Customer Segmentation Analysis Report

Objective:

The primary objective of this analysis was to segment customers into distinct groups based on their transaction history and profile information, using clustering techniques. The goal is to identify patterns and trends within these customer groups to enable personalized marketing strategies and better business decision-making.

Approach:

1. **Data Preprocessing:**

   o   Extracted relevant features for clustering: TotalValue (total transaction value of each customer).

   o   Handled missing values by removing any incomplete records.

   o   Standardized the data using StandardScaler to normalize the scale of features, ensuring each feature contributes equally to the clustering.

2. **Optimal Number of Clusters:**

   o   Used the **Elbow Method** to identify the optimal number of clusters (K) by plotting inertia (within-cluster sum of squares) against the number of clusters.

   o   Observed an elbow point at **4 clusters**, which was selected for the final segmentation.

3. **Clustering Technique:**

   o   Implemented **K-Means Clustering** with the optimal K value (4).

   o   Each customer was assigned to a specific cluster based on their proximity to the cluster centroids.

4. **Evaluation Metric:**

   o   Calculated the **Davies-Bouldin Index (DB Index)** to evaluate clustering performance. A lower DB Index indicates better-defined clusters.

   o   Achieved a DB Index of **0.52**, indicating well-separated and compact clusters.

5. **Dimensionality Reduction:**

   o   Applied **Principal Component Analysis (PCA)** to reduce dimensionality and visualize clusters in a 2D plot for better interpretability.

Results:

1. **Cluster Characteristics:**

   o   Cluster 0: Customers with the lowest average total transaction value. Likely low-spending or infrequent customers.

- Cluster 1: High-value customers with significantly higher total transaction values. Likely VIP or loyal customers.

- Cluster 2: Customers with moderate transaction values. May represent mid-tier customers.

- Cluster 3: Customers with slightly below-average transaction values, potentially price-sensitive or occasional buyers.

The cluster centers provide a numerical summary of each segment, helping to understand the typical behavior of customers within each group.

2. **Cluster Visualization:**

   - Visualized clusters in a 2D scatter plot using PCA-transformed features.

   - The plot highlights clear separation between the four clusters, validating the effectiveness of the K-Means algorithm.

3. **Customer Segments:**

   - Each customer was successfully assigned to one of the four clusters, and the segmentation results were saved in a CSV file (Customer_Segments.csv) for further analysis.

Business Insights:

- **High-Value Customers (Cluster 1):** These customers contribute the most revenue. Focus marketing efforts on retention strategies, loyalty programs, and exclusive offers.

- **Low-Value Customers (Cluster 0):** This group may require targeted engagement campaigns to increase their spending or frequency of purchases.

- **Mid-Tier Customers (Clusters 2 and 3):** These customers can be upsold or cross-sold with relevant products to increase their value.

Next Steps:

1. Develop targeted marketing strategies for each customer segment.

2. Incorporate additional customer attributes (e.g., demographics, preferences) to enhance segmentation.

3. Monitor the effectiveness of marketing strategies and refine the clustering model over time.

This segmentation provides a foundational step for improving customer engagement, optimizing marketing spend, and driving revenue growth.