Lead Scoring Case Study using Logistic Regression

Contents:

- Problem Statement
- Problem Approach
- Business objective
- Observations after EDA
- Model Evaluation
- Model Statistics
- Conclusion

Problem Statement:

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around 30%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Business Objective:

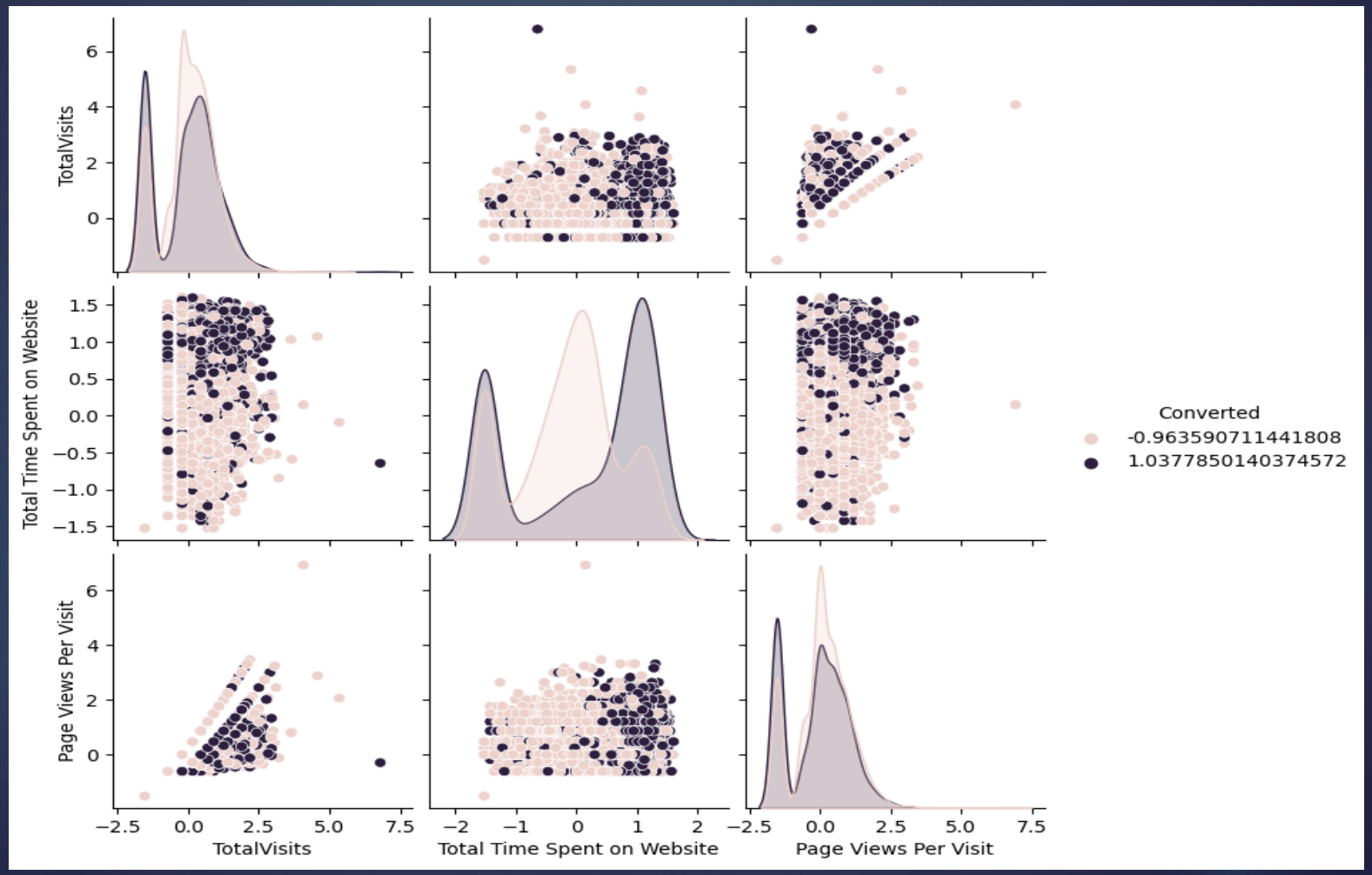
- Lead X wants us to build a model to give every lead a lead score between 0 -100, So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches.

