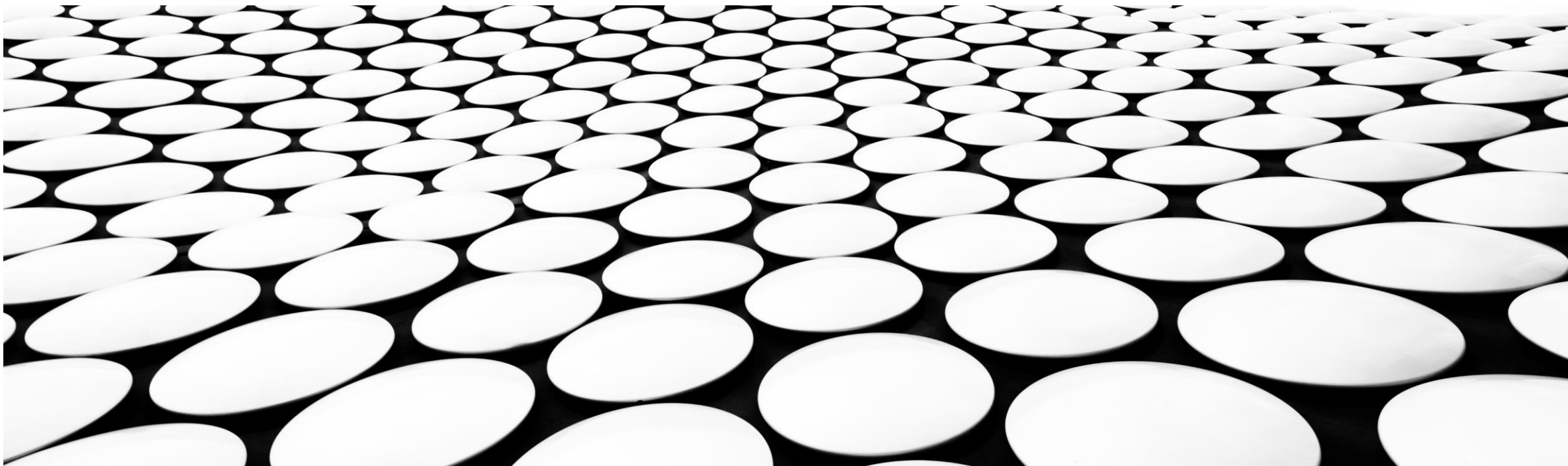# LEAD SCORING CASE STUDY

JAYALEKSHMI K

MOHAN SWARUP VANA

SYED MOHAMMAD RAZA NAQVI

AGENDA

To build a model 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.

PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. Company makes its courses available on various several websites and search engines and once people browse through, they fill up a form providing their email address or phone number, they are classified to be a lead. Also company gets leads through past referrals. Once the leads are acquired, there is a 30% chance of them getting converted. Company wants to find hot leads (most potential leads)

GOALS OF THE CASE STUDY

Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

APPROACH

*Reading and Understanding the data

*Data Cleaning and Manipulation

*Exploratory Data Analysis

*Data Transformation

*Dummy Variable Creation

*Classification Technique

*Test-Train Split

*Validation of models -Standard scaling used

*Model Building

*Implemented learning to test model as well and calculated conversion probability.

