Customer Segmentation Clustering Report

Overview

This report provides an in-depth analysis of customer segmentation derived from the customer transaction dataset. The clustering approach aims to identify groups of customers with similar purchasing behaviours based on total spending and transaction frequency. The insights can support data-driven marketing strategies and customer engagement initiatives.

Key Results

1. Number of Clusters Formed

- **Optimal Number of Clusters**: **2**
    - The optimal cluster count was determined using the Davies-Bouldin (DB) Index, which evaluates cluster compactness and separation.

2. Clustering Metrics

- **Davies-Bouldin Index**:
    - Optimal clustering DB Index: **0.73**
    - Lower DB Index values indicate better clustering quality (compact and well-separated clusters).
- **Silhouette Score**:
    - Silhouette Score for optimal clustering: **0.49**
    - This moderate score indicates reasonably well-defined clusters.

Cluster Analysis

Distribution of Customers Across Clusters

- **Cluster Sizes**:
    - Cluster 0: **89 customers**
    - Cluster 1: **111 customers**
- These clusters represent distinct customer behaviours, detailed below.

Average Spending and Transaction Frequency

- **Cluster 0:** Customers with high spending and moderate transaction frequency, indicating premium shoppers.
- **Cluster 1**: Customers with lower spending and less frequent transactions, likely representing budget-conscious or casual shoppers.

Cluster	Average Spending ($)	Average Transaction Frequency
Cluster 0	5033.04	6.89
Cluster 1	2180.68	3.49

Behavioural Insights

- **Cluster 0**: High-value customers who may prioritise quality or premium services. Engaging these customers with loyalty programs or exclusive offerings can help sustain their high-value behaviour.
- **Cluster 1**: Price-sensitive customers who may respond to promotions or discounts to encourage more frequent purchases.

Visualizations

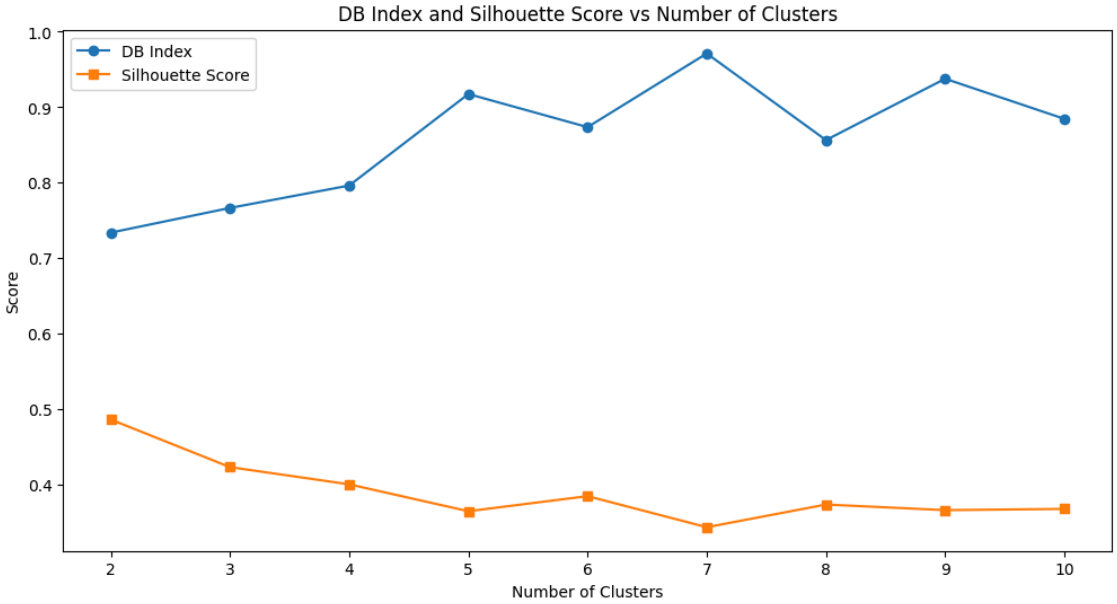
1. Cluster Distribution

A PCA scatter plot was used to visualize clusters based on reduced dimensions:

- Cluster 0: Fewer customers but with significantly higher spending and transaction frequency.
- Cluster 1: Larger group with relatively lower spending and frequency.