Problem Approach:

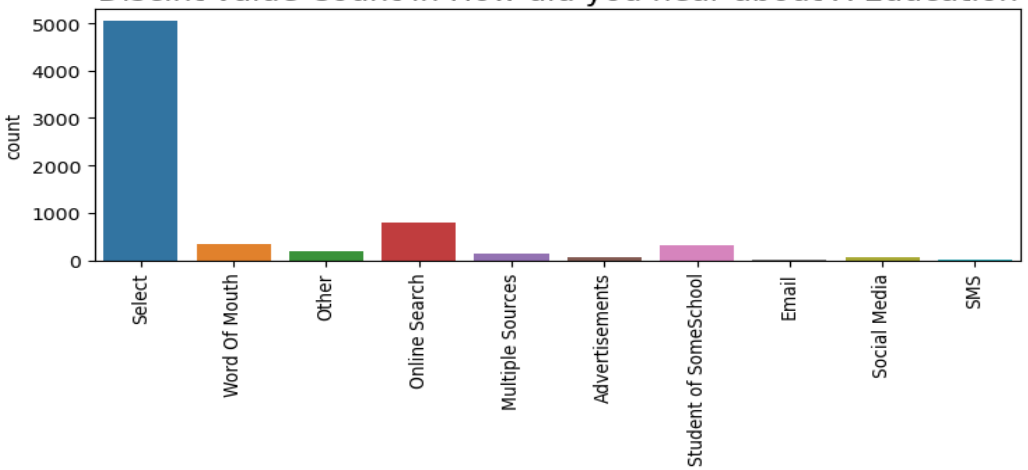
- Importing the Dataset
- Data preparation and Cleaning
- Exploratory data analysis
- Dummy variable creation
- Train test split
- Feature Scaling
- Model building
- Model Evaluations
- Testing and making predictions against test set

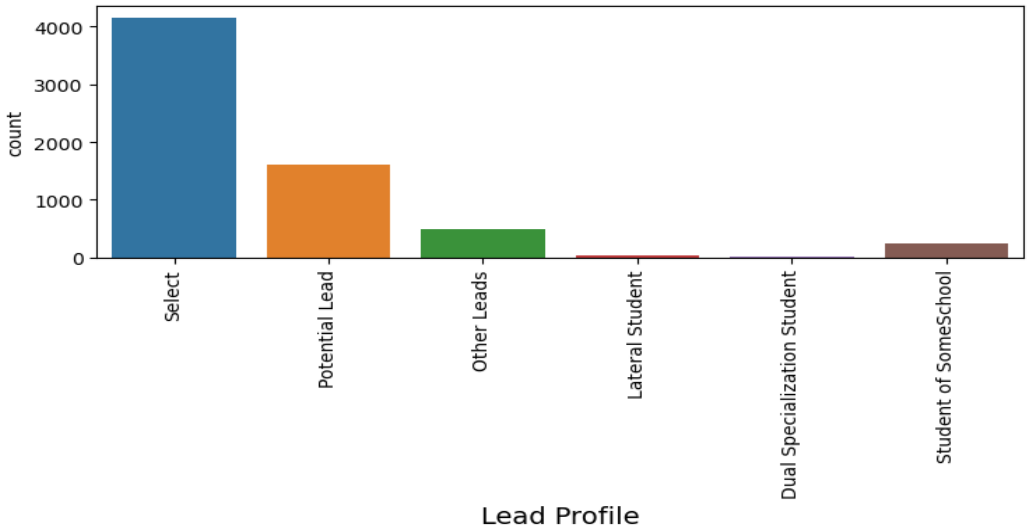
Observations after EDA:

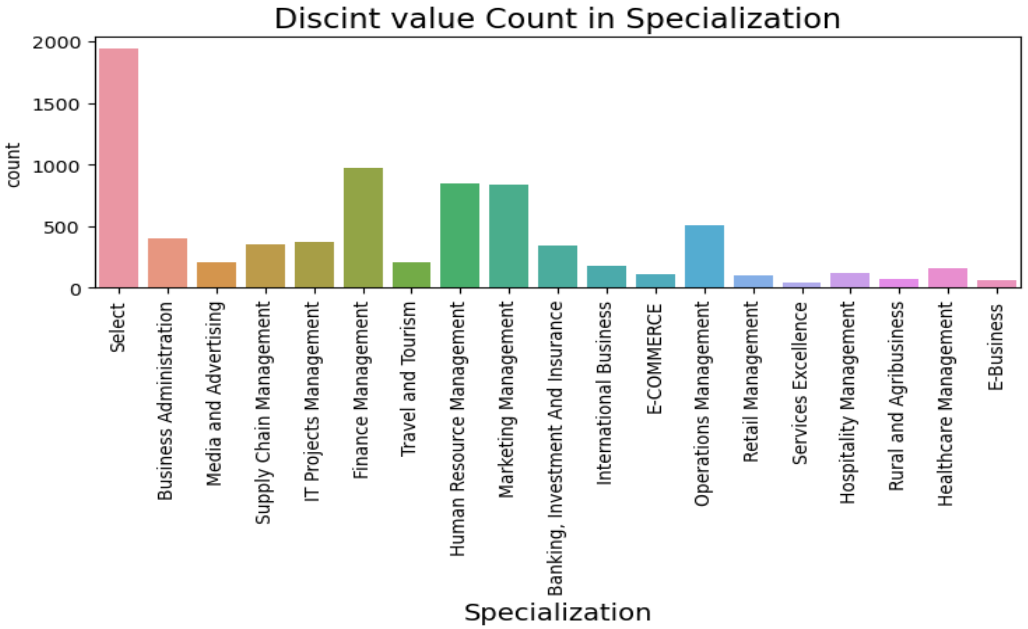
- Following seems to be the transformed conversion plots for the numerical variables



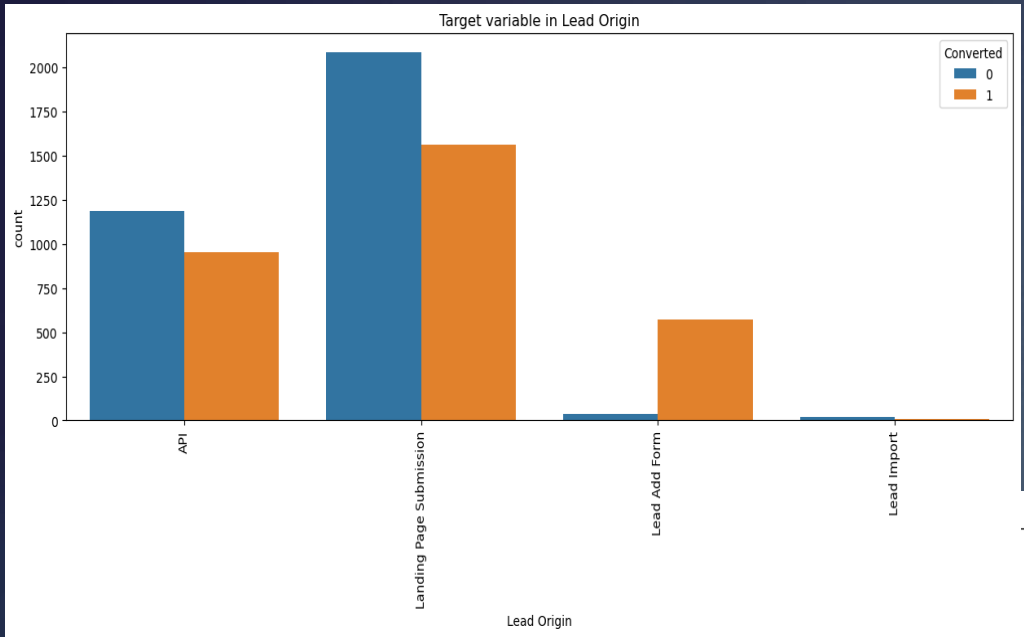
- Most leads prefer not to share their personal information like lead profiles, their specializations and source through which they heard about the course



