



Lead Scoring Case Study



Problem Statement

X Education Company provides online sources for Industry professional.

Currently, they are receiving leads through multiple channels for these courses. Since these leads are all over the place resulting the conversion rate low which is just 30%.

Instead of following up on all these leads, company wants to find promising leads.

X Education, wants to build a model with provided data and identify leads which will have 80% conversion rate.

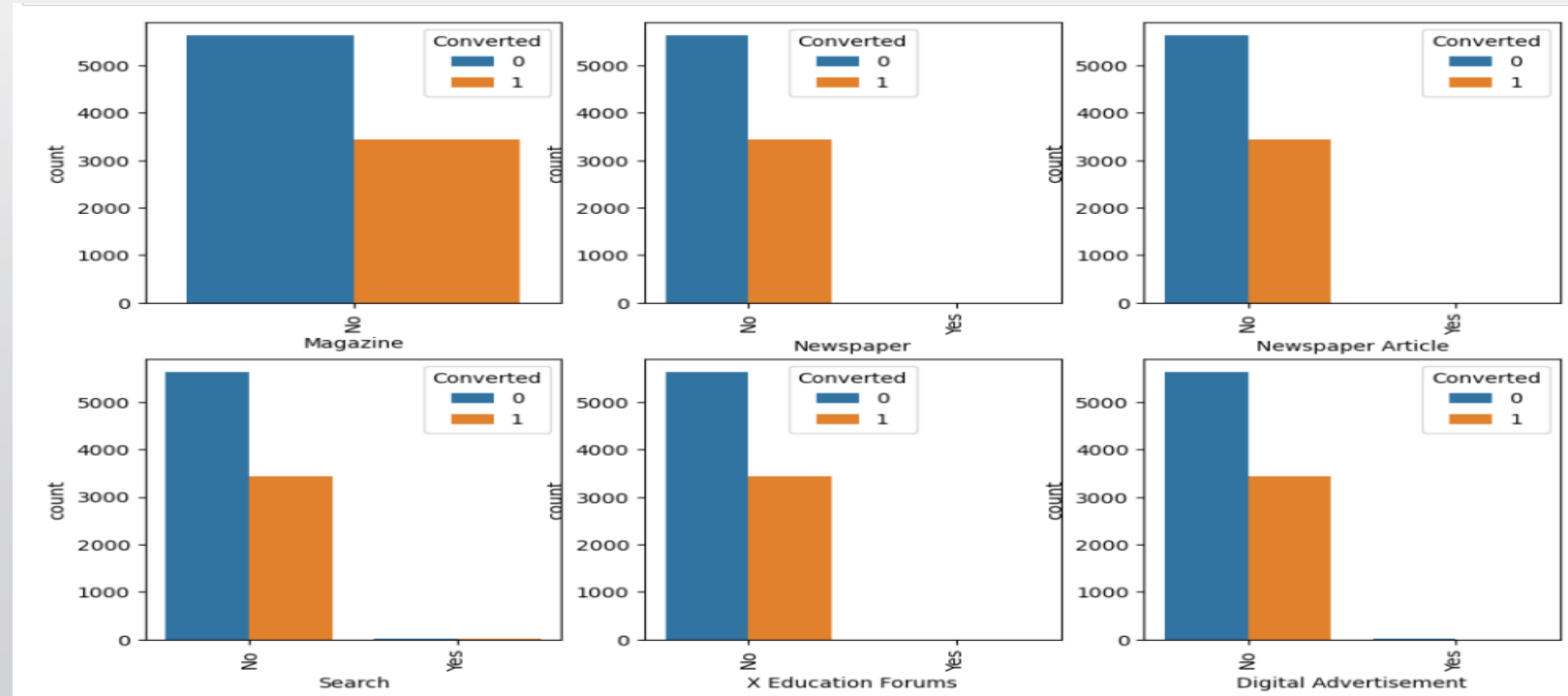


Analysis Approach

- ▶ Understanding data and what each feature means based on data dictionary provided.
- ▶ Handle missing values
- ▶ Remove unnecessary columns or row not showing variation or contributing much to the conversion rate based on EDA.
- ▶ Preparing data for model building
- ▶ Building model. Eliminating features with high VIF or P value and rebuilding model
- ▶ Analyse model with optimal cutoff to have better accuracy, sensitivity and specificity
- ▶ Make prediction on test data
- ▶ Compare train vs. test data results if meeting the criteria of 80% conversion rate

