Lead Scoring Case Study

Submitted by :

Sandeep Santhosh

Sangeetha Manoharan

Sanchita Deb

Problem statement

- **Background:** X Education, an online educational platform, actively markets its courses to industry professionals across various digital platforms, including search engines and websites.

- **Challenge:** Despite generating a significant volume of leads daily through website interactions and referrals, the company's lead conversion rate remains low at approximately 30%. This means that for every 100 leads generated, only about 30 convert into paying customers.

- **Current Inefficiencies:** The sales team expends substantial effort on leads with low conversion potential, which dilutes their focus and reduces overall efficiency.

- **Opportunity:** There is a critical need to optimize the lead management process by identifying the most promising leads, referred to as 'Hot Leads'. By targeting these leads, X Education aims to increase its conversion rate, allowing the sales team to concentrate their efforts on those most likely to enroll.

- **Strategic Goal:** Enhance the lead conversion process to boost the conversion rate to the target of 80%, thereby maximizing resource utilization and increasing return on investment.

Business Objective

- **Primary Objective:** Develop a predictive model to assign a lead score ranging from 0 to 100 to each lead, based on their likelihood to convert into paying customers. The aim is to enhance the lead conversion rate significantly.

- **Target Lead Conversion Rate:** Achieve a lead conversion rate of 80%, a substantial increase from the current rate of 30%, as specified by the CEO.

- **Hot Leads Identification:** The model should enable X Education to identify 'Hot Leads'—those most likely to convert, allowing the sales team to focus their efforts more effectively and efficiently.

- **Adaptability for Future Needs:** The model must be versatile enough to adapt to future changes in business strategies or market conditions. It should accommodate various scenarios, such as peak times, optimal manpower utilization, and strategic shifts after achieving initial conversion targets.

- **Long-Term Business Impact:** By focusing on the most promising leads, the company expects not only to increase direct sales efficiency but also to optimize overall marketing and operational strategies, thereby maximizing ROI and sustaining business growth.

Problem approach

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis
- Feature Scaling
- Splitting the data into Test and Train dataset
- Building a logistic Regression model and calculate Lead Score
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics

Methodology

Data Sourcing , Cleaning and Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization

Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set

Model Building

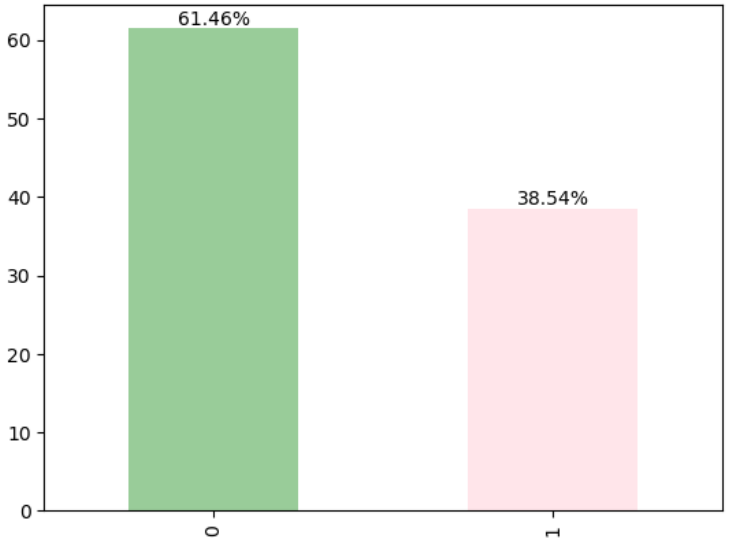
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model

Result

- Determine the lead score and check if target final predictions amounts to 80% conversion rate
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

