

Lead Scoring Project

LOGISTIC REGRESSION

BY:- Shaikh Sara Naaz

PROBLEM STATEMENT

X Education, an online education company, faces a challenge with its lead conversion process. Despite generating a significant number of leads through website visits, form submissions, and referrals, the current lead conversion rate is only 30%. The company aims to enhance efficiency by identifying and prioritizing 'Hot Leads'—those with the highest potential for conversion. By focusing the sales team's efforts on engaging with these high-potential leads, X Education aims to improve its overall lead conversion rate and optimize the sales process.

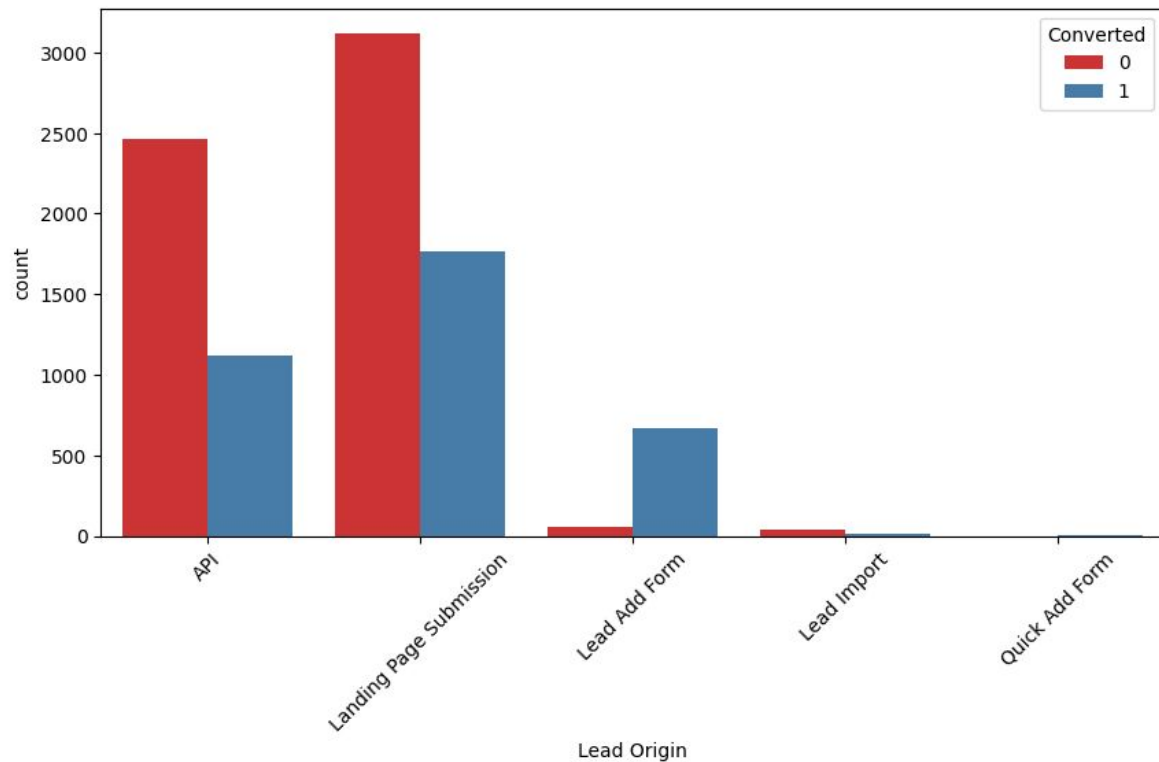
BUSINESS GOAL

The business goal in this problem statement is to enhance the lead conversion process efficiency for X Education. Specifically, the company aims to identify and prioritize 'Hot Leads'—leads with the highest potential for conversion. By doing so, the overarching objective is to increase the lead conversion rate, thereby optimizing the efforts of the sales team and improving the overall effectiveness of the company's sales and marketing strategies.

