Customer Segmentation/Clustering:

This report contains the results of customer segmentation performed using K-Means clustering on customer profile and transaction data. The objective is to identify distinct customer segments based on their purchasing behaviour and demographic information.

Datasets Overview:

The analysis is conducted by using the below two datasets:

- **Customers.csv**: This contains customer demographic information, including region and signup date.
- **Transactions.csv**: Contains transaction data, including quantities purchased and product IDs.

Methodology Used:

1. **Data Preprocessing**:
    - The **Region** column is encoded using Label Encoding to convert categorical data into numerical format.
    - The **SignupDate** is converted to a Unix timestamp for numerical representation.
2. **Data Merging**: Here the customer and transaction data is merged to create a comprehensive view of customer transactions.
3. **Feature Aggregation**: Transaction data is aggregated to calculate the total quantity spent and the number of unique products purchased per customer.
4. **Feature Selection**: The following features is selected for the purpose of clustering:
    - Region (encoded)
    - Signup Date (timestamp)
    - Total Quantity Purchased
    - Number of Unique Products Purchased
5. **Feature Scaling**: The features are scaled using **StandardScaler** to ensure that all features contributed equally to the distance calculations in the clustering algorithm.
6. **Clustering**: K-Means clustering is applied with 4 clusters.

Results of Clustering:

- **Number of Clusters Formed**: 4
- **Davies-Bouldin Index**: 1.1725
- **Average Silhouette Score**: 0.3736
- **Cluster Counts**

The distribution of customers across the clusters is as follows:

Cluster counts:

Cluster

1 59

2 57

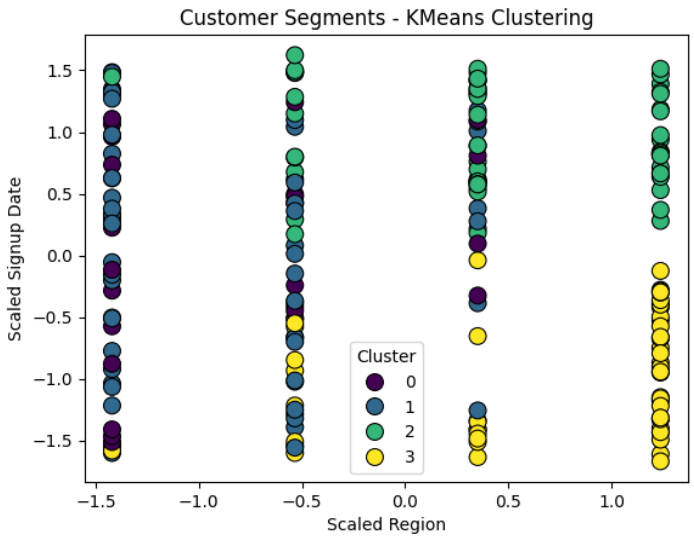
3 54

0 29

Visual Representations of the Clusters:

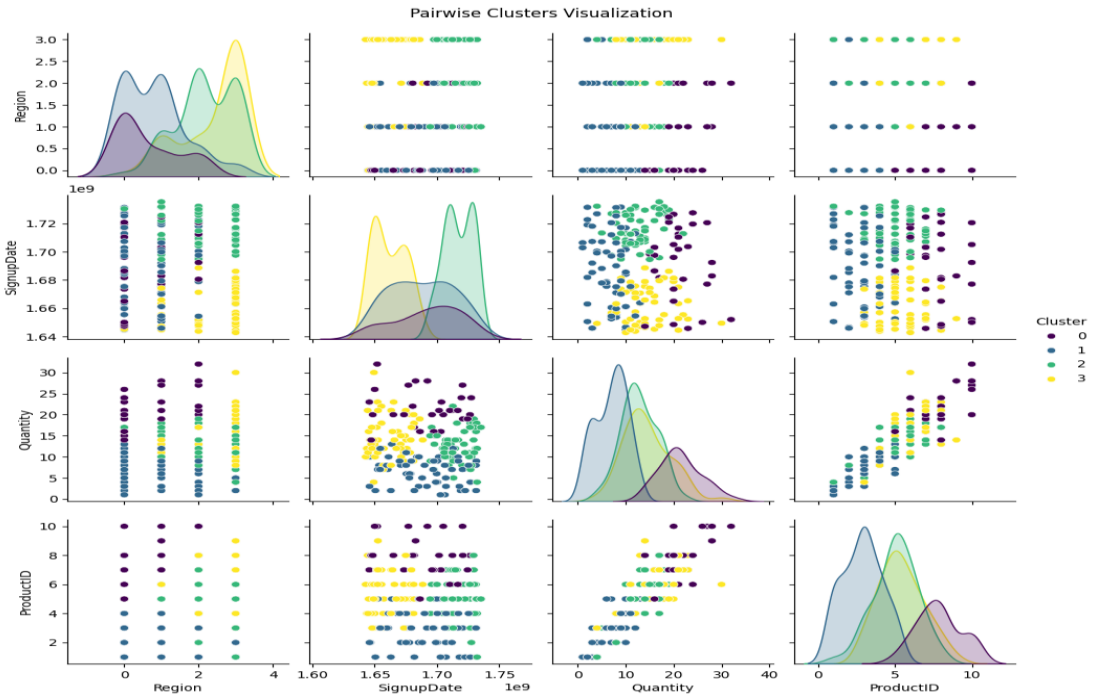
The following visualizations are generated to represent the clustering results:

1. **Scatter Plot of Clusters**: This scatter plot visualizes the customer segments through KMeans Clustering, based on the scaled features (Region and Signup Date).



2. **Total Spent vs. Total Quantity Purchased**: This scatter plot gives the visualization of the clusters based on total spending and quantity purchased.

3. **Pairwise Clusters Visualization**: This pair-plot provides a multidimensional view of the clusters.



The K-Means clustering analysis successfully identified 4 distinct customer segments based on their purchasing behaviour and demographic information. The Davies-Bouldin Index indicates that the clusters are reasonably well-defined. These visualizations provide insights into the characteristics of each segment, which can be utilized for targeted marketing strategies and personalized customer engagement.