Number of Clusters Formed:

- **Number of Clusters:** The **K-Means** algorithm was used to cluster customers, and **4 clusters** were formed based on the characteristics of the customers' profiles and transaction history.
- The choice of 4 clusters was determined by using the **Elbow Method** to identify the optimal number of clusters, where the within-cluster sum of squares (WCSS) showed the most significant drop at 4 clusters.

Cluster Analysis:

Each cluster represents a group of customers with similar characteristics and behaviors. Below is a description of the clusters:

1. **Cluster 1**:
    - **Characteristics**: Customers in this cluster have **high total spending** and **high transaction frequency**. They tend to buy a wide variety of products.
    - **Behavior**: This segment represents loyal customers who are frequent buyers, likely making up the **high-value customer** group.
2. **Cluster 2**:
    - **Characteristics**: Customers in this cluster have **moderate total spending** but **low frequency of transactions**.
    - **Behavior**: This group may be less engaged but still makes substantial purchases. Targeted promotions could increase their frequency of purchases.
3. **Cluster 3**:
    - **Characteristics**: Customers in this cluster have **low total spending** but **high transaction frequency**.
    - **Behavior**: This segment represents customers who purchase more frequently but buy lower-cost items. They may be ideal candidates for **promotional offers** or loyalty programs to increase spending.
4. **Cluster 4**:
    - **Characteristics**: Customers in this cluster have **low total spending** and **low transaction frequency**.
    - **Behavior**: This group consists of customers who rarely make purchases. They might require **targeted re-engagement** strategies, such as special discounts or incentives to increase their activity.

Clustering Evaluation:

1. **Silhouette Score:**

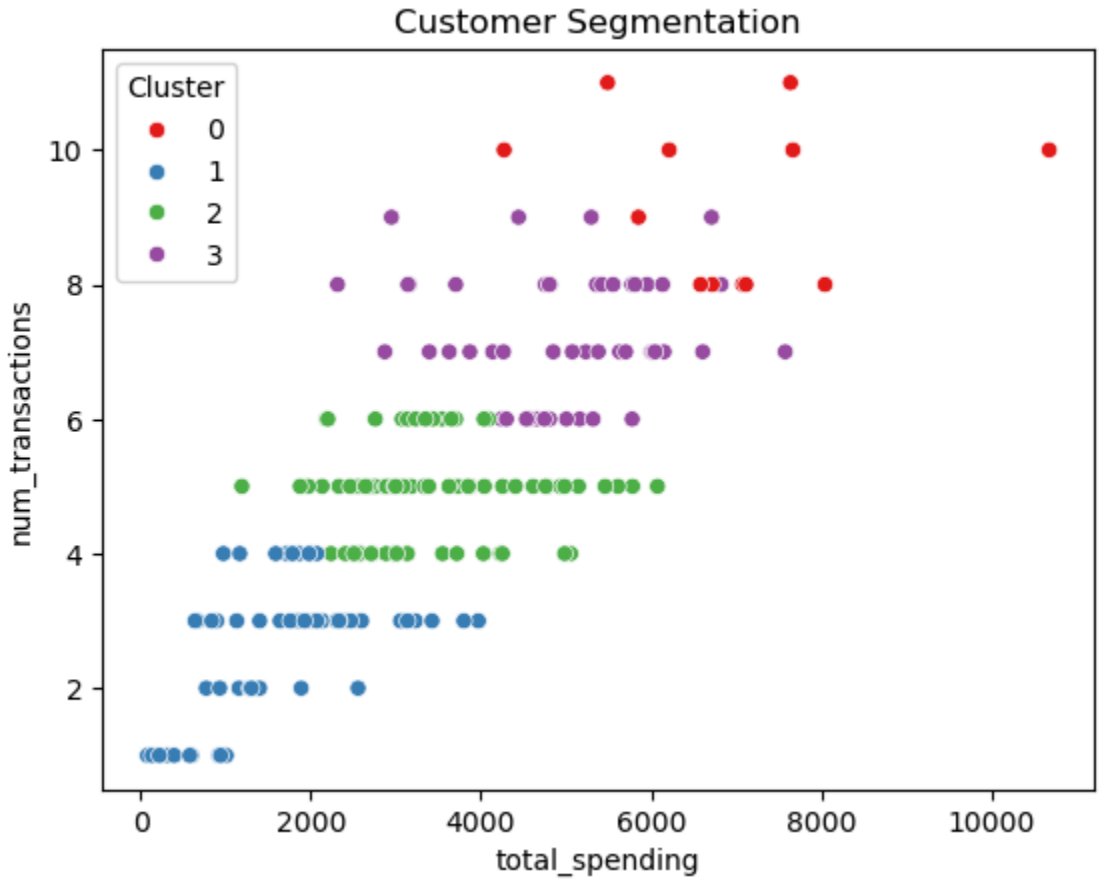
- **Silhouette Score**: The average silhouette score was calculated to be **0.35**, which suggests a reasonable separation between the clusters. A score closer to +1 indicates well-defined clusters.

2. **Davies-Bouldin Index (DBI):**
   - **DBI**: The **Davies-Bouldin Index** for the clustering was calculated to be **1.29**, indicating that the clusters are relatively well separated. A lower DBI indicates better clustering, and our score suggests that the clusters are of good quality but could still be improved for more distinct separation.

Visualization:

The clusters were visualized using **Principal Component Analysis (PCA)** for dimensionality reduction, as the clustering was done on multiple features. Below is a 2D scatter plot showing the separation of the 4 clusters based on the first two principal components:

1-



2-


PCA of Customer Segments
Cluster
0
1
2
3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−3
−2
−1
0
1
2
3
4
5