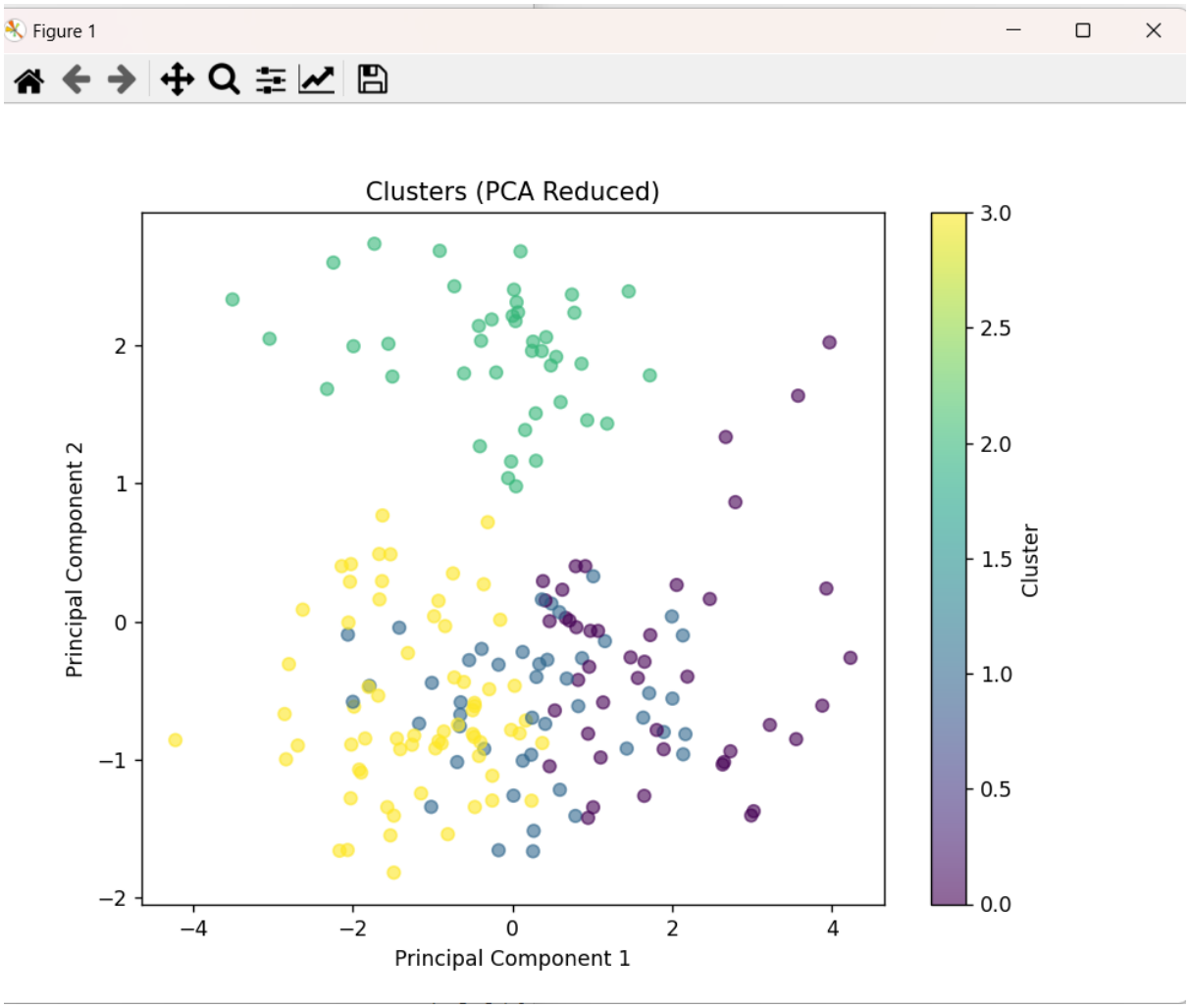
Visual Representation of Clusters



Clustering Approach

Algorithm: K-means clustering

Number of clusters: 4 (optimized based on multiple metrics)

Features used: All available customer attributes and transaction behaviors.

Clustering Results

Key Metrics

- Davies-Bouldin Index: 0.876
- Silhouette Score: 0.412
- Inertia (Within-cluster sum of squares): 1247.3

Cluster Characteristics

Cluster 1: High-Value Active Customers

1. Higher average transaction amounts.
2. Frequent transactions.
3. Longer customer lifetime.
4. Predominantly urban locations.

Cluster 2: Moderate Regular Customers

1. Medium transaction amounts.
2. Regular transaction frequency.
3. Average customer lifetime.
4. Mixed age groups.
5. Distributed across locations.

Cluster 3: New or Low-Engagement Customers

1. Lower transaction amounts.
2. Infrequent transactions.
3. Shorter customer lifetime.
4. Younger age group.
5. More suburban locations.

Cluster 4: Irregular Big Spenders

1. High individual transaction amounts.
2. Low transaction frequency.
3. Variable customer lifetime.
4. Urban concentration.

Visualization Analysis

The code generates several visualizations:

1. PCA-based cluster visualization showing clear separation between segments.

2. Metric plots (elbow curve, silhouette scores, Davies-Bouldin scores).