LEAD SCORE CASE STUDY

Group Members:
Riya Bothra
Mansi Walunj
K.Abdul Wadood

Problem Statement

- X Education is an organisation that provides online courses for industrial professional. The company marks it's courses on several popular websites and search engine like Google.

- The company wants to select the most promising leads that can be converted to paying customers.

- The company has had the conversion rate of 30% through the whole process of turning leads into customers. Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion.

- Leads come through different modes like email, advertisements on website, google searches etc. The process of turning the leads into customers can be done by approaching those leads which are to be found having interest in taking the course. To make this process more efficient, the company wishes to identify the most potential leads, also knows as 'HOT LEADS'.

Business Goals

- The company requires to build a model to know the most promising leads.
- The company needs a model wherein 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 the lower lead score have lower conversion chance.
- The model to be built in lead conversion rate around 80% rate or more.

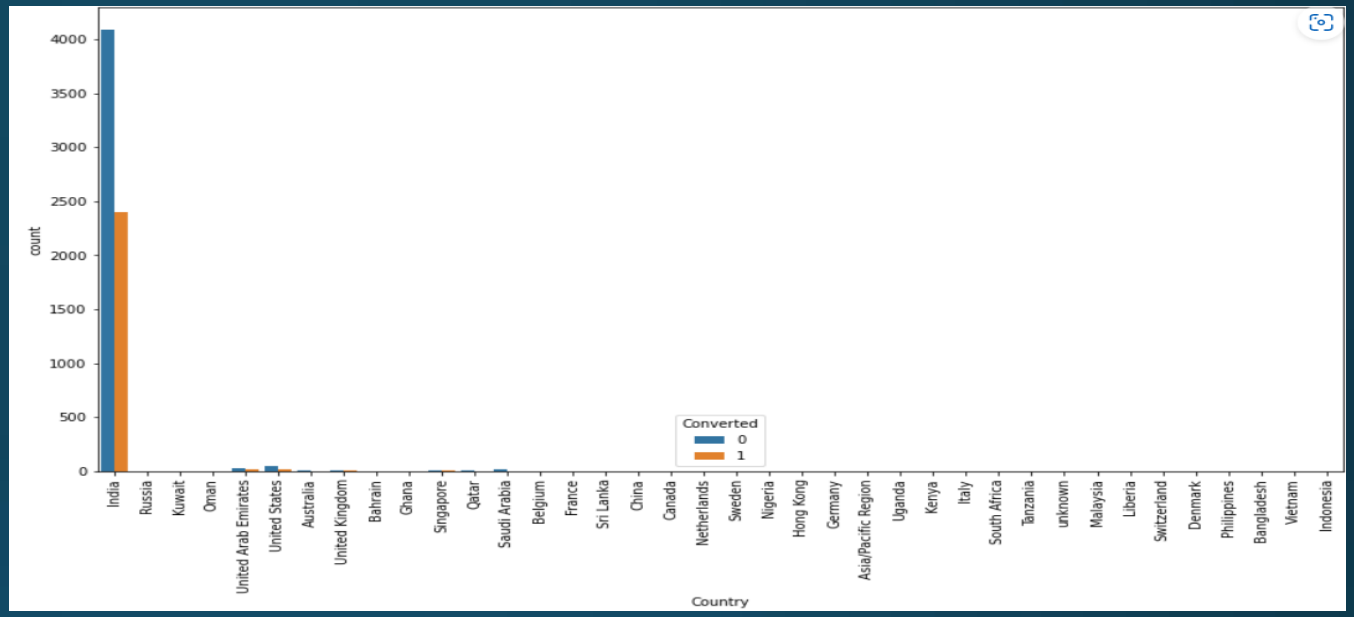
Solution Methodology

- Import the data for analysis.
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out the attributes for conversion.
- Dummy variable creation and Feature scaling
- Splitting the data into Test and Train dataset.
- Building a logistic regression model and calculate lead score.
- Test the model on train dataset.
- Evaluating the model by different measures and metrics.
- Test the model on test dataset.
- Measure the accuracy of the model and other metrics for evaluation.

Exploratory Data Analysis

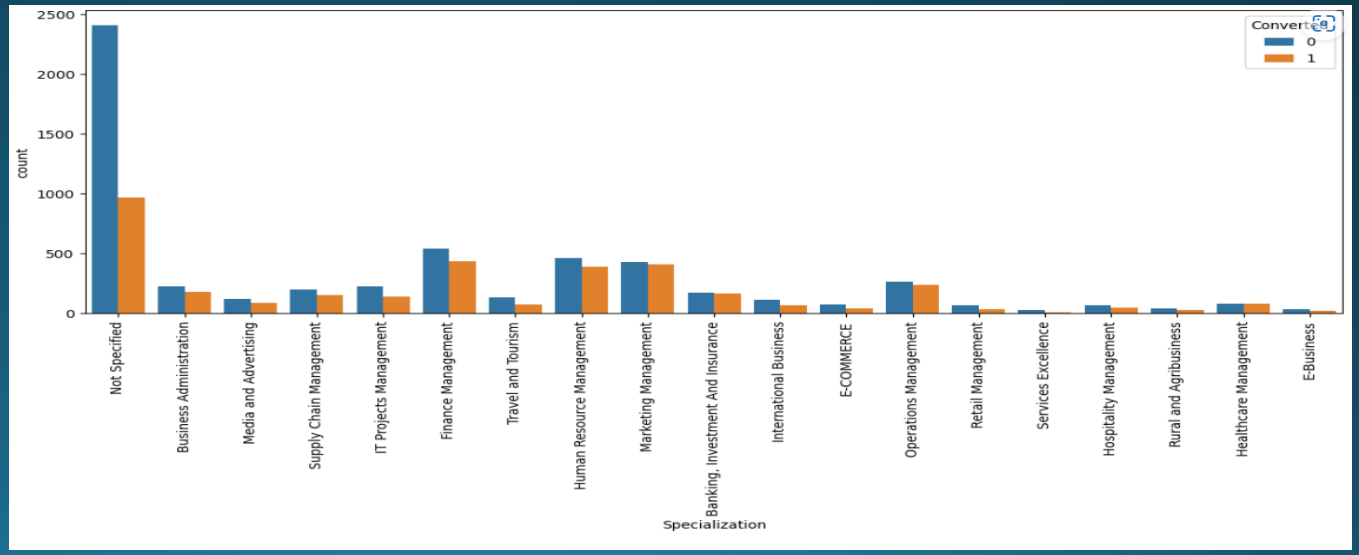
Country VS Converted

India has the highest conversions compared to other countries.



