

X EDUCATION

Lead Score Case Study

PROBLEM STATEMENT:

An education company named X Education sells online courses to industry professionals. 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%

BUISNESS GOALS:

X Education needs help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

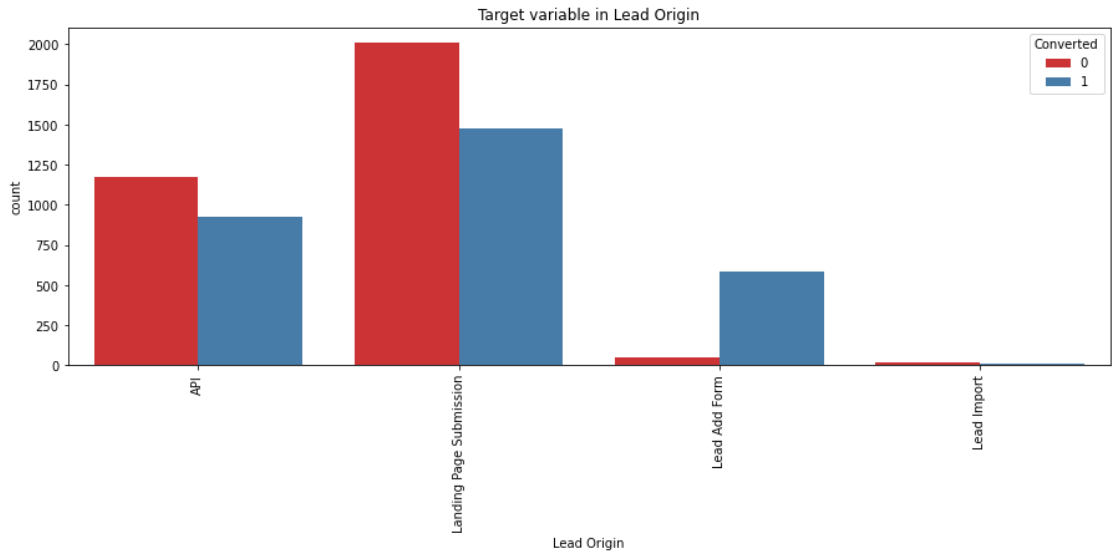
The company needs a model where a lead score is assigned to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

Goal is to increase lead conversion rate to be around 80% from 30% to increase the revenue of the company.

STRATEGY:

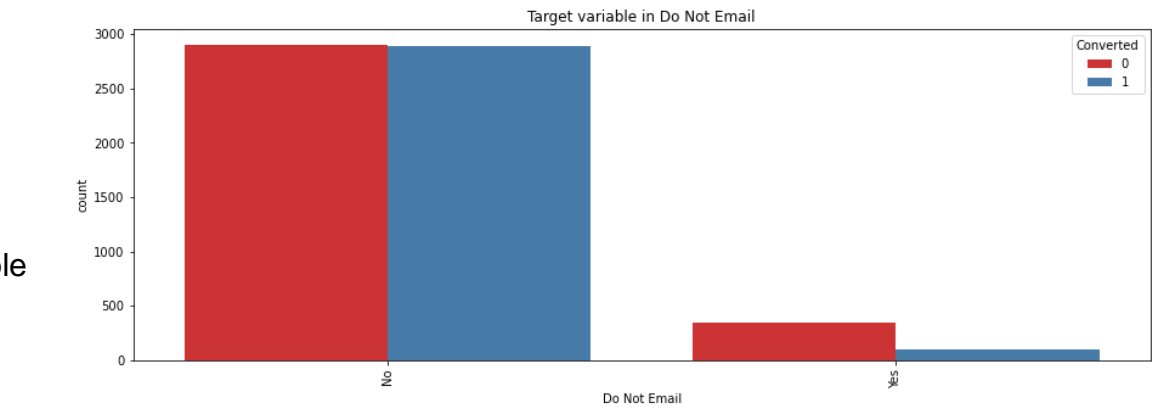
- Importing data
- Data Cleaning and Manipulation
- Explanatory Data Analysis
- Scaling Features
- Prepare the data for Model Building
- Splitting the data into Test and Train Dataset
- Building a Logistic Regression Model and assign Lead Scores
- Test the model on Train set
- Evaluate the model by different Measures and Metrics
- Test the model on Test set
- Measure the accuracy of the model
- Evaluate and execute the perfect fit model

