

LEAD SCORING CASE STUDY

Problem Statement :

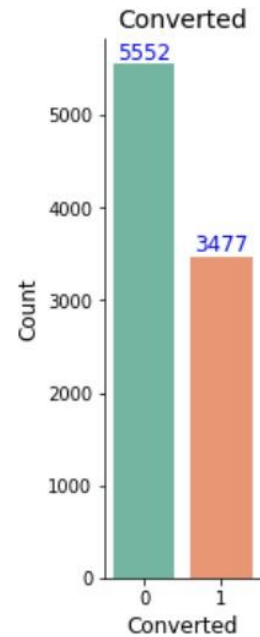
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form 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:

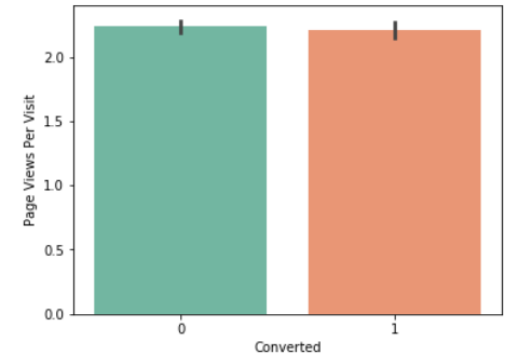
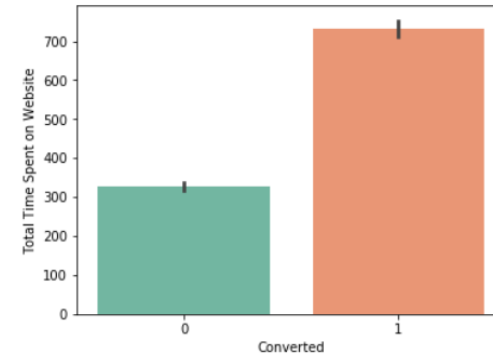
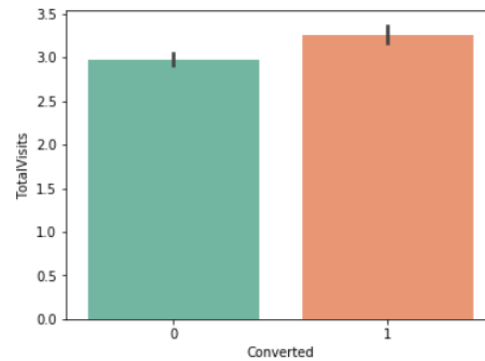
- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Exploratory Data Analysis

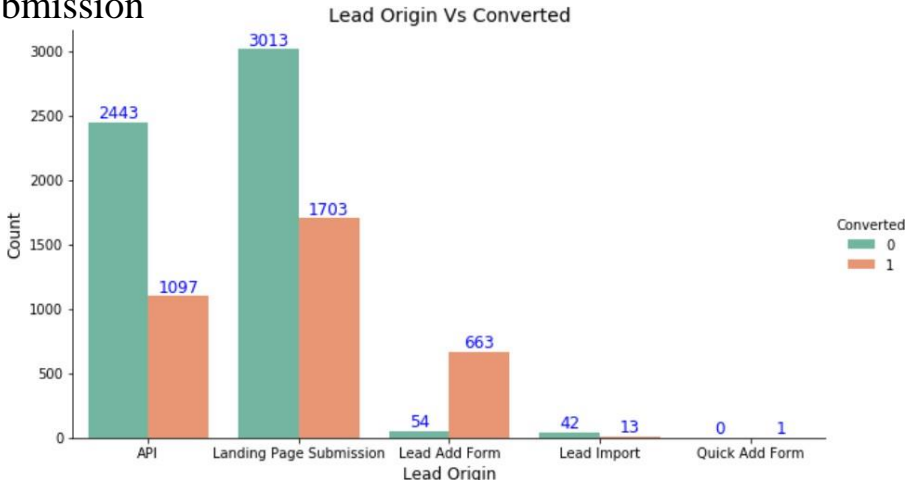
We have around 39% Conversion rate in Total



