Customer Segmentation: Clustering Results Report

Methodology

Feature Engineering

The analysis uses a combination of customer profile and behavioral features:

Customer Profile Features:

- Customer Age (days since signup)
- Geographic Region (one-hot encoded)

Behavioral Features:

- Recency (days since last purchase)
- Frequency (total number of transactions)
- Monetary Value:
    - Total Spend
    - Average Purchase Value

Clustering Approach

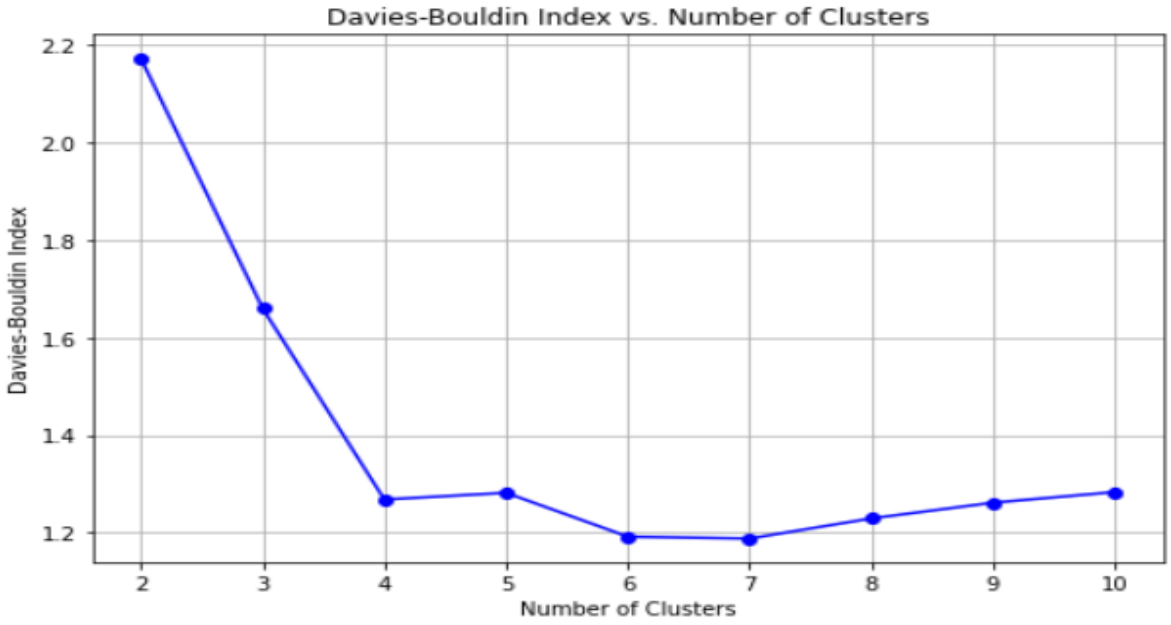
We employed K-means clustering with the following steps:

1. Feature standardization using StandardScaler
2. Testing cluster numbers from 2 to 10
3. Evaluation using multiple clustering metrics