INFERENCES FROM EXPLANATORY DATA ANALYSIS:



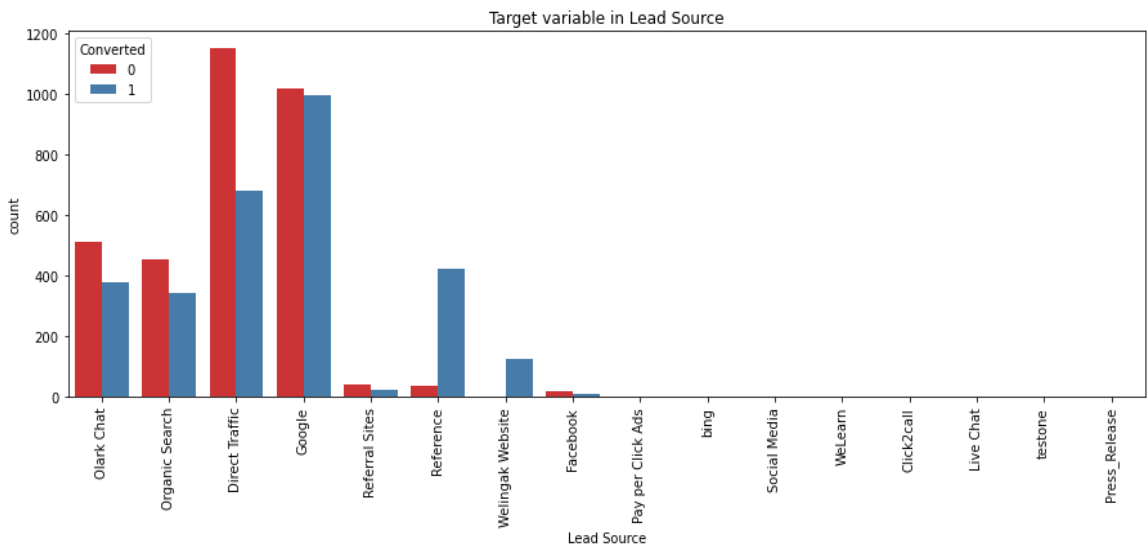
LEAD ORIGIN VS CONVERTED

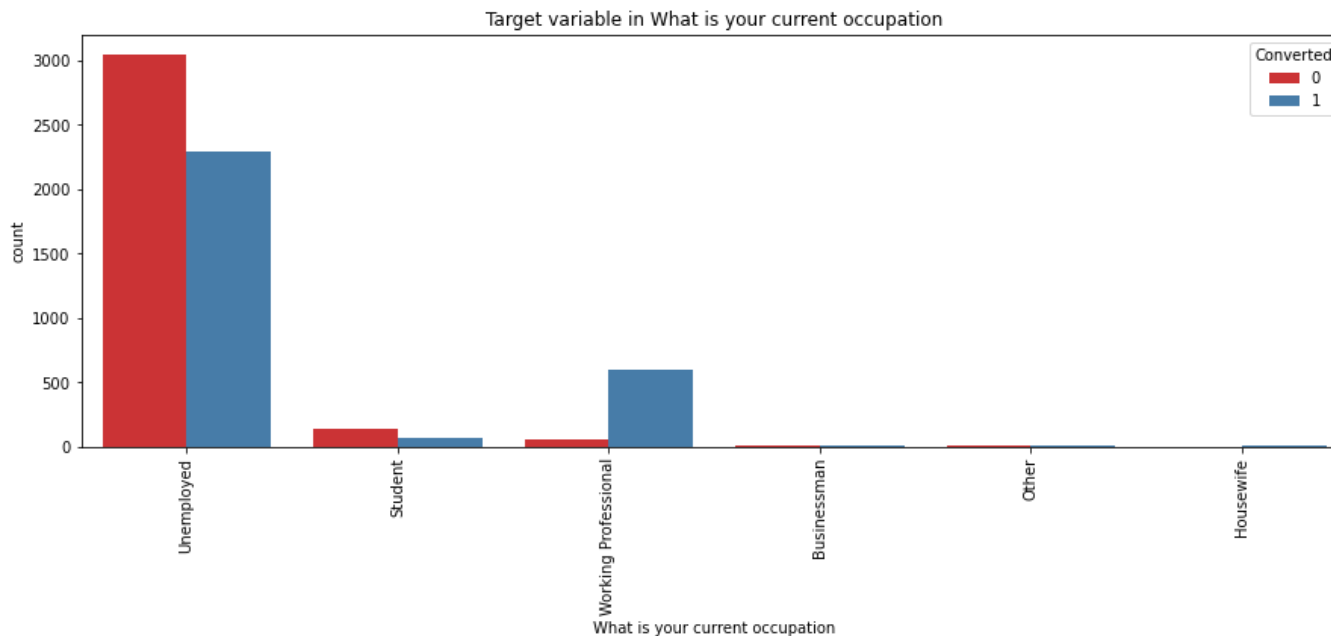
Landing page submission constitutes of majority of leads
Lead add form consists of high probability of lead conversions.
API and Lead import has least conversion rate though leads from lead import are very less in number.



