LEAD CASE STUDY

Atul Tiwari

Piyush Arya

Kajal Shah

Problem statement

- X Education is an organization which sells online courses for industry professional. The company marks its courses on several popular websites including 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 conversions.

Case Study 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 more promising the lead to get converted.
- The model to be built in lead conversion rate around 80% or more.
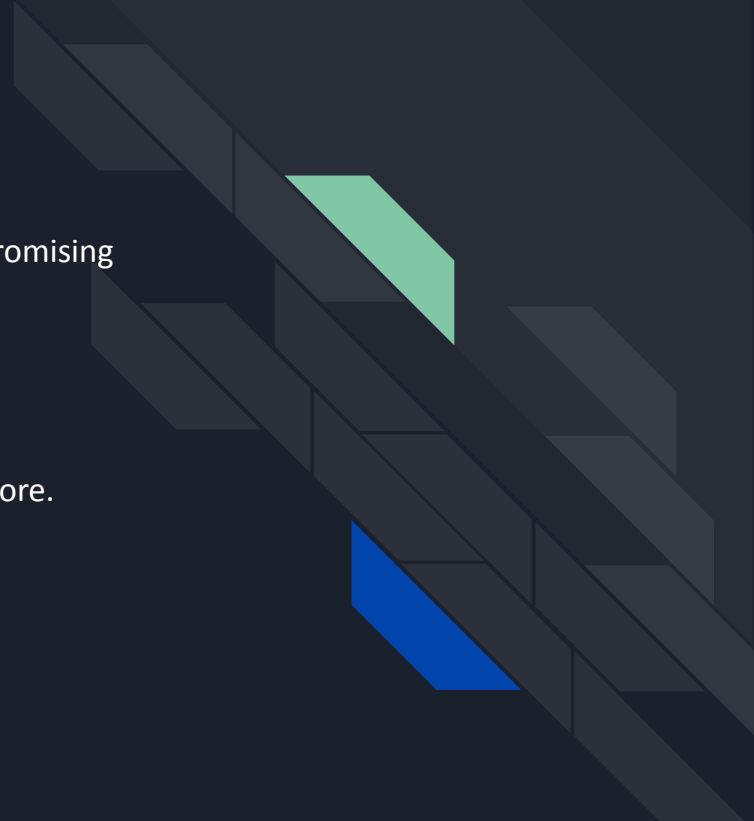
Lead Source Analysis

- 'Google' and 'Direct Traffic' generate the greatest number of leads and lead conversions while some sources like 'Welingkar Website', 'Reference' and 'Others' are having maximum conversion ratio of leads.

- To improve overall lead conversion rate, focus should be on improving lead conversion of 'direct traffic' and 'google leads' and efforts should be put to generate more leads from 'reference' and 'welingak website' as they are conversion rate is very strong.
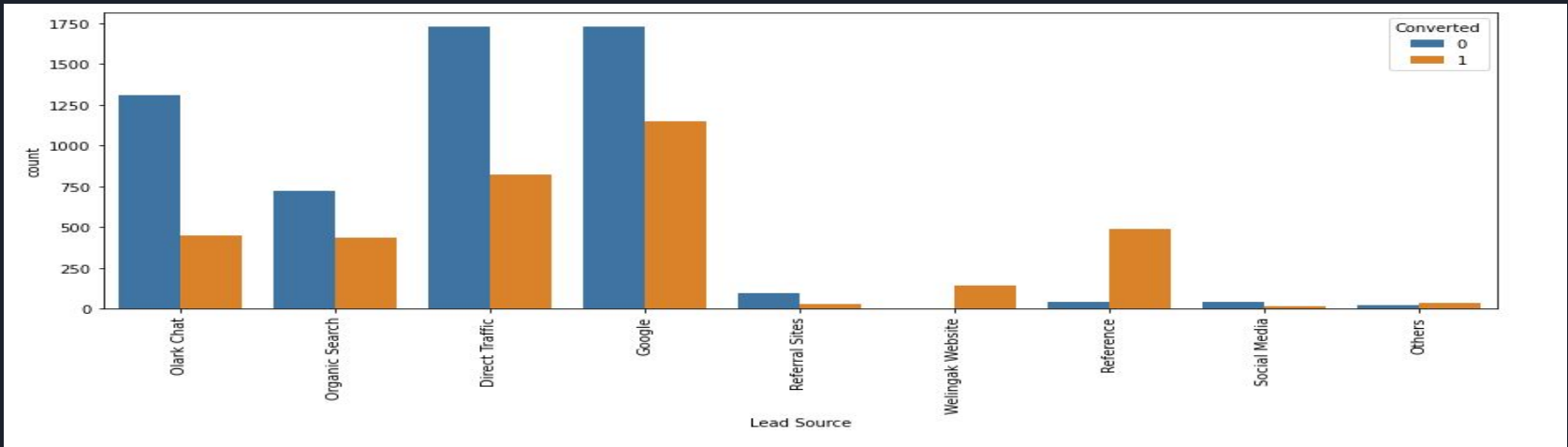
Lead Origin Analysis