STEPS

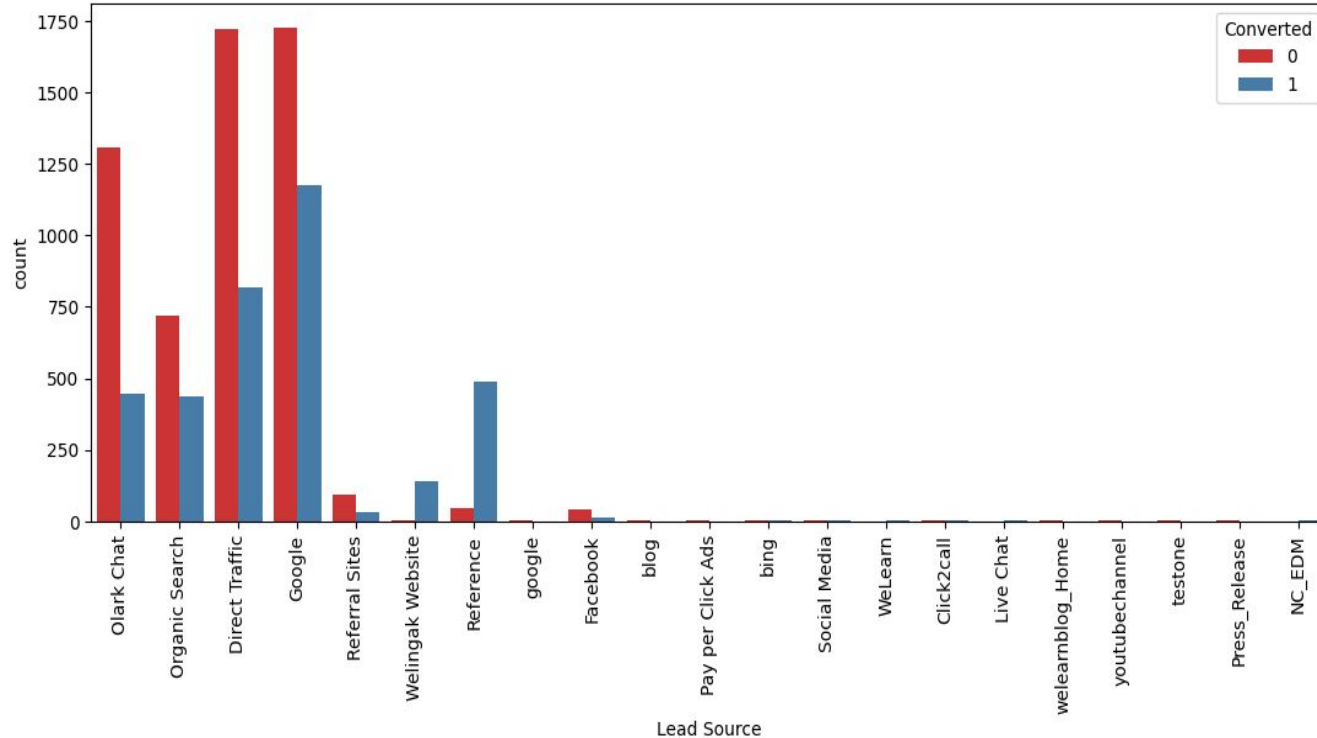
1. Read and understand the data.
2. Clean the data.
3. Prepare the data for Model Building.
4. Model Building.
5. Model Evaluation.
6. Making Predictions on the Test Set.

