LEAD SCORING CASE STUDY

Group Members :

1. Pokala Preethi Shivani
2. Akshaya C
3. Richa

Problem Statement

X Education is a organization which sells online courses for industry professionals, they markets its courses on several websites and search engines like Google.

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.

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. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

Requirements of the Company

The company requires to build a model for selecting most promising leads.

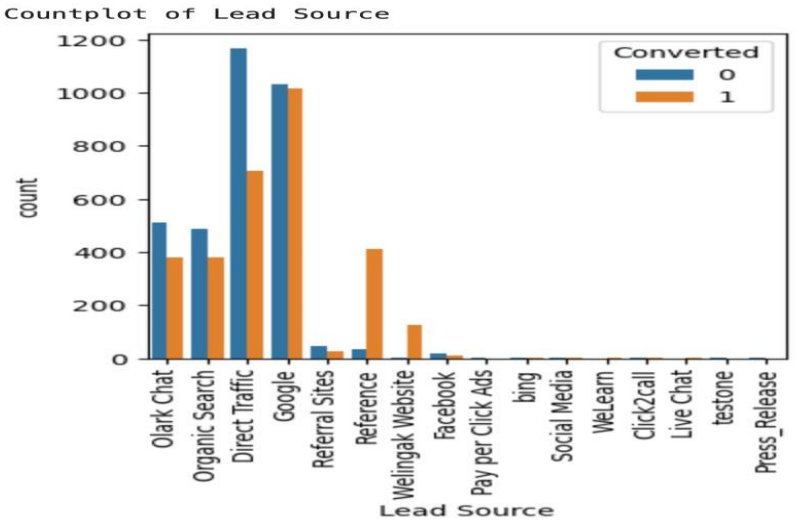
Lead score to be given to each leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

The model to be built in lead conversion rate around 80% or more.

Strategy

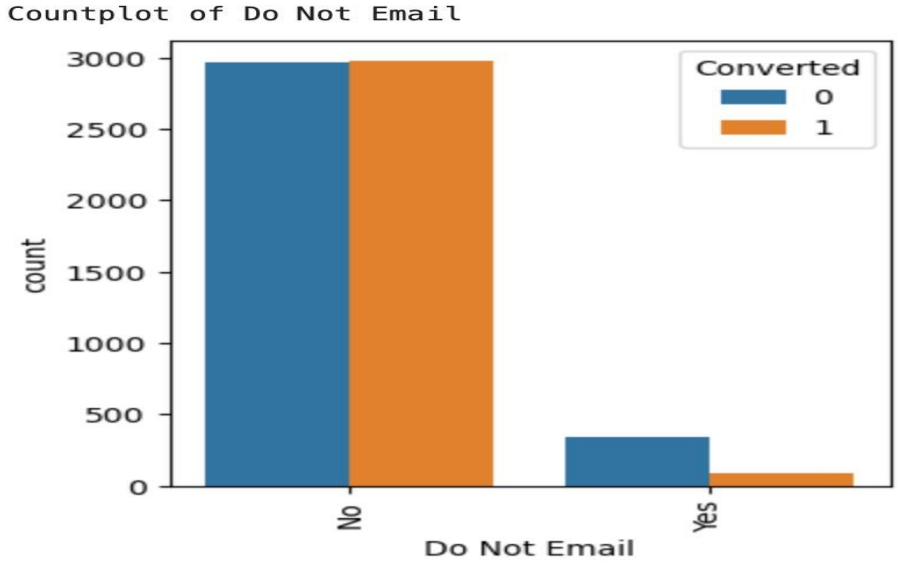
- Important data
- Clean and prepare the acquired data for further details
- Scaling Features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model

