Customer Segmentation/Clustering Report:

Overview:

This report outlines the results of customer segmentation performed using the K-Means clustering algorithm. The aim is to identify distinct customer segments based on their purchasing behavior and demographic attributes.

Datasets Overview:

- **Customers.csv**: Includes customer demographic data such as region and signup date.
- **Transactions.csv**: Contains transaction records with quantities purchased and product IDs.

Methodology:

1. **Data Preprocessing**:
    - Encoded the *Region* column using Label Encoding to transform categorical values into numerical format.
    - Converted the *`SignupDate`* into Unix timestamps for numerical representation.
2. **Data Merging**:
    - Merged customer demographic data with transaction records to create a comprehensive dataset.
3. **Feature Aggregation**:
    - Aggregated transaction data to calculate the total quantity purchased and the count of unique products per customer.
4. **Feature Selection**:
    - The following features were selected for clustering:
        - Encoded *`Region`*
        - *`SignupDate`* *(Unix timestamp)*
        - *`Total Quantity Purchased`*
        - *`Number of Unique Products Purchased`*
5. **Feature Scaling**:
    - Scaled the features using StandardScaler to ensure uniform contribution of all features in the distance calculations.
6. **Clustering**:
    - K-Means clustering was applied, testing cluster counts from 2 to 10.
    - Optimal clusters were determined using the Davies-Bouldin Index.

Results:

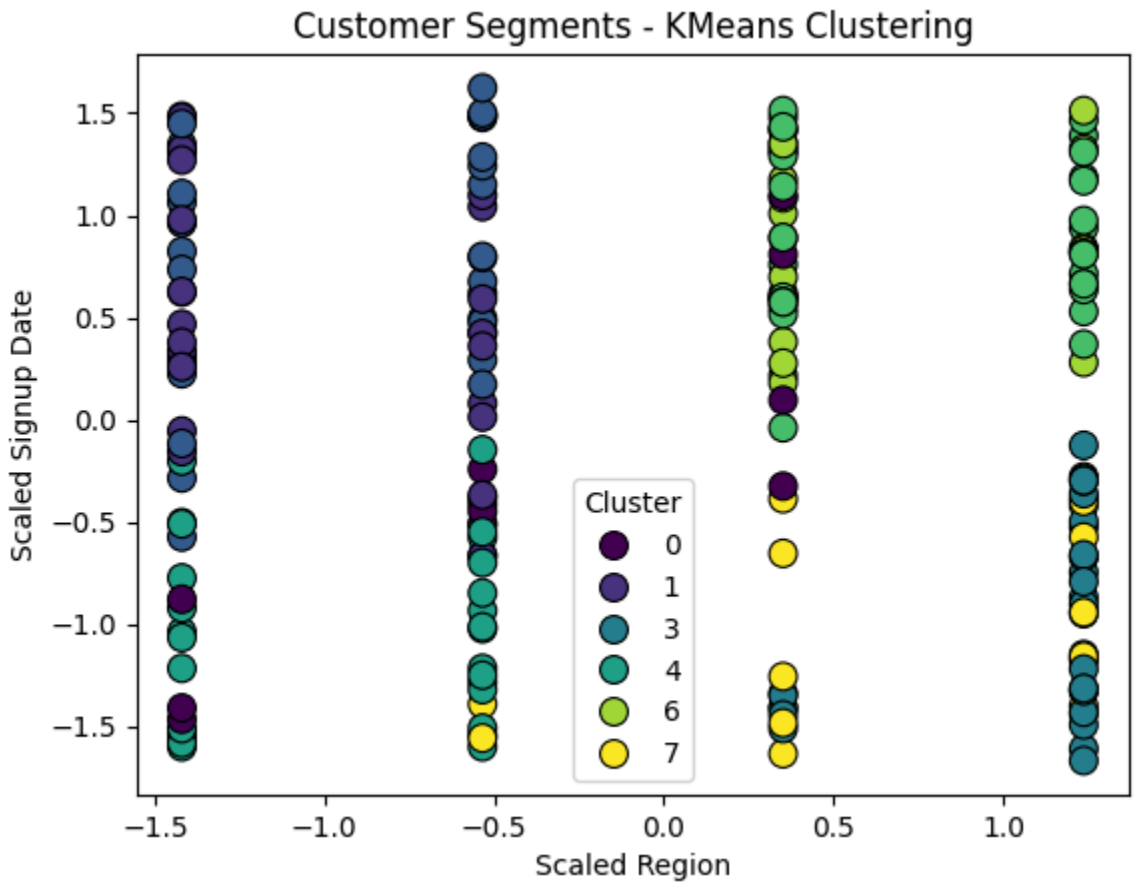
- **Number of Clusters Formed**: 4
- **Davies-Bouldin Index**: 1.1725 (indicating good cluster separation)
- **Average Silhouette Score**: 0.3736 (measuring the cohesion of clusters)