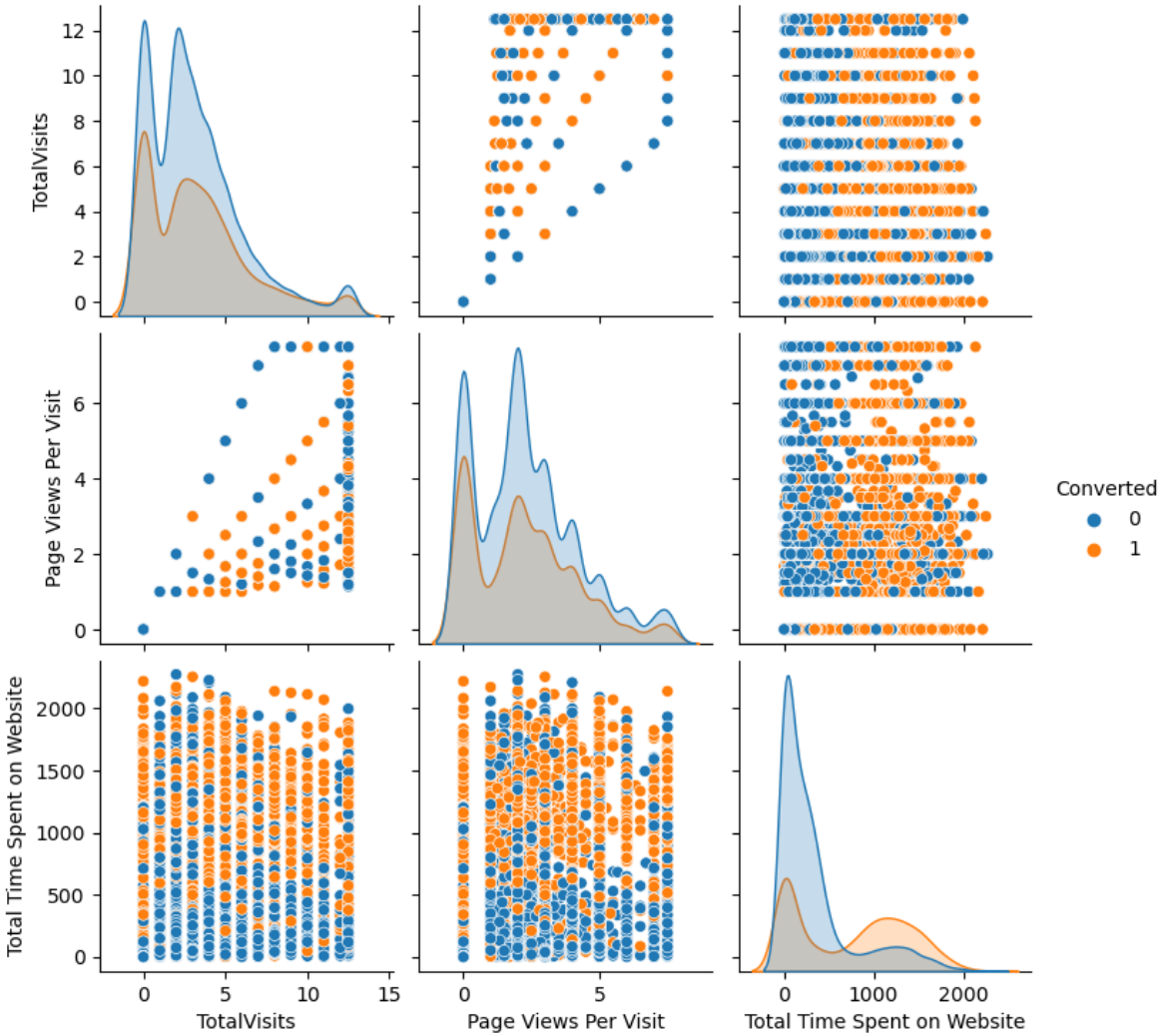
Exploratory Data Analysis

Conversion rate is 38.5% in total

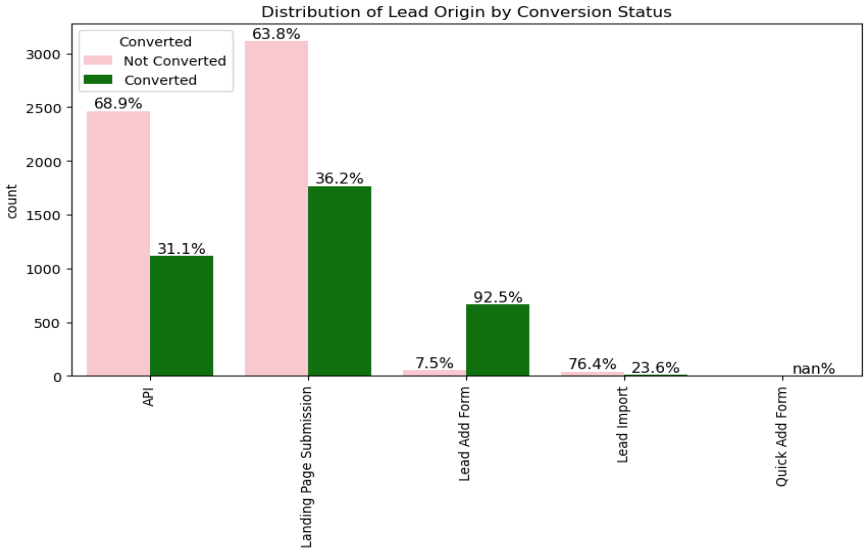


.

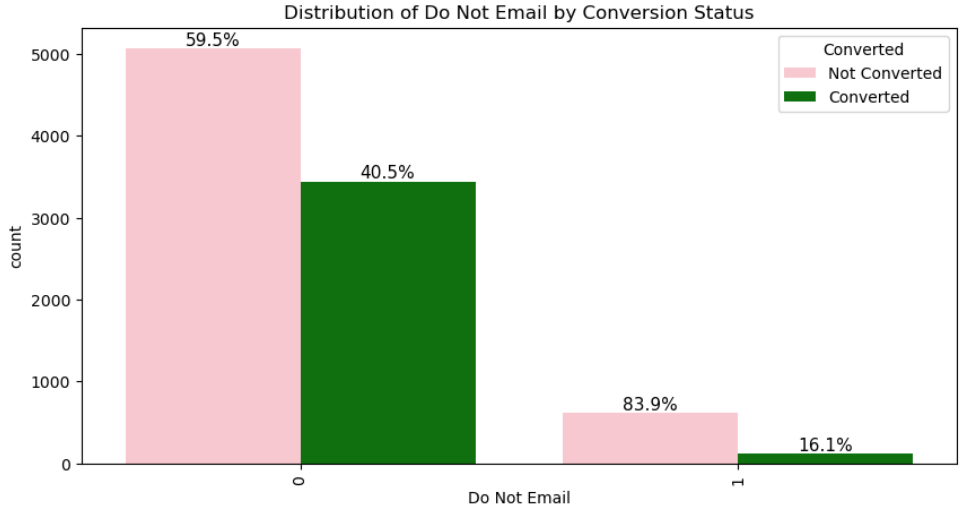
