LEAD SCORE CASE STUDY

ARYA NANDAKUMAR
LAVISH M SAHU

Problem statement

X Education, an organization that offers online courses for industry professionals, promotes its courses on various well-known platforms, including Google.

X Education wants to select most promising leads that can be converted to paying customers.

While the company successfully generates numerous leads, only a small percentage transitions into paying customers. The company aims for a higher conversion rate and acquires leads through various channels such as email, website advertisements, and Google searches.

The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

Case Study Goal

The company requires a model to be built for selecting most promising leads.

Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score more promising the lead to get converted.

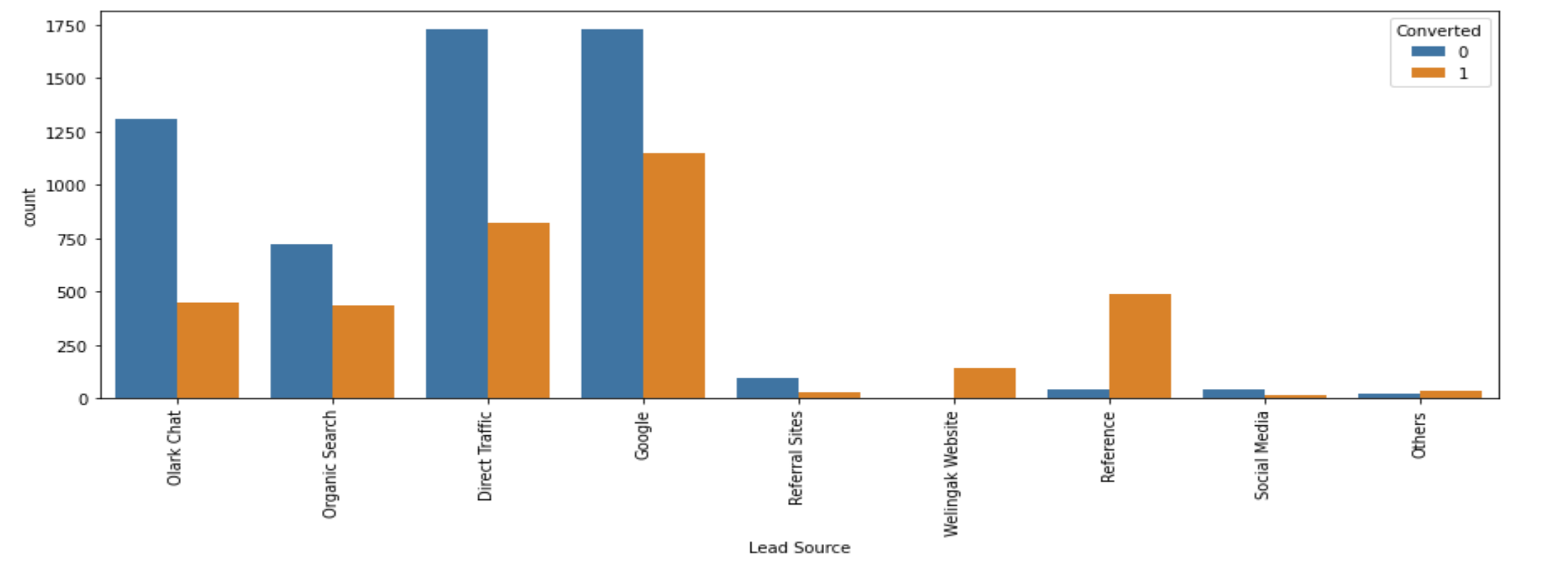
The model to be built in lead conversion rate around 80% or more.

Solution Methodology

- Reading And Understanding The Data
- Checking Data Quality
- Exploratory Data Analysis
  - Data Cleaning and Treatment
    - 1.Check and handle NA values and missing values.
    - 2.Drop columns, if it contains large number of missing values and not useful for the analysis.
    - 3. Imputation of the values, if necessary.
    - 4.Check and handle outliers in data.
- Categorical Variables Analysis
- Identifying Categorical Variables Columns and Creating Dummy Variables
- Logistic Regression Model Building
- Making Predictions On Test Dataset With Model Built From Training Dataset
- Documenting Final Observations including measuring and comparing Accuracy, Sensitivity and Specificity between Train and Test Data sets.

