# Lead Scoring Case Study

Predicting Potential Customers for X Education

## Problem Statement

Challenge:  
- X Education's current lead conversion rate is approximately 30%.  
Objective:  
- Develop a logistic regression model to assign a lead score (0-100) to each lead, indicating their likelihood of conversion.

## Data Overview

Dataset:  
- ~9,000 leads with attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc.  
Target Variable:  
- 'Converted' (1: Lead converted; 0: Lead not converted)

## Data Preparation

Steps Taken:  
- Handled missing values and categorical variables.  
- Performed exploratory data analysis to understand variable distributions.  
- Created dummy variables for categorical data.  
- Scaled numerical features for model compatibility.

## Modeling Approach

Model Used:  
- Logistic Regression  
Feature Selection:  
- Recursive Feature Elimination (RFE) to identify significant predictors.  
Model Evaluation Metrics:  
- Accuracy, Sensitivity (Recall), Specificity, Precision, ROC-AUC Curve

## Model Performance

Training Set Results:  
- Accuracy: 79.5%  
- Sensitivity: 80.2%  
- Specificity: 78.9%  
Test Set Results:  
- Accuracy: 78.8%  
- Sensitivity: 79.5%  
- Specificity: 78.2%

## Lead Scoring Implementation

Lead Score Calculation:  
- Model assigns a probability score (0-1) to each lead, multiplied by 100 to get a lead score between 0 and 100.  
Interpretation:  
- Higher scores indicate a greater likelihood of conversion, enabling the sales team to prioritize efforts.

## Business Impact

Improved Efficiency:  
- By focusing on leads with higher scores, the sales team can allocate resources more effectively.  
Potential Increase in Conversion Rate:  
- Targeting 'hot' leads could elevate the conversion rate towards the desired 80%.

## Conclusion

Summary:  
- The logistic regression model provides a systematic approach to scoring leads, assisting X Education in identifying and prioritizing potential customers.  
Next Steps:  
- Implement the model in the sales process and monitor performance to make data-driven adjustments.