

# Customer Segmentation Report

## Overview

Customer segmentation was performed using clustering techniques to group customers based on transactional and profile data. The clustering process aimed to identify distinct customer groups to help target marketing efforts and improve business strategies.

---

## Methodology

### Data Sources

1. Customers.csv: Contains customer profile information (CustomerID, Region, SignupDate).
2. Transactions.csv: Includes transactional details (TransactionID, ProductID, TransactionDate, Quantity, TotalValue, Price).

### Feature Engineering

#### Features used for clustering:

- TotalSpending: Sum of transaction values per customer.
- TransactionCount: Number of transactions per customer.
- TotalQuantity: Total quantity of products purchased by the customer.
- Region: Encoded as numerical values based on customer regions.

### Clustering Technique

- Algorithm: KMeans
- Number of Clusters: 5 (configurable between 2 and 10)
- Scaling: StandardScaler was used to normalize the data before clustering.

### Evaluation Metrics

- Davies-Bouldin Index: Measures clustering quality (lower values indicate better clusters).

### Visualization

Pairplots were used to visualize the clusters in a multi-dimensional space.

---

## Results

### Clustering Summary

- Number of Clusters: 5
- Davies-Bouldin Index: 1.0175

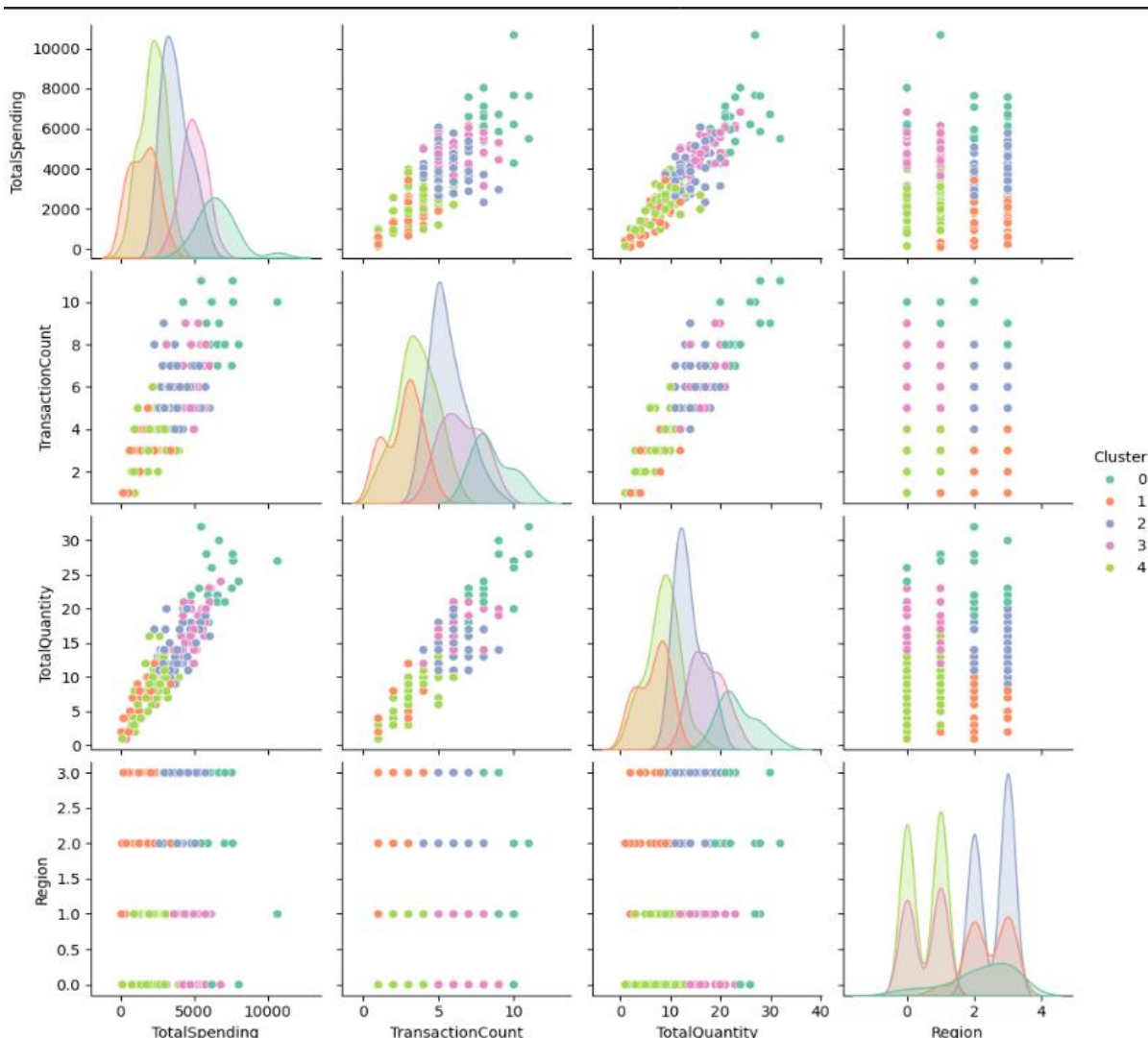
### Cluster Characteristics

Clusters were analyzed based on average spending, transaction count, and quantity:

- **Cluster 0:** Customers with moderate spending (AvgSpending: \$2214.85), low transaction counts (AvgTransactionCount: 3.57), and moderate product quantities (AvgQuantity: 8.51). These customers are distributed evenly across regions (AvgRegion: 0.51).
- **Cluster 1:** High-spending customers (AvgSpending: \$6316.50) with a high number of transactions (AvgTransactionCount: 8.22) and significant product quantities purchased (AvgQuantity: 22.52). These customers are primarily from urban regions (AvgRegion: 2.48).
- **Cluster 2:** Customers with relatively high spending (AvgSpending: \$5085.74), moderate transaction counts (AvgTransactionCount: 6.67), and moderately high product quantities purchased (AvgQuantity: 17.83). These customers are mainly in suburban regions (AvgRegion: 0.50).
- **Cluster 3:** Low-spending customers (AvgSpending: \$1436.36) with very low transaction counts (AvgTransactionCount: 2.52) and low product quantities (AvgQuantity: 6.10). These customers are generally from urban areas (AvgRegion: 2.39).
- **Cluster 4:** Customers with moderate-to-high spending (AvgSpending: \$3566.34), moderate transaction counts (AvgTransactionCount: 5.43), and moderate product quantities (AvgQuantity: 13.16). These customers are mostly located in urban areas (AvgRegion: 2.55).

## Visual Representation

A pairplot visualization of the clusters is included to show how customers are grouped based on the selected features.



## **Conclusion**

Customer segmentation revealed distinct groups based on their purchasing behavior and profile information. These insights can be leveraged for targeted marketing strategies and personalized customer engagement.