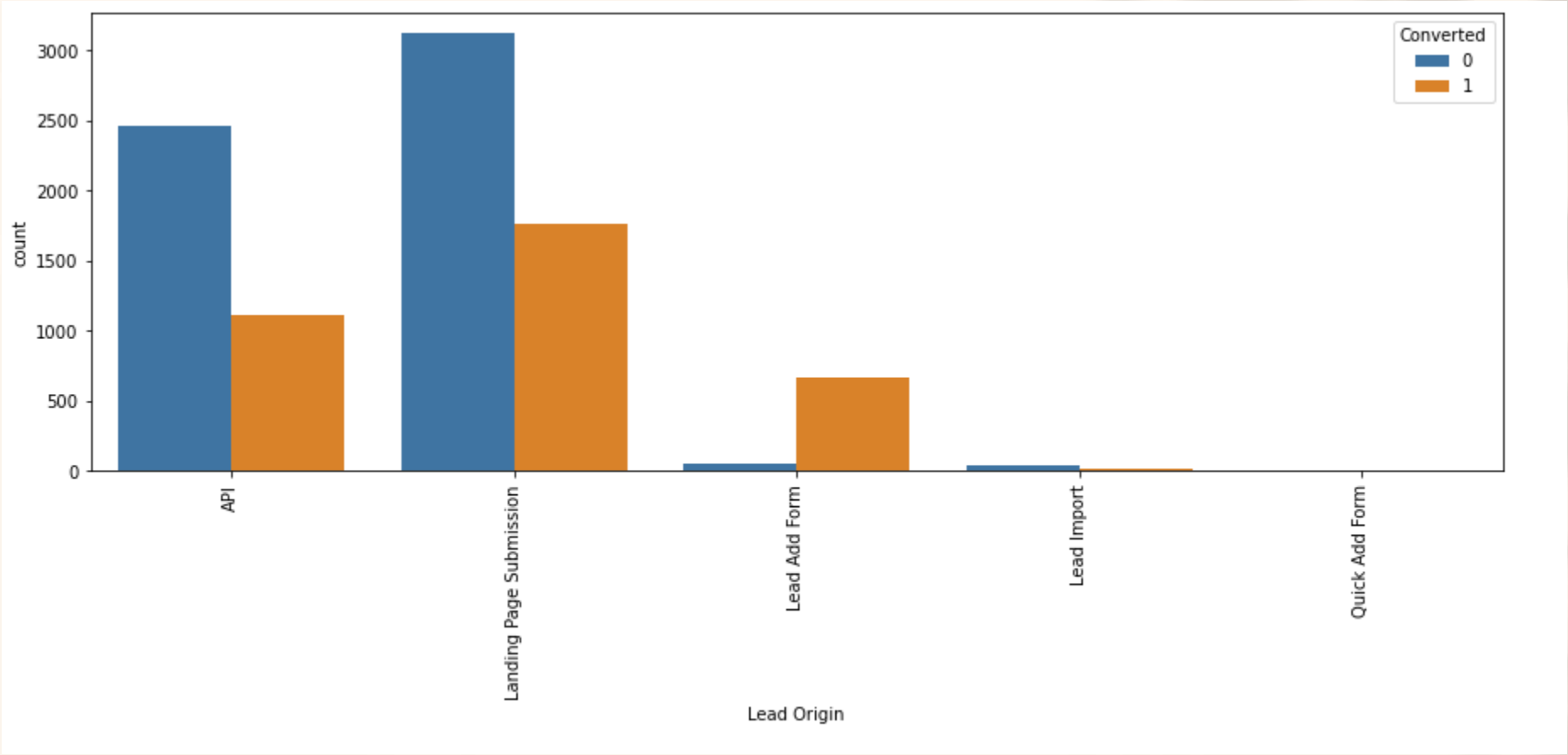
Lead Source Analysis

- The highest number of leads and lead conversions are generated by 'Google' and 'Direct Traffic,' whereas sources such as 'Welingak Website,' 'Reference,' and 'Others' exhibit the highest conversion ratio for leads.
- To enhance the overall lead conversion rate, prioritize the improvement of lead conversions from 'direct traffic' and 'Google leads.' Additionally, allocate efforts towards generating more leads from 'reference' and the 'Welingak website' due to their strong conversion rates.



