Customer Segmentation/Clustering Report

This report presents the findings from a customer segmentation analysis conducted using K-Means clustering on customer profile and transaction data. The primary goal is to identify distinct customer segments based on their purchasing behavior and demographic characteristics.

Datasets Overview

The analysis utilizes the following two datasets:

- **Customers.csv**: This dataset includes demographic information about customers, such as their region and signup date.
- **Transactions.csv**: This dataset contains transaction details, including quantities purchased and product IDs.

Methodology

1. **Data Preprocessing**:
    - The **Region** column is encoded using Label Encoding to transform categorical data into a numerical format.
    - The **Signup Date** is converted into a Unix timestamp for numerical representation.
2. **Data Merging**: The customer and transaction datasets are merged to create a comprehensive view of customer transactions.
3. **Feature Aggregation**: The transaction data is aggregated to compute the total quantity spent and the number of unique products purchased by each customer.
4. **Feature Selection**: The following features are selected for clustering:
    - Encoded **Region**
    - **Signup Date** (timestamp)
    - **Total Quantity Purchased**
    - **Number of Unique Products Purchased**
5. **Feature Scaling**: The features are scaled using Standard Scaler to ensure that all features contribute equally to the distance calculations in the clustering algorithm.
6. **Clustering**: K-Means clustering is applied, resulting in the formation of 4 clusters.

Clustering Results

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

Cluster Distribution

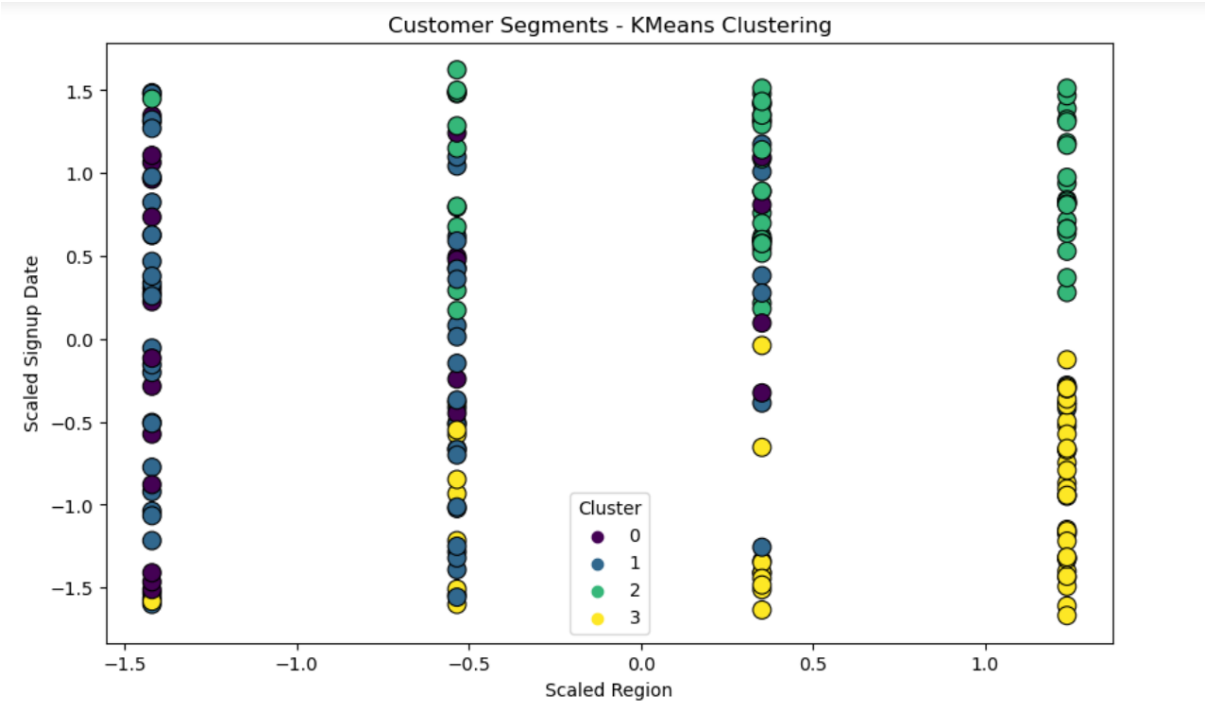
The distribution of customers across the clusters is as follows:

