Lead Score Case Study

Submitted by : - Avinash Patra
- Ashwani Gupta

Lead Score Case Study

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%.

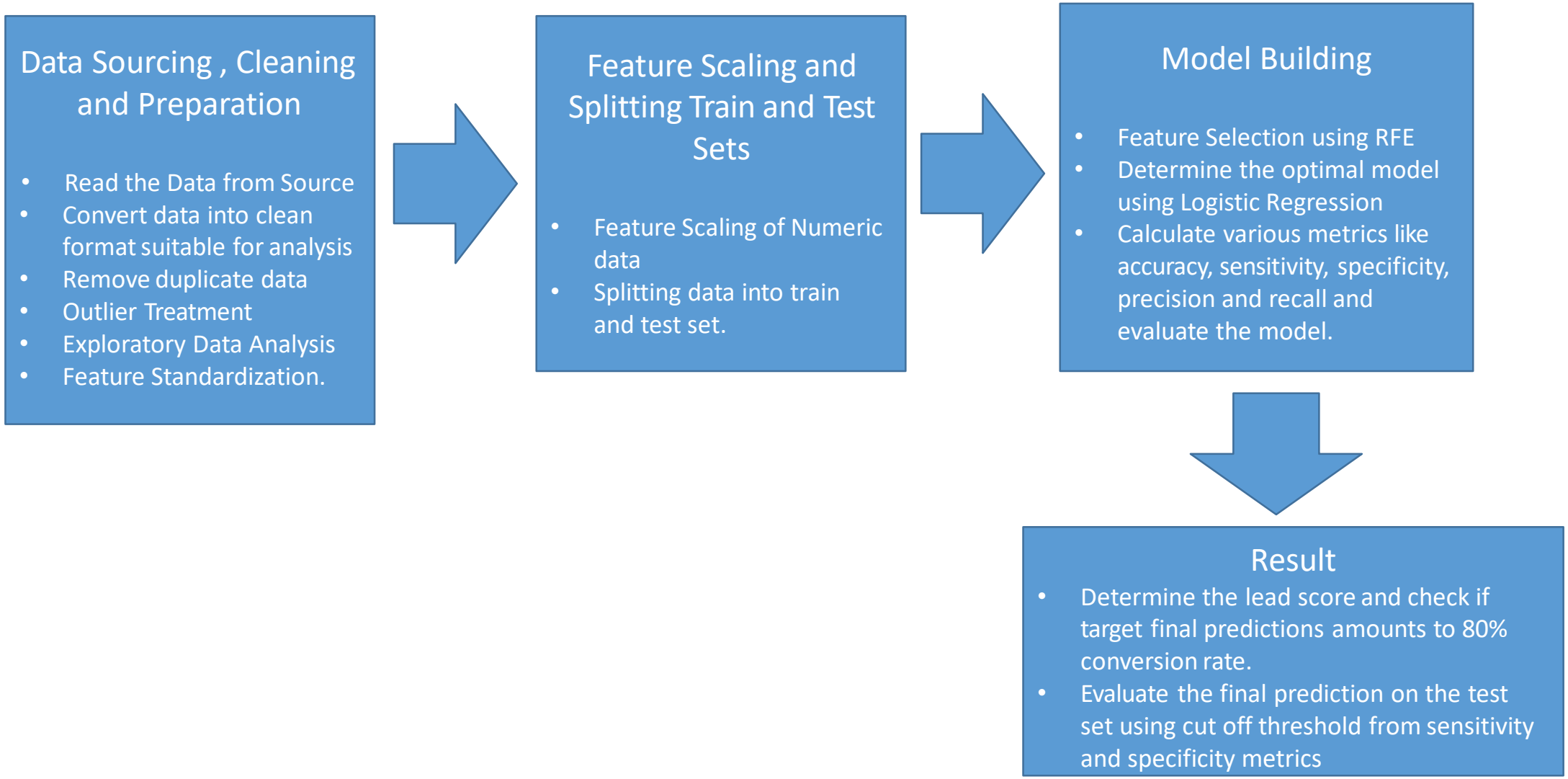
Business Goal:

- X Education requires assistance in identifying promising leads with a high likelihood of conversion to paying customers.
- A model is sought where each lead is assigned a lead score, with higher scores indicating a greater chance of conversion.
- The goal is to prioritize leads based on their lead scores, ensuring that those with higher scores are more likely to convert.
- The CEO has specified a target lead conversion rate of approximately 80%, providing a benchmark for the model's effectiveness.