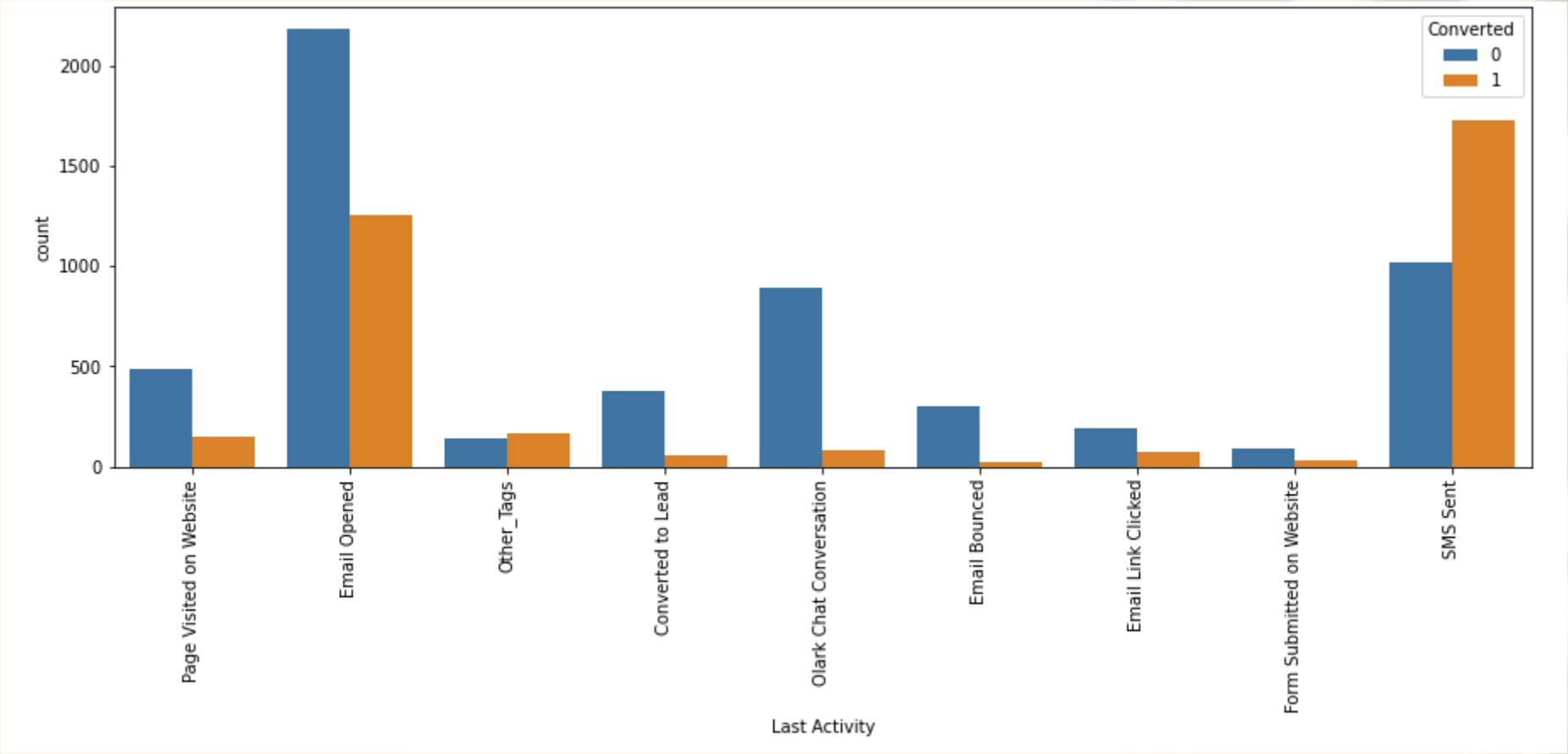
Lead Origin Analysis

- 'Landing Page Submissions' and 'API' bring a higher number of leads and see more lead conversion as well.
- Overall, “Lead Add Form” has high lead conversion ratio, but low volume and Landing page submission has high lead conversion ratio with enough volume to increase number of leads converted. Lead Import and Quick Add Form get very few leads.
- To improve overall lead conversion rate, we must improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



