Customer Segmentation using K-Means Clustering

**Code created by:S.Govind Sai Charan**

This project performs customer segmentation using **K-Means clustering** in Python. It includes data preprocessing, clustering analysis, visualization, and a GUI for interactive insights.

**Features**

✅ **Synthetic Data Generation** – Creates a dataset with customer behaviors like money\_spent, products\_purchased, and complains.  
✅ **Preprocessing & Standardization** – Scales data for improved clustering.  
✅ **K-Means Clustering** – Determines optimal clusters using **Elbow Method** & **Silhouette Score**.  
✅ **Cluster Labeling** – Categorizes customers based on spending, purchasing habits, and satisfaction.  
✅ **Data Visualization** – Generates elbow plots, PCA-based cluster visualizations, and box plots.  
✅ **Interactive GUI** – Built with Tkinter to display segmentation results and marketing recommendations.

**1. Data Generation**

* The script generates synthetic customer data with attributes:
  + customer\_id
  + products\_purchased
  + complains
  + money\_spent
* The data is randomly generated using log-normal and binomial distributions to simulate real-world customer behaviors.

**2. Data Preprocessing**

* The features (products\_purchased, complains, money\_spent) are standardized using StandardScaler for better clustering performance.

**3. K-Means Clustering**

* The script determines the optimal number of clusters using:
  + **Elbow Method** (Inertia values)
  + **Silhouette Score**
* It selects an optimal number of clusters (default is 4) and assigns each customer to a cluster.

**4. Cluster Naming**

* The script labels each cluster based on:
  + Spending behavior (High-Spending, Medium-Spending, Low-Spending)
  + Purchasing habits (Diverse-Buyer, Multi-Buyer, Single-Item)
  + Customer satisfaction (Satisfied, Dissatisfied)

**5. Visualization**

* The script generates plots to visualize the clustering results:
  + **Elbow method & silhouette scores** to determine the best number of clusters.
  + **PCA scatter plot** to reduce dimensionality and visualize clusters.
  + **Box plots** showing how features vary across clusters.
  + **Cluster sizes** to understand distribution.

**6. GUI Interface**

* A Tkinter GUI is built to provide:
  + A summary of customer segmentation.
  + Detailed cluster insights.
  + Graphical visualizations.
  + Marketing recommendations based on customer segmentation.

**7. Marketing Recommendations**

* Based on cluster characteristics, the script suggests marketing strategies:
  + Loyalty programs for high-spending customers.
  + Upselling & cross-selling for different purchase behaviors.
  + Special discounts for low-spending customers.
  + Customer service improvements for dissatisfied customers.