- 'Landing Page Submissions' and 'API' bring a higher number of leads and see more lead conversion as well.

- Overall, "Lead Add Form" has high lead conversion ratio, but low volume and Landing page submission has high lead conversion ratio with enough volume to increase number of leads converted. Lead Import and Quick Add Form get very few leads.

- To improve overall lead conversion rate, we must improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
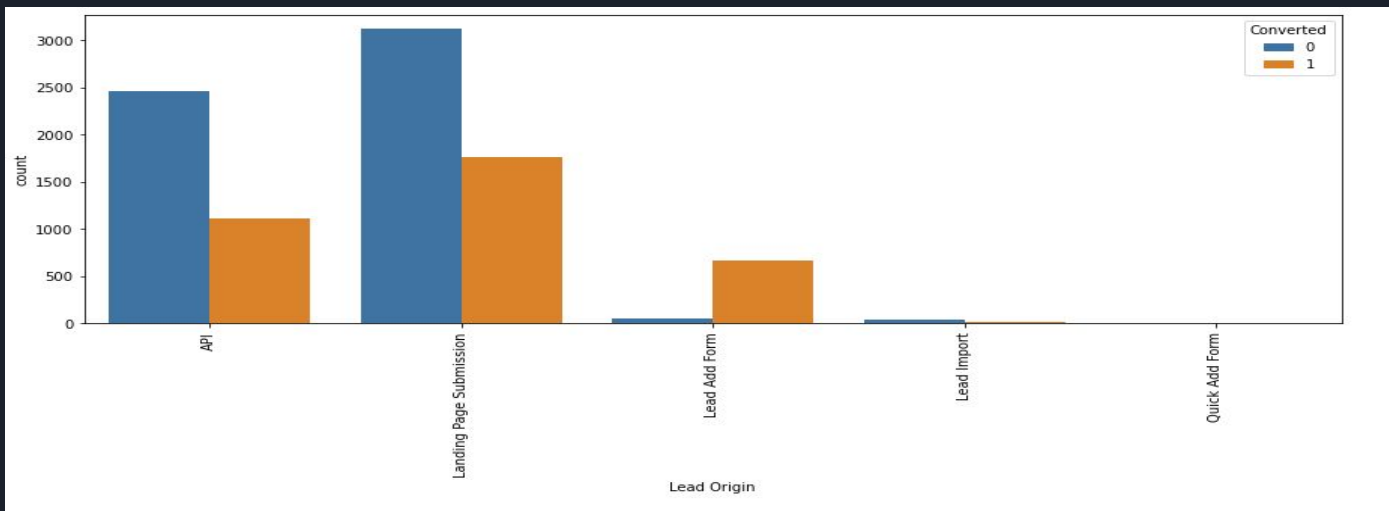
Last activity conversion Analysis

- Leads having last activity as 'SMS Sent' have the most conversion rate.

- 'Email Opened' brings maximum number of leads and has second most conversion as well.