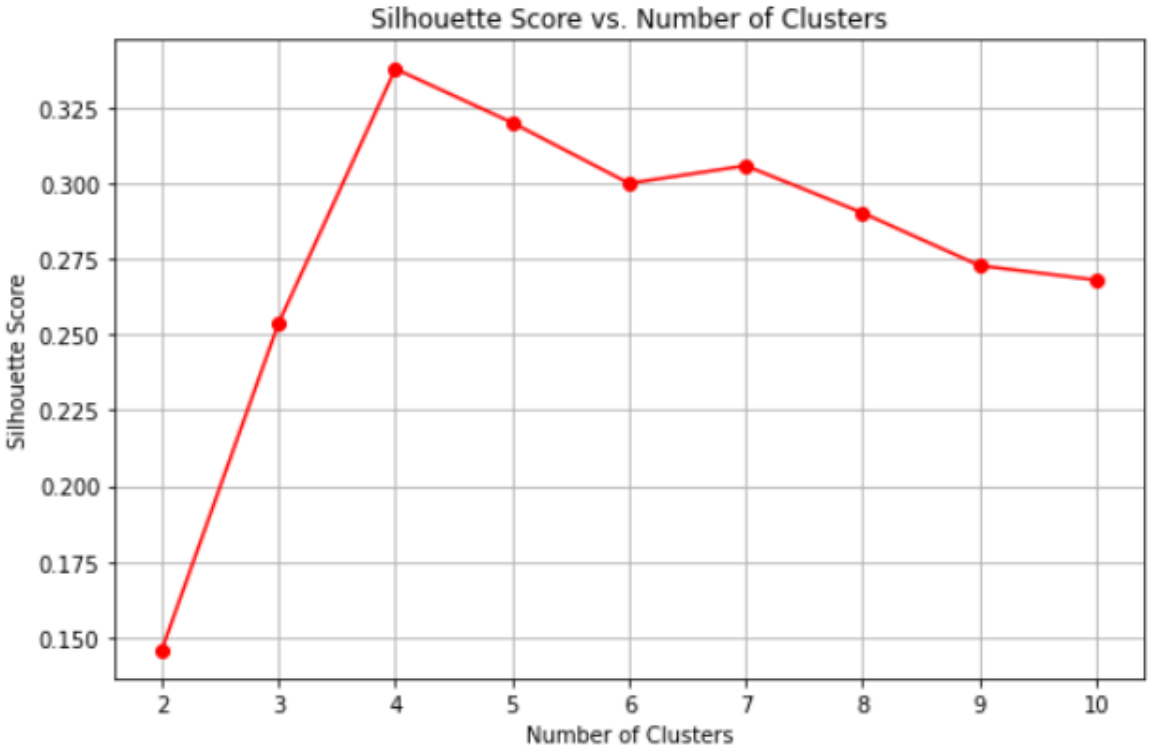
Clustering Metrics

Our analysis uses four key metrics to determine the optimal number of clusters:

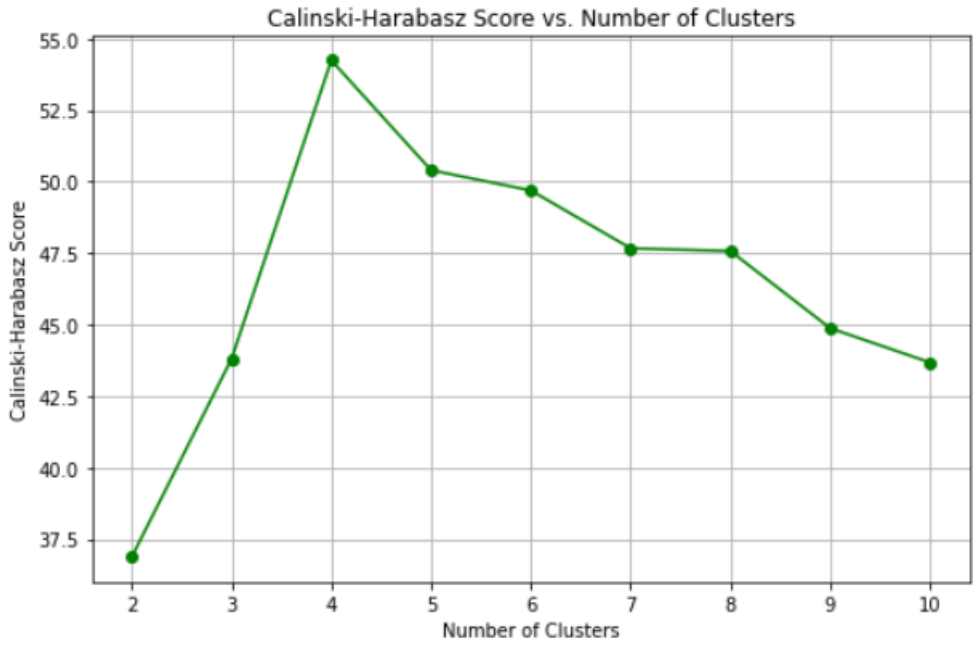
1. **Davies-Bouldin Index (Primary):**
    - Measures average similarity between clusters
    - Lower values indicate better clustering
    - Particularly useful for evaluating cluster separation



