Lead Score Case Study

Submitted by :

1. Muqeet

2.Khushal Kharvante

Lead Score Case Study for X Education

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%.

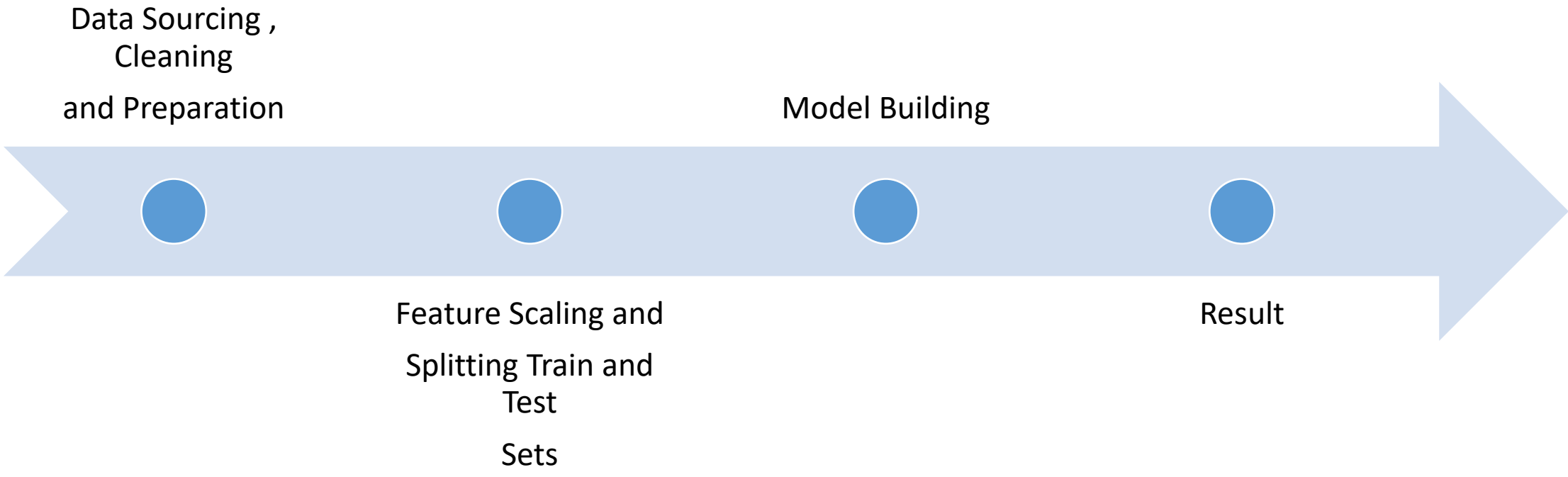
Business Goal:

- 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%.

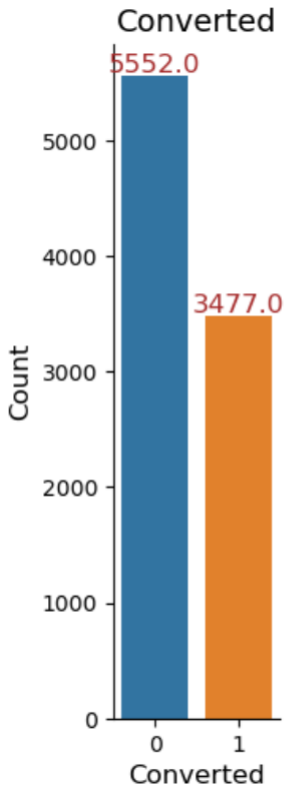
Strategy

1. Source the data for analysis.
2. Clean and preprocess the data.
3. Conduct exploratory data analysis (EDA).
4. Scale the features.
5. Split the data into training and testing datasets.
6. Build a logistic regression model and calculate the lead score.
7. Evaluate the model using metrics such as specificity, sensitivity, precision, and recall.
8. Apply the best-performing model to the test data based on sensitivity and specificity metrics.