*Found top 3 features which can contribute to lead conversion

LEAD ORIGIN VS CONVERTED
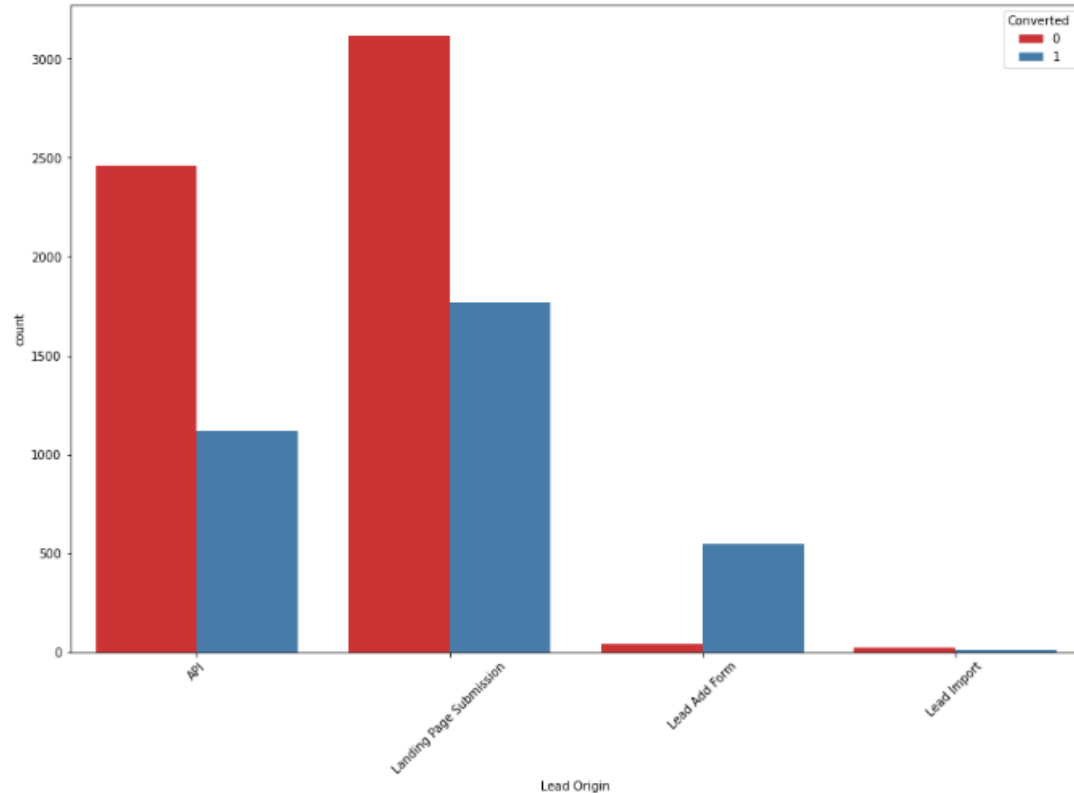
- API and Landing Page Submissions have a higher number of leads as well as conversions.
- Lead Add Form has high conversion rate(around 90%) but lead count is not as high
- Lead Import and Quick Add Form has few leads.
- To improve the overall lead conversion rate, we need to improve the lead conversion