Exploratory data analysis



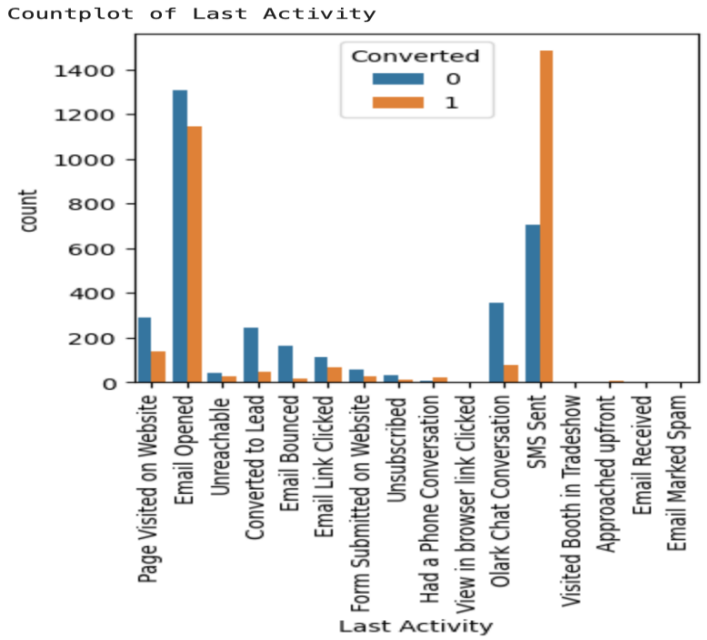
Lead source vs converted

Google search has the highest conversions compared to other modes. References has most conversion rate.



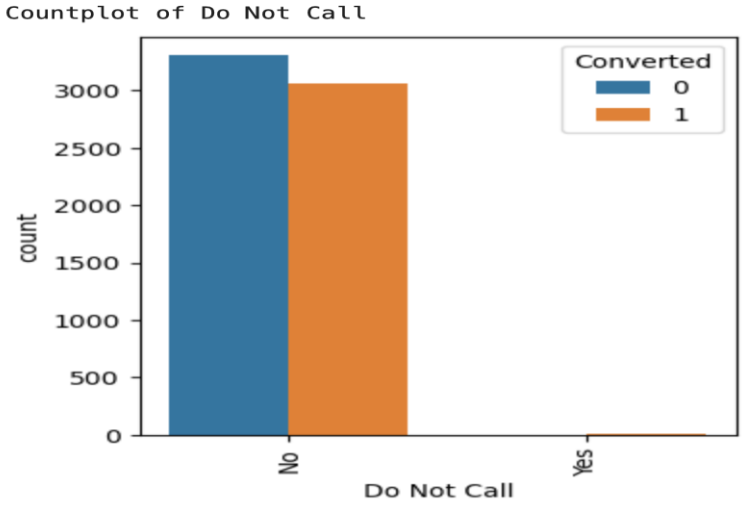
Do not email v/s conversion

'Do not email' selection has not majorly affected conversion



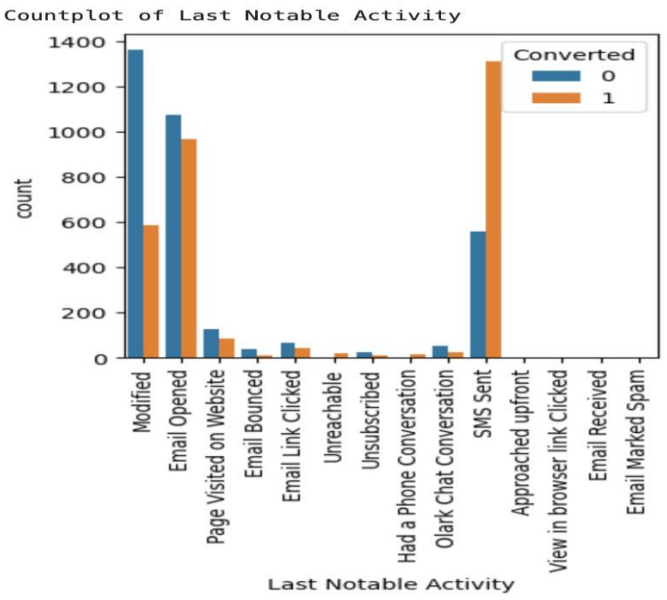
Last activity vs converted

SMS has shown to be a promising method for getting higher confirmed leads, emails also has high conversion.



