Lead-Scoring Case Study

Rohit Kumar & Rohini Parthasarathy

Problem Statement

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The business wants us to create a model in which every lead must be given a lead score, and based on the lead score, the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. Specifically, the CEO has stated that an approximate 80% lead conversion rate is the goal.

Business Objective

- The business objective is to find the prospective 'Hot Leads' that eventually gets converted to a customer who takes up a course.

- Also, the business wants the model to be designed to achieve the conversion ratio of 80%.
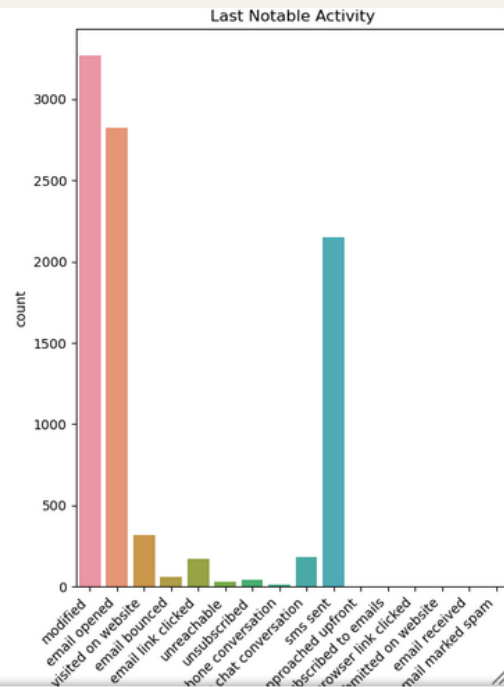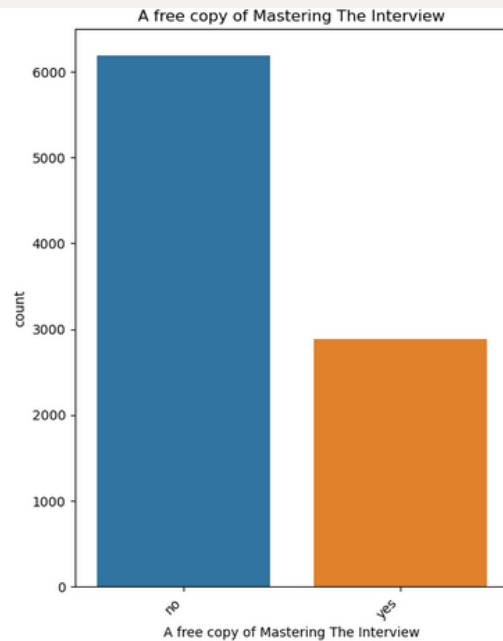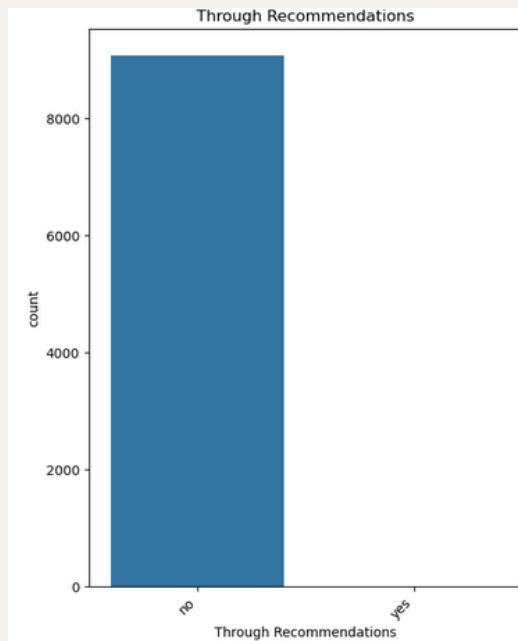
Steps involved in the Assignment

1. Reading the data
2. Cleaning the data
3. EDA
4. Creating Dummy variables
5. Splitting data into train and test set
6. Building Model
7. Making Predictions
8. Model Evaluation(sensitivity and specificity)
9. ROC Curve
10. Prediction on test set
11. Precision- Recall tradeoff