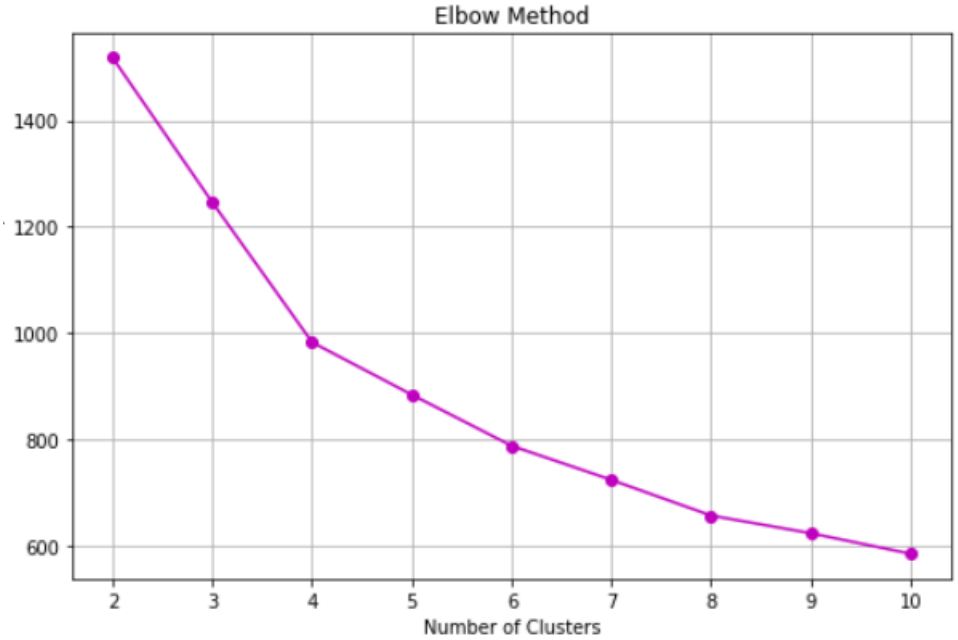
2. **Silhouette Score:**
    - Measures how similar objects are to their own cluster compared to other clusters
    - Range: -1 to 1 (higher is better)
    - Values near 1 indicate well-defined clusters



3. **Calinski-Harabasz Index:**
    - Ratio of between-cluster to within-cluster dispersion
    - Higher values indicate better-defined clusters
    - Useful for dense, well-separated clusters

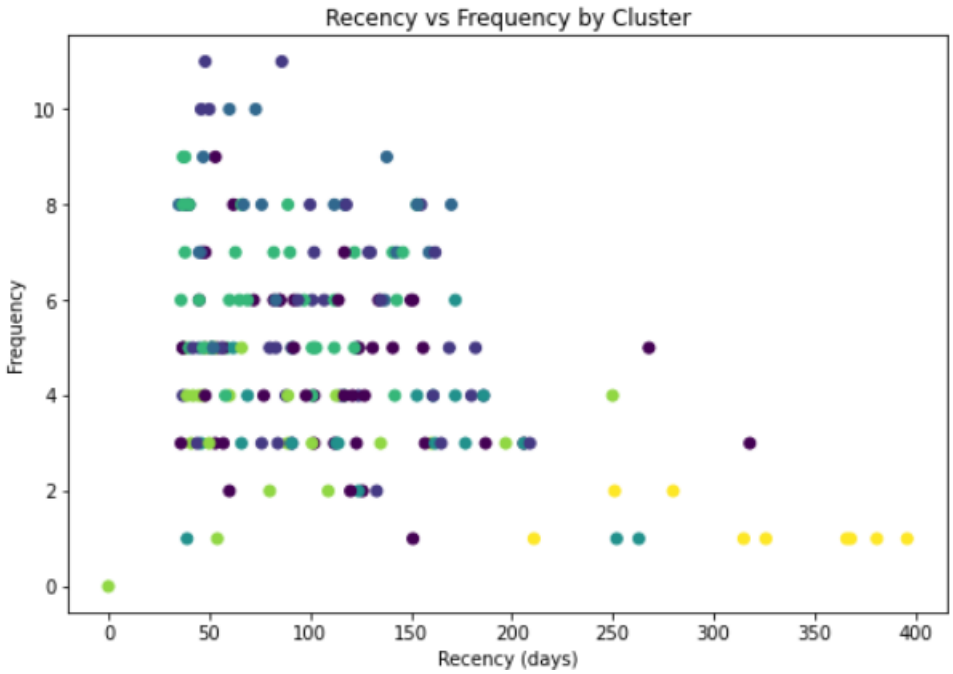


4. **Elbow Method (WCSS):**
    - Shows within-cluster sum of squares
    - Helps identify the point where adding more clusters provides diminishing returns

