Lead Score - Case Study

Submitted by

Abhinav Patel

Parvathy J

Parvathy Suresh

Problem statement

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%.

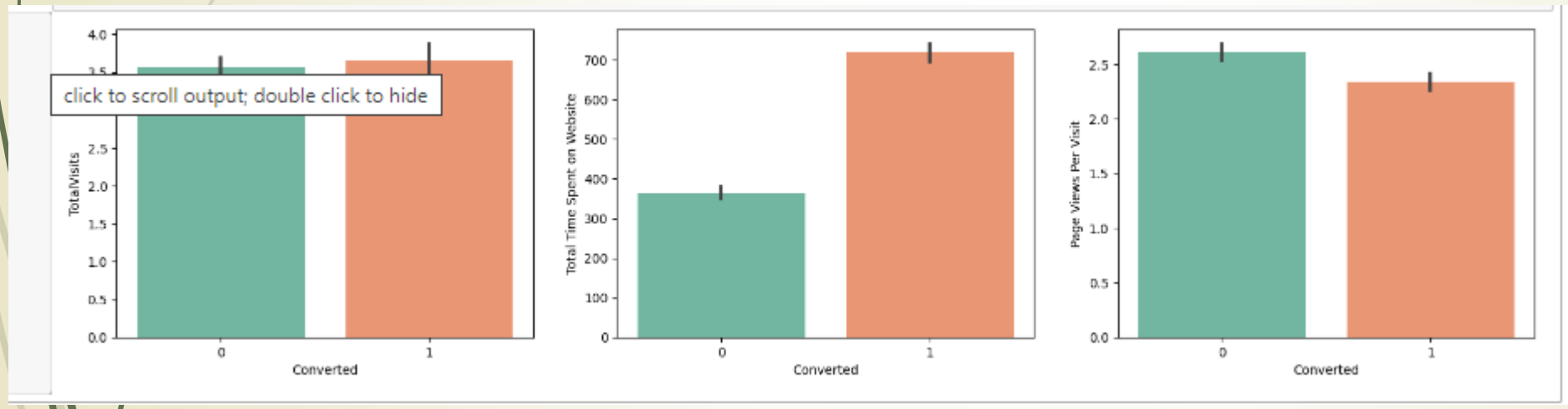
The steps :

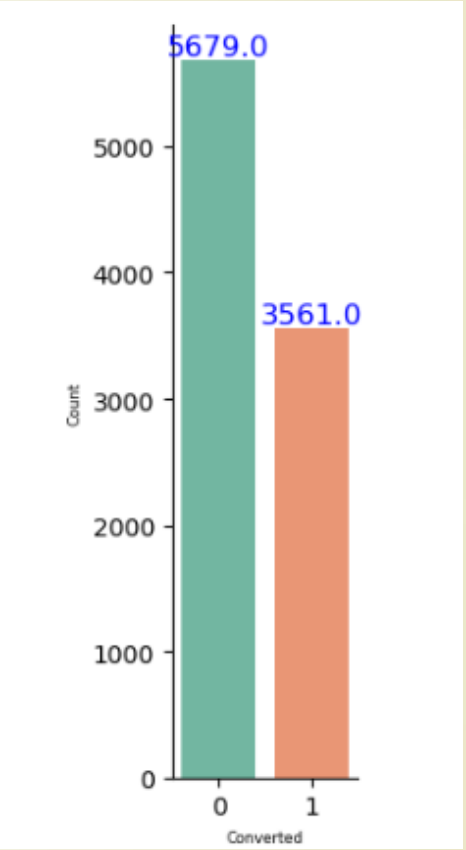
1. Read and understand the data
2. Clean the data
3. Prepare the data for Model Building
4. Model Building
5. Model Evaluation
6. Making Predictions on the Test Set