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit



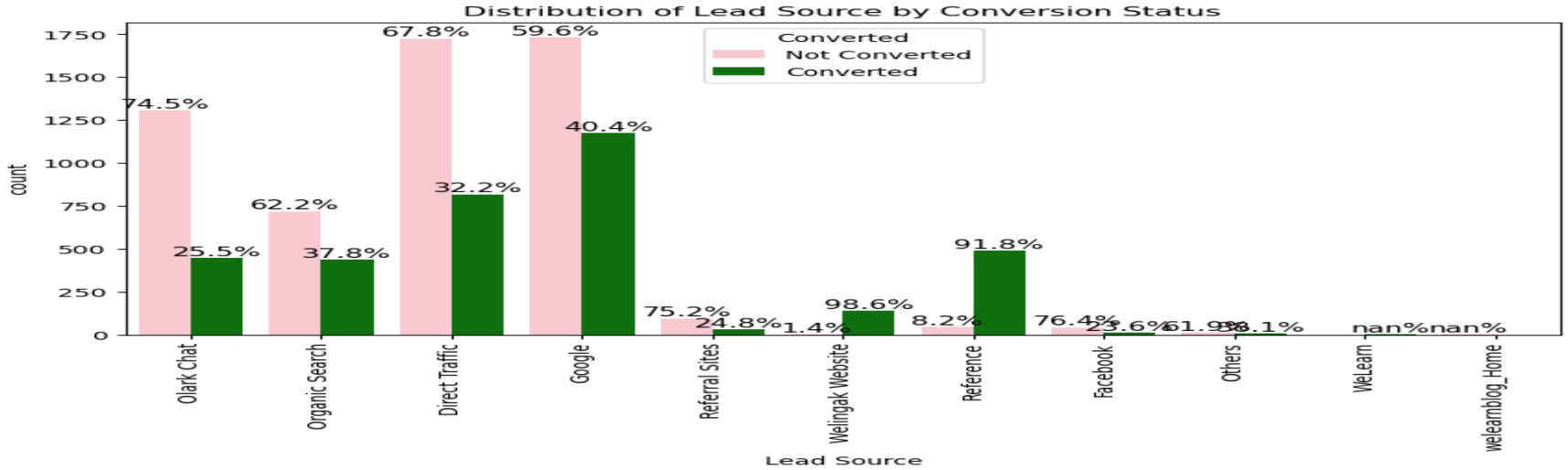
In Lead Origin, maximum conversion happened from Landing Page Submission



