LEAD SCORE CASE STUDY

Submitted By:
Kiran Kumawat
Sandeep Chaudhary
Batch: DS C43

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

- to identify the most potential leads ie hot leads
- to build a logistic regression model to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and vice versa
- Target lead conversion rate to be around 80 %

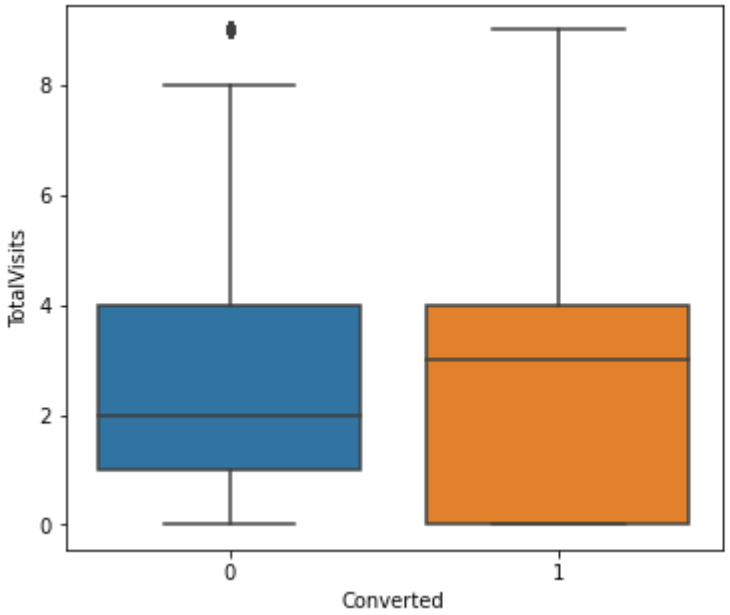
ASSUMPTIONS

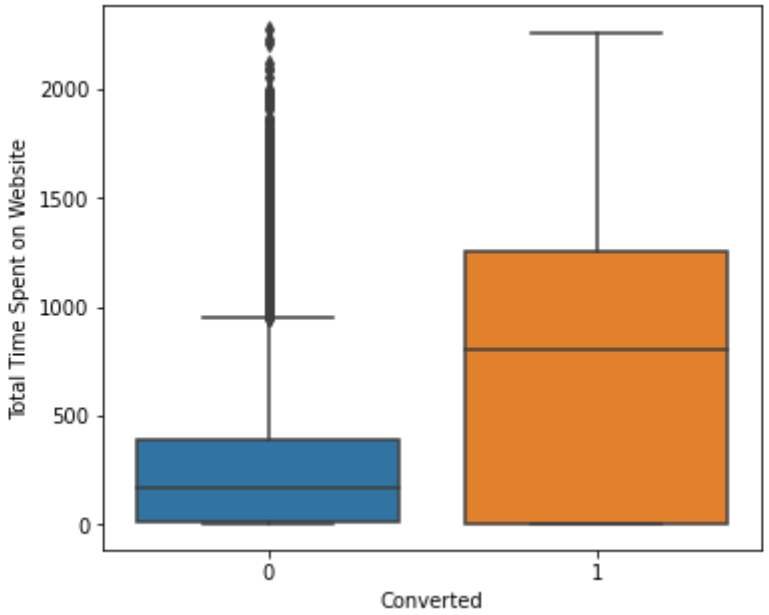
- We have assumed that variables having values as ‘Select’ are equivalent to null values
- Removing columns with more than 45 % of missing values as most of the data is missing and if we impute it with mean/ median/ mode then, it may lead to bias
- Missing values can be imputed with mode (for categorical variables) and median (for continuous variables)

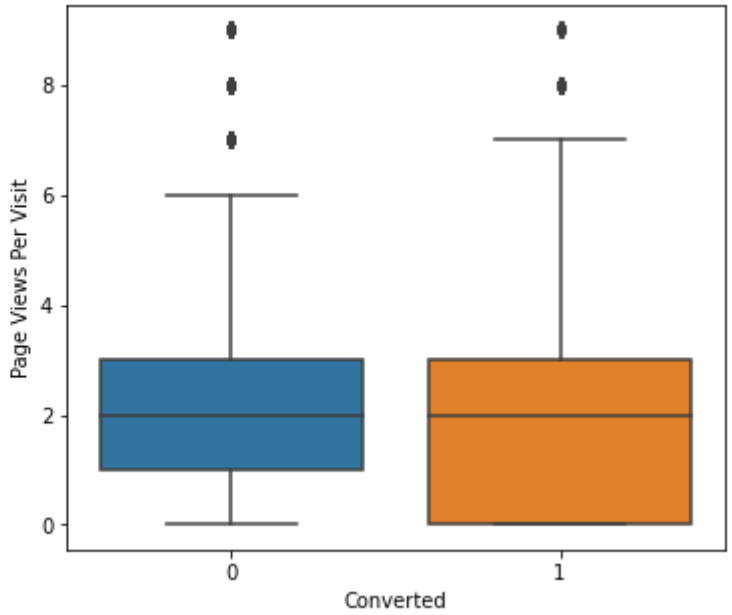
OVERALL APPROACH

- Data Cleaning and EDA
- Data Preparation
- Test – Train Split
- Feature Scaling using MinMax Scaler
- Model Building using RFE
- Finding Optimal Cutoff Point – accuracy, sensitivity and specificity curve and using precision – recall trade off
- Making predictions on the test set
- Calculating the lead Score
- Making conclusions

