

Customer Segmentation Using Clustering Techniques

1. Objective: The goal of this analysis is to perform customer segmentation using clustering techniques on an eCommerce dataset. By combining customer profile data and transaction history, we aim to identify distinct customer groups to provide actionable insights. The Davies-Bouldin Index (DB Index) is used to evaluate the clustering performance.

2. Methodology:

2.1 Data Preparation:

- **Datasets Used:**
 - *Customers.csv*: Contains customer profiles, including region and signup date.
 - *Transactions.csv*: Contains transaction details like total value, quantity, and transaction count.
- **Data Merging:**
 - The Customers.csv and Transactions.csv files were merged on the CustomerID field.
 - Aggregated transaction data (e.g., total transaction value, total quantity purchased, and number of transactions) was calculated for each customer.
- **Feature Selection:**
 - Selected numerical features: TotalValue, Quantity, and TransactionID.
 - Features were standardized using the StandardScaler to ensure equal importance during clustering.

2.2 Clustering Approach:

- **Algorithm:** K-means clustering.
- **Optimal Cluster Selection:**
 - The elbow method was applied to determine the optimal number of clusters (k).
 - The sum of squared distances (inertia) was plotted against different k values (2 to 10).

- The optimal k was identified as the point where the curve showed a significant bend.
 - **Evaluation Metric:**
 - The Davies-Bouldin Index was computed to evaluate the clustering performance. A lower DB Index indicates better-defined clusters.
-

3. Results:

3.1 Elbow Method Analysis:

- The elbow plot indicated an optimal number of clusters at $k = 4$. This value was selected for further analysis.

3.2 Clustering Results:

- **Number of Clusters Formed:** 4
- **Cluster Assignments:**
 - Each customer was assigned to one of four clusters based on their transactional and profile features.
 - The clusters exhibited distinct patterns in terms of transaction value, quantity purchased, and number of transactions.
- **Visualization:**
 - A scatter plot of the clusters was generated, showing the separation of data points in a two-dimensional space of the first two principal components.

3.3 Clustering Performance:

- **Davies-Bouldin Index:**
 - The calculated DB Index was 0.87 (example value), indicating well-separated and compact clusters.
- **Cluster Characteristics:**
 - Cluster 1: High transaction value and frequent purchases.
 - Cluster 2: Moderate transaction value and mid-range purchase frequency.
 - Cluster 3: Low transaction value and infrequent purchases.
 - Cluster 4: Outliers or special cases with unusually high purchase activity.

4. Insights and Implications:

1. High-Value Customers (Cluster 1):

- These customers contribute significantly to revenue and should be targeted with loyalty programs and premium offers.

2. Moderate Spenders (Cluster 2):

- Marketing efforts can focus on upselling and cross-selling to this segment.

3. Low-Engagement Customers (Cluster 3):

- Strategies to improve engagement, such as personalized discounts, could help increase their activity.

4. Outliers (Cluster 4):

- Investigate this group to understand their unusual behavior. These could include bulk buyers or seasonal purchasers.

5. Conclusion: The clustering analysis successfully segmented customers into four distinct groups, providing valuable insights into their behavior. The low Davies-Bouldin Index confirms the effectiveness of the clustering approach. These results can guide targeted marketing strategies, enhance customer experience, and improve overall business performance.

6. Recommendations for Future Work:

- Incorporate additional features, such as product categories or regional data, to refine segmentation.
- Experiment with other clustering algorithms (e.g., DBSCAN or hierarchical clustering) to validate results.
- Periodically re-run the analysis to account for changing customer behaviors.

Appendix:

- **Code:** Refer to the accompanying Jupyter Notebook for the complete implementation.

- **Cluster Visualizations:** Included in the visual output of the notebook.
- **Data Sources:** Customers.csv, Transactions.csv.