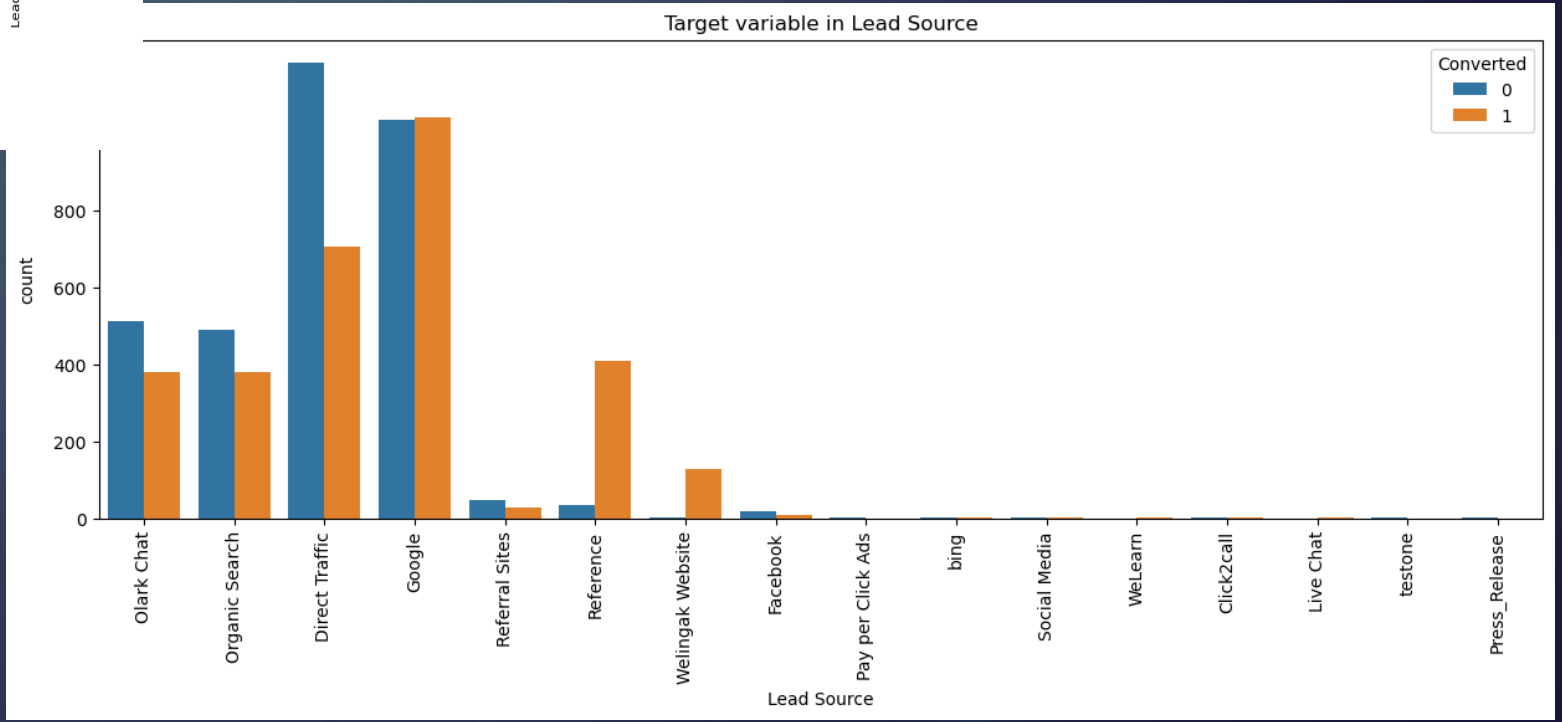




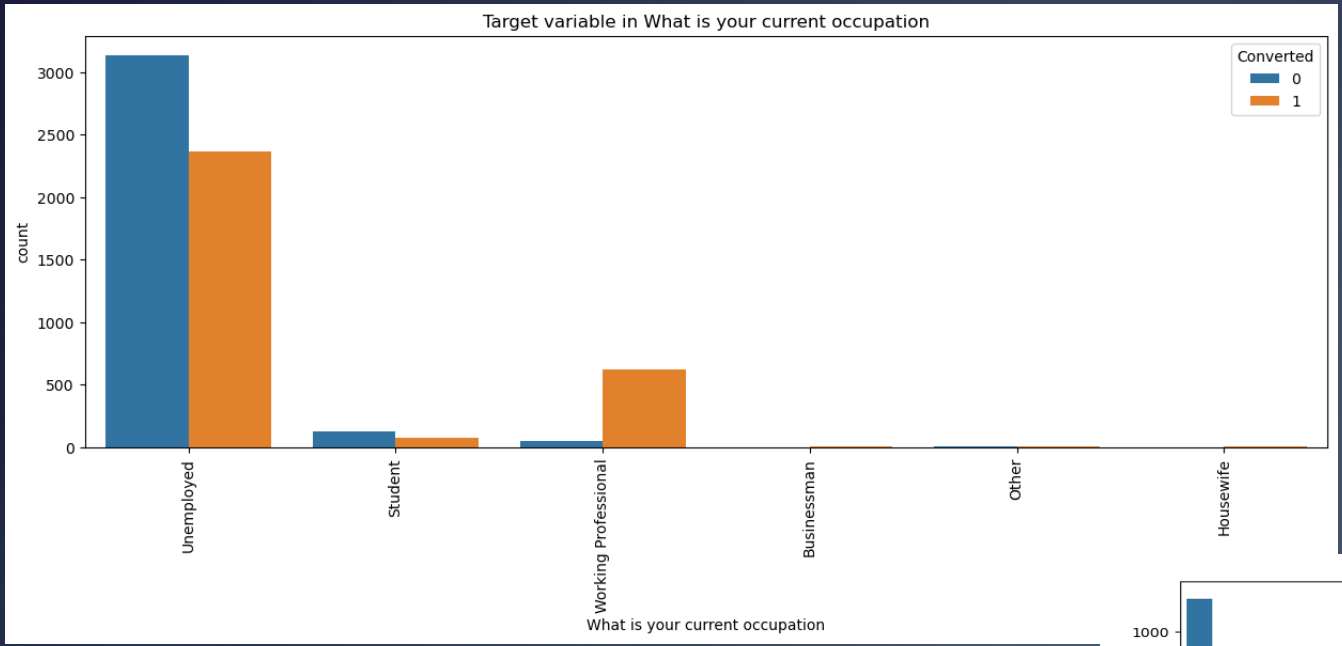
- Following seems to be the ratio of lead conversions for lead origin and lead source



There seems to be a high conversion rate for leads that Add Form and had sources as Reference, Welingak, Google and Organic search

