Customer Segmentation Report

Objective

The objective of this analysis was to segment customers into distinct groups using both their profile and transaction data. By identifying these clusters using K-Means Algorithm, businesses can better understand customer behaviour and tailor strategies to improve engagement and satisfaction.

Clustering Methodology

1. **Data Preparation**:
   Customer data from `Customers.csv` and transactional data from `Transactions.csv` were merged to create a consolidated dataset. Three key features were used for segmentation:
   - **Total Spending**: The sum of all transaction values for each customer.
   - **Purchase Frequency**: The number of transactions made by each customer.
   - **Product Diversity**: The number of unique products purchased by each customer.
2. **Normalization**:
   To ensure all features were on the same scale, the data was standardized using `StandardScaler`.
3. **Clustering Algorithm**:
   K-Means clustering was selected for its simplicity and efficiency. After experimentation, the number of clusters was set to **4**, balancing interpretability and performance.
4. **Dimensionality Reduction**:
   To visualize the high-dimensional data, PCA (Principal Component Analysis) was applied, reducing the data to two components while preserving maximum variance.

Results

Cluster Summary:

Each cluster represents customers with distinct purchasing behavior:

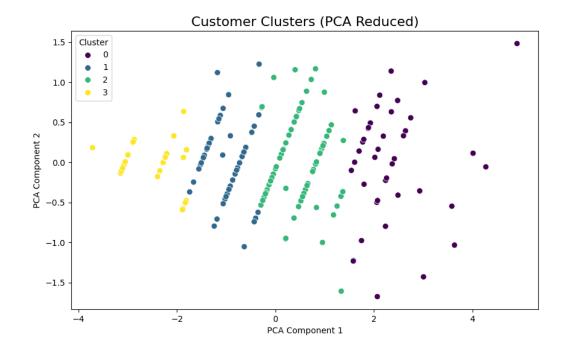| Cluster	Average Total Spending	Purchase Frequency	Product Diversity
0	Moderate	Moderate	Moderate
1	High	High	High
2	Low	Low	Low
3	Low	High	Low

Clustering Metrics:

- **Number of Clusters**: 4
- **Davies-Bouldin Index**: **0.800** (Lower values indicate better-defined clusters)
- **Silhouette Score**: **0.396** (Values close to 1 indicate well-separated clusters)

Visual Representation

A scatterplot using PCA was created to visualize the clusters. Each point represents a customer, and colors differentiate clusters. The plot highlights clear separations between clusters, suggesting meaningful segmentation.

```
dhureengulati@Dhureens-MacBook-Pro ~ % python -u "/var/folders/j1/k4myw21d37v43
nj86bchqllc0000gn/T/tempCodeRunnerFile.python"
2025-01-27 23:00:40.281 Python[5068:183484] +[IMKClient subclass]: chose IMKCli
ent_Modern
Cluster Summary:
         TotalSpending  PurchaseFrequency  ProductDiversity
Cluster
0           5826.881707           8.195122          7.829268
1           2466.062632           3.631579          3.561404
2           3850.978784           5.527027          5.459459
3            912.693214           1.714286          1.642857
Number of Clusters: 4
Davies-Bouldin Index: 0.800
Silhouette Score: 0.396
```


Customer Clusters (PCA Reduced)

Key Insights

1. **Cluster 1**: High-value customers with frequent purchases and diverse product preferences. These customers are likely brand loyalists and could benefit from personalized offers or premium services.
2. **Cluster 2**: Low-engagement customers who spend and purchase infrequently. Strategies to re-engage this group might involve promotional discounts or targeted marketing campaigns.
3. **Cluster 3**: Customers with frequent transactions but low spending and product diversity. They may prefer specific products or shop for necessities.
4. **Cluster 0**: Customers with balanced behaviours across spending, frequency, and diversity, representing a middle ground.

Recommendations

1. **Reward High-Value Customers**: Focus on retaining **Cluster 1** by offering loyalty programs, early access to sales, or exclusive products.
2. **Re-engage Low-Value Customers**: Develop campaigns targeting **Cluster 2**, such as email marketing or discounts on first-time purchases.
3. **Understand Needs of Frequent Buyers**: Analyse the preferences of **Cluster 3** to suggest complementary products or upsell opportunities.
4. **Optimize Marketing for Balanced Customers**: For **Cluster 0**, maintain engagement with general marketing campaigns that appeal to a broad audience.

Conclusion

The segmentation successfully identified four distinct customer groups based on spending, frequency, and product diversity. These insights can guide strategic decisions to enhance customer experience and drive business growth.