Univariate Analysis
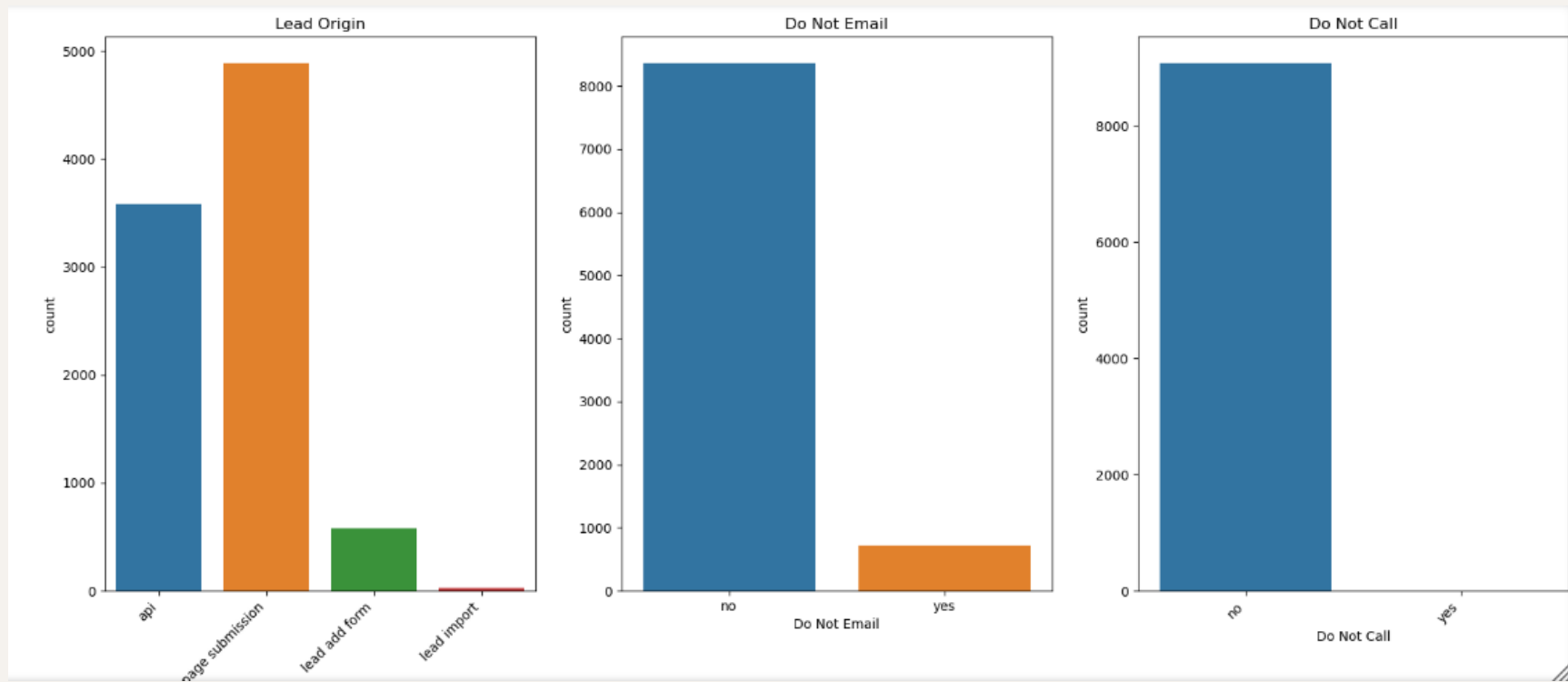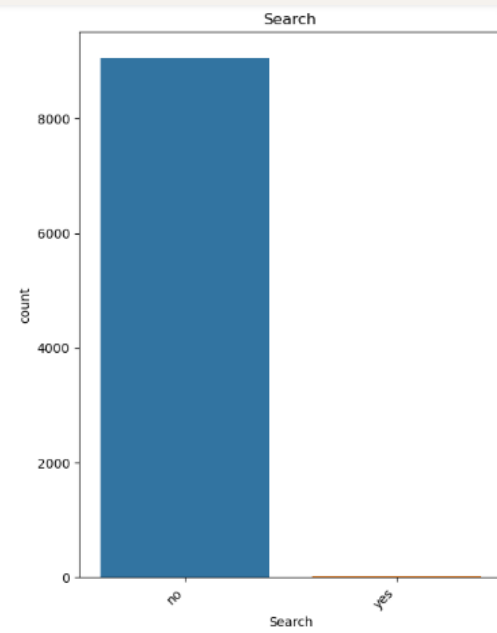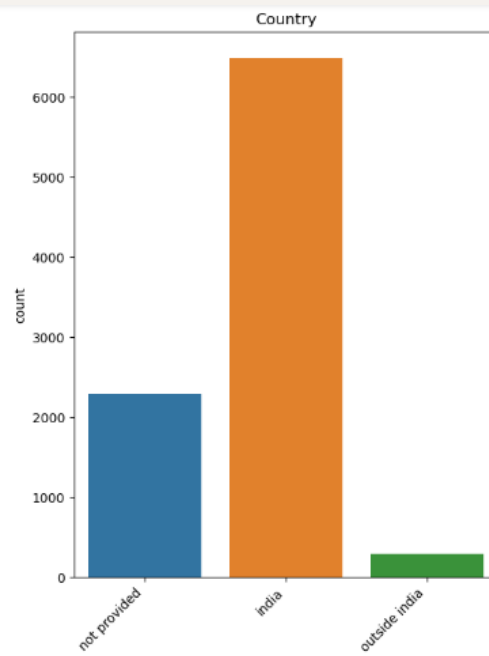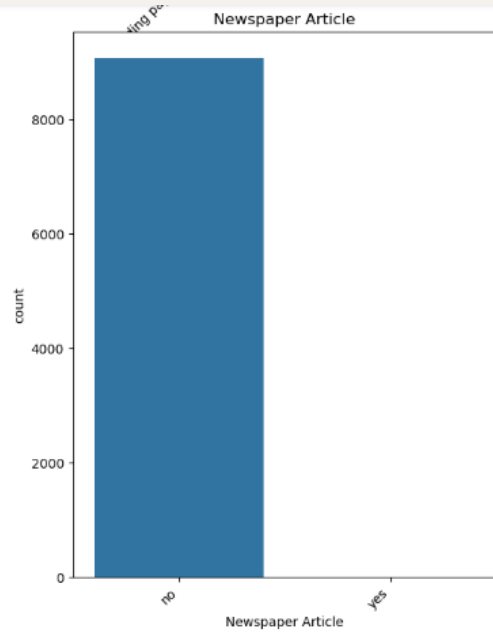
Observations