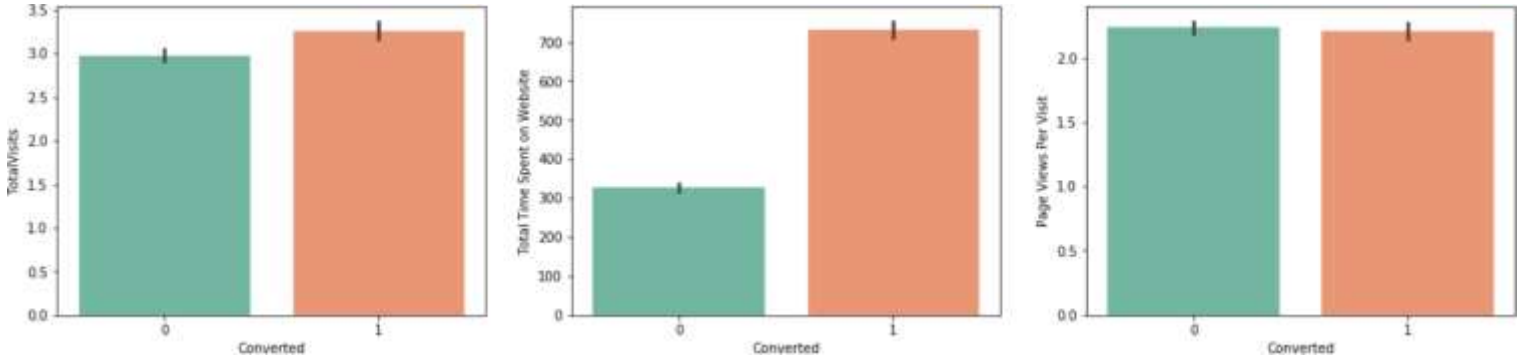
Strategy

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

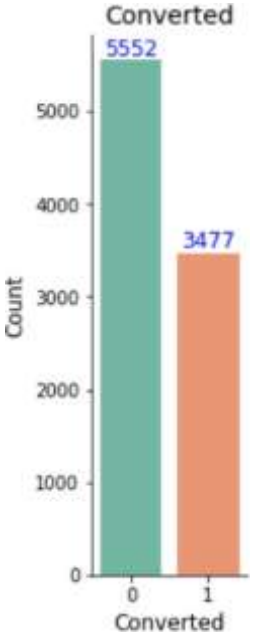
Methodology



Exploratory Data Analysis

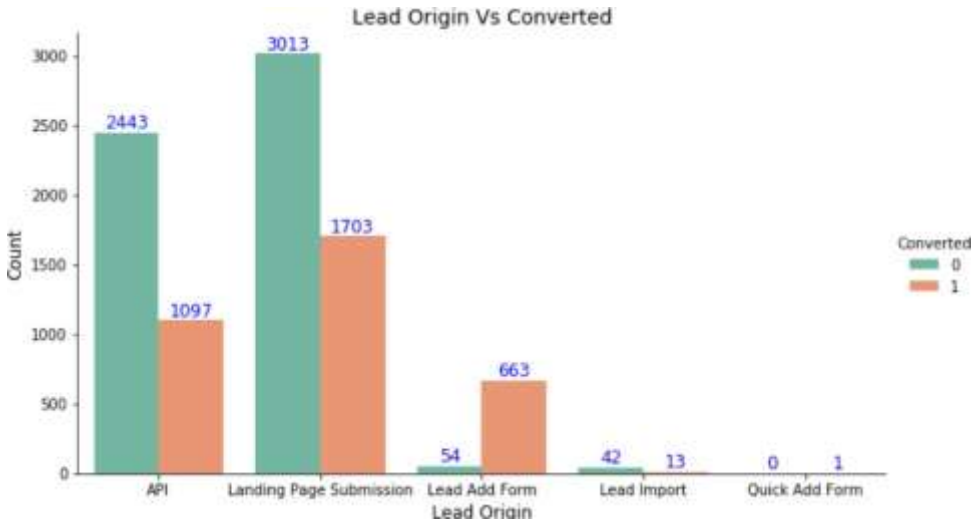