Exploratory Data Analysis

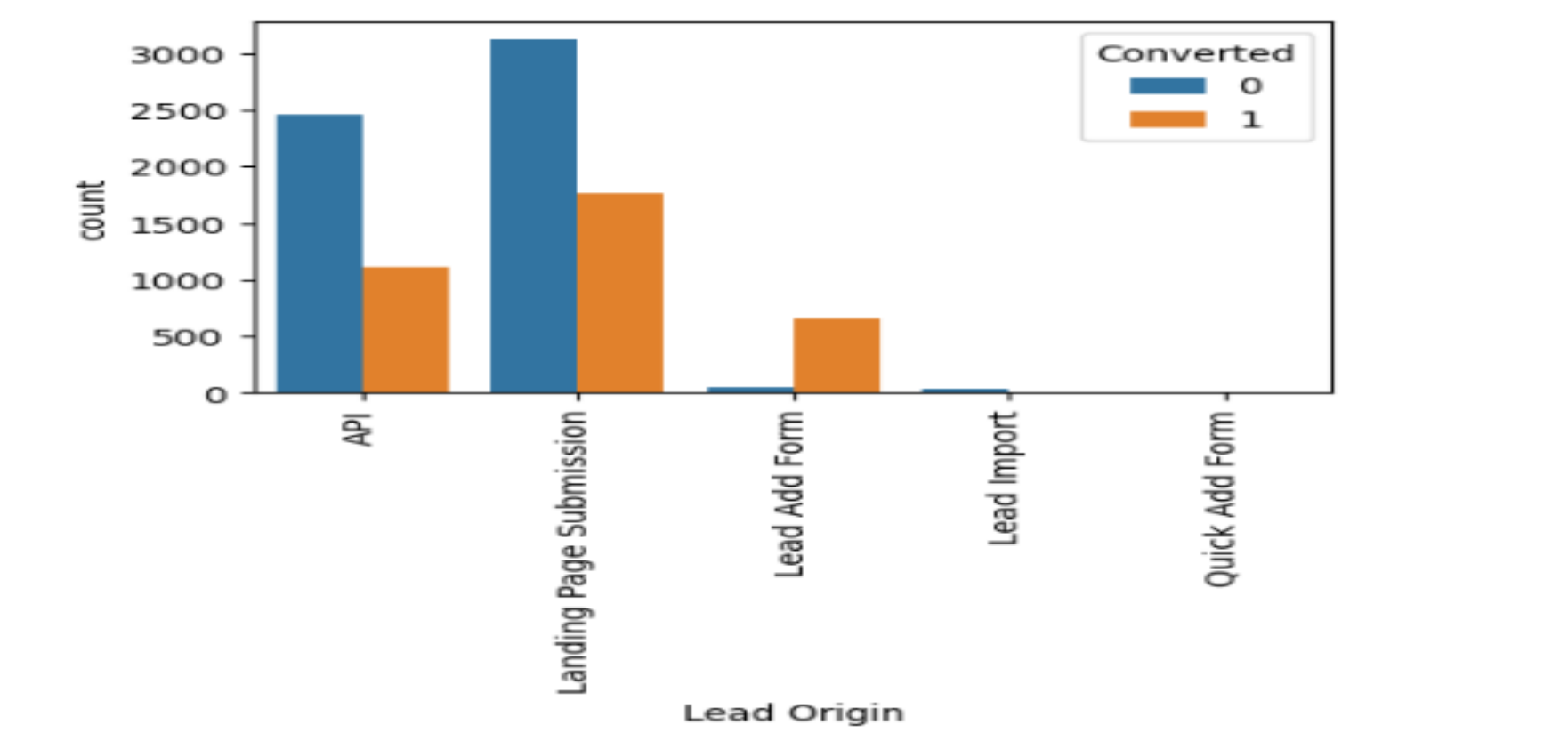
We have around 39% Conversion rate in Total

The conversion rates were high for Total Visits, Total Time Spent on the Website and Page Views Per Visit