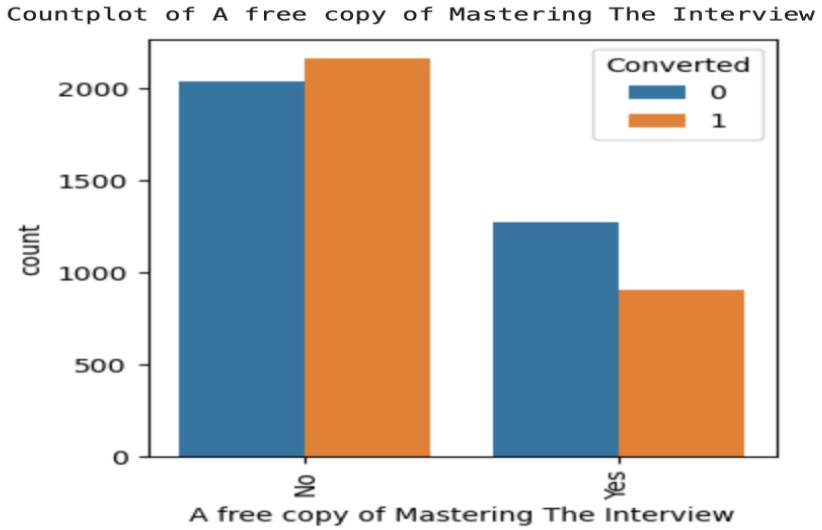
Do not call vs converted

Most leads prefer not to informed through phone



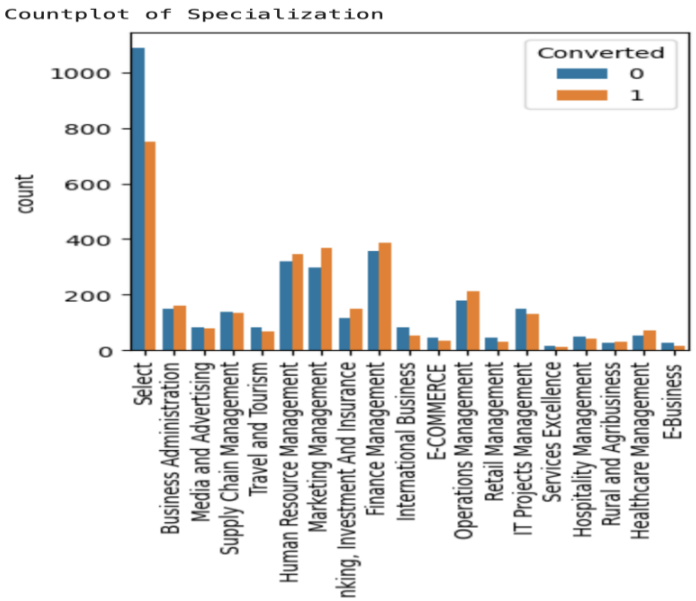
Last notable activity vs converted

Most leads are converted with messages. Emails also include leads.



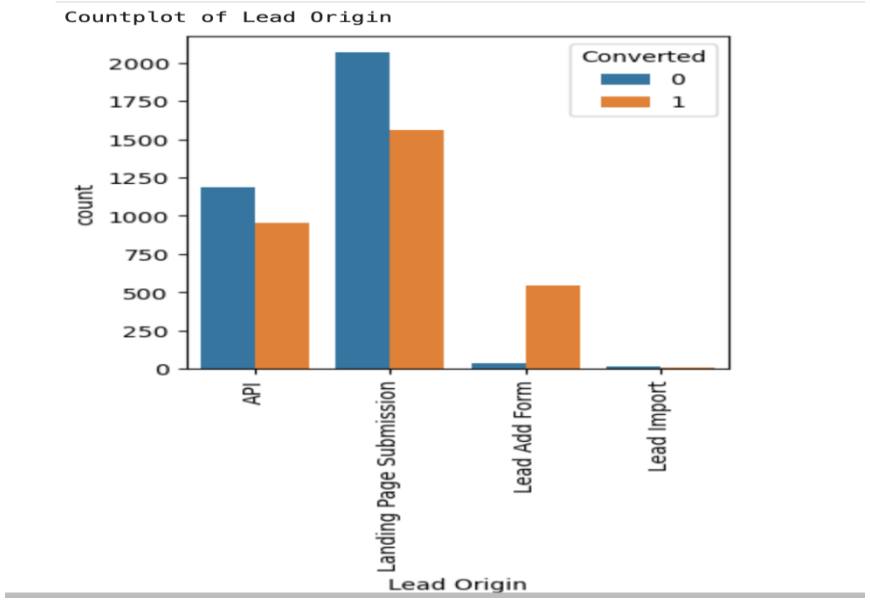
A Free copy of Mastering the interview vs converted