The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

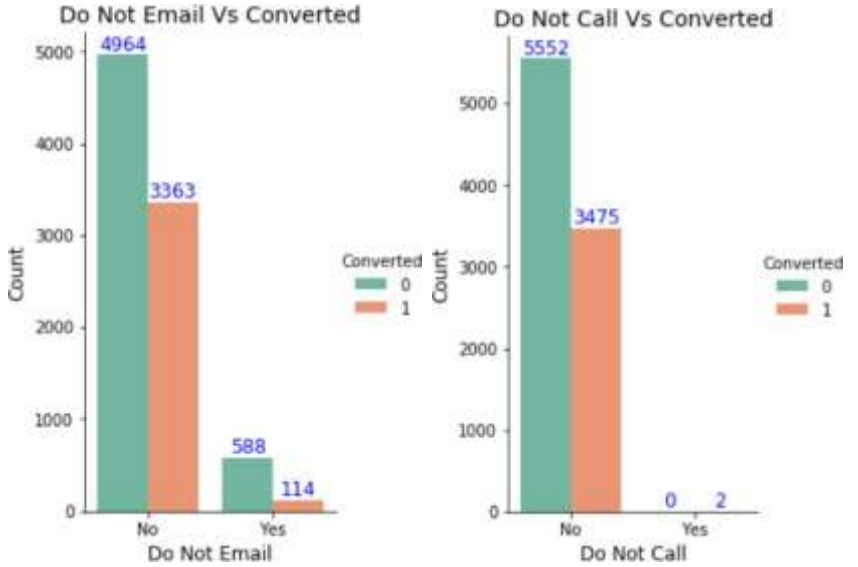


We have around 39% Conversion rate in Total

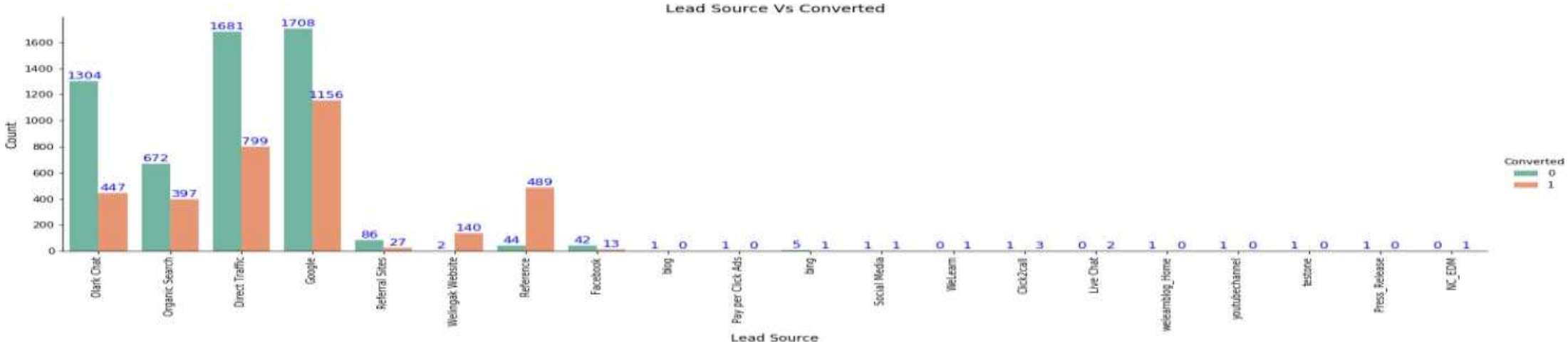
.



