Clustering Report

1. Objective

The primary goal of clustering is to segment customers or transactions into distinct groups based on similar purchasing behaviors, using features such as Quantity, TotalValue, and Price. These clusters can help businesses understand customer profiles, improve targeting, and optimize marketing strategies.

2. Data Preparation

- **Dataset**: Merged dataset of customers, transactions, and products.
- **Features Used for Clustering**:
    - Quantity: Number of items purchased in a transaction.
    - TotalValue: Total value of each transaction.
    - Price: Unit price of the product.
- **Scaling**: Features were normalized using StandardScaler to ensure equal contribution to clustering.

3. Clustering Technique

- **Algorithm Used**: K-Means Clustering
- **Number of Clusters**: Predefined as **3 clusters** (n_clusters=3).
- **Implementation**: The clustering was performed on the scaled dataset.

4. Clustering Results

- **Number of Clusters Formed**: 3
- **Davies-Bouldin Index**: **0.82**
    - The DB Index evaluates cluster quality, with lower values indicating better-defined clusters. A score of 0.82 suggests that the clusters are compact and well-separated.
- **Cluster Labels**: Each data point was assigned a cluster label (0, 1, or 2).

5. Insights from Clustering

- **Cluster Size Distribution**:
  The clusters varied in size, indicating that customer behaviors are not uniform and distinct purchasing patterns exist.
- **Cluster Characteristics**:
  - Cluster 0: Represents transactions with high prices and moderate quantities. Likely premium purchases.
  - Cluster 1: Includes transactions with moderate prices and quantities, possibly average or regular buyers.
  - Cluster 2: Captures transactions with low prices and high quantities, suggesting bulk or budget purchases.

6. Visualizations

- **Scatter Plot of Clusters**: A 2D plot of features (Quantity, TotalValue) was used to visually inspect cluster separation.
- **Cluster Centers**: Visualized to show typical transaction profiles for each cluster.

7. Business Recommendations

1. **Cluster-Specific Strategies**:
   - For high-value transactions (Cluster 0): Introduce premium loyalty programs or exclusive offers.
   - For regular transactions (Cluster 1): Target with personalized marketing to encourage repeat purchases.
   - For bulk/low-value transactions (Cluster 2): Focus on promotions or bulk discounts to increase volume.
2. **Resource Allocation**: Allocate inventory and marketing resources based on cluster sizes and characteristics.
3. **Future Scope**: Extend clustering with additional features like customer demographics or time-based purchasing trends for deeper segmentation.

8. Limitations

- The clustering is limited to the selected features (Quantity, TotalValue, Price). Adding more dimensions may yield better insights.
- Predefined clusters (3) may not reflect the natural grouping of the data; further experimentation is required to optimize the cluster count.