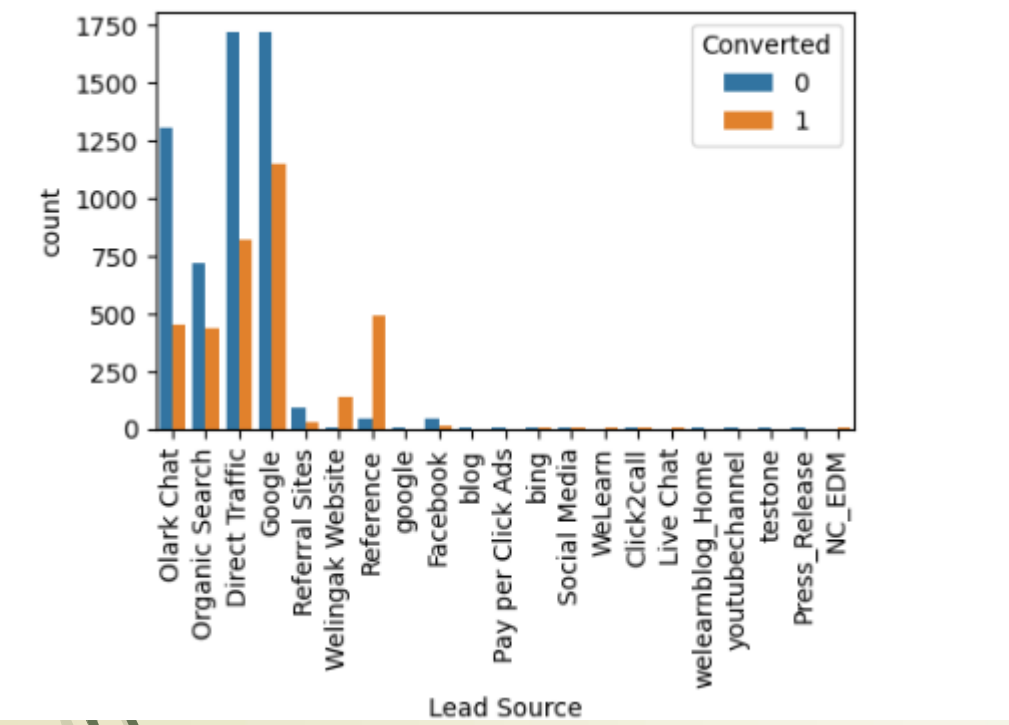




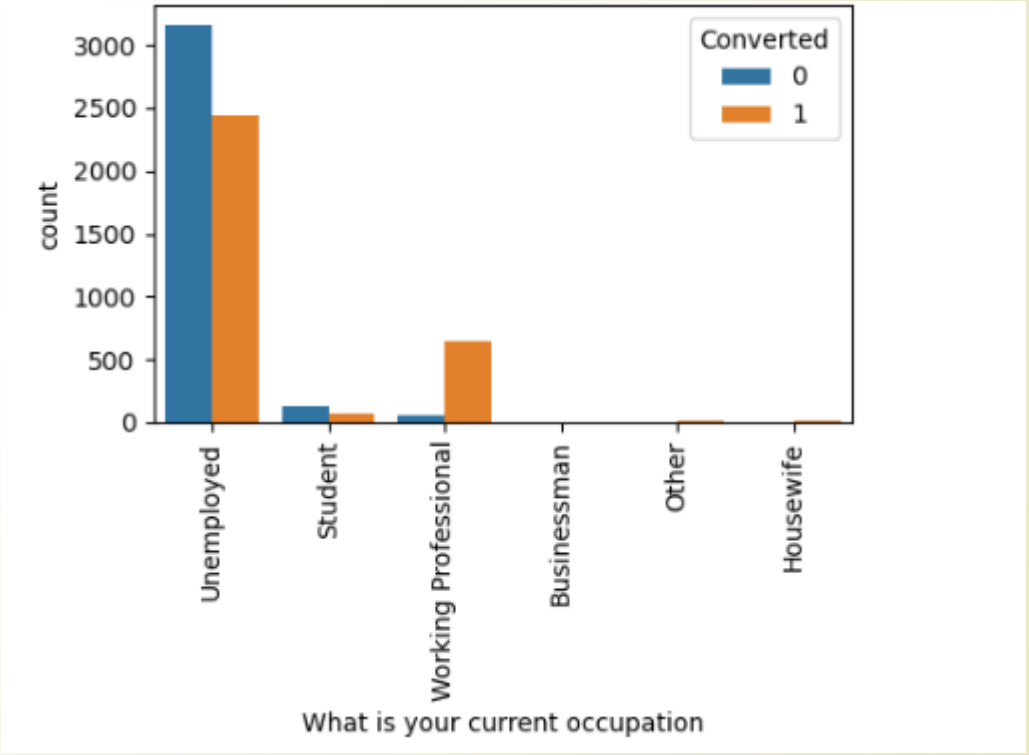
In Lead Origin, maximum conversion happened from Landing Page Submission



