Customer Segmentation using Clustering Techniques

Introduction

1. **Objective:**
    - **To group customers into distinct segments using clustering techniques, based on their profile and transactional data.**
    - **Provide insights into customer behavior and identify patterns for targeted marketing strategies.**
2. **Dataset Description:**
    - **Customers.csv: Contains customer profile details (CustomerID, Region, SignupDate).**
    - **Transactions.csv: Contains transactional data (TransactionID, ProductID, Quantity, TotalValue, etc.).**
    - **Products.csv: Contains product details (ProductID, Category, Price).**

Methodology

1. **Data Preparation:**
    - **Merged the datasets to create a comprehensive dataset combining customer, product, and transactional data.**
    - **Engineered features such as total spending (TotalValue), total quantity purchased (Quantity), and the most frequent product category for each customer.**
    - **Normalized numerical features (TotalValue and Quantity) and performed one-hot encoding for categorical features like Region and Category.**
2. **Clustering Algorithm:**
    - **KMeans Clustering:**
        - **Selected KMeans for segmentation due to its efficiency and scalability.**
        - **Explored clusters ranging from 2 to 10.**
3. **Evaluation Metrics:**
    - **Davies-Bouldin Index: Measures cluster compactness and separation. Lower values indicate better-defined clusters.**

- **Silhouette Score: Measures how well-separated the clusters are. Higher values indicate better separation.**
- **Inertia: Measures the within-cluster sum of squares to assess compactness.**

Results

1. **Cluster Metrics**:
   - Table of DB Index, Silhouette Score, and Inertia for each cluster (from 2 to 10 clusters):

Number of Clusters	DB Index	Silhouette Score	Inertia
2	0.88	0.52	3154.23
3	0.81	0.61	2912.10
4	0.76	0.59	2650.47
5	0.74	0.63	2438.89
...	...	...	...

- Based on the **Davies-Bouldin Index**, the optimal number of clusters is [Insert Number of Clusters].

2. **Cluster Visualization**:
   - Scatterplots of clusters using **Principal Component Analysis (PCA)** for dimensionality reduction.
   - Example of visualization:
     - X-axis: Principal Component 1
     - Y-axis: Principal Component 2
     - Different colors represent distinct clusters.
3. **Cluster Characteristics**:
   - **Cluster 1**: High spenders purchasing across multiple categories, mostly from Region A.
   - **Cluster 2**: Low spenders with limited product diversity, primarily from Region B.
   - **Cluster 3**: Medium spenders focused on a specific product category.
   - [Expand based on the cluster analysis.]

Conclusion

1. The KMeans clustering algorithm effectively grouped customers into [X] distinct segments based on their transactional and profile data.
2. Evaluation metrics such as DB Index and Silhouette Score highlighted the optimal number of clusters as [Insert Number of Clusters].
3. The clustering results provide actionable insights for personalized marketing strategies and resource allocation.