

LEAD SCORING CASE STUDY

LOGISTIC REGRESSION

Contents

- **Problem statement**
- **Business Objective**
- **Problem approach**
- **EDA**
- **Model Evaluation**
- **Observations**
- **Recommendations**

Problem Statement

- X Education, an online course provider for industry professionals, experiences a steady influx of visitors to their website daily. These visitors often explore available courses and engage with the website by filling out forms, thus becoming leads.
- Once acquired, the sales team initiates outreach efforts through calls, emails, and other channels. However, while some leads convert, a majority do not. The typical lead conversion rate hovers around 30% at X Education. This means that out of 100 acquired leads, only approximately 30 result in conversions.
- To streamline this process and boost efficiency, X Education aims to identify the most promising leads, commonly referred to as Hot Leads.
- By accurately pinpointing these high-potential leads, the company anticipates a significant increase in the lead conversion rate. This strategic shift allows the sales team to concentrate efforts on engaging with these promising prospects, rather than casting a wide net and contacting every lead indiscriminately.

Business Objective

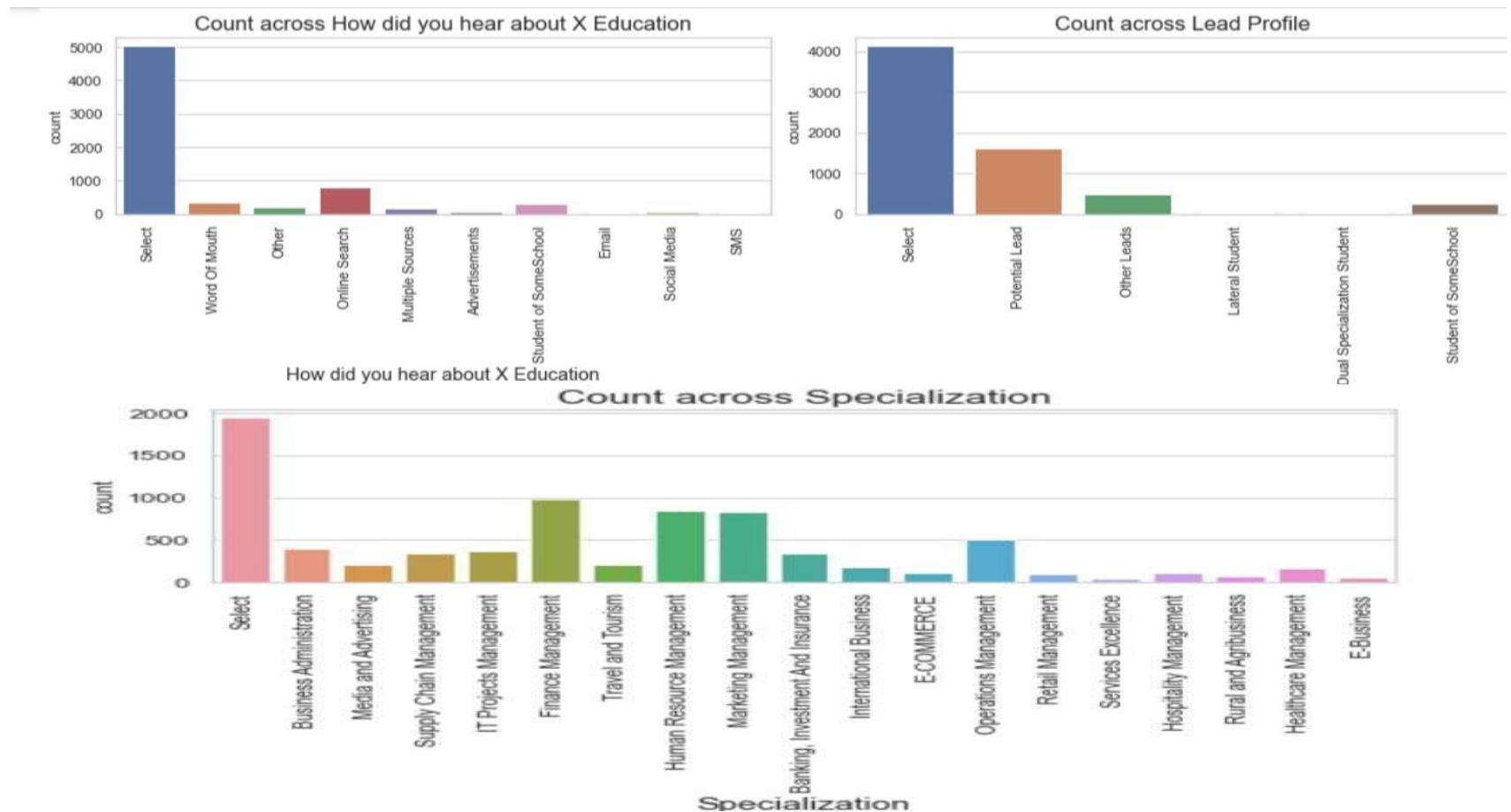
- Lead X has tasked us with creating a model to assign a score ranging from 0 to 100 to each lead.
- This initiative aims to aid in the identification of hot leads and enhance their conversion rate. The CEO's goal is to achieve a lead conversion rate of 80%.
- Additionally, Lead X wants the model to address future constraints such as peak time actions, maximizing manpower utilization, and post-target achievement approaches.

