**CUSTOMER SEGMENTATION USING DATA SCIENCE**

**Phase 5:** Project Documentation & Submission

**Problem Definition:**

The key aim of this project is to configure marketing strategies and improve customer satisfaction via data science methodologies, resulting in an array of benefits such as enhanced customer engagement, higher conversion rates, customer loyalty and contentment, increased revenue, and so on. This segmentation empowers organizations to manipulate their marketing tactics and services to distinct consumer segments, thus enhancing patron satisfaction and driving business prosperity.

Dataset Link:  <https://www.kaggle.com/datasets/akram24/mall-customers>

**Data Preprocessing:**

**Data Collection and Data Cleaning:**

* + Here we are employed with the process of gathering the dataset that includes CustomerID, Genre, Age, Spending Score(1-100), and Income per Age for each customers.
  + This process is followed by imputing and removing incompletes records.
  + As we didn’t find any of the missing value in the dataset, no necessary action is needed to perform.

**Data Transformation:**

* + Here we haven’t performed any Encode categorical variables for "Genre" (e.g., one-hot encoding or label encoding). If it is necessary then we can perform the following operation.
  + To bring the attributes like "Age," "Spending Score," and "Income per Age" to a common scale we perform Normalization and standardization.

**Data Splitting:**

Here we divide the dataset into training and testing subsets by importing the necessary python libraries, such as scikit-learn, pandas etc to evaluate the model's performance.

**Featured Extraction Technique:**

The above process is proceeding by performing feature engineering to create new features and to modify existing ones to improve segmentation. Some feature engineering ideas used in this project are:

* Age group
* Income per age group
* Customer LifeTime Value (CLV)
* Income Quartiles
* And so on.

**Machine Learning Algorithm, Model Training and Evaluation:**

As we have performed the necessary data data preprocessing steps, including data cleaning, feature engineering, normalization or standardization, Now we can proceed with the machine learning algorithms. Some of the machine learning algorithm that are performed on the dataset are:

* Principal Component Analysis (PCA)
* Linear Discriminant Analysis
* Hiererchical Clustering
* Gaussian Mixture Model (GMM)
* K-Means Clustering