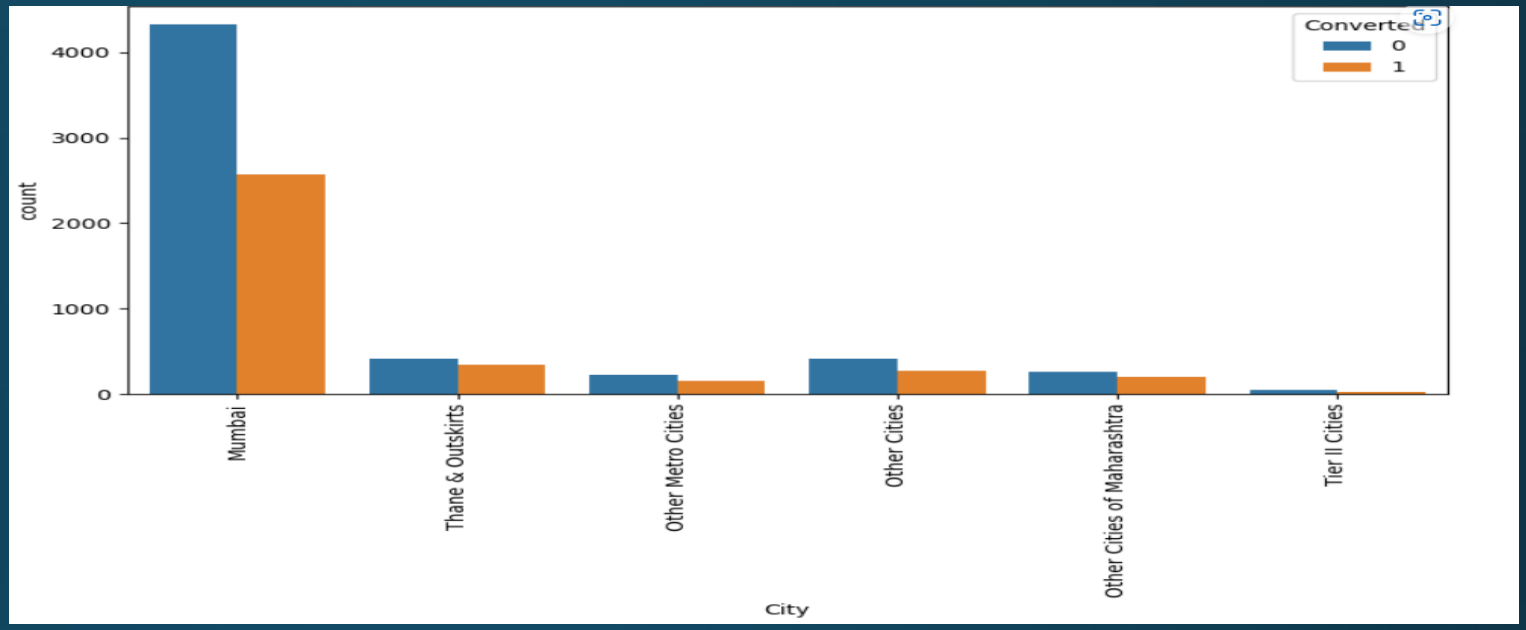
Specialization VS Converted

Most of the leads are Not Specified on the specialization variable. On the other hand, leads from Finance Management, Human Resource Management and Marketing Management can be promising leads as they have high conversion rate.



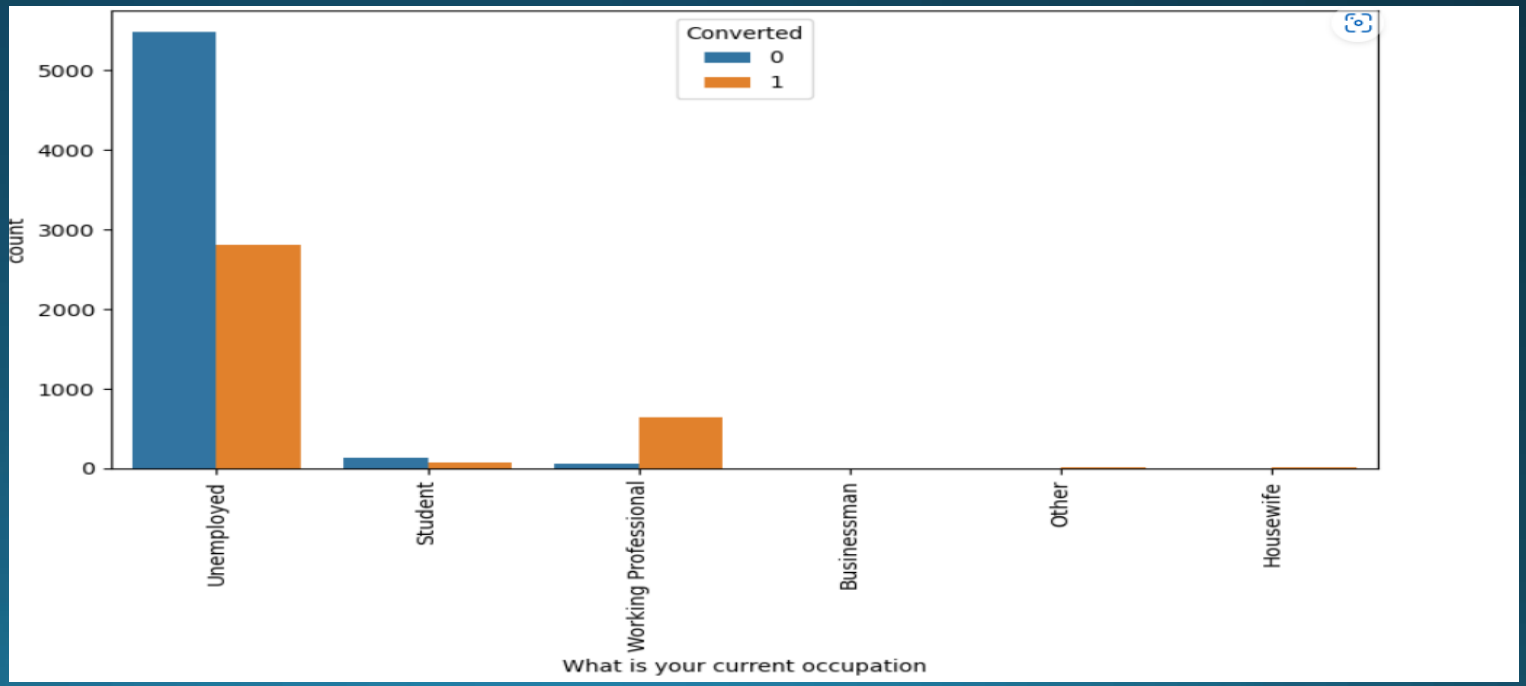
City VS Converted

Mumbai City has the highest conversion rate among all the cities.