Problem Solving Methodology

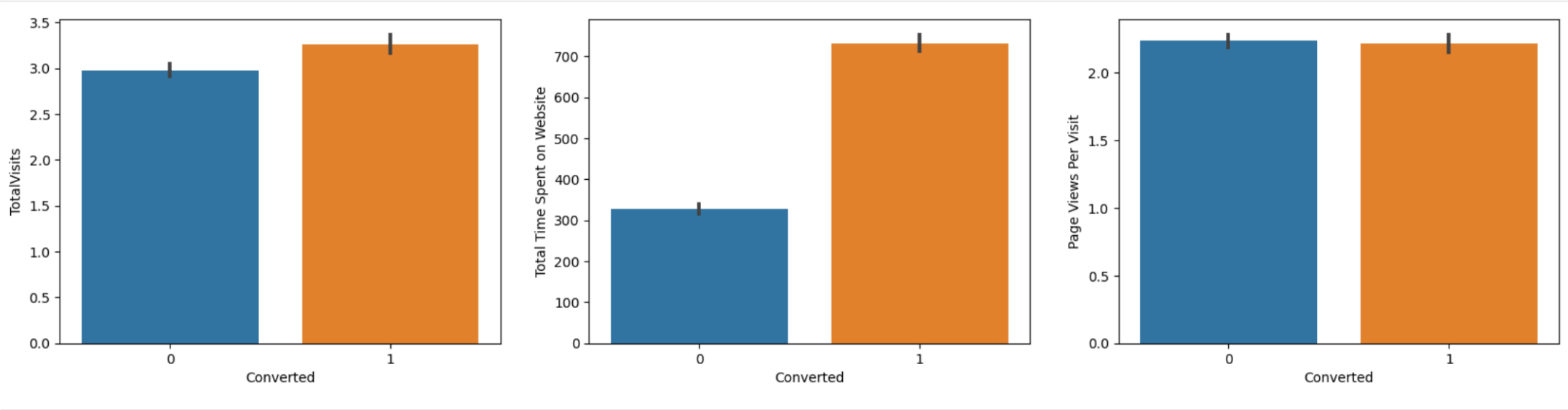


Exploratory Data Analysis

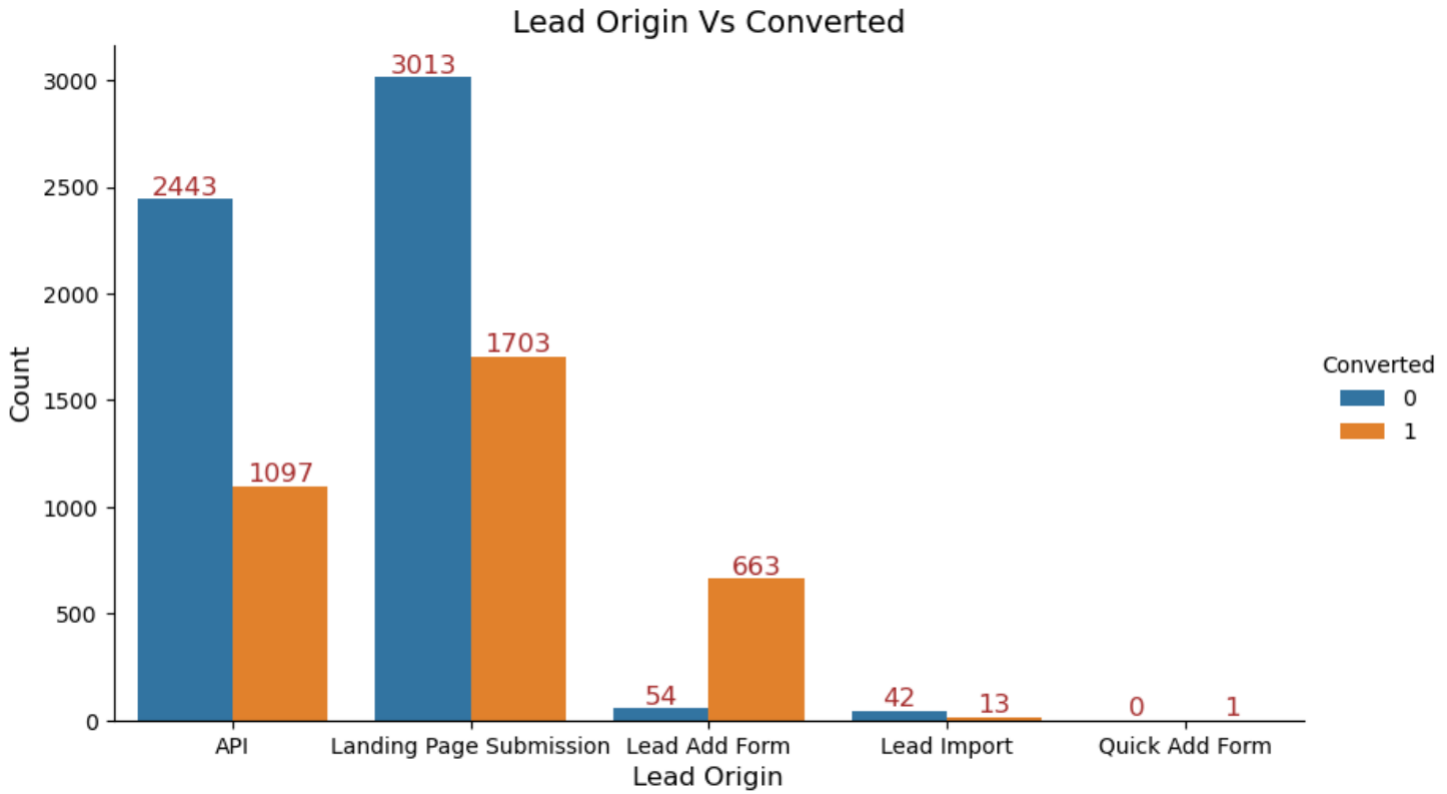