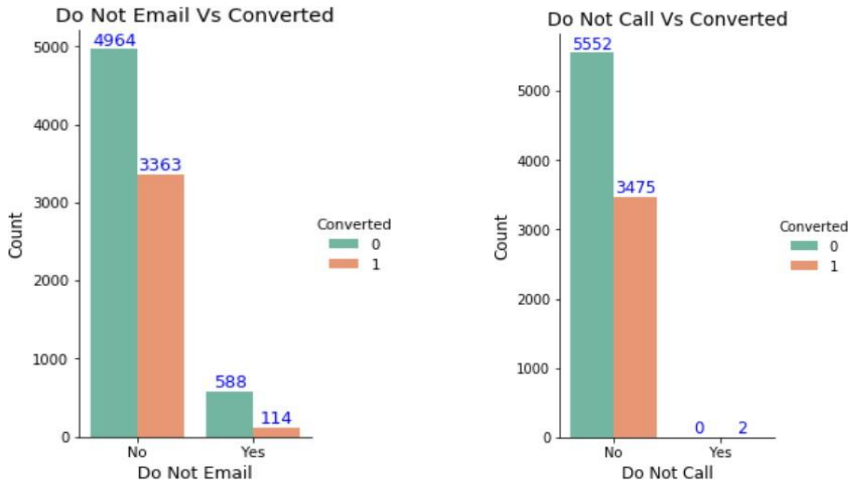
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit



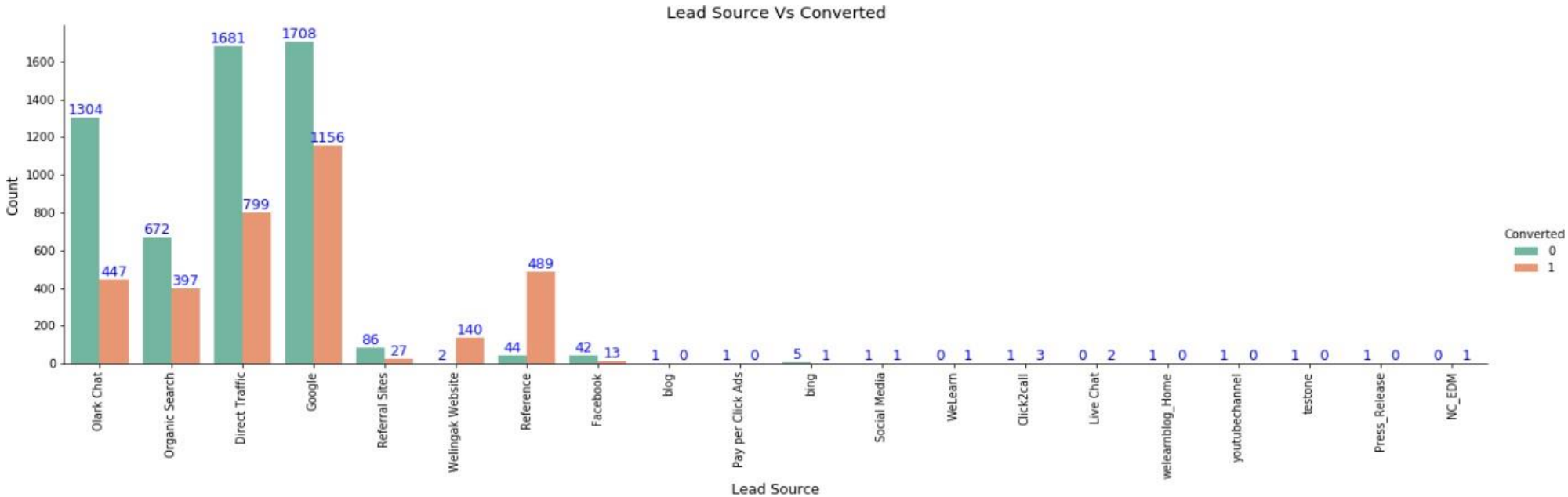
In Lead Origin, maximum conversion happened from Landing Page Submission