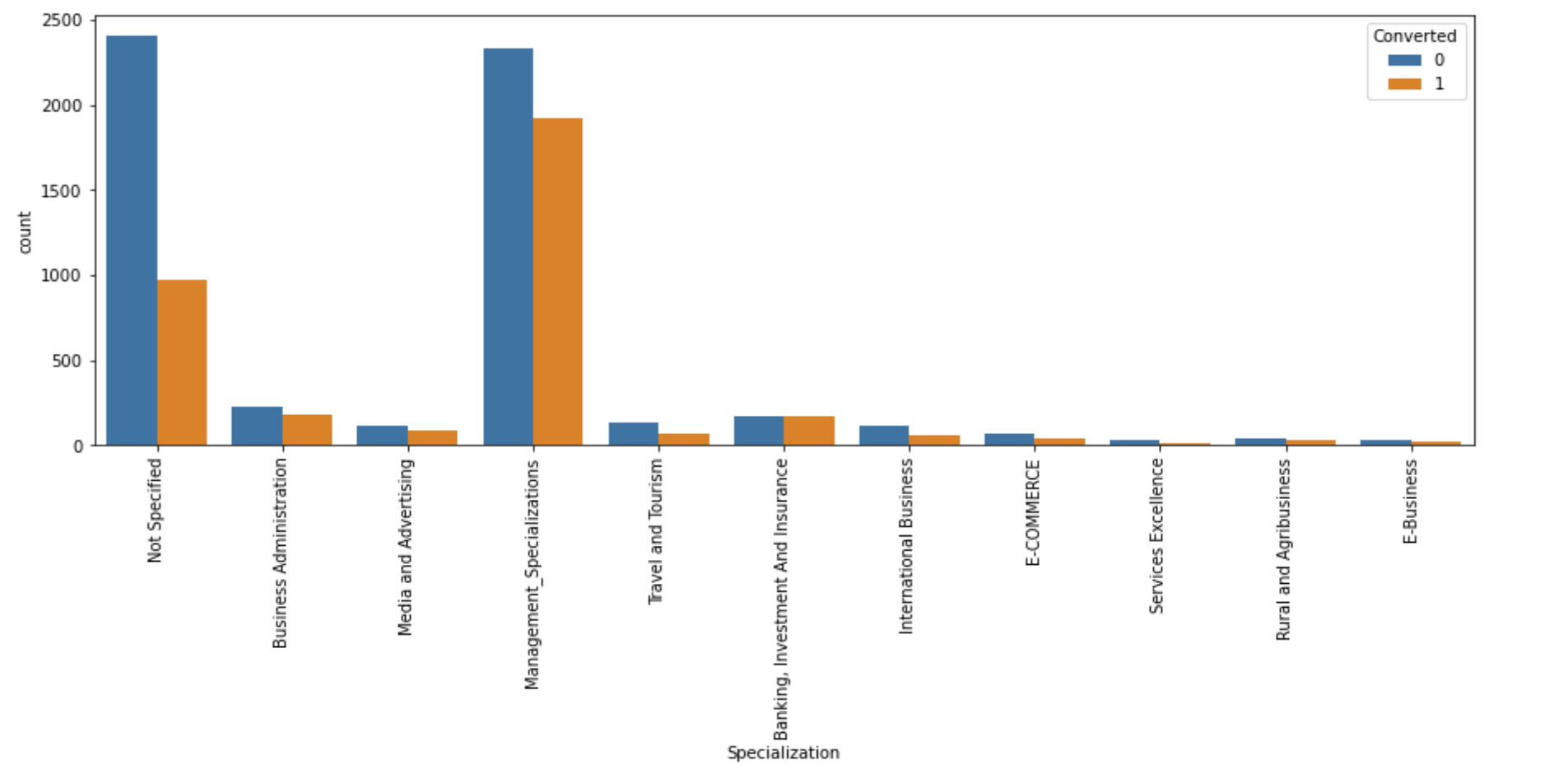
Last activity conversion Analysis

- Leads having last activity as 'SMS Sent' have the most conversion rate.
- 'Email Opened' brings maximum number of leads and has second most conversion as well.
- Focus should be more on leads where last activity will be SMS sent as they see highest conversion ratios.



