Customer Segmentation and Clustering Report

1. Introduction

This report covers customer segmentation using K-Means clustering on an eCommerce transactions dataset. The objective was to identify distinct customer segments based on their purchase behavior and transaction history, providing actionable insights for business strategies.

2. Data Preparation

The following datasets were merged and preprocessed:

- Customers.csv: Provided customer profile information.
- Transactions.csv: Contained transactional details, including product purchases and transaction values.

Feature Engineering:

The following features were derived for clustering:

- Total Purchase Value: Sum of all transaction values per customer.
- Average Transaction Value: Mean value of transactions per customer.
- Number of Transactions: Count of transactions per customer.
- Unique Product Count: Number of unique products purchased.

All features were normalized using StandardScaler for clustering.

3. Clustering Process

Algorithm Selection:

K-Means was selected due to its efficiency and interpretability for customer segmentation tasks.

Optimal Number of Clusters:

The Elbow Method was used to determine the optimal number of clusters by plotting WCSS (Within-Cluster Sum of Squares) against the number of clusters.

- The optimal number of clusters was identified as 4.

Davies-Bouldin Index:

The clustering quality was evaluated using the Davies-Bouldin Index.

- DB Index: 0.78 (Lower values indicate better clustering performance)

4. Results and Insights

1. Number of Clusters Formed: 4 distinct customer segments.
2. Cluster Characteristics:
    - Cluster 0: High transaction frequency and diverse product purchases.
    - Cluster 1: Moderate transaction value and repeat purchases of specific products.
    - Cluster 2: Low transaction frequency and low average purchase value.
    - Cluster 3: High-value customers with fewer but expensive transactions.

5. Visualization

- A pair plot was generated to visualize the distribution of clusters across features. Distinct groupings validated the effectiveness of the segmentation.

6. Conclusion

The customer segmentation revealed meaningful clusters that can help the eCommerce platform tailor marketing strategies and product recommendations. The K-Means algorithm with four clusters provided actionable insights, and the Davies-Bouldin Index of 0.78 confirmed a reasonable clustering performance.

7. Deliverables:

- **CSV File:** Kiran_B_Clustering.csv
- **Code:** Jupyter Notebook for clustering logic.
- **Visuals:** Elbow Plot and Pair Plot included in the code output.