We Have around 39% conversion rate in Total.

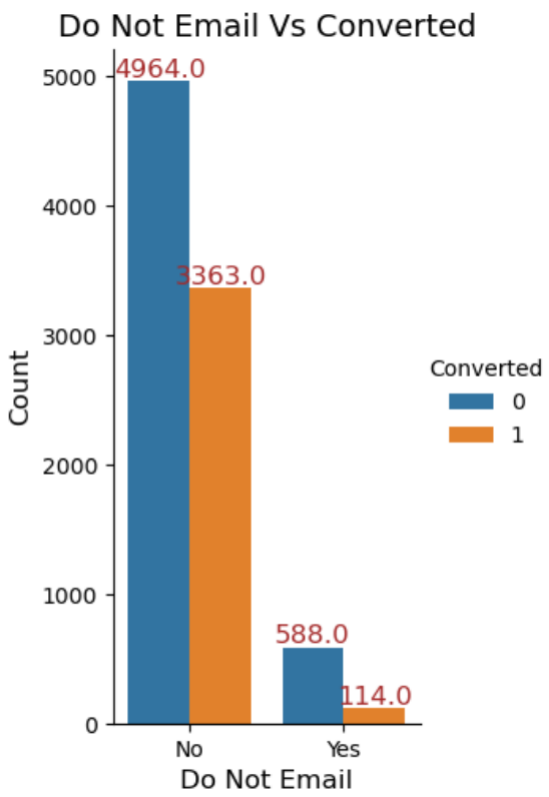
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit.