EXPLORATORY DATA ANALYSIS



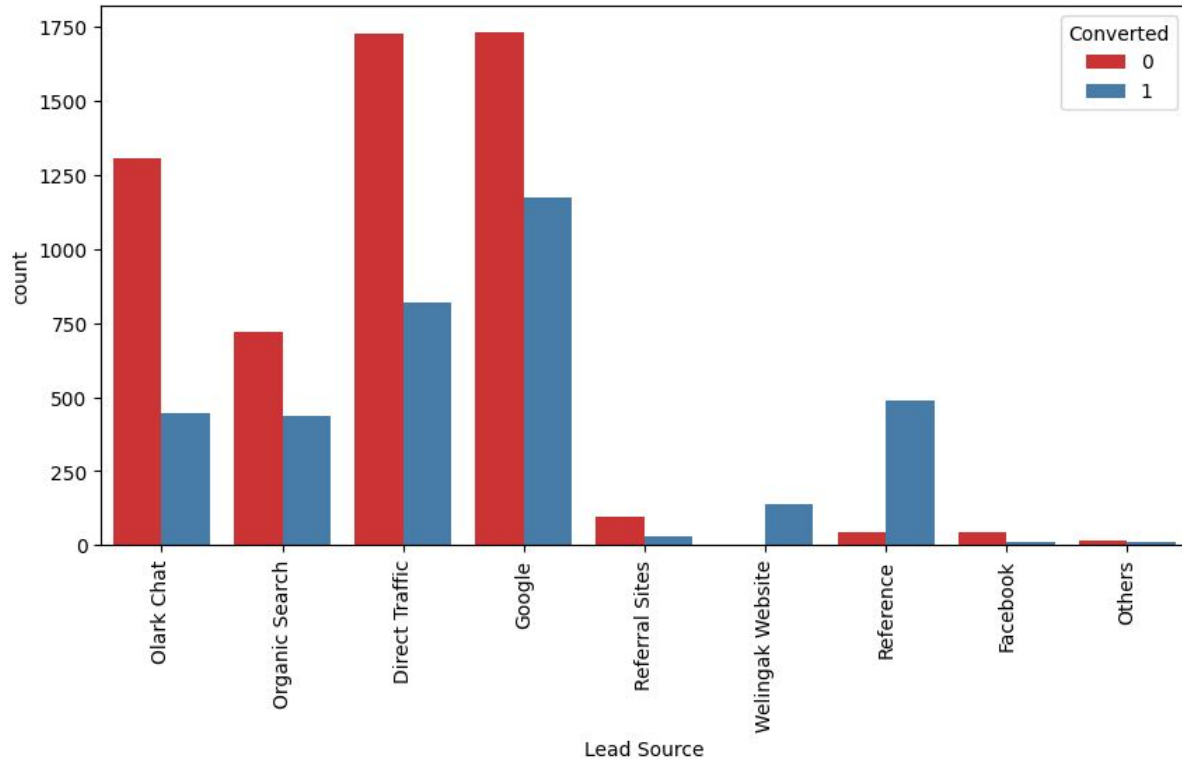
Landing Page Submissions yield a conversion rate of 30-35%, resulting in a substantial number of generated leads.

EXPLORATORY DATA ANALYSIS



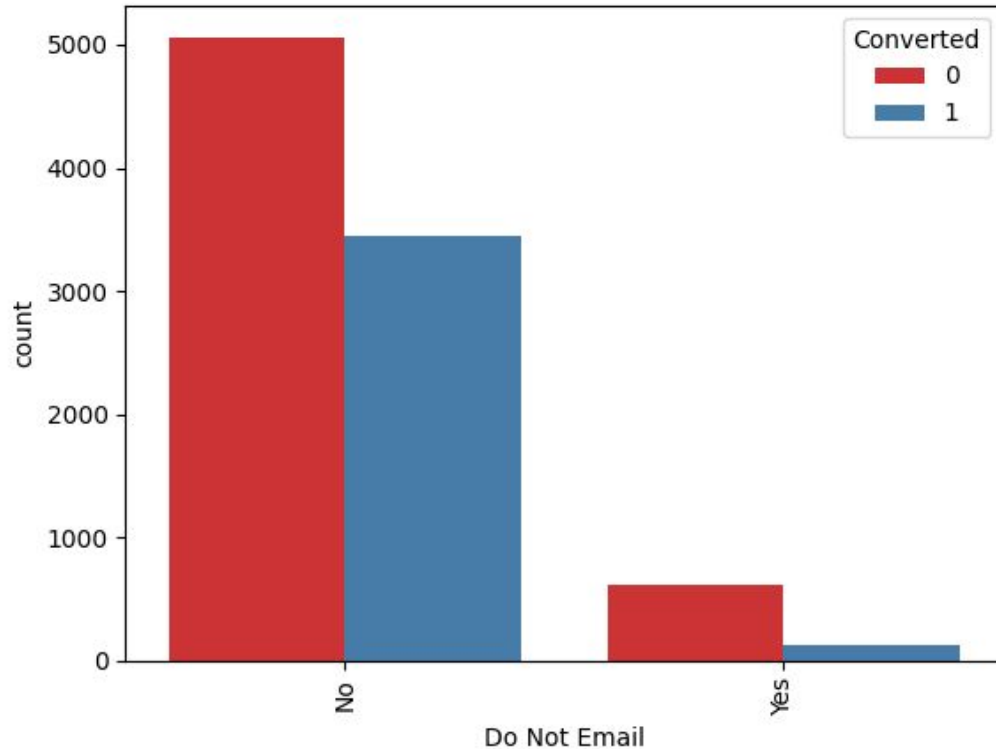
Google searches has the highest conversion compare to the other modes, which references to the higher conversion rate.