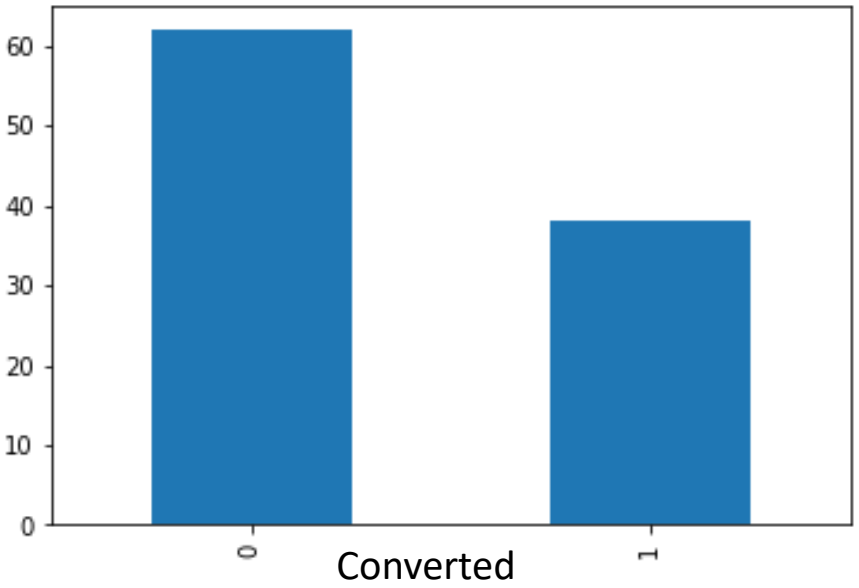
INSIGHTS FROM THE DATA



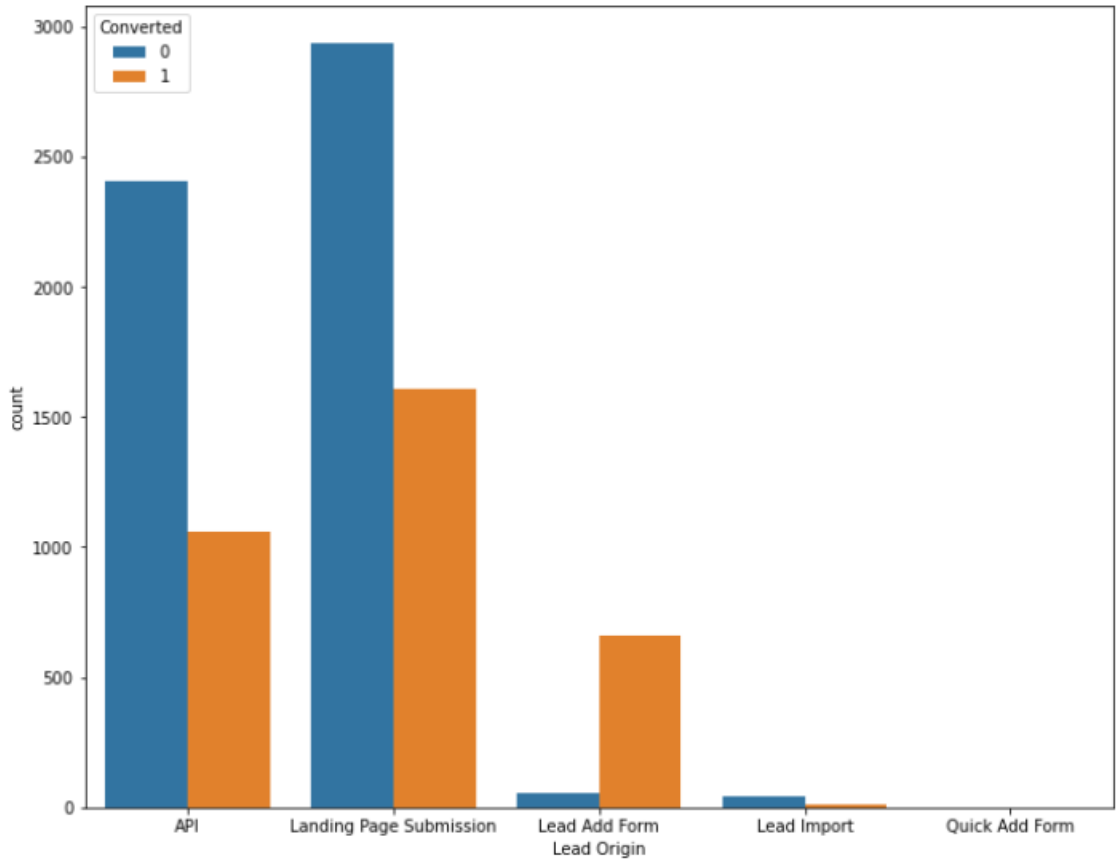




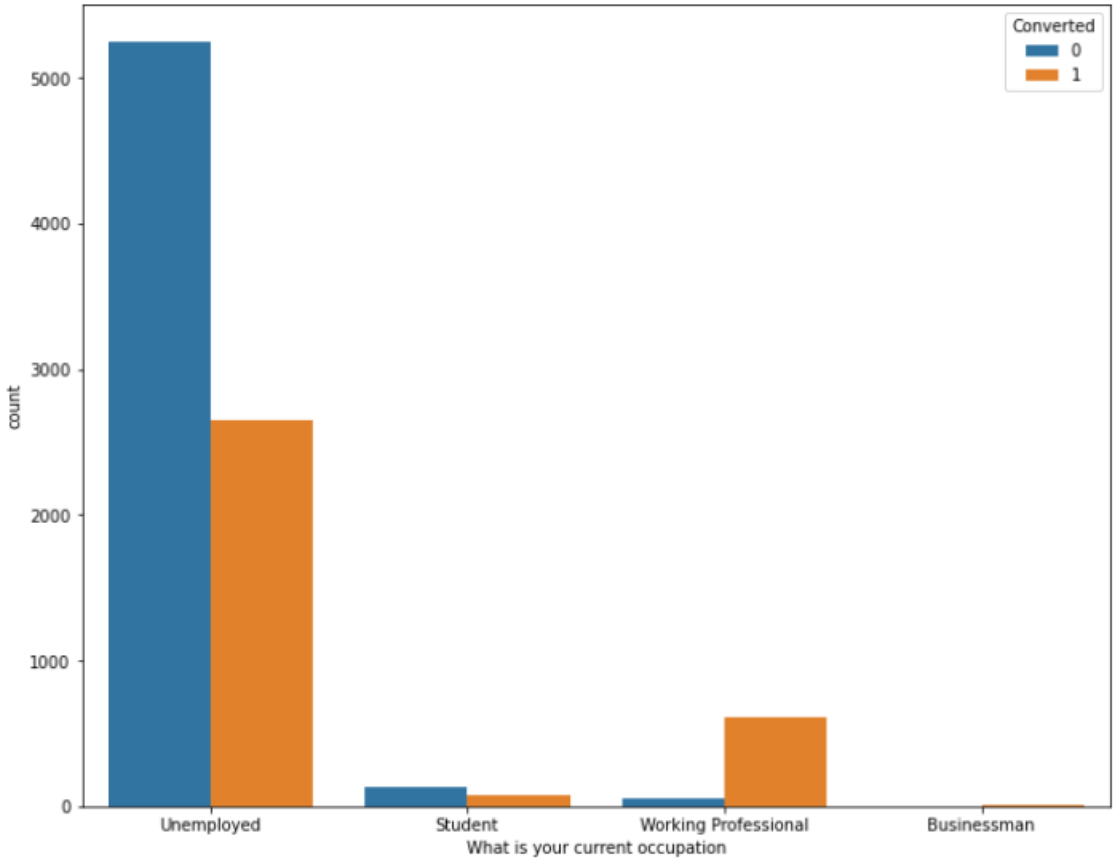
- Median of Total Visits is more for converted ones. People visiting the site more have higher chances of conversion.
