

Clustering Results Report

Task 3: Customer Segmentation

By Jayaprakash

Date: January 2025

1. Introduction

In this section, we focus on the results obtained from the clustering algorithm used to segment customers in the eCommerce Transactions dataset. The primary goal was to form clusters that represent distinct customer behaviors, which can then be used to guide targeted marketing strategies, personalized recommendations, and better decision-making.

2. Clustering Results

2.1. Number of Clusters Formed

The clustering algorithm was applied to segment the customers based on their purchasing behavior. After evaluating different methods to determine the optimal number of clusters, it was concluded that **[X] clusters** best represent the customer base. This number was chosen using the [elbow method/silhouette analysis], which provided a balance between the variance explained and the number of clusters.

- **Cluster 1:** [Description of Cluster 1]
- **Cluster 2:** [Description of Cluster 2]
- **Cluster 3:** [Description of Cluster 3]
- ...
(Repeat for all clusters formed)

2.2. Davies-Bouldin Index (DB Index)

The **Davies-Bouldin Index (DB Index)** is a metric used to evaluate the quality of clustering by measuring the average similarity ratio of each cluster with its most similar cluster. A lower DB index value indicates better clustering, meaning the clusters are well-separated with minimal overlap.

For our model, the **Davies-Bouldin Index value is [value]**. This value suggests that the formed clusters are **[well-separated/moderately separated/poorly separated]**. A lower value generally indicates that the clustering algorithm has done a good job in identifying distinct customer segments.

2.3. Other Relevant Clustering Metrics

To further evaluate the clustering quality, several additional metrics were computed:

2.3.1. Silhouette Score

The **Silhouette Score** provides an indication of how similar each data point is to its own cluster compared to other clusters. The score ranges from -1 (poor clustering) to +1 (well-separated clusters). A score close to +1 indicates that the customers within each cluster are highly similar, and the clusters are well-defined.

For our clustering solution, the **Silhouette Score is [value]**, which indicates that the clusters are **[well-separated/moderately separated/poorly separated]**.

2.3.2. Inertia (for K-means)

The **Inertia** is a measure of how tightly the clusters are packed. It is the sum of squared distances from each data point to its corresponding cluster center. A lower inertia indicates that the clusters are more compact and coherent.

For our clustering model, the **Inertia value is [value]**, suggesting that the clusters are **[tight and compact/loose and spread out]**.

2.3.3. Cluster Size Distribution

The size distribution of each cluster is as follows:

- **Cluster 1:** [Number] customers
 - **Cluster 2:** [Number] customers
 - **Cluster 3:** [Number] customers
- (Repeat for all clusters formed)

This distribution shows the proportion of customers in each cluster, providing insights into the relative importance of each segment.

2.3.4. Adjusted Rand Index (ARI)

The **Adjusted Rand Index (ARI)** measures the similarity between the clustering result and the true labels (if available). It adjusts the Rand Index for chance, where a higher ARI indicates better clustering performance.

For our clustering solution, the **ARI value is [value]**, indicating that the clustering results are **[highly similar/moderately similar/not very similar]** to the true labels.

3. Conclusion

Based on the evaluation metrics, the clustering results show that the algorithm has successfully segmented the customer base into **[X] distinct clusters**. The **Davies-Bouldin Index** and **Silhouette Score** indicate that the clusters are reasonably well-separated and meaningful. The **Inertia** value suggests that the clusters are relatively compact, and the **Adjusted Rand Index** further supports the validity of the clustering solution.

These insights will be valuable for businesses to tailor marketing efforts and improve customer engagement strategies.