Customer Clustering Report

Objective

The objective of this analysis is to segment customers into distinct clusters based on their profile and transaction history. This segmentation helps understand customer behaviour, improve targeting, and enhance business strategies.

Methodology

1. **Data Preparation**:
    - Customer and transaction data were merged to create a unified dataset.
    - Advanced features were engineered from transaction data, such as:
        - Total amount spent (total_spent)
        - Average spending (avg_spent)
        - Total quantity purchased (total_quantity)
        - Number of transactions (transaction_count)
        - Spending frequency, and more.
    - Categorical data (e.g., region) was encoded using One-Hot Encoding.
2. **Feature Scaling**:
    - RobustScaler was used to scale features for better handling of any extreme values and to standardize the feature ranges.
3. **Clustering Approach**:
    - The KMeans algorithm was used to segment customers into clusters.
    - Optimal clusters were determined using a combination of:
        - **Davies-Bouldin Index (DB Index)**
        - **Silhouette Score**
    - PCA was applied for dimensionality reduction to visualize the clusters

Results

Optimal Number of Clusters:

The optimal number of clusters was determined to be **3** based on a multi-metric evaluation.

Clustering Metrics:

- **Davies-Bouldin Index (DB Index)**: **0.3195** (lower is better; indicates compact and well-separated clusters).
- **Silhouette Score**: **0.8984** (closer to 1 indicates well-defined and cohesive clusters).