1. Most of the customers were unemployed.

2. Google was the highest lead source followed by Direct traffic.

3. "Landing page submission' accounted for maximum lead origin.

4. Most of the callers opted not to email or call them about the course.

5. The last activity noted was Email opened and SMS sent.

Bivariate Analysis
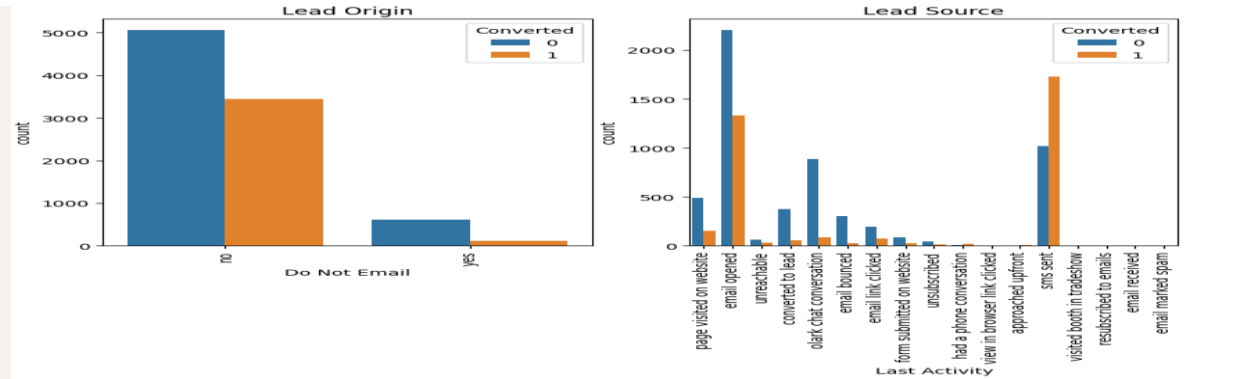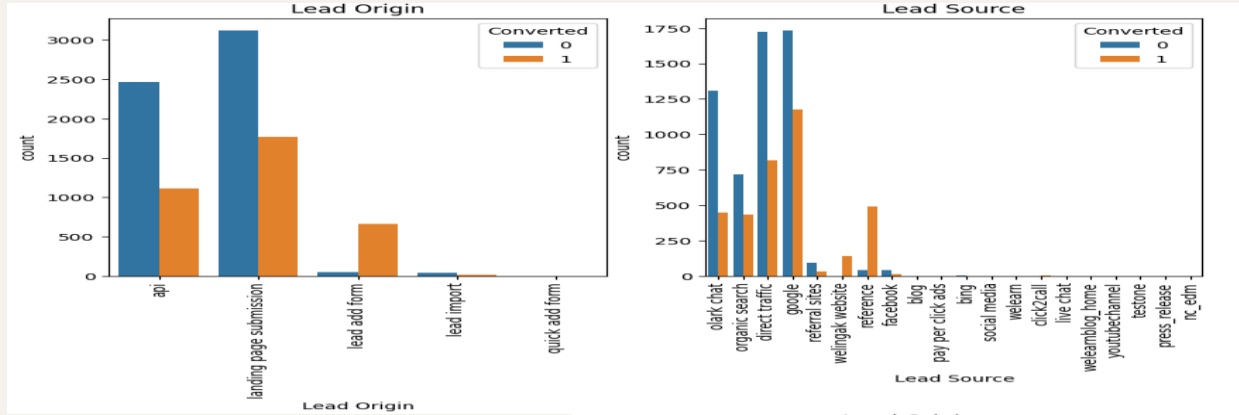
Observations