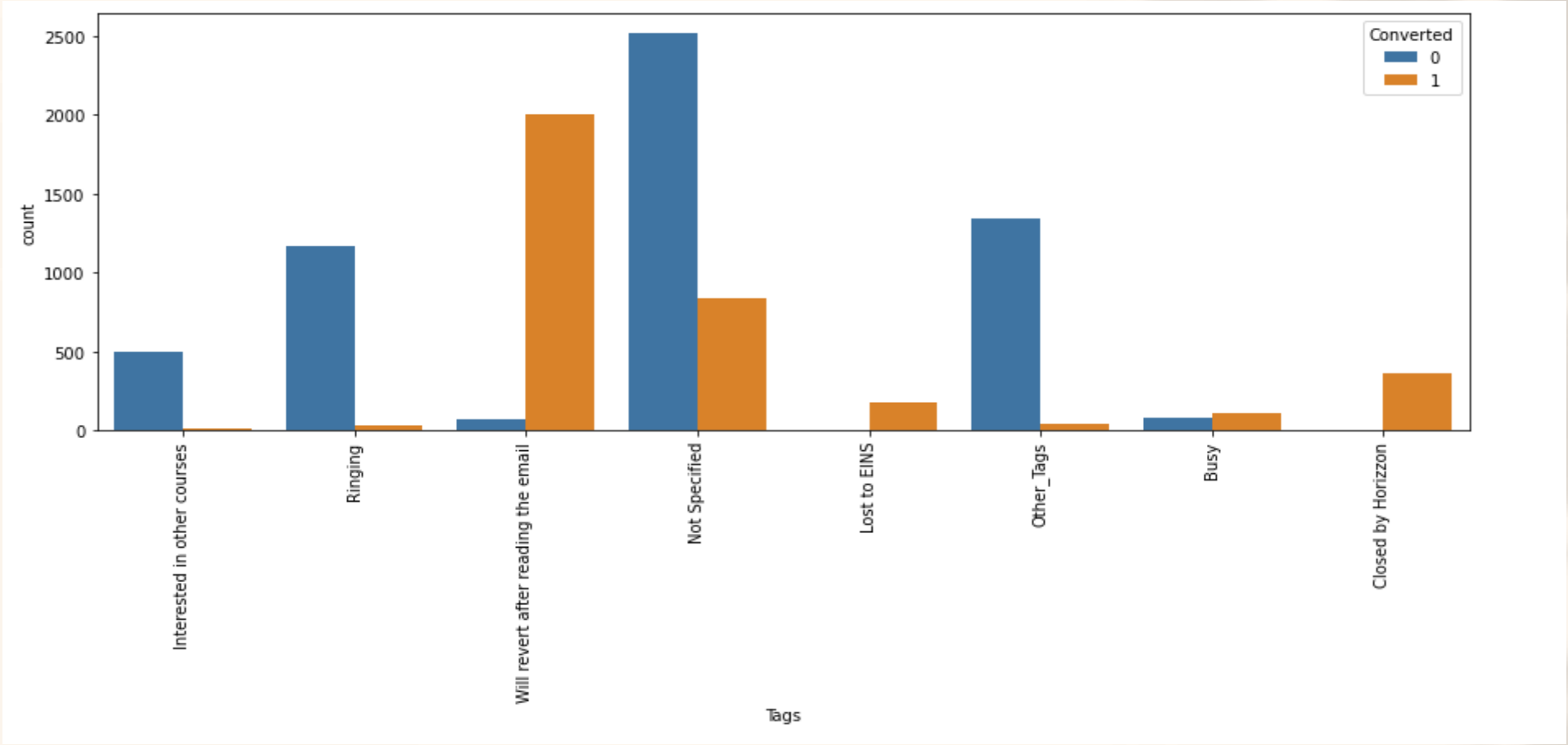
Specialization conversion Analysis

- We see that leads having different types of 'Management' specializations are more in number as well as have higher chances of getting converted.
- Hence lead score is higher for management professions and should be focused on to get more lead conversions.