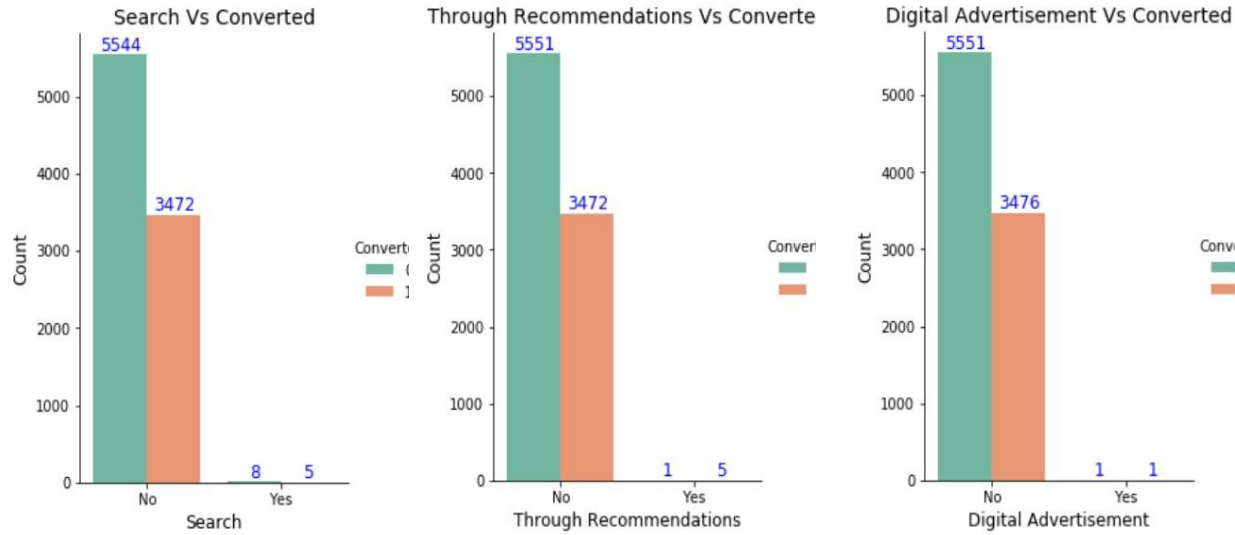
Major conversion has happened from Emails sent and Calls made



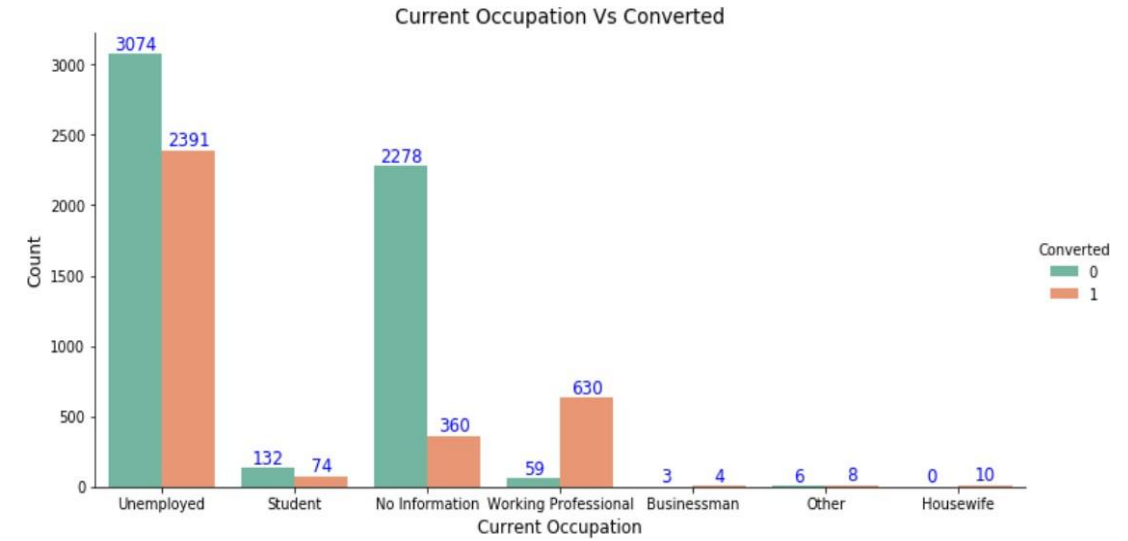
Major conversion in the lead source is from Google



