Customer Segmentation Report: Clustering Analysis

This report summarizes the customer segmentation process using clustering techniques. Key results and insights are detailed below.

Methodology

Clustering Algorithm: K-Means **Optimal Number of Clusters:** 8 (determined using the Davies-Bouldin Index)

1.Data Preprocessing:

- Aggregated transactional and profile data into customer-level features.
- Applied one-hot encoding for categorical features (e.g., Region).
- Standardized numerical features to ensure uniform scaling.

2. Evaluation Metrics

- **Davies-Bouldin Index (DB Index):** 1.3553 (optimal configuration with K=8).
- **Silhouette Score:** 0.2434 (provides additional context for cluster cohesion).
- **Within-Cluster Sum of Squares (WCSS):** Utilized in the Elbow Method to evaluate the appropriate K.

3. Results

Optimal Clusters:

- **Number of Clusters:** 8

Cluster Distribution:

- Cluster 4: 36 customers
- Cluster 5: 32 customers
- Cluster 3: 27 customers
- Cluster 1: 25 customers
- Cluster 2: 25 customers

- Cluster 0: 22 customers
- Cluster 7: 21 customers
- Cluster 6: 11 customers

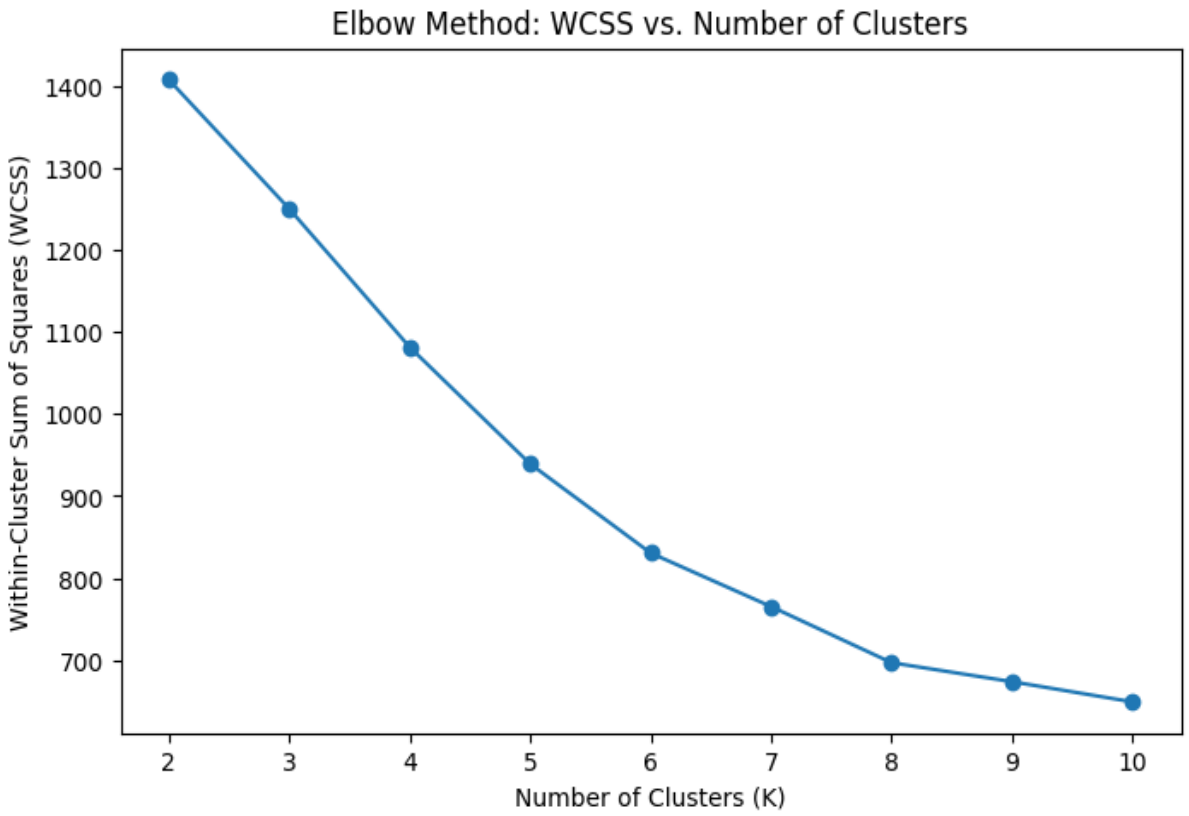
Cluster Characteristics: Clusters were formed based on features such as transaction frequency, total spending, recency, and tenure. Key characteristics include:

- **Cluster 4:** The largest cluster, consisting of high-spending customers with moderate recency.
- **Cluster 6:** The smallest cluster, likely representing high-value, low-frequency customers.
- **Cluster 7:** Represents customers with recent activity and moderate spending patterns.