EXPLORATORY DATA ANALYSIS



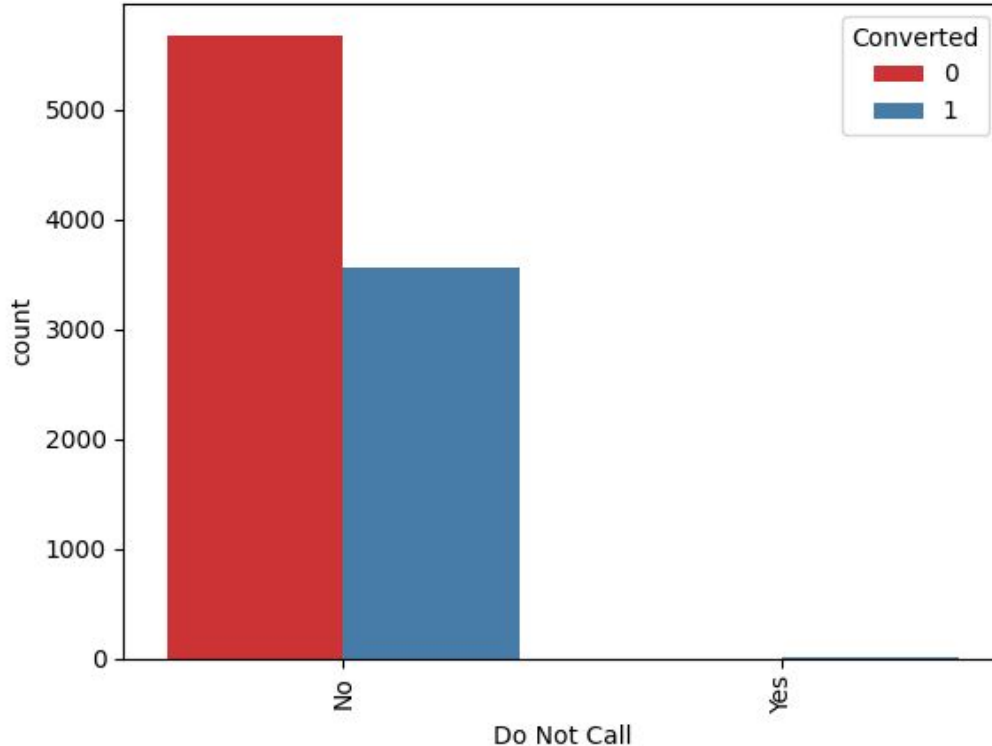
lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingkar website.