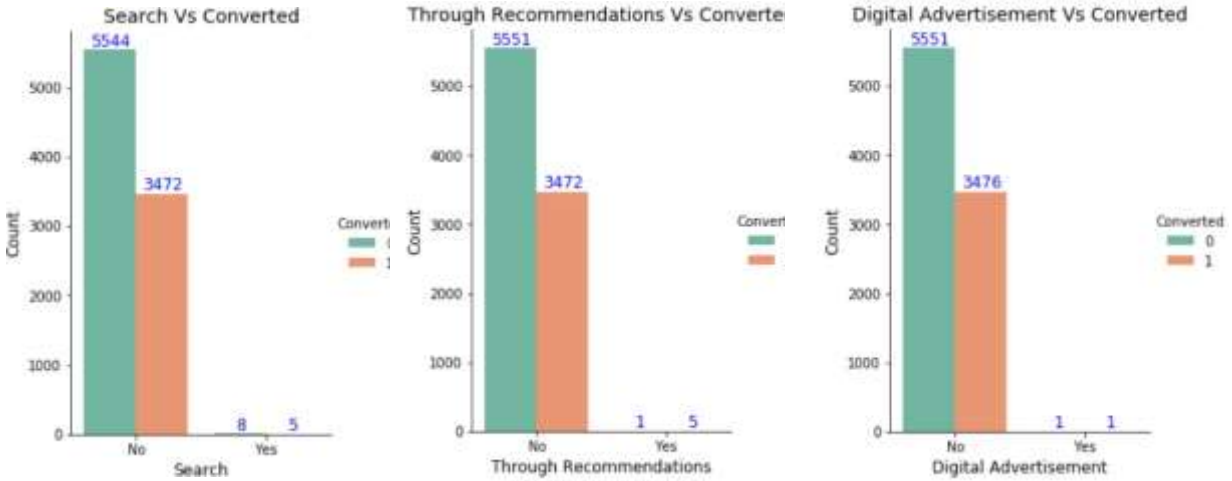
In Lead Origin, maximum conversion happened from Landing Page Submission



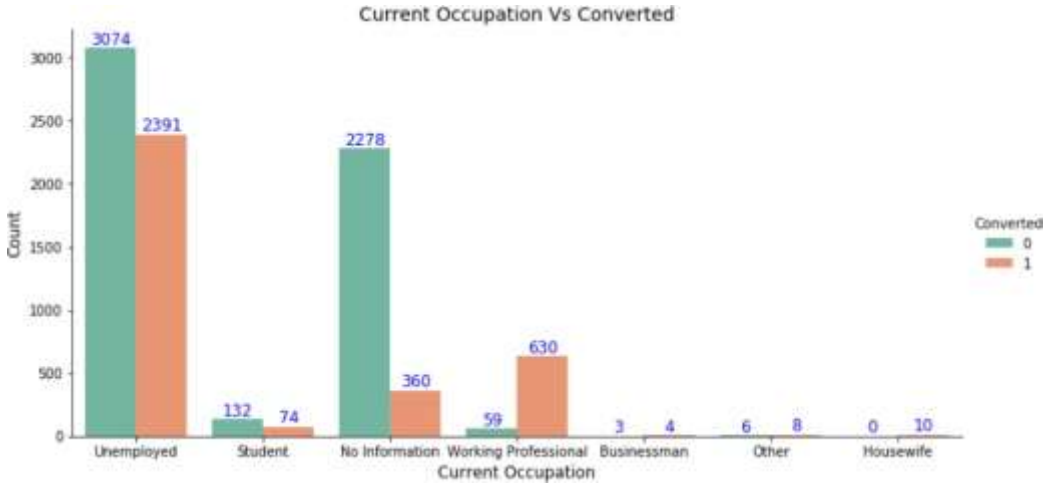
Major conversion has happened from Emails sent and Calls made