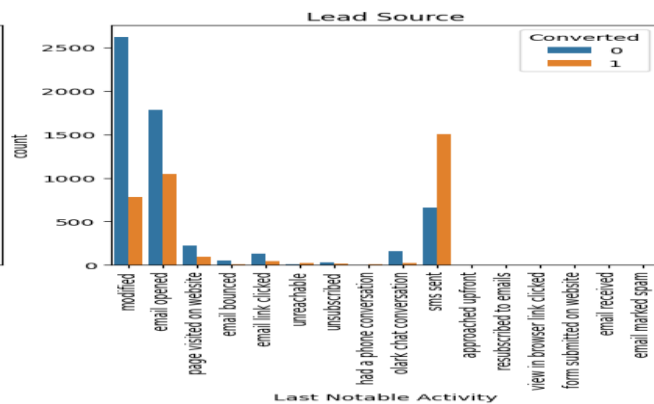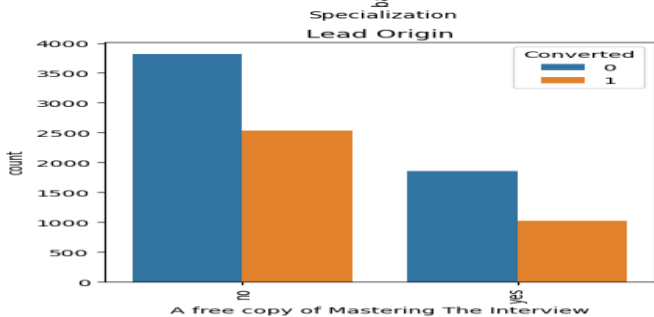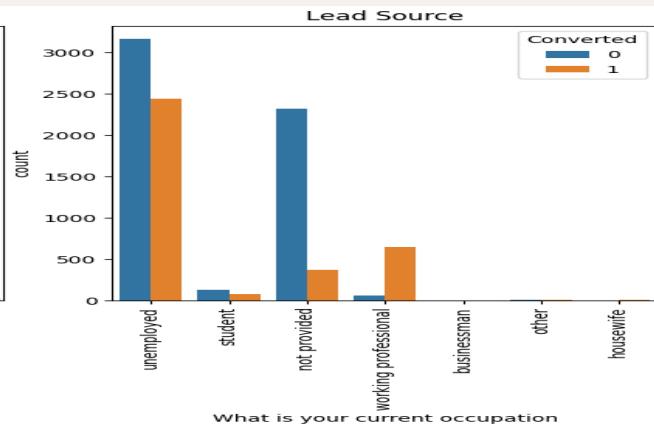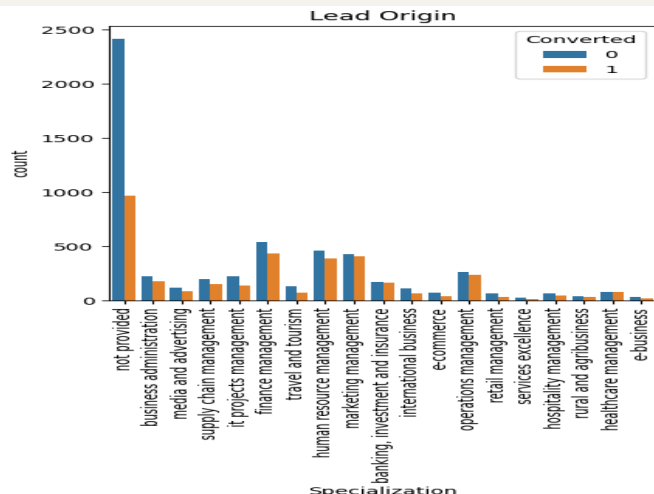1. We find that the conversion rate is high for those whose last notable activity was SMS sent and email opened.
2. We find that the conversion rate is very high in working professionals and unemployed customers.
3. Also, we find that the conversion rate is good for finance management, human resource management, marketing management categories.
4. We find that the conversion rate is very high for 'SMS sent' and 'Email Opened' customers.
5. We find that the conversion is high for 'Lead add form' and 'Landing page submission'.
6. We also find that the conversion rate is very high for 'references' followed by 'Google'.

