# Customer Segmentation with Clustering

**GitHub Repository:** https://github.com/dsuarez99/Customer-Segmentation

## Business Problem Being Solved

The objective of this project will be to perform clustering analysis to divide customers into segments and help marketing teams prioritize high-potential customers and optimize campaign targeting strategies, ultimately improving return on marketing investment.

## Methodology and Approach

1. Data Source:

* This assignment was developed using the Customer Personality Analysis dataset obtained from Kaggle (<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>). This dataset contains information about customers, including attributes such as age, income, education, marital status, number of children, and their spending across various product categories. It also includes information on their response to past marketing campaigns.

2. Data Preparation:

* Cleaned missing values and derived new variables.
* Encoded categorical variables using one-hot encoding.
* Removed highly correlated or redundant features.
* Standardized numerical features for clustering.
* Applied PCA to reduce multicollinearity and capture 90% of the variance in 6 components.

3. Modeling:

* K-Means Clustering: Selected 4 clusters based on the Elbow Method and silhouette scores.
* DBSCAN Clustering: Tested as an alternative for density-based clustering with outlier handling.

4. Evaluation:

* Compared clustering performance using silhouette score and business interpretability.

## Key Findings and Insights

- K-Means Clustering produced four meaningful segments:

* Cluster 0: High-income, high-spending, responsive to campaigns.
* Cluster 1: Low-income families with high digital activity and low responsiveness.
* Cluster 2: Older customers, moderate income, digitally active.
* Cluster 3: Large households with kids, low to moderate engagement.

- DBSCAN Clustering offered a simpler, strategic split:

* Cluster 0: Budget-conscious families (low income, many children, high online activity).
* Cluster 1: High-value, older customers (high income, high spending, low digital activity).

## Instructions for Running the Code

1. Clone the repository from GitHub

2. Install required libraries

3. Open the Jupyter Notebook

4. Run the notebook