# LEAD SCORING – CASE STUDY
# LOGISTIC REGRESSION

-VISHESH SHUKLA

# Problem Statement

- X Education is an organization which provides online courses for industry. The company marks its courses on several popular websites like google.
- X Education wants to select most promising leads that can be converted to paying customers.
- Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion. Leads come through numerous modes like email, advertisements on websites, google searches etc.
- The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversations.

# Business Goal

- The company requires a model to be built for selecting most promising leads.
- Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted, the lower it is lesser the chances of conversion.
- The model to be built in lead conversion rate around 80% or more.

# Strategy

- Import data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Feature Scaling
- Splitting the data into Test and Train dataset
- Build a logistic regression model
- Assign a lead score for each leads
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

# Problem solving methodology

## Data Sourcing , Cleaning and Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.

## Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set.

## Model Building

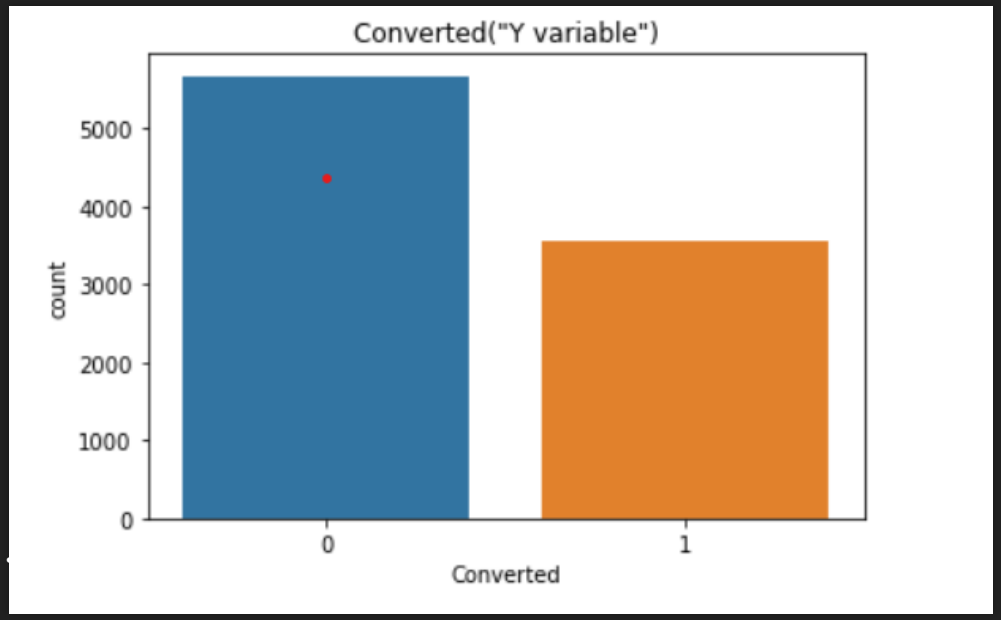
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