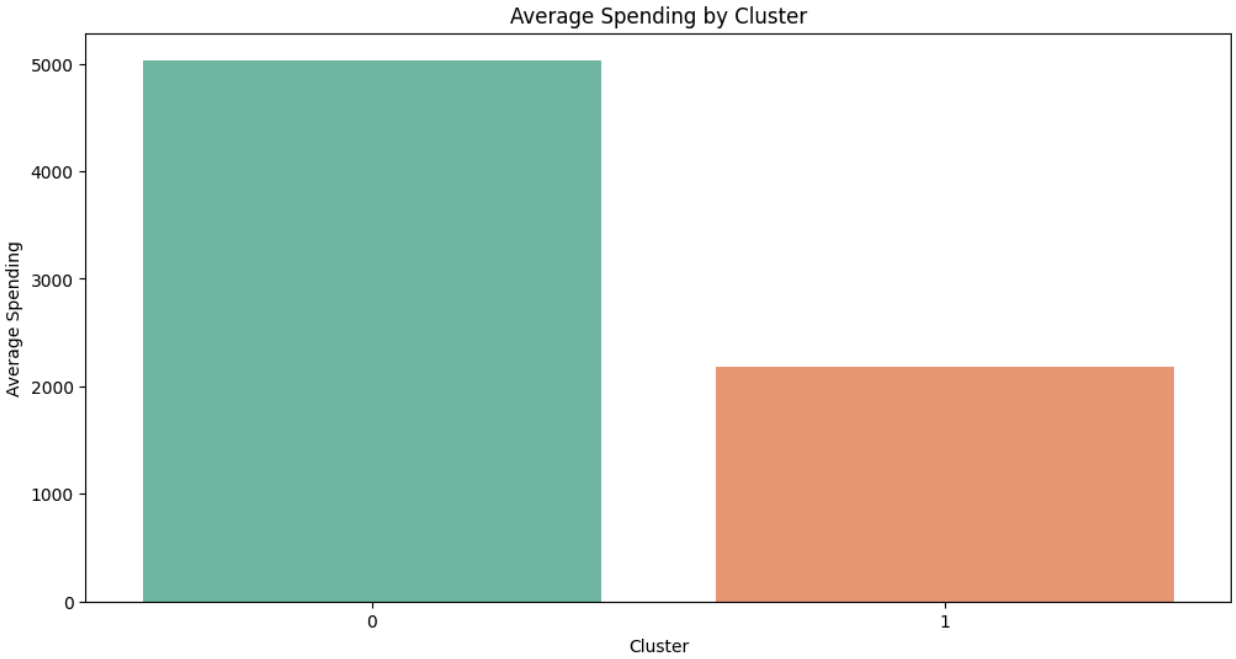
2. DB Index and Silhouette Score vs. Number of Clusters

- The DB Index ranged from **0.73 to 1.0**, with **2 clusters** achieving the lowest value of **0.73**.
- Silhouette scores peaked at **2 clusters** with a score of **0.49**, validating the choice of 2 as the optimal number of clusters.

