



# Lead Scoring Case Study

**By: Bidisha Bose Mukherjee**



# Problem Statement:

- An Education Company named X Education, sell online coursed to industry Professionals.
- Now, Although X Education gets a lots of leads, its lead conversion rate is very poor, only 30%
- Now the Company wants to increase the conversion Rate and wants to know how to increase the 'hot leads.



# Goals of the Case Study:

- Build a logistic Regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
- A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted
- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.



# Approach

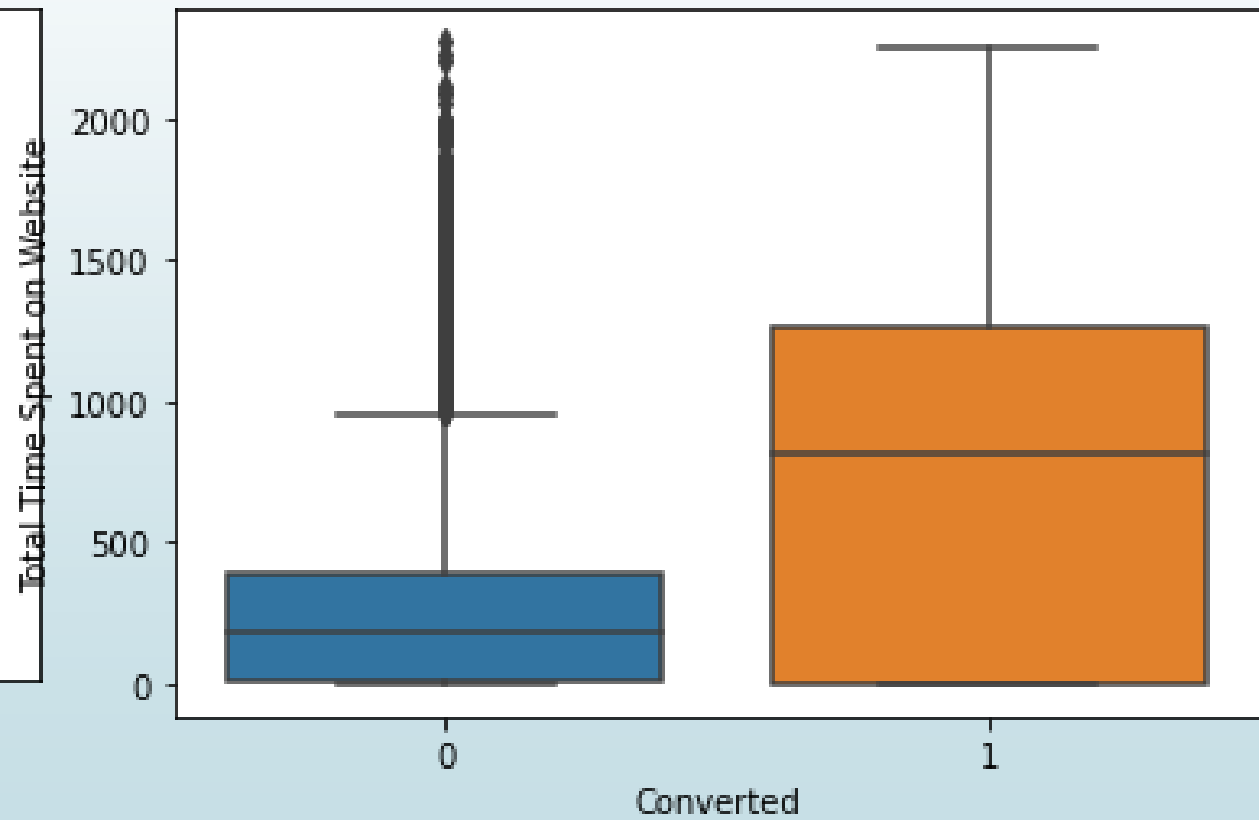
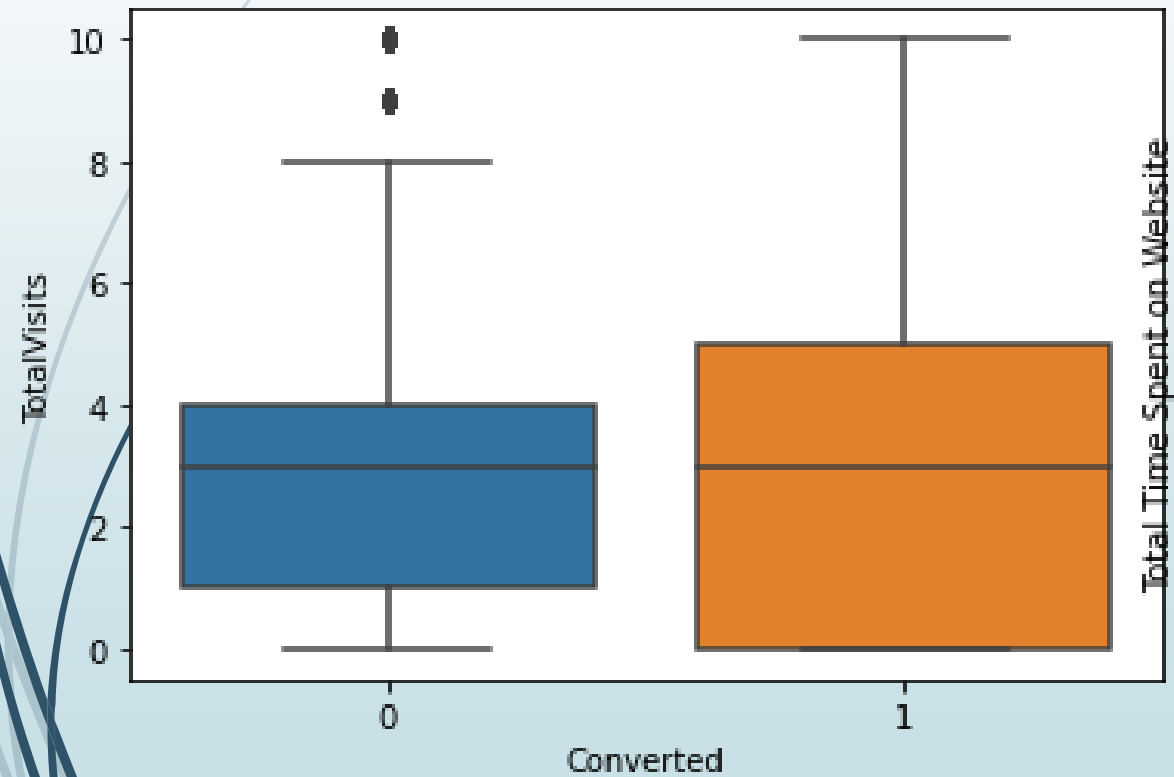
- Source the data For Analysis
- Reading & Understanding the data
- Data Cleaning
- EDA
- Feature Scalling
- Splitting the data into test & train dataset
- Prepare the data for Modeling
- Model Building
- Model Evaluation-specificity & sensitivity or precision recall
- Making Predictions on the test set