4. Visualizations

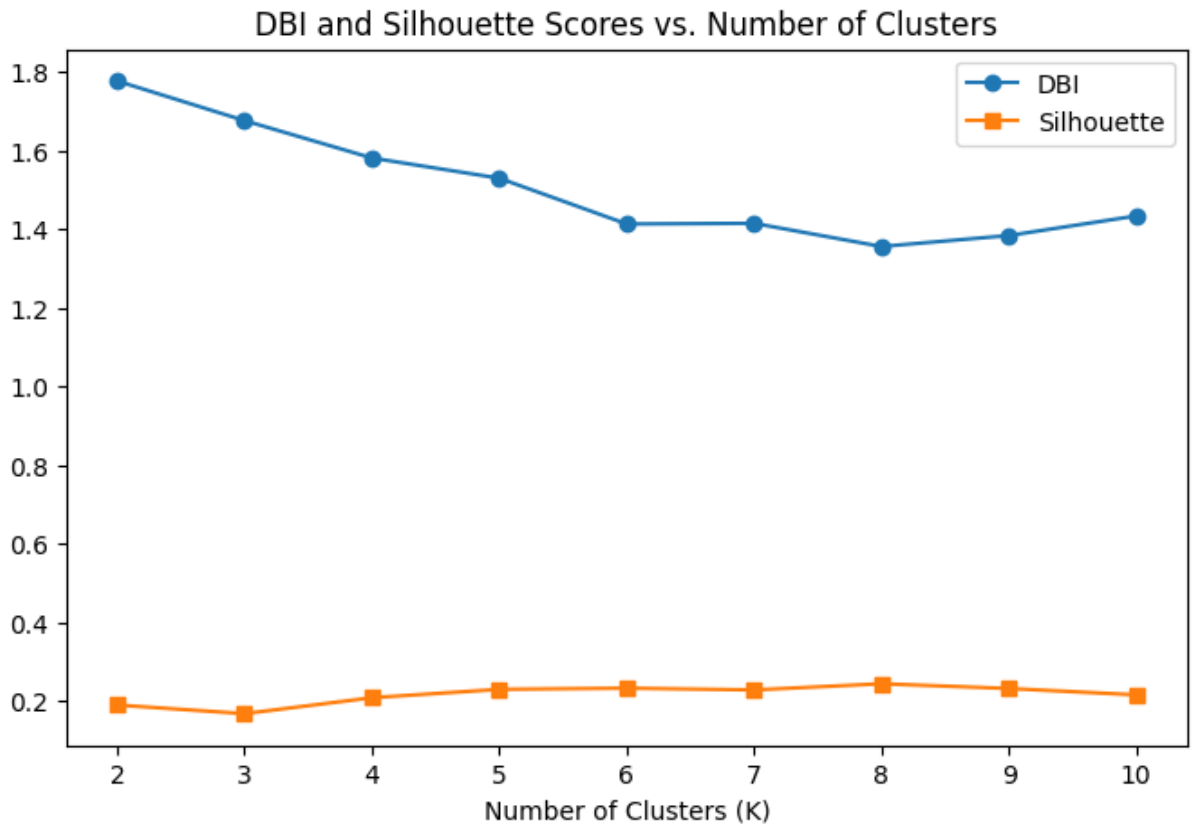
Elbow Method:

- Visualized WCSS vs. Number of Clusters to confirm K=8 as optimal.



DBI and Silhouette Analysis:

- Plots comparing DBI and Silhouette Scores across K=2 to K=10 reinforced the choice of K=8.



Cluster Distribution:

- A scatter plot using PCA (2D projection) displayed distinct customer clusters and areas of overlap, offering insights into cluster separability.

5. Conclusion

The clustering analysis effectively segmented the customer base into 8 clusters, achieving a low Davies-Bouldin Index (1.3553) while maintaining reasonable cluster separation (Silhouette Score: 0.2434). These clusters provide valuable insights for strategic decision-making:

- **Cluster 4:** Largest group, consisting of high-spending customers who warrant attention for loyalty programs.
- **Cluster 6:** Smallest group, representing high-value customers who may benefit from personalized engagement.