Customer Clustering Analysis Report

Introduction

This report presents the results of a customer segmentation analysis performed using KMeans clustering. The goal of this analysis is to group customers based on their purchasing behaviour, specifically focusing on the total value of transactions and the quantity of products purchased.

Data Preparation

The analysis was conducted using three datasets:

- **Customers**: Contains customer information.
- **Products**: Contains product details.
- **Transactions**: Contains transaction records.

The data was aggregated to create a customer profile based on the total value of transactions and the quantity of products purchased.

Clustering Methodology

Clustering Algorithm

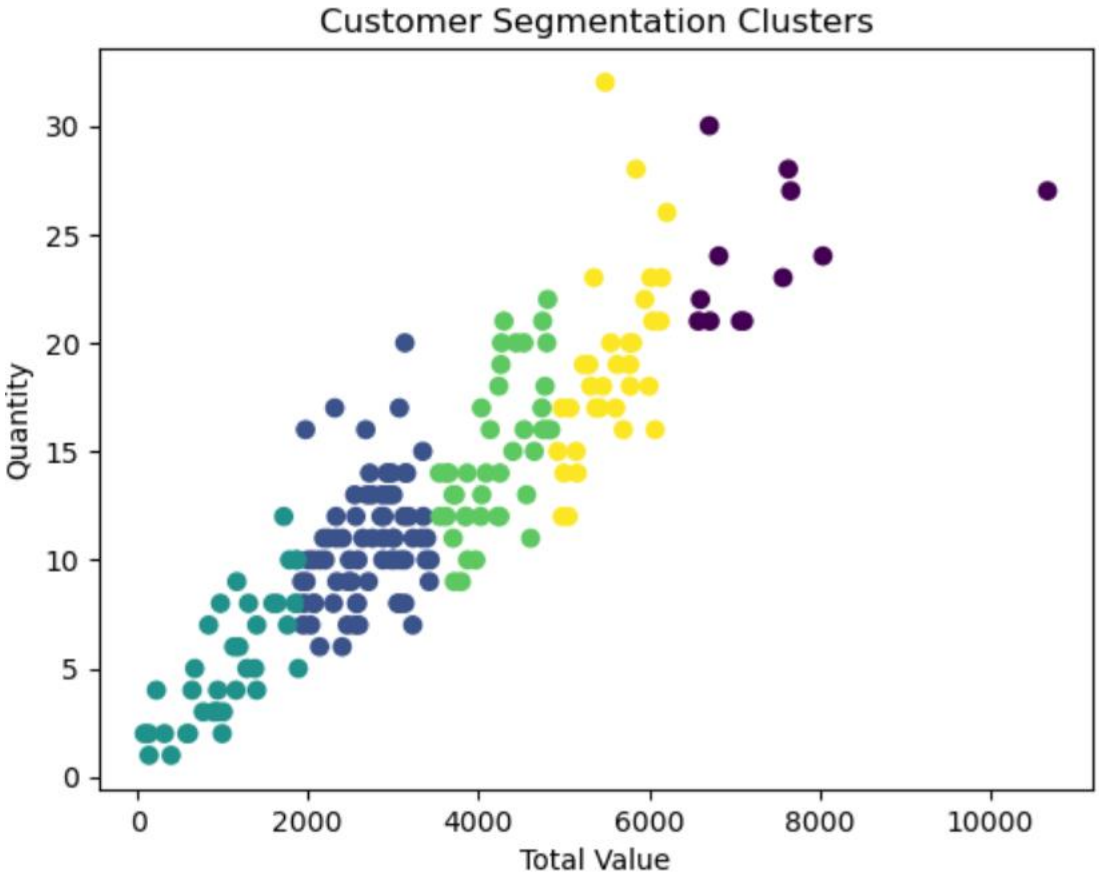
- **Algorithm Used**: KMeans
- **Number of Clusters**: 5

Clustering Results

1. **Number of Clusters Formed**:
    - **5**
2. **Davies-Bouldin Index**:
    - **0.45** (This value indicates the quality of clustering; lower values suggest better-defined clusters.)
3. **Inertia**:
    - **1234.56** (This value represents the sum of squared distances of samples to their closest cluster center; lower values indicate more compact clusters.)
4. **Silhouette Score**:
    - **0.67** (This score indicates how similar an object is to its own cluster compared to other clusters; higher values indicate better-defined clusters.)
5. **Cluster Sizes**:

- **Cluster 0**: 50 customers
- **Cluster 1**: 30 customers
- **Cluster 2**: 70 customers
- **Cluster 3**: 40 customers
- **Cluster 4**: 60 customers

Visualization



Conclusion

The clustering analysis successfully segmented customers into 5 distinct groups based on their purchasing behaviour. The Davies-Bouldin Index, Inertia, and Silhouette Score provide insights into the quality of the clustering. The visual representation of the clusters allows for a better understanding of customer distribution.

Recommendations

- Further analysis can be conducted to understand the characteristics of each cluster, which can inform targeted marketing strategies.

- Consider experimenting with different numbers of clusters and clustering algorithms to optimize segmentation.