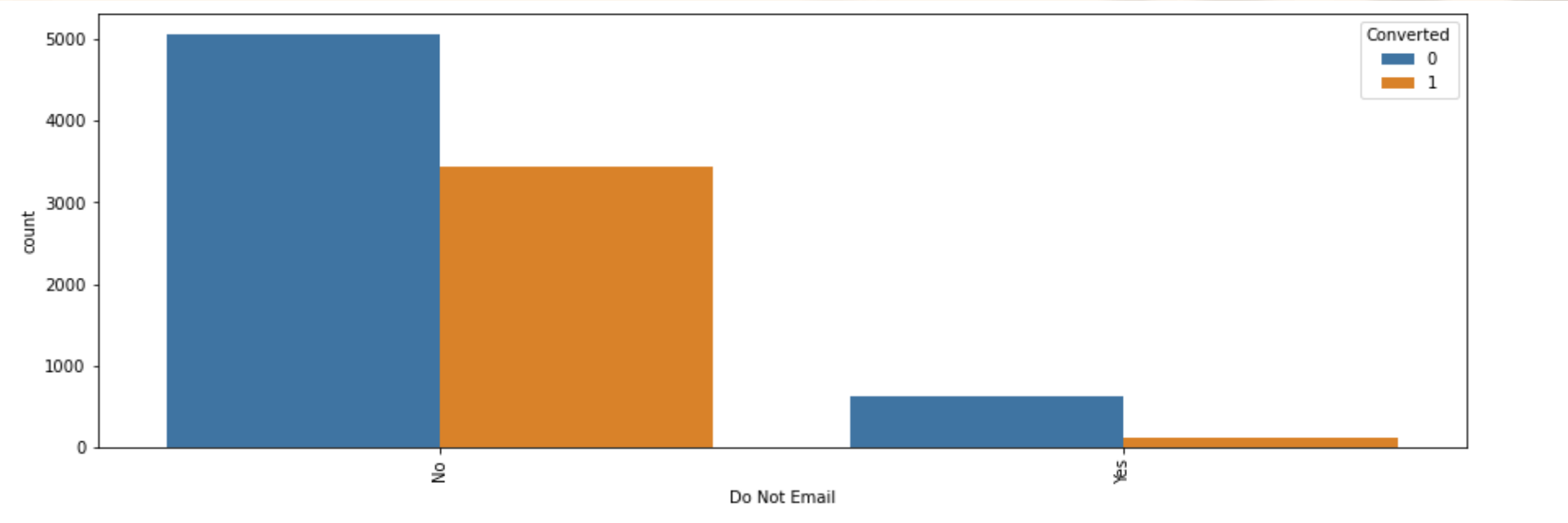
Tags vs Conversion Analysis

- There is a higher quantity of leads labeled as 'Ringing' and 'Not Specified,' so it is advisable to focus on optimizing conversions from these tags.
- Certainly! "Emphasizing efforts on leads labeled 'Will revert after reading mail' is recommended, as they exhibit the highest likelihood of conversion. Following closely are leads categorized as 'Lost To EINS,' 'Closed By Horizon,' and 'Busy.' Prioritizing attention on these specific leads can enhance overall conversion rates..



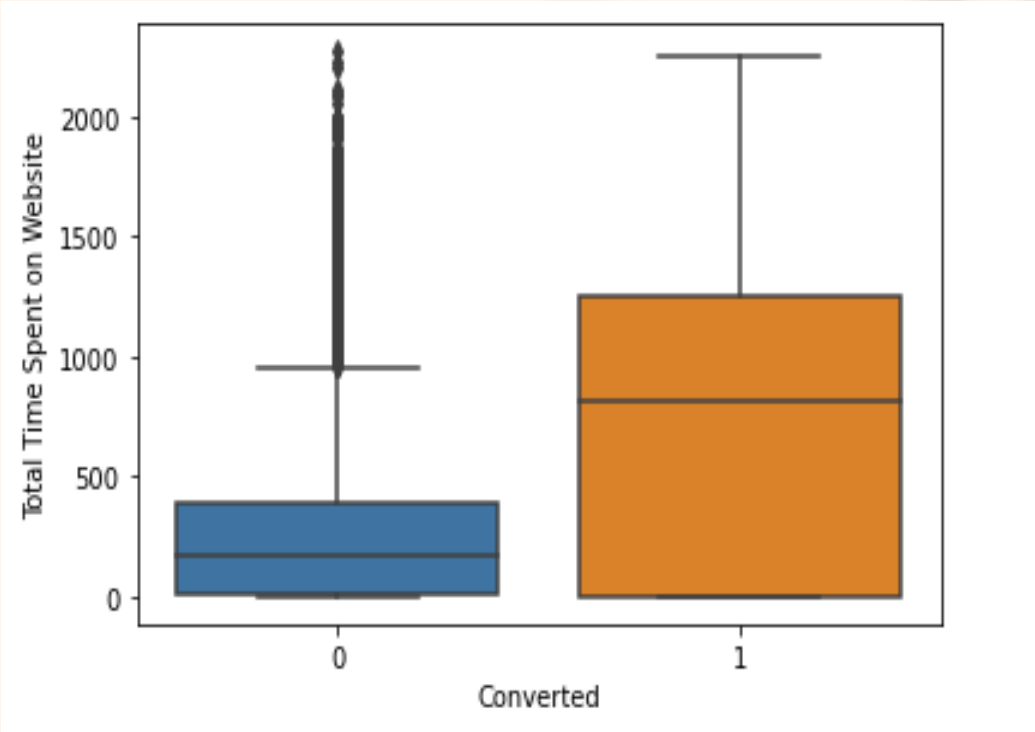
Do not Email vs Conversion Analysis

- Leads which do not object to an email communication have higher lead score than with leads which say yes to “Do not Email” option.
- More Focus should be put on pursuing the leads that say no to “Do not Email” to have a greater number of conversions.