Problem Approach

- **Importing the data and inspecting the data frame**
- **Data preparation**
- **EDA**
- **Dummy variable creation**
- **Test-Train split**
- **Feature scaling**
- **Correlations**
- **Model Building (RFE Rsquared VIF and p- values)**
- **Model Evaluation**
- **Making predictions on test set**

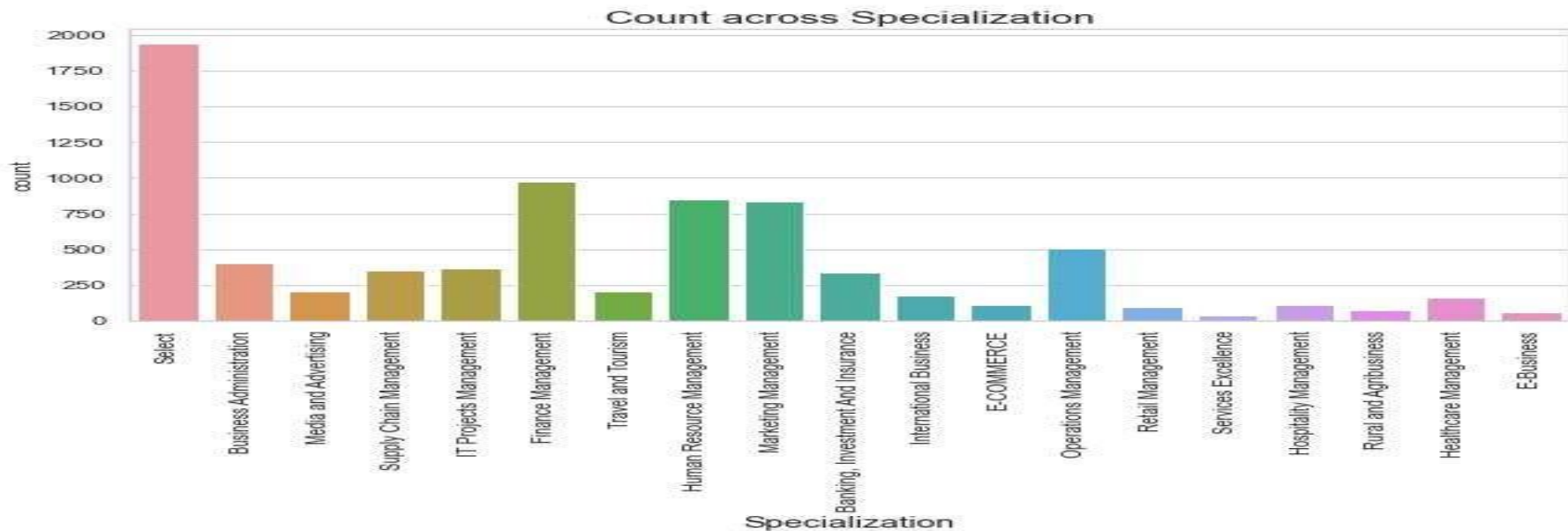
EDA – Data Cleaning

There are a few columns in which there is a level called 'Select' which is taking care