LEAD SOURCE VS CONVERTED

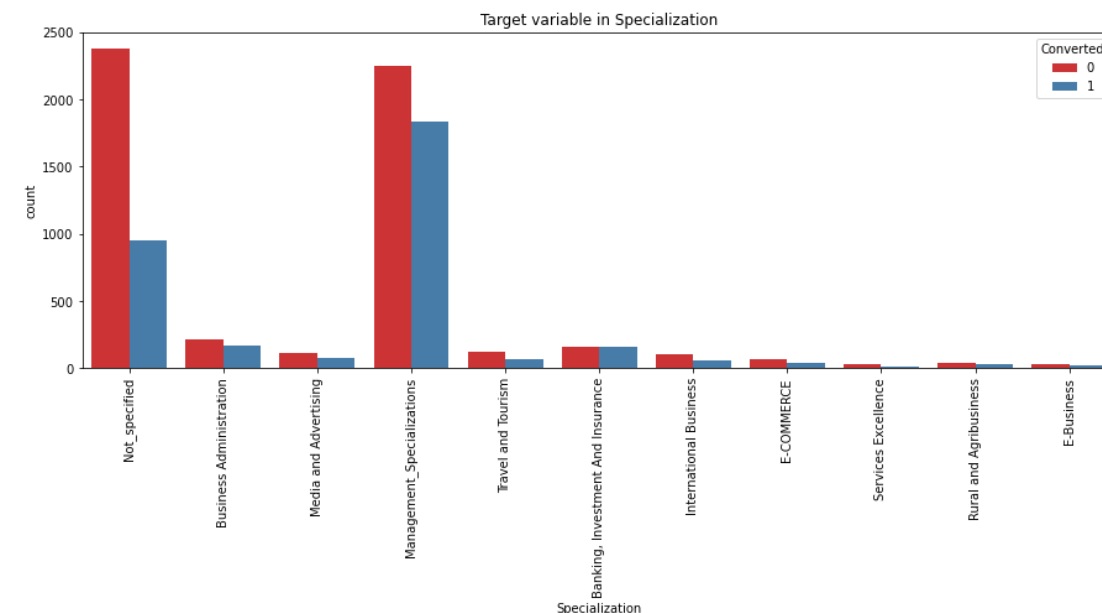
Majority of leads are from Google and direct traffic.
Reference have good conversion rate.