- People who have converted have spent more time on the website and the median time spent on website is more in case of converted as compared to non-converted.
- Median for no. of pages viewed per visit is same for both converted and non-converted



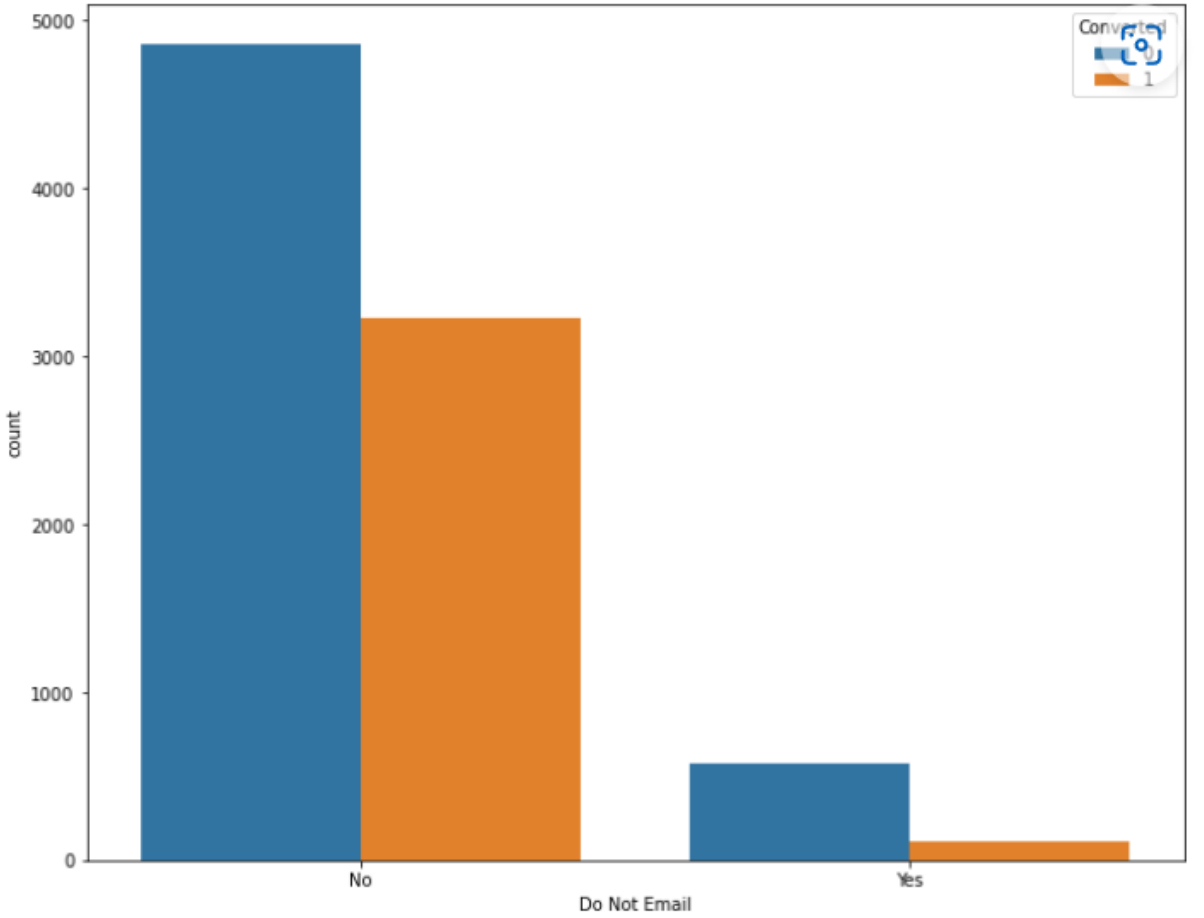
Originally, lead conversion rate is approximately 38 %