Analysis

- Features 'What matters most to you in choosing a course', 'Search', 'X Education Forums', 'Digital Advertisement', 'Magazine', 'Newspaper', 'Newspaper Article', 'Through Recommendations', 'Country', 'City', 'Receive More Updates About Our Courses', 'I agree to pay the amount through cheque', 'Update me on Supply Chain Content', 'Tags', 'Get updates on DM Content' are not considered for analysis as they are having monotonous values and would not contribute to the final results.



Analysis continue..

1. Lead Origin - Leads originated from API and Landing submission page are high but conversion ratio is poor. Need to focus here for conversion rate.

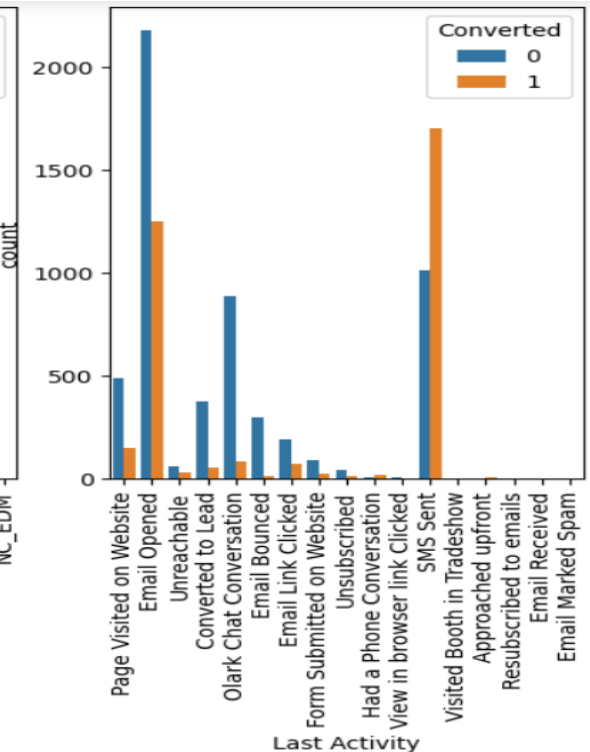
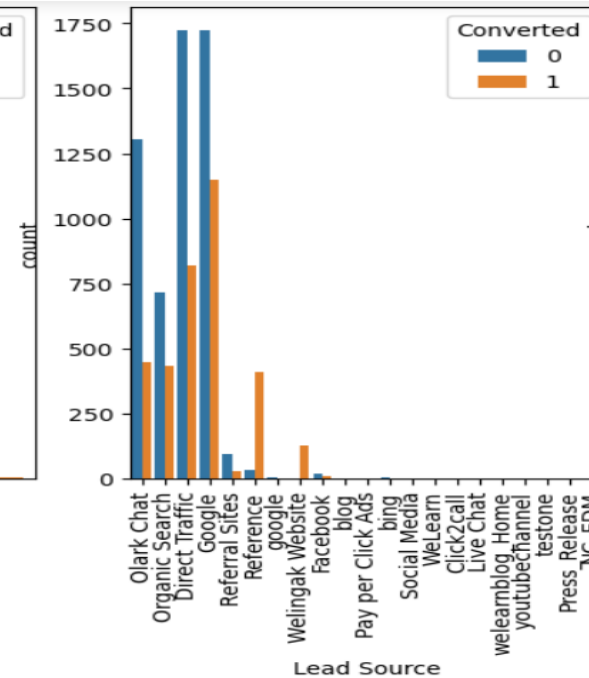
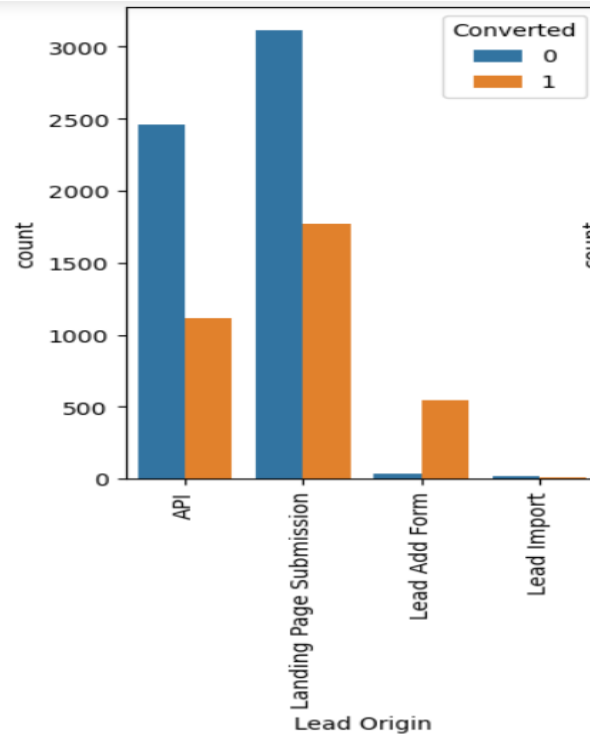
Between 30-40% getting converted for these 2 category.

