

Lead Scoring Case Study

Nitu Kumari
Phong Nyugen

Problem Statement

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. A typical lead conversion process can be represented using the following funnel:

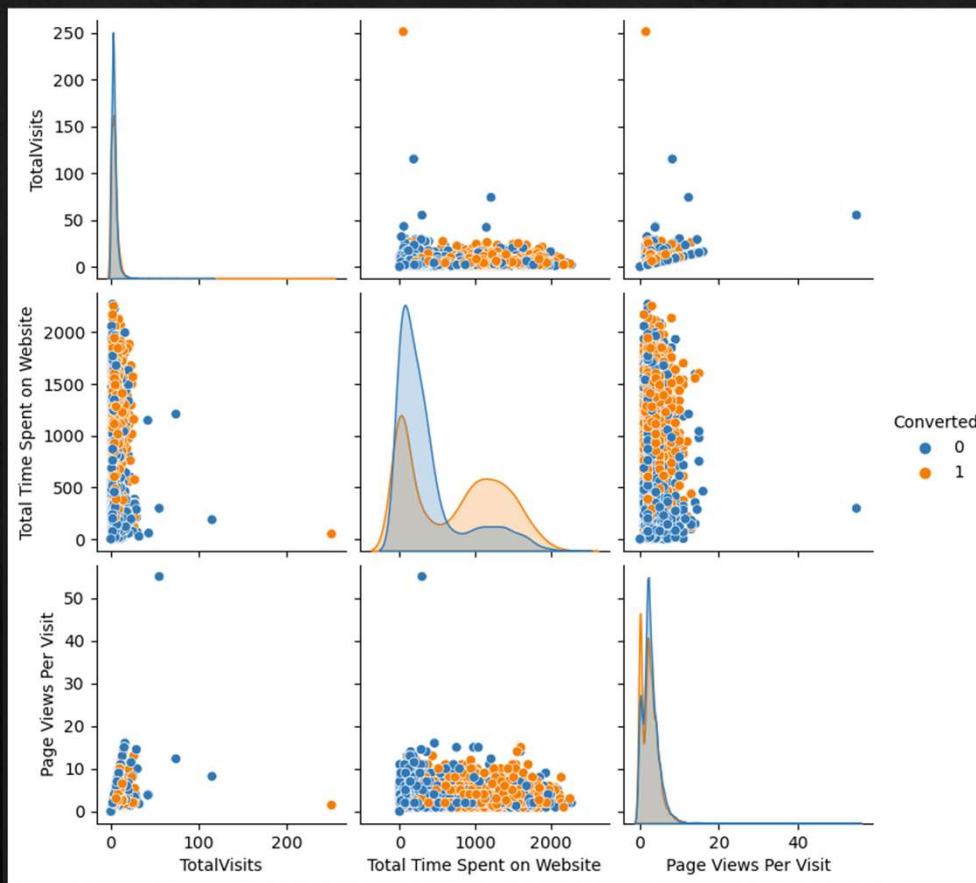
Objectives

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

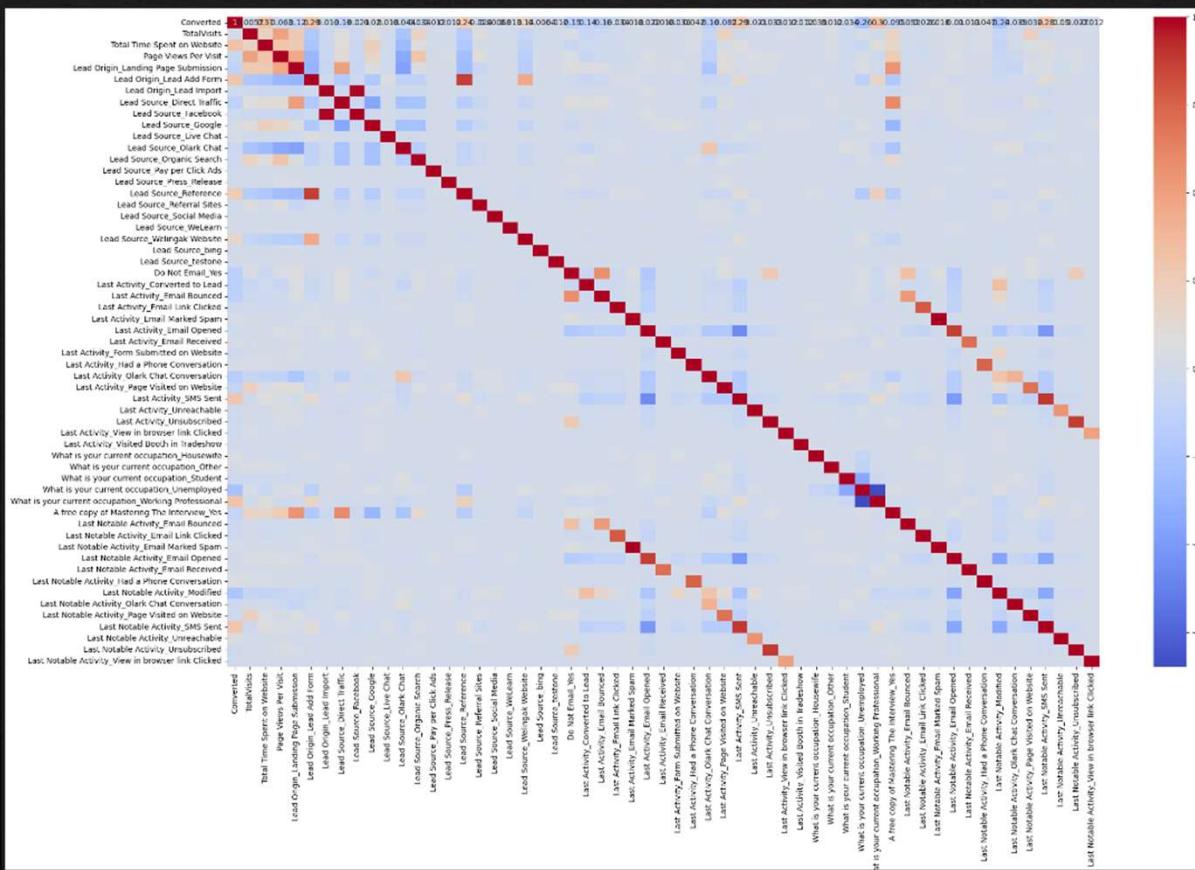
Data Understanding and Preparation

- ❖ Using df.shape we observe that the dataset had 9240 rows and 37 columns.
- ❖ It has Object , Int64 and float type data .
- ❖ There are 6 columns with missing values count more than 3000. These columns are not useful , we eliminated these six columns.
- ❖ Since it is online platform , we can drop City and Country column from the dataset .
- ❖ There are few columns where user did not select anything – they appear with “Select”. They are as good as missing values. Eg Lead profile, How did you hear about X education. We can eliminate them.

