Customer Segmentation Report

Overview

Customer segmentation is the process of dividing a customer base into distinct groups of individuals who share similar characteristics. This helps businesses tailor their marketing strategies, improve customer satisfaction, and optimize resource allocation.

Data Preparation

We used three datasets:

- Customers.csv: Contains profile information of customers, including ID, name, region, and signup date.
- Transactions.csv: Contains transaction details such as transaction ID, customer ID, product ID, and transaction date.
- Products.csv: Contains product information including product ID, name, category, and price.

These datasets were merged to create a comprehensive customer profile with transaction history and product information.

Feature Engineering

We performed the following steps:

1. Encoded categorical features: Converted categorical variables such as Region and Category into numerical values using one-hot encoding.
2. Standardized numerical features: Standardized features like Age, Income, Total Amount Spent, and Total Number of Transactions to ensure they contribute equally to the clustering process.

Customer Segmentation Report

Clustering Algorithm

We used the K-Means clustering algorithm to segment the customers. We evaluated different numbers of clusters (between 2 and 10) and selected the optimal number based on the Davies-Bouldin Index (DB Index).

Evaluation Metrics

- Davies-Bouldin Index (DB Index): Measures the average similarity ratio of each cluster with the cluster most similar to it. Lower values indicate better-defined clusters.

Results

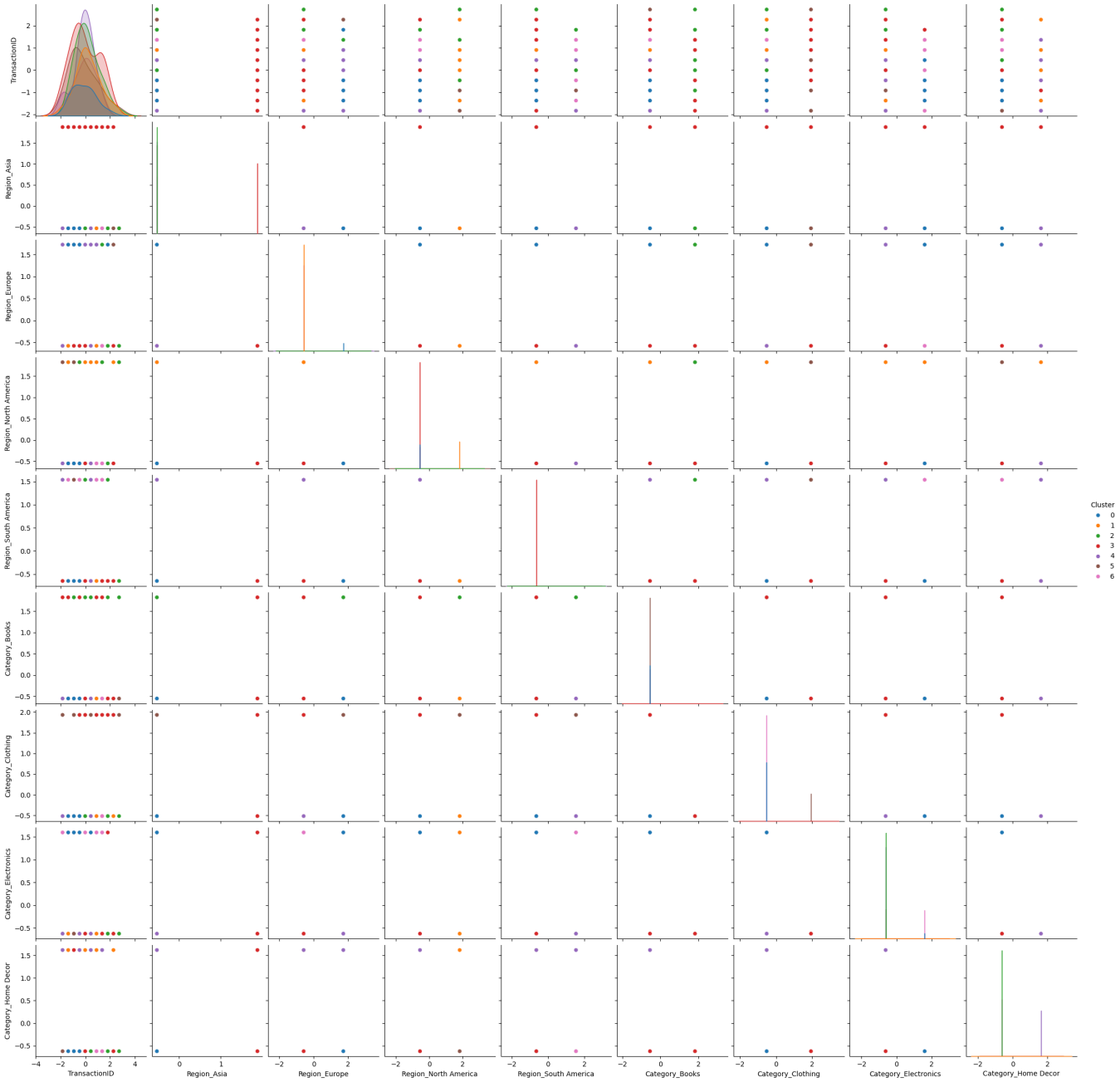
1. Number of Clusters:
   - The optimal number of clusters was determined to be 7, as it produced the lowest DB Index.

2. Davies-Bouldin Index:
   - The DB Index for the optimal clustering solution was 1.0995971990467308.

3. Other Metrics:
   - Inertia: Sum of squared distances of samples to their closest cluster center.
   - Silhouette Score: Measures how similar an object is to its own cluster compared to other clusters. (Optional for inclusion in the report)

Customer Segmentation Report

Pair Plot



PCA Plot

Customer Segmentation Report