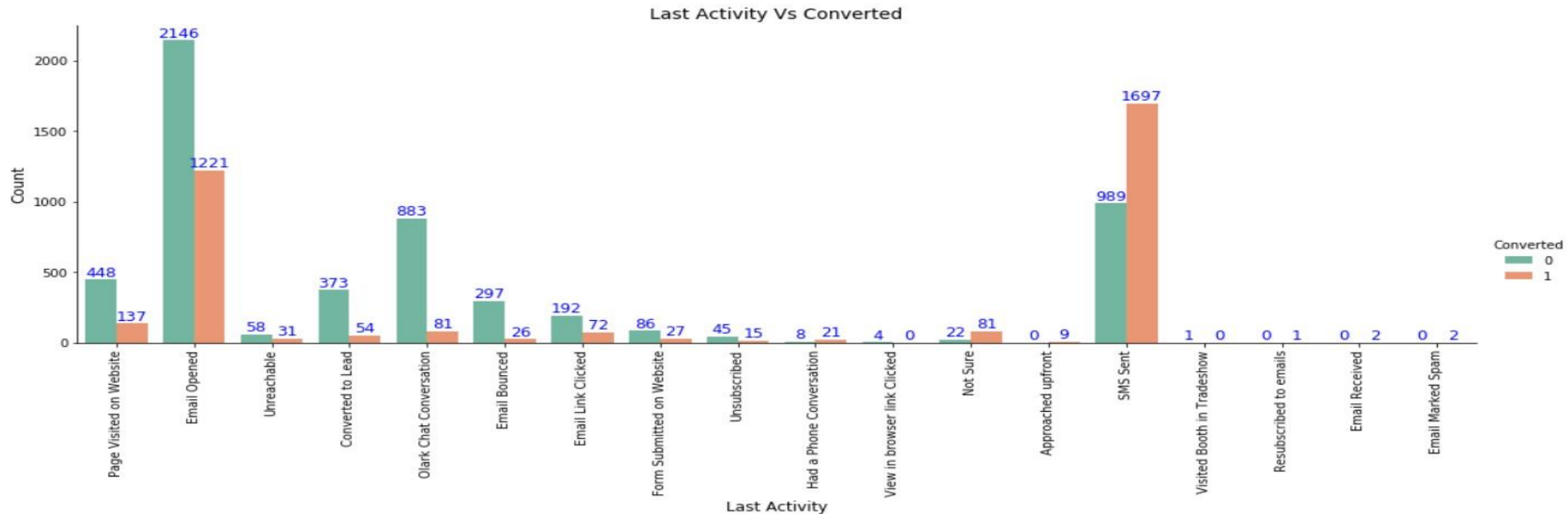
Not much impact on conversion rates through Search, digital advertisements and through recommendations



More conversion happened with people who are unemployed

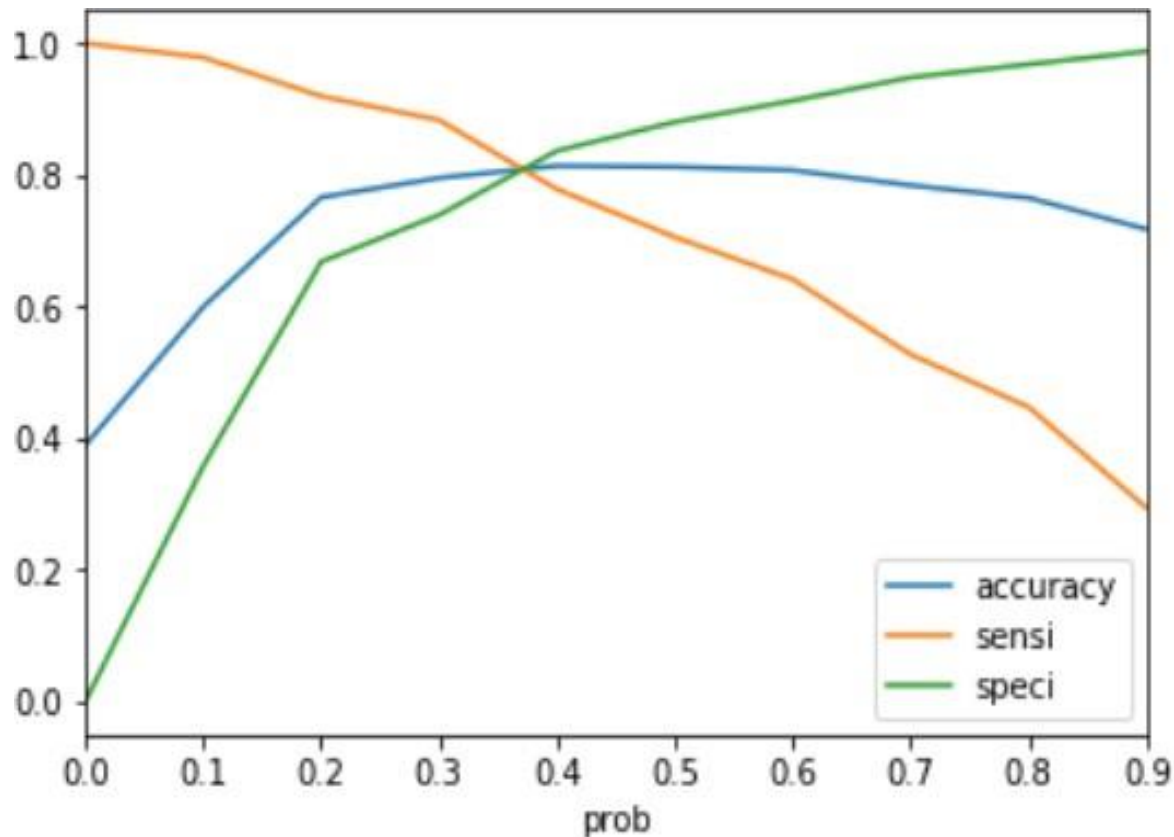


Last Activity value of SMS Sent' had more conversion.