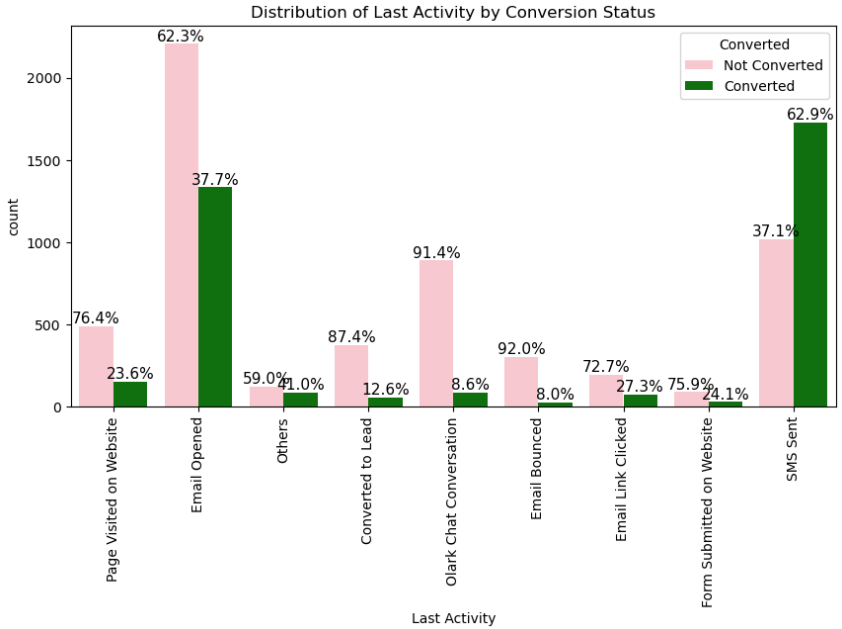
Major conversion has happened from Emails sent



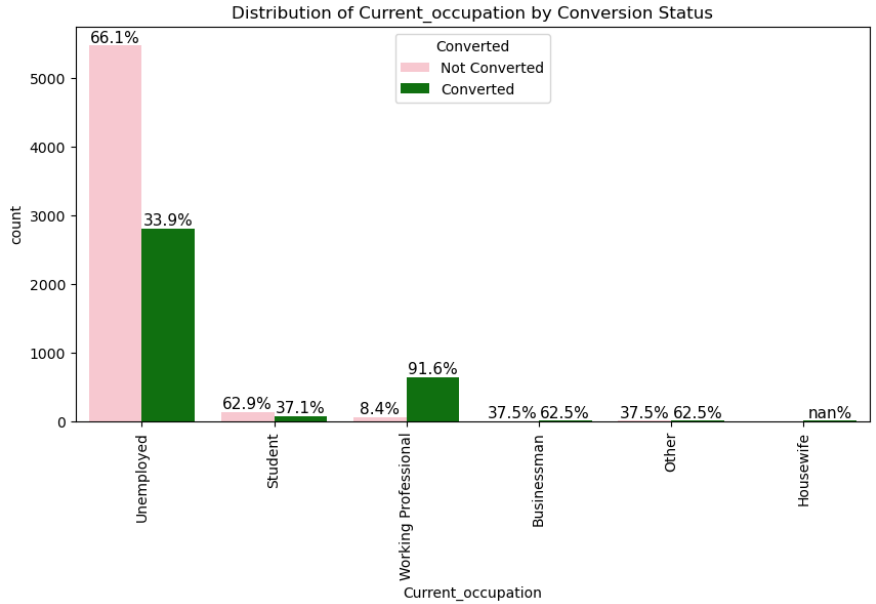
Major conversion in the lead source is from Google

