Customer Segmentation Report

Overview

Clustering analysis was conducted on customer transaction data to identify distinct customer segments based on total spending and transaction behavior. Various clustering techniques, including K-Means, Hierarchical Clustering, and DBSCAN, were applied to the dataset.

Clustering Methodology

- Clustering was performed both with and without outliers.
- Keeping outliers resulted in improved clustering quality.
- Some high-spending customers were initially misclassified as outliers but were retained due to their business significance.
- The t-SNE visualization confirmed the presence of 4 distinct groups.

Clustering Results

- **Number of Clusters:** 4
- **DB Index Value:** 0.93
- **Silhouette Score:** 0.45

Cluster Descriptions

- **Cluster 0:** Moderate spenders with balanced transaction behavior.
- **Cluster 1:** High-transaction customers, predominantly from North America.
- **Cluster 2:** European customers with moderate spending patterns.
- **Cluster 3:** Highest-spending customers, primarily from South America.

Business Insights

- Spending and transaction frequency vary significantly across regions.
- Clustering enables targeted marketing strategies based on customer segmentation.
- Retaining outliers improved clustering quality, as evidenced by the DB Index and Silhouette Score.
- The 4-group segmentation provides meaningful insights that can be leveraged for business decision-making.

Conclusion

The analysis confirms that a **4-group segmentation** is the most effective approach. This segmentation can be utilized for personalized marketing, customer engagement, and strategic decision-making to enhance business growth.