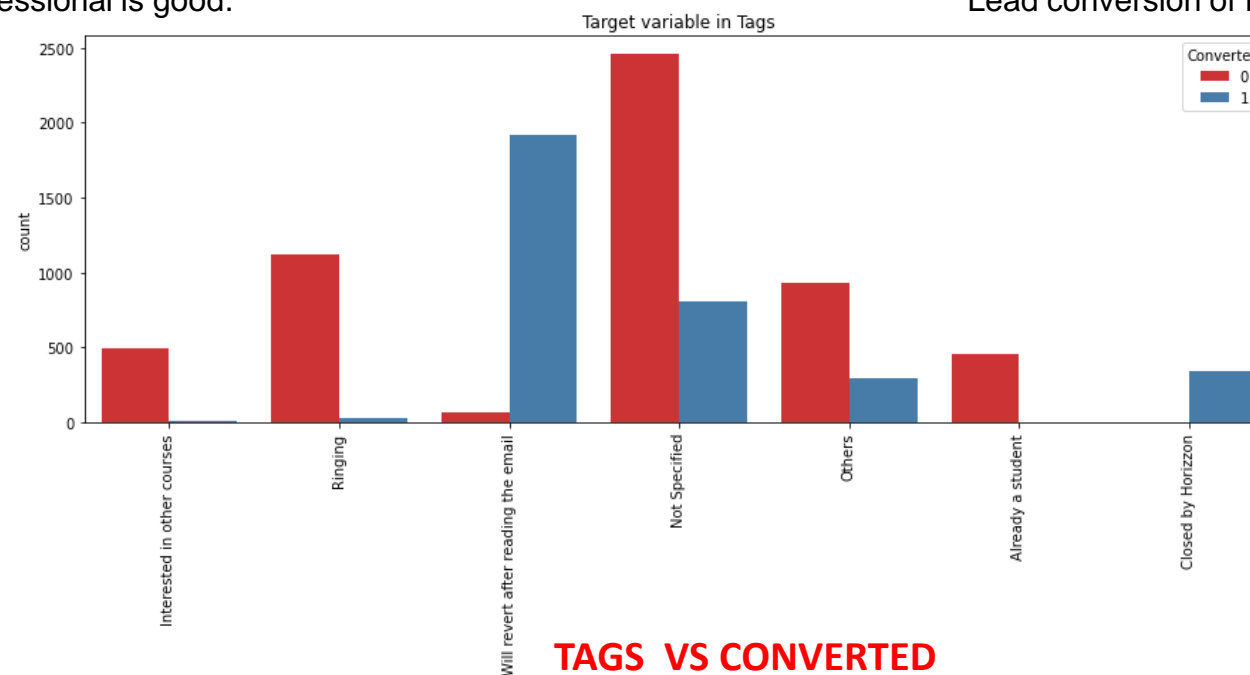
WHAT IS YOUR CURRENT OCCUPATION VS CONVERTED

Maximum leads have occupation as Unemployed,
Conversion rate of working professional is good.