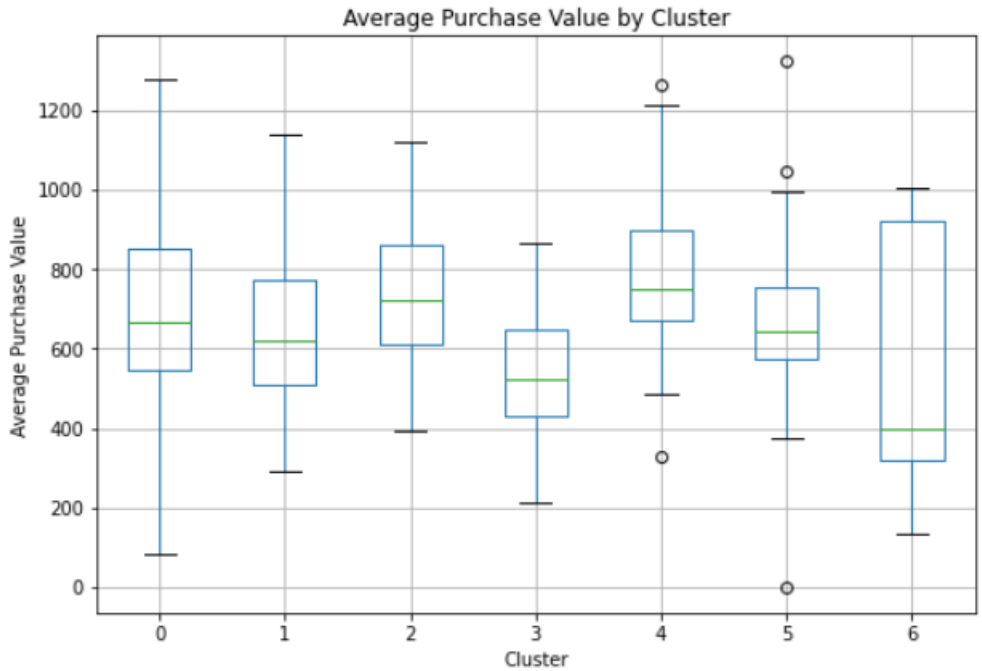


Clustering Results & Insights

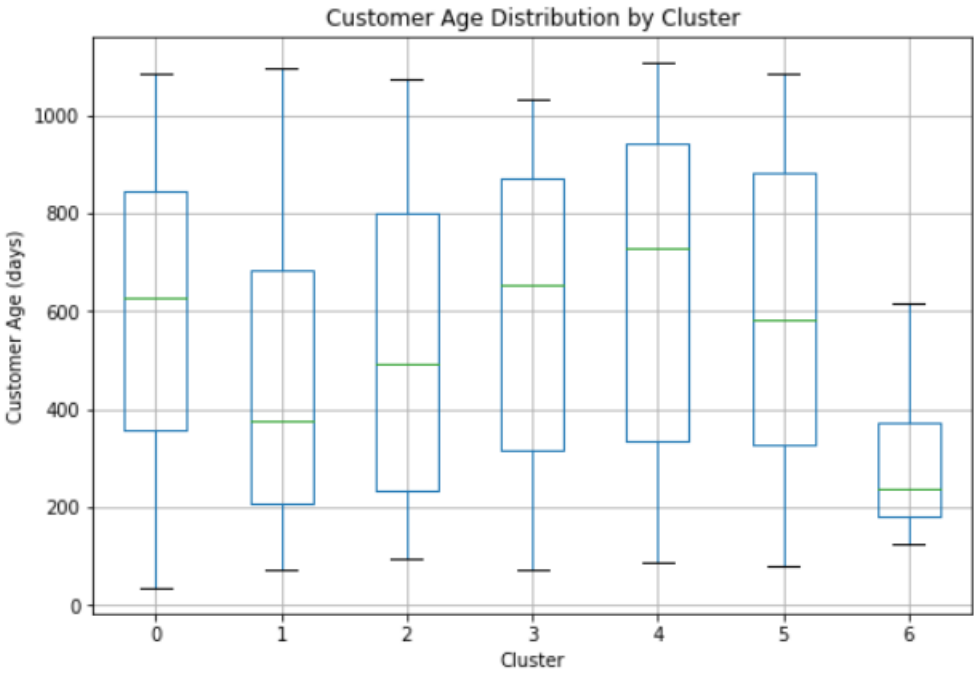
- **Number of Clusters**: 7
    - **DB Index**: 1.188
    - **Silhouette Score:** 0.306
    - **Calinski-Harabasz Index:** 47.672
    - **Elbow Method (WCSS):** 725.214
- **Key Observations from Cluster Analysis**:
    - **Recency vs. Frequency Scatter Plot**: Showed distinct purchasing patterns across clusters.



    - **Average Purchase Value by Cluster**: Indicated variations in spending behaviour.



- **Customer Age Distribution by Cluster**: Revealed demographic trends.



- **Cluster Centre's Heatmap**: Highlighted key distinguishing features across clusters.