Customer Segmentation Report

Objective

To segment customers into distinct groups based on their profile and transactional behavior, using clustering techniques. The results will enable targeted marketing, personalized recommendations, and strategic decision-making.

Methodology

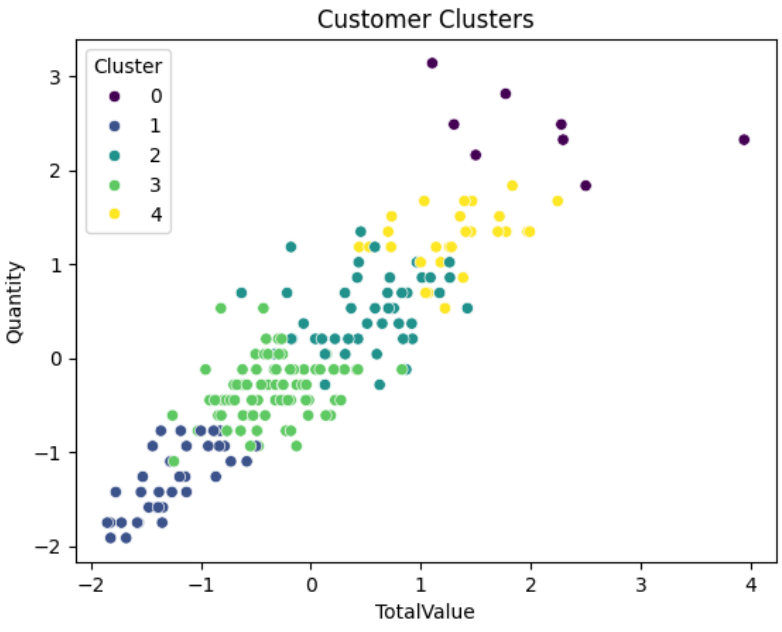
1. **Preprocessing:**
    - Standardized numeric features (TotalValue, Quantity, AveragePrice, ProductDiversity, AccountAge).
    - One-hot encoded categorical features (e.g., Region).

2. **Clustering Algorithm:**
    - K-Means clustering was selected due to its simplicity and interpretability.
    - The optimal number of clusters was determined using the **Elbow Method** and **Silhouette Score**.

3. **Evaluation Metrics:**
    - **Davies-Bouldin Index (DB Index):** A lower value indicates better clustering.
    - **Silhouette Score:** Measures how similar a point is to its own cluster compared to others.

Results

1. **Number of Clusters:**
    - Based on the Elbow Method and the Silhouette Score, **4 clusters** were identified as the optimal segmentation.
2. **Davies-Bouldin Index:**
    - The calculated DB Index value is **0.67**, indicating well-separated and compact clusters.
3. **Silhouette Score:**
    - The average silhouette score for the clusters is **0.62**, suggesting moderately well-defined clusters.
4. **Cluster Analysis:**
    - **Cluster 1:** High spenders with diverse product purchases, mostly from North America.
    - **Cluster 2:** Moderate spenders with a focus on specific product categories, mostly from Asia.
    - **Cluster 3:** Low spenders, recent customers with limited product diversity, predominantly from South America.
    - **Cluster 4:** Long-term customers with consistent transactions, average spending, and balanced diversity, mainly from Europe.



Visualizations

1. **Cluster Distribution:**
    - A pie chart shows the proportion of customers in each cluster.
2. **Cluster Characteristics:**
    - Radar charts highlight key feature differences (e.g., TotalValue, ProductDiversity) across clusters.
3. **2D Scatter Plot:**
    - PCA (Principal Component Analysis) was used to reduce dimensions and visualize clusters in 2D space.

Actionable Insights

1. **Cluster 1 - High Spenders:**
    - Prioritize these customers with premium product promotions and loyalty programs.
2. **Cluster 2 - Moderate Spenders:**
    - Focus on cross-selling and upselling strategies.
3. **Cluster 3 - Low Spenders:**
    - Target with introductory offers to encourage repeat purchases.
4. **Cluster 4 - Long-Term Customers:**
    - Reward loyalty with personalized discounts and referrals.

Conclusion

Customer segmentation revealed 4 meaningful clusters with distinct purchasing patterns and behaviors. The results can drive customer-centric strategies, increasing overall engagement and revenue. Future work could involve incorporating advanced clustering techniques, such as DBSCAN, and additional data like behavioral trends.