of API and Landing Page Submission and generate more leads from Lead Add Form
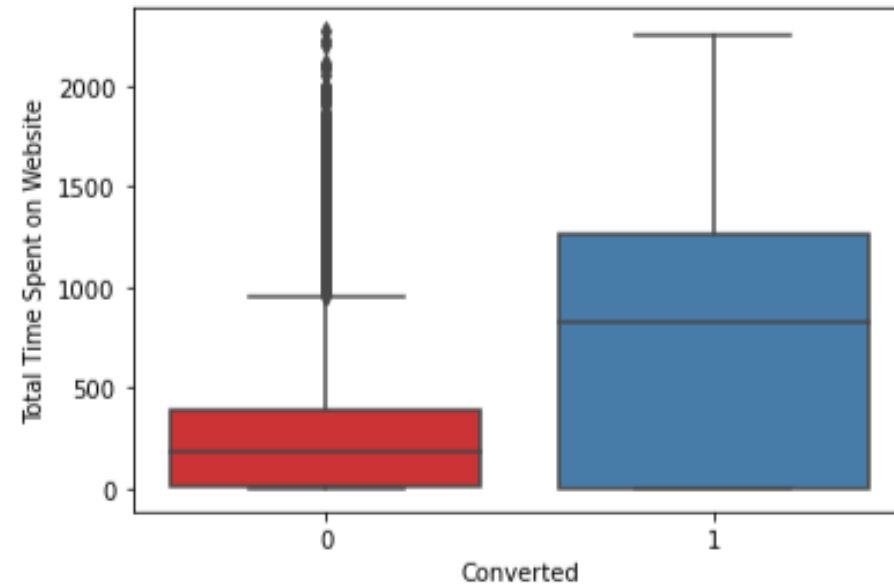
LEAD SOURCE VS CONVERTED
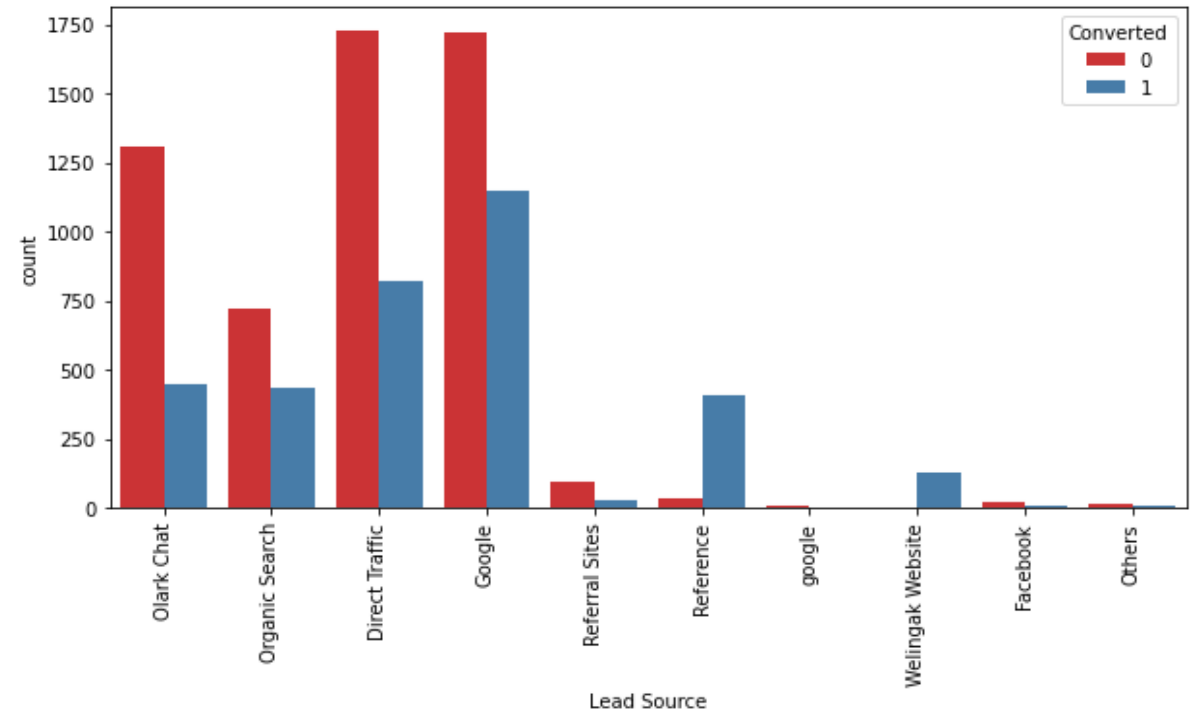
- Google Searches has the high conversion rates followed by direct traffic, organic search and the Olark chat.
- Olark chat has a high rejection rate than the conversion rate
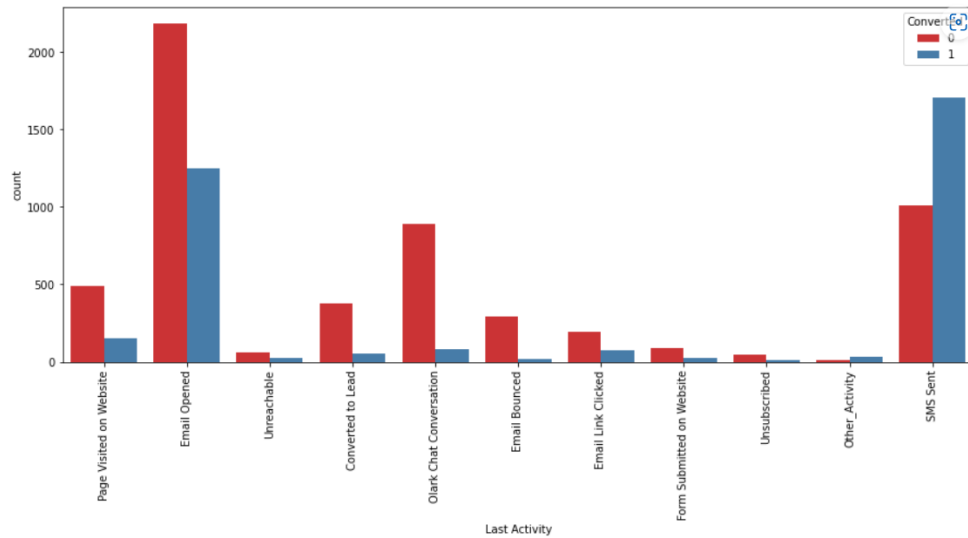
TOTAL VISITS VS CONVERTED

