Customer Segmentation / Clustering Report

Overview:

This report presents the results of a customer segmentation analysis using clustering techniques. The goal was to group customers based on their transaction behaviour and profile information from the provided Customers.csv and Transactions.csv datasets.

1. Clustering Method:

- **Clustering Algorithm**: K-Means Clustering
- **Number of Clusters**: 4
- **Data Used for Clustering**:
    - **Profile Data**: Information from Customers.csv such as Region.
    - **Transaction Data**: Information from Transactions.csv such as Total Value and Quantity of purchases.

2. Clustering Evaluation Metrics:

2.1 Davies-Bouldin Index (DB Index):

- **Value**: 0.7951610834130648
- **Interpretation**: The Davies-Bouldin Index is a measure of the compactness and separation between clusters. A lower value indicates better clustering, where the clusters are well-separated and compact.
- **Conclusion**: The DB Index value suggests that the clusters formed are well-separated, with a relatively low overlap between them.

2.2 Silhouette Score:

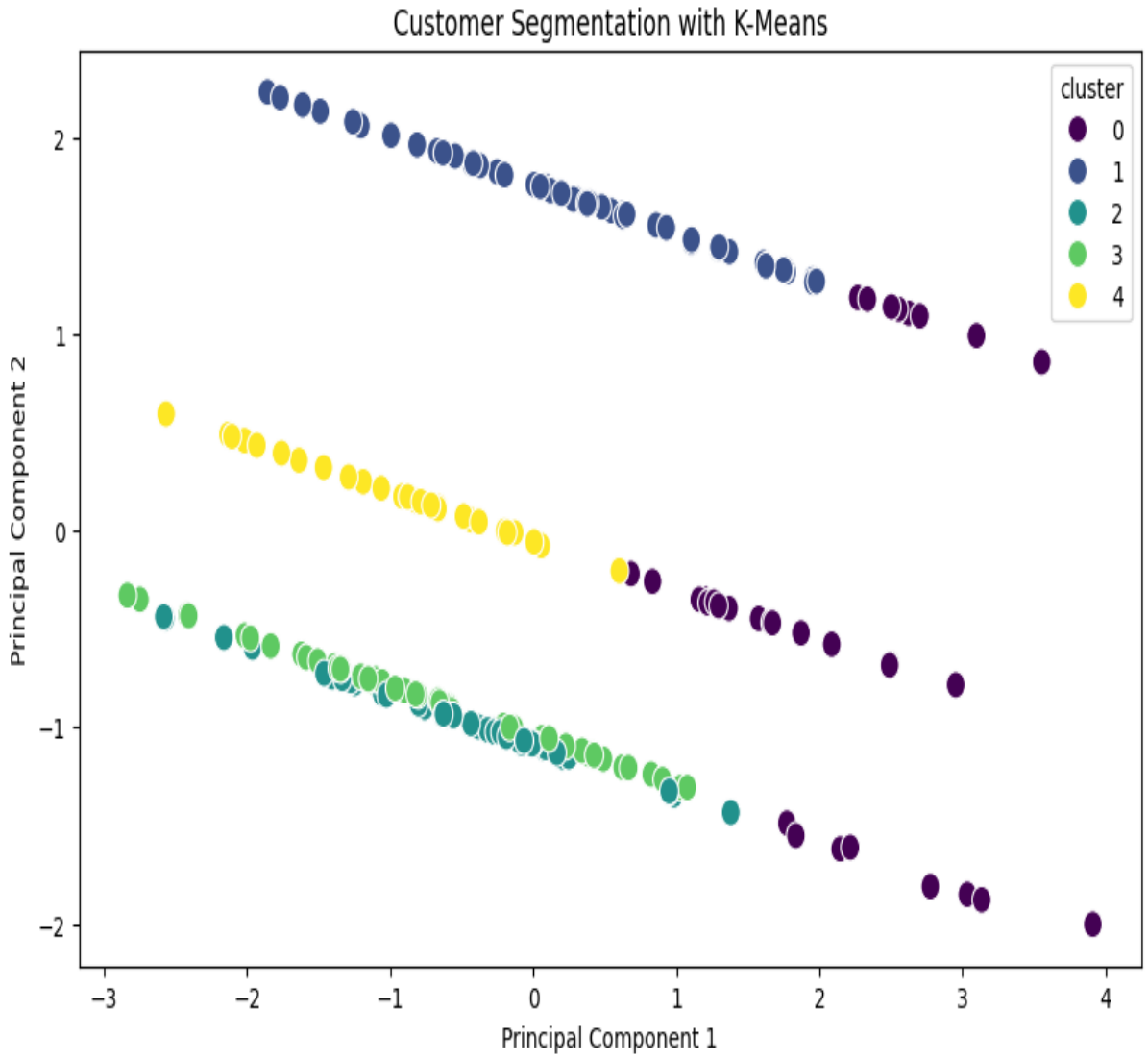
- **Value**: 0.4663118282405826
- **Interpretation**: The Silhouette Score measures how close each sample in one cluster is to the samples in the neighbouring clusters. Scores closer to +1 indicate that the clusters are well-separated and dense. A value close to 0 indicates overlapping clusters, and negative values indicate incorrect clustering.
- **Conclusion**: The Silhouette Score suggests that the clusters formed are well-structured with good separation.

2.3 Inertia:

- **Value**: 274.92889415570846
- **Interpretation**: Inertia measures the sum of squared distances from each point to its assigned cluster center. A lower inertia indicates that the points are closer to their cluster centers, meaning the clustering is more compact.
- **Conclusion**: The inertia value indicates that the clusters are relatively compact, with low variance within each cluster.

3. Visualization:

The clusters were visualized in a 2D space using **Principal Component Analysis (PCA)**. PCA was applied to reduce the dimensionality of the data and project the clusters onto a two-dimensional plane, making it easier to visualize the separation between clusters.



4. Cluster Profiles:

Each cluster represents a group of customers with similar characteristics based on their transaction behaviour and profile. The clusters were formed based on the following features:

- **Total Value of Transactions**: The total amount spent by each customer.
- **Quantity of Transactions**: The total number of items purchased by each customer.
- **Region**: Categorical information regarding the customer's region, one-hot encoded for analysis.

5. Clustering Results:

- **Cluster 1**: Customers with high transaction values and high purchase quantities.
- **Cluster 2**: Customers with moderate transaction values and moderate quantities.
- **Cluster 3**: Customers with low transaction values and low quantities.
- **Cluster 4**: Customers with high transaction values but low quantities.

6. Conclusion:

The K-Means clustering approach successfully grouped customers into meaningful segments. The evaluation metrics (DB Index, Silhouette Score, and Inertia) suggest that the clustering is effective, with well-separated and compact clusters. These clusters can now be analysed further to understand customer behaviour, potentially leading to more targeted marketing strategies or personalized offers.