## Fundamentals of ML - Lab 3

**OBJECTIVE**: To perform exploratory data analysis to understand the nature of relation between the different features in the given data. Based on the inferences, perform further clustering to identify clusters in the data.

**DATA**: You are given data of customers visiting a store. You have some basic information about the customers through memberships. Each customer is also assigned a spending score. Spending Score is something you assign to the customer based on defined parameters like customer behavior and purchasing data.

### Step 1: Load the data

Download data given to you in csv format. Load it into pandas.

### Step 2: Data Exploration and Visualisation

* Create plots to understand the distribution of the each feature in the data
* Create plots to understand the distribution of data with respect to the other features

### Step 3: Feature Preprocessing

* Fill any missing values in the data
* Identify any outliers and remove them if necessary
* Perform feature scaling if necessary
* Encode categorical features into numerical data - if any categorical features.

Now your data is ready to be used for clustering

### Step 4: Perform clustering using DBSCAN

* Range of epsilons : (8,13). Range of min\_samples: (3,9)
* For the above given ranges of hyperparameters, fit DBSCAN models to identify the best epsilon and min\_samples value.
* You can use silhouette scores to identify the optimal value. Plot a heatmap of epsilons, min\_samples and silhouette scores to identify optimal hyperparameters
* Using the optimal hyperparameters, fit a DBSCAN model on your data.

### 

### Step 5: Visualize clusters

* Visualize the clusters formed by plotting points in 2D for ‘silhouette score’ and ‘annual income’
* Visualize clusters formed by plotting points in 3D for ‘silhouette score’ and ‘annual income’ and ‘age’