# Lead Score - Case Study - Summary Report

## Problem Statement:

X Education, an e-learning platform for professionals, is creating a logistic regression model to assign lead scores (0-100) for effective lead targeting. Higher scores signify better conversion chances (hot leads), while lower scores imply lower conversion likelihood.

The goal is to achieve an 80% conversion rate, aligning with the CEO's vision for sustainable business growth through valuable e-learning experiences.

## Summary:

### Step 1: Data Exploration :

Begin by delving into and thoroughly comprehending the dataset, followed by a detailed analysis.

#### Step 1.1: Data Cleaning :

- Checked for duplicate values.
- Replaced the **'Select'** placeholder with NaN values.
- Dropped columns with over **40% null values**.
- Imputed missing values for numerical columns.
- Transformed categorical variables by creating new classification variables.

### Step 2: Exploratory Data Analysis :

- Utilized box plots to detect outliers in numerical columns.
- Applied capping techniques up to the **95th** percentile to manage outliers.
- Conducted an analysis of both categorical and numerical columns related to the target variable **'Converted,'** resulting in valuable insights and findings.

### Step 3: Linear Regression :

#### Step 3.1: Creating dummy variables :

- We proceeded by generating synthetic data for the categorical variables.
- Dropped the columns for which we have created dummy variables

#### Step 3.2: Splitting data into 'Test' and 'Train' set :

- - Divided the dataset into training and testing subsets, following a **70-30%** ratio.

### Step 3.3: Rescaling using MinMax :

- Implemented Min-Max Scaling for numerical variables.
- Utilized the stats model to construct an initial model, providing a thorough statistical summary of model parameters.
- 

### Step 3.4: Feature Selection - RFE :

- Utilized Recursive Feature Elimination to select the top **20** important features, iteratively examining P-values to retain significant ones.

## Step 4: Model Building :

### Step 4.1: Constructing the models :

- Our initial model incorporated all **RFE**-identified columns. We improved it by excluding columns with high **P-values** and **VIF** values, resulting in a final model with **13** highly significant features and **VIF** values below **5**.

## Step 5: Model Evaluation :

### Step 5.1: Evaluating the final model :

- Produced forecasts for the training dataset.
- Formed a DataFrame pairing actual conversion outcomes with predicted probabilities.
- Employed an initial threshold probability of **0.5** for predicted labels.
- Constructed a **confusion matrix** and generated a **classification report**.
- Computed supplementary metrics such as **'Accuracy,' 'Sensitivity,' 'Specificity,' 'False Positive Rate,' 'Positive Predictive Value (Precision),' and 'Negative Predictive Value'** to evaluate model reliability.

### Step 5.2: Plotting 'ROC' Curve :

- Plotted the ROC curve for the features, resulting in a strong 88% area under the curve, enhancing the model's reliability.

### Step 5.3: Determining the Ideal Threshold :

- Plotted **'Accuracy,' 'Sensitivity,' and 'Specificity'** for different probabilities, identifying the optimal cutoff at **0.35** where these curves intersected. After this adjustment, about **81%** of values were predicted accurately, resulting in updated metrics: **'Accuracy = 81%', 'Sensitivity = 80.8%', 'Specificity = 80.4%'.**

### Step 5.4: Determining the 'Precision' & 'Recall' :

- We computed Precision and Recall metrics, yielding **79%** and **70%** on the training dataset. Balancing **Precision** and **Recall**, we determined an **optimal cutoff value** of around **0.42.**

## Step 6: Making Predictions on the 'Test' data set :

Next, we applied these insights to the test model, achieving an **80% Accuracy** with approximately **80%** for both **Sensitivity** and **Specificity**.

## Step 7: Conclusion :

In the conclusion, we identified **'Prospect ID'** records with lead scores exceeding 85, totaling **360** records. We also provided recommendations for the company/CEO.