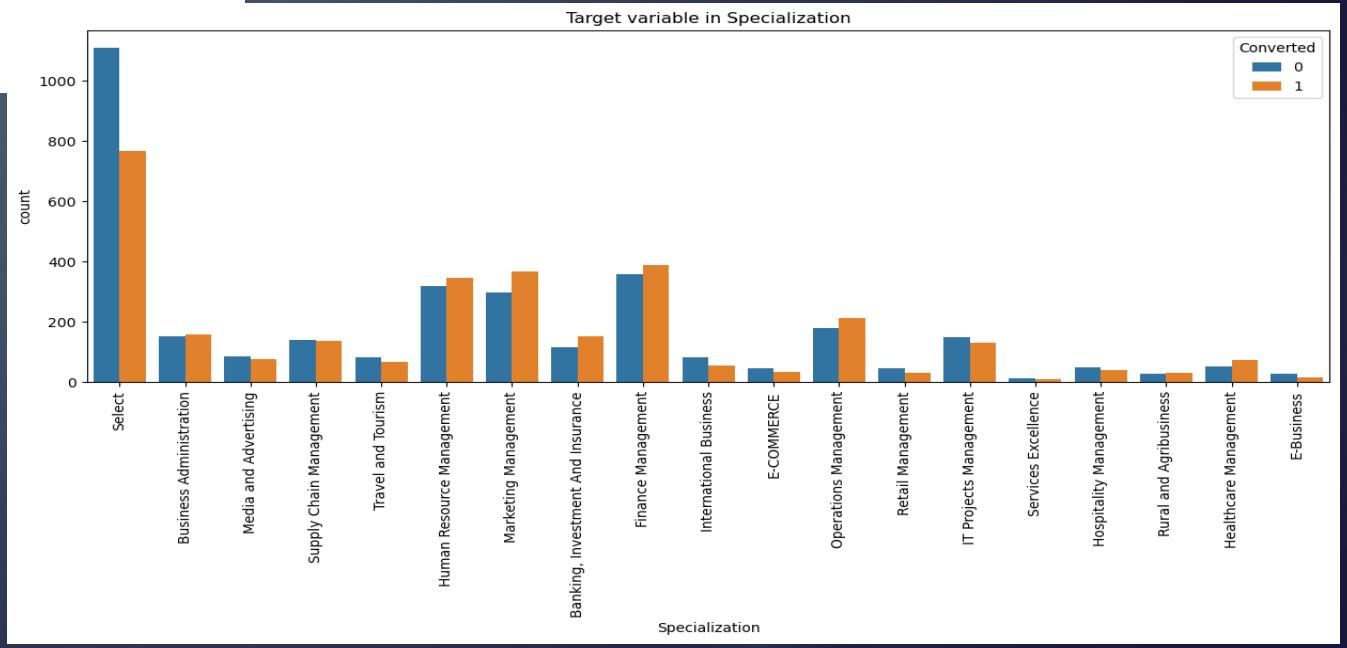


- Following seems to be the ratio of lead conversions for current occupation and specializations



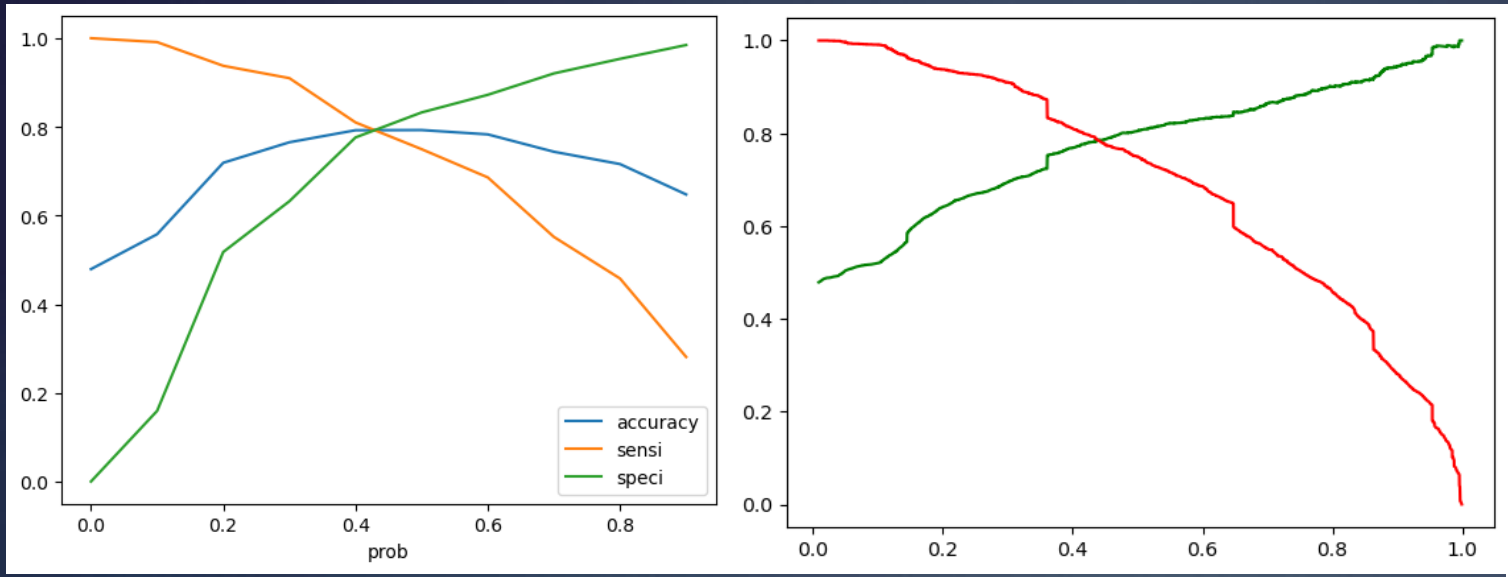
There seems to be a high conversion rate for Working professionals followed by unemployed and students

