Customer Segmentation Report

Objective

The goal of this project was to perform customer segmentation using clustering techniques to group customers based on their transaction and profile information. The segmentation aimed to identify distinct customer clusters that could be used for targeted marketing strategies and business insights.

Data Preparation

1. **Datasets Used**:
    - **Customers.csv**: Contains customer profile information.
    - **Transactions.csv**: Contains transaction data.
2. **Data Integration**:
    - Both datasets were merged on `CustomerID` to create a comprehensive dataset for analysis.
3. **Feature Engineering**:
    - Aggregated transaction data to compute:
        - **Total Transactions**: The total monetary value of transactions per customer.
        - **Average Price**: The average price of items purchased.
        - **Transaction Count**: The number of transactions for each customer.
    - Converted the categorical `Region` variable into dummy variables for clustering.
4. **Data Preprocessing**:
    - Handled missing values by filling them with the mean for numeric columns.
    - Standardized features using **StandardScaler** to ensure uniform scaling.

Clustering Methodology

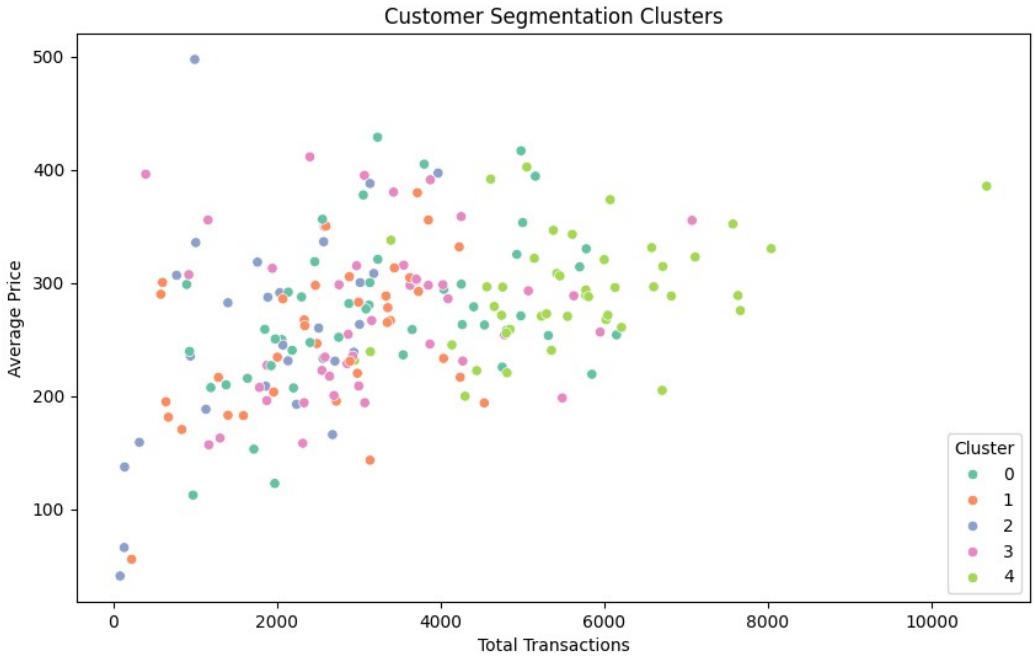
1. **Clustering Algorithm**:
    - **KMeans Clustering** with 5 clusters (`n_clusters=5`), chosen based on experimentation within the range of 2–10 clusters.
    - The clustering process grouped customers based on their transaction behavior and regional information.
2. **Evaluation Metric**:
    - **Davies-Bouldin Index (DB Index)**: A value of **1.1275** was achieved, indicating well-defined clusters. Lower DB Index values suggest better cluster separation and compactness.

Results

1. **Number of Clusters**:
    - A total of **5 clusters** were formed, labeled as `0`, `1`, `2`, `3`, and `4`.

2. **Cluster Characteristics**:
    - The clusters vary in terms of total transaction values, average price, and transaction counts:
        - **Cluster Sizes**:
            - Cluster `0`: 47 customers.
            - Cluster `1`: 36 customers.
            - Cluster `2`: 29 customers.
            - Cluster `3`: 43 customers.
            - Cluster `4`: 44 customers.
        - **Feature Ranges**:
            - Total Transactions: 82.36 to 10,673.87.
            - Average Price: 41.18 to 497.76.
            - Transaction Counts: 1 to 11.
3. **Cluster Visualization**:
    - The clustering results were visualized in a scatter plot using `Total Transactions` and `Average Price` as axes. The plot below highlights the clusters with distinct color coding:

Cluster Visualization:



4. **Data Output**:
    - The clustering results, including cluster labels and key features, are saved in `Customer_Segmentation_Results.csv`.