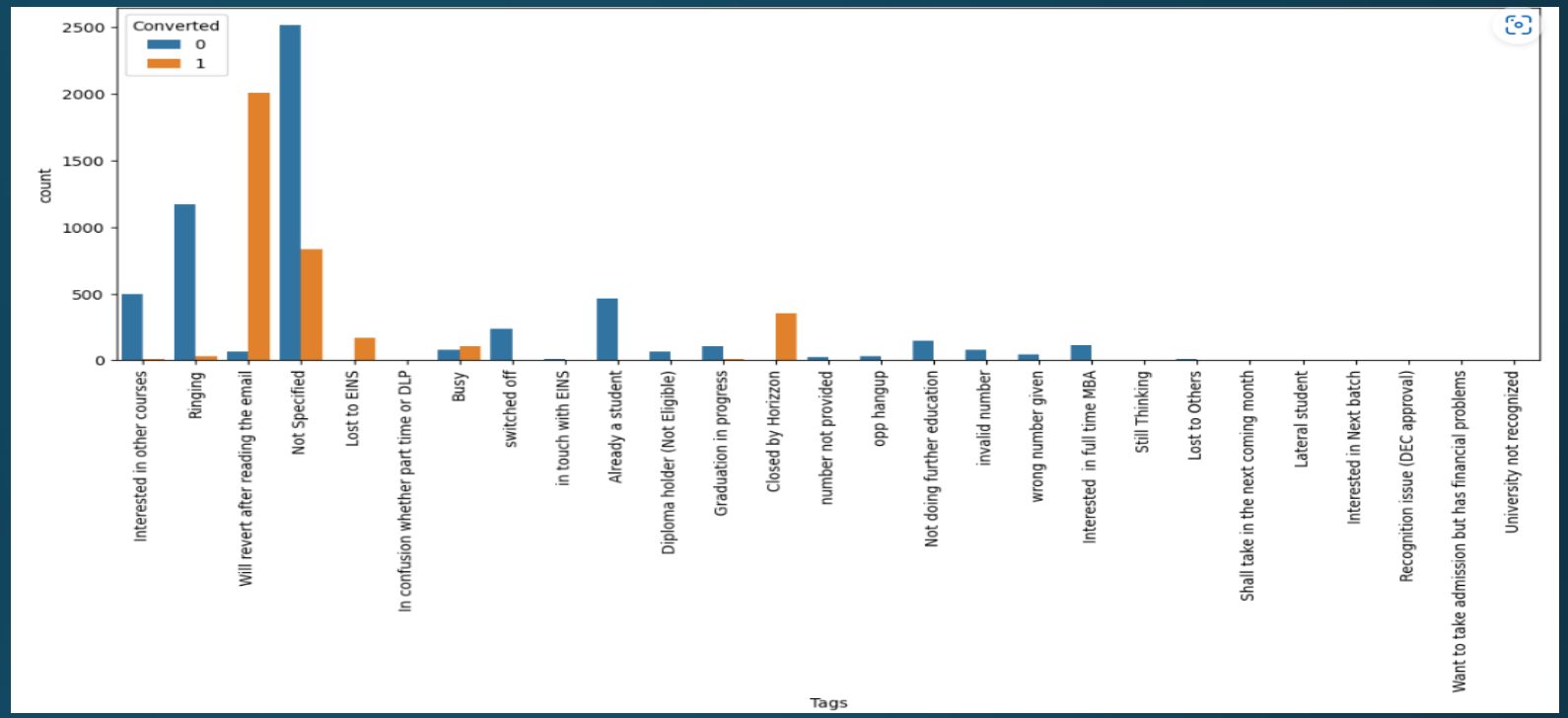
What is your current occupation VS Converted

Most of the people are Unemployed on the conversion. On the contrary, Working Professional also have the highest conversion rate from the current occupation variable.



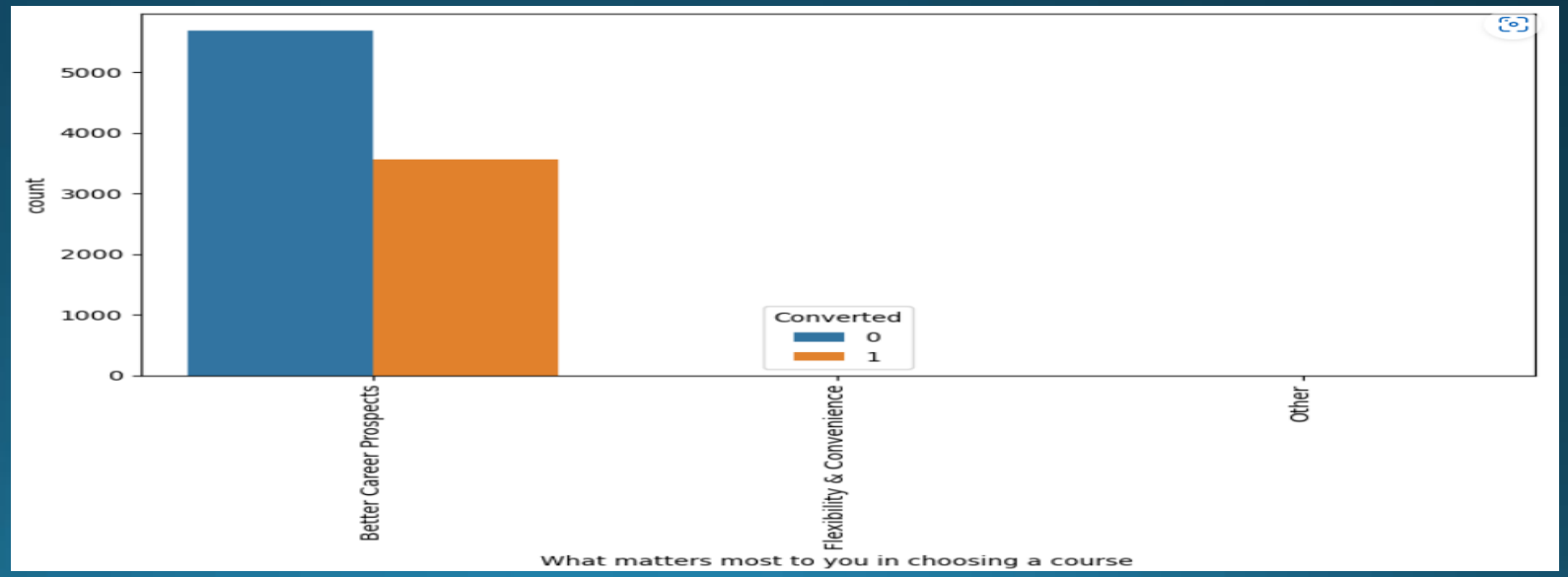
Tags VS Converted

Tags such as Not Specified, Will revert after reading the email and Ringing are used widely among the people with more conversion.