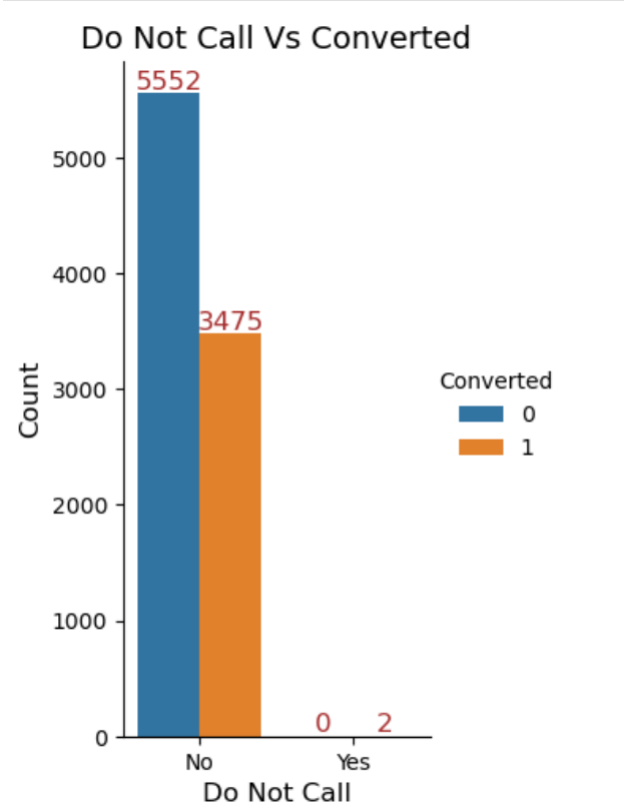


In Lead Origin, maximum conversion happened from Landing Page Submission

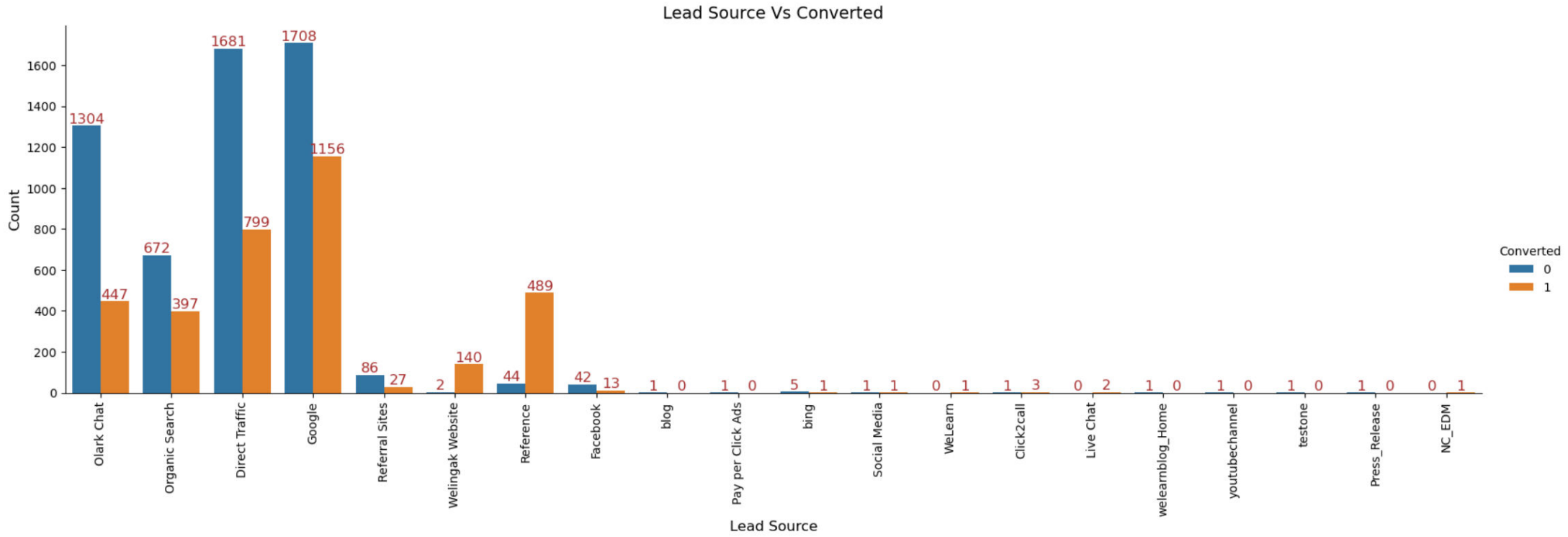


Major conversion has happened from Emails sent and Calls made

