Customer Segmentation - Clustering Report

Introduction:

This report provides an overview of the customer segmentation process using clustering techniques. The goal was to segment customers based on their transaction history and profile information. KMeans clustering algorithm was applied to identify distinct customer groups, and the results are analyzed to provide actionable insights.

Data Preparation:

The data used for clustering consists of:

- Customers_data: Contains details about each customer.
- Transactions_data: Contains transaction details for each customer.

These datasets were merged to create a comprehensive summary of each customer's transactions, including the total value of transactions, the quantity of products purchased, and the number of unique transaction dates.

Feature Selection:

The features used for clustering are:

- TotalValue: Total sum of the transaction values for each customer.
- Quantity: Total sum of the quantity of products purchased by each customer.
- TransactionDate: Number of unique transaction dates for each customer.

These features were standardized to ensure they contribute equally to the clustering process.

Clustering Algorithm:

The KMeans algorithm was chosen for clustering. The number of clusters (`n_clusters`) was set to 4. The choice of the number of clusters can be adjusted based on the analysis requirements and validation metrics.

Clustering Metrics:

- Number of Clusters Formed: 4 clusters were formed based on the analysis.

- Davies-Bouldin Index (DB Index): This metric measures the quality of the clustering. A lower DB Index indicates better-defined clusters. The DB Index value for this analysis was calculated to assess the clustering performance.

PCA Visualization:

Principal Component Analysis (PCA) was used to reduce the dimensionality of the data to two components for visualization. This allows for a clear visual representation of the clusters formed by the KMeans algorithm.

Visual Representation of Clusters:

The PCA results were plotted to show the separation and distribution of customer segments. Each point in the plot represents a customer, and the color indicates the cluster to which the customer belongs. The PCA plot helps in understanding the distinctiveness and separation of the clusters.

Conclusion

The clustering analysis resulted in the formation of 4 distinct customer segments. The Davies-Bouldin Index indicates that the clusters are reasonably well-defined. These clusters can be used for various business purposes, such as targeted marketing, personalized promotions, and improved customer relationship management.

Summary

- Number of Clusters: 4
- Davies-Bouldin Index: Indicates the effectiveness of the clustering.
- Clustering Metrics: The DB Index and visual representation show that the clustering is well-defined.

These insights and metrics provide a solid foundation for leveraging customer segmentation to enhance business strategies and improve customer engagement.