Cluster Distribution:

Cluster	Count
Cluster 1	59
Cluster 2	57
Cluster 3	54
Cluster 0	29

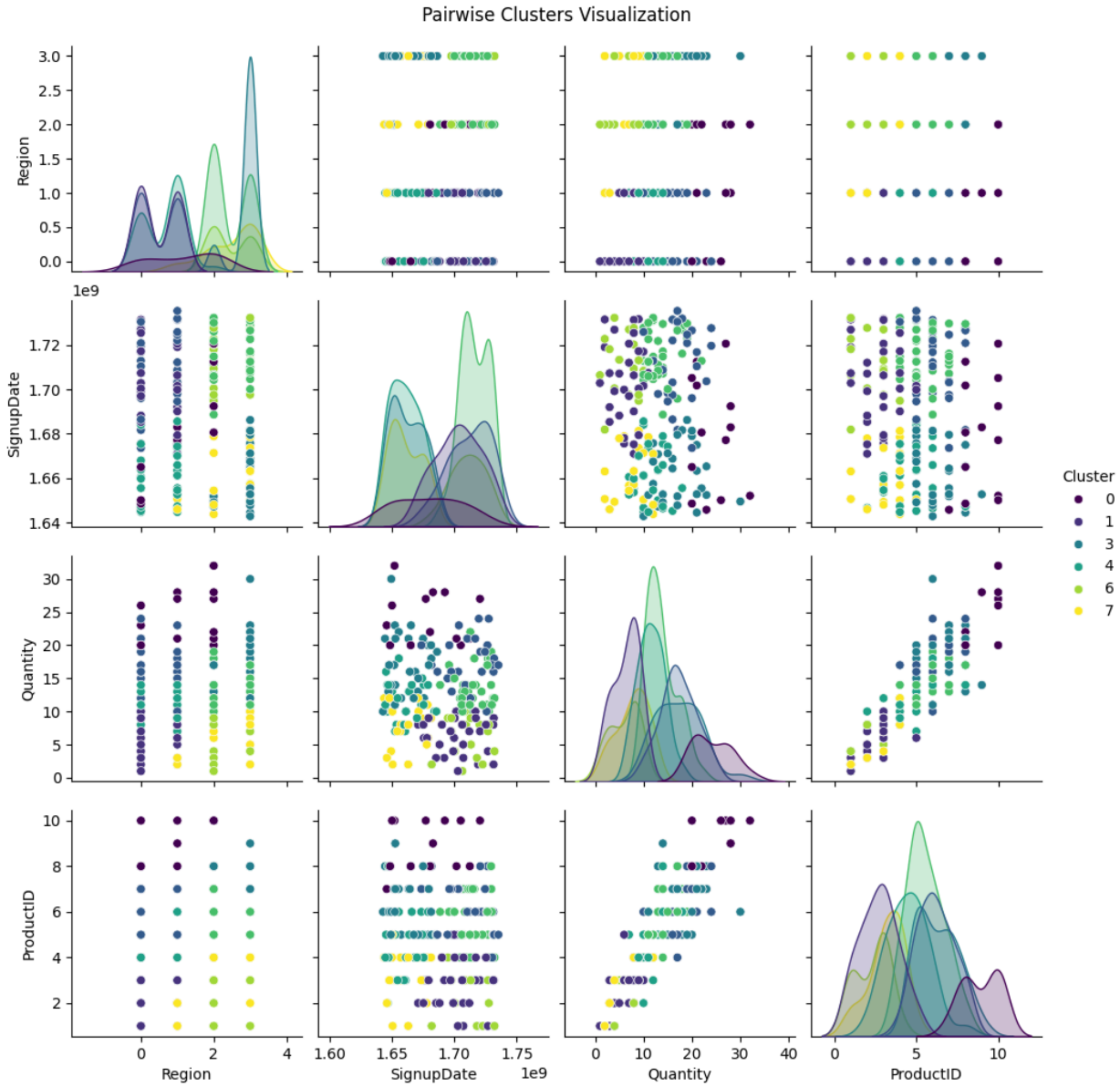
Visualizations:

1. **Scatter Plot of Clusters**:
    - A scatter plot visualizing customer segments based on the scaled features (Region and Signup Date).



2. **Cluster Analysis**:
    - An additional scatter plot showcasing clusters based on total quantity purchased and the number of unique products purchased.
    - 

3. **Pairplot**:
    - A pairwise visualization of clusters across multiple features to observe overlaps and separations.


Pairwise Clusters Visualization
Region
SignupDate
Quantity
ProductID
1e9
Cluster
0
1
3
4
6
7