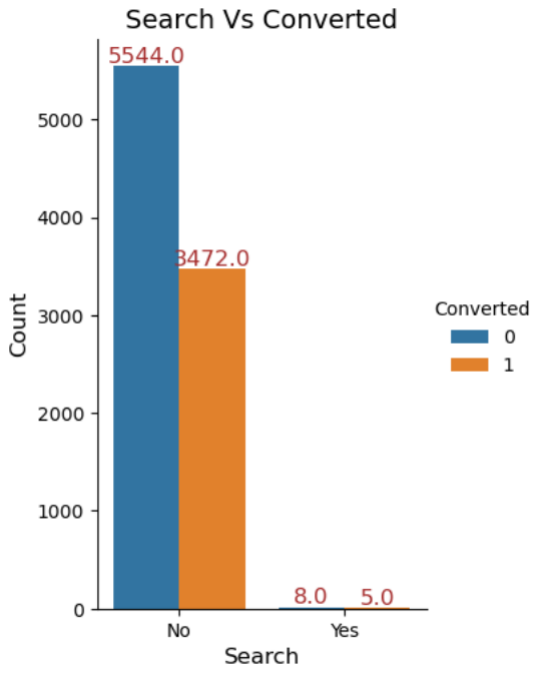


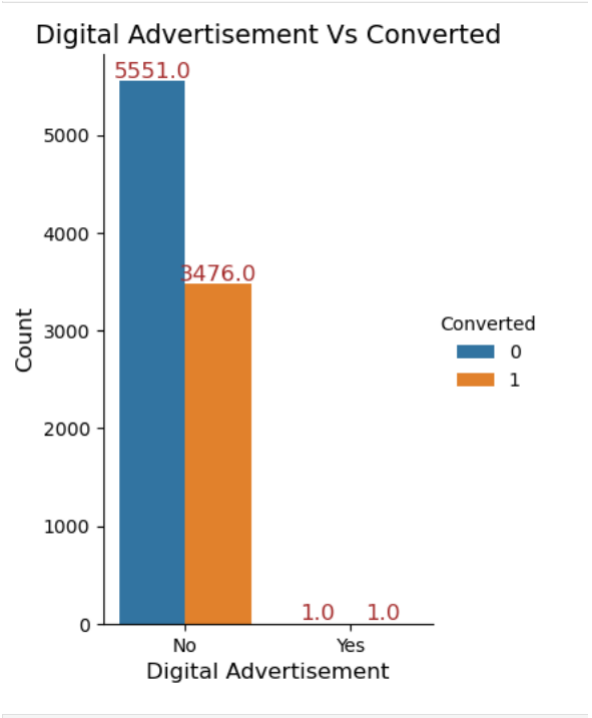


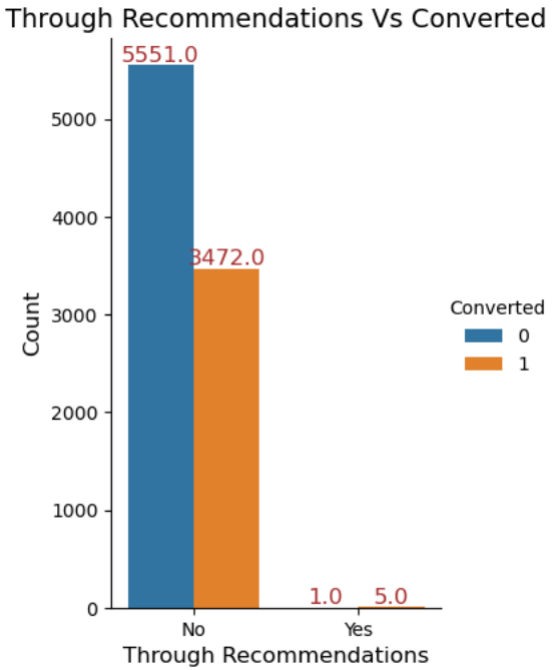
Major conversion in the lead source is from Google.