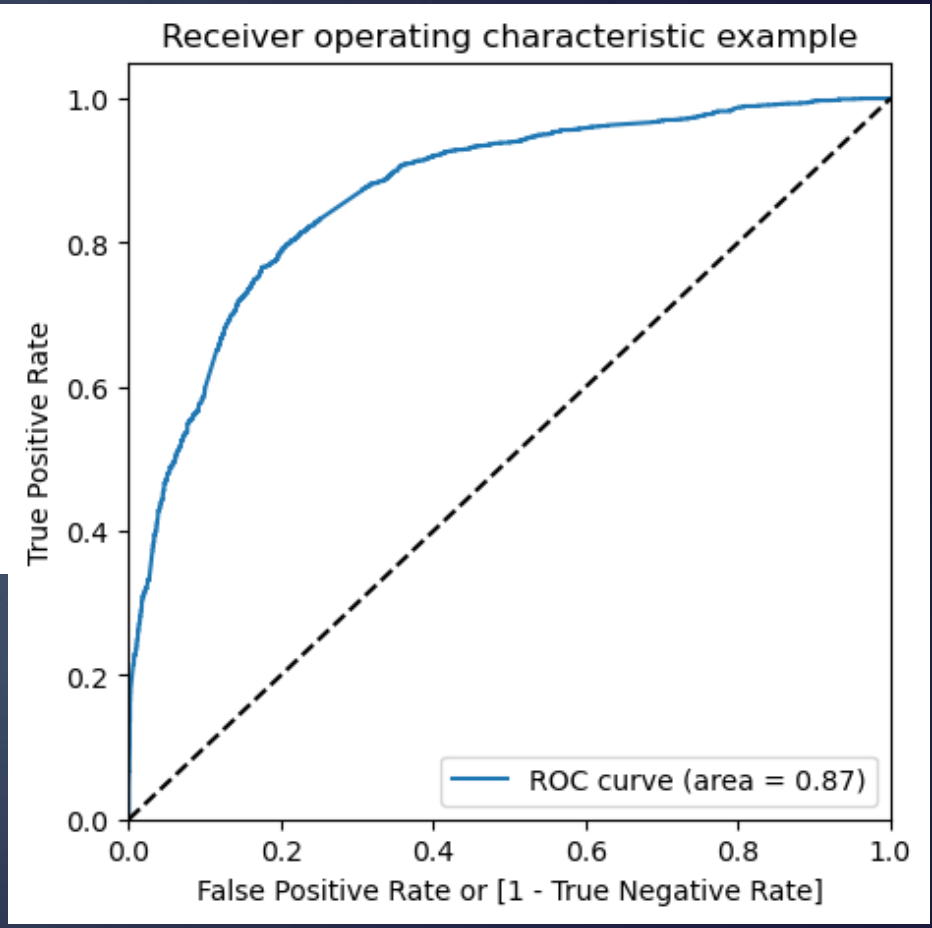
Leads belonging to Finance Management , banking and marketing tend to enroll more



Model Evaluation:

- 0.41 seems to be the optimal cutoff point for the model with ROC being 87%





Model Statistics:

Train Data:

1. Accuracy : 79.23%
2. Sensitivity : 80.35%
3. Specificity : 78.21%

Test Data:

1. Accuracy : 77.78%
2. Sensitivity : 79.95%
3. Specificity : 75.71%

Final Feature list:

- Total Visits
- Total Time Spent on Website
- Page Views per Visit
- Lead Origin_Lead Add Form
- Lead Source_Olark Chat
- Lead Source_Welingak Website
- Do Not Email_Yes
- Last Activity_Conveted to Lead
- Last Activity_Email Bounced
- Last Activity_Had a Phone Conversation
- Last Activity_Olark Chat Conversation
- Last Activity_SMS Sent
- What is your current occupation_Working Professional
- Last Notable Activity_Unreachable

Conclusion:

- There seems to be a high conversion rate for leads that Add Form and had sources as Reference, Welingak, Google and Organic search
- There seems to be a high conversion rate for Working professionals followed by unemployed and students
- Leads belonging to Finance Management , banking and marketing tend to enroll more
- Total time spent on website , Lead Origin add form and Total Visits are the top variables that seem to be contributed most towards the lead getting converted, all three are showing positive contribution towards the leads getting successfully converted into enrolling to the program
- Be proactive on sending the relevant SMS's, they impact the lead conversion probability positively.
- Ignore leads that either do not email or revert on emails.