- Focus should be more on leads where last activity will be SMS sent as they see highest conversion ratios.
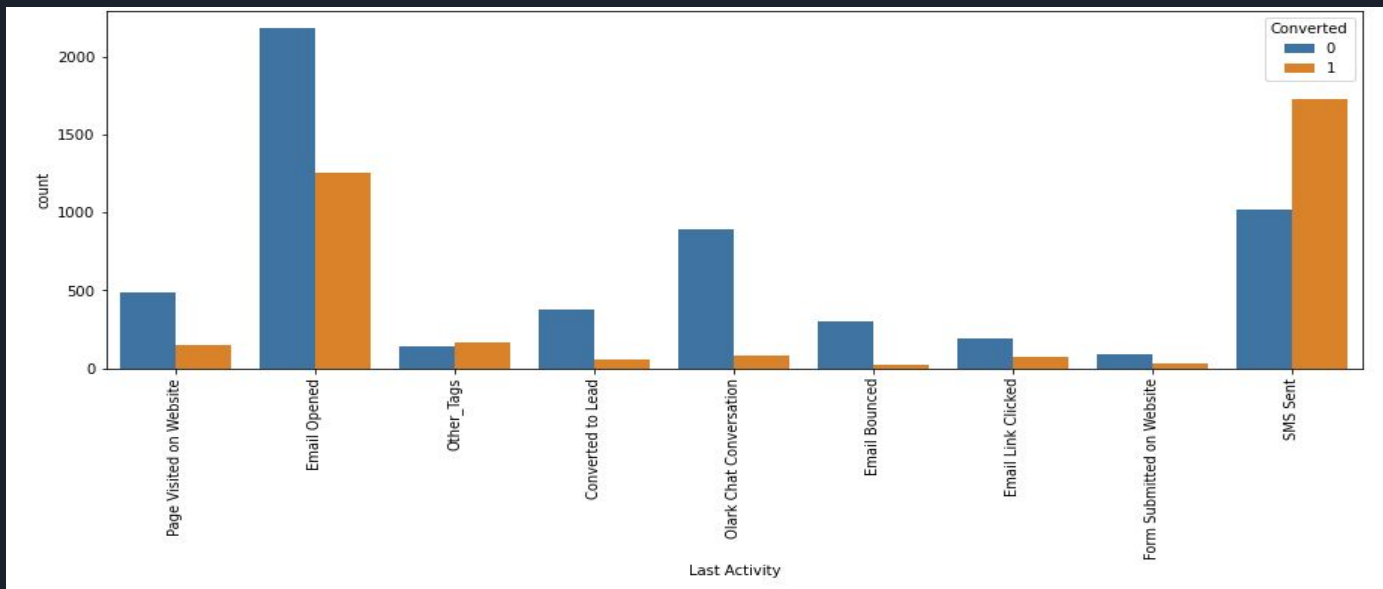
Specialization conversion Analysis