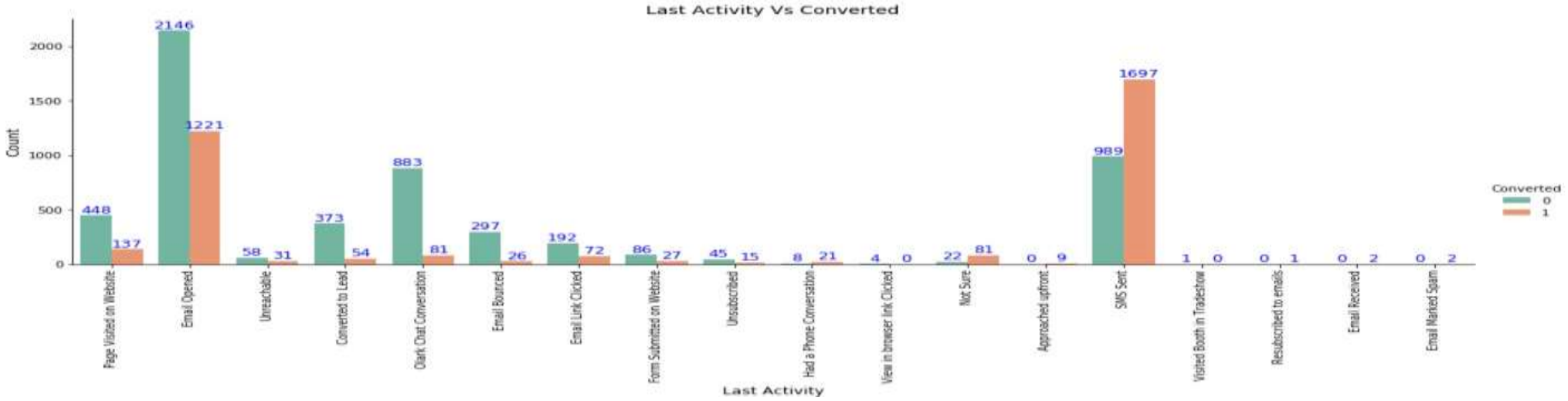
Major conversion in the lead source is from Google



Not much impact on conversion rates through Search, digital advertisements and through recommendations