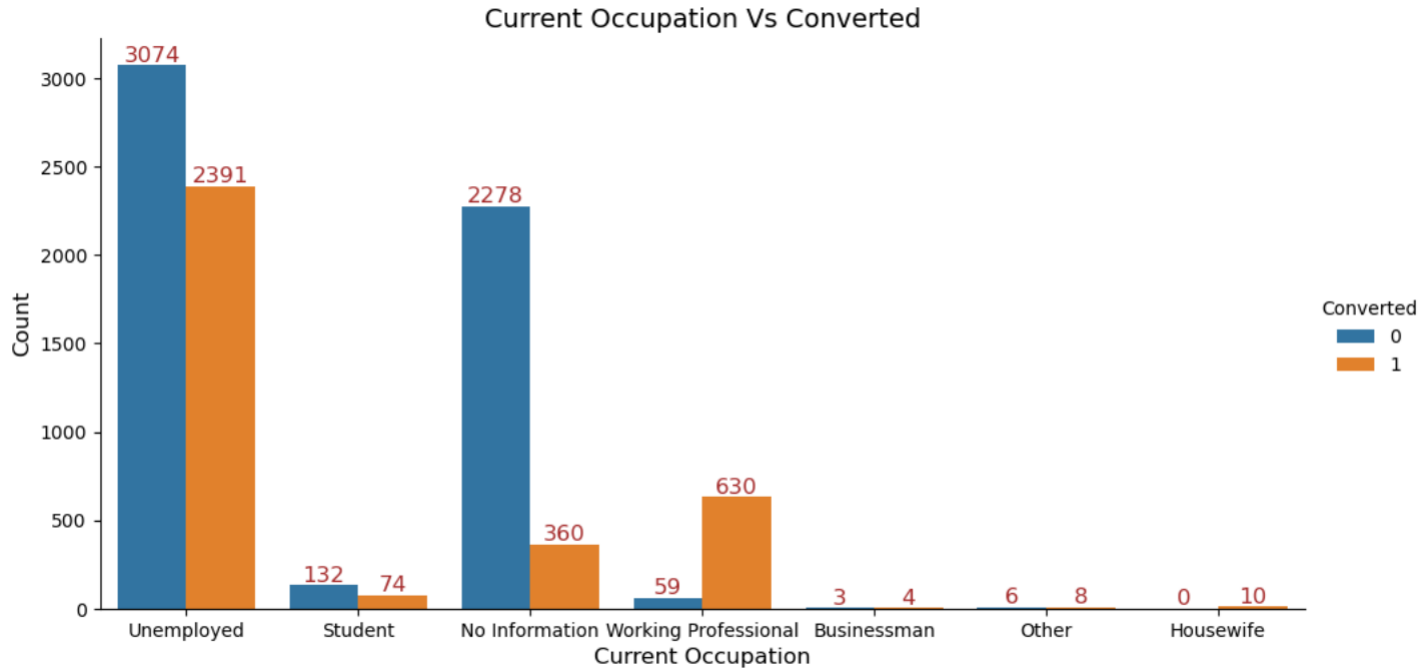
Not much impact on conversion rates through Search, digital advertisements and through recommendations