Lead add form has highest conversion (ration >90%) but the count is substantially low against overall count.

2. Lead Source - Olark chat, Direct traffic and Google are generating most leads.

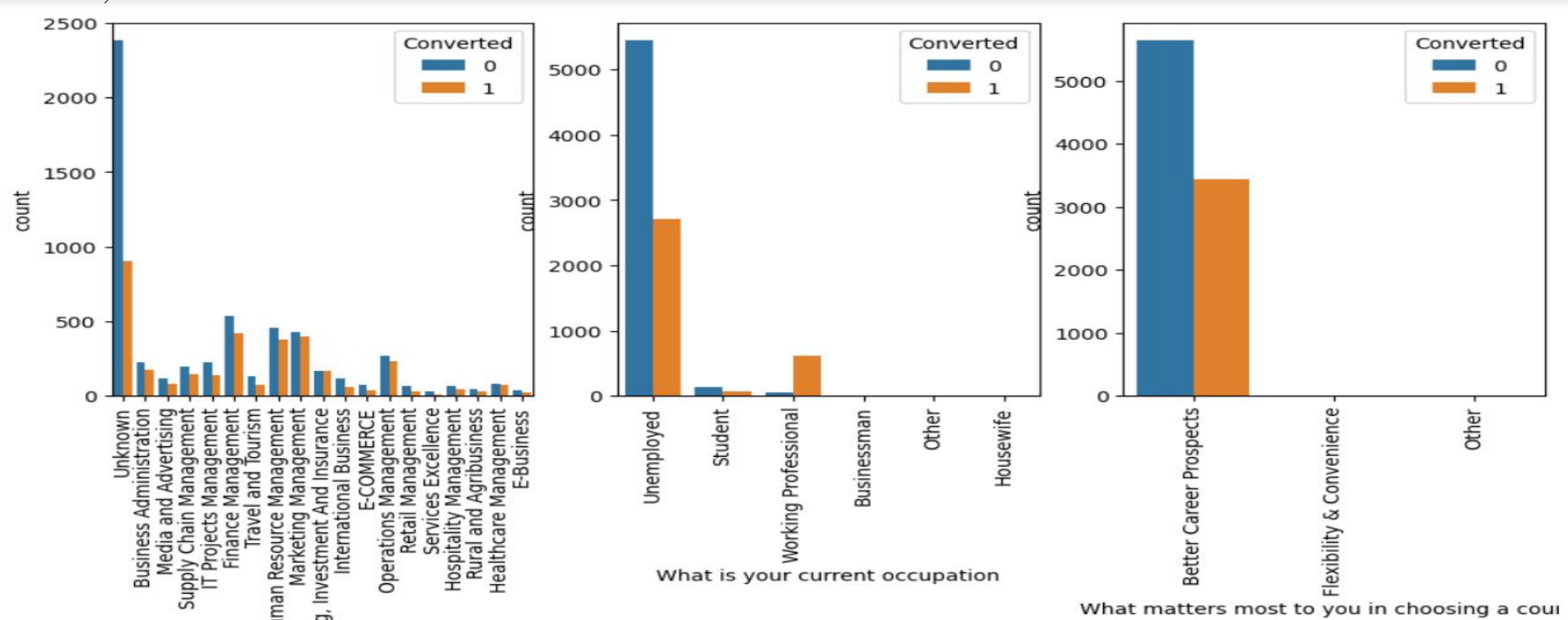
Reference and Welingak wesite have highest conversion rate but not huge traffic from these 2.