- We see that leads having different types of 'Management' specializations are more in number as well as have higher chances of getting converted.
- Hence lead score is higher for management professions and should be focused on to get more lead conversions.
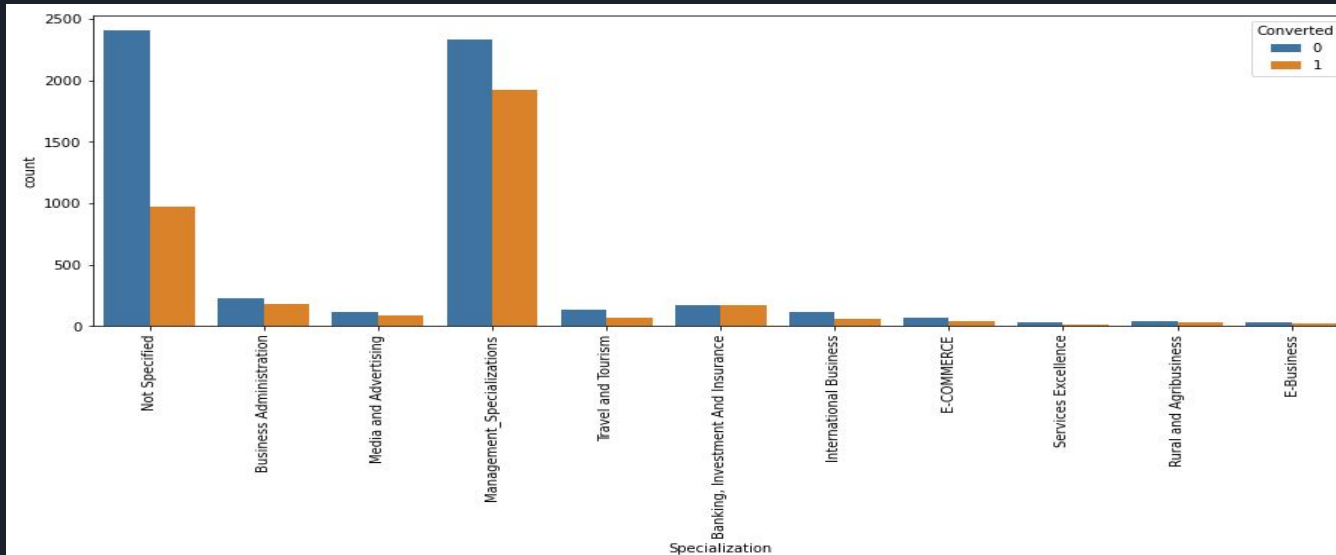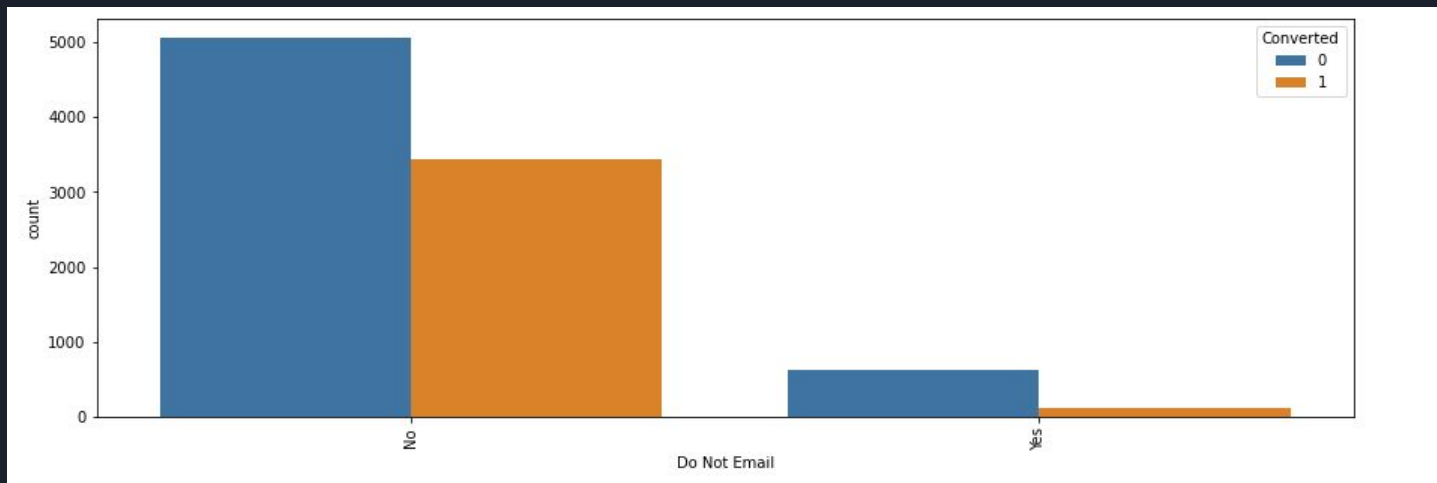
Tags vs Conversion Analysis

- 'Ringing' and 'Not Specified' tagged leads are more in number and hence efforts should be made to maximize conversion from these tags.