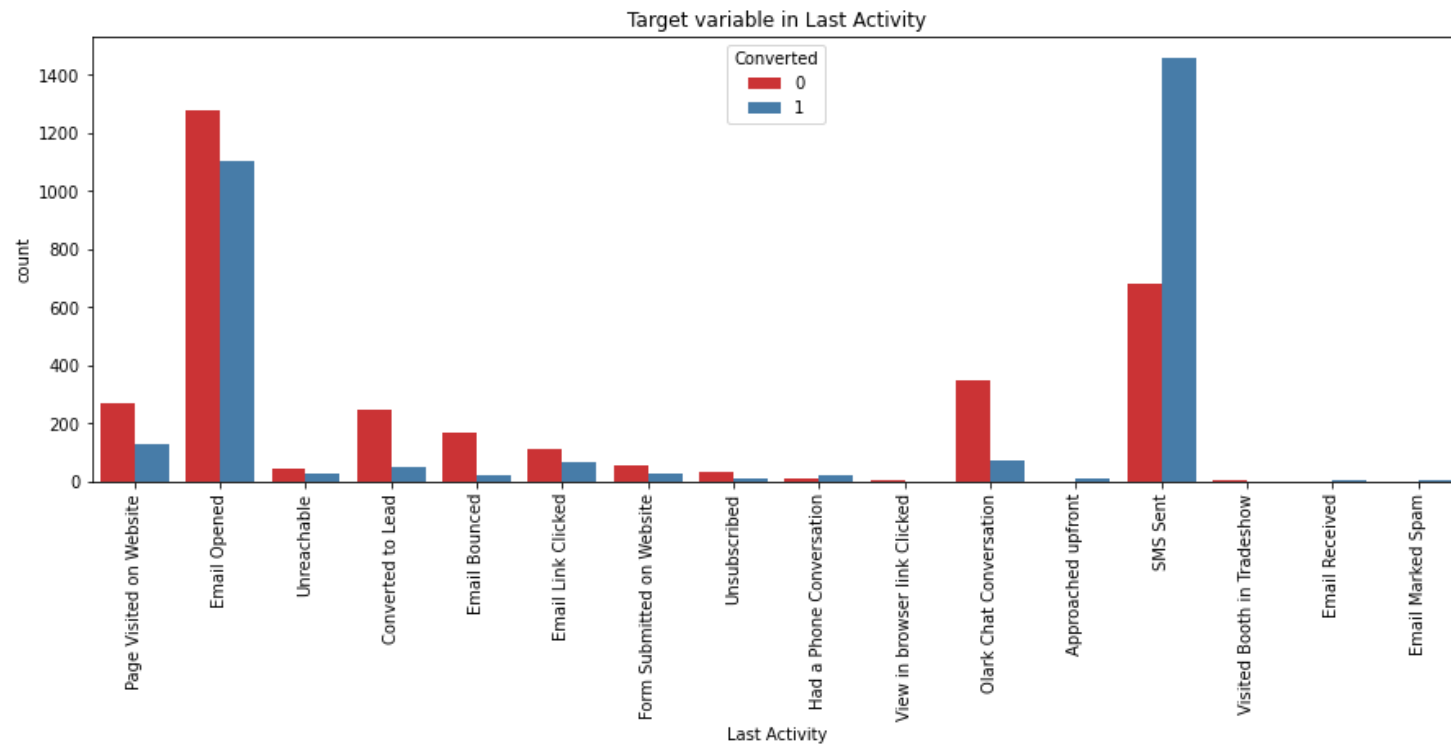


SPECIALIZATION VS CONVERTED

Not specified has maximum number of leads
Lead conversion of management specialization is high..

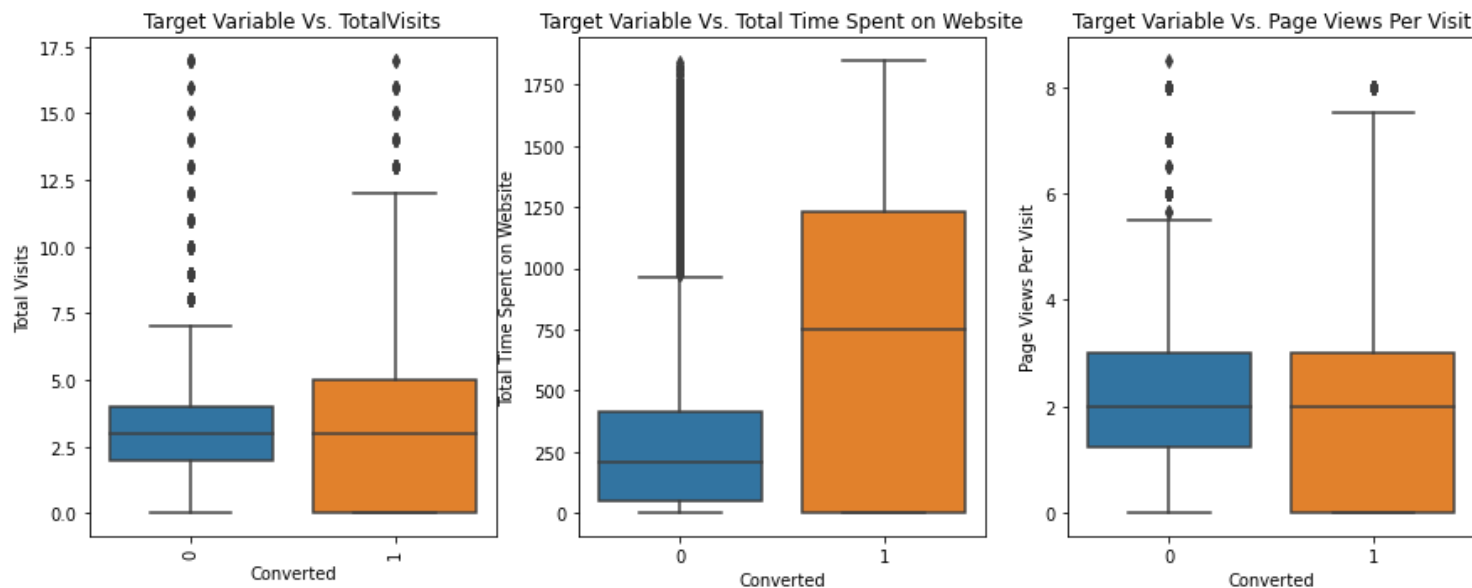


TAGS VS CONVERTED



LAST ACTIVITY VS CONVERTED

SMS sent and Email Opened as their last activity has maximum leads and lead conversions.



