

# Contents

---

- Problem Statement
- Problem Solving Approach
- EDA
- Correlations
- Model Evaluations
- Observations
- Conclusion

# Problem Statement

---

- An education company named X Education sells online courses to industry professionals. On any given day, many professions who are interested in the courses land on their website and vrowse 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%. 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

---

- To build a model to give every lead a lead score between 0 – 100. So that Hot Leads can be identified, and conversion rate can be increased.
- The CEO want to achieve a lead conversion rate of 80%.
- Model is to be build to handle future constraints as well like Peak time actions required, how to utilize full manpower and after achieving target what should be the approaches.

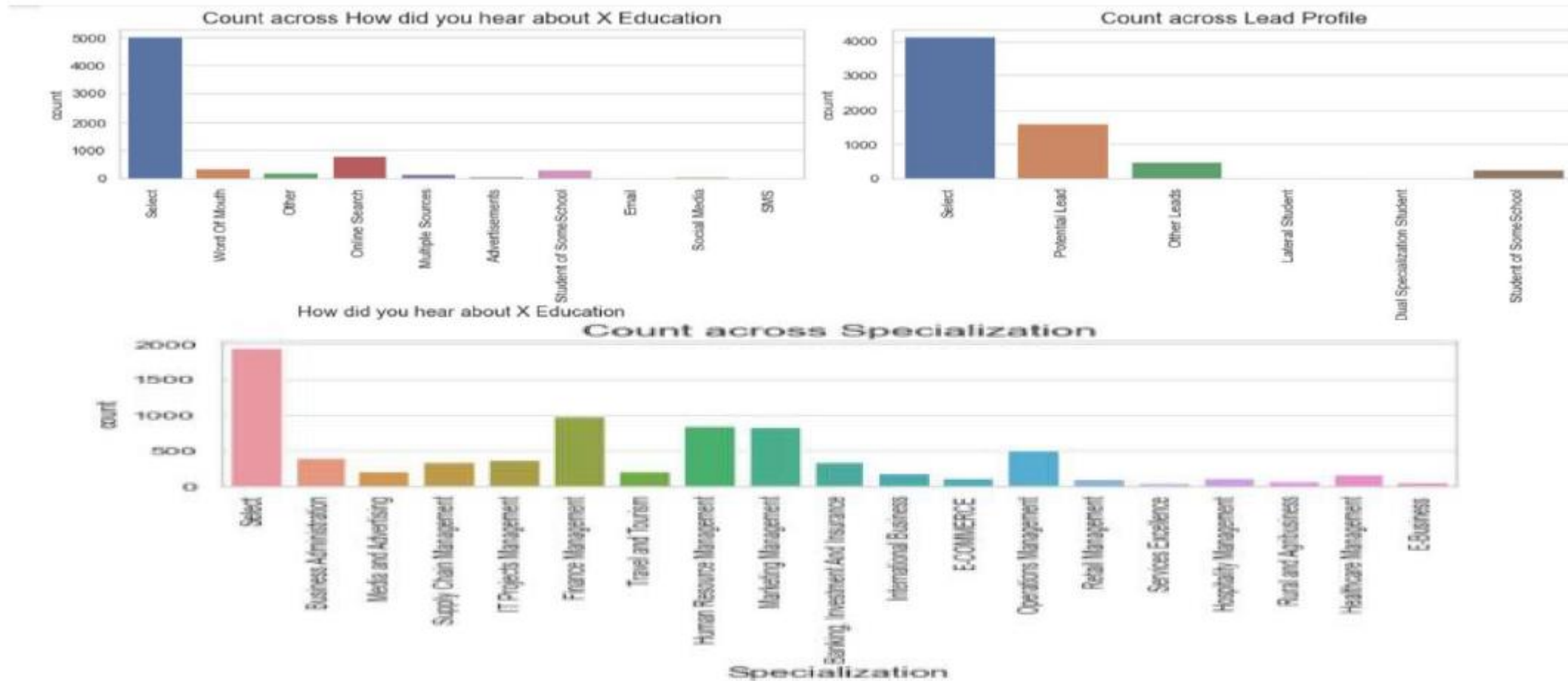
# Problem Solving Approach

---

- Importing the data and inspecting the data frame.
- Data Cleaning & Data Preparation
- EDA
- Dummy variable creation
- Test-Train Split
- Feature Scaling
- Correlations
- Model Building (RFE R-squared VIF and p-values)
- Model Evaluation
- Making Predictions on Test Set

