Clustering Methodology

The customer segmentation was performed using the **K-Means** clustering algorithm, combining both customer profile data (e.g., region, tenure) and transaction behavior

Feature Engineering:

- **Region**: One-hot encoded geographic regions.
- **Product Categories**: Count of transactions per category.
- **Recency , Frequency , activity duration**

Preprocessing:

- Features normalized using `StandardScaler` to ensure equal weighting.

Optimal Cluster Selection:

- Evaluated clusters for K=2 to 10 using the **Davies-Bouldin Index (DB Index)**.

- Silhouette Score used to validate cluster density and separation.

Clustering Results

1. **Optimal Number of Clusters**:
    - **4 clusters** were selected based on the lowest DB Index.
2. **Evaluation Metrics**:
    - **DB Index**: **1.25** (lower values indicate better separation).
    - **Silhouette Score**: **0.35** (values closer to 1 denote well-separated clusters).
3. **Cluster Visualization**:
    - **PCA (Principal Component Analysis)** reduced features to 2D for visualization (Figure 1).
    - Clusters show distinct groupings, indicating meaningful segmentation.