TOTAL VISIT VS CONVERTED

Total visits - more the number of visits, more chances of lead conversion

TOTAL TIME SPENT ON WEBSITE VS CONVERTED

Total Time Spent on website concludes if person is spending more time on website , has more chances of lead conversion.

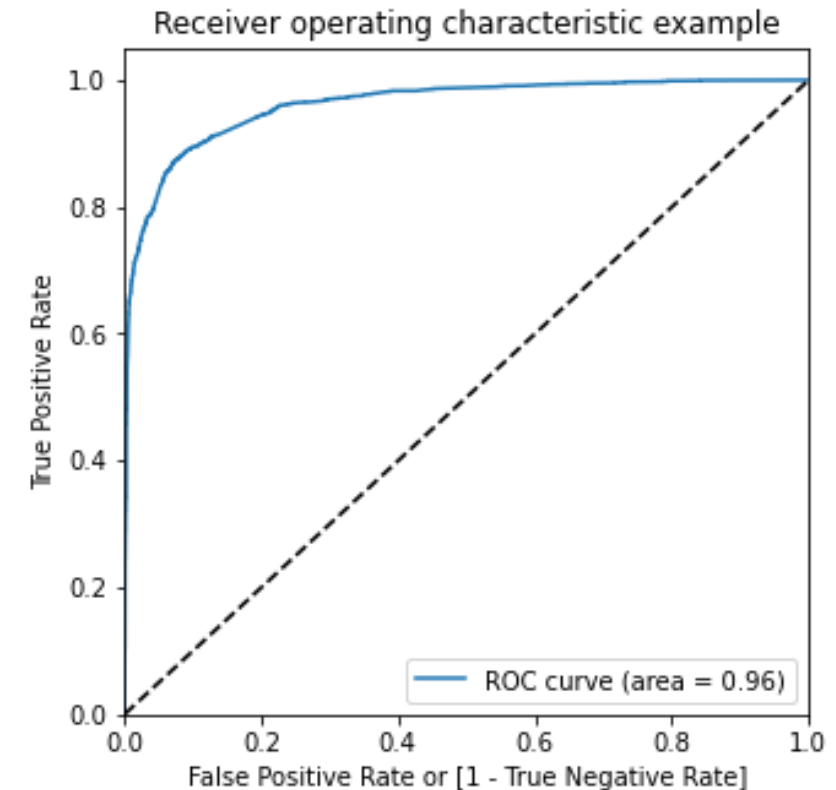
PAGE VIEWS PER VISIT VS CONVERTED

Page Views Per Visit - Median for both types of Leads i.e converted and non converted are similar so Nothing can be concluded.

MODEL BUILDING:

Variable impacting the conversion rate:

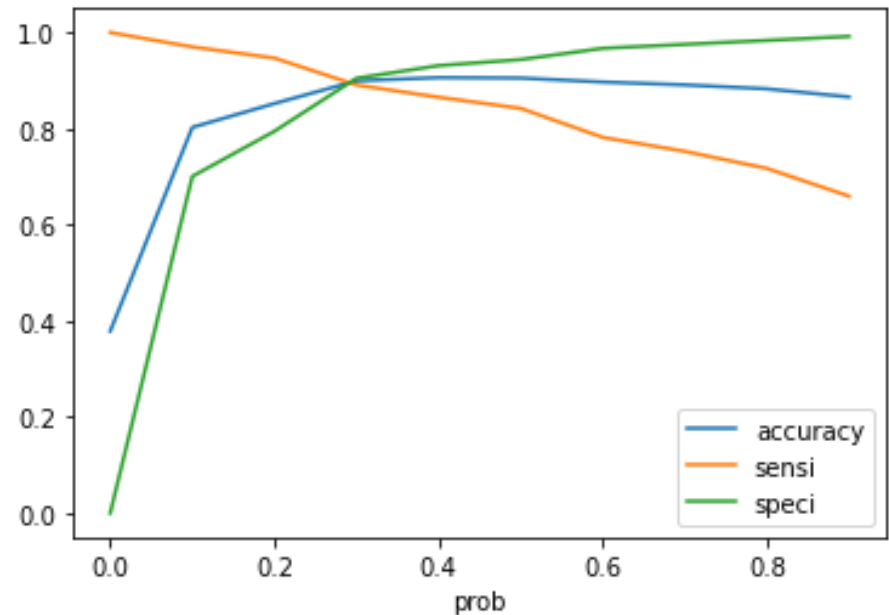
- Total time spent on website
- Lead Origin_Lead Add Form
- What is your current occupation_Working Professional
- Last Activity_Email Bounced
- Last Activity_Email Opened
- Last Activity_Olark Chat Conversation
- Last Activity_SMS Sent
- Lead Source_Olark Chat
- Lead Source_Welingak Website
- Specialization_Travel and Tourism
- Tags_Already a student
- Tags_Closed by Horizon
- Tags_Interested in other courses
- Tags_Ringing
- Tags_Will revert after reading the email



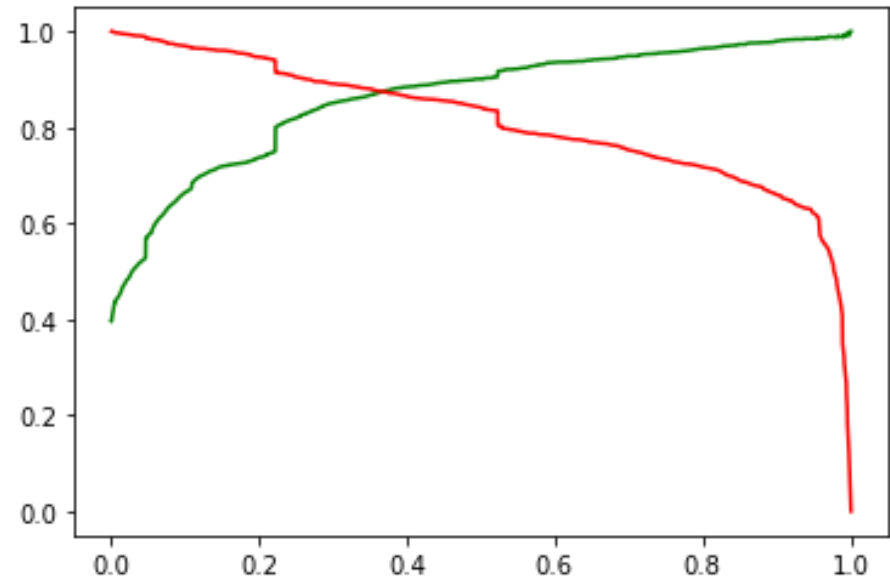
The ROC curve has a value of 0.96

MODEL EVALUATION:

TRAIN



TEST



Confusion matrix

3668	219
374	1995

Accuracy : 89.94%
Sensitivity : 89.06%
Specificity : 90.48%

Confusion matrix

3517	370
259	2110

Accuracy : 89.59%
Sensitivity : 88.52%
Specificity : 90.28%

CONCLUSION:

EDA:

- Landing page submission can help in finding more positive leads.
- Leads from lead add from has high conversion rate.
- Google is major lead source and with maximum conversions.
- References is good source for leads as its conversion rate is high, should run offers for referring.
- Customers willing to see the mails can be considered as positive sign for conversions
- More number of calls, more chances of lead conversion.
- Unemployed customers have more chances of lead conversion
- SMS sent and Email Opened as their last activity has maximum leads and lead conversions.
- Not specified has maximum number of leads in specialization and Lead conversion of management specialization is high.

LOGISTIC REGRESSION MODEL:

- The model shows high close to 89% accuracy
- The threshold has been selected from Accuracy, Sensitivity, Specificity measures and precision, recall curves.
- The model shows 89% sensitivity and 90% specificity.
- The model finds correct promising leads and leads that have less chances of converting
- Overall this model proves to be accurate.