Clustering Report

Overview

This report presents the results of customer segmentation performed using clustering techniques. The analysis utilized profile data and transactional information to group customers into clusters based on similarities in purchasing behavior and preferences.

Methodology

1. **Data Preparation**:
    - Data was aggregated to create customer profiles with features such as total transaction value, total quantity purchased, and average price per transaction.
    - Data was normalized to ensure uniform scaling.
2. **Dimensionality Reduction**:
    - Principal Component Analysis (PCA) was applied to reduce the data to two dimensions for visualization while retaining maximum variance.
3. **Clustering Algorithm**:
    - **KMeans Clustering** was chosen for its simplicity and efficiency.
    - The number of clusters was determined to be 5 based on experimentation and evaluation metrics.
4. **Evaluation Metrics**:
    - Davies-Bouldin Index (DBI) was used to assess the quality of clustering. Lower DBI values indicate better-defined clusters.

Results

Cluster Metrics

- **Number of Clusters**: 5
- **Davies-Bouldin Index**: 0.5375

Cluster Characteristics

1. **Cluster 0**:

    - High total transaction value and high average price.
    - Represents premium customers.
2. **Cluster 1**:
    - Medium transaction value and quantity.
    - Customers likely to purchase consistently but not high-value items.
3. **Cluster 2**:
    - Low transaction value and quantity.
    - Customers with infrequent or low-value purchases.
4. **Cluster 3**:
    - High quantity but moderate transaction value.
    - Likely bulk buyers of lower-priced items.
5. **Cluster 4**:
    - Moderate transaction value and high average price.
    - Customers interested in quality over quantity.

Visualization

Cluster Distribution

- A scatter plot with PCA components as axes was used to visualize the clustering results.
- Each cluster is represented by a distinct color.

Key Observations:

- Clusters are well-separated, indicating distinct customer groups.
- Cluster 0 and Cluster 4 show high spending patterns, suggesting a focus group for premium offerings.

Visualization: (Include scatter plot here with labeled clusters.)

Cluster Sizes

- A bar chart showing the number of customers in each cluster.
- Cluster 2 (low-value customers) is the largest group, highlighting an opportunity for engagement strategies to increase their spending.

Business Interpretations

1. **Premium Customer Engagement**:
    - Clusters 0 and 4 represent high-value customers. Consider exclusive rewards or loyalty programs to retain these customers.
2. **Targeted Promotions for Low-Value Customers**:
    - Cluster 2 includes customers with low spending. Implement strategies such as personalized discounts or product recommendations to increase engagement.
3. **Bulk Buyer Optimization**:
    - Cluster 3 represents bulk buyers. Offer tiered discounts or bundle deals to maximize their transactions.
4. **Product Strategy**:
    - Cluster characteristics suggest diverse customer needs. Tailor product offerings to meet the preferences of each cluster, especially for premium and bulk-buying customers.
5. **Marketing Campaigns**:
    - Use clustering results to design segmented campaigns. Focus on the unique attributes of each cluster for effective targeting.

Conclusion

Customer segmentation provided actionable insights into the eCommerce customer base. By understanding the characteristics of each cluster, the business can implement targeted strategies to optimize revenue, improve customer satisfaction, and enhance overall engagement.