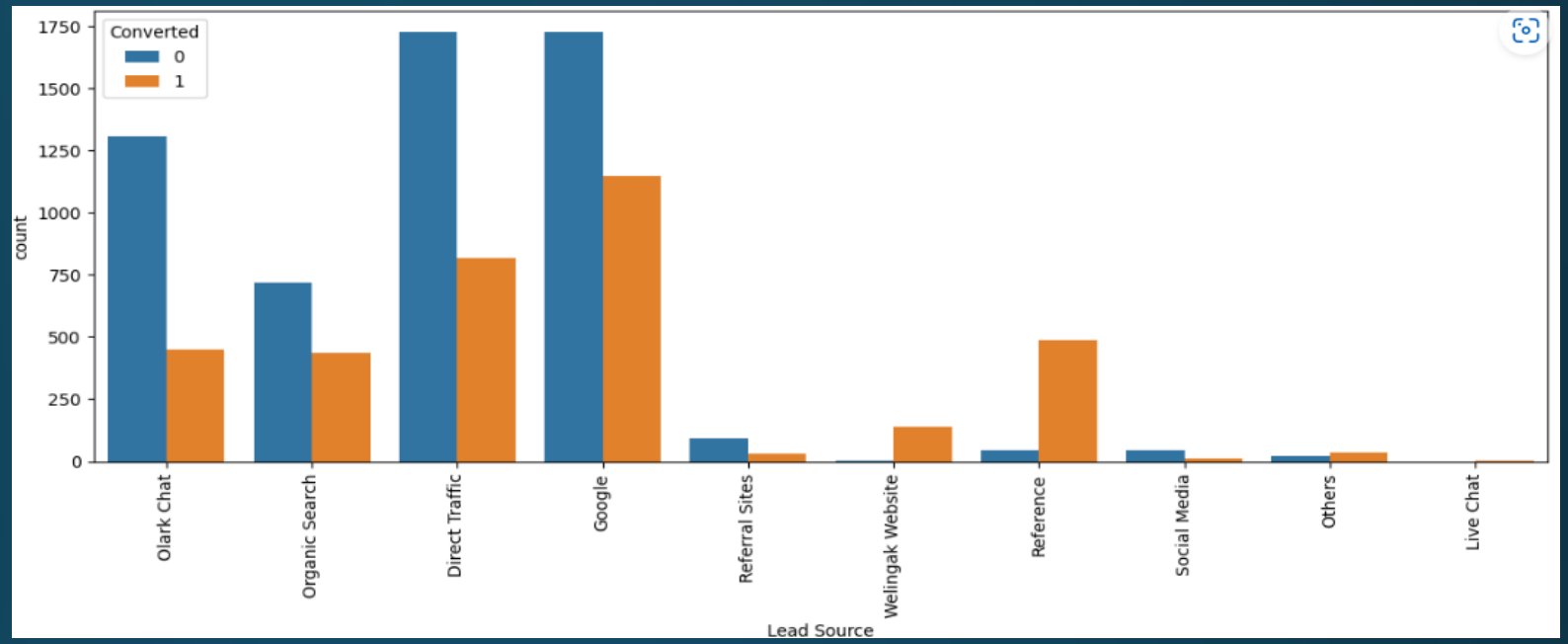
What matters most to you in choosing a course VS Converted

Better Career Prospects are the major reason for the people in choosing a course .



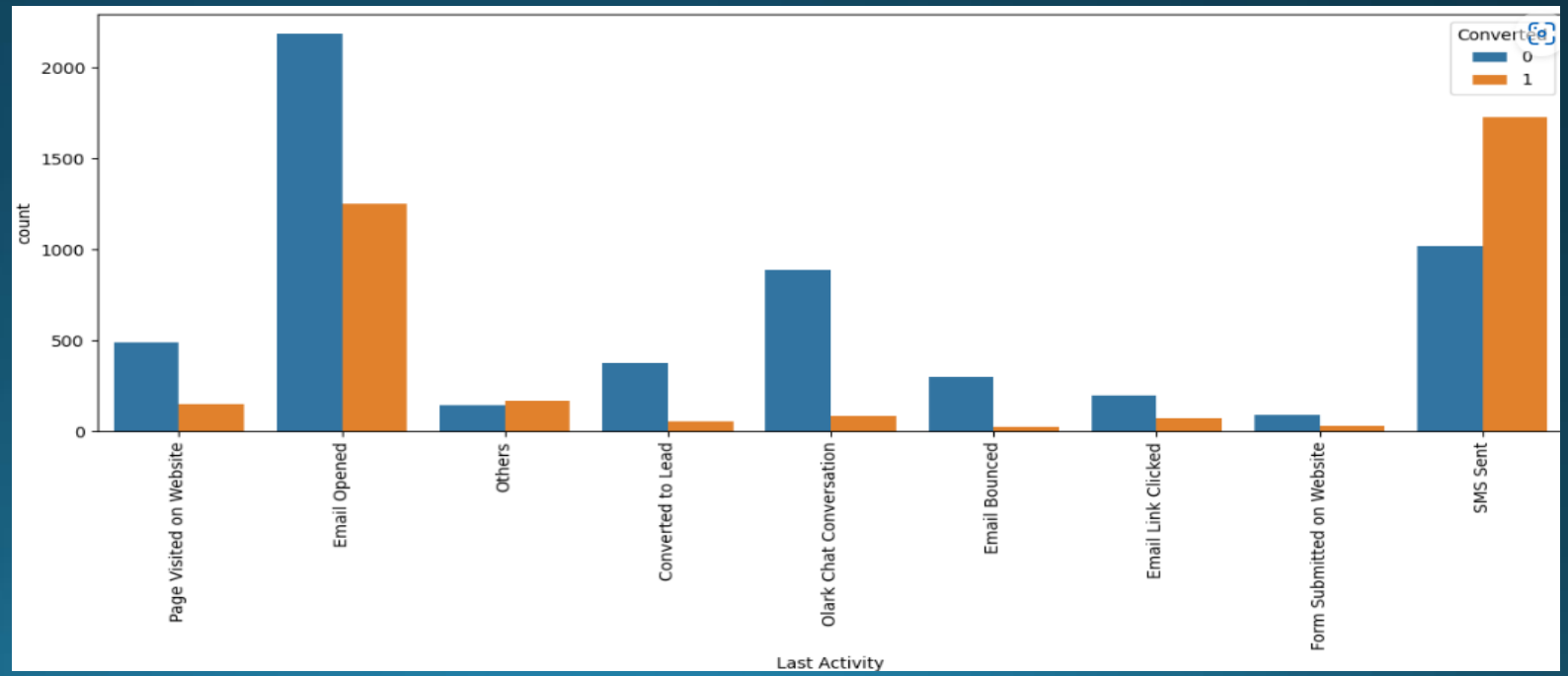
Lead Source VS Converted

Google searches has the highest conversion compared to other modes, whilst Direct Traffic has high conversion rate.



