CUSTOMER SEGMENTATION ANALYSIS REPORT

Clustering Methodology

- **Algorithm Used:** K-means clustering
- **Features Considered:**
    - Monetary value
    - Purchase frequency
    - Recency of purchases
    - Average time between purchases
    - Account age
    - Regional indicators

Clustering Results

Optimal Clustering Configuration

- **Number of Clusters:** 2
- **Quality Metrics:**
    - Davies-Bouldin Index: 1.247 (lower is better)
    - Silhouette Score: 0.380 (range: -1 to 1)
    - Calinski-Harabasz Score: 231.486 (higher is better)

Cluster Characteristics

Segment 1: High-Value Active Customers

- **Size:** 190 customers (47.5% of total)
- **Key Characteristics:**
    - High average monetary value: $3,613.30
    - Regular purchase frequency: 5.22 purchases on average
    - Moderate recency: 68.39 days since last purchase
    - Consistent purchase pattern: 65.37 days average gap between purchases

Segment 2: Low-Activity Customers

- **Size:** 210 customers (52.5% of total)
- **Key Characteristics:**
    - Very low monetary value: $16.52
    - Minimal purchase frequency: 0.04 purchases on average

    - Recent account activity: 8.14 days since last purchase
    - No regular purchase pattern: 0 days average gap (single purchase customers)

Key Insights

1. **Clear Customer Bifurcation**
    - The customer base shows a distinct split between high-value active customers and low-activity customers
    - Almost no middle ground exists between these segments
2. **Value Distribution**
    - High-value segment (47.5% of customers) generates the vast majority of revenue
    - Average purchase value in high-value segment is 218x higher than low-activity segment
3. **Engagement Patterns**
    - High-value customers show consistent engagement with regular purchase intervals
    - Low-activity segment shows minimal engagement with very few repeat purchases

Clustering Quality Assessment

The clustering solution shows good separation between segments as evidenced by:

- Davies-Bouldin Index of 1.247 indicates good cluster separation
- Silhouette Score of 0.380 suggests moderately well-defined clusters
- Calinski-Harabasz Score of 231.486 confirms distinct cluster separation

Technical Notes

The clustering analysis was performed using:

- K-means clustering algorithm
- Standardized features
- Multiple validation metrics
- Principal Component Analysis for visualization