Task 3: Customer Segmentation / Clustering

Mantripragada VSP Praneeth

Number of Clusters Formed

- **Optimal Number of Clusters:** The notebook used **5 clusters**, determined explicitly in the code (optimal_clusters = 5). The clustering method employed was K-Means.

Davies-Bouldin Index (DB Index)

- **Value:** 0.9915
- **Interpretation:** A Davies-Bouldin Index close to 0 indicates better clustering performance. The score suggests reasonably compact and well-separated clusters.

Other Relevant Clustering Metrics

1. **Clustering Approach:**
    - **Algorithm:** K-Means clustering was used with an optimal number of 5 clusters.
    - **Preprocessing Steps:**
        - StandardScaler was applied to normalize numerical features, ensuring features like TotalSpent, TotalQuantity, TransactionCount, and AveragePrice had consistent scales.
    - **Similarity Metric:** The clustering relies on Euclidean distance after scaling.
2. **Dimensionality Reduction:**
    - PCA (Principal Component Analysis) was mentioned, potentially used for visualization or dimensionality reduction, though its detailed role is not clear from the extracted data.

3. **Clustering Results Storage:**
    - The final cluster assignments were saved in a file named Clustering_Results.csv, including the CustomerID and respective cluster labels.