Task 3 : Customer Segmentation Clustering Analysis Report

- **Overview :**
  The analysis employed K-means clustering to segment customers based on their purchasing behavior and characteristics. The model considered multiple features including total transactions, total quantity purchased, total amount spent, average transaction value, days since signup, and regional indicators.

- **Clustering Algorithm Details :**
  1. **Algorithm**: K-means clustering
  2. **Features:**
     - Total transactions
     - Total quantity
     - Total spent
     - Average transaction value
     - Days since signup
     - Region (one-hot encoded)
  3. **Data Preprocessing:** StandardScaler applied to normalize features
  4. **Random State:** 42 (for reproducibility)

- **Clustering Results :**

  1. **Number of Clusters :**
     The optimal number of clusters was determined to be 5, based on the evaluation of Davies-Bouldin Index and Silhouette Scores across different cluster numbers (k=2 to k=10).

  2. **Key Metrics :**
     -Davies-Bouldin (DB) Index : `1.3547088760955934`
     - Silhouette Score : `0.2549935635490153`

- **Interpretation from Metrics :**
  1. **Davies-Bouldin Index:**
     - The DB Index value of `1.35` indicates moderate cluster separation
     - Lower DB Index values indicate better clustering (range: 0 to ∞)
     - The score suggests that while clusters are distinct, there is some overlap between segments.

2. Silhouette Score :

- The Silhouette Score of `0.25` indicates reasonable cluster cohesion
- Scores range from -1 to 1, with higher values indicating better-defined clusters.
- The positive score suggests that customers within each segment are relatively similar to each other.

- **Additional Observations :**
    1. The elbow curve analysis and DB Index trends were used to determine the optimal number of clusters
    2. The visualization shows distinct customer segments based on total spent vs. total transactions
    3. The standardization of features ensures that all variables contribute equally to the clustering process