Clustering Report: Customer Segmentation

Overview:

This report presents the results of customer segmentation based on both customer profile data and transaction history from an eCommerce dataset. The objective was to classify customers into distinct segments using clustering techniques, calculate clustering evaluation metrics (Davies-Bouldin Index and Silhouette Score), and visualize the clusters for actionable business insights.

Clustering Technique:

The **KMeans clustering algorithm** was applied to segment customers into groups based on their transaction behavior and profile information. The algorithm was trained using normalized customer features, including total transaction value, purchase quantity, transaction frequency, and region.

1. Number of Clusters Formed:

After analyzing the results of the **Elbow Method**, it was determined that 4 clusters provided the most meaningful and stable segmentation. The Elbow Method revealed a point where the Within-Cluster Sum of Squares (WCSS) showed a sharp decrease, followed by diminishing returns, indicating the optimal number of clusters.

- **Number of Clusters**: 4

2. Clustering Evaluation Metrics:

Davies-Bouldin Index (DB Index):

The **Davies-Bouldin Index** is a measure of cluster separation where a lower score indicates better-defined clusters. The index was calculated as follows:

- **DB Index**: 1.306315421417168

This value suggests that the clusters are reasonably well-separated, with a low degree of overlap between them.

Silhouette Score:

The **Silhouette Score** quantifies how similar each sample is to its own cluster compared to other clusters. The higher the silhouette score, the more distinct and well-defined the clusters are. A value close to 1 indicates well-separated clusters, while values closer to -1 suggest poor cluster quality.

- **Silhouette Score**: 0.23682403624632775

The silhouette score indicates a strong clustering structure, with well-defined segments that are distinct from each other.

3. Visual Representation of Clusters:

Scatter Plot:

A scatter plot of **Total Value (Normalized)** vs. **Quantity (Normalized)** was created to visualize the clusters. The plot shows distinct groupings, with each cluster represented in a different color, indicating the segmentation based on customer purchase behavior.

Cluster Distribution:

The distribution of customers across the 4 clusters was visualized through a count plot, showing how the customers are distributed within each segment. This helps to understand the relative size and balance of each cluster.

4. Insights and Business Recommendations:

The clustering results revealed different types of customers, which can be leveraged for targeted marketing strategies:

- **Cluster 1**: Customers with high transaction value but low frequency. These customers are likely high-value, low-frequency buyers. Marketing strategies can focus on personalized offers to encourage more frequent purchases.

- **Cluster 2**: Customers with moderate transaction value and frequency. These are moderately engaged customers. Offering loyalty rewards or volume discounts could enhance customer retention and lifetime value.

- **Cluster 3**: Customers with low transaction value but high frequency. These customers are price-sensitive and engage in frequent smaller purchases. Offering bundling promotions or volume discounts may increase their spending.

- **Cluster 4**: Customers with low transaction value and frequency. These customers might be new or less engaged. Re-engagement strategies, such as special discounts or targeted advertisements, could help convert them into active buyers.

5. Conclusion:

The KMeans clustering model has successfully segmented the customer base into four distinct clusters, each with its own unique characteristics. By analyzing these segments, businesses can develop targeted strategies to improve customer engagement, retention, and revenue generation.