TOTAL TIME SPENT ON WEBSITE VS CONVERTED
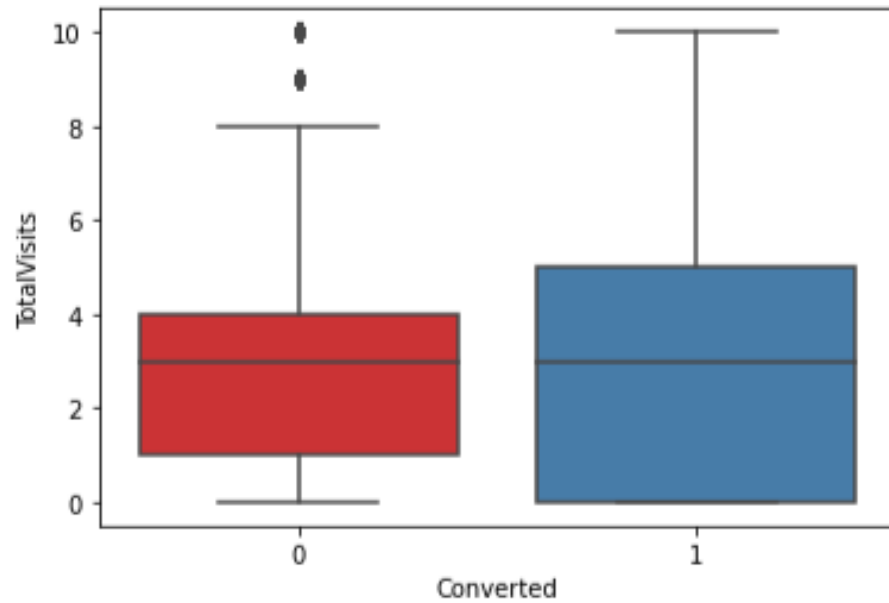
* Median for converted and not converted leads are the same.