Leads prefer less copies of interviews



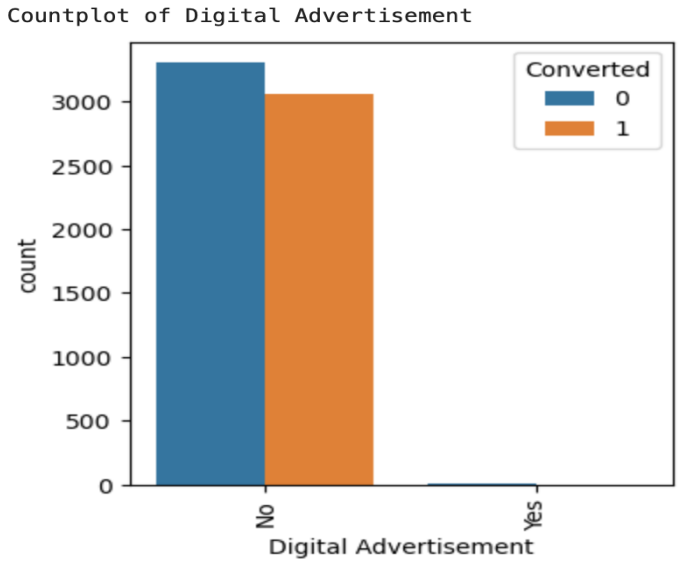
Specialization vs converted

Most of the leads have no information about specialization. On the other had, marketing management, human resources management has high conversion rates, people from these specialization can be promising leads



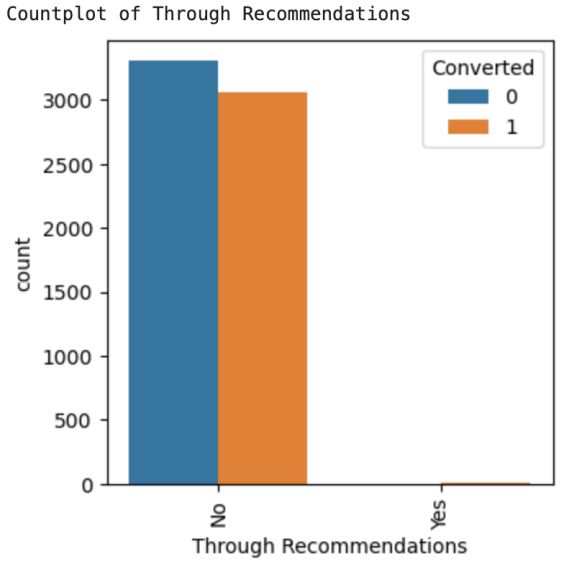
Lead origin vs converted

API, Landing Page Submission, Lead Add Form brings good no of leads as well as conversion.