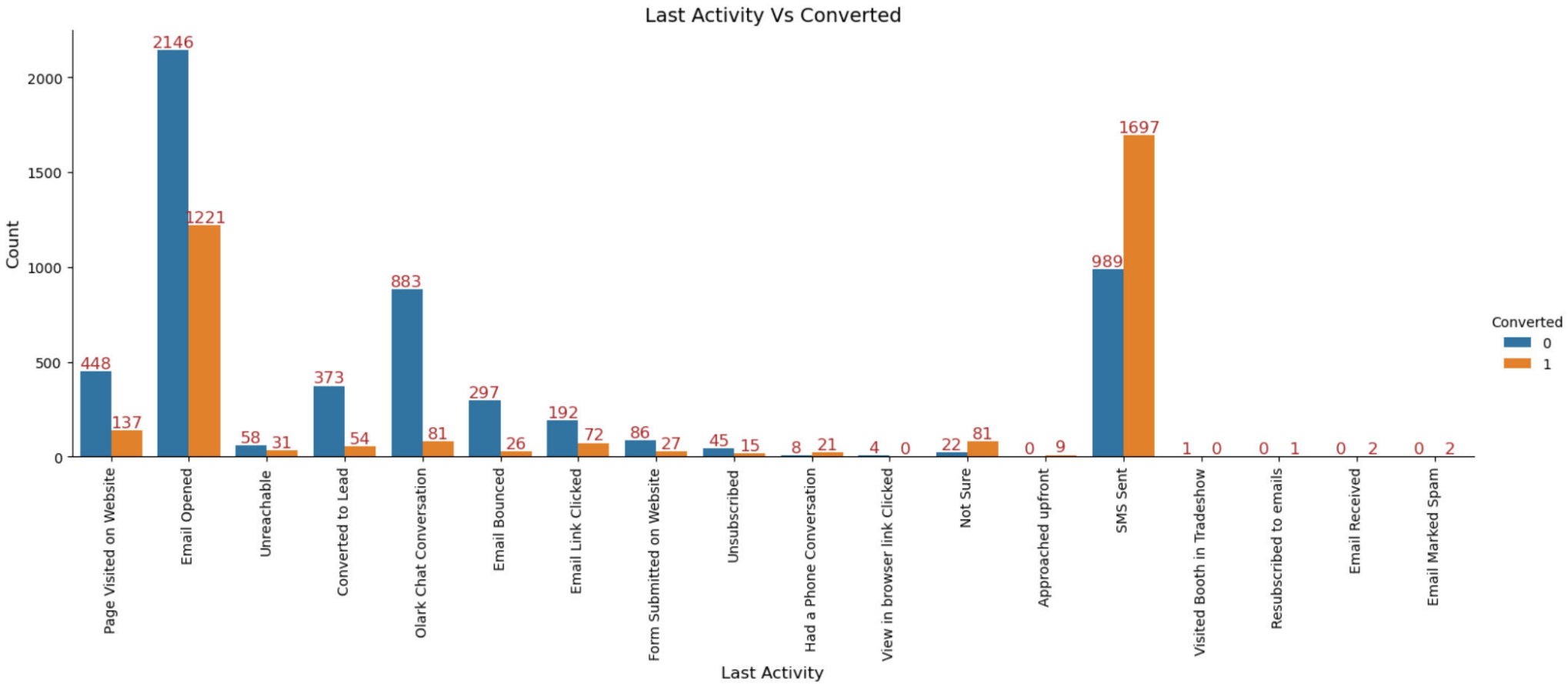




More conversion happened with people who are unemployed



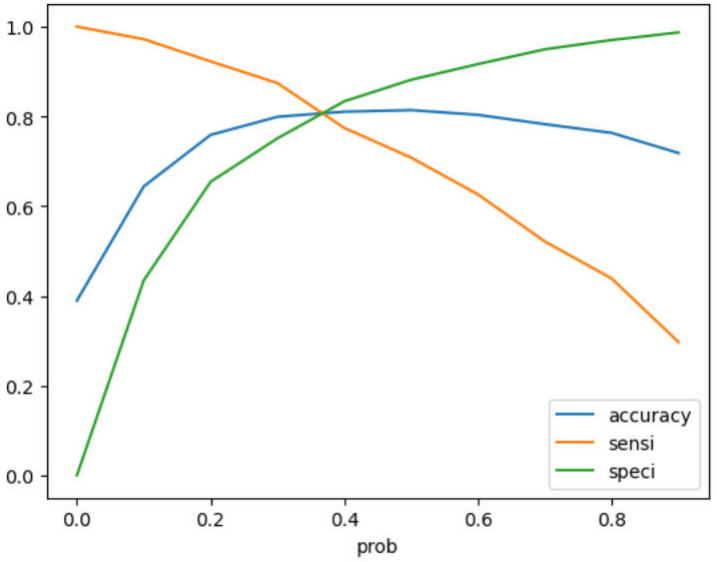
Last Activity value of SMS Sent' had more conversion.



Variables Impacting the Conversion Rate

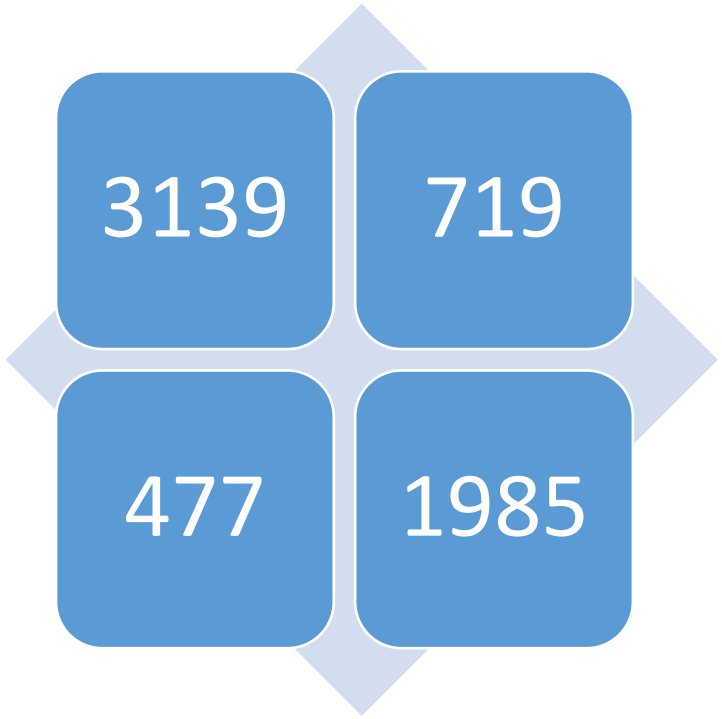
- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – 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

