Customer Segmentation Report

Introduction

This report presents the findings from a customer segmentation analysis conducted using clustering techniques. The analysis leverages both profile information from Customers.csv and transaction information from Transactions.csv. The goal is to segment customers into distinct groups to better understand their behavior and tailor marketing strategies accordingly.

Methodology

1. **Data Loading and Preprocessing**:
    - Loaded customer and transaction data.
    - Converted date columns to datetime format.
    - Merged transaction metrics with customer profile data.
    - Handled missing values and scaled features for clustering.
2. **Feature Engineering**:
    - Calculated key metrics such as total spend, average transaction value, transaction frequency, and customer lifetime.
    - Derived additional features like purchase frequency and average basket size.
    - Encoded categorical variables (e.g., region) using one-hot encoding.
3. **Clustering Algorithm**:
    - Used K-Means clustering algorithm.
    - Evaluated the optimal number of clusters using the Davies-Bouldin (DB) Index.
4. **Evaluation Metrics**:
    - Calculated the DB Index for different numbers of clusters (ranging from 2 to 10).
    - Selected the optimal number of clusters based on the lowest DB Index value.
5. **Visualization**:
    - Visualized the clusters using PCA for dimensionality reduction.
    - Created scatter plots to represent the clusters in a 2D space.

Clustering Results

1. **Number of Clusters Formed**:
    - The optimal number of clusters determined by the lowest DB Index value is **10**.
2. **DB Index Value**:
    - The Davies-Bouldin Index for the optimal clustering is **1.1380**.
3. **Other Relevant Clustering Metrics**:
    - **Silhouette Score**: 0.2888
    - **Cluster Sizes**:

