CLUSTERING REPORT

Objective

The goal of this clustering exercise was to segment customers based on their profile and transaction data to uncover patterns and create actionable groups. This involved:

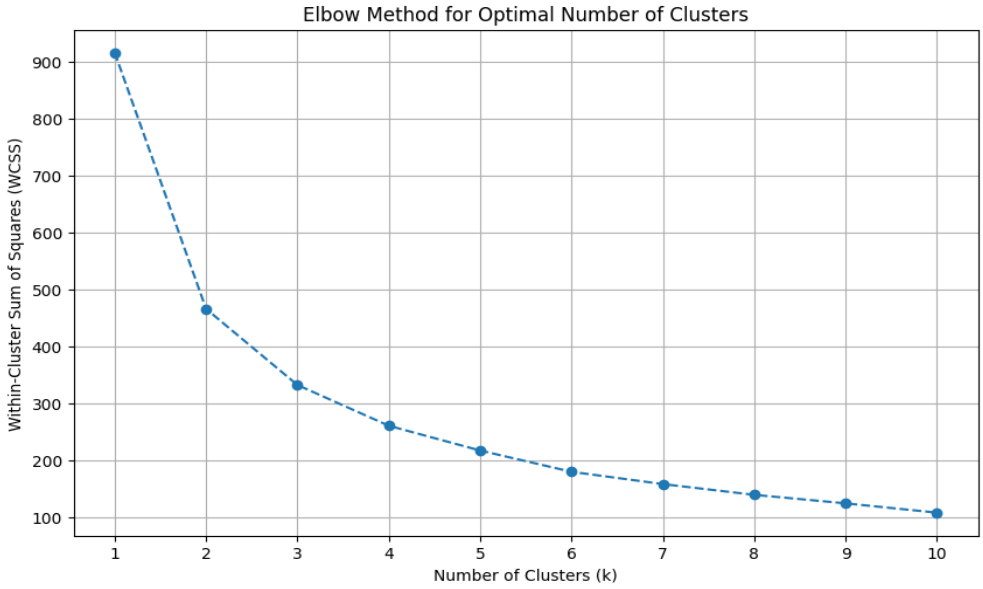
1. Determining the optimal number of clusters.
2. Evaluating clustering quality using metrics such as the **Davies-Bouldin Index (DB Index)** and **Silhouette Score**.
3. Visualizing the clusters and analyzing their characteristics.

Data Preparation

- **Datasets Used**:
    - Customers.csv
    - Transactions.csv
- **Feature Engineering**:
    - Aggregated features included:
        - Total spending.
        - Average transaction value.
        - Total quantity purchased.
        - Number of unique product.
        - Number of transactions.
    - Features were normalized using **StandardScaler** to ensure compatibility across metrics.

Methodology

1. **Clustering Algorithm**:
    - K-Means clustering was selected for its efficiency and effectiveness with numerical data.
2. **Elbow Method**:
    - The Elbow Method was used to determine the initial range of potential cluster numbers. The optimal cluster range was identified as **4 to 6** clusters.



3. **Clustering Evaluation**:
    - Clustering was evaluated using:
        - **Davies-Bouldin Index (DB Index)**: Lower values indicate better clustering.
        - **Silhouette Score**: Higher values indicate better-separated and more compact clusters.

Results

Number of Clusters Formed

After evaluating metrics and visualizing clusters, **4 clusters** were identified as the optimal number. This number balanced compactness, separation, and interpretability.

Evaluation Metrics

- **Davies-Bouldin Index**: **0.898**
    - Indicates well-separated clusters with minimal overlap.
- **Silhouette Score**: **0.348**
    - Suggests moderately compact and well-separated clusters.

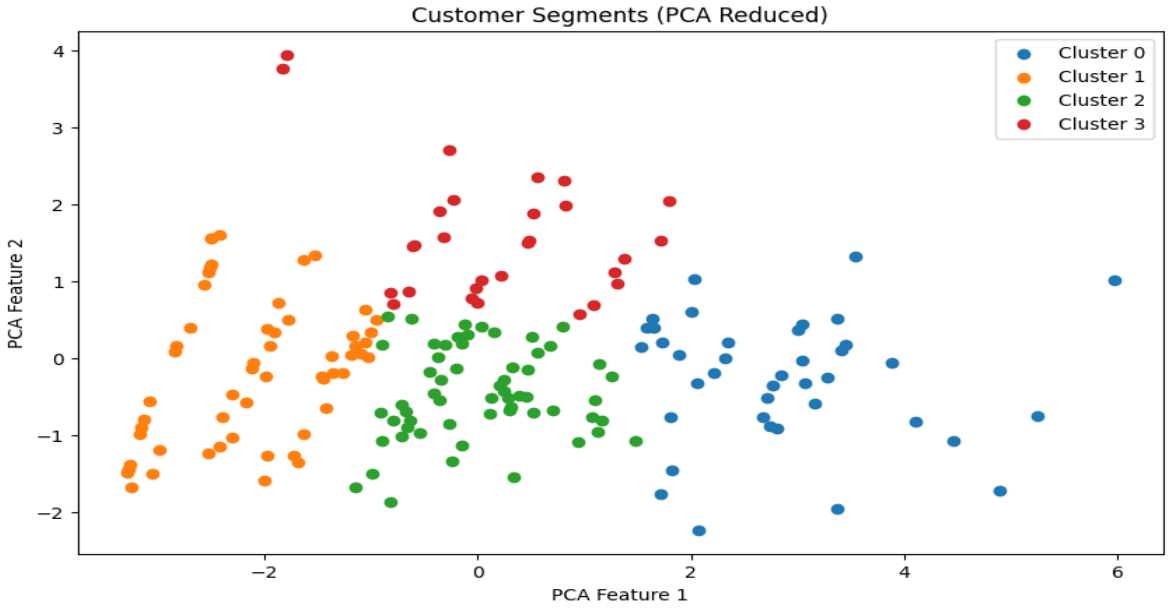
Cluster Composition

Each cluster exhibited distinct characteristics:

1. **Cluster 0:** High-value, frequent buyers with diverse product preferences.
2. **Cluster 1:** Low-value, infrequent buyers with minimal product diversity.
3. **Cluster 2:** Moderate spenders with steady transaction frequency and decent variety.
4. **Cluster 3:** High spenders with lower purchase frequency and limited product variety.

Visualization

The clusters were visualized using **PCA-reduced 2D scatter plots** for interpretability. Each cluster was clearly distinguishable, indicating the effectiveness of the clustering process. Below is an example of the cluster visualization:



Actionable Insights

1. **Cluster 0 (High-Value, Engaged Customers)**
    - Key Traits: Highest spending, most frequent transactions, largest product diversity.
    - Action: Focus on retention with loyalty programs and exclusive offers.

2. **Cluster 1 (Low-Value, Infrequent Customers)**
    - Key Traits: Lowest spending, least transactions, minimal product diversity.
    - Action: Target with promotions or personalized campaigns to increase engagement.

3. **Cluster 2 (Moderate Spenders)**
    - Key Traits: Moderate spending, steady transaction frequency, decent product variety.
    - Action: Encourage higher spending via upselling or cross-selling strategies.

4. **Cluster 3 (High-Spenders with Lower Frequency)**
    - Key Traits: High transaction value but less frequent purchases and lower product variety.
    - Action: Offer subscription models or discounts to encourage repeat purchases.