Customer Segmentation Report

Introduction

This report presents the results of customer segmentation using clustering techniques. The analysis aims to group customers based on their profile and transaction information to identify distinct segments for targeted marketing strategies.

Methodology

Data Preparation

- **Data Sources**: The analysis uses data from **Customers.csv**, **Transactions.csv**, and **Products.csv**.
- **Feature Engineering**: Key features were created, including:
    - Average Transaction Value
    - Total Spend
    - Purchase Count
    - Region

Clustering Approach

- **Algorithm**: K-Means clustering was chosen for its simplicity and effectiveness in partitioning data into distinct groups.
- **Preprocessing**: Features were standardized to ensure comparability across different scales.

Determining Optimal Clusters

- **Elbow Method**: Used to identify the point where adding more clusters yields diminishing returns. The elbow was observed around 4 or 5 clusters.
- **Davies-Bouldin Index**: Calculated for each K value to evaluate clustering quality. A lower DB Index indicates better-defined clusters.

Results

Number of Clusters

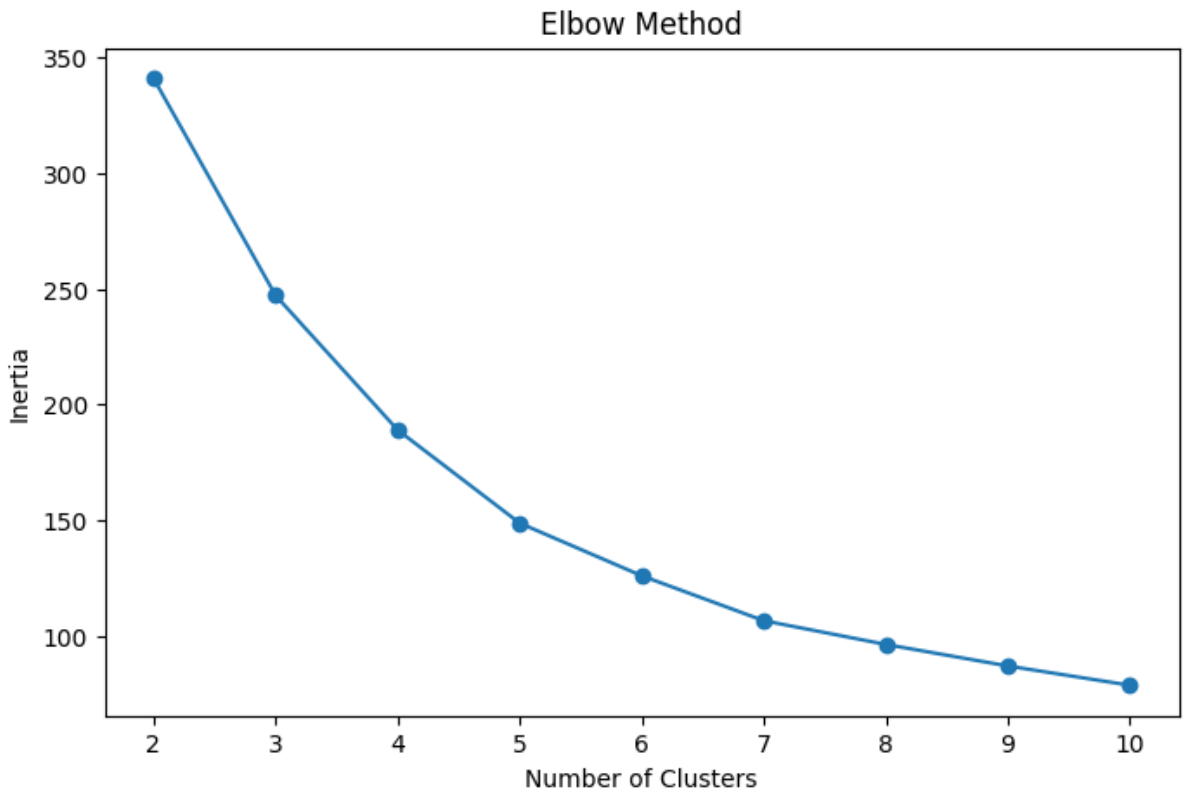
- The optimal number of clusters was determined to be 4, based on the Elbow Method and DB Index.

