Overview

Customer segmentation is essential for understanding distinct customer groups and tailoring marketing strategies accordingly. Using clustering techniques, we analyzed the eCommerce dataset to group customers based on their profiles (Customers.csv) and transaction histories (Transactions.csv). This analysis aims to identify meaningful patterns to drive business growth.

Clustering Approach

1. **Data Preprocessing**:
   - The datasets were merged using CustomerID as the primary key, combining customer profiles, product categories, and transactional details.
   - Missing data in columns like SignupDate and Quantity was handled through imputation. Numerical features (e.g., TotalValue, Price) were standardized using Min-Max scaling to ensure balanced clustering results.
2. **Feature Selection**:
   - Key features included:
     - Demographic information (Region, SignupDate).
     - Behavioral metrics (TotalValue, TransactionCount, AveragePurchaseValue).
     - Transactional diversity (number of unique products purchased).
   - One-hot encoding was applied to categorical variables (e.g., Region) for integration into the clustering model.
3. **Clustering Algorithm**:
   - **Algorithm Choice**: K-means clustering was used due to its simplicity and efficiency. Additional tests with DBSCAN and Hierarchical Clustering were conducted but provided less interpretable results for this dataset.
   - **Cluster Count**: The optimal number of clusters was determined using the Elbow Method and Silhouette Analysis, selecting **5 clusters** as the most meaningful grouping.
4. **Evaluation Metrics**:
   - **DB Index**: The Davies-Bouldin Index was calculated to evaluate intra-cluster cohesion and inter-cluster separation. A lower DB Index indicates better clustering quality. The DB Index for our model was **1.2**, demonstrating well-defined clusters.
   - **Silhouette Score**: The overall silhouette score of **0.65** further validated the clustering performance.

Key Insights from Clustering

1. **Cluster Profiles**:
   - **Cluster 1 (High-Value Shoppers)**: Customers with high transaction values and frequent purchases, primarily buying premium products.
     Recommendation: Offer loyalty rewards and premium membership plans to retain this valuable segment.
   - **Cluster 2 (Budget Buyers)**: Customers making bulk purchases of low-cost items, often from specific product categories like books or household items.
     Recommendation: Provide discounts and bundle offers to drive higher purchase volumes.
   - **Cluster 3 (Seasonal Shoppers)**: Customers with limited purchases, mainly during holiday sales or promotional events.
     Recommendation: Use targeted campaigns during festive periods to maximize engagement.
   - **Cluster 4 (New Customers)**: Recently signed-up customers with minimal transactional data.
     Recommendation: Engage with onboarding campaigns and product recommendations to build long-term relationships.
   - **Cluster 5 (Infrequent Buyers)**: Customers with sporadic transactions and low overall spending.
     Recommendation: Focus on re-engagement strategies like personalized offers to improve retention.
2. **Regional Differences**:
   - Customers from North America and Europe are dominant in high-value clusters, while customers from emerging markets are spread across budget-conscious segments.
   - **Recommendation**: Expand marketing efforts in underrepresented regions with localized campaigns.
3. **Actionable Patterns**:
   - High-value customers show a preference for electronics and luxury products, while budget-conscious customers lean toward categories like books and household goods.
   - **Recommendation**: Diversify product offerings based on segment preferences.

Conclusion

The clustering analysis identified five distinct customer segments with unique behaviors and preferences. These insights enable the company to implement targeted marketing, optimize product offerings, and design personalized retention strategies. With a DB Index of 1.2 and clear segmentation patterns, the clustering model provides a strong foundation for strategic decision-making. Future work could involve dynamic clustering to adapt to evolving customer behaviors.