EXPLORATORY DATA ANALYSIS



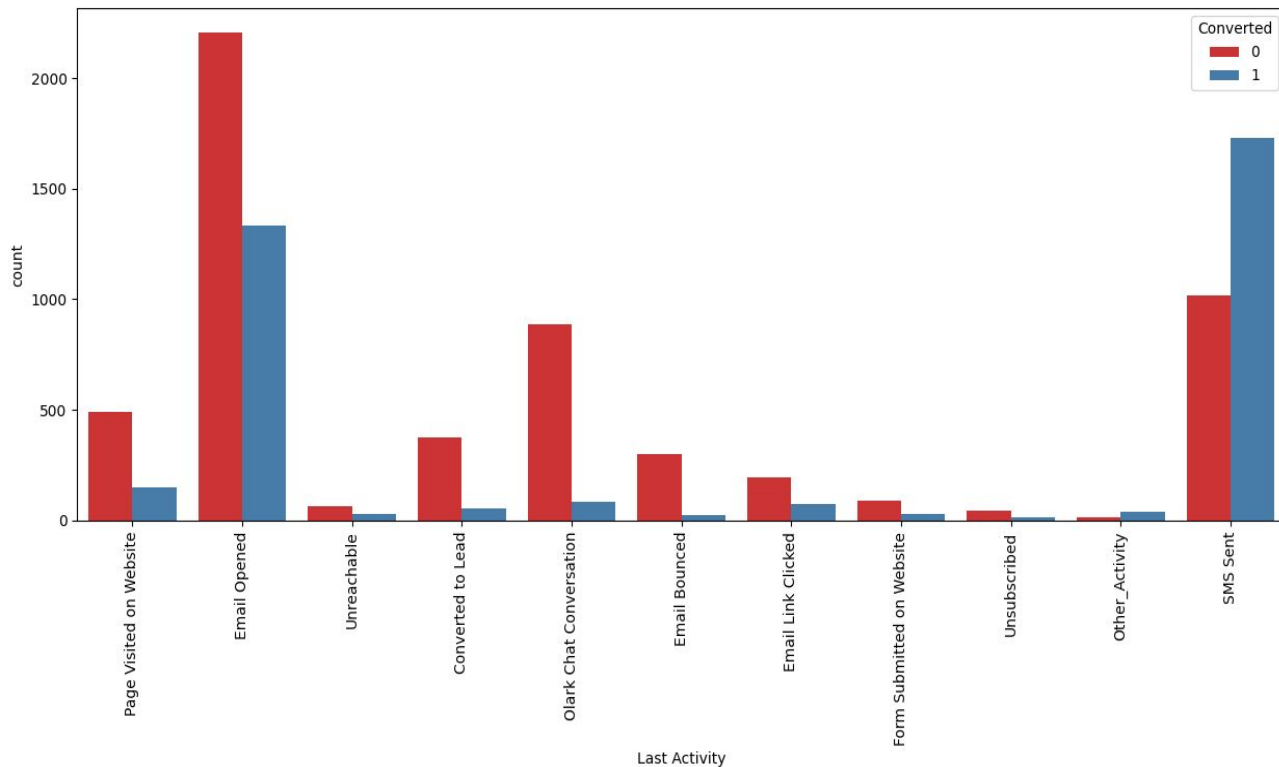
Most of the entries are 'NO' No Inference can be drawn with this parameter.

EXPLORATORY DATA ANALYSIS



Most entries are 'No'. No Inference can be drawn with this parameter.