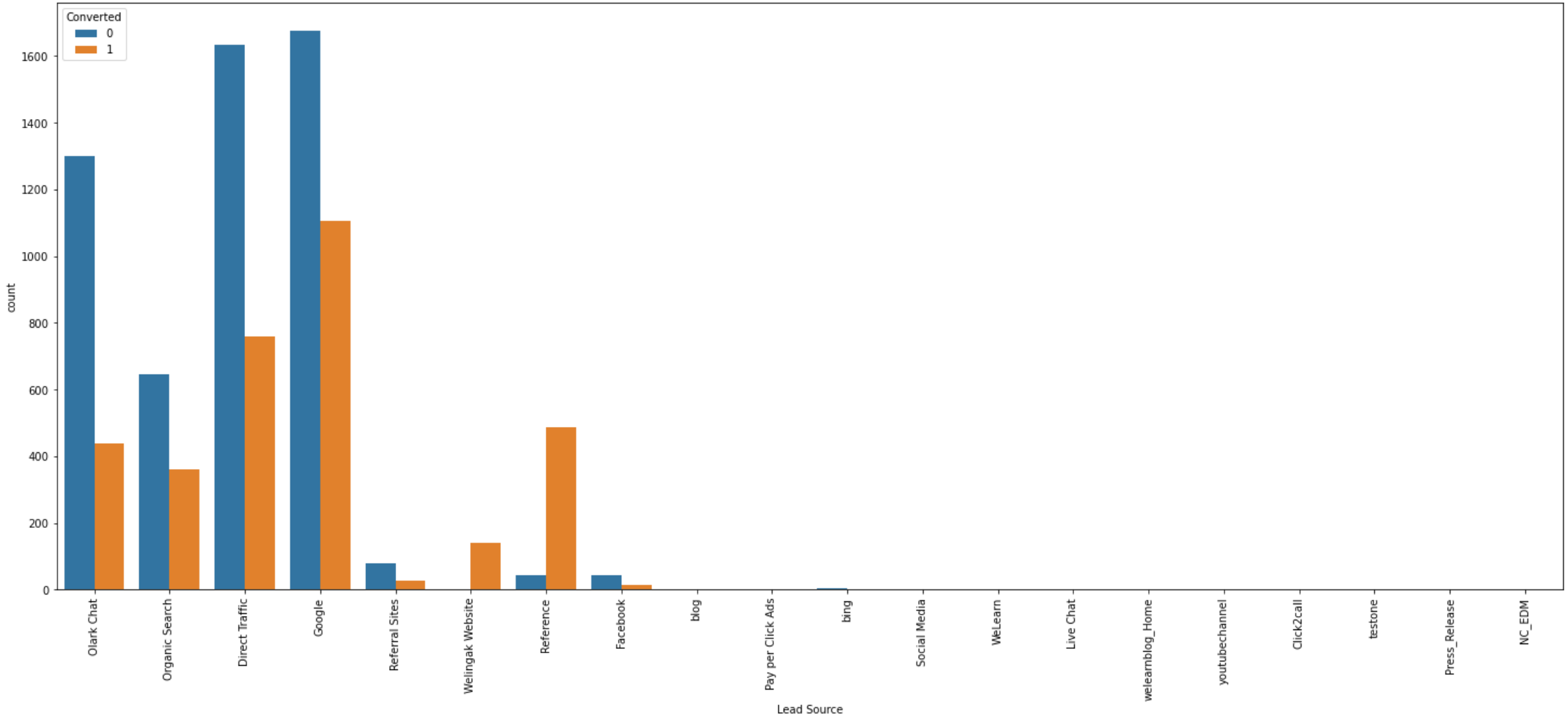
Lead conversion is highest for Lead Add Form