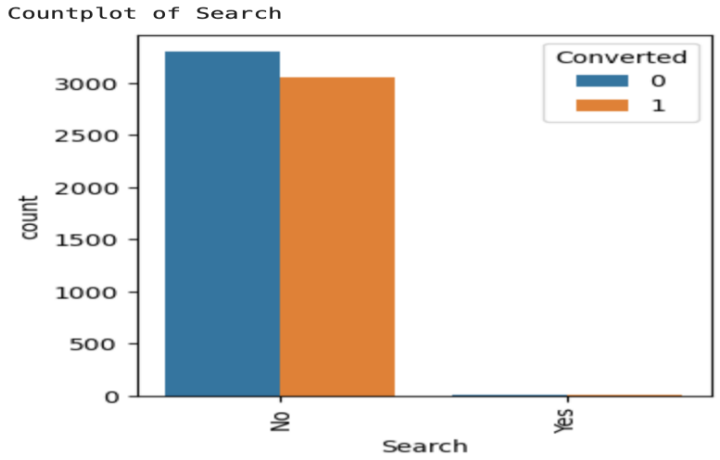
Digital advertisement vs converted

Based on the above graph digital advertisements do not have promising leads



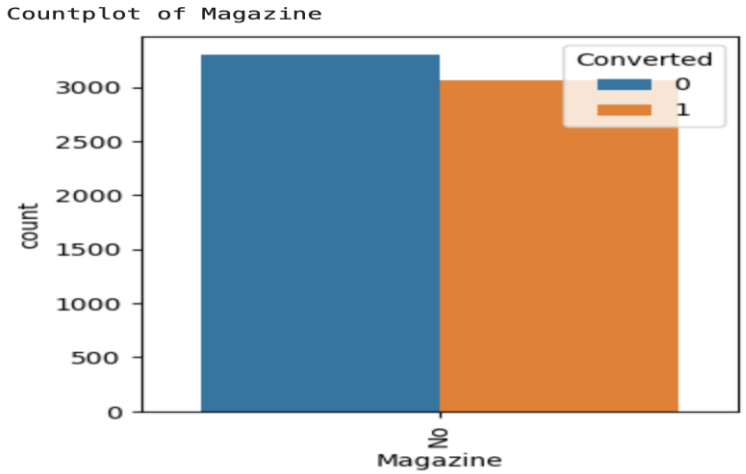
Through recommendations vs converted

From the above graph, recommendations are not a good source for promising leads



Search vs converted

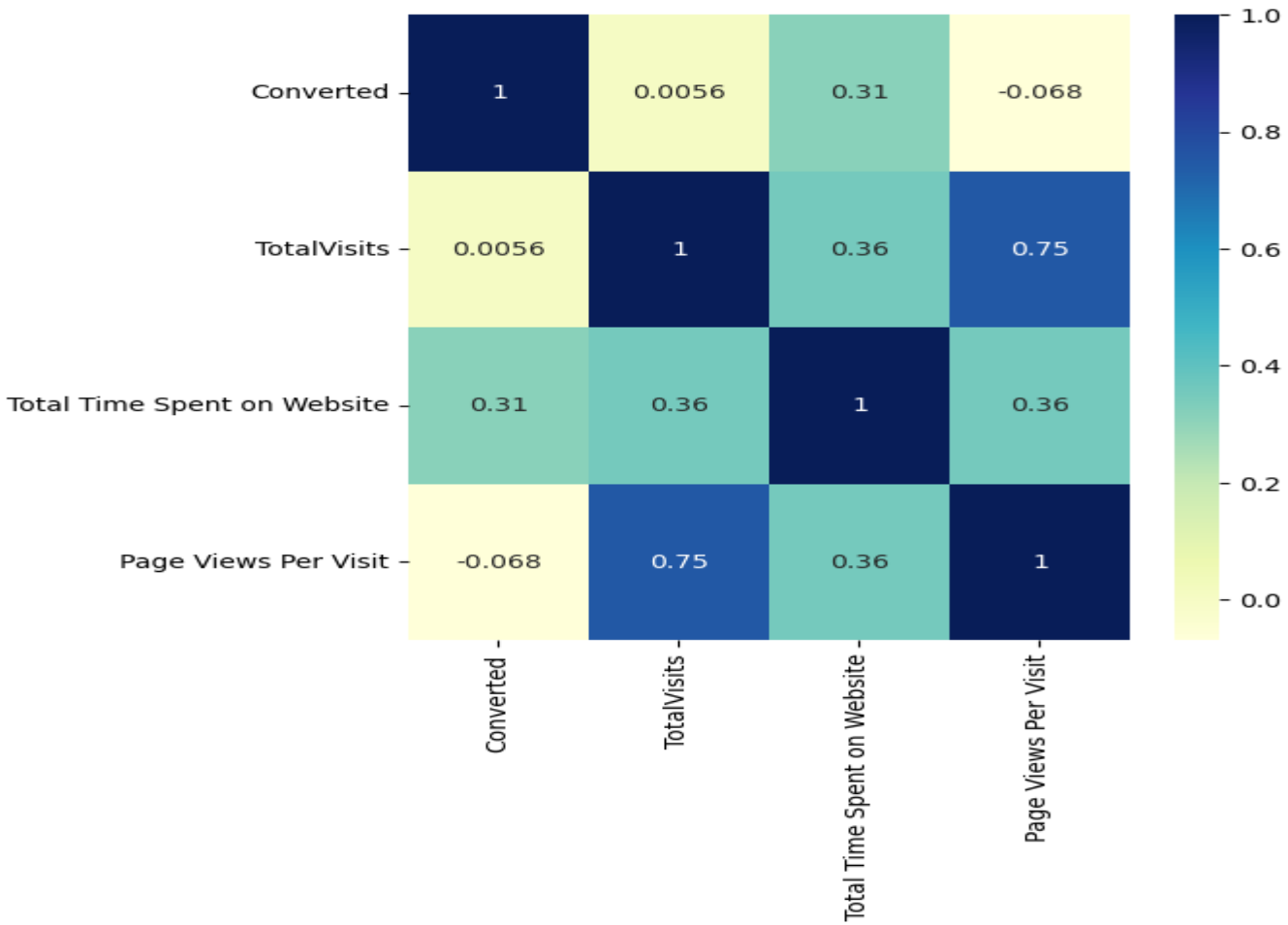
The above graph shows searches are not good sources of leads.



