eCommerce Tranactions

-BY MANJUNATH
Manjunathmaktal49@gmail.com

1. Clustering Methodology

1.1 Data Preparation

- Customer profiles from Customers.csv were merged with transaction data from Transactions.csv.
- Key features used for clustering included:
  - **TotalValue**: Total transaction value for each customer.
  - **Quantity**: Total number of products purchased.
  - **ProductDiversity**: Number of unique products purchased.
- Missing values were filled with 0 to ensure completeness.

1.2 Clustering Algorithm

- K Means Clustering was selected for its simplicity and efficiency.
- The number of clusters was set to 3, based on exploratory evaluation

1.3 Evaluation Metrics:

- The **Davis-Bouldin Index (DB Index)** was calculated to assess clustering quality.
- A lower DB Index indicates better-defined clusters.

2. Result

2.1 Number of Clusters:

- **3 Clusters** were formed, representing distinct customer groups.

2.2 Cluster Characteristics:

- **Cluster 0**:
  - High spenders with a large quantity of purchases.
  - Customers in this cluster contribute significantly to revenue.
- **Cluster 1**:
  - Low spenders with low product diversity.
  - Likely occasional buyers or price-sensitive customers.
- **Cluster 2**:
  - Medium spenders with moderate product diversity.
  - Represent a balanced customer segment with potential for growth.

2.3 DB Index:

- **Davis-Bouldin Index**: **0.78**
  - Indicates good separation and compactness of clusters.

2.4 Visualization:

- Scatter plots of customer data revealed clear separation between the clusters, supporting the segmentation.

3. Business Insights

1. **High Revenue Drivers**:
    - Customers in Cluster 0 are the primary revenue drivers. Targeting them with premium offers and loyalty programs could enhance retention and profitability.
2. **Growth Potential**:
    - Cluster 2 represents a group with moderate spending and diversity. Cross-selling and up-selling campaigns could increase their lifetime value.
3. **Cost-Efficient Strategies**:
    - Cluster 1 contains low spenders who may respond well to discounts or introductory offers to increase engagement.
4. **Product Preferences**:
    - Understanding product preferences within each cluster can guide personalized marketing strategies.
5. **Regional Variations**:
    - By analyzing region-specific patterns within clusters, businesses can tailor campaigns to different markets effectively.

Conclusion

The clustering analysis successfully segmented the customer base into three distinct groups, each with unique characteristics. The insights derived can guide targeted marketing, customer retention efforts, and strategic decision-making to maximize business outcomes.

Key Metrics

- **Number of Clusters**: 3
- **Davis-Bouldin Index**: 0.78
- **Features Used**: TotalValue, Quantity, ProductDiversity