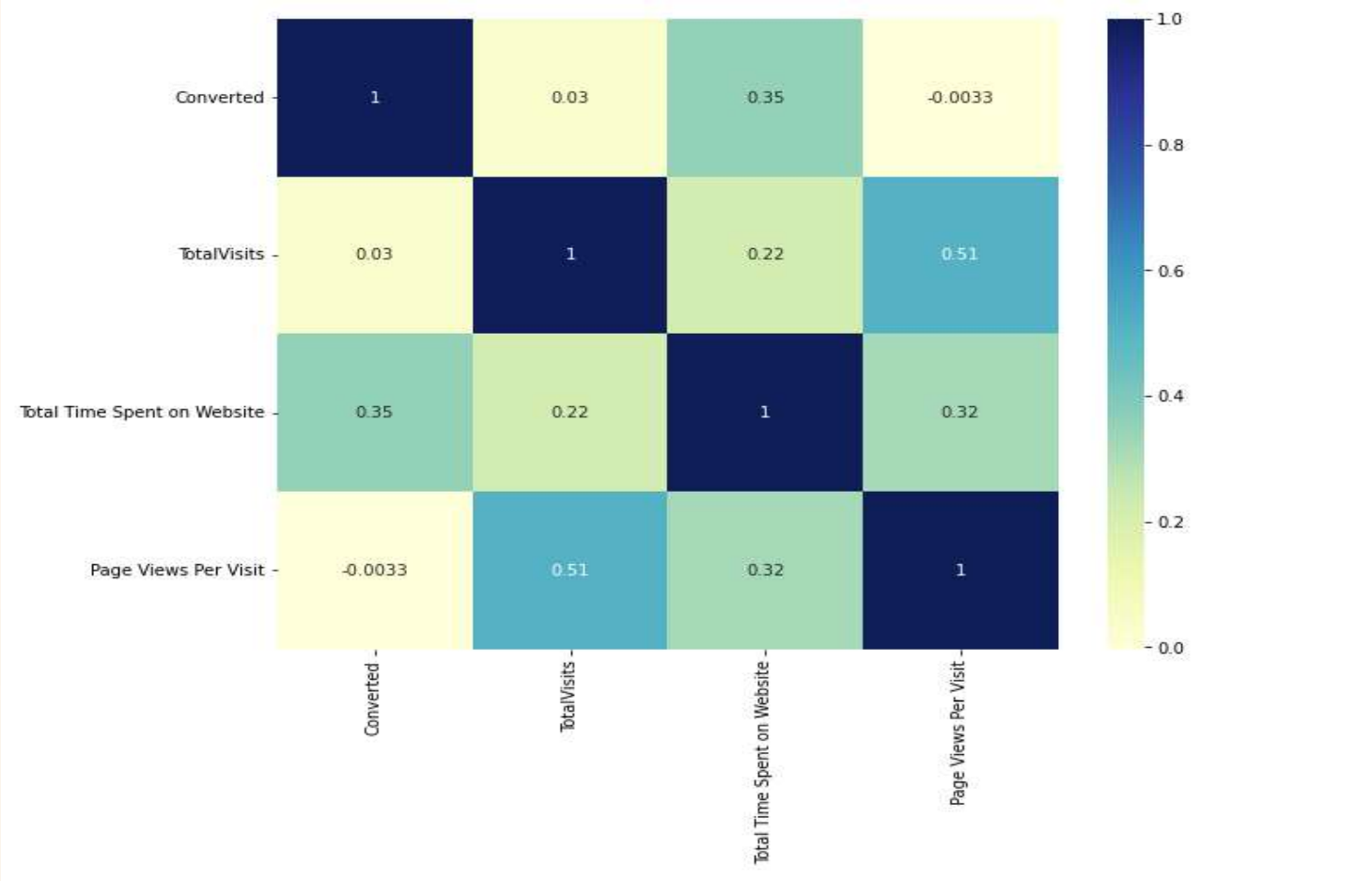
Total Time Spent On Websites vs Conversion Analysis

People spending higher than average time are promising leads. Hence will have higher lead score.



Correlation between various numerical variable using heatmap

'Total Visits' and 'Page Views Per Visit' have the most correlation with each other.

