Customer Segmentation Report

Overview

The customer segmentation analysis was performed using the **KMeans clustering algorithm**, leveraging both numerical and categorical customer profile data. The dataset included features such as total spending, quantity purchased, average product price, most purchased category, and customer region. The clustering results are evaluated using the **Davies-Bouldin Index (DBI)** and additional metrics like cluster sizes and visualization insights.

Clustering Results

1. **Number of Clusters Formed**:
   The clustering algorithm grouped customers into **5 distinct clusters**. These clusters represent segments of customers with similar purchasing behavior and characteristics.

2. **Davies-Bouldin Index (DBI)**:
   The DB Index value for the clustering model is **1.93**.
   - A lower DBI indicates better clustering quality.
   - This value suggests moderately good separation and compactness of clusters.

3. **Cluster Sizes**:
   The sizes of each cluster are as follows:
   - **Cluster 1**: 67 customers (largest cluster)
   - **Cluster 4**: 39 customers
   - **Cluster 3**: 36 customers
   - **Cluster 2**: 33 customers
   - **Cluster 0**: 24 customers (smallest cluster)

Insights from Clustering

1. **Cluster Characteristics**:
   Each cluster represents a unique group of customers with distinct purchasing patterns based on total spending, quantity purchased, product pricing, and preferred categories. For example:
   - Larger clusters (e.g., Cluster 1) might represent general customers with average spending behavior.
   - Smaller clusters (e.g., Cluster 0) might include high-value or niche customers.
2. **Cluster Visualization**:
   - Using **PCA** for dimensionality reduction, the clusters were visualized in a 2D scatter plot.
   - The plot shows clear segmentation of clusters, indicating that the customers exhibit diverse behavioral patterns that were effectively captured by the clustering algorithm.