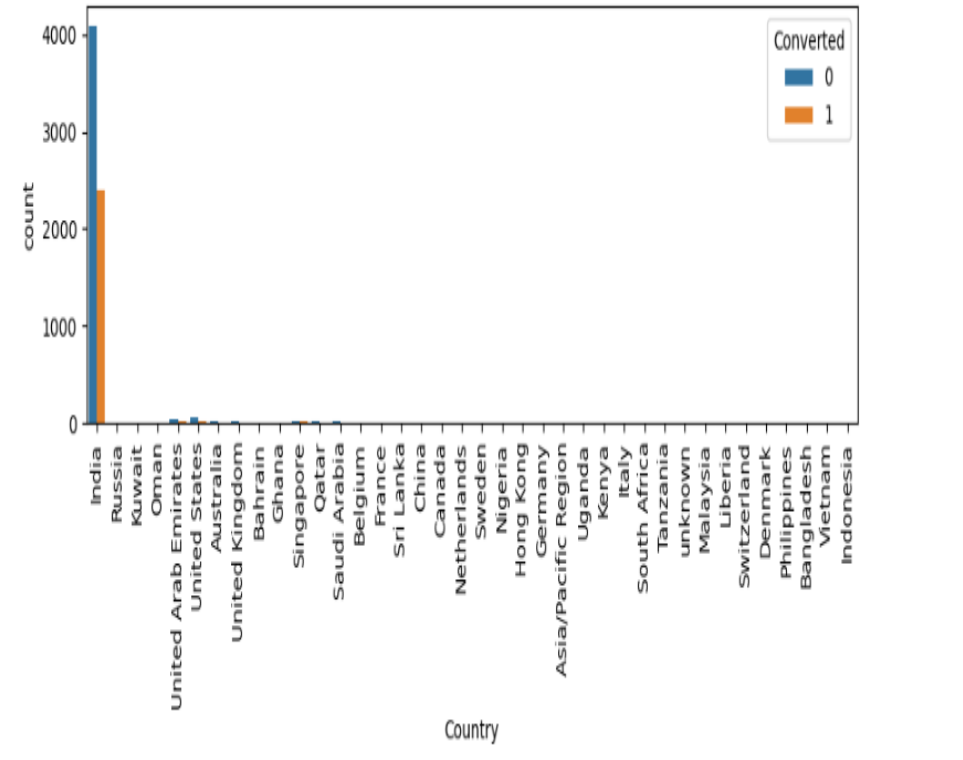
Major conversion in the lead source is from Google



