
- **Elbow Method**: The optimal number of clusters was determined using the **Elbow Method**, which identified **k = 2** as the best choice.
- **Cluster Visualization**: A **3D plot of customer clusters** was created to assess the separation and accuracy of clustering visually.

Clustering Metrics

To evaluate the clustering performance, several key metrics were computed:

- **Davies-Bouldin Index**: `0.63` (Lower values indicate better-defined clusters)
- **Silhouette Score**: `0.547` (Closer to 1 indicates well-separated clusters)
- **Calinski-Harabasz Index**: `333.18` (Higher values indicate better clustering)
- **Adjusted Rand Index (ARI)**: `1.0` (Perfect agreement with ground truth)
- **Mutual Information (MI)**: `1.0` (Indicates high information similarity)

Conclusion

The clustering results demonstrate **well-separated and well-defined clusters**, with **k=2** being the optimal choice. The high **Silhouette Score** and **Calinski-Harabasz Index** indicate strong cluster separation, while the **low Davies-Bouldin Index** confirms cluster compactness. The **ARI and MI values of 1.0** suggest that the clustering model aligns perfectly with an ideal classification.

Overall, this segmentation provides valuable insights into customer purchasing behavior and can be used for targeted marketing and personalized customer engagement.