Customer Segmentation Clustering Report

This report documents the results of a customer segmentation study using K-Means clustering. The dataset includes customer profiles **(Customers.csv)** and transactional data **(Transactions.csv)**. The study aimed to group customers into distinct clusters based on their transaction behavior and profile information.

2. Methodology

1. **Data Preprocessing:**
   - Merged customer and transaction data.
   - Aggregated transactional data to compute:
     - Total spending.
     - Average transaction value.
     - Number of transactions.
   - Filled missing values with 0 and normalized features using StandardScaler.
2. **Elbow Method:**
   - Used the Sum of Squared Distances (SSD) to identify the optimal number of clusters (k) based on the "elbow point."
3. **Clustering:**
   - Performed K-Means clustering with the optimal k.
   - Reduced dimensions using Principal Component Analysis (PCA) for visualization.
4. **Evaluation:**
   - Assessed clustering quality using **Davis-Bouldin (DB) Index**.
   - Analyzed the clusters using visualizations and statistics.

3. Results

1. **Optimal Number of Clusters:**
   - Based on the Elbow Method, the optimal number of clusters was determined to be **{optimal_k}**.
2. **DB Index:**
   - The **Davis-Bouldin (DB) Index** was calculated as **{db_index:.2f}**.
     - The DB Index is a measure of cluster compactness and separation; lower values indicate better clustering performance.
3. **Cluster Statistics:**
   - Distribution of customers across clusters:
   - Cluster 0: {n_customers_0} customers
   - Cluster 1: {n_customers_1} customers
   - Cluster 2: {n_customers_2} customers
   - ...
4. **PCA Visualization:**
   - The clusters were visualized in 2D using PCA, showing distinct separations between groups.

4. Interpretation of Clusters

Each cluster represents a unique customer segment based on transaction behavior and profile metrics:

- **Cluster 0:**
  - High total spending, high average transaction value, frequent transactions.
  - Represents "high-value customers."
- **Cluster 1:**
  - Moderate spending, average transaction value, infrequent transactions.
  - Represents "occasional shoppers."
- **Cluster 2:**
  - Low total spending, low average transaction value, rare transactions.
  - Represents "low-value customers."

5. Recommendations

1. **Personalized Marketing:**
   - Target high-value customers (Cluster 0) with exclusive rewards and loyalty programs.
   - Encourage occasional shoppers (Cluster 1) with personalized discounts.
   - Re-engage low-value customers (Cluster 2) with promotional offers.
2. **Strategic Insights:**
   - Focus on growing the high-value customer base by identifying similar potential customers.
3. **Further Refinements:**
   - Include additional features like product categories or customer demographics for deeper insights.

6. Visualizations

1. **Elbow Method:**
   - (Include the Elbow Method plot here.)
2. **Cluster Visualization:**
   - (Include the PCA scatter plot here.)