OBSERVATIONS
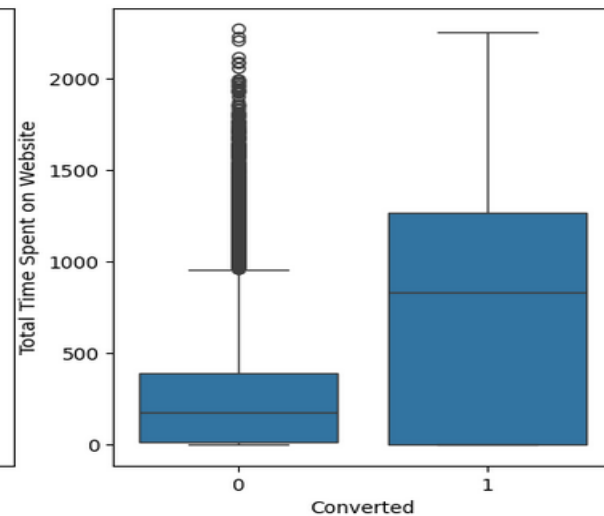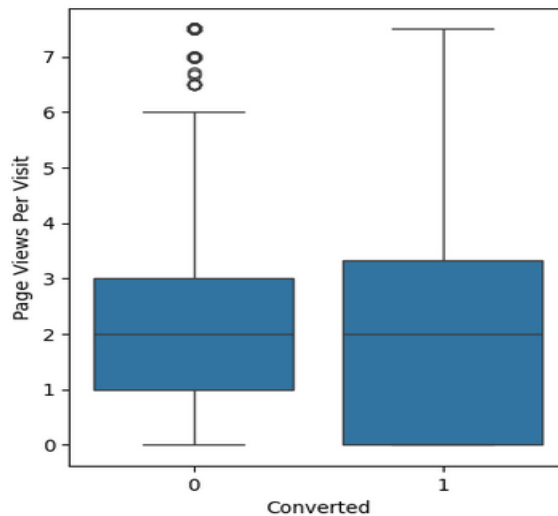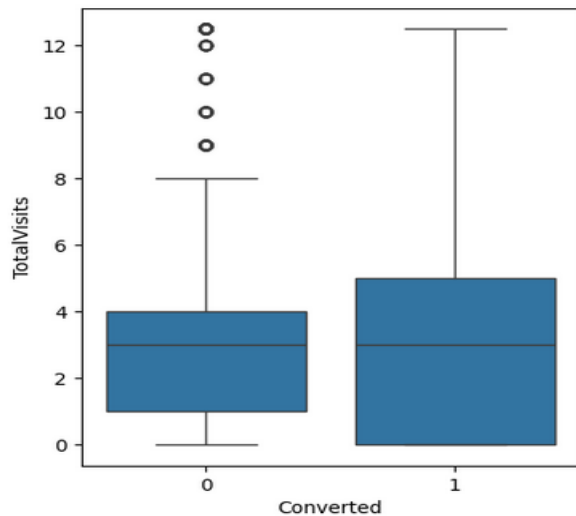
We find that the customers who spend more time on website has better conversion rate.