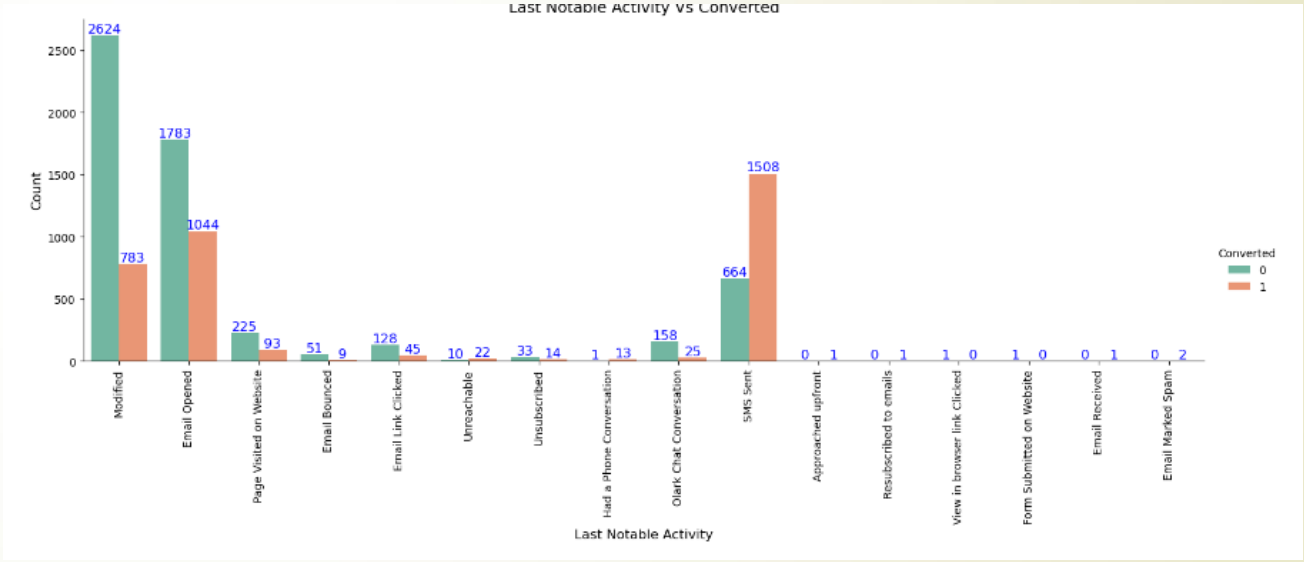
More conversion happened with unemployed people