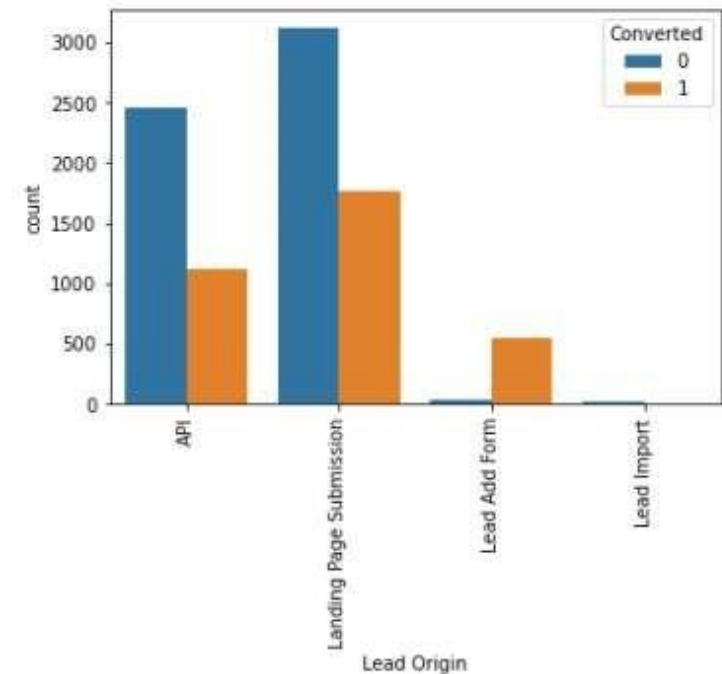
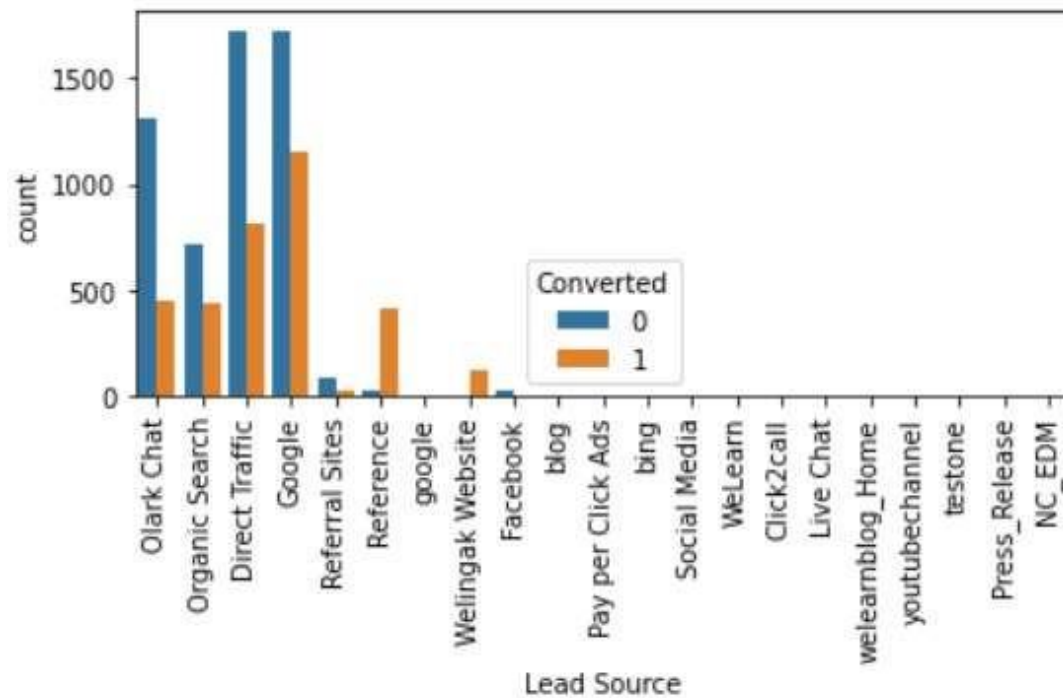
Specialization