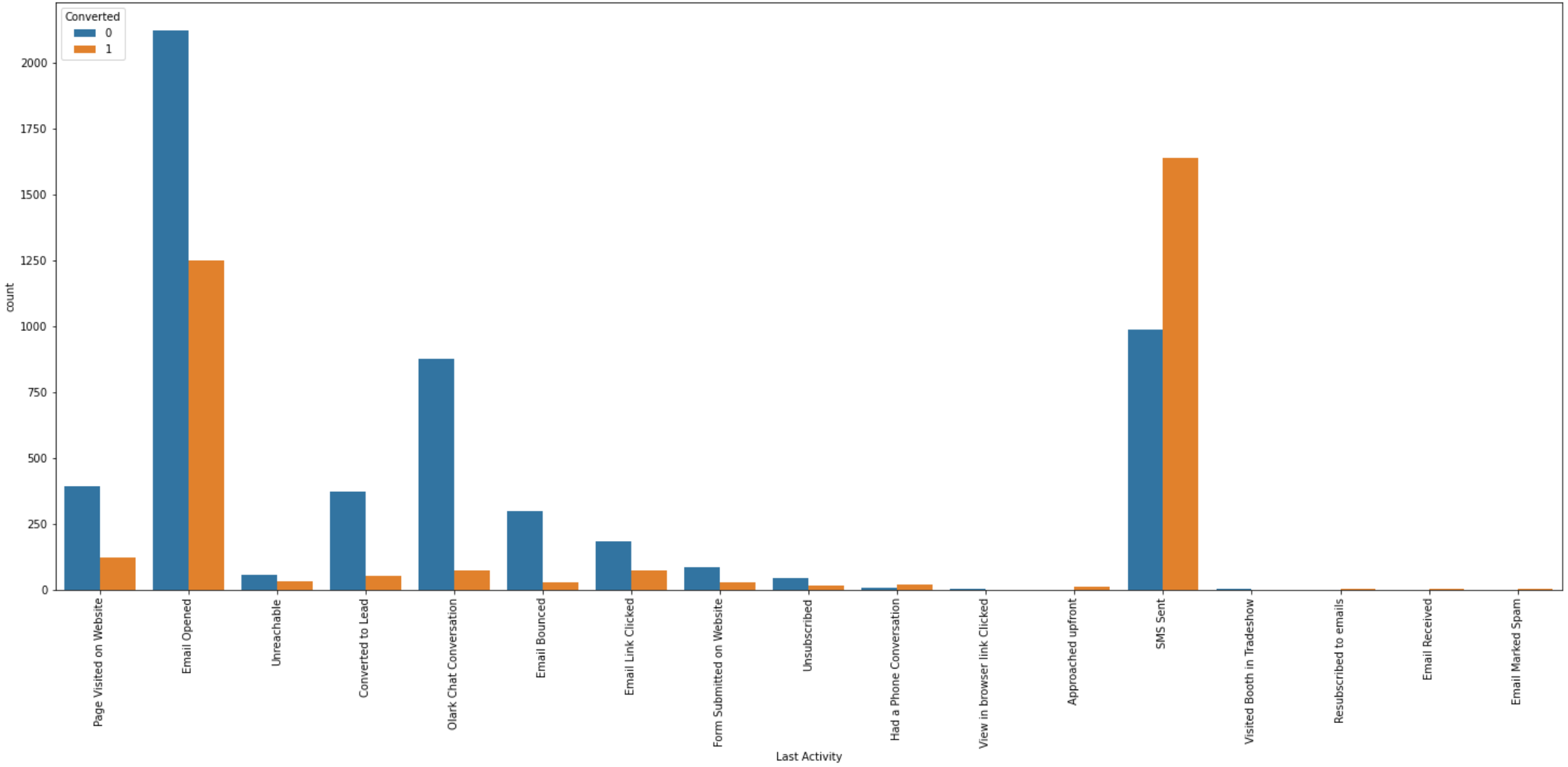
Working Professionals have higher conversion rate



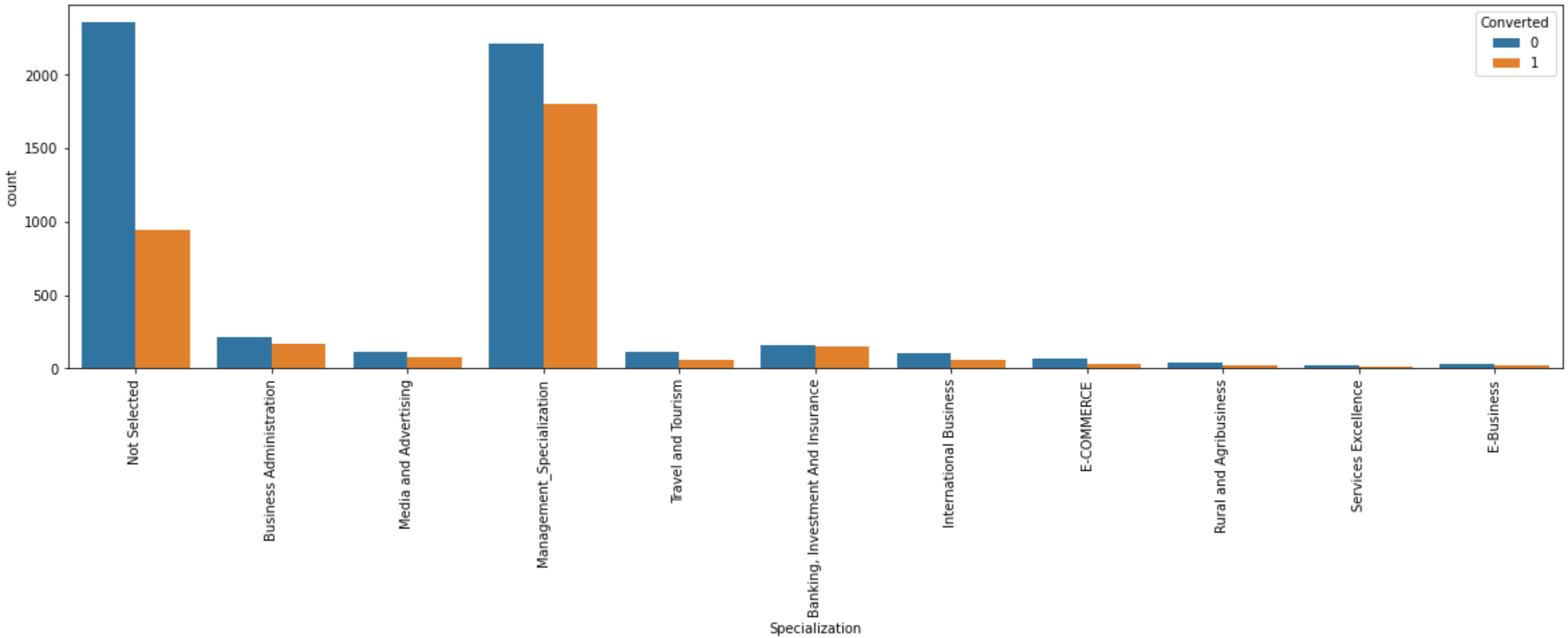
Higher conversion is for people responding as 'No' for 'Do Not Email'