The majority of the conversions are from India

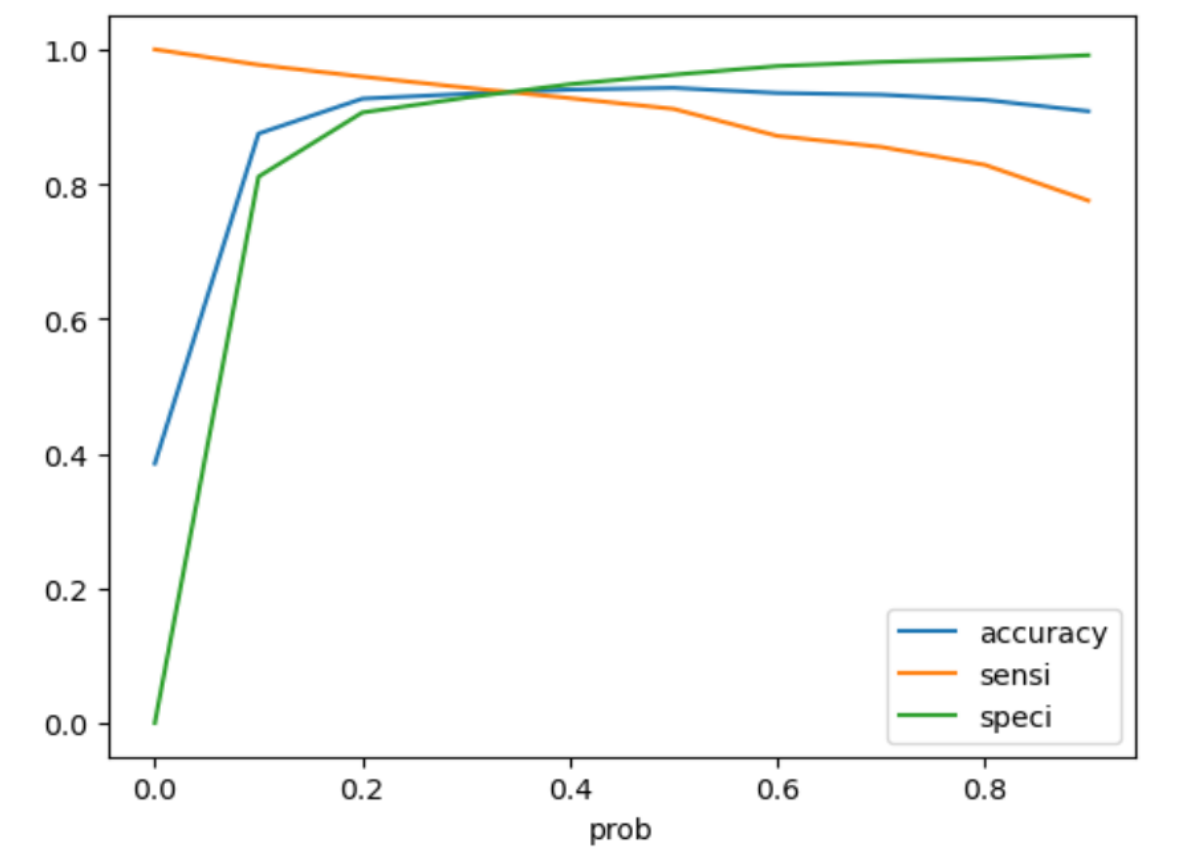


The last Activity value of SMS Sent' had more conversion.



Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.42 based on Accuracy, Sensitivity and Specificity



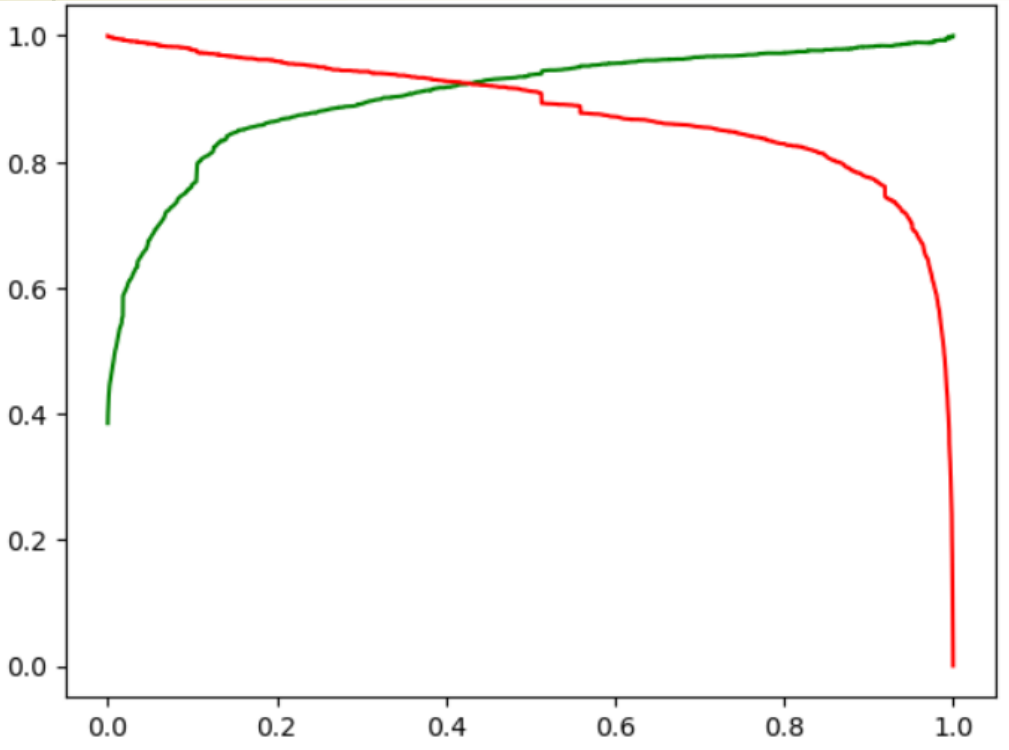
Confusion Matrix

3758	147
216	2230

- Accuracy -94%
- Sensitivity - 91 %
- Specificity - 96 %

Model Evaluation- Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.42 based on Precision and Confusion Matrix Recall



Confusion Matrix

3703	202
177	2269

- Accuracy - 94 %
- Precision - 92 %
- Recall – 93%

Model Evaluation – Sensitivity and Specificity on Test Dataset

- Accuracy -93%
- Sensitivity - 91 %
- Specificity - 93 %

Confusion Matrix

1617	117
87	902

Conclusion

While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut-off based on
Sensitivity and Specificity for calculating the final prediction.

- The top 3 variables that contribute for lead getting converted in the model are

a) Total Time Spent on Website
b) Total Visits
c) Page Views Per Visit