EXECUTIVE REPORT
ON
"Leads Scoring"

AUTHORS
Abhinay Choday
Abhishek Gandham
Abhijit Ingale

PROBLEM STATEMENT

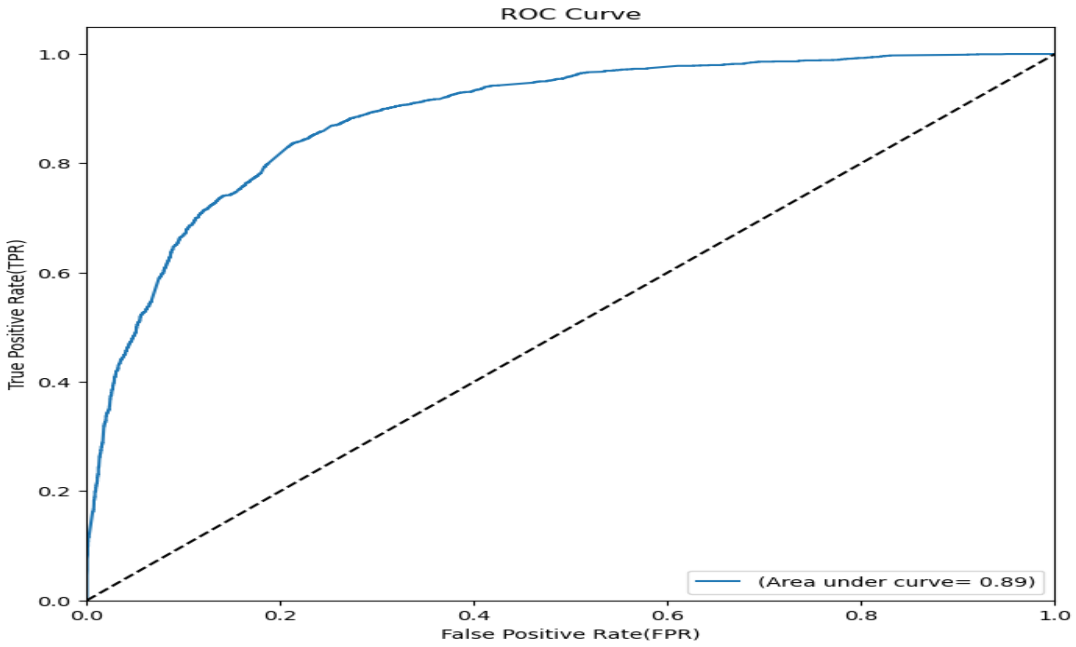
The main aim of X-Education is to increase their lead conversion rate from 30% to 80%, to increase the sale and revenues generated by the same. X-Education envisages to use previous conversion data and model a classification algorithm to identify hot leads which have a higher probability of the lead getting converted and crucial insights benefiting the business.

MAJOR INSIGHTS DURING EDA

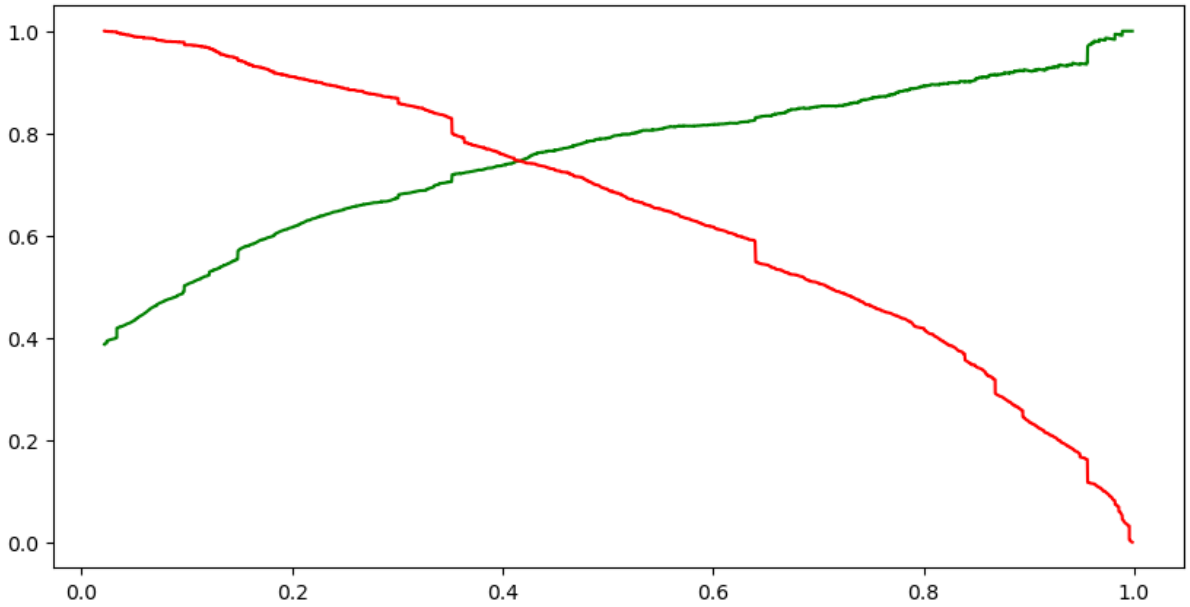
Variables	Insights During EDA
TotalVisits	The median value for Converted Leads is higher than not converted leads
Total Time Spent on Website	The median and IQR between the 25 percentile to 75 Percentile of Total time spent on a website is significantly higher. Time spent on a website could be a significant indicator for conversion.
What is your current occupation	Most of the leads are Unemployed but Working Professionals are more likely to convert since they have a source of income as well as motivation to upskill themselves.
What matters most to you in choosing this course	Better career prospects is the main motivating factor for most people looking to upksill on X education platform. So providing links to better job opportunities and assistance in placements would go a long way in converting leads to paying customers.

MODEL BUILDING

Based on the available data and the business problem at hand, Logistic Regression model is best suited for our binary classification of leads as Hot Leads or not. Logistic Regression model from Sklearn module of python was used to build a model and RFE is used for initial selection of 20 features. Model was optimized based on p-values and VIF for all selected features. The obtained model was verified to be close to ideal with the help of ROC curve.



Now we needed to set a threshold value for deciding Hot Leads based on the probabilities. Since CEO wanted a conversion rate of 80% we used Precision vs Recall view to find the optimum threshold. Based on the Precision vs Recall graph we need a Precsion of 0.8,which has a threshold close to 0.52. With a threshold of 0.52 we get a Precision of 0.8 and False Positive Rate was only 0.1 or 10%. So we decided that a threshold of 0.52 is ideal for our requirements of reaching a target of 80% conversion.



So our final model has an accuracy of 83% , was able to have a False Positive Rate of 10% and a Precision of 80%. In business terms, for a lead classified as Hot Lead by the model there is an 80% chance to convert. The Lead Score would be a number between 0-100 with a higher chance of conversion, higher the number. For our target of 80% conversion, we can decide 52 Lead Score as a threshold for a lead to be considered as Hot Lead.

POINTS OF ACTION FOR BUSINESS

1. Based on the Logistic Regression Model used, Total Time Spent on Website, Lead Originated through Add Forms, and Occupation-Working Professionals have a higher probability to drive conversion.

2. So the company should work towards making website's UI intuitive, interesting and adding engaging content. The aim of the calling executive during initial calls should be to make the lead engage and spend time on website

3. Ensure most leads fill out the Add Forms for getting more details from them.

4. Try to contact working professionals after working hours, always take appointment for calling, have a friendly conversation by asking about their work, recommend and explain about courses available on X education which would improve their career. This needs to be done in multiple calls and emails.