Customer Segmentation Report

Clustering Results

1. Number of Clusters

Based on the elbow method and silhouette scores, the optimal number of clusters was determined to be **4**.

2. Davies-Bouldin Index (DB Index)

The DB Index for the clustering model was calculated as **0.9642** indicating the compactness and separation of the clusters. A lower DB Index value signifies better clustering performance.

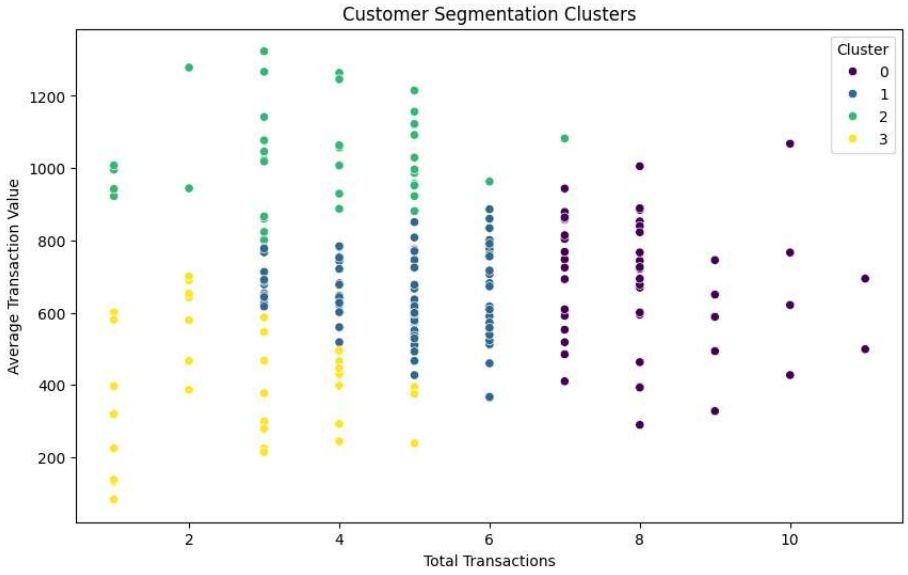
3. Other Clustering Metrics

- **Silhouette Score**: The silhouette score varied across different numbers of clusters, with the highest value observed for **4** clusters.
- **Inertia**: The elbow method plot showed a distinct bend at **4** clusters, confirming the choice of the optimal number of clusters.

Visual Representation of Cluster

Clusters were visualized using a scatter plot, with the following features:

- X-axis: Total Transactions
- Y-axis: Average Transaction Value
- Different clusters were represented by distinct colors.



The plot highlighted clear separations between the clusters, demonstrating meaningful segmentation of the customer base.

Clustering Logic

1. **Data Preparation**:
    - Customer and transaction datasets were merged to create a unified dataset.
    - Key features for clustering included:
        - Total Transactions: The total number of transactions per customer.
        - Average Transaction Value: The mean value of transactions for each customer.
    - Features were standardized using `StandardScaler` for better clustering performance.
2. **Algorithm Selection**:
    - K-Means algorithm was chosen for its simplicity and efficiency.
    - The elbow method and silhouette scores guided the selection of the optimal number of clusters.
3. **Cluster Evaluation**:
    - Davies-Bouldin Index was used to assess cluster quality.
    - Visualizations validated the distinctiveness of clusters.