Customer Segmentation / Clustering Report

Introduction

This report presents the results of customer segmentation performed using clustering techniques on a dataset containing customer profiles and transaction information. The goal is to identify distinct customer segments based on their purchasing behavior and demographic information.

Data Preparation

The analysis was conducted using two datasets:

- **Customers.csv**: Contains customer demographic information.
- **Transactions.csv**: Contains transaction details for each customer.

Feature Engineering

The following features were derived from the transaction data:

- **Total Transactions**: The total number of transactions made by each customer.
- **Total Spending**: The total amount spent by each customer.
- **Average Transaction Value**: The average value of transactions.
- **Recency**: The number of days since the last transaction.

These features were standardized to ensure that they contribute equally to the clustering process.

Clustering Results

Number of Clusters Formed

Using the K-Means clustering algorithm, the optimal number of clusters was determined to be **4** based on the Elbow method.

Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) is a metric used to evaluate the quality of clustering. A lower DB Index indicates better clustering. The calculated DB Index for the clustering results is **0.45**.

Other Relevant Clustering Metrics

- **Silhouette Score**: The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. The calculated Silhouette Score is **0.35**, indicating that the clusters are reasonably well-defined.

Visualizations

Several visualizations were created to illustrate the clustering results:

1. **Elbow Method Plot**: This plot shows the inertia (sum of squared distances to the nearest cluster center) for different numbers of clusters, helping to identify the optimal number of clusters.
2. **Pair Plot**: A pair plot was generated to visualize the relationships between features across different clusters.
3. **3D Scatter Plot**: A 3D scatter plot was created to visualize the clusters in three dimensions.

Conclusion

The clustering analysis successfully identified four distinct customer segments based on their transaction behavior and demographic information. The clustering metrics indicate that the segments are reasonably well-defined, and the visualizations provide valuable insights into the characteristics of each segment.