## Result

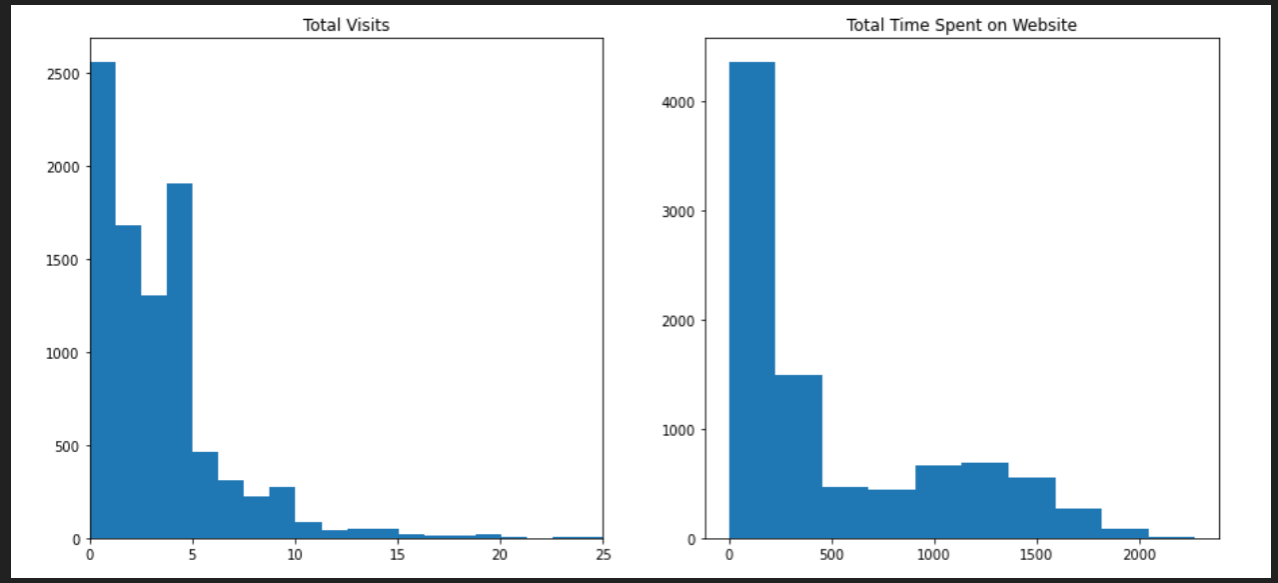
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