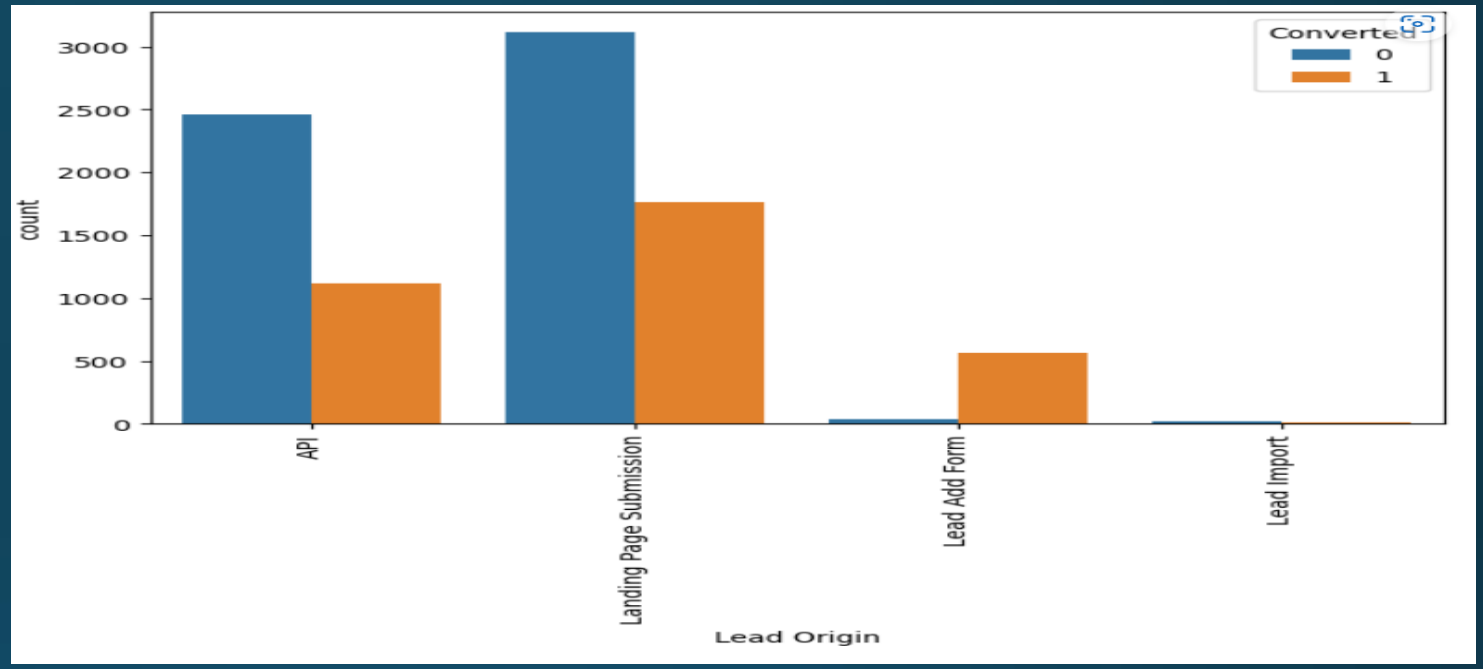
Last Activity VS Converted

Email opened has shown to be a promising method for getting higher confirmed leads, SMS Sent also has high conversion.



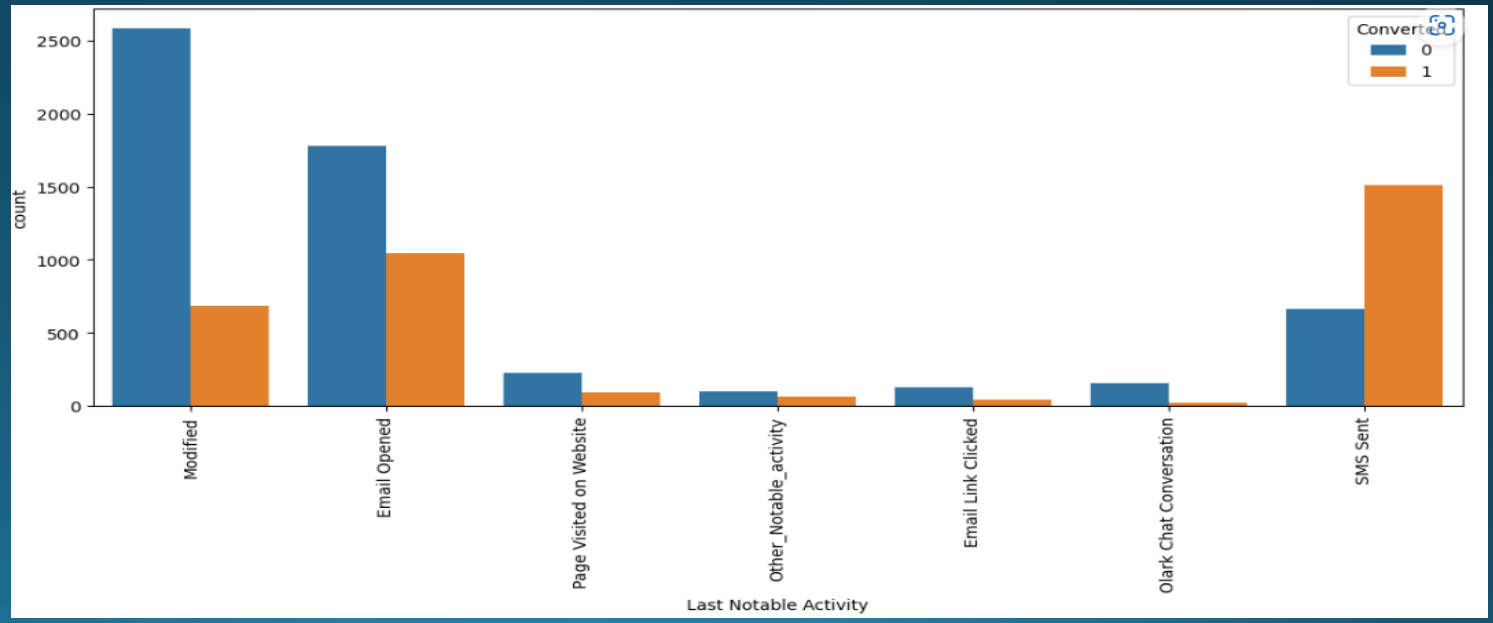
Lead Origin VS Converted

Landing Page Submission has high lead conversion than other Lead Origin.

