Customer Segmentation Report

This report details the customer segmentation performed using clustering techniques on an eCommerce transactions dataset. The objective is to group customers based on their transactional and profile information to derive actionable insights for targeted marketing and better customer management.

Clustering Algorithm: K-Means

The K-Means algorithm was chosen for its simplicity and effectiveness in grouping numerical data into distinct clusters. The number of clusters was determined through experimentation and evaluation using the Davies-Bouldin (DB) Index, aiming for well-separated and compact clusters.

Clustering Methodology

Features Used:

- TotalSpend: Total spending by each customer.
- TransactionCount: Number of transactions made.
- AverageQuantity: Average quantity purchased per transaction.
- DaysBetweenPurchases: Recency of transactions.
- DaysSinceSignup: Time since the customer signed up.

Steps:

1. Aggregated transactional data at the customer level.
2. Normalized numerical features to ensure equal importance during clustering.
3. One-hot encoded the 'Region' column to incorporate geographic information.

Customer Segmentation Report

4. Applied K-Means clustering with 4 clusters, chosen based on iterative evaluation using the DB Index.

Clustering Results

Number of Clusters: 4

Davies-Bouldin Index (DB Index): 1.331

The DB Index quantifies cluster compactness and separation, with lower values indicating better clustering. In this analysis, a DB Index of 1.331 reflects well-separated and compact clusters.

Cluster Characteristics:

Cluster 0: Average spenders and moderately active, higher representation from Europe and South America.

Cluster 1: Lowest spenders and least active, primarily from Europe.

Cluster 2: Moderately active customers, highest representation from North America.

Cluster 3: High spenders and frequent buyers, diverse geographic representation.

Cluster Visualization

The scatter plot below visualizes the customer clusters based on normalized total spending (x-axis) and transaction count (y-axis). Each point represents a customer, colored according to its assigned cluster. Key observations:

- Cluster 3 contains high spenders and frequent buyers, distinct from other clusters.
- Cluster 1 includes the least active and lowest spending customers.

Customer Segmentation Report

- Clusters 0 and 2 represent moderate spending and activity levels, with varying geographic representation.