Customer Segmentation Report

Overview:

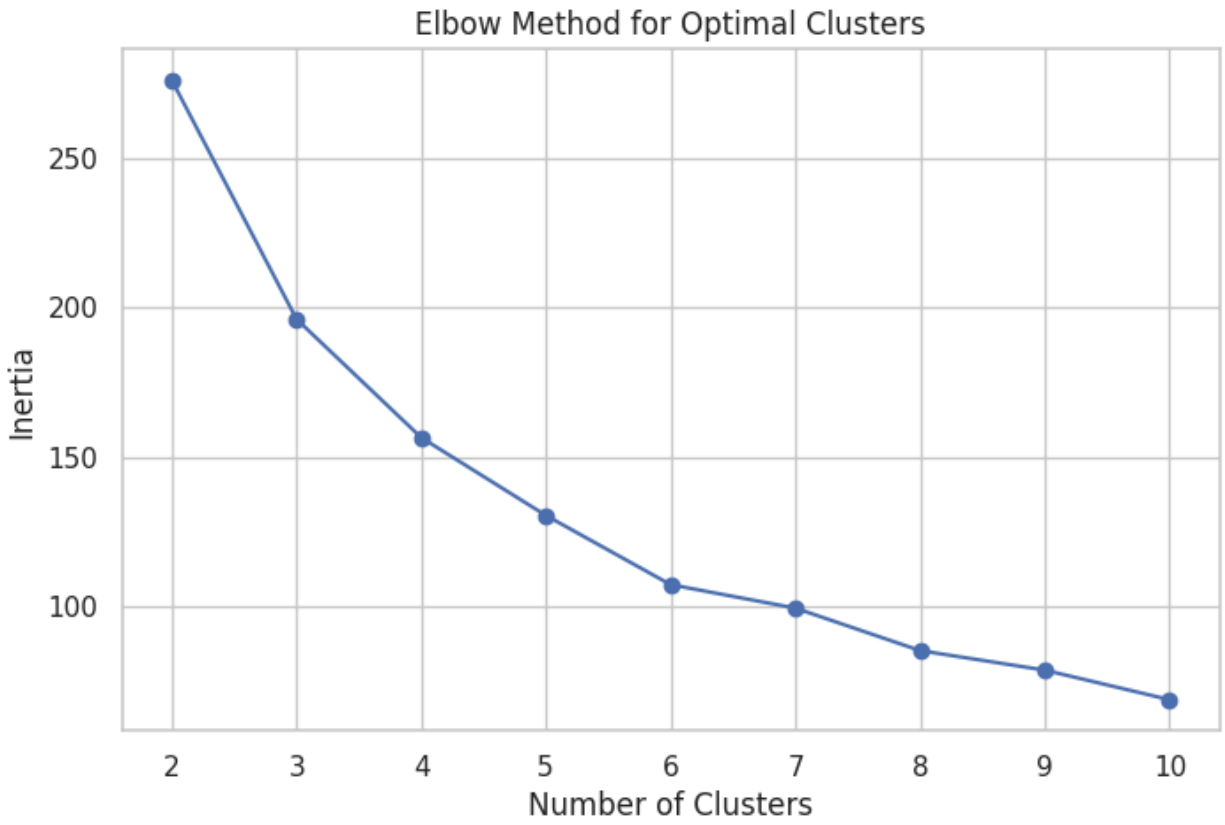
The goal of this analysis was to segment customers based on their transaction behavior to gain actionable insights. K-Means clustering was selected due to its simplicity, speed, and suitability for large datasets. The segmentation utilized customer frequency, total spending, and unique product categories purchased as the key features. Although the elbow method suggested 6 clusters initially, the evaluation metrics (Davies-Bouldin Index and Silhouette Score) indicated that 2 clusters were optimal.

