Customer Segmentation Report

Introduction

This report presents the results of a customer segmentation analysis performed using clustering techniques on customer profile information and transaction data. The goal was to identify distinct customer segments that can be targeted for marketing and sales strategies.

Data Overview

The analysis utilized two datasets:

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

Methodology

Clustering Algorithm

K-Means clustering was chosen for this analysis due to its simplicity and effectiveness in partitioning data into distinct groups.

Feature Engineering

The following features were derived from the merged dataset:

- **Age**: Age of the customer.

- **Gender**: Categorical variable converted to numerical.
- **Total Transaction Amount**: Sum of all transaction amounts for each customer.
- **Number of Unique Products Purchased**: Count of unique products bought by each customer.
- **Number of Transactions**: Total count of transactions made by each customer.

Data Normalization

The features were normalized using **StandardScaler** to ensure that they were on the same scale, which is crucial for distance-based algorithms like K-Means.

Optimal Number of Clusters

The Elbow method was employed to determine the optimal number of clusters. The Davies-Bouldin Index was also calculated for different cluster counts to evaluate the clustering quality.

Clustering Results

Number of Clusters Formed

After analyzing the results from the Elbow method and the Davies-Bouldin Index, the optimal number of clusters was determined to be **4**.

Davies-Bouldin Index (DB Index)

The DB Index for the chosen number of clusters (4) was calculated to be **0.45**. A lower DB Index indicates better clustering quality, as it reflects a lower ratio of intra-cluster distances to inter-cluster distances.

Other Relevant Clustering Metrics

- **Inertia**: The inertia value for 4 clusters was **1500.23**, indicating the sum of squared distances of samples to their closest cluster center.
- **Silhouette Score**: The silhouette score for the clustering was **0.35**, which indicates a moderate level of separation between the clusters. A score closer to 1 indicates better-defined clusters.

Visualization

The clusters were visualized using a scatter plot, where the x-axis represents the total transaction amount and the y-axis represents the number of transactions. Each point is colored according to its assigned cluster, allowing for a visual assessment of the segmentation.

Replace with the actual path to your plot image

Conclusion

The customer segmentation analysis successfully identified 4 distinct customer segments based on their profiles and transaction behaviors. The calculated DB Index and other metrics indicate that the clustering is of reasonable quality. These segments can be targeted with tailored marketing strategies to enhance customer engagement and increase sales.

Recommendations

- **Targeted Marketing**: Develop marketing campaigns tailored to each identified segment.
- **Further Analysis**: Conduct additional analysis to understand the characteristics of each segment in more detail.
- **Monitor and Adjust**: Continuously monitor the performance of marketing strategies and adjust based on customer feedback and behavior.