Leads from HR, Finance & Marketing management specializations are high probability to convert



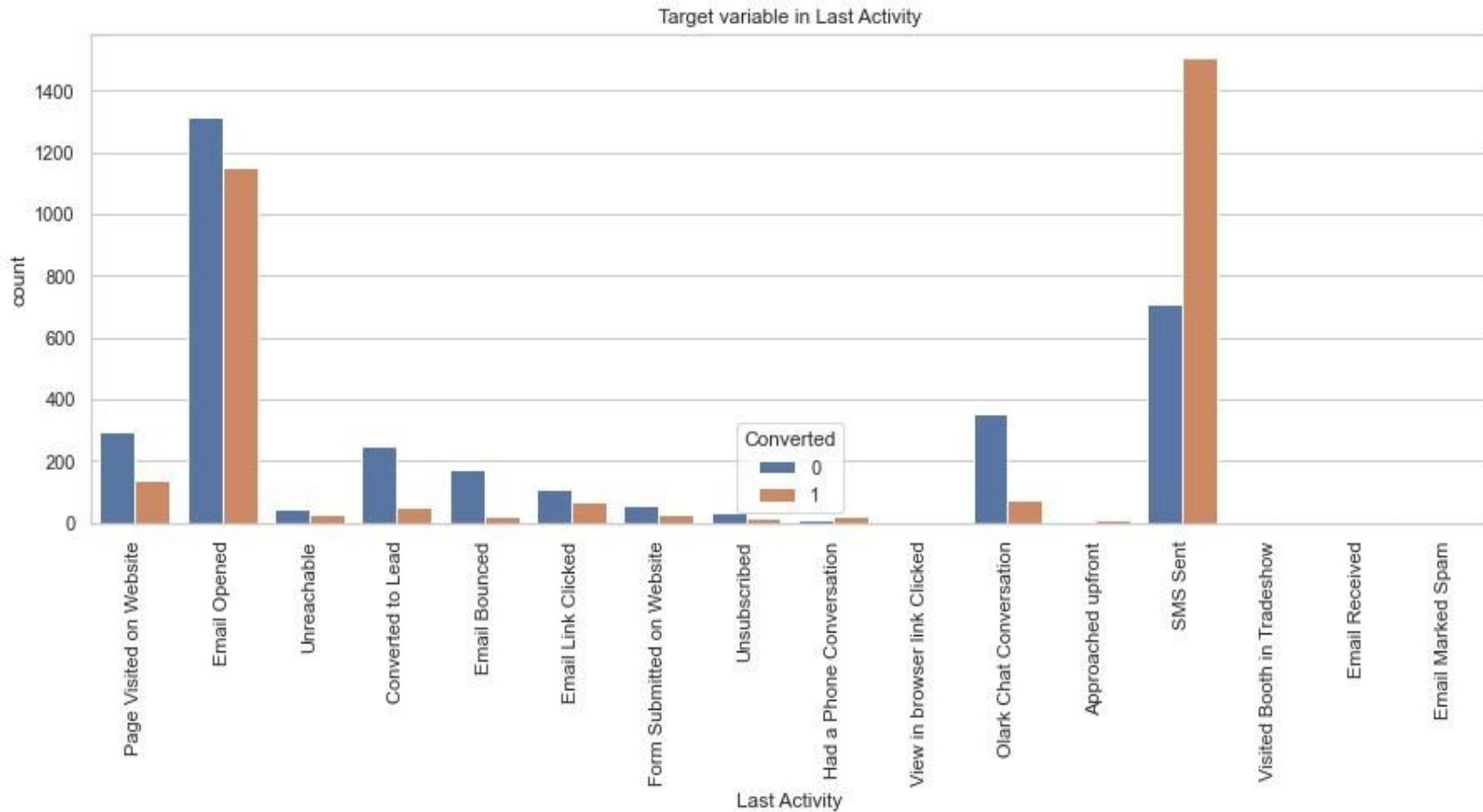
Lead Source & Lead origin

In lead source the leads through google Whereas in Lead origin most number of leads are landing on submission