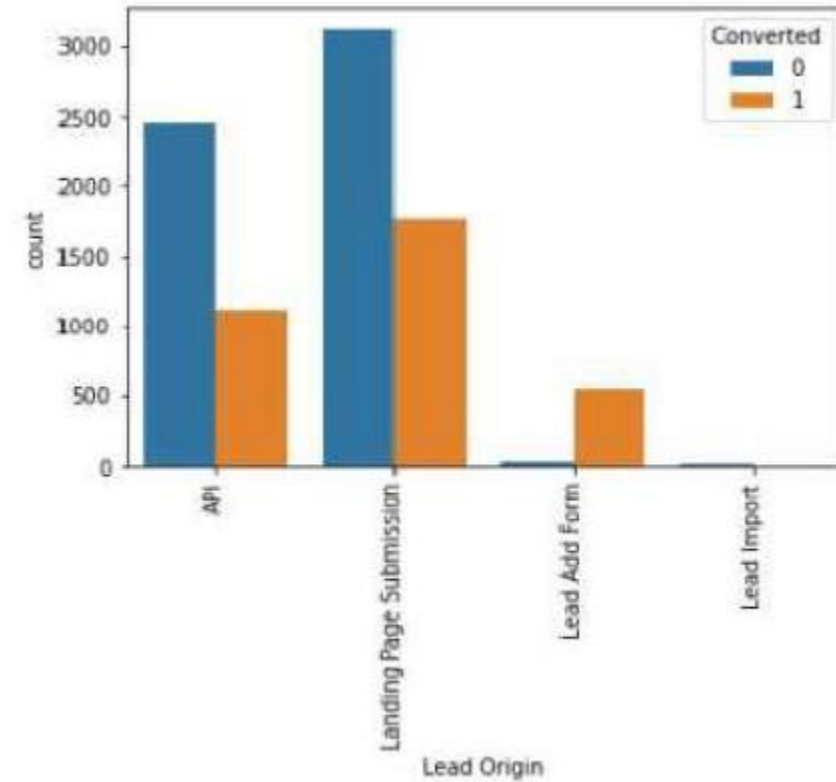
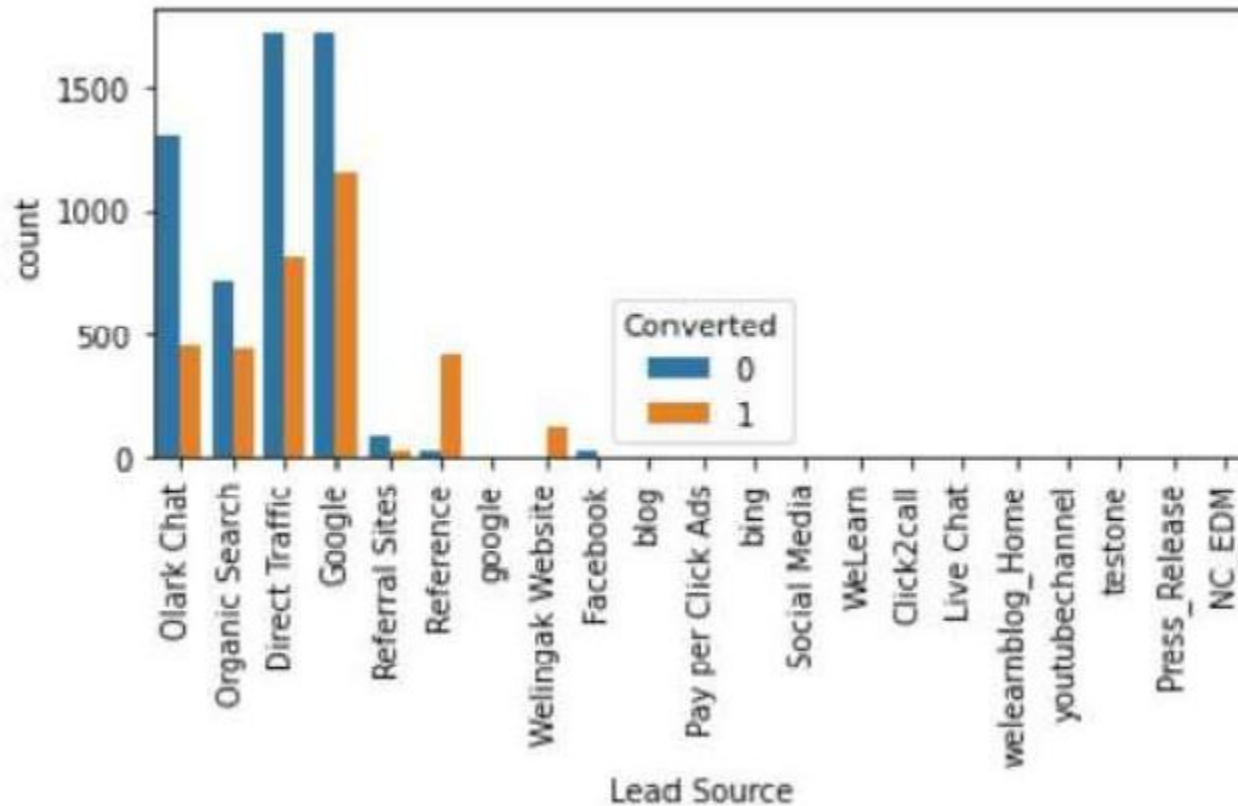
# EDA – Data Cleaning

- Identified the columns in which there is a level called 'Select' and handled the values



