Customer Segmentation through Clustering Analysis

This report presents the outcomes of customer segmentation performed using the K-Means clustering algorithm. The analysis aims to identify unique customer groups based on purchasing behaviors and demographic data.

Overview of Datasets

The analysis utilized two datasets:

1. **Customers.csv**: Contains demographic details such as region and signup date.
2. **Transactions.csv**: Provides transaction information, including quantities purchased and product IDs.

Methodology

1. **Data Preprocessing**:
    - Categorical data in the "Region" column was converted to numerical values using Label Encoding.
    - The "Signup Date" was transformed into Unix timestamps for numerical representation.
2. **Data Integration**: Customer and transaction data were merged for a complete view of customer behavior.
3. **Feature Aggregation**: Transaction data was summarized to compute total quantities purchased and the count of unique products per customer.
4. **Feature Selection**: Selected variables for clustering:
    - Encoded region
    - Signup date (timestamp)
    - Total quantity purchased
    - Number of unique products bought
5. **Feature Scaling**: StandardScaler was used to normalize features, ensuring equal contribution to clustering calculations.
6. **Clustering Approach**: The K-Means algorithm was applied, generating 4 clusters.

Clustering Results

- **Davies-Bouldin Index**: 1.4496
- **Average Silhouette Score**: 0.2335
- **Number of Clusters**: 4

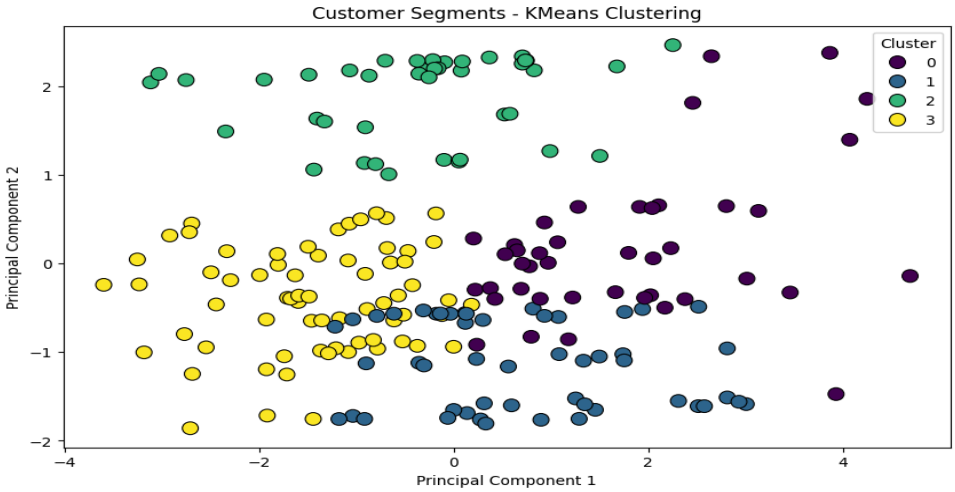
The distribution of customers across clusters.

Cluster counts:

Cluster	
3	67
1	50
0	42
2	41

Visual Insights

1. **Cluster Scatter Plot**: Displays customer groups based on encoded region .



2. **Spending vs. Quantity Scatter Plot**: Highlights variations in total spending and quantity purchased across clusters.