*Leads spending more time on the website are more likely to be converted.
*Website should be made more engaging to spend more time.

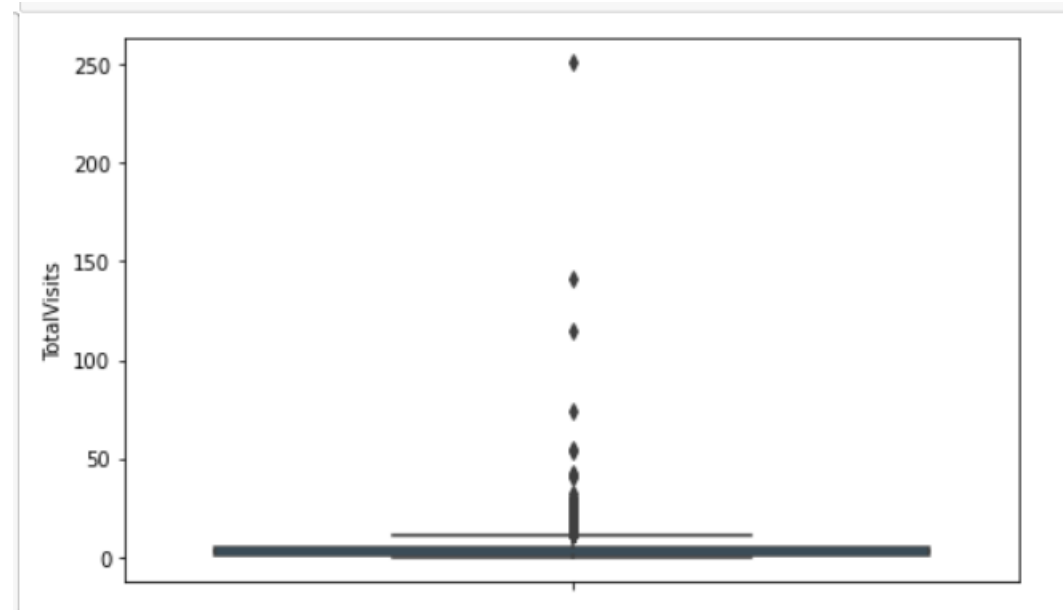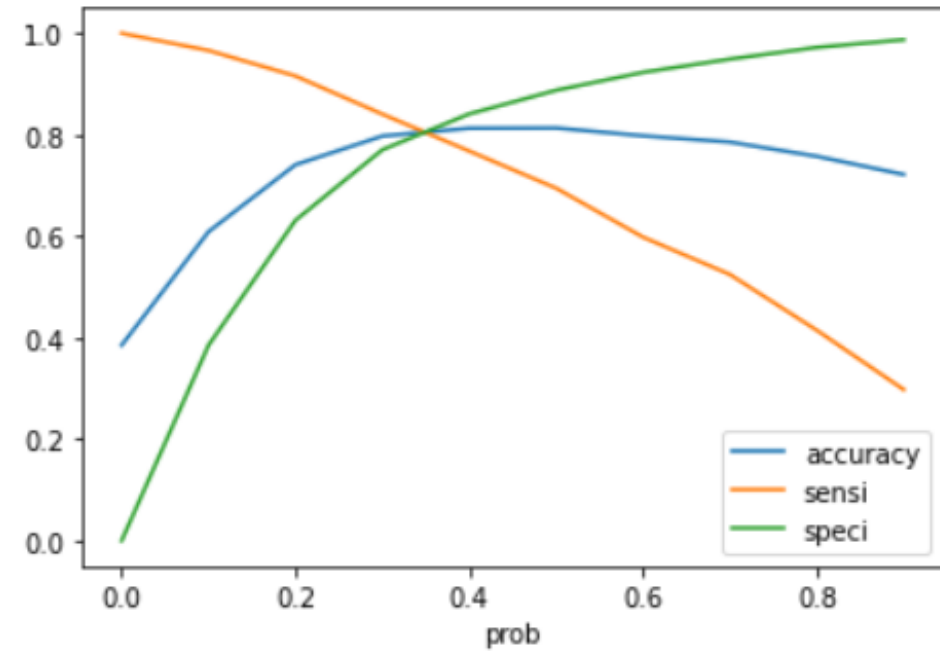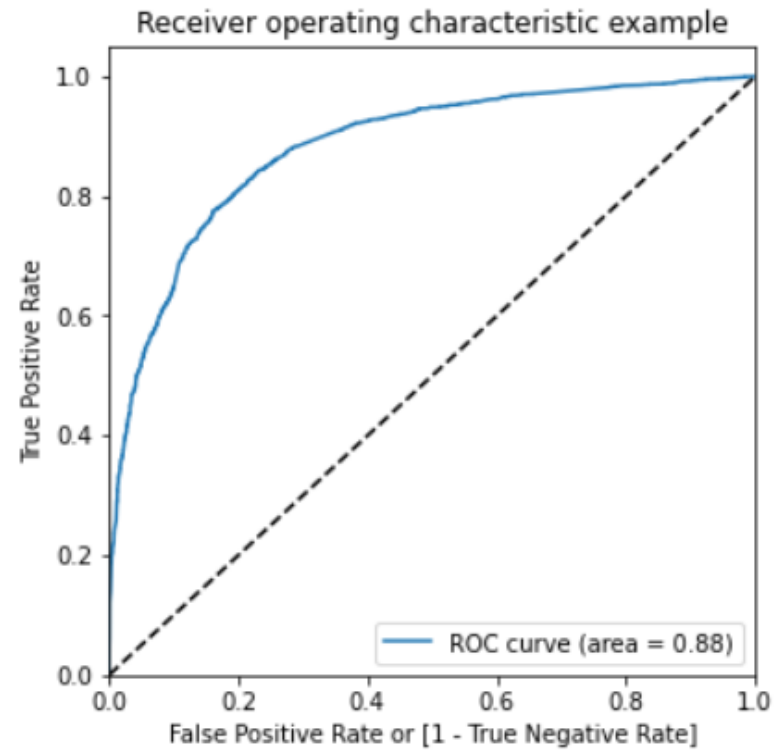# LAST ACTIVITY VS CONVERTED

FINDING OUTLIERS IN TOTAL VISTS

* Most of the lead have their Email opened as their last activity.
*Conversion rate for last activity as SMS Sent is almost 60%.

