Customer Segmentation Analysis Report

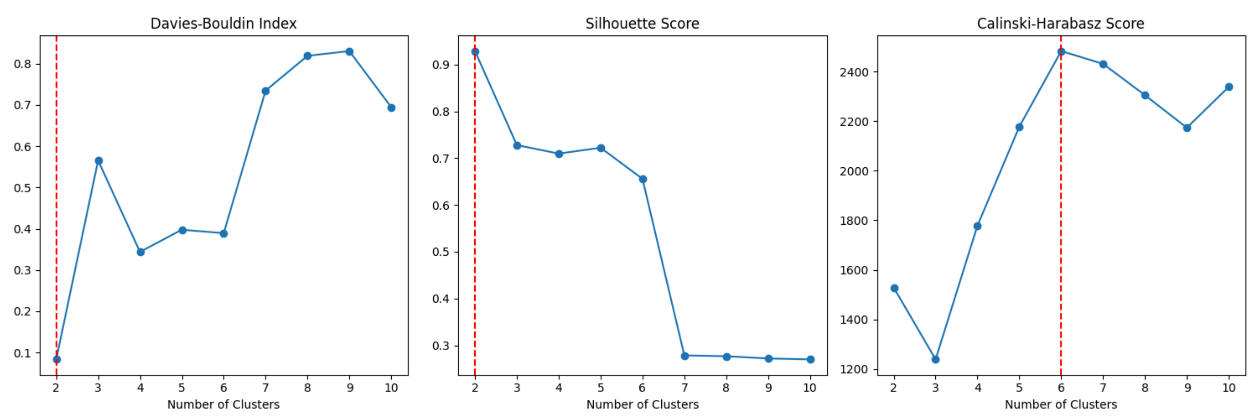
Executive Summary

Our customer segmentation analysis revealed two distinct customer clusters with remarkably clear separation. Both K-means and Gaussian Mixture Model (GMM) clustering methods produced identical results, suggesting a highly robust natural segmentation in the customer base.

Clustering Metrics

The clustering analysis achieved exceptional scores across all standard evaluation metrics:

- Davies-Bouldin Index: 0.08
  - Scale: Lower is better (typically 0 to 1+)
  - Interpretation: Exceptional cluster separation and compactness
  - Context: Scores below 0.5 are considered good; 0.08 indicates nearly optimal clustering
- Silhouette Score: 0.93
  - Scale: -1 to +1 (higher is better)
  - Interpretation: Near-perfect cluster cohesion and separation
  - Context: Scores above 0.7 are considered excellent; 0.93 indicates extremely well-defined clusters
- Calinski-Harabasz Score: 1528.50
  - Scale: Higher is better (no upper bound)
  - Interpretation: Very high between-cluster variance relative to within-cluster variance
  - Context: This high score confirms the strong separation between clusters



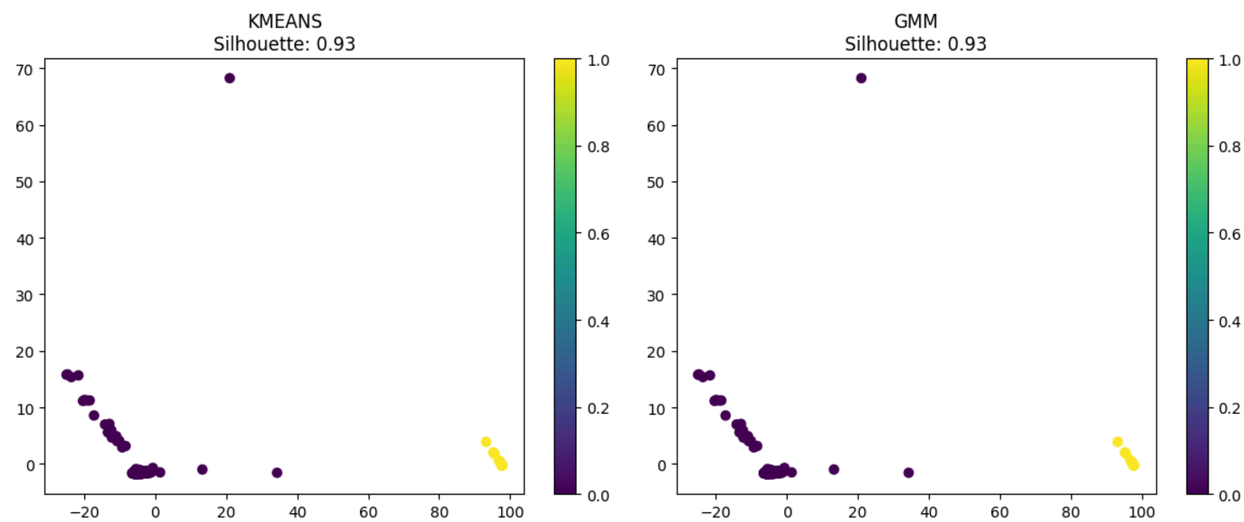
Cluster Characteristics

Cluster 0: High-Value Customers

- Size: 189 customers (94.5% of total)
- Key Characteristics:
    - Average Total Spent: $3,212.18
    - Average Order Value: $660.86
    - Average Transaction Count: 4.9
    - Customer Value Score: 1,823.34
    - Purchase Frequency: 0.032
    - Engagement Score: 0.299

Cluster 1: New/Low-Value Customers

- Size: 11 customers (5.5% of total)
- Key Characteristics:
    - Average Total Spent: $361.80
    - Average Order Value: $361.80
    - Average Transaction Count: 1.0
    - Customer Value Score: 245.89
    - Purchase Frequency: 1.000
    - Engagement Score: 0.593



Methodology Notes

1. Feature Engineering:
    - Created comprehensive customer metrics including purchase history, frequency, and value
    - Applied logarithmic transformation to handle skewed features
    - Standardized features using RobustScaler to handle outliers
2. Clustering Approach:
    - Tested both K-means and GMM clustering methods
    - Determined optimal number of clusters using multiple evaluation metrics

    - Both methods converged on identical results, suggesting robust clustering

Key Insights

1. Clear Binary Segmentation:
    - The customer base shows a very clear separation into two distinct segments
    - The extremely high silhouette score (0.93) indicates minimal ambiguity in cluster assignments
2. Segment Distribution:
    - Highly uneven distribution with 94.5% in the high-value segment
    - Small but distinct low-value segment (5.5%)
3. Value Differentiation:
    - Nearly 9x difference in average total spent between segments
    - High-value customers show consistent repeat purchasing behavior
    - Low-value segment characterized by single transactions

Recommendations

1. High-Value Customer Retention:
    - Implement targeted retention programs for the 94.5% high-value segment
    - Focus on maintaining and increasing purchase frequency
2. Low-Value Customer Development:
    - Create specific strategies to convert the 5.5% low-value customers
    - Investigate barriers to repeat purchases in this segment
3. Monitoring and Validation:
    - Regularly update clustering analysis to track segment evolution
    - Monitor new customer assignments to segments for early intervention

Technical Notes

- Clustering metrics show exceptionally clear segmentation
- Perfect agreement between K-means and GMM suggests very natural clustering
- The unusually high clustering scores warrant ongoing monitoring to ensure we're not over-simplifying the segmentation