More conversion happened with people who are unemployed

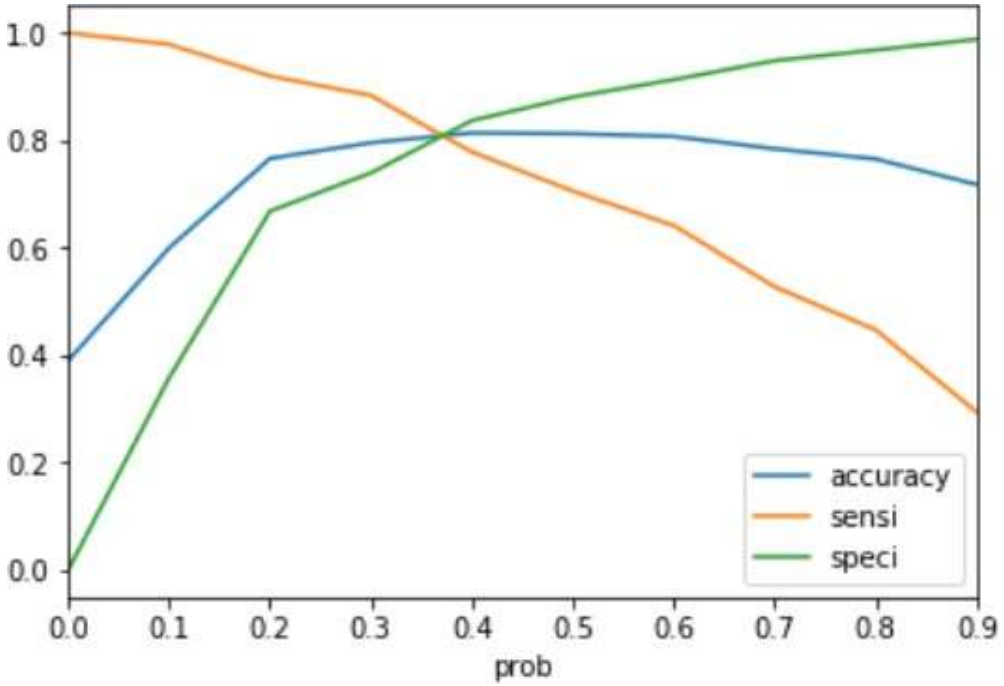


Last Activity value of SMS Sent' had more conversion.

Variables Impacting the Conversion Rate :

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

Model Evaluation - Sensitivity and Specificity on Train Data Set



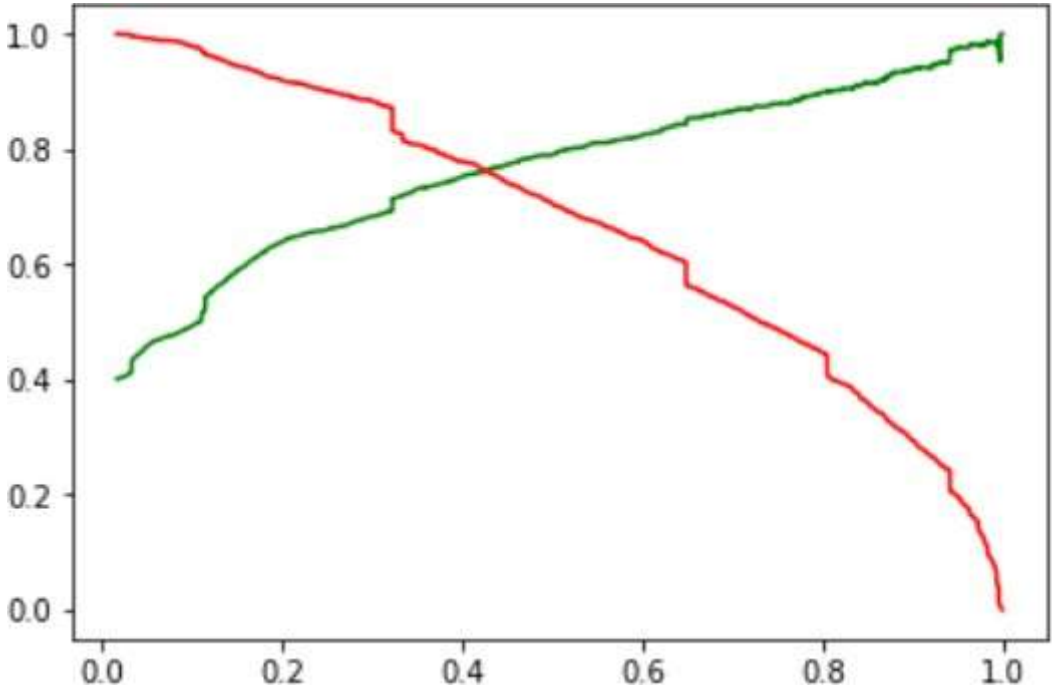
The graph depicts an optimal cut off of 0.37 based on Accuracy,
Sensitivity and Specificity

Confusion Matrix

3161	697
974	1965

- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset



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

Confusion Matrix

3397	461
725	1737

- Precision - 79 %
- Recall - 71 %

Conclusion

- The model's performance was assessed using both Sensitivity-Specificity and Precision-Recall metrics.
- Optimal cutoff for predictions was determined based on Sensitivity and Specificity considerations.
- Test set metrics, including Accuracy (81%), Sensitivity (79%), and Specificity (82%), closely mirror those calculated on the trained set.
- The lead score, indicating the conversion rate based on the final predicted model, is approximately 80% in the training set.
- Overall, the model demonstrates favorable performance across various evaluation criteria.