Exploratory Data Analysis



Exploratory Data Analysis



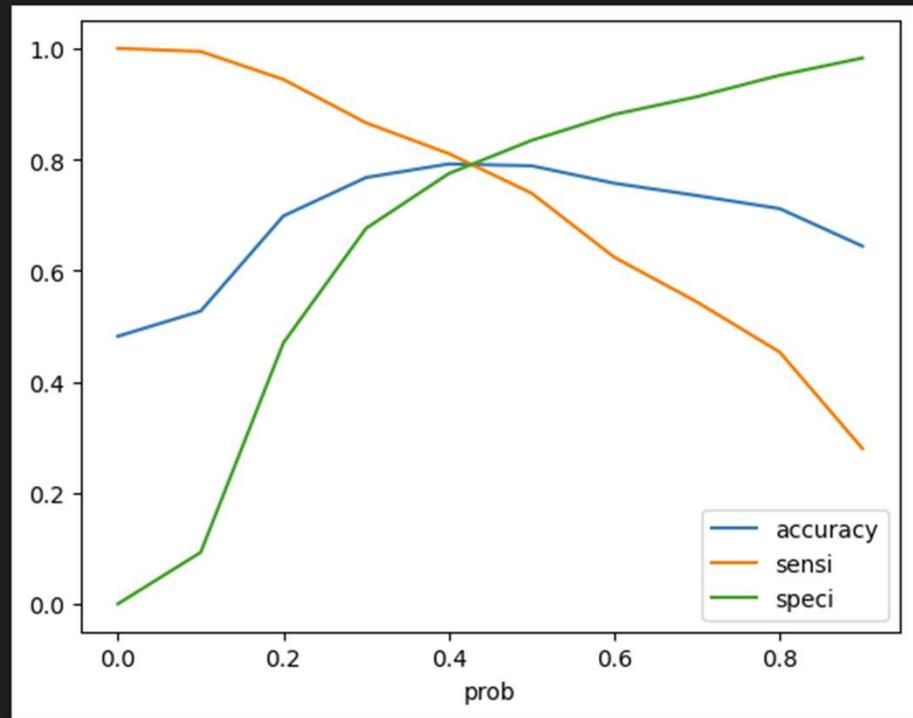
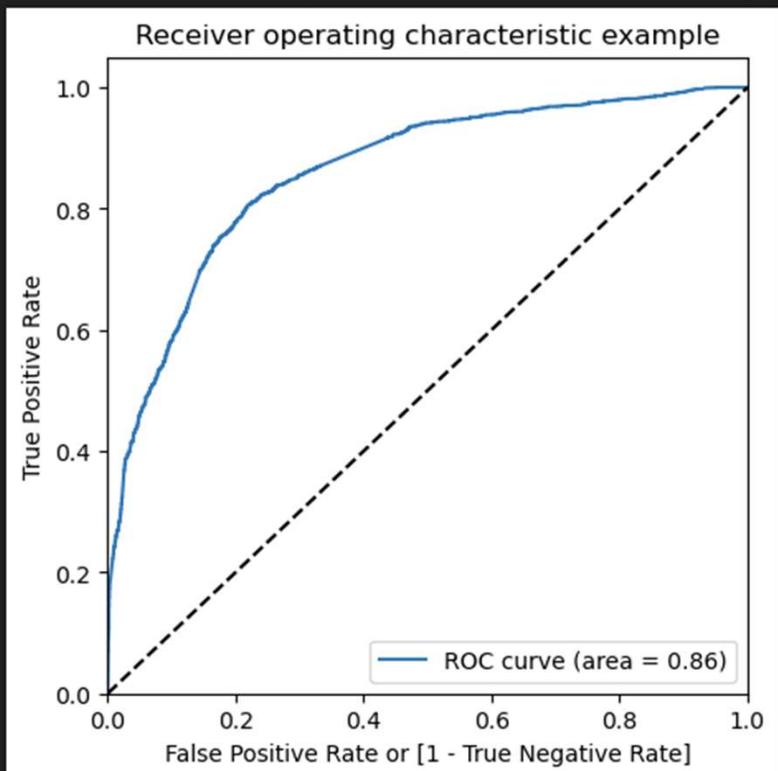
Model Approach

- ❖ Import 'LogisticRegression' and create a LogisticRegression object
- ❖ Import RFE and select 15 variables
- ❖ running RFE with 15 variables as output
- ❖ Let's take a look at which features have been selected by RFE
- ❖ Fit a logistic Regression model on X_train after adding a constant and output the summary
- ❖ Import 'variance_inflation_factor'
- ❖ Make a VIF dataframe for all the variables present
- ❖ Let's first drop the variable example. `Lead Source_Reference` since it has a high p-value as well as a high VIF.
- ❖ Refit the model with the new set of features. Repeat .

Model Evaluation

- ❖ Once both the p-values and VIFs seem decent enough for all the variables. Then we make predictions using this final set of features.
- ❖ Use 'predict' to predict the probabilities on the train set
- ❖ Create a new dataframe containing the actual conversion flag and the probabilities predicted by the model
- ❖ Create confusion matrix
- ❖ Check Accuracy , sensitivity and specificity
- ❖ Find the optimal cut off using ROC curve.

Model Evaluation



Key variables and Business Insights

- ❖ Total Visits,
- ❖ Total Time spent on the website,
- ❖ Lead Origin_Lead Add Form
- ❖ Lead Origin_Lead Add Form,
- ❖ Last Activity_Had a phone conversation,
- ❖ Lead Score_Wellingak Website

Strategies

- ❖ >Lower the Lead Score Threshold
- ❖ >Segment Leads
- ❖ >Use Engagement Metrics
- ❖ >Train interns
- ❖ >Monitor and Adjust

Recommendations and Action Plan

- ❖ >Increase the Lead Score Threshold
- ❖ >Automate Low-Priority Leads
- ❖ >Reassess Lead Activity
- ❖ >Sales Team Allocation
- ❖ >Follow-Up Automation
- ❖ >Periodic Lead Scoring Update