3. Last Activity - Email and SMS are major activities done by visitors . Visitor with SMS activity have high conversion ration >50%



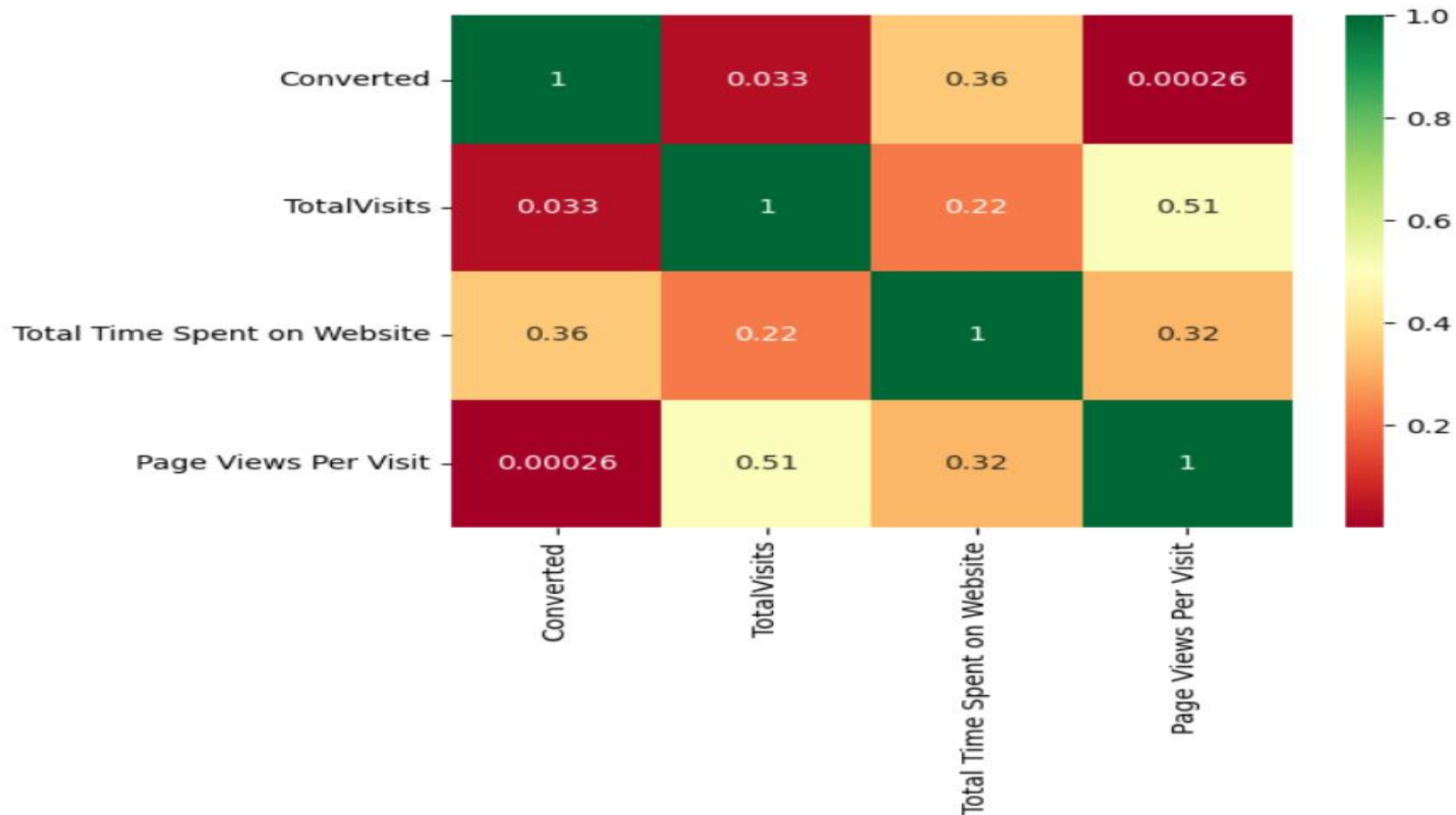
Analysis continue..

