Customer Segmentation

Objective:

The objective of this task is to perform customer segmentation by identifying distinct customer groups using clustering techniques. This segmentation will help derive actionable insights to improve targeted marketing strategies, optimize business operations, and enhance customer engagement.

Methodology:

1. **Data Preparation**:
    - Combined **Customers.csv** and **Transactions.csv** datasets.
    - Engineered features such as:
        - **TotalValue**: Total customer spend.
        - **Quantity**: Total number of products purchased.
        - **TransactionCount**: Number of transactions.
    - Included profile information like customer **Region** (one-hot encoded for analysis).
2. **Feature Scaling**:
    - Standardized all numeric features using **StandardScaler** to ensure uniform contribution across variables.
3. **Clustering Algorithm**:
    - Used the **K-Means** clustering algorithm to group customers based on transactional and profile data.
    - Experimented with 2 to 10 clusters and selected the optimal number using the **Elbow Method** and **Davies-Bouldin Index**.
4. **Evaluation Metrics**:
    - **Davies-Bouldin Index**: Evaluated cluster compactness and separation. Lower values indicate better clustering.
    - **Silhouette Score**: Measured how well each data point fits within its cluster. Higher scores indicate better-defined clusters.

Results:

- Number of Clusters: 4
- Davies-Bouldin Index: 0.988
- Silhouette Score: 0.396

Cluster Profiles:

1. **Cluster 0**:
    - **Characteristics**: Moderate spenders with average purchase frequency across multiple regions.
    - **Action**: Target with general promotions and product bundles to increase engagement.
2. **Cluster 1**:
    - **Characteristics**: High spenders who make frequent purchases, contributing significantly to overall revenue.
    - **Action**: Focus on loyalty programs, premium product recommendations, and exclusive offers.
3. **Cluster 2**:
    - **Characteristics**: Customers with occasional purchases and low total spending.
    - **Action**: Design campaigns to re-engage these customers with discounts and referral incentives.
4. **Cluster 3**:
    - **Characteristics**: Regional customers with consistent spending patterns.
    - **Action**: Implement region-specific campaigns to maximize their value.

Visualizations:

1. **Elbow Curve**:
    - The Elbow Method identified 4 as the optimal number of clusters, as the **Within-Cluster Sum of Squares (WCSS)** flattens significantly at this point.
2. **Scatter Plot**:
    - A scatter plot of **Total Spend (TotalValue)** vs. **Total Quantity Purchased** clearly shows the separation between clusters.
    - Color-coded points represent customers in different clusters.
3. **Cluster Distribution**:
    - A bar chart demonstrates the number of customers in each cluster, providing insights into the relative sizes of the groups.

Recommendations:

1. **Cluster 1** customers should be prioritized as high-value clients, with personalized offers to maintain loyalty.
2. Design campaigns for **Cluster 0** to boost their spending through targeted promotions or upselling.
3. Re-engage **Cluster 2** with campaigns to increase their activity and spending frequency.
4. Focus on region-specific products and campaigns for **Cluster 3**, as they exhibit consistent spending patterns in their regions.