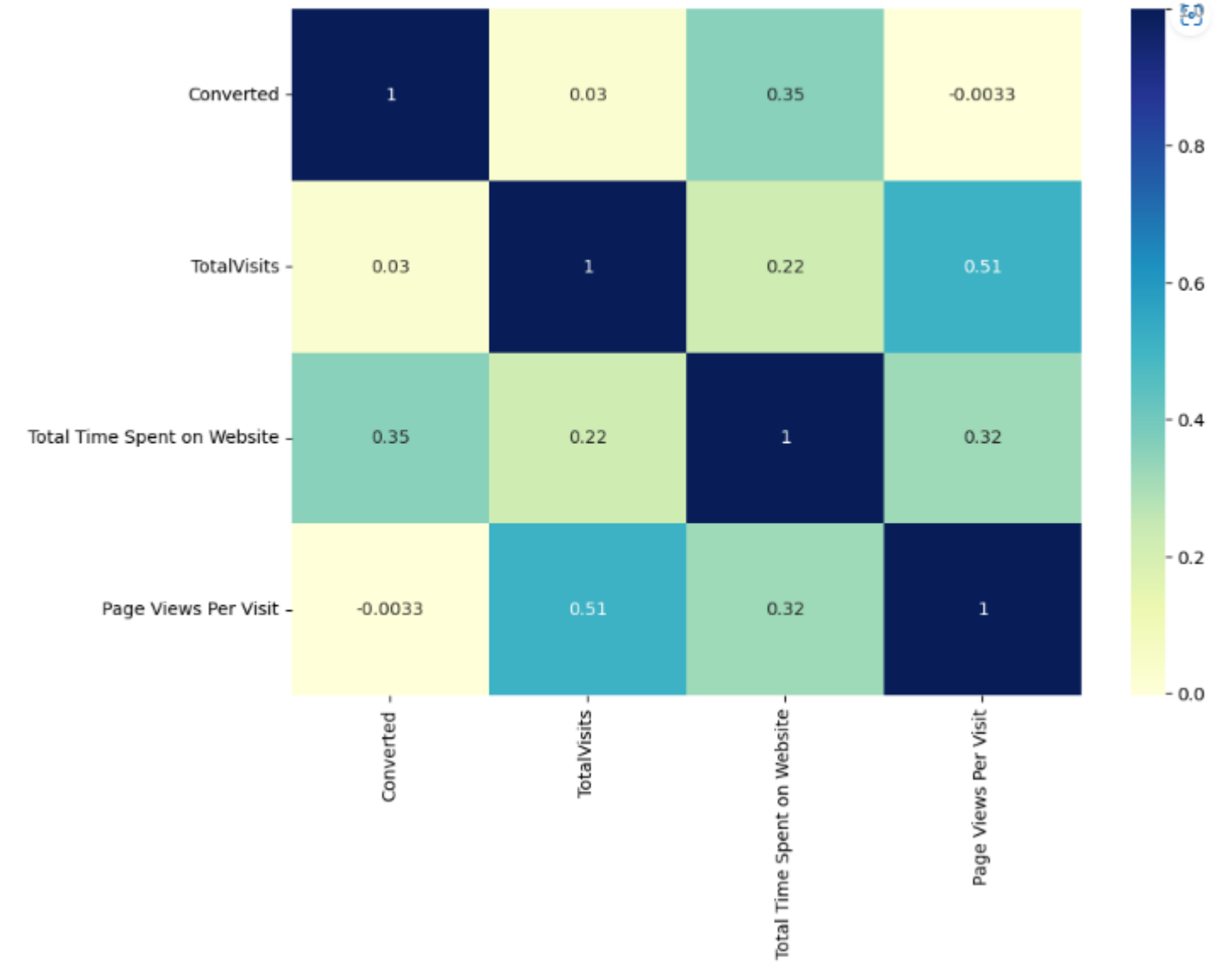


Last Notable Activity VS Converted

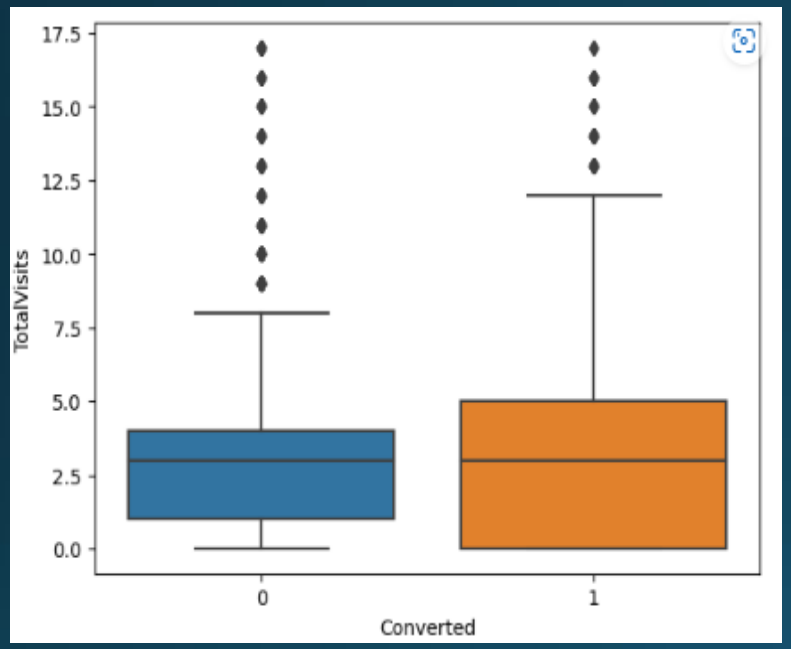
Most leads are converted with Modified Last Notable Activity. Email opened also induce leads.