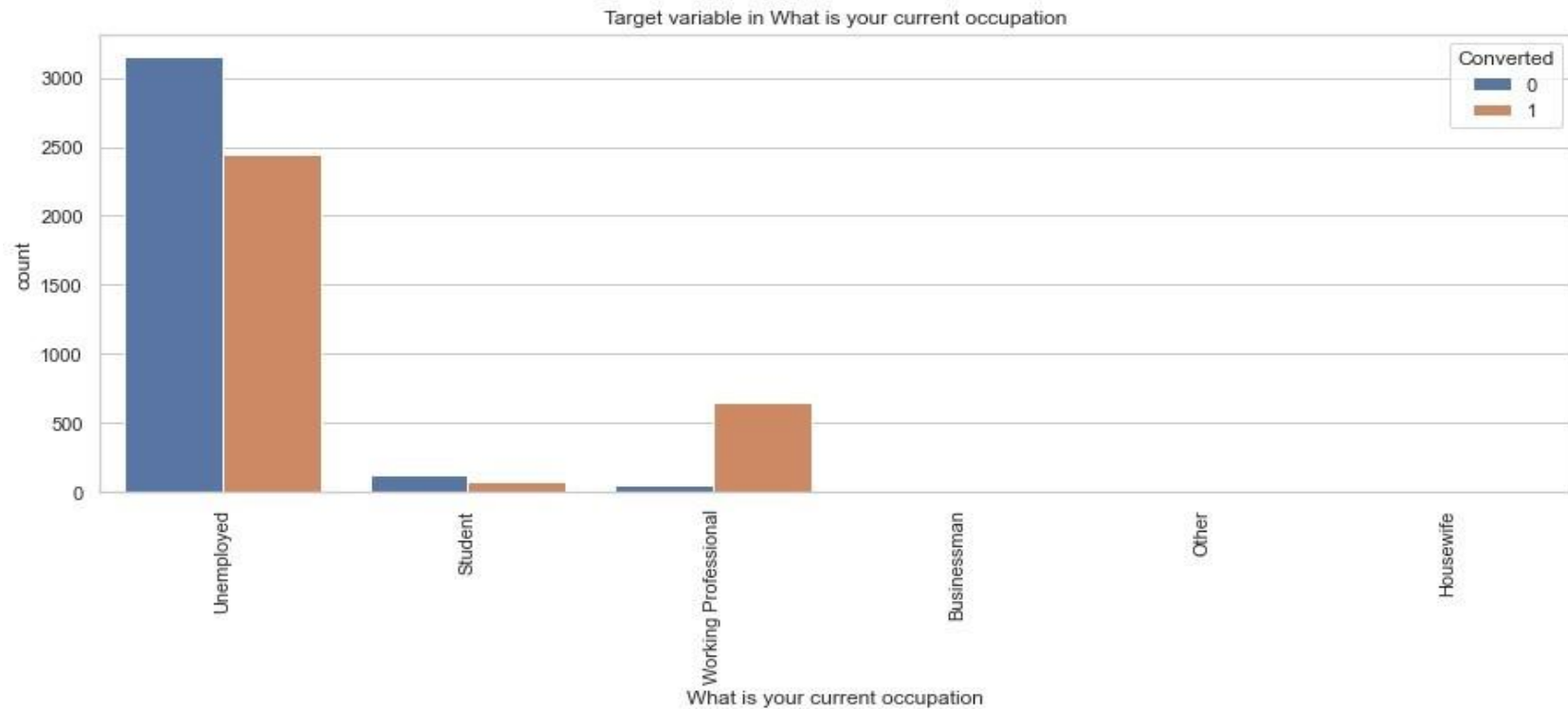


Last lead Activity

Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.