ROC CURVE
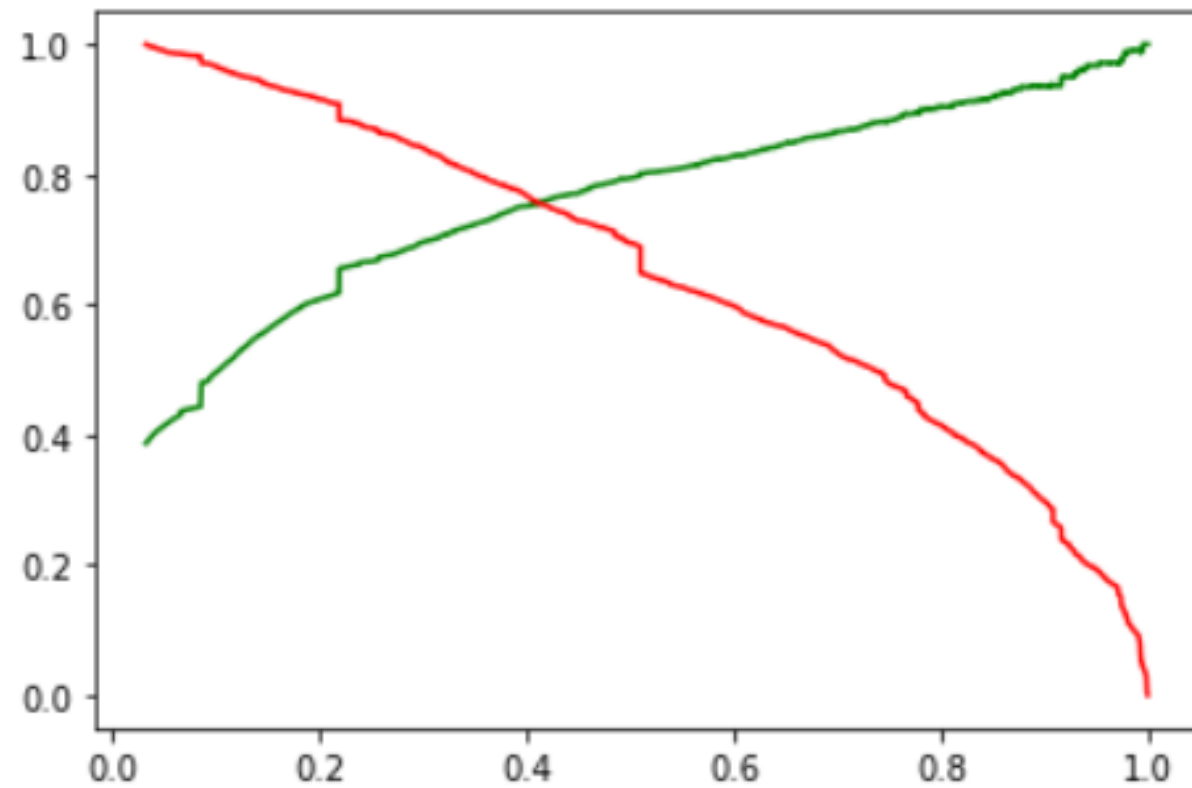
PRECISION AND RECALL CURVE

CONCLUSION

.The lead score calculated in the trained set of data shows the Conversion rate on the final predicted model is around 80%

.Perfect value of Sensitivity help us to find "hot leads"

.Top 3 important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted in decreasing order are :

#1)Lead Source_Welingak Website
#2)Lead Source_Reference
#3)What is your current occupation_Working Professional