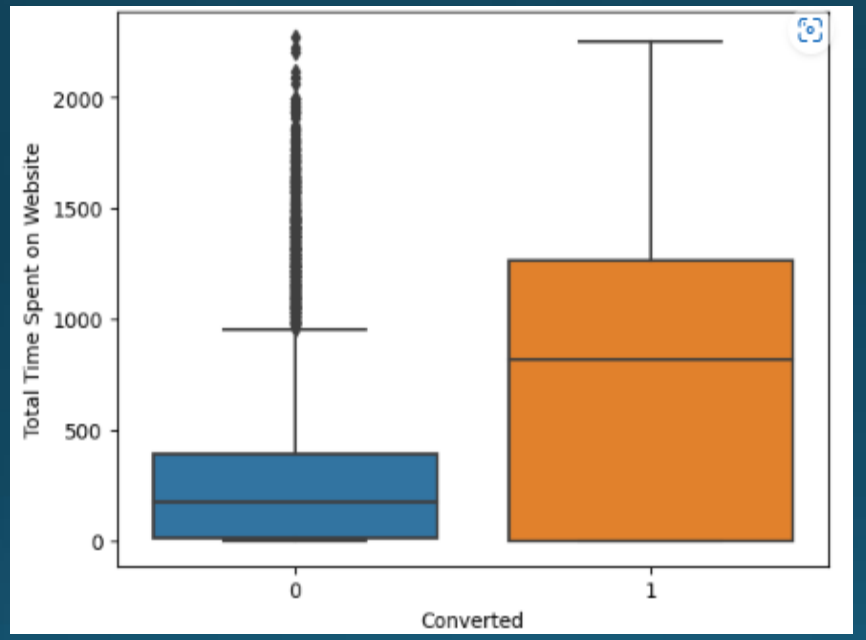
Heatmap of Numerical Variables



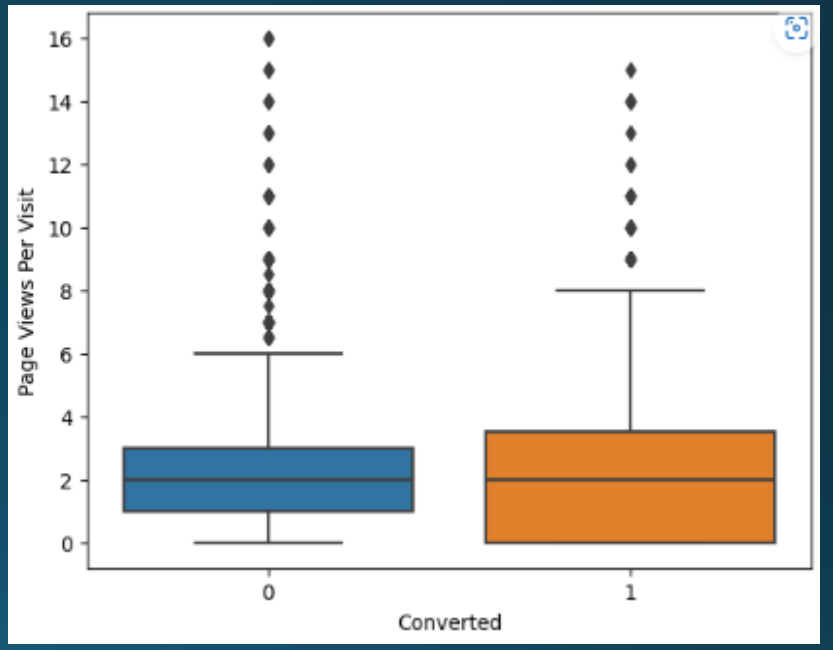
Visualizing Numerical Variables



Total Visits VS Converted



Total Time Spent on Website VS Converted



Page Views Per Visit VS Converted

Model Building

- Splitting the data into train and test data set.
- Scaling variables in train dataset.
- Build the logistic regression model.
- Use RFE to eliminate less relevant variable.
- Predict using train set.
- Evaluate accuracy and other metric.
- Predict using test set.
- Evaluation precision, recall and other metrics.

Model Evaluation (Train)

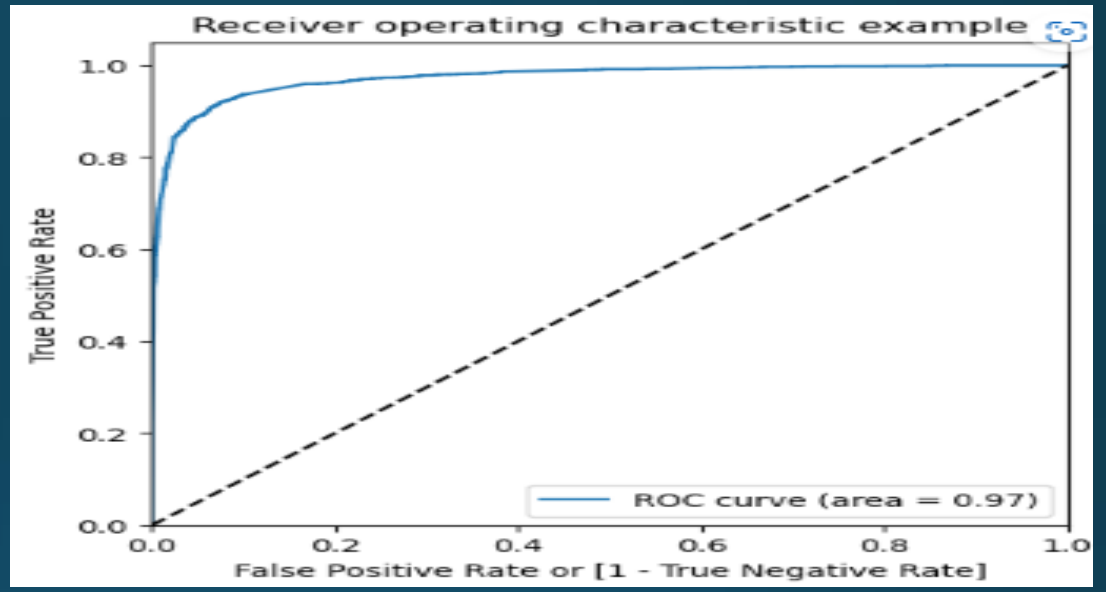
Plotting The ROC Curve and Evaluating Values:

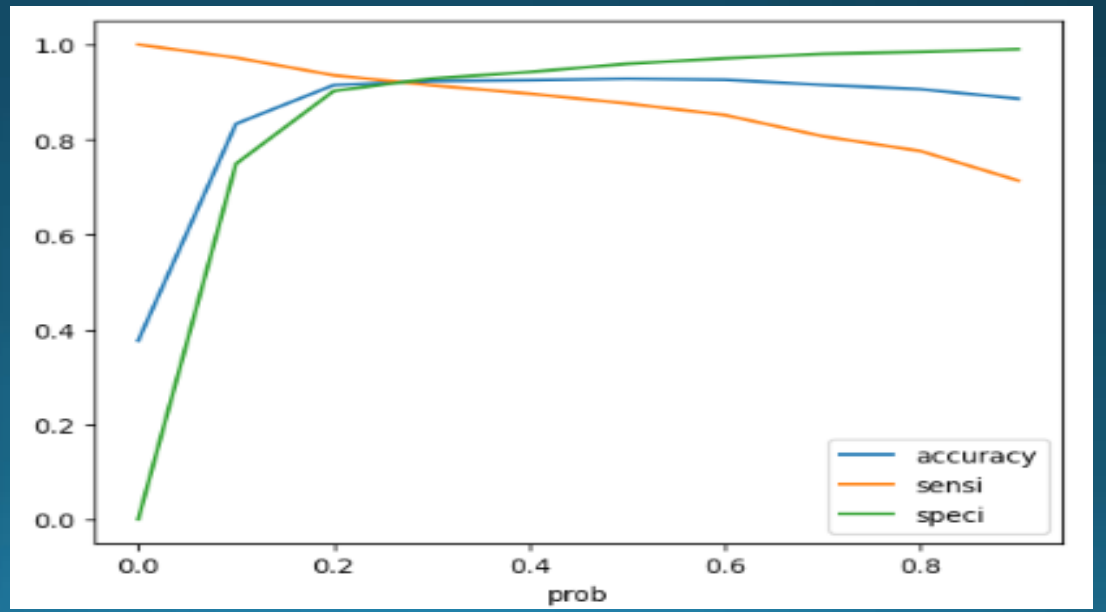
ACCURACY SENSITIVITY AND SPECIFICITY

- 92% Accuracy
- 88% Sensitivity
- 92% Specificity

3655	282
204	2173

Confusion Matrix

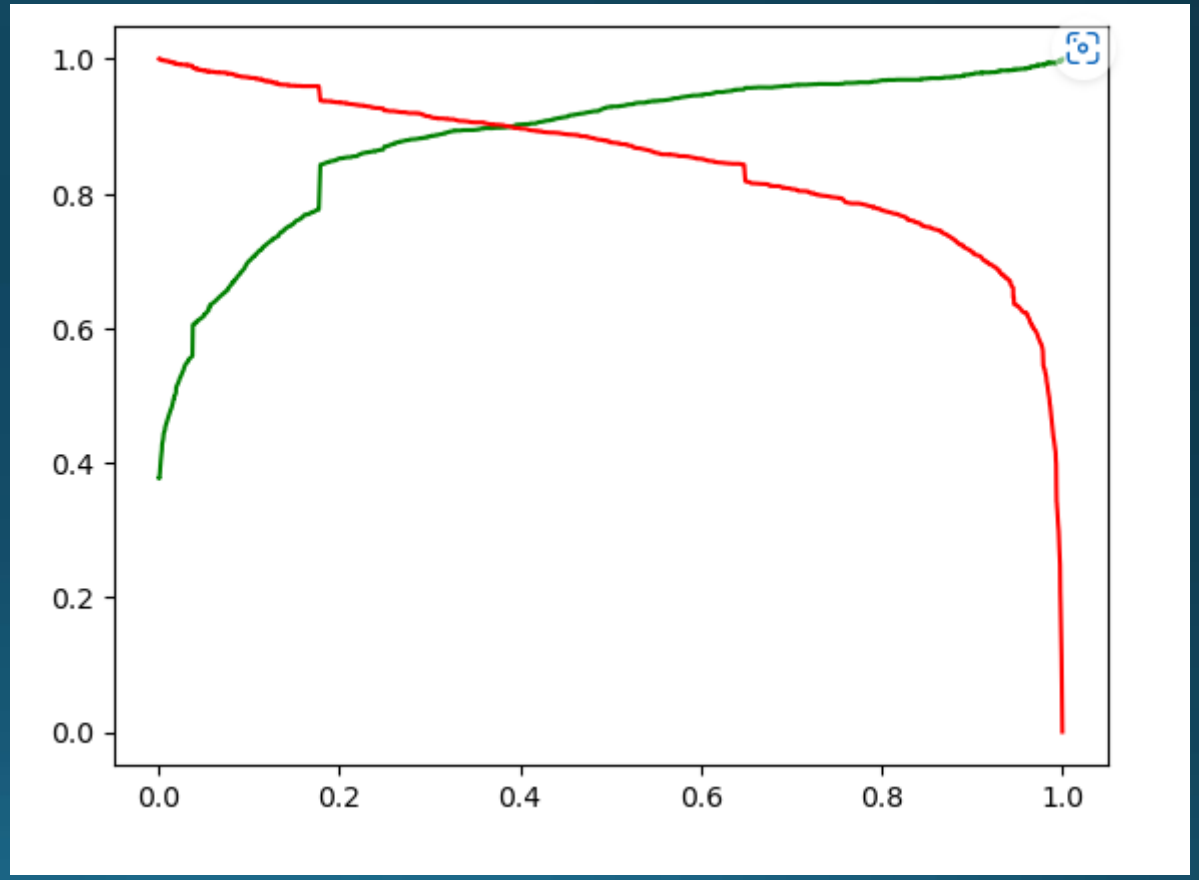




Model Evaluation (Train)

Precision And Recall:

- 89% Precision
- 92% Recall
- 89% Positive Predictive Value
- 95% Negative Predictive Value



Model Evaluation (Test)

- 92% Accuracy
- 88% Sensitivity
- 92% Specificity
- 88% Precision
- 92% Recall

1523	135
87	961

Confusion Matrix

Conclusion:

- The model shows the accuracy, sensitivity and specificity of test set around 92% , 88% and 92% which are approximately closer to the respectively values calculated using train dataset.
- The variables that contribute for lead getting converted in the model are,
  - Total Time Spent on Website.
  - When their current occupation is a Working Professional.
  - Lead Add Form from the Lead Origin.
  - When the SMS Sent from Last Activity.
  - When the Lead Course was Google, Organic Search, Direct Traffic and Welingak Website.
- The model finds correct promising leads and leads that have less chances of getting converted.
- Hence overall this model seems to be accurate.