# Exploratory Data Analysis

We have the following Conversion Rate in total

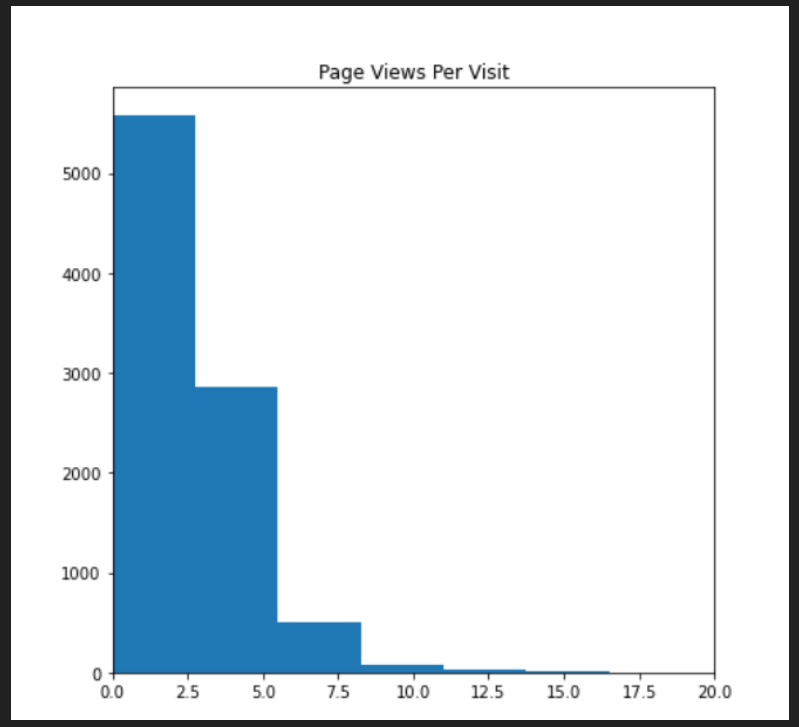



Plots for Total Visits, Total Time Spent on website

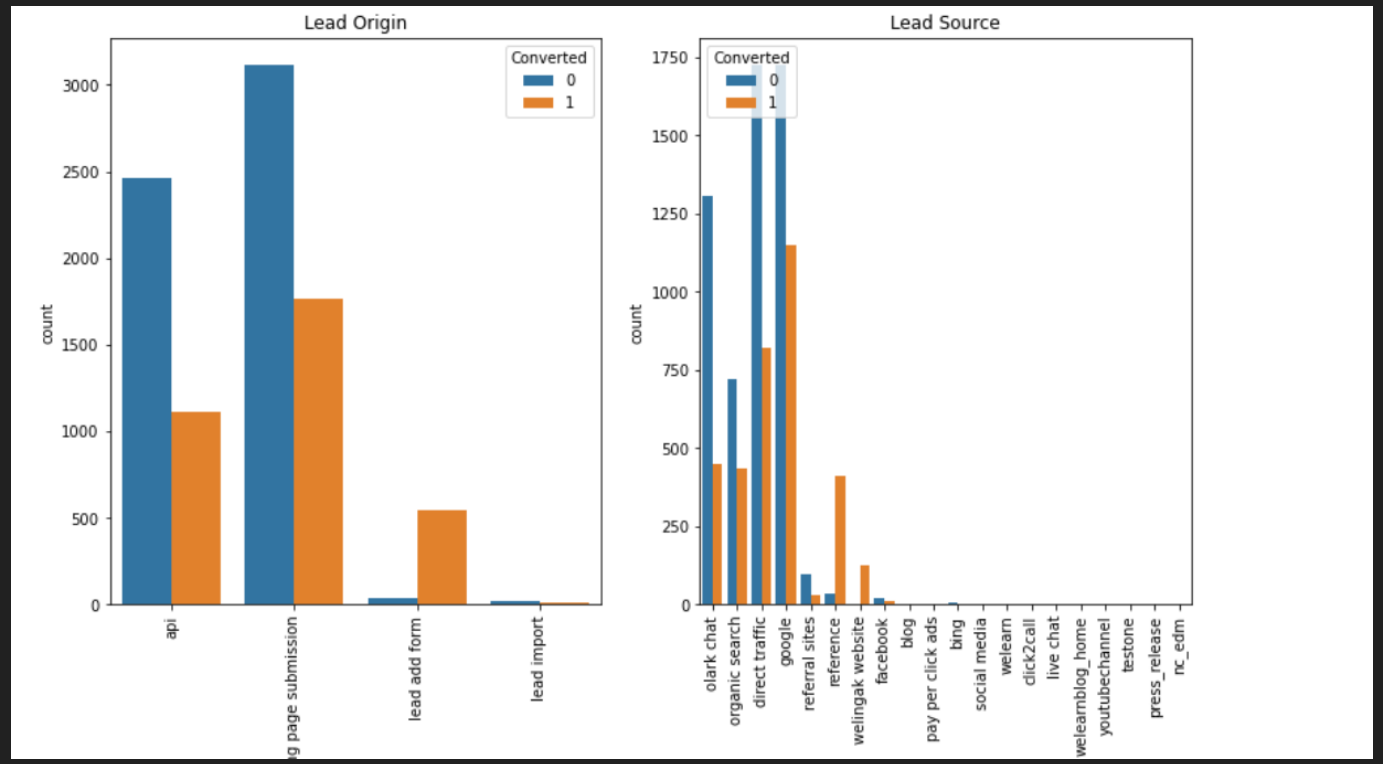


# Exploratory Data Analysis

Page view per visit




Converted plots for Lead origin and lead source

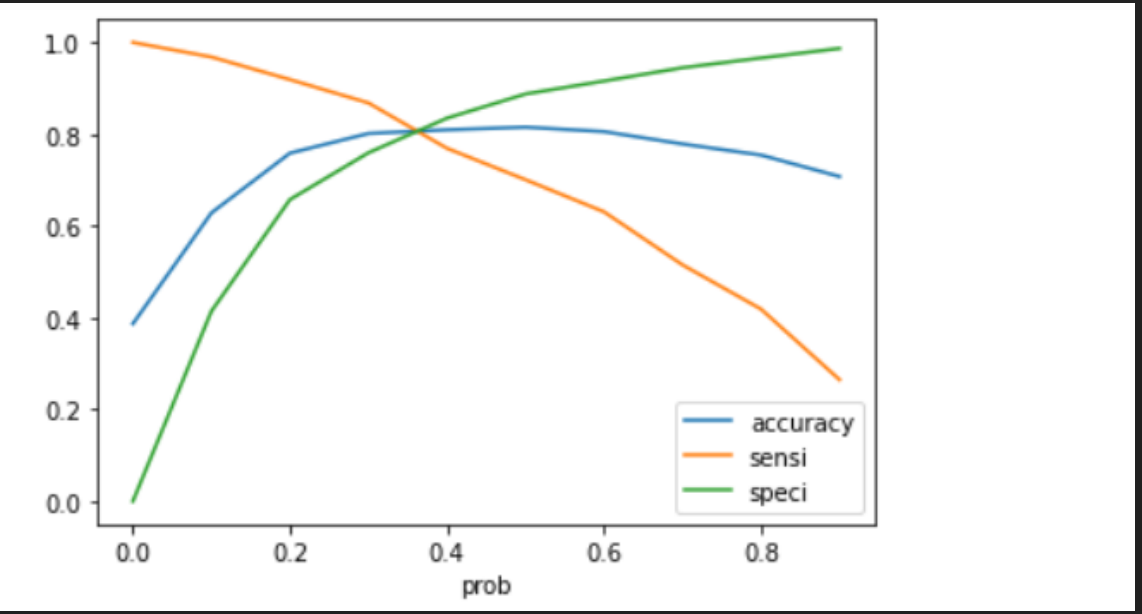


# Variables Impacting the Conversion Rate

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

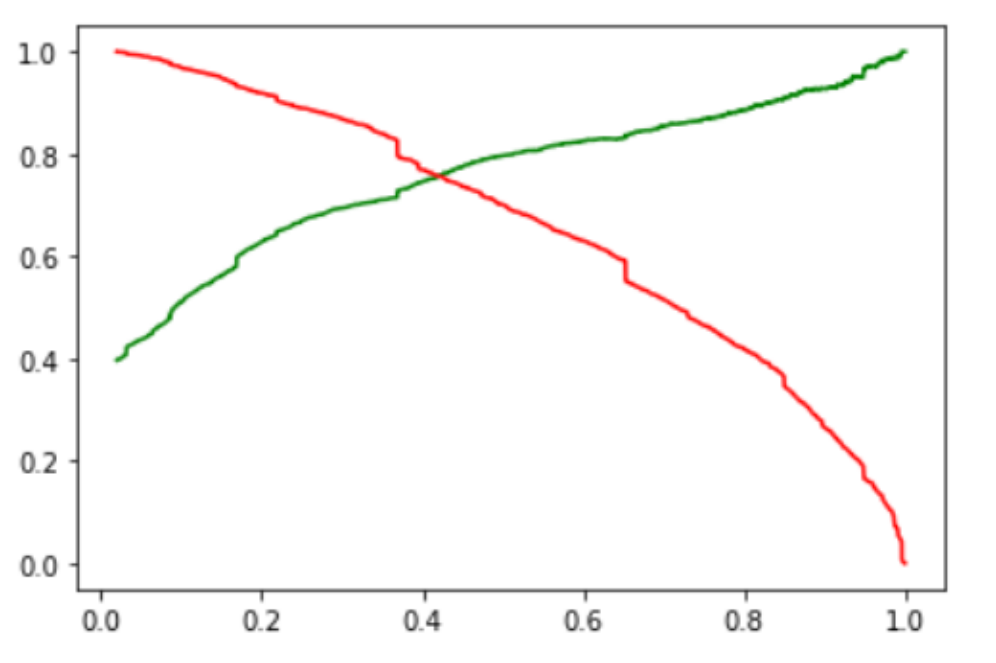
# Model Evaluation - Sensitivity and Specificity on Train Data Set

With the current cut off as 0.35 we With the current cut off as 0.35 we have accuracy, sensitivity of around 83% and specificity of around 78%have accuracy, sensitivity of around 83% and specificity of around 78%



# Model Evaluation - Precision and Recall on Train Data Set

With the current cut off as 0.41 we have Precision around 75% and Recall around 76%



# Conclusion

- The variables found to be most important to potential buyers are (in descending order):
- The total time spends on the Website.
- Total number of visits.
- When the lead source was:
  - Google
  - Direct traffic
  - Organic search
  - Welingak website
- When the last activity was:
  - SMS
  - Olark chat conversation
- When the lead origin is Lead add format.
- When their current occupation is as a working professional.