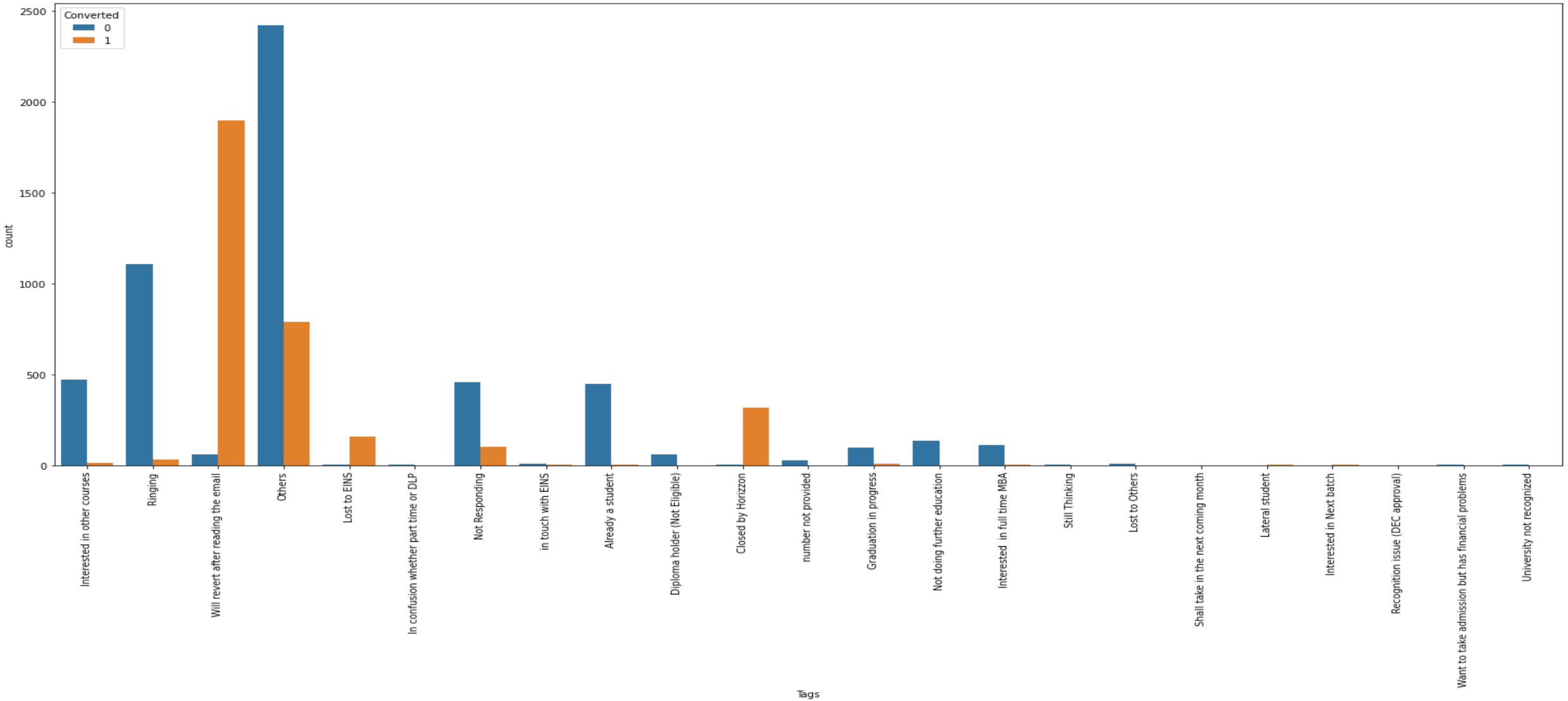
Lead conversion is higher through Reference, Welingak Website, Google



For Last Activity of the customer, conversion rate is higher for those whom SMS has been sent followed by Email Opened.

