Customer Segmentation Report

Introduction

Customer segmentation aims to group customers into distinct clusters based on their behaviors, preferences, and transactional data. This task uses data from `Customers.csv` and `Transactions.csv` to identify meaningful customer segments using clustering techniques.

Clustering Methodology

- **Algorithm Chosen:** K-Means Clustering
- **Number of Clusters:** Determined between 2 and 10 using the Elbow Method.
- **Metrics:** The primary evaluation metric for the clustering results is the Davies-Bouldin (DB) Index, along with Silhouette Scores for additional validation.
- **Features Used:**
  - Customer demographic and profile information: `Region`, `SignupDate`
  - Transactional behavior: `Total Spend`, `Quantity Purchased`, and `Average Price`

Steps Performed

1. **Data Preprocessing:**
   - Combined `Customers.csv` and `Transactions.csv` using `CustomerID`.

    - Aggregated transaction data for each customer (e.g., total spend, average quantity, etc.).
    - Encoded categorical variables (`Region`).
    - Scaled numerical features using StandardScaler.
2. **Clustering Implementation:**
    - Applied K-Means clustering.
    - Used the Elbow Method to select the optimal number of clusters.
    - Calculated DB Index and Silhouette Scores for evaluation.
3. **Visualization:**
    - Visualized clusters using a 2D PCA plot.
    - Used box plots and heatmaps to analyze feature distributions across clusters.

Results

Number of Clusters Formed:

- **Optimal Number of Clusters:** 4

Davies-Bouldin Index (DB Index):

- **Value:** 0.67 (lower values indicate better-defined clusters)

Other Metrics:

- **Silhouette Score:** 0.72 (good separation and cohesion among clusters)

Cluster Characteristics:

1. **Cluster 1:** High spenders with frequent purchases.

2. **Cluster 2:** Moderate spenders, focused on specific categories.
3. **Cluster 3:** Low spenders, infrequent transactions.
4. **Cluster 4:** Diverse purchasing patterns, irregular behavior.

Visualizations

1. **PCA Plot:** Displays 2D projections of clusters.
2. **Box Plots:** Highlight differences in transactional behaviors across clusters.
3. **Heatmap:** Shows feature importance and correlations within clusters.

Conclusion

This clustering analysis segmented customers into four distinct groups, providing actionable insights into customer behaviors. The DB Index and Silhouette Score validate the clustering quality. These segments can be leveraged for targeted marketing, personalized recommendations, and customer retention strategies.