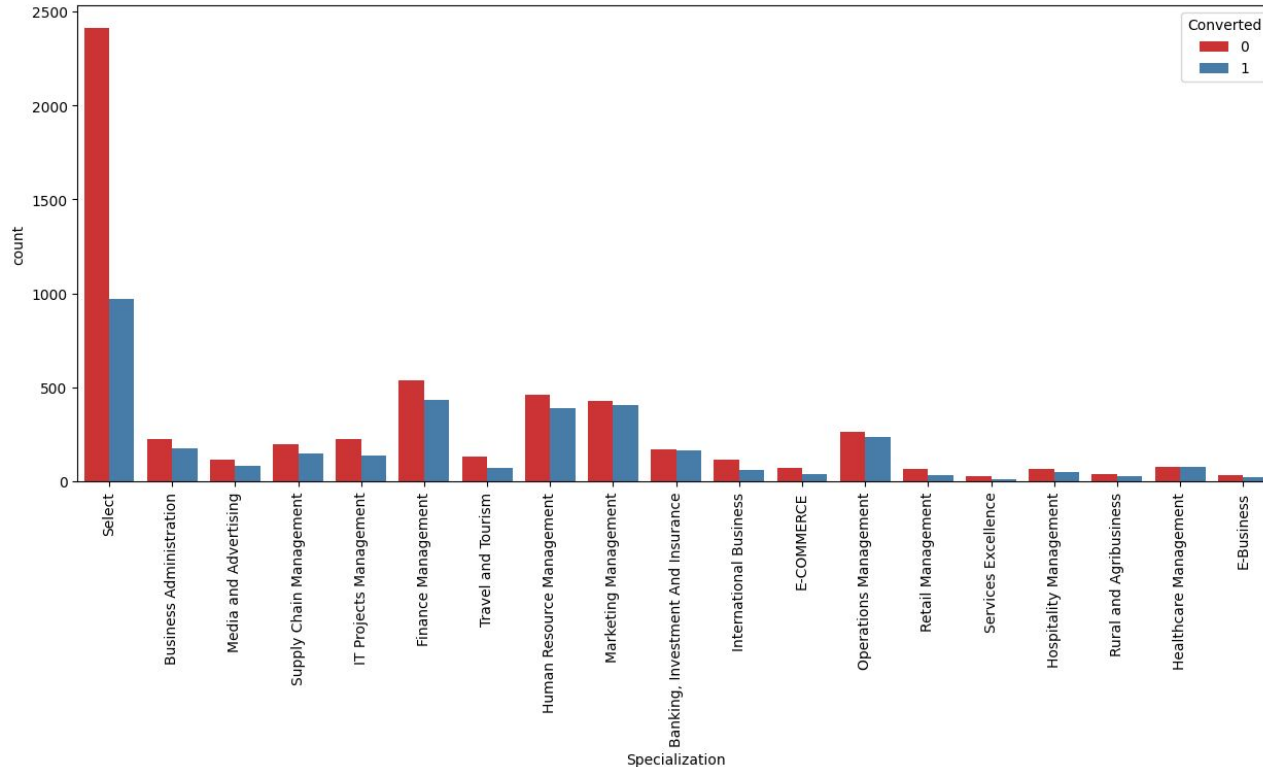
EXPLORATORY DATA ANALYSIS



Most of the lead have their Email opened as their last activity.

Conversion rate for leads with last activity as SMS Sent is almost 60%.