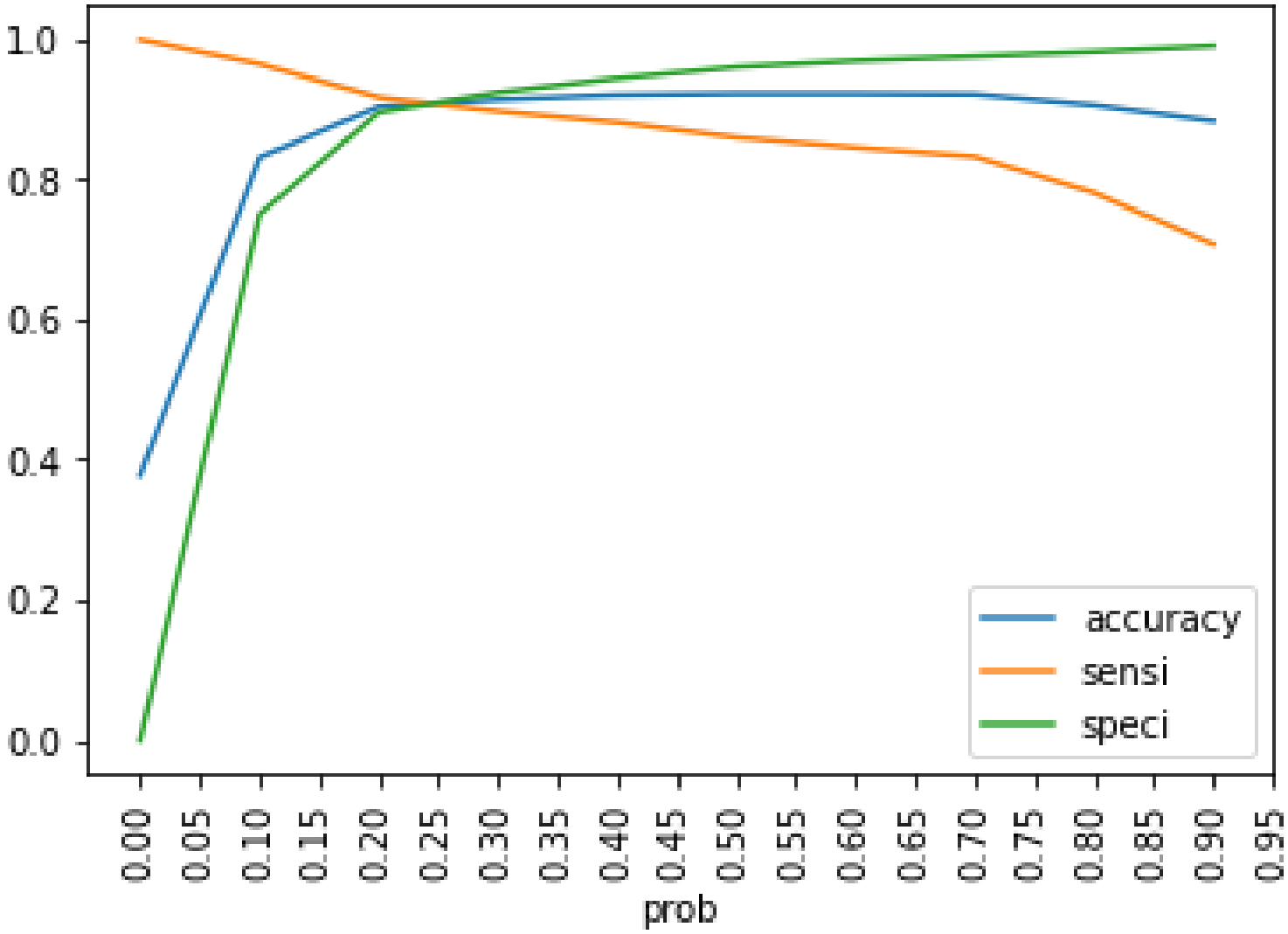


Customers working in the field of Banking, Investment & Insurance, Business Administration, Media & Advertising, Management_Specialization seems to have good conversion rate.



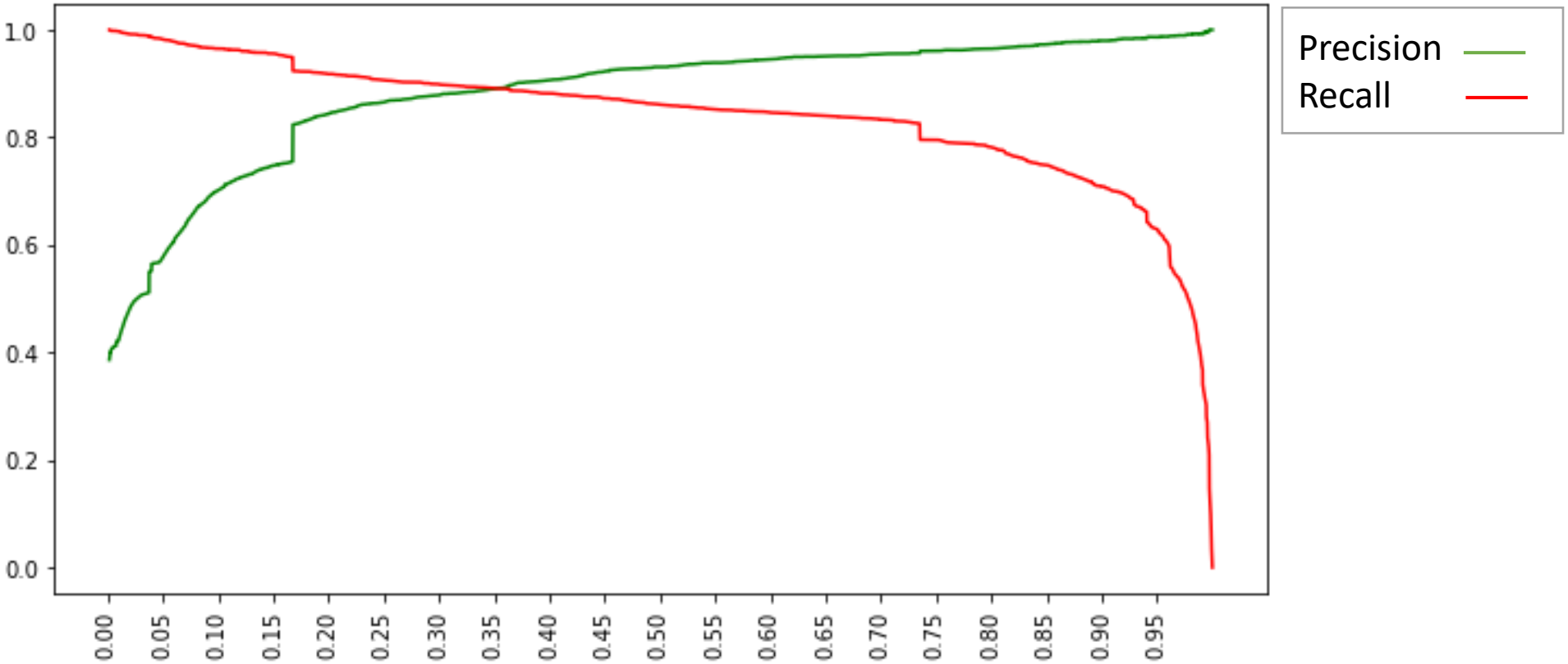
Tags assigned as Will revert after reading the email, Lost to EINS, Closed by Horizon seems to have higher conversion rates.