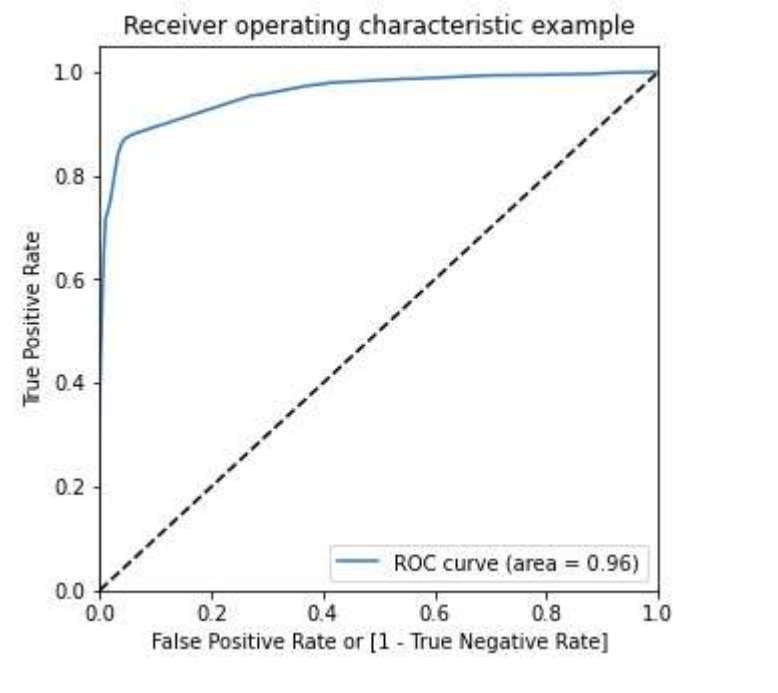


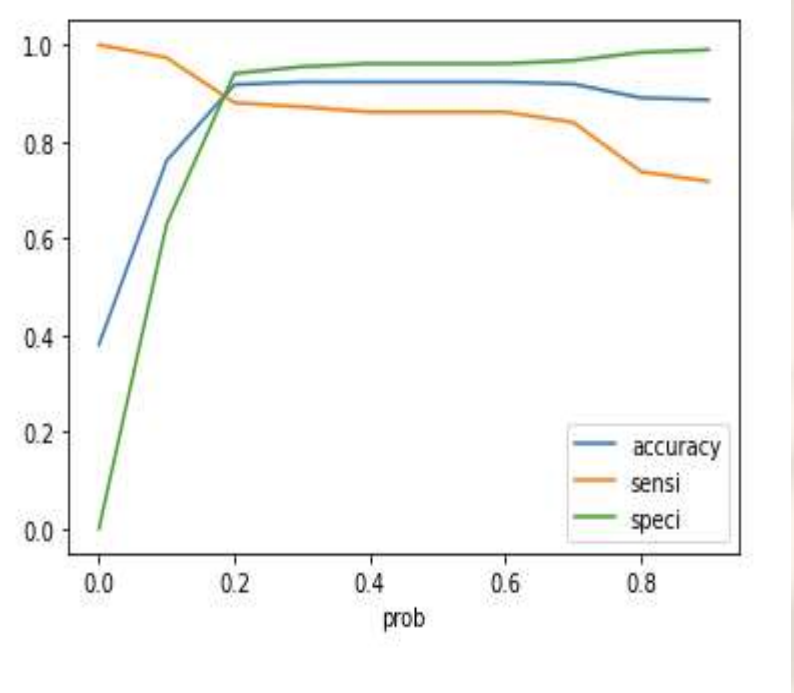
Model building steps and predictions

- Logistic Regression Model Building
  - Splitting The Dataset into train and test sets (70:30)
  - Rescaling the numerical variables
  - Model Building using Stats models and RFE (Total 5 models built)
  - Deriving Probabilities, Predictions and Lead Score on Train Data
  - Plotting confusion matrix and plotting ROC curve
  - Finding Optimal Cut-Off
- Making Predictions On Test Dataset With Model Built From Training Dataset

Final analysis on training data set

- ROC curve should be a value closer to 1 for a good model. We have got a value of 0.96 which is extremely good.
- Some important statistics of our model:
  - Accuracy: 91.75%
  - Sensitivity: 88%
  - Specificity: 94.04%





Final analysis on test data set

- Upon running the trained model on the test dataset, we obtain the following figures:
  - Accuracy: 92.14%
  - Sensitivity: 88.40%
  - Specificity: 94.39%

Final Observations

- Important variables that need to be focused on with high lead score include
  - Lead Source_Direct Traffic
  - Lead Source_Welingak Website
  - Last Activity_Email Bounced
  - Last Activity_Olark Chat Conversation
  - Tags_Busy
  - Tags_Closed By Horizon
  - Tags_Lost To EINS
  - Tags_Not Specified
  - Tags_Ringing
  - Tags_Will revert after reading the mail
  - Last Notable Activity_SMS Sent