# Data Sourcing, Cleaning and Preparation:

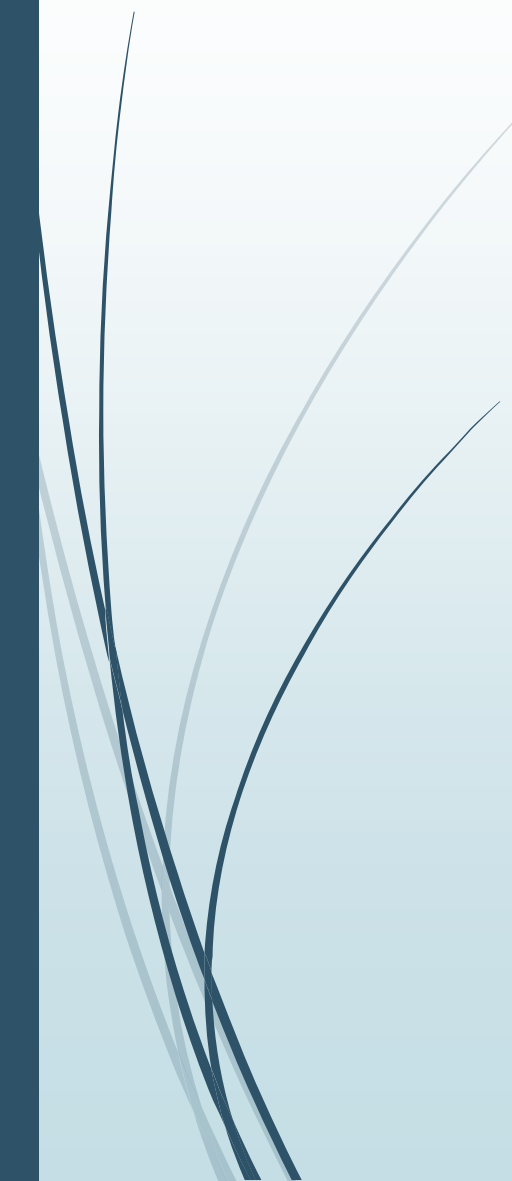
- Read the CSV file
- Outlier Treatment
- Data Cleaning, Handling null values and removing higher null values data
- Removing Redundant columns in the data
- Imputing Null Values
- Exploratory data analysis-approx. conversion rate is 38%
- Feature Standardization