Last What is Your Occupation

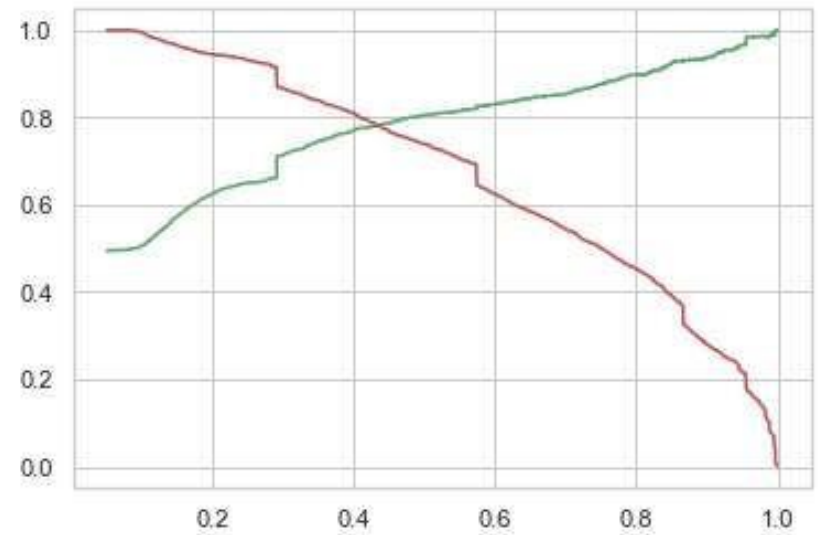
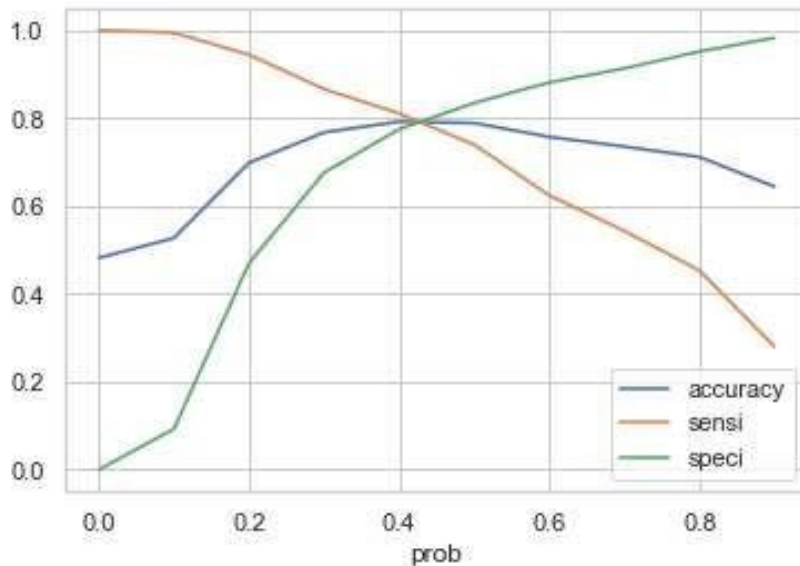


Model Evaluation

ROC Curve

0.42 is the tradeoff between Precision and Recall -

Thus we can safely choose to consider any Prospect Lead with Conversion **Probability higher than 42 %** to be a hot Lead



Observations

Train Data:

Accuracy : 80%

Sensitivity : 80%

Specificity : 81%

Test Data:

Accuracy : 81%

Sensitivity : 81%

Specificity : 80%

Final Features list:

- Lead Source_Olark Chat
- Specialization_Others
- Last Activity_SMS Sent
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- What is your current occupation_Working Professionals
- Do Not Email

Recommendations

- We observe that the conversion rate is 30-35% (close to average) for leads originating from API and Landing page submissions. However, it is significantly lower for leads from the Lead Add form and Lead import. Therefore, we should focus more on the leads originating from API and Landing page submissions.
- The maximum number of leads are generated through Google or direct traffic. The highest conversion ratio is achieved through referrals and the Welingak website.
- Leads spending more time on the website are more likely to convert.
- The prevailing final engagement tends to be email openings, while the most frequent action is sending SMS. The majority are currently without employment, with the most conversions occurring with individuals in professional roles.