Cluster	Count
0	29
1	59
2	57
3	54

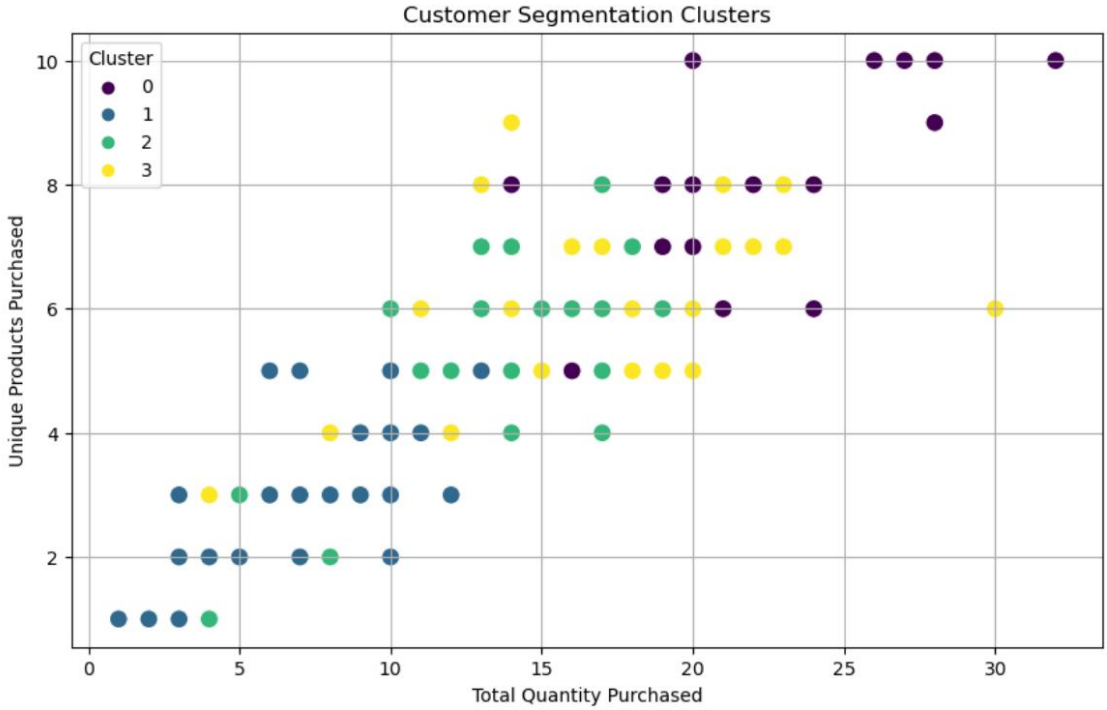
Visual Representations of the Clusters

Several visualizations have been generated to illustrate the clustering results:

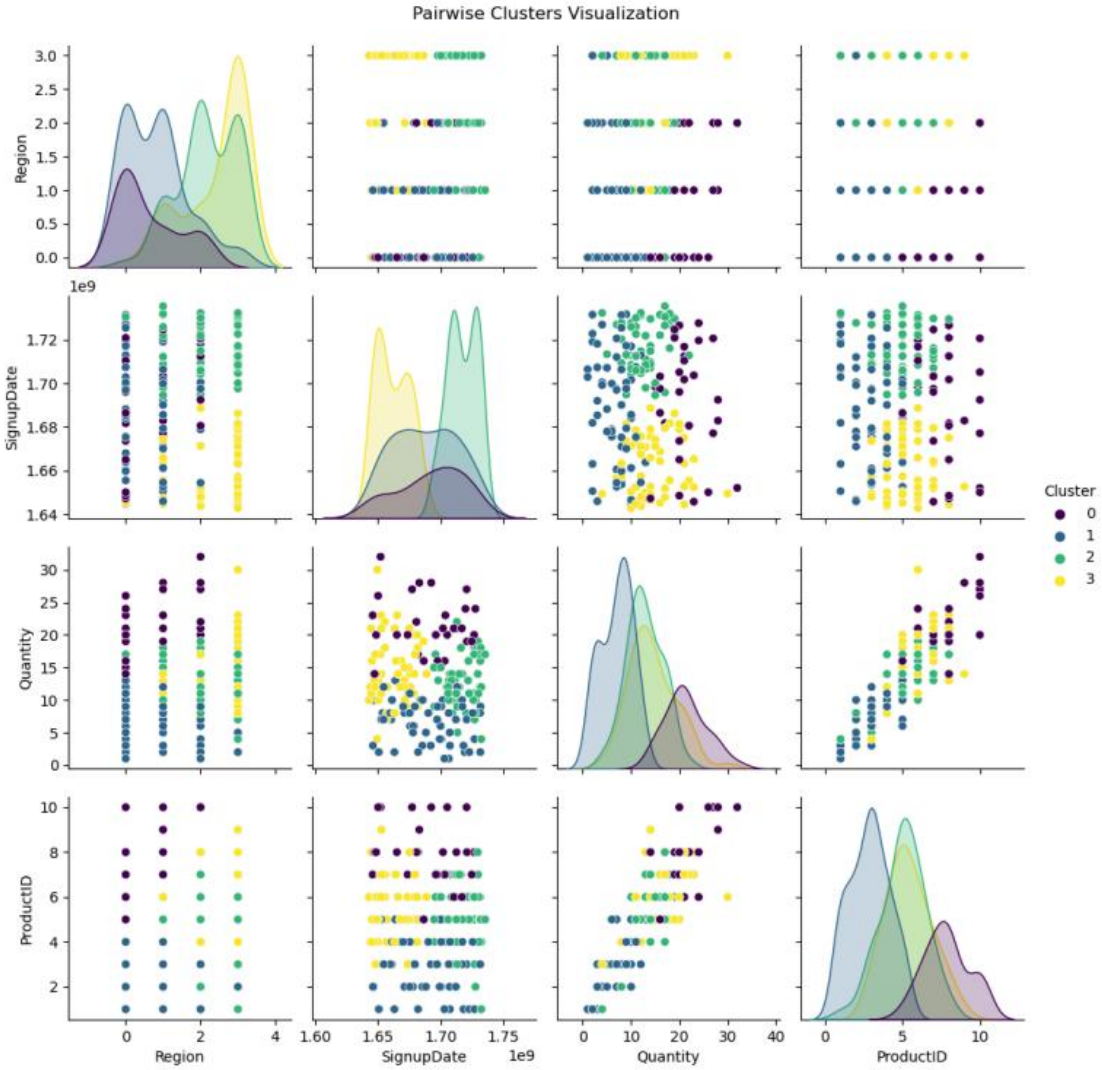
1. **Scatter Plot of Clusters**: This scatter plot visualizes customer segments identified through K-Means clustering, based on the scaled features (Region and Signup Date).



2. **Total Spent vs. Total Quantity Purchased**: This scatter plot visualizes 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 four distinct customer segments based on their purchasing behavior and demographic information. The Davies-Bouldin Index suggests that the clusters are reasonably well-defined. These visualizations offer valuable insights into the characteristics of each segment, which can be leveraged for targeted marketing strategies and personalized customer engagement.