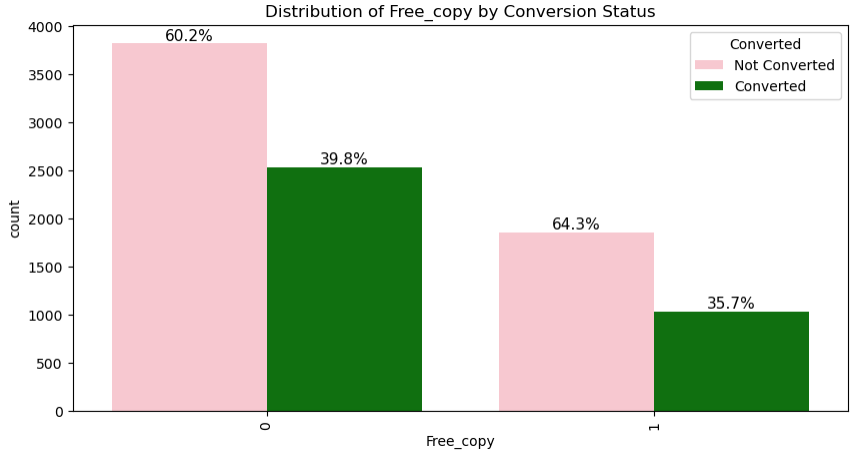


Last Activity value of SMS Sent' had more conversion.