1. Specialization- Finance management, human resource management, marketing management, operations management , banking, investment and insurance specialized leads have high conversion ratio.
2. Current Occupation - Unemployed generate most lead but working professionals have high conversion ratio.
3. Matters most in choosing course- Better career prospect have most leads generated.



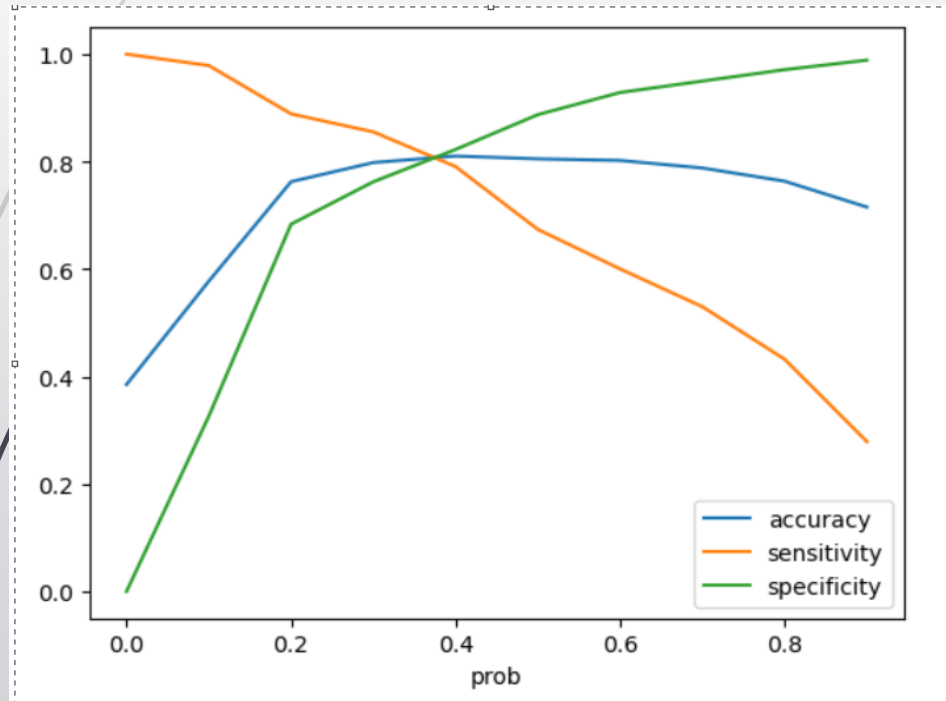
Analysis continue..

Here we cannot see any strong linear relation between variables. 'Total visit' has some collinearity with 'Page Views Per Visits'.