Target variable – 'Converted'
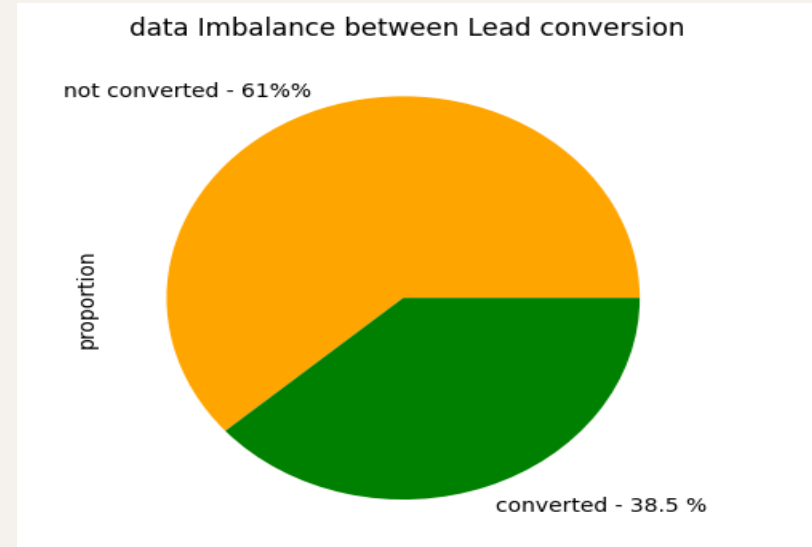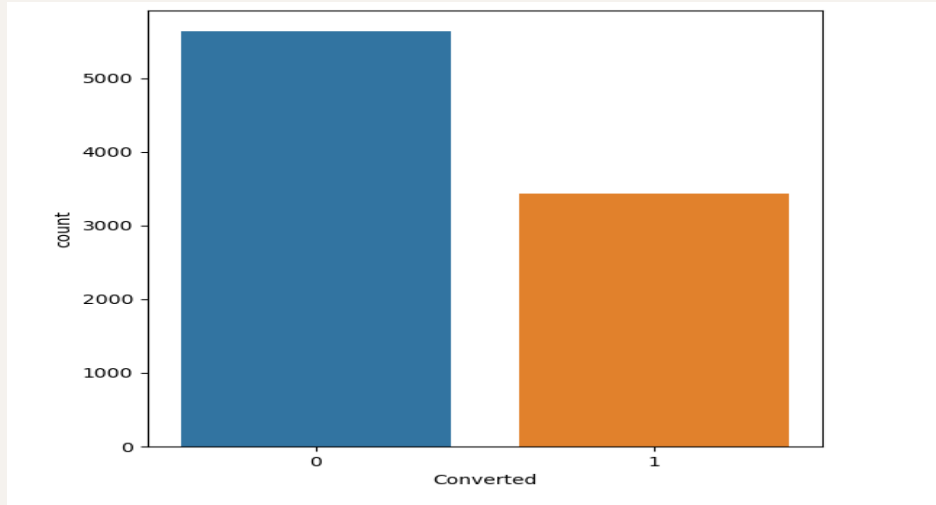
Here the conversion rate is 38.5%,which means 38.5% of the people got converted to leads remaining 61% couldn't get converted to leads.

Evaluation Metrics

Training and testing data values with cutoff as 0.35

Training data

Accuracy – 79.67%

Sensitivity – 79.92%

Specificity – 79.51%

Testing data

Accuracy – 80.05%

Sensitivity – 80.28%

Specificity – 79.93%

Sensitivity in this case indicates how many leads the model identify correctly out of all potential leads which are converting. The company has set a target of 80% conversion rate and as per our model the sensitivity is around 80%,which meets the target.

Conclusion and Recommendation

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
   a. Google
   b. Direct traffic
   c. Organic search
   d. Welingak website
4. When the last activity was:
   a. SMS
   b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.
   Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Actions that can be taken

- Focus on features with positive coefficients for increasing the conversion rate.
- Develop strategies to attract leads from top-performing lead sources.
- Contact working professionals with specific tailor-made messages.
- More budget/spend can be done on Welingak Website in terms of advertising, etc.
- Incentives/discounts for providing reference that convert to lead, encourage providing more references.
- Working professionals to be aggressively targeted as they have high conversion rate and will have better financial situation to pay higher fees too.

Areas of improvement:
- Work on features that have negative coefficients.