- However, leads tagged 'Will revert after reading mail' have highest chances of being converted followed by 'Lost To EINS', 'Closed By Horizon and 'Busy' and efforts should be made to focus more on these leads to improve conversions.
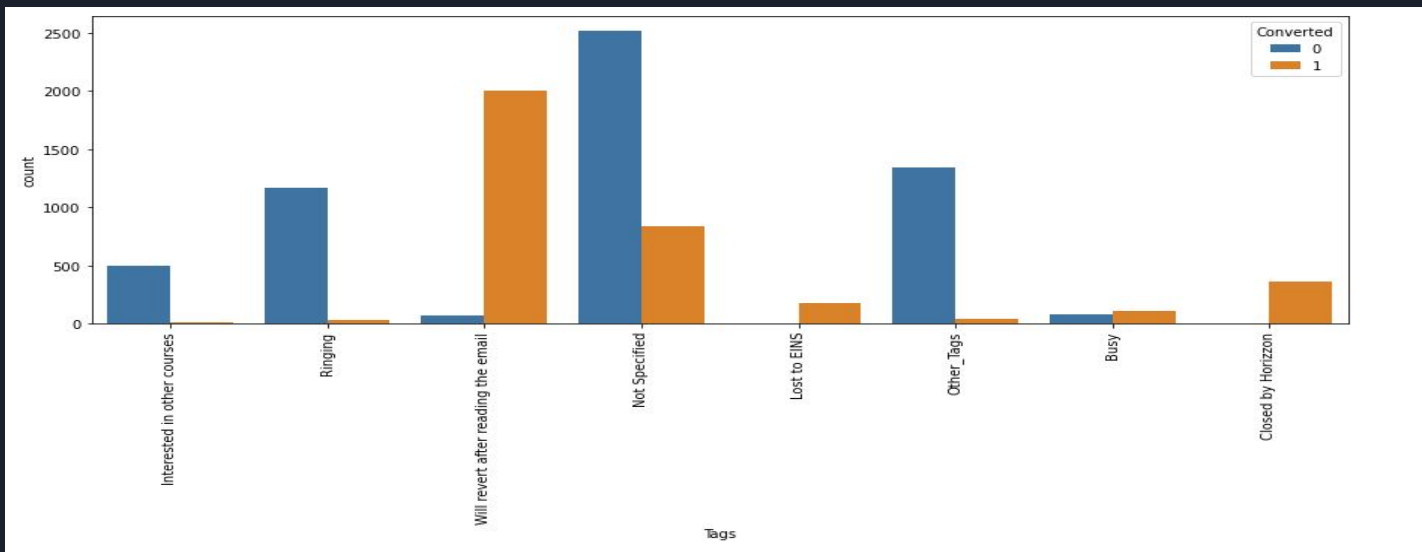
Do not Email vs Conversion Analysis

- Leads which do not object to an email communication have higher lead score than with leads which say yes to "Do not Email" option.

- More Focus should be put on pursuing the leads that say no to "Do not Email" to have a greater number of conversions.
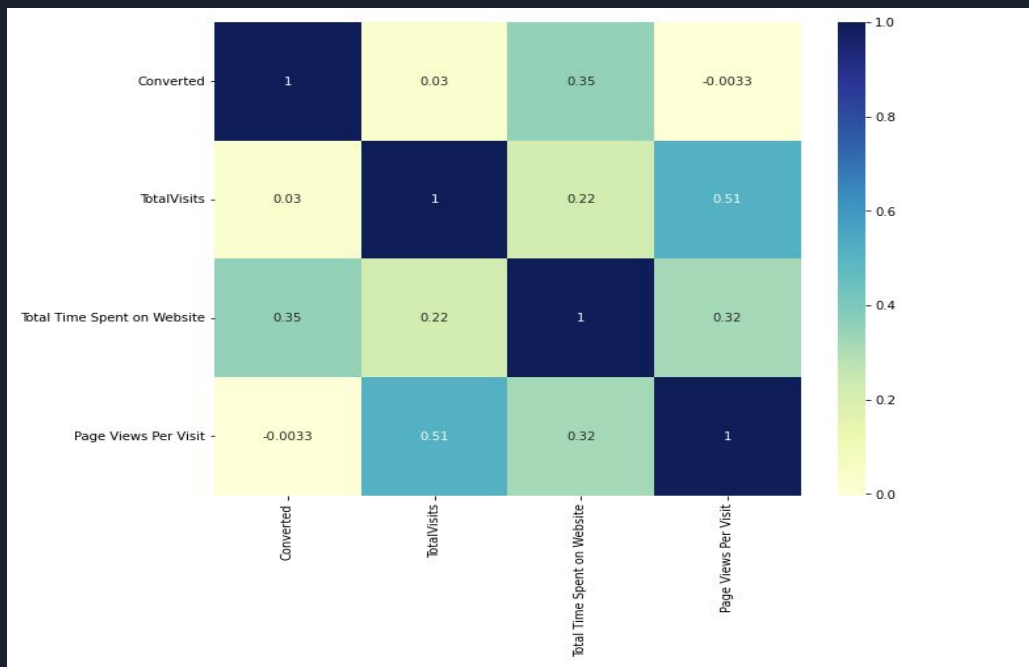
Total Time Spent On Websites vs Conversion Analysis