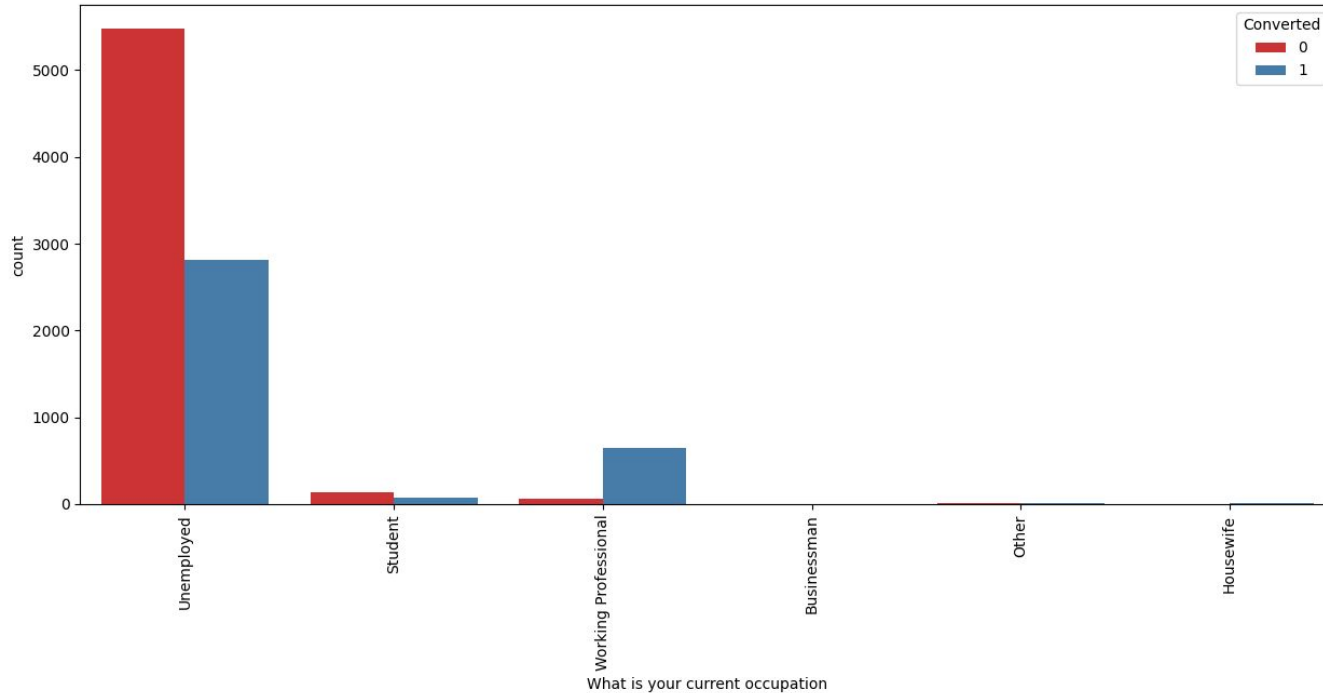
EXPLORATORY DATA ANALYSIS



Focus should be more on the Specialization with high conversion rate.

The industry domain in which the customer worked before. Includes the level 'Select Specialization' which means the customer had not selected this option while filling the form.

