## Documentation Evaluation Parameters (15 points)

The following need to be completed as a separate document (pdf or doc format) for each project, in addition to the code (py or ipynb format) :

1. Explanation of your understanding of the data based on EDA, in the form of both text and visualizations (4 points).
2. Explanation of Data Wrangling and Feature Engineering steps used, if any. If not used, explain why (2 points).
3. Explanation of the choice of candidate algorithms - min 2 algorithms (2 points).
4. Explanation of model training including hyper-parameter tuning using grid search, evaluation using cross validation and selection (4 points).
   1. For hyper-parameter tuning, explain the range of values for each parameter.
   2. Use both text and visualizations to summarize model evaluation for all the models. Use the following metrics.:
      1. BAC, REC and AUC for classification
      2. RMSE and R2 for regression
5. Explanation of innovation used in the project (3 points).

## Code Evaluation Parameters (35 points)

1. Coding Style: PEP 8 to be followed (10 points).
2. Efficiency of solution - towards more optimal solution (5 points).
3. Correctness / accuracy of solution (15 points).
4. Innovation (5 points).

Project 2

### Problem Description

A retail company wants to understand the customer purchase behaviour against various products of different categories. They have shared the purchase summary of various customers for selected high volume products from last month.

The data set also contains customer demographics (age, gender, marital status, city\_type, stay\_in\_current\_city), product details (product\_id and product category) and Total purchase\_amount from last month.

They want you to build a model to predict the purchase amount of customers against various products which will help them to create personalized offers for customers against different products.

The output schema required is as follows (in a csv file):

User\_ID, Product\_ID, Purchase\_Amount

### 

### Dataset

1 Training Dataset and 1 Testing Dataset