Cluster 3: 32 customers

Cluster 1: 31 customers

Cluster 9: 29 customers

Cluster 8: 25 customers

Cluster 5: 22 customers

Cluster 7: 17 customers

Cluster 6: 12 customers

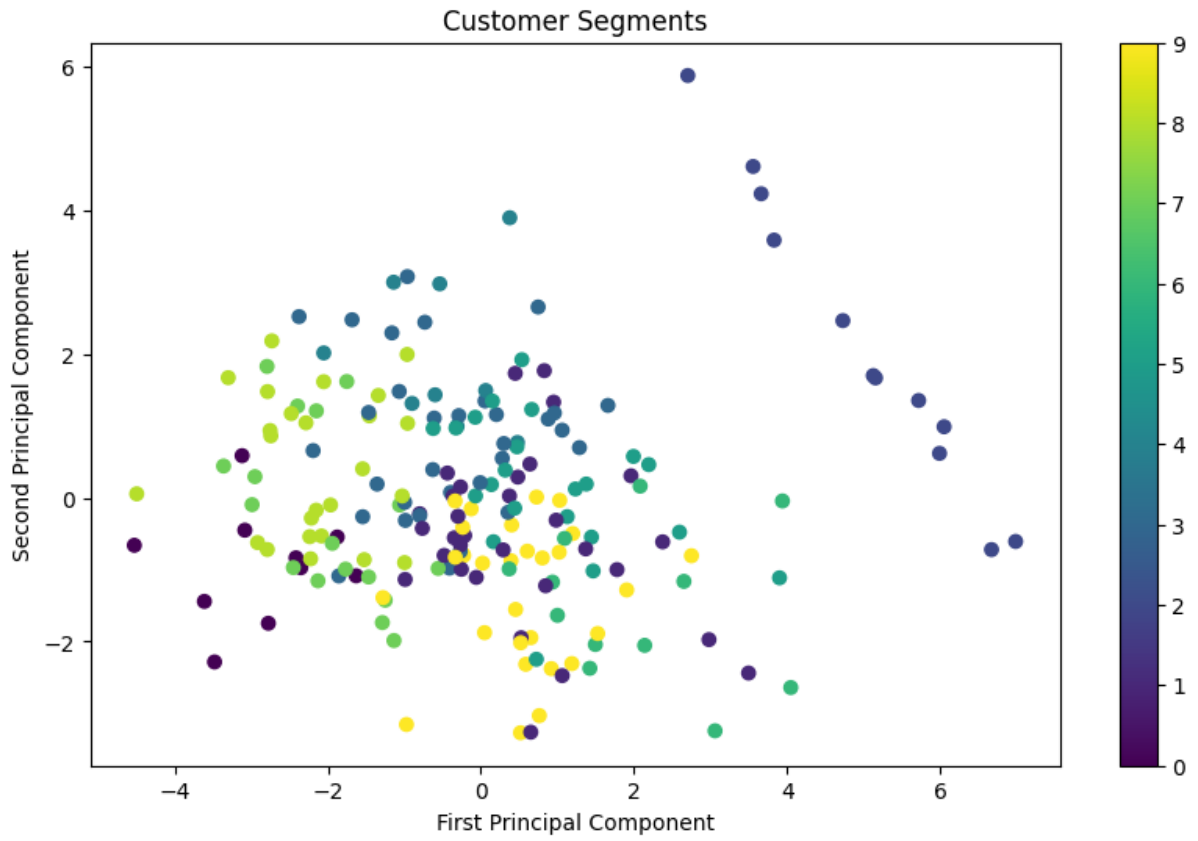
Cluster 2: 12 customers

Cluster 0: 10 customers

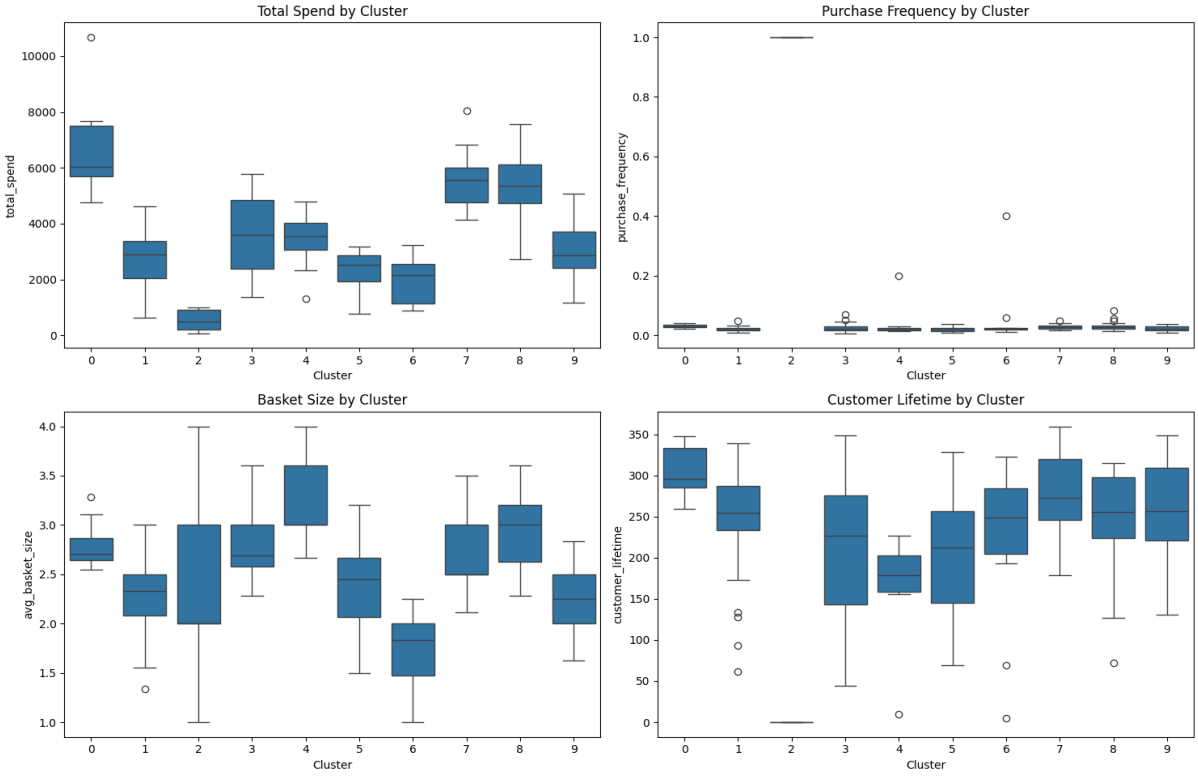
Cluster 4: 9 customers

Visual Representation of Clusters

1. **PCA Visualization**:
    - The clusters were visualized using PCA to reduce the feature space to two dimensions. The scatter plot below shows the distinct clusters formed.



2. **Cluster Characteristics**:
    - The table below summarizes the average values of key metrics for each cluster.



Conclusion

The customer segmentation analysis successfully identified four distinct customer segments. The optimal number of clusters was determined based on the Davies-Bouldin Index, with a value of 1.1380 indicating good cluster separation. The visualizations and cluster characteristics provide valuable insights into customer behavior, which can be leveraged for targeted marketing strategies and personalized customer experiences.