Hands on Assessment - **Supervise Learning**

• **DOMAIN***:* Banking, Marketing

• **CONTEXT***:* A bank X is on a massive digital transformation for all its departments. Bank has a growing customer base where majority of them are liability customers (depositors) vs borrowers (asset customers). The bank is interested in expanding the borrowers base rapidly to bring in more business via loan interests. A campaign that the bank ran in last quarter showed an average single digit conversion rate. Digital transformation being the core strength of the business strategy, marketing department wants to devise effective campaigns with better target marketing to increase the conversion ratio to double digit with same budget as per last campaign.

• **DATA DICTIONARY:**
1. Team: Team's name
2. ID: Customer ID
3. Age: Customer's approximate age.
4. CustomerSince: Customer of the bank since. [unit is masked]
5. HighestSpend: Customer's highest spend so far in one transaction. [unit is masked]
6. ZipCode: Customer's zip code.
7. HiddenScore: A score associated to the customer which is masked by the bank as an IP.
8. MonthlyAverageSpend: Customer's monthly average spend so far. [unit is masked]
9. Level: A level associated to the customer which is masked by the bank as an IP.
10. Mortgage: Customer's mortgage. [unit is masked]
11. Security: Customer's security asset with the bank. [unit is masked]
12. FixedDepositAccount: Customer's fixed deposit account with the bank. [unit is masked]
13. InternetBanking: if the customer uses internet banking.
14. CreditCard: if the customer uses bank's credit card.
15. LoanOnCard: if the customer has a loan on credit card.

• **PROJECT OBJECTIVE***:* Build a Machine Learning model to perform focused marketing by predicting the potential customers who will convert using the historical dataset.

• **STEPS AND TASK:**
1. Data Understanding and Preparation: (A : 5M, B : 5M, C : 5M, D : 5M) = 20 Marks
     A. Read both the Datasets 'Data1' and 'Data 2' as DataFrame and store them into two separate variables.
     B. Print shape and Column Names and DataTypes of both the Dataframes.
     C. Merge both the Dataframes on 'ID' feature to form a single DataFrame
     D. Change Datatype of below features to 'Object' 'CreditCard', 'InternetBanking', 'FixedDepositAccount', 'Security', 'Level', 'HiddenScore'.
     [Reason behind performing this operation: - Values in these features are binary i.e. 1/0. But Datatype is 'int'/'float' which is not expected.]
2. Data Exploration and Analysis: (A : 10M, B : 5M, C : 10M) = 25 Marks
     A. Visualize distribution of Target variable 'LoanOnCard' (5Marks) and clearly share insights (5Marks)
     B. Check the percentage of missing values and impute if required
     C. Check for unexpected values in each categorical variable and impute with best suitable value
3. Data Preparation and model building: (A : 5M, B : 5M, C : 5M, D : 10M) = 25 Marks
     A. Split data into X and Y
          [Recommended to drop ID & ZipCode. LoanOnCard is target Variable]
     B. Split data into train and test. Keep 25% data reserved for testing
     C. Train a Supervised Learning Classification base model - Logistic Regression
     D. Print evaluation metrics for the model and clearly share insights