OPTIMAL CUT OFF POINT



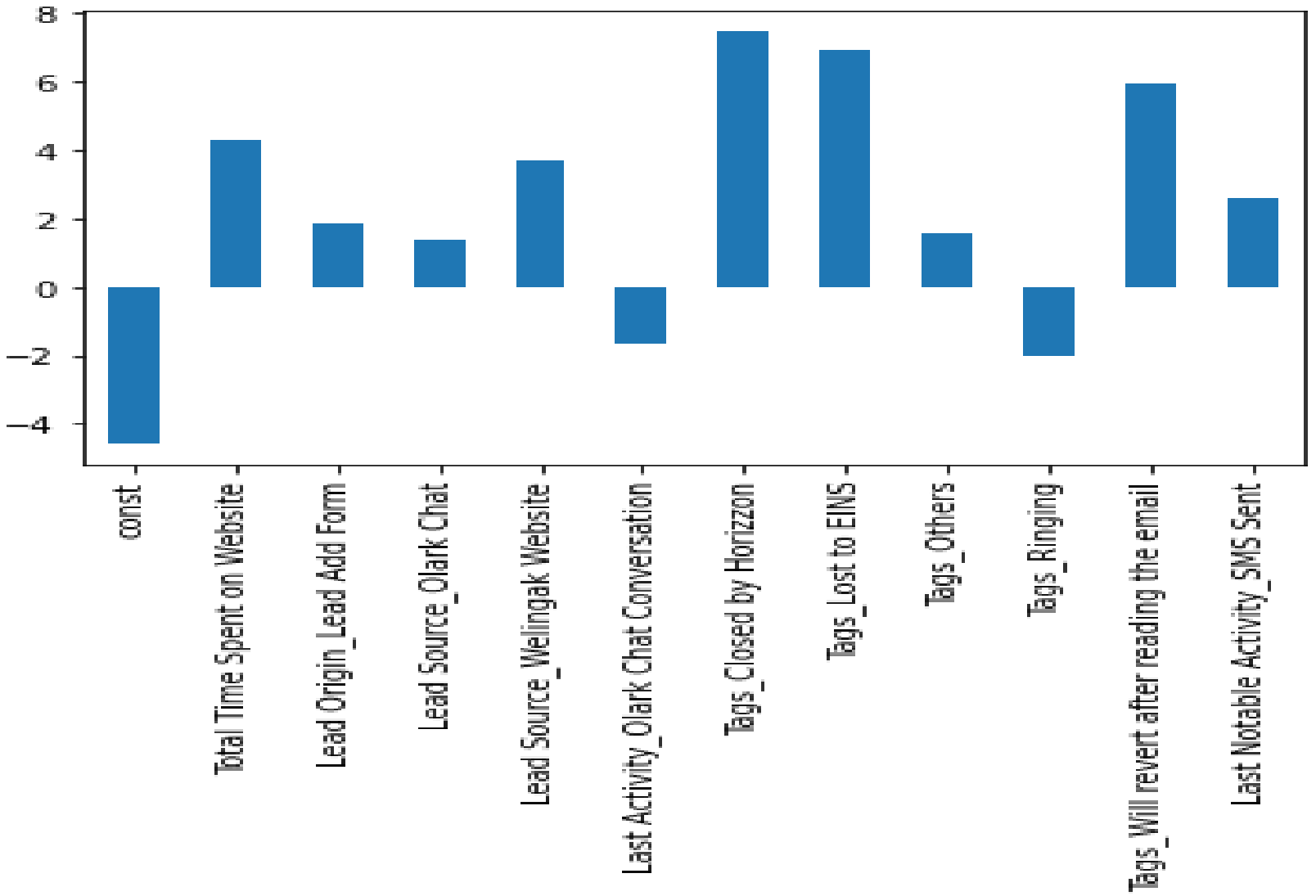
The graph depicts an optimal cut off of 0.25 based on Accuracy, Sensitivity and Specificity.

PRECISION – RECALL TRADE OFF



From the curve above, 0.35 is the optimum point based on recall-precision tradeoff but at 0.25 we are getting approximately similar results in terms of conversion probability so we will take 0.25 as the cut off as we want the model to have higher sensitivity.

CONCLUSION



Top 3 predictors for higher conversions are:

- Tags_Closed by horizzon
- Tags_Lost to EINS
- Tags_Will revert after reading the email

Confusion Metrics Results for Train and Test data

For Train Data:

Accuracy = 91.10 %

Sensitivity or Recall = 90.61 %

Specificity = 91.40 %

Precision = 86.52 %

F1 Score = 88.52 %

For Test Data:

Accuracy = 91.29 %

Sensitivity or Recall = 91.52 %

Specificity = 91.16 %

Precision = 86.66 %

F1 Score = 89.02 %

The model is performing pretty well on both test and train data. It is having good & comparable accuracy, sensitivity, specificity, precision and F1 scores.

RECOMMENDATIONS

- Company should focus on the leads having high Lead Score i.e lead score > 80 as these leads have high probability of conversion
- Company should focus on the leads having Tags as Closed by horizons, Lost to EINS, will revert after reading email as they are the top 3 predictors for target variable

THANK YOU