# Outliers:

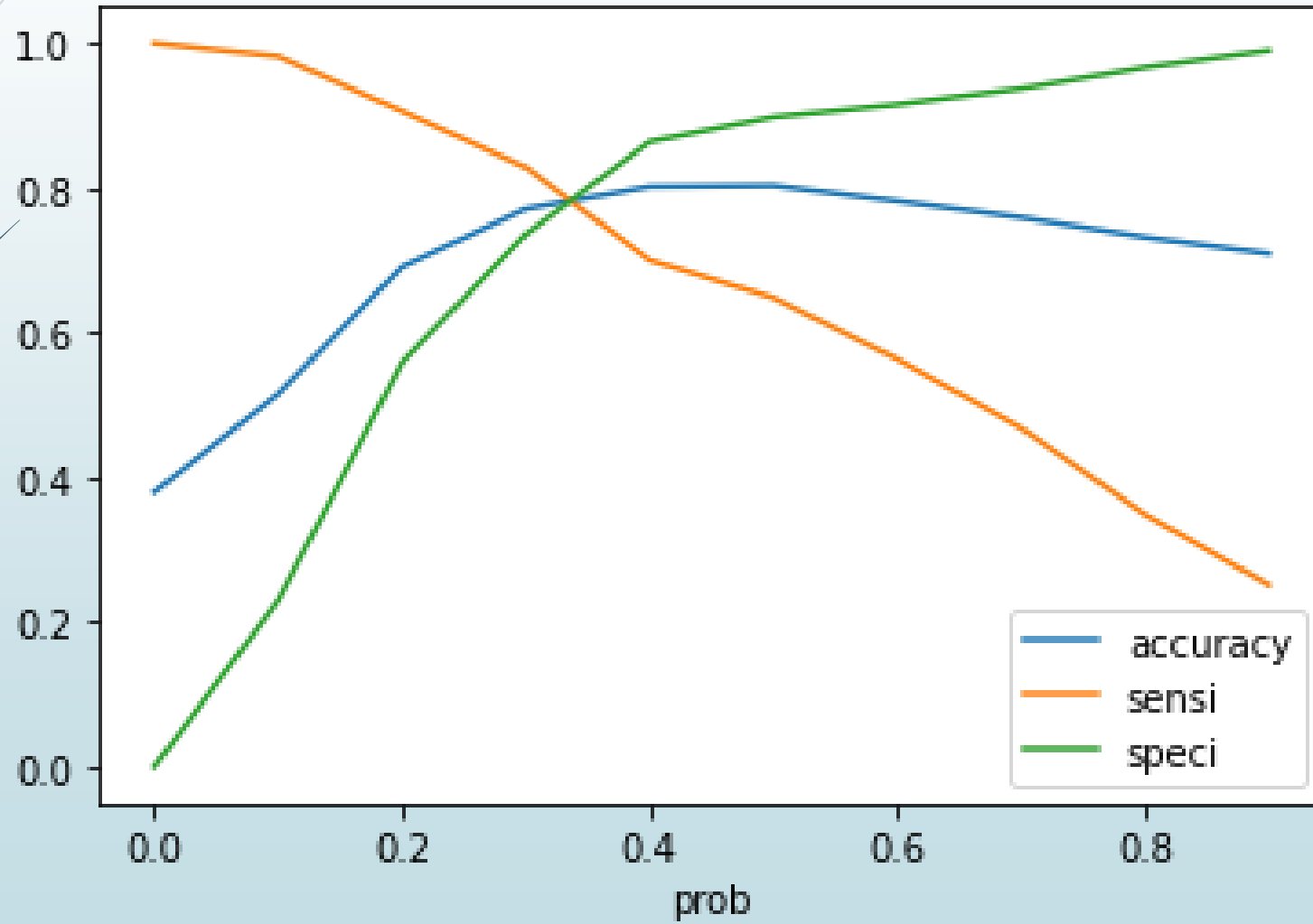




# Features Scaling & Splitting Train & Test Sets:

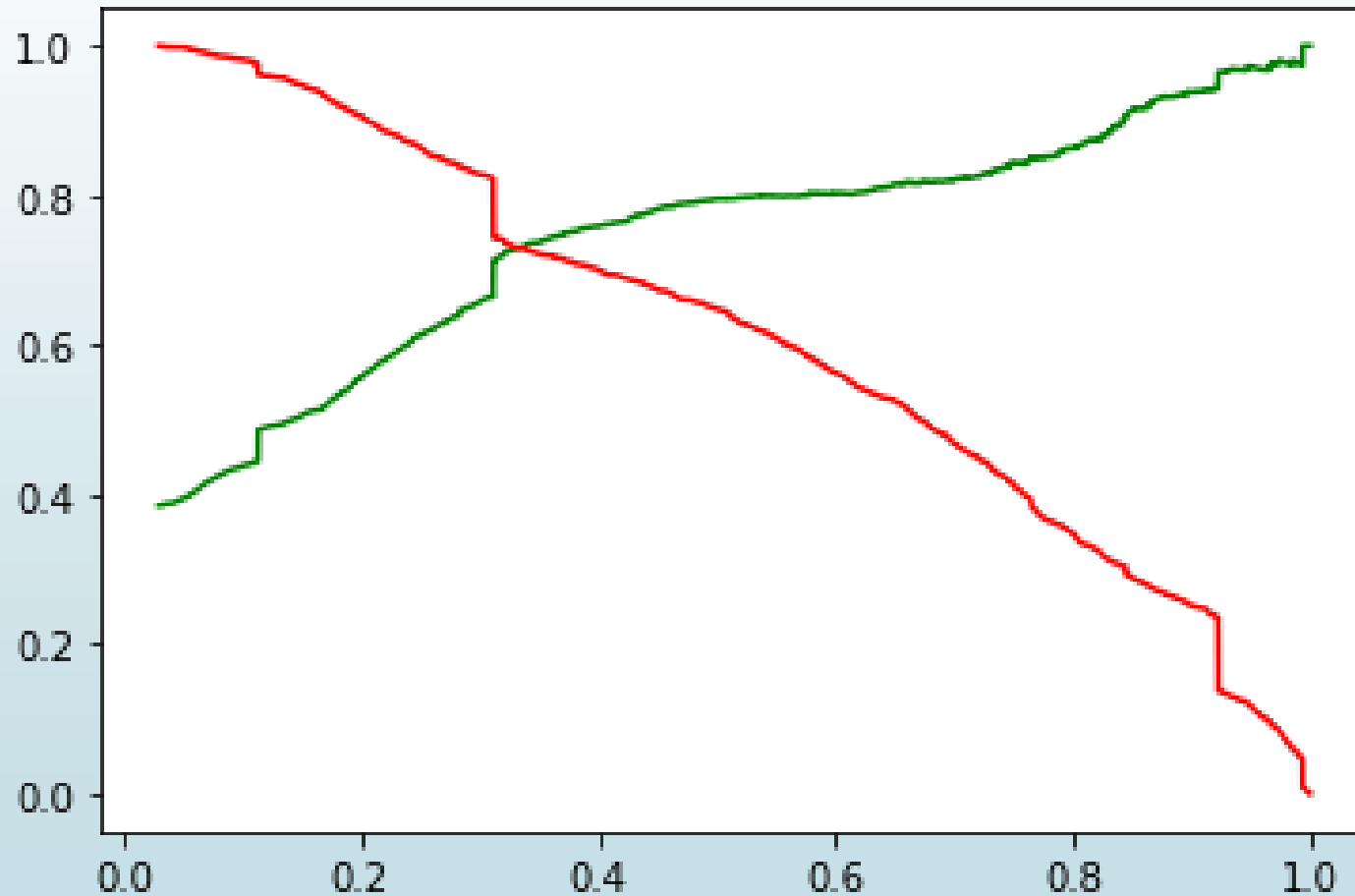
- Feature scaling of Numeric Data
  - Splitting data into Train & Test set
- 