EXPLORATORY DATA ANALYSIS

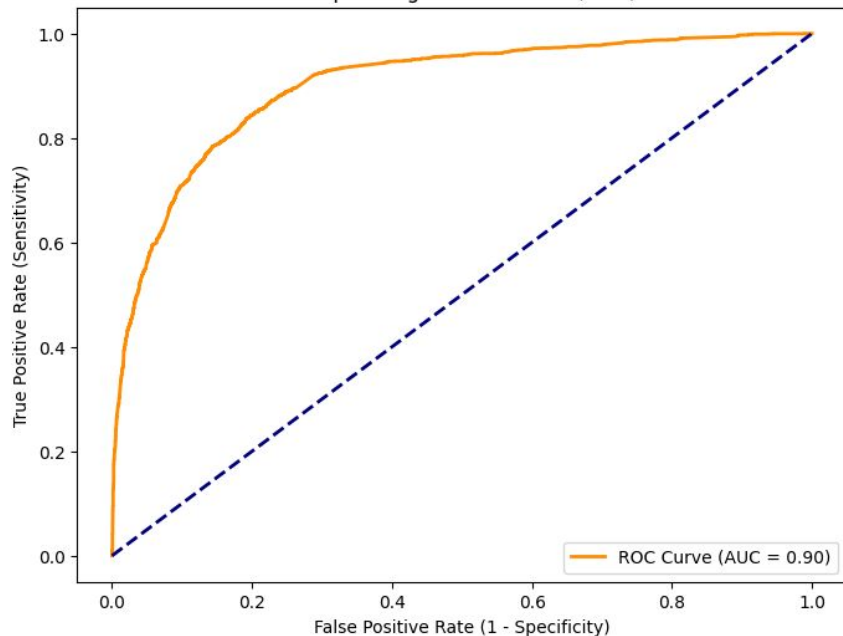


1. Working Professionals going for the course have high chances of joining it.

2. Unemployed leads are the most in numbers but has around 30-35% conversion rate.

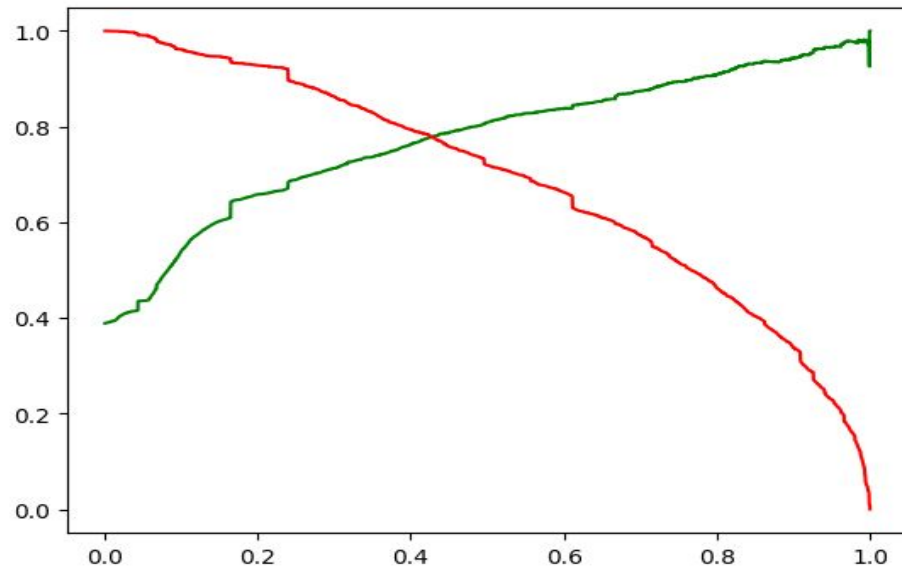
MODEL EVALUATION

Receiver Operating Characteristic (ROC) Curve



ROC [Curve]

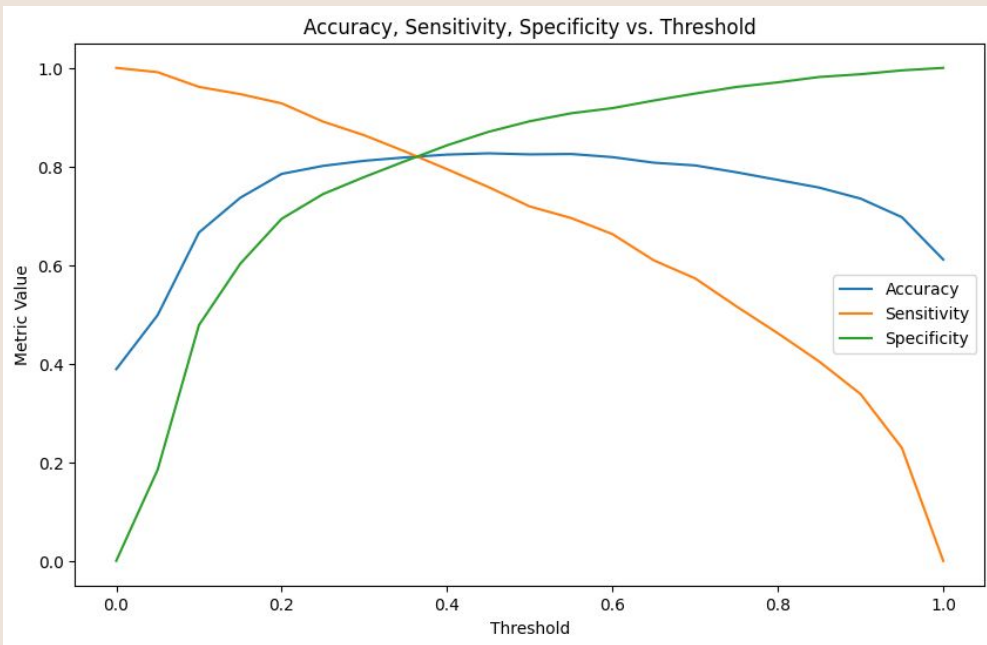
Precision vs Recall tradeoff



Precision : 72.2%

Recall : 74.6%

MODEL EVALUATION



Sensitivity (True Positive Rate): 0.81
Specificity (True Negative Rate): 0.83
Precision: 0.74
Accuracy: 0.82

EDA :-

1. Email opened and SMS send have high impact on lead conversion.
2. Organic search, direct traffic, and google leads and generate more leads.
3. The level 'Select Specialization' which means the customer had not selected this option while filling the form.
4. Working Professionals going for the course have high chances of joining it.
5. Most leads are from mumbai with around 50% conversion rate.

MODEL :-

1. The model shows the 82% accuracy.
2. The threshold has been selected from the Accuracy, Sensitivity, Specificity measures.
3. The model shows 76% sensitivity and 83% specificity.
4. The model showed it could accurately identify hot leads.
5. The classification report at the end gives a quick overview confirming that the model is pretty good at determining if a lead is hot or not.