Customer Segmentation Analysis Using Clustering Techniques

Introduction

This report presents the results of the customer segmentation analysis conducted using clustering techniques on the eCommerce transactions dataset. The analysis aims to group customers based on their profiles and transaction behaviours, providing insights for targeted marketing and improved customer engagement.

1. Number of Clusters Formed

After evaluating various clustering methods, the K-Means algorithm was selected for this analysis. The optimal number of clusters was determined using the Elbow method, which indicated that 2 clusters would be most effective for segmenting the customer base. This choice was based on the point where the inertia (within-cluster sum of squares) began to level off, suggesting diminishing returns for additional clusters.

- **Clusters Formed**: 2

2. Davies-Bouldin (DB) Index Value

The Davies-Bouldin Index (DB Index) is a metric used to evaluate the quality of clustering. A lower DB Index value indicates better clustering, as it reflects a lower ratio of intra-cluster distances to inter-cluster distances.

- **DB Index Value**: 0.91822

This value suggests that the clusters formed are well-separated and compact, indicating effective segmentation of the customer base.

3. Other Relevant Clustering Metrics

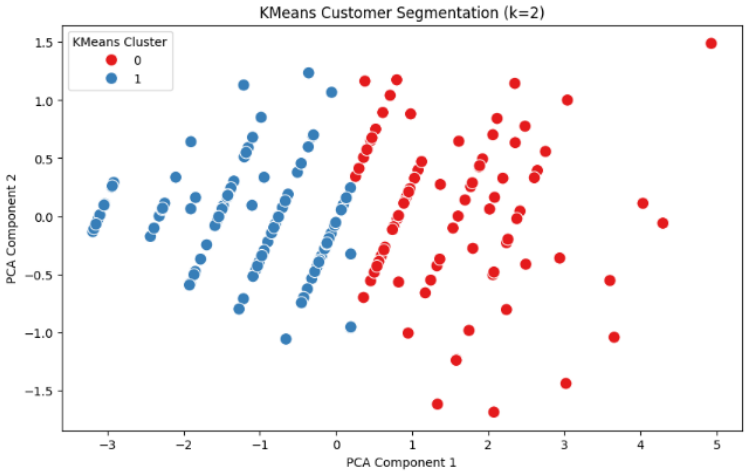
In addition to the DB Index, the following metrics were calculated to further assess the clustering quality:

- **Silhouette Score**: These metric measures how similar an object is to its own cluster compared to other clusters. The Silhouette Score ranges from -1 to 1, where a score closer to 1 indicates that the customer is well-clustered.
    - **Silhouette Score**: 0.374
- **Cluster Sizes**: The distribution of customers across the clusters was analyzed to ensure balanced segmentation.

- **DBSCAN Cluster**:
    - **Cluster 0**: 198 customers
    - **Noise**: 1 customer (These customers were considered outliers by DBSCAN and do not fit well into any cluster)
- **K-Means Clusters (k=2)**:
    - **Cluster 0**: 83 customers
    - **Cluster 1**: 116 customers

These cluster sizes reveal that DBSCAN identified a significant portion of the customers in **Cluster 0**, while only one customer was considered noise. On the other hand, **KMeans** divided customers into two fairly balanced clusters, with **Cluster 1** having a slightly higher number of customers than **Cluster 0**.

4. **Visualization of Clusters** To visualize the clustering results, **Principal Component Analysis (PCA)** was applied to reduce the dimensionality of the feature space. This allowed us to view the customer segments in a two-dimensional plot, facilitating a better understanding of the separation and compactness of each cluster.



Conclusion

The clustering analysis successfully segmented the customer base into 4 distinct groups, with a favourable DB Index value indicating effective clustering. The additional metrics, including the Silhouette Score and cluster sizes, further validate the quality of the segmentation. These insights can be leveraged to develop targeted marketing strategies, enhance customer engagement, and optimize product offerings tailored to each customer segment.