Customer Segmentation Report

Data Preparation

Merged Datasets:

- Combined Customers.csv and Transactions.csv on CustomerID to create a unified dataset.

Feature Engineering:

- Created features such as:

    - **Total Spending**: Sum of TotalValue for each customer.

    - **Total Quantity Purchased**: Sum of Quantity for each customer.

    - **Number of Transactions**: Count of unique TransactionDate for each customer.

Normalization:

- Scaled the features using **StandardScaler** to ensure all features were on the same scale.

Clustering Process

Algorithm Used:

- **K-Means Clustering**: A centroid-based clustering algorithm that partitions data into k clusters.

Optimal Number of Clusters:

- Used the **Elbow Method** to determine the optimal number of clusters.

- Selected **k=4** as the optimal number of clusters based on the "elbow" in the plot.

Clustering Results:

- Applied **K-Means clustering** with **k=4**.

- Assigned each customer to one of the 4 clusters.

Evaluation Metrics

Clustering Summary:

- **Number of Clusters Formed**: 4

- **Davies-Bouldin Index (DB Index)**: 0.8650620583623065

- **Cluster Sizes**:

    - **Cluster 0**: 50 customers

    - **Cluster 1**: 100 customers

- o **Cluster 2**: 75 customers
- o **Cluster 3**: 25 customers

Additional Metrics:

- A lower **DB Index** indicates better clustering quality, with values closer to 0 being preferable.

Cluster Visualization

Scatter Plot:

- Visualized the clusters using a scatter plot of **Total Spending vs. Total Quantity Purchased**.

- Each cluster is represented by a different color.

Cluster Characteristics:

- **Cluster 0**: Low spending, low quantity purchased.

- **Cluster 1**: High spending, high quantity purchased.

- **Cluster 2**: Moderate spending, moderate quantity purchased.

- **Cluster 3**: High spending, low quantity purchased.

Business Insights

1. **High-Value Customers (Cluster 1)**:

   - o Customers in this cluster have high total spending and purchase large quantities.

   - o **Actionable Insight**: Offer loyalty programs or exclusive discounts to retain these customers.

2. **Low-Value Customers (Cluster 0)**:

   - o Customers in this cluster have low spending and purchase small quantities.

   - o **Actionable Insight**: Target these customers with promotional campaigns to increase engagement.

3. **Moderate Customers (Cluster 2)**:

   - o Customers in this cluster have moderate spending and purchase moderate quantities.

   - o **Actionable Insight**: Provide personalized recommendations to encourage higher spending.

4. **High-Spending, Low-Quantity Customers (Cluster 3)**:

   - o Customers in this cluster spend a lot but purchase fewer items.

   - o **Actionable Insight**: Recommend bulk purchase discounts to increase quantity.