# Model Evaluation – Sensitivity & Specificity on Train Data Set



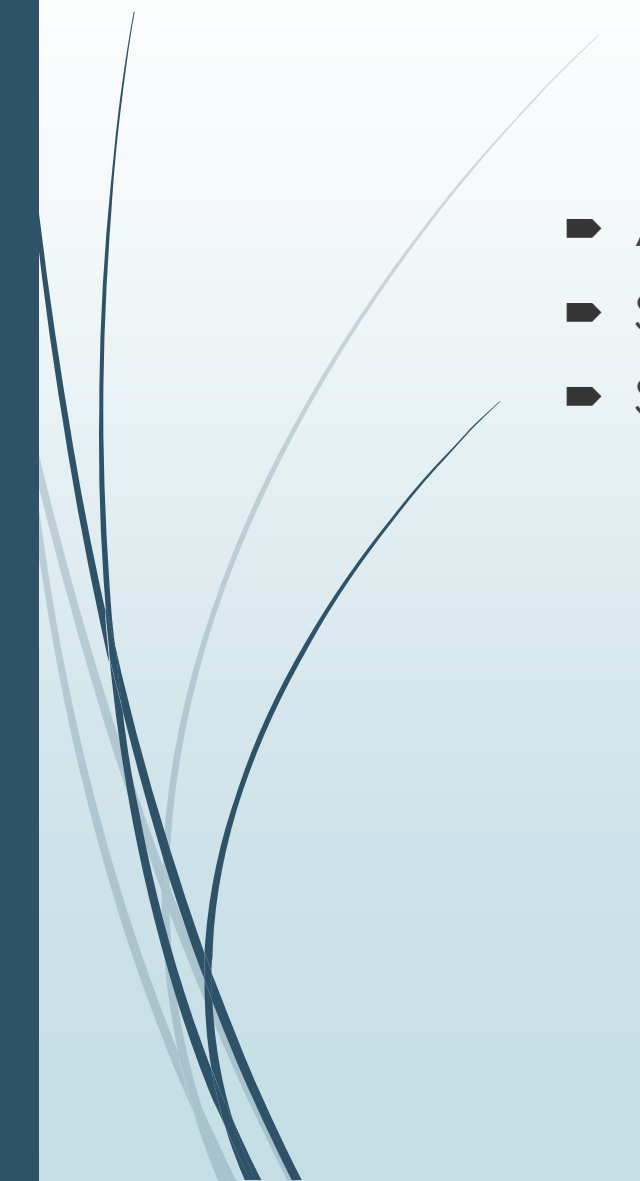


# Model Evaluation(Precision & Recall on Train Dataset)





# Model Evaluation

- Accuracy: 77.52%
  - Sensitivity : 83.01%
  - Specificity: 74.13%
- 



# Conclusion:

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 77%, 83% and 74% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%
- Hence overall this model seems to be good.