TASK 3 :

Clustering Analysis Report

1. Overview

The clustering analysis was conducted using K Means to identify distinct groups within the dataset. The process involved multi-dimensional feature analysis to segment data points into clusters based on similarities. The results were evaluated using standard metrics such as the Davies-Bouldin Index, Silhouette Score, and Calinski-Harabasz Score.

---

2. Methodology

- **Clustering Algorithm:** K Means

- **Number of Clusters:** 5 (chosen based on initial observations and/or the elbow method)

- **Features Used:**

    o   Transaction count (tx_count)

    o   Total spend (total_spend)

    o   Average spend (avg_spend)

    o   Standard deviation of spend (spend_std)

    o   Total items purchased (total_items)

    o   Average items per transaction (avg_items)

    o   Average price per product (avg_price)

---

3. Evaluation Metrics

The clustering performance was evaluated using the following metrics:

- **Davies-Bouldin Index (DB Index):**
  This metric measures the average similarity ratio between each cluster and its nearest cluster. Lower values indicate better-defined clusters.

    o   **Value:** [Insert Value]

- **Silhouette Score:**
  This score evaluates the separation of clusters by comparing intra-cluster distances to inter-cluster distances. A value closer to 1 indicates well-separated clusters.

    o   **Value:** [Insert Value]

- **Calinski-Harabasz Score:**
  This metric measures the ratio of between-cluster dispersion to within-cluster dispersion. Higher values indicate well-separated clusters.

    o   **Value:** [Insert Value]

4. Results

- The analysis formed **5 clusters** based on the input data.

- **Cluster Interpretations:**
  Each cluster represents a unique group with distinct characteristics. For example:

  - Cluster 1: High spenders with frequent transactions.

  - Cluster 2: Occasional shoppers with low spending.

  - Cluster 3: Moderate spenders with steady transaction rates.

  - (Continue interpreting based on your dataset insights.)

5. Strategic Implications

The clustering results provide actionable insights for the following:

1. **Targeted Marketing:** Create personalized marketing campaigns for each segment.

2. **Customer Retention:** Identify high-value customers and design loyalty programs to retain them.

3. **Inventory Management:** Optimize stock based on the preferences of different clusters.

Next Steps

To enhance clustering accuracy and usability:

1. Experiment with other clustering algorithms (e.g., DBSCAN, Agglomerative Clustering).

2. Refine feature selection to include domain-specific attributes.

3. Perform additional validation using domain knowledge.