CUSTOMER SEGMENTATION REPORT

THE OBJECTIVE:

The goal of this project was to group customers into similar categories (clusters) based on their transaction and profile data. These groups can help businesses understand customer behaviour and plan better marketing strategies.

DATA PREPARATION:

- **Files Used**:
    - **Customers.csv:** It Contains the details about customers.
    - **Transactions.csv:** It Contains the records of customer purchases.
    - **Products.csv:** It Contains the information about products.
- **Steps Taken**:
    - First Combined Transactions.csv with Customers.csv using the CustomerID.
    - Then Created two new features:
        - **TotalAmount**: The Total money spent (calculated as Quantity × Price).
        - **Frequency**: The Number of transactions made by each customer.
    - After that Standardized (scaled) these features to prepare for clustering.

CLUSTERING RESULTS:

- **Method Used**: K-Means Clustering
- **Clusters Tested**: 2 to 10 groups.
- **Best Result**: 2 clusters (groups) were found to be the most meaningful based on the **Davies-Bouldin Index** (a measure of how well the clusters are formed).
- **Davies-Bouldin Index for 2 Clusters**: 0.7341 (lower is better).

CLUSTER DETAILS:

The two customer groups have these characteristics:

1. **Cluster 0**:
    - Customers who buy frequently but spend less overall.
    - Likely regular shoppers who make small purchases often.
2. **Cluster 1**:
    - Customers who buy less often but spend more per transaction.

- Likely premium or big-ticket shoppers.

VISUALIZATION:

A scatter plot was created to show the clusters:

- **X-axis**: Total Amount Spent.
- **Y-axis**: Transaction Frequency.
- Each cluster is marked with a different color, showing clear differences between the two groups.

CONCLUSIONS:

1. **The Learnings**
    - Found 2 meaningful customer groups based on their spending and buying habits.
    - **Cluster 0**: Best for discounts and frequent deals to keep them engaged.
    - **Cluster 1**: Best for personalized offers or loyalty rewards to maximize their spending.
2. **Cluster Quality**:
    - The clusters are well-separated, as shown by the **Davies-Bouldin Index (0.7341)**, which means the clustering worked well.