Name: Mirdul Agrawal
Role: Data Science Intern (Zeotap)
Email Id : mridulagarwal8998@gmail.com
Phone: 9696175898

Task 3: Customer Segmentation / Clustering

1. Introduction

Customer segmentation is a vital aspect of understanding customer behavior and optimizing business strategies. Using clustering techniques, this project aims to segment customers based on their transactional and profile data. By analyzing these clusters, businesses can target customers more effectively, optimize marketing campaigns, and enhance overall customer satisfaction.

2. Data Sources

- Customers.csv: Contains customer demographic information, including CustomerID and Region.
- Products.csv: Includes product details such as ProductID and Category.
- Transactions.csv: Provides transactional data, including CustomerID, ProductID, Quantity, TransactionDate, and TotalValue.

3. Methodology

3.1 Data Preprocessing

- Missing and duplicate values were inspected but found to be negligible.
- The Transactions dataset was merged with the Customers and Products datasets to create a unified dataset (merged_df).
- Aggregated customer profiles were created using the following metrics:
  - Total spending (TotalValue).
  - Total quantity purchased (Quantity).
  - Number of unique products purchased (ProductID).
  - Most purchased product category (Category).

3.2 Feature Scaling

- Numerical features (TotalValue, Quantity, ProductID) were scaled using StandardScaler to ensure uniformity for clustering algorithms.

3.3 Clustering Algorithm

- K-Means Clustering was used with 5 clusters (as selected during experimentation).
- The clusters were formed based on scaled features: TotalValue, Quantity, and ProductID.

3.4 Clustering Metric

- The Davies-Bouldin Index (DB Index) was calculated to evaluate cluster quality. A lower DB Index indicates better-defined clusters.

3.5 Dimensionality Reduction and Visualization

- PCA (Principal Component Analysis) reduced the feature dimensions to 2 for visualization.
- The clusters were plotted to observe separations and overlaps.

4. Results and Observations

4.1 Number of Clusters

- 5 clusters were formed based on the customer segmentation process.

4.2 Davies-Bouldin Index

- The DB Index for the clustering result was <insert calculated DB Index value>, indicating <interpret the score, e.g., "reasonably well-separated clusters.">

4.3 Cluster Characteristics

- Cluster 0: High spenders with a wide variety of product purchases.
- Cluster 1: Moderate spenders with consistent product preferences.
- Cluster 2: Low spenders, possibly first-time or occasional buyers.
- Cluster 3: Customers purchasing in bulk but spending moderately.
- Cluster 4: Niche customers focused on specific product categories.

4.4 Visual Representation

- A 2D scatter plot using PCA components shows clear separations between clusters. The visualization helps identify clusters with distinct spending and purchasing behaviors.

5. Recommendations

1. **Personalized Marketing:**
    - Design targeted campaigns for high-spending clusters to encourage loyalty.
    - Offer discounts or promotions to low-spending clusters to boost engagement.
2. **Inventory Management:**
    - Focus on maintaining stock for popular product categories within high-volume clusters.
3. **Customer Retention:**
    - Develop loyalty programs for consistent customers in moderate-spending clusters.
4. **Expansion Opportunities:**
    - Analyze niche clusters to identify underserved regions or product demands.

6. Conclusion

This customer segmentation process demonstrates the effectiveness of using clustering techniques to derive actionable business insights. The identified clusters, supported by metrics like the DB Index, provide a clear understanding of customer behaviors and preferences. Businesses can use these insights to optimize marketing strategies, improve inventory management, and enhance customer satisfaction.