Model Evaluation – Sensitivity and Specificity on Train Data

- The optimal cutoff is 0.37 based on Accuracy, Sensitivity and Specificity from the below graph .



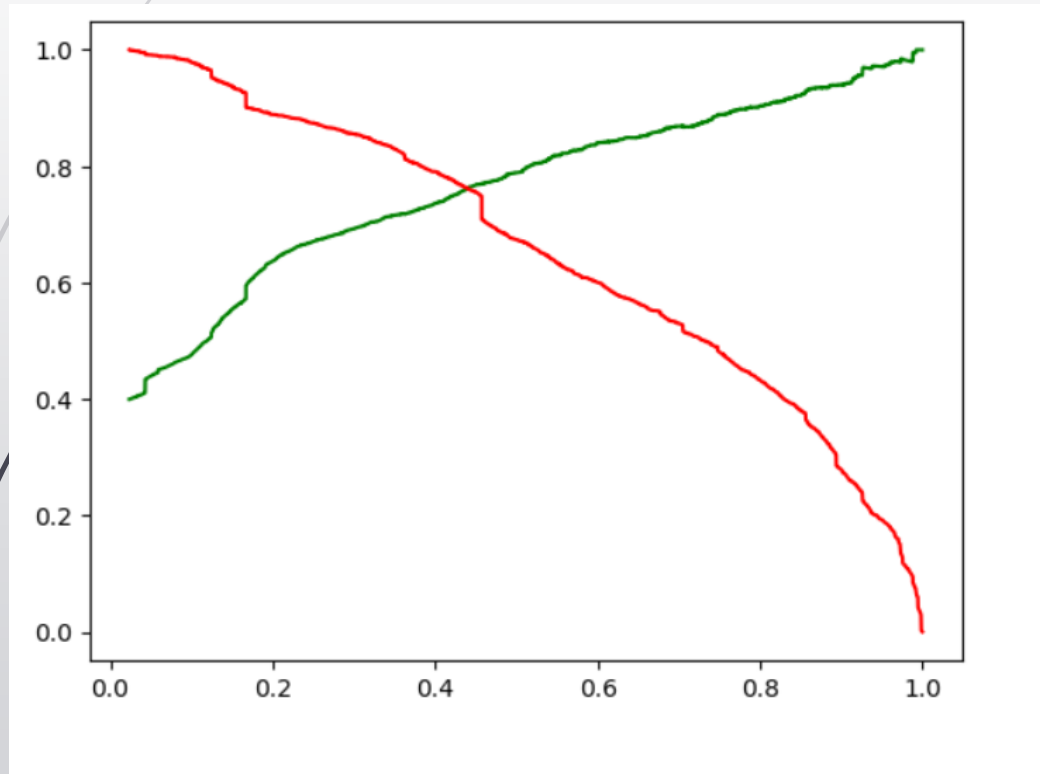
Confusion Matrix

```
[[3141, 764],  
 [ 470, 1976]]
```

- Accuracy – 80.5%
- Sensitivity – 80.7%
- Specificity – 80.4%
- False positive rate – 19.5%
- Positive predictive value – 72.1%
- Negative predictive value – 86.9%

Model Evaluation – Recall and Precision on Train Data

- The optimal cutoff is 0.42 based on Precision and Recall from the below graph .



- Precision – 78.9%
- Recall – 67.3%

Model Evaluation – Sensitivity and Specificity on Test Data

Confusion Matrix

```
[[1400, 334],  
 [ 213, 776]]
```

- Accuracy – 79.9%
- Sensitivity – 68.4%
- Specificity – 80.7%

Recommendation for Business

- Consider working professionals leads as they have higher chance of conversion.
- Welingak website generated leads seems authentic and should be followed up for conversion.
- Higher the time spent on official website contributing positively to conversion rate
- Phone conversation activity and Add form lead origin also contributing positively
- Avoid following up any activity related to Email as they are contributing negatively.
- Avoid unemployed , students crowd as they have most contribution in lead generation but low conversion rate.
- Avoid leads from Olark chat.