Magazine vs converted

Magazines do not have higher conversion rates

Correlation Between Variables



- Total Time Spent on Website has a good Correlation with Target Variable Converted.
- Total Time Spent on Website and TotalVisits has correlation with each other.

Model building:

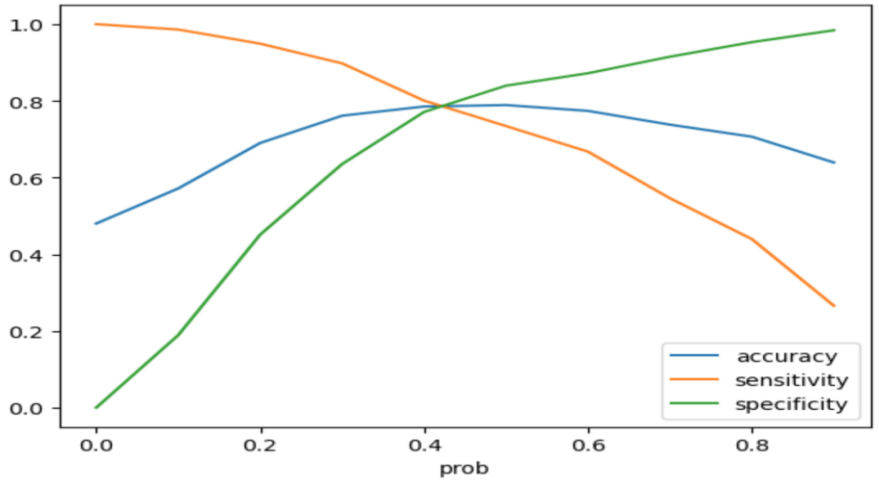
Before model building, we do the following:

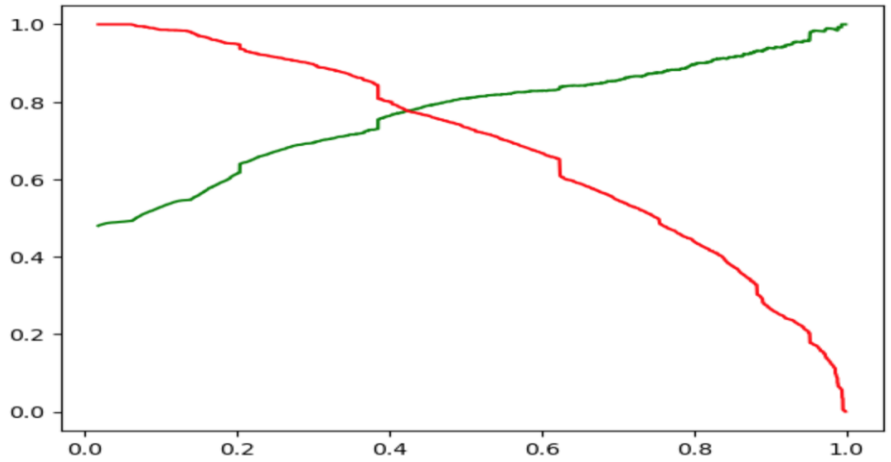
- Dropping the columns which are not necessary
- Creating Dummy Variables
- Splitting data into Training and Testing sets
- Rescaling Variables

Next steps:

- Build the first model
- Use RFE to eliminate less relevant variables
- Build next models
- Eliminate variables based on high p-values
- Check VIF value for all the existing columns
- Predict using train data
- Evaluate accuracy and other matrices
- Plot the ROC Curve
- Find Optimal Cutoff point

Model Evaluation:





Final Observation:

Train Data:

- Accuracy: 78.58%
- Sensitivity : 79.07%
- Specificity : 78.13%

Test Data:

- Accuracy : 79.28%
- Sensitivity : 79.87%
- Specificity : 78.74%

- Thus we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% .
- The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 80%

Conclusion

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

1. When current occupation is :
    a. Student
    b. Unemployed
2. When the last activity was :
    a. SMS Sent
    b. Had a Phone Conversation
3. The total time spend on the Website.
4. When the lead source was:
    a. Olark Chat
    b. Reference
    c. Welingak Website
5. Do Not Email_Yes

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.