Clustering Report

Introduction

Customer segmentation is essential for businesses to categorize their customers based on purchasing behavior and demographics. This report presents the clustering analysis of the eCommerce Transactions dataset, using transaction details such as total purchase value and quantity purchased. The clustering technique helps businesses identify distinct customer segments to develop targeted marketing and retention strategies.

Data Preparation

The clustering analysis is performed using the Transactions dataset. The following preprocessing steps were applied:

1. **Data Cleaning**: No missing values were found in the dataset, ensuring complete data availability.
2. **Feature Selection**: TotalValue (total spending) and Quantity (number of items purchased) were chosen as key clustering features.
3. **Feature Scaling**: Standardized the features using StandardScaler to ensure proper clustering.

Clustering Methodology

- **Algorithm Used**: K-Means clustering was selected for segmentation due to its efficiency in handling large datasets.
- **Number of Clusters**: Three clusters were formed as the optimal number based on exploratory trials.
- **Evaluation Metric**: The Davies-Bouldin Index (DBI) was used to evaluate clustering performance, with a final DBI score of **0.7086**, indicating well-separated clusters.
- **Other Metrics**:
    - **Silhouette Score**: Provided additional validation of the clustering structure.
    - **Inertia (Within-Cluster Sum of Squares - WCSS)**: Used to measure the compactness of clusters.

Cluster Analysis and Insights

Cluster 0: High-Value Customers

- Customers in this group have the highest total spending and purchase frequency.
- They contribute significantly to overall revenue and should be prioritized for loyalty programs and personalized offers.

Cluster 1: Moderate-Value Customers

- This segment includes customers with average spending and purchase frequency.
- They are potential candidates for upselling and cross-selling strategies to increase revenue.

Cluster 2: Low-Value Customers

- These customers make infrequent and low-value purchases.

- Targeted promotional campaigns and incentives can encourage higher engagement from this segment.

Visualization of Clusters

- A scatter plot was used to visualize customer segmentation, showing clear distinctions between clusters.
- The high-value customers formed a distinct group with significantly higher total spending and purchase volume.
- Moderate-value and low-value customers were more spread out, indicating variability in their purchasing behaviors.

Business Recommendations

1. **High-Value Customers**: Implement loyalty programs, exclusive discounts, and premium memberships to retain and encourage further spending.
2. **Moderate-Value Customers**: Offer personalized recommendations and special promotions to convert them into high-value customers.
3. **Low-Value Customers**: Engage with targeted email marketing and discount offers to increase their purchasing frequency.
4. **Future Enhancements**: Further segmentation using demographic and behavioral data can refine customer insights and improve business strategies.

Conclusion

The clustering analysis successfully segmented customers into three distinct groups based on spending patterns and purchasing behavior. The DB Index value of **0.7086** confirms the effectiveness of clustering, and additional metrics like silhouette score and inertia provide further validation. These insights provide a foundation for strategic marketing initiatives, customer retention efforts, and revenue optimization. Future research could integrate additional features such as product preferences and time-based spending habits for more refined segmentation.