Clustering Metrics

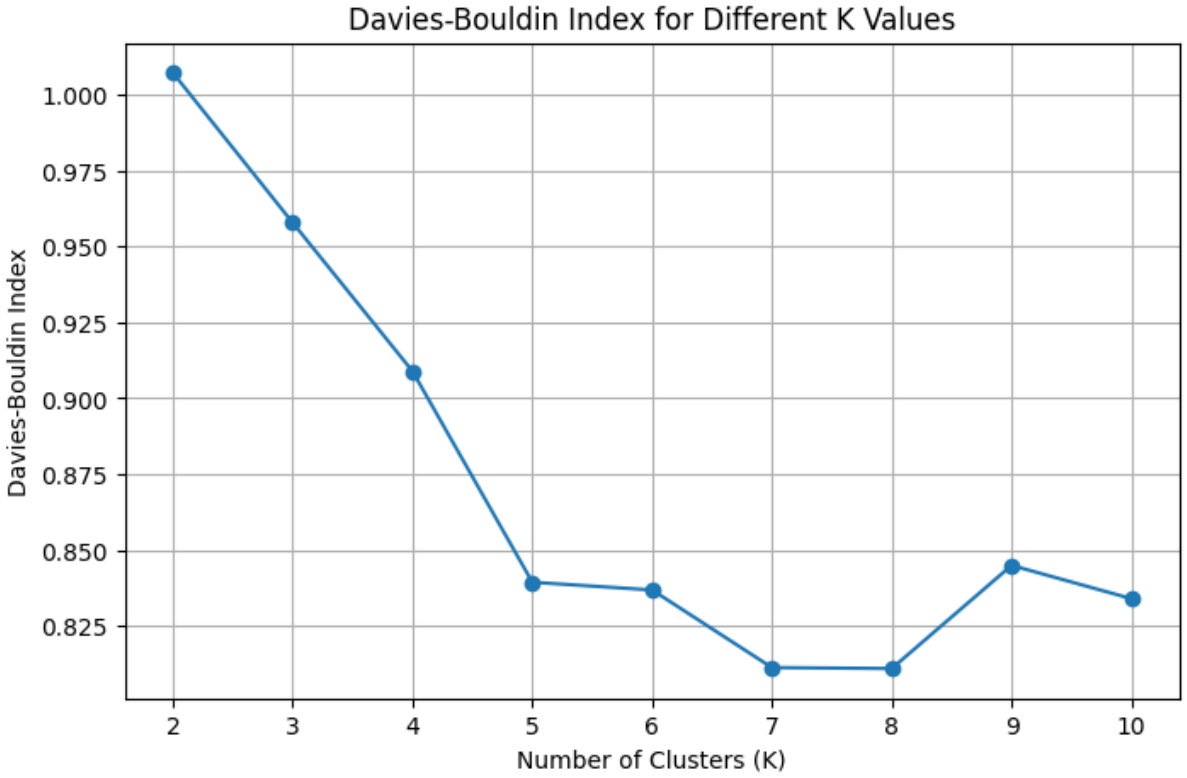
- **Davies-Bouldin Index**: 0.909

Visualizations

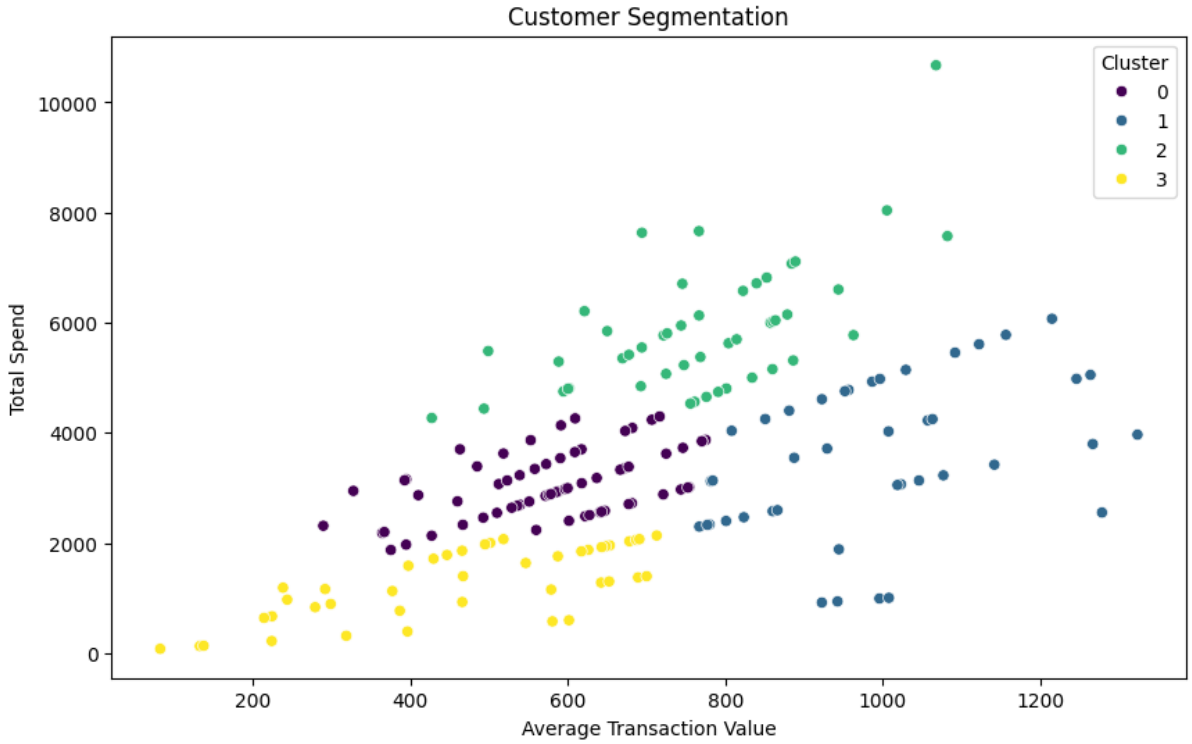
- **Elbow Plot**: The plot below shows the inertia for different K values, highlighting the optimal clustering.



- **DB Index Plot**: The plot below shows the DB Index values for different K values, highlighting the optimal clustering.



- **Cluster Plot**: The scatter plot below visualizes the customer segments based on Average Transaction Value and Total Spend.



Conclusion

The clustering analysis segmented customers into 4 distinct groups. These segments can be leveraged for targeted marketing and personalized customer engagement strategies. The use of both the Elbow Method and Davies-Bouldin Index ensured the selection of an optimal number of clusters, providing a robust framework for customer segmentation.