- People spending higher than average time are promising leads. Hence will have higher lead score.
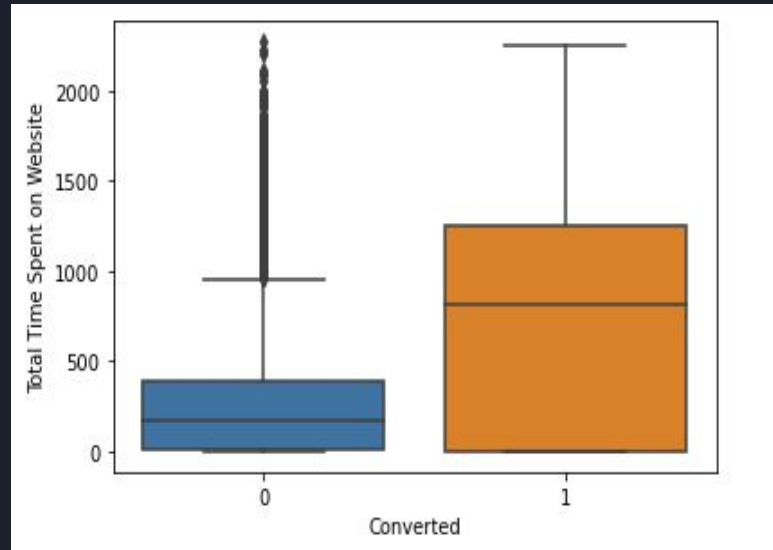
Correlation between various numerical variable using heatmap

- 'Total Visits' and 'Page Views Per Visit' have the most correlation with each other.

Model building steps and predictions

Logistic Regression Model Building

- Splitting The Dataset into train and test sets (80:20)
- Rescaling the numerical variables
- Model Building using Stats models and RFE (Total 5 models built)
- Deriving Probabilities, Predictions and Lead Score on Train Data
- Plotting confusion matrix and plotting ROC curve
- Finding Optimal Cut-Off

Making Predictions On Test Dataset With Model Built From Training Dataset
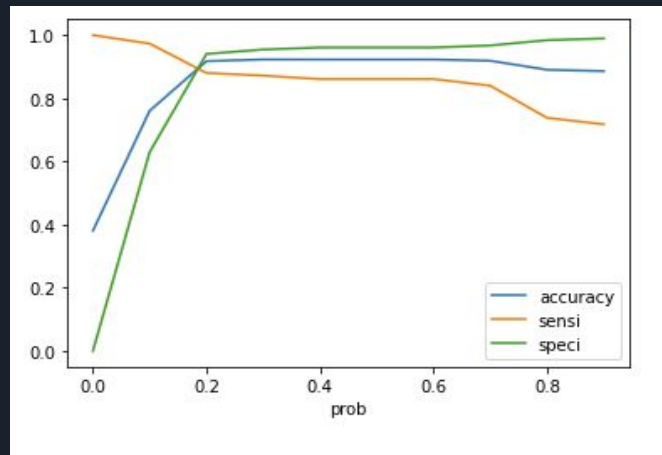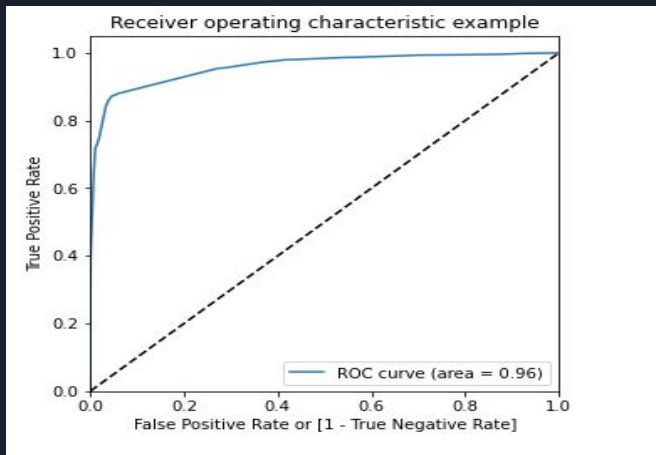
Final analysis on training data set

ROC curve should be a value closer to 1 for a good model. We have got a value of 0.96 which is extremely good.

Some important statistics of our model:

Accuracy: 91.75%

Sensitivity: 88%

Specificity: 94.04%

Final Observations

Important variables that need to be focused on with high lead score include

- Lead Source_Direct Traffic
- Last Activity_Email Bounced
- Last Activity_Olark Chat Conversation
- Tags_Busy
- Tags_Not Specified
- Tags_Ringing
- Tags_Will revert after reading the mail
- Last Notable Activity_SMS Sent

Thank you

SUBMITTED BY :-

Atul Tiwari

Piyush Arya

Kajal Shah