Clustering Metrics and Scores

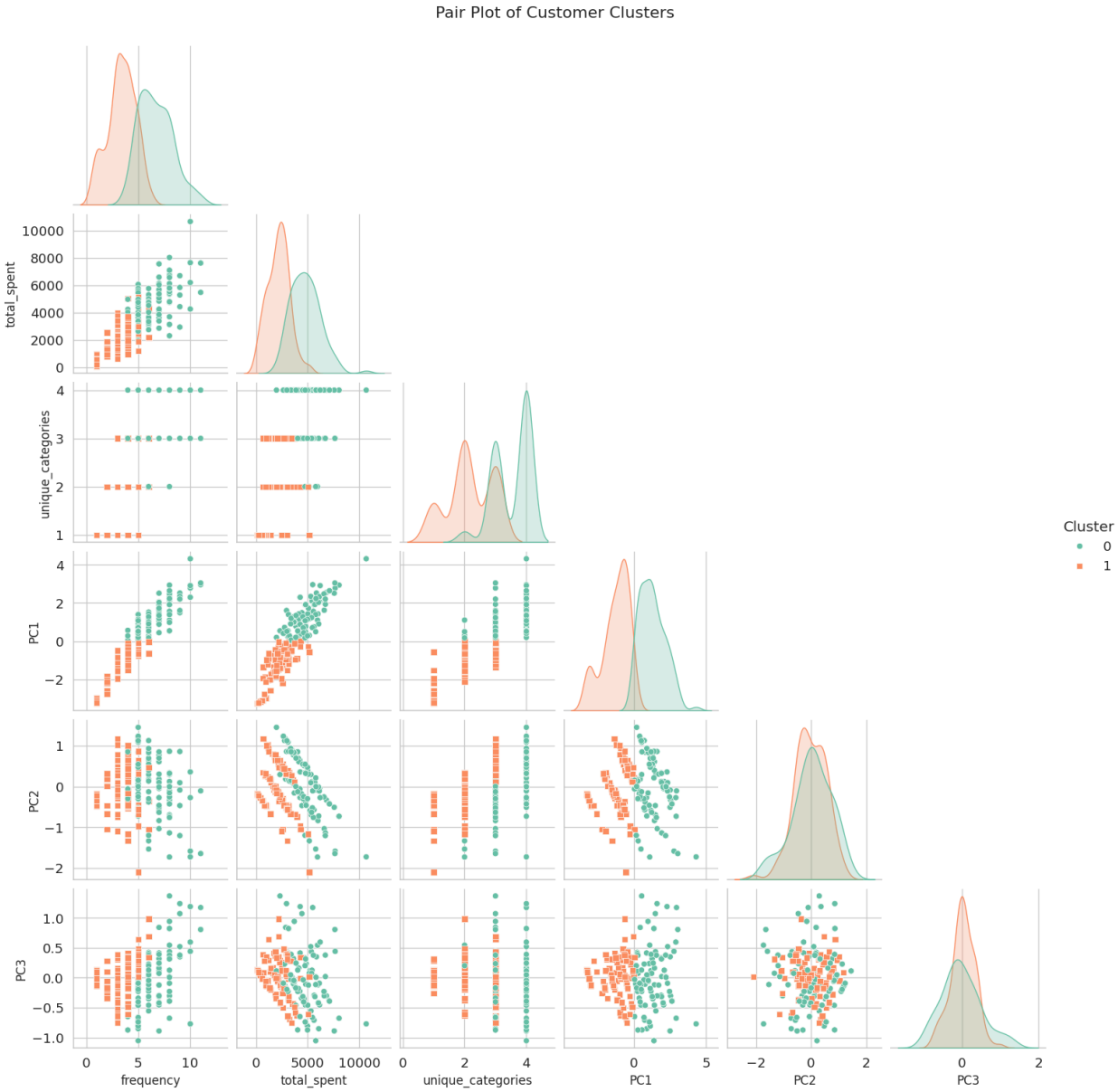
1. **Initial Observations**:
    - Using the elbow method, 6 clusters seemed appropriate as the inertia significantly decreased up to this point.
    - However, the clustering quality metrics guided the final decision to use 2 clusters.
2. **Evaluation Scores**:
    - **2 Clusters**:
        - Davies-Bouldin Index (DB): **0.8453**
        - Silhouette Score: **0.4330**
    - **6 Clusters**:
        - Davies-Bouldin Index (DB): **0.8946**
        - Silhouette Score: **0.3848**

Visual Representation

1. **Elbow Method Plot**:


Elbow Method for Optimal Clusters
Inertia
250
200
150
100
2
3
4
5
6
7
8
9
10
Number of Clusters

2. **Cluster Visualization**:



Conclusion By adopting K-Means clustering with 2 clusters, we achieved a balance between cluster quality and interpretability. Cluster 0 represents high-value customers, while Cluster 1 comprises low-engagement customers. These insights can guide targeted marketing strategies, such as premium offers for high-value customers and re-engagement campaigns for low-frequency buyers. Further exploration of cluster characteristics can refine these strategies for maximum impact.