More conversion happened with people who are unemployed

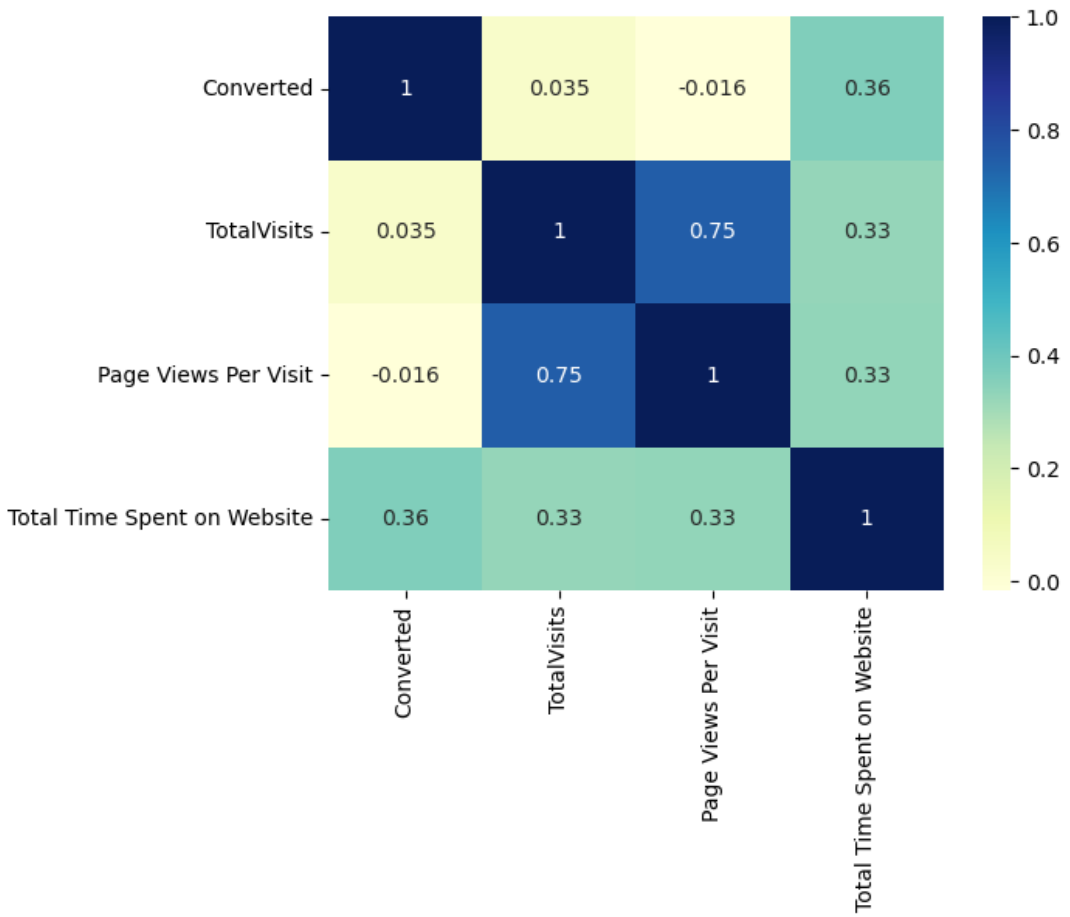




More conversion happened with free copy not distributed

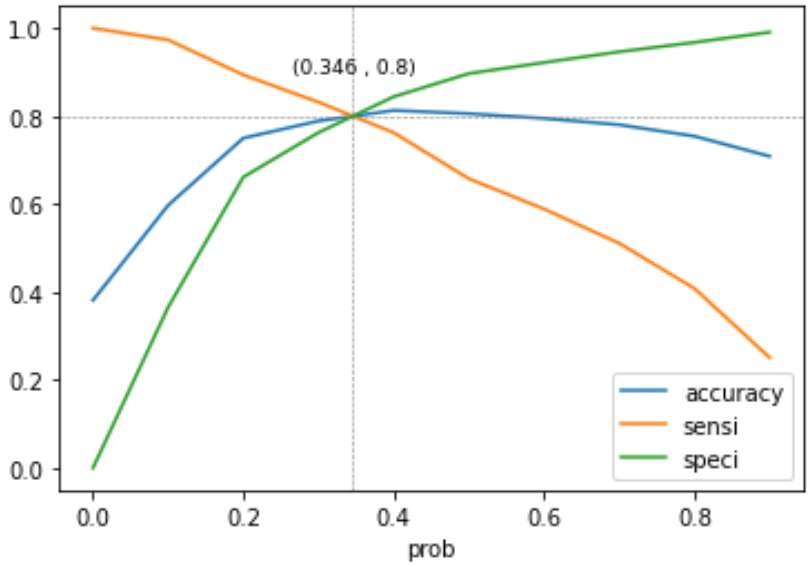
Correlation

Total visits and page views per visit have a strong positive correlation

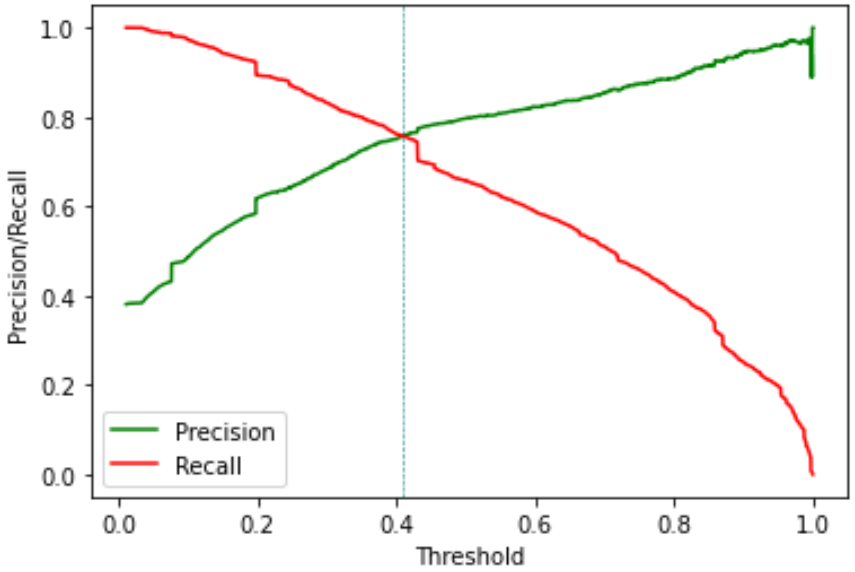


Model Evaluation

The graph depicts an optimal cut off of 0.346 based on Accuracy, Sensitivity and Specificity

