Introduction

This analysis seeks to conduct customer segmentation based on the use of clustering techniques. It takes into consideration the profile of customers and information contained in the Customers.csv, and it includes details from Transactions.csv. This procedure identifies various clusters of customers in an effort to create actionable insight.

Methodology

Data Aggregation

1. **Sources**:
   - Customers.csv: Contains customer profile data, including region and signup date.
   - Transactions.csv: Includes transaction-level data such as quantity and total value.
   - Products.csv: Provides product details, including category and price.
2. **Features Extracted**:
   - **Customer-Level Metrics**:
     - Total Transactions: Number of purchases made by a customer.
     - Total Quantity: Total items purchased by a customer.
     - Total Spend: Total monetary value spent by a customer.
     - Average Transaction Value: Average amount spent per transaction.
     - Days Since Signup: Days since the customer joined (as of Jan 1, 2025).
   - **Categorical Features**:
     - Region: Customer region (one-hot encoded).

Data Preprocessing

- **Numerical Features**: Standardized using **StandardScaler** to ensure uniform scaling.
- **Categorical Features**: Encoded using OneHotEncoder to convert regions into numerical form.

Clustering Technique

- **Algorithm**: KMeans clustering.
- **Number of Clusters**: Tested cluster numbers from 2 to 10.
- **Evaluation Metric**: Davies-Bouldin (DB) Index.

Result

Insights

1. Cluster 0: {Describe Key Traits, e.g., "High spenders with frequent transactions"}.
2. Cluster 1: {Describe Key Traits, e.g., "New customers with low spending habits"}.

Visualization

- The scatter plot below represents the clusters after dimensionality reduction using PCA.
- Each color represents a distinct cluster.

Conclusion

The analysis successfully segmented customers into {2} clusters. These clusters reveal distinct customer behaviors, such as spending patterns, transaction frequency, and longevity.

Recommendations

1. **Marketing Strategies**: Tailor campaigns for high spenders (e.g., loyalty rewards) and low spenders (e.g., discounts).
2. **Customer Retention**: Focus on retaining customers in {Insert Cluster} who show high churn risk (e.g., low frequency, recent signup).
3. **Upselling Opportunities**: Target {Insert Cluster} for upselling based on their high average transaction value.