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



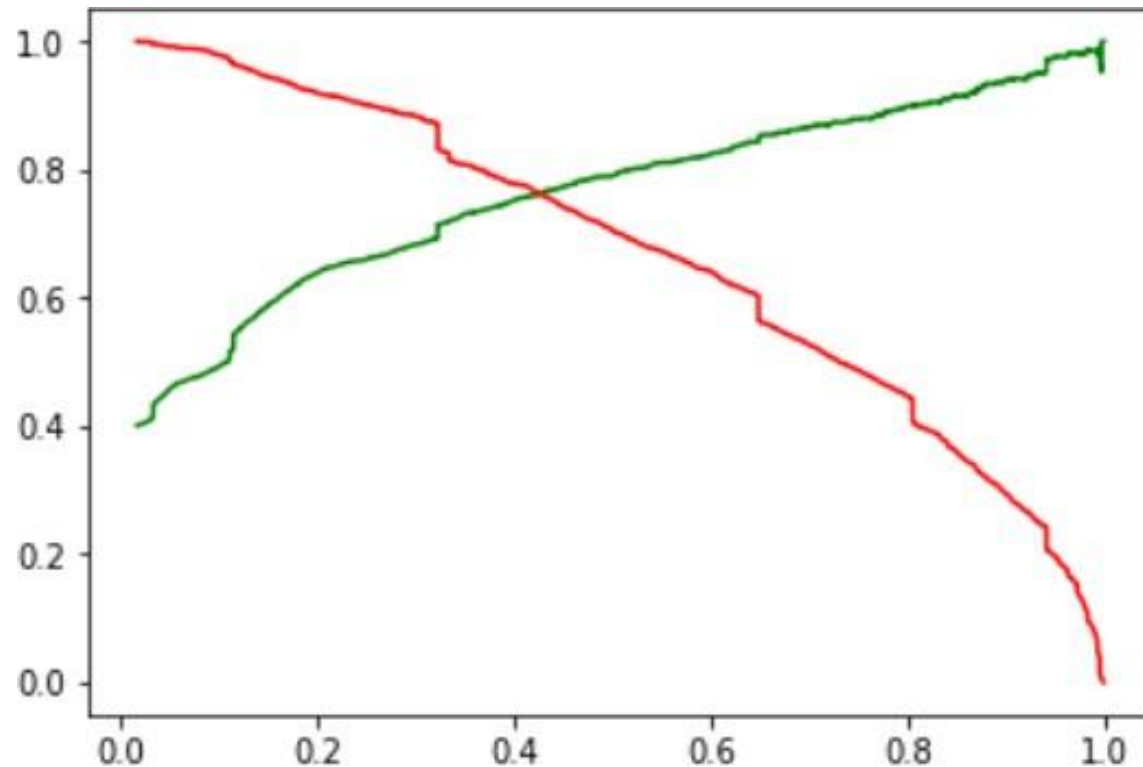
Confusion Matrix

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 %

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix

A 2x2 grid of blue rounded squares. The top-left square contains the number 1394, the top-right contains 300, the bottom-left contains 218, and the bottom-right contains 797. A light gray crosshair is centered in the middle of the four squares.

1394	300
218	797

- Accuracy - 81 %
- Sensitivity - 79 %
- Specificity - 82 %

Conclusion

- While checking the metrics for precision and recall as well as sensitivity-specificity, we also evaluated the best cutoff based on sensitivity and specificity when determining the final forecast.
- The test set's accuracy, sensitivity, and specificity values are around 81%, 79%, and 82%, which are closer to the corresponding figures derived using the training set.
- Additionally, the lead score calculation indicates that the final projected model's conversion rate is about 80% (in the train set) and 79% (in the test set).
- The top 3 factors in the model that influence how many leads convert are :
 - Total time spent on website
 - Lead Add Form from Lead Origin
 - Had a Phone Conversation from Last Notable Activity
- Thus, this strategy appears to be effective overall.