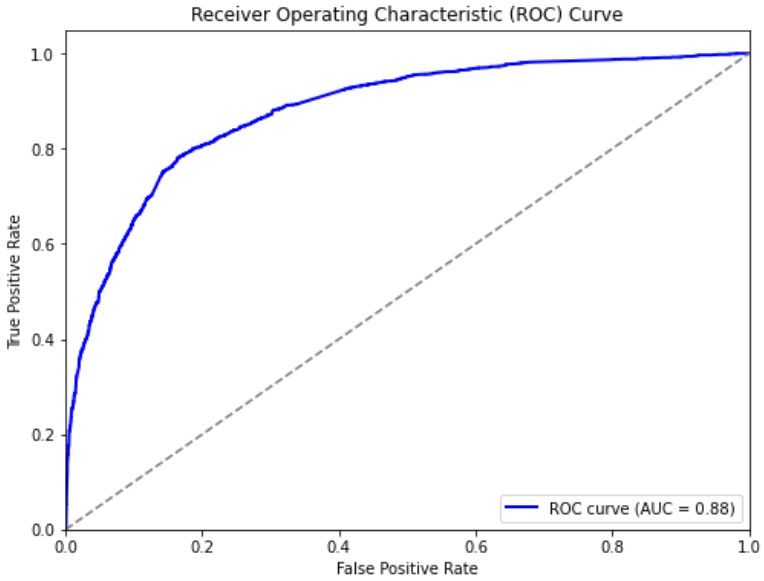


The graph depicts an optimal cut off of 0.42 based on Precision and Recall

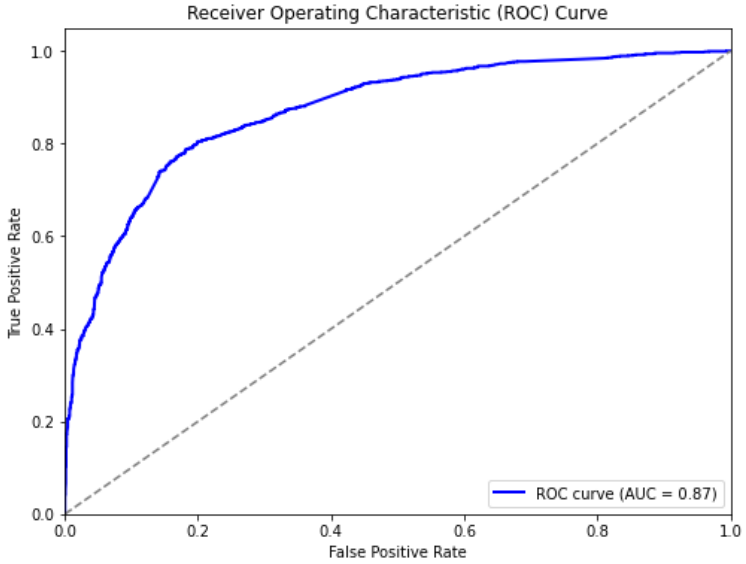


Model Evaluation

ROC for train data



ROC for test data



Confusion matrix – Train Dataset

3589	413
844	1622

Confusion Matrix- Test Dataset

1351	326
222	873

Observations

Train Data Set:
- Accuracy: 81.57%
- Sensitivity: 75.75%
- Specificity: 85.16%

Test Data Set:
- Accuracy: 80.23%
- Sensitivity: 79.73% ≈ 80%
- Specificity: 80.56%

Recommendations to increase our Lead Conversion Rates:

- *Focus on features with positive coefficients for targeted marketing strategies.*
- *Develop strategies to attract high-quality leads from top-performing lead sources.*
- *Engage working professionals with tailored messaging.*
- *Optimize communication channels based on lead engagement impact.*
- *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.*
- *To identify areas of improvement:*
- *Analyze negative coefficients in specialization offerings.*
- *Review landing page submission process for areas of improvement*

Conclusion

- The evaluation matrics are pretty close to each other so it indicates that the model is performing consistently across different evaluation metrics in both test and train dataset.
- The model achieved a sensitivity of 81.57% in the train set and 79.73% in the test set, using a cut-off value of 0.345.
- Sensitivity in this case indicates how many leads the model identify correctly out of all potential leads which are converting
- The CEO of X Education had set a target sensitivity of around 80%.
- The model also achieved an accuracy of 80.23%, which is in line with the study's objectives.
- Hence overall this model seems to be good.