TASK-3

Customer Segmentation / Clustering

Introduction:

The objective of Task 3 is to perform **customer segmentation** using clustering techniques. This involves grouping customers based on their profile attributes (e.g., region, signup date) and transaction behavior (e.g., total spending, purchase frequency). By segmenting customers, the company can gain a deeper understanding of customer behavior, enabling targeted marketing strategies, personalized offers, and better resource allocation.

The task requires selecting a clustering algorithm, determining the optimal number of clusters (between 2 and 10), calculating clustering metrics (including the Davies-Bouldin Index), and visualizing the clusters to ensure meaningful segmentation.

Steps for Clustering :

1. **Data Preprocessing:**
    - Merge Customers.csv and Transactions.csv on CustomerID to create a unified dataset.
    - Clean and handle missing data, if any.
    - Standardize or normalize numeric features (e.g., TotalValue, Quantity) for uniform scaling.
    - Encode categorical variables (e.g., Region, ProductCategory) using one-hot encoding or label encoding.
2. **Feature Selection:**
    - Select features that influence customer behavior, such as:
        - Total spending (TotalValue)
        - Transaction frequency
        - Preferred product categories
        - Region or location
    - Reduce dimensionality if needed (e.g., using PCA) to enhance clustering performance.

3. **Clustering Algorithm:**
    - Choose a clustering algorithm based on the data's structure and properties:
        - **K-Means Clustering**: Popular for simplicity and speed.
        - **DBSCAN**: Suitable for identifying clusters of varying density.
        - **Agglomerative Clustering**: Useful for hierarchical cluster formation.
    - Experiment with different cluster counts (2 to 10) to identify the optimal number of clusters.
4. **Clustering Metrics:**
    - Calculate the **Davies-Bouldin Index (DB Index)** to evaluate cluster compactness and separation.
    - Additional metrics like **Silhouette Score** and **Within-Cluster Sum of Squares (WCSS)** can also be calculated for validation.
5. **Visualization:**
    - Visualize clusters using scatter plots, pair plots, or 3D plots to observe group separations.
    - Use color coding to differentiate clusters and ensure interpretability.

Conclusion:

By segmenting customers into distinct clusters, businesses can optimize marketing efforts, improve customer engagement, and enhance overall profitability. This task leverages data-driven clustering techniques to generate actionable insights, providing a foundation for effective decision-making and strategy development.