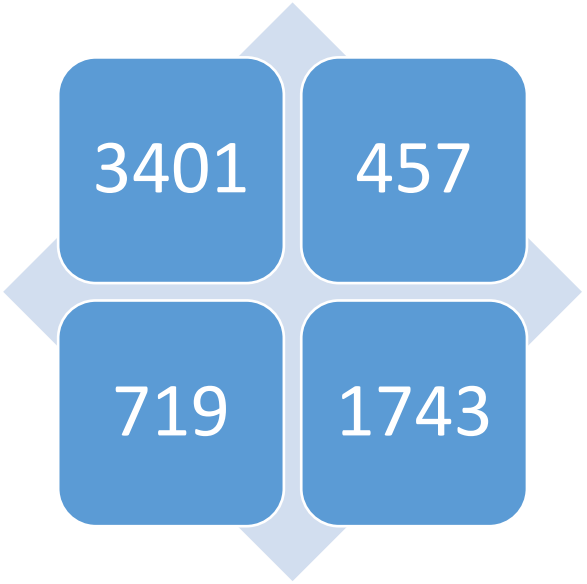
- Confusion Matrix



Accuracy – 81%
Sensitivity – 80%
Specificity – 81%
False Positive Rate – 18%
Positive Predictive Value – 73%
Negative Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset

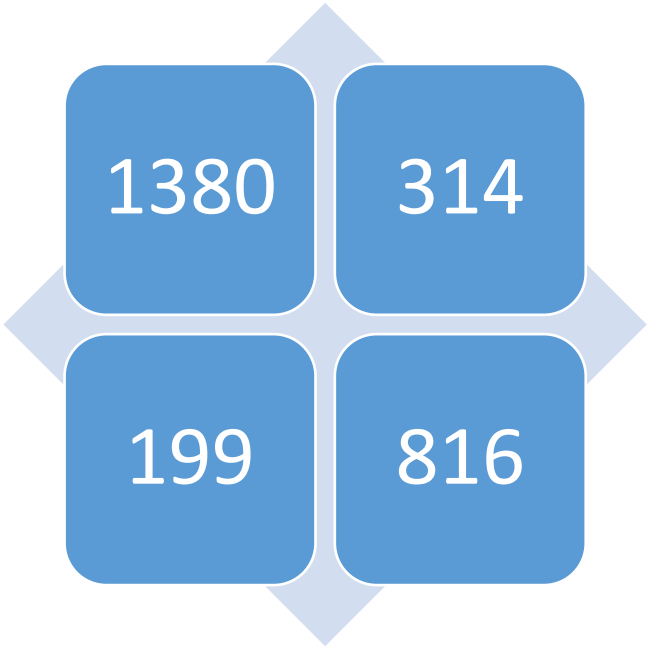
- The graph depicts an optimal cut off of 0.42 based on Precision and Recall
- Confusion Matrix



- Precision – 79%
- Recall – 70%

Model Evaluation – Sensitivity and Specificity on Test Dataset

- Confusion Matrix



- Accuracy – 81%
- Sensitivity – 80%
- Specificity – 81%

Conclusion

After evaluating both Sensitivity-Specificity and Precision-Recall metrics, we determined the optimal cutoff based on Sensitivity and Specificity for the final prediction.

The accuracy, Sensitivity, and Specificity values for the test set are approximately 81%, 80%, and 81%, respectively, which are consistent with the values obtained from the training set.

The lead scoring model indicates a conversion rate of around 80% in the training set and 79% in the test set, reflecting the model's reliability in predicting lead conversions.

The top three variables contributing significantly to lead conversion are:

1. Total Time Spent on Website

2. Lead Add Form from Lead Origin

3. Had a Phone Conversation from Last Notable Activity

Overall, the model performs well, effectively predicting lead conversions with high accuracy and consistency across both training and test sets.