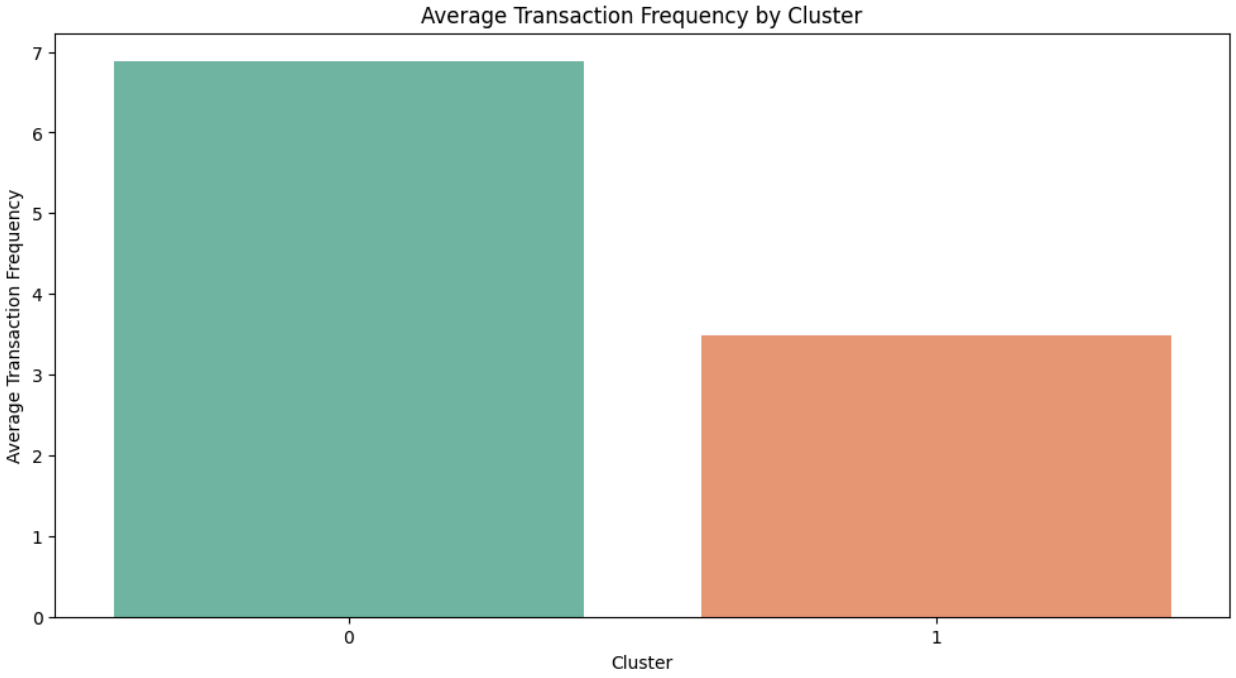


3. Cluster-Wise Metrics

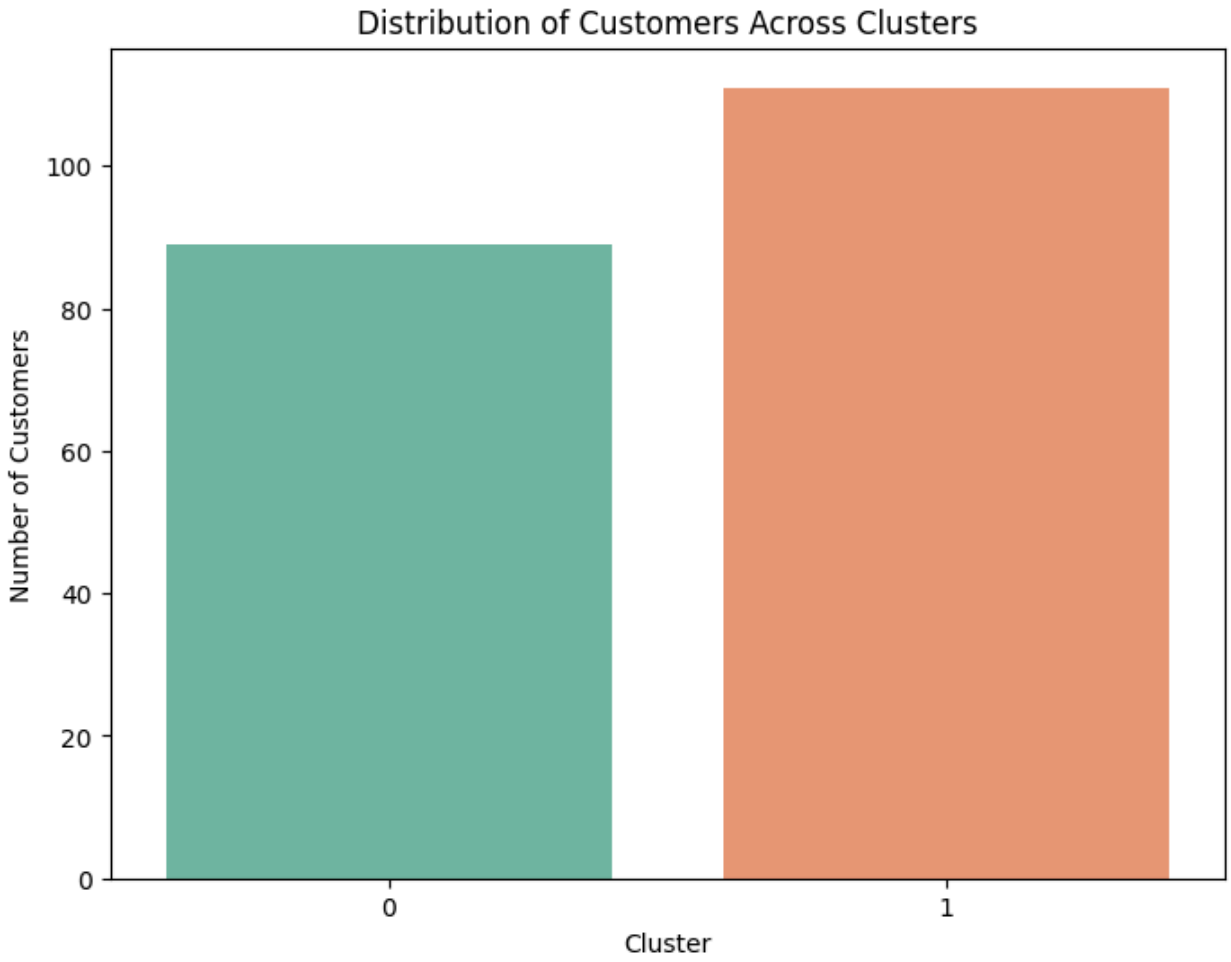
- **Average Spending by Cluster**: Visualized using bar charts, showing the significant spending disparity between clusters.



- **Transaction Frequency by Cluster**: Bar charts highlight the difference in transaction frequency between clusters.



- **Customer Count by Cluster**: A count plot reveals the distribution of customers across clusters.



Observations

1. **Distinct Clusters**:
    - Cluster 0 comprises high-value customers with moderate frequency, representing premium shoppers.
    - Cluster 1 includes budget-conscious or infrequent shoppers.
2. **Segmentation Validation**:
    - The DB Index of **0.73** indicates compact and well-separated clusters.
    - The moderate silhouette score (**0.49**) reflects reasonably defined clusters.
3. **Business Opportunities**:
    - Cluster segmentation highlights opportunities for personalized marketing strategies.

Recommendations

1. Targeted Marketing

- **Cluster 0**:
    - Develop exclusive loyalty programs, premium service offerings, and high-value rewards.
    - Utilize direct communication to build long-term relationships.
- **Cluster 1**:
    - Offer promotions, discounts, or referral incentives to encourage higher spending and frequency.
    - Focus on accessible, value-based messaging.

2. Data Enrichment

- Incorporate additional customer data (e.g., preferred product categories, geographical regions) for deeper profiling.
- Analyze seasonal trends to align marketing campaigns with customer behaviour.

3. Advanced Clustering Techniques

- Explore alternative clustering algorithms such as DBSCAN or hierarchical clustering to validate and refine segmentation.
- Experiment with more features (e.g., product preferences, demographics) to uncover nuanced patterns.

Conclusion

The clustering analysis successfully segmented customers into two meaningful groups based on spending and transaction frequency. These insights, supported by the DB Index and Silhouette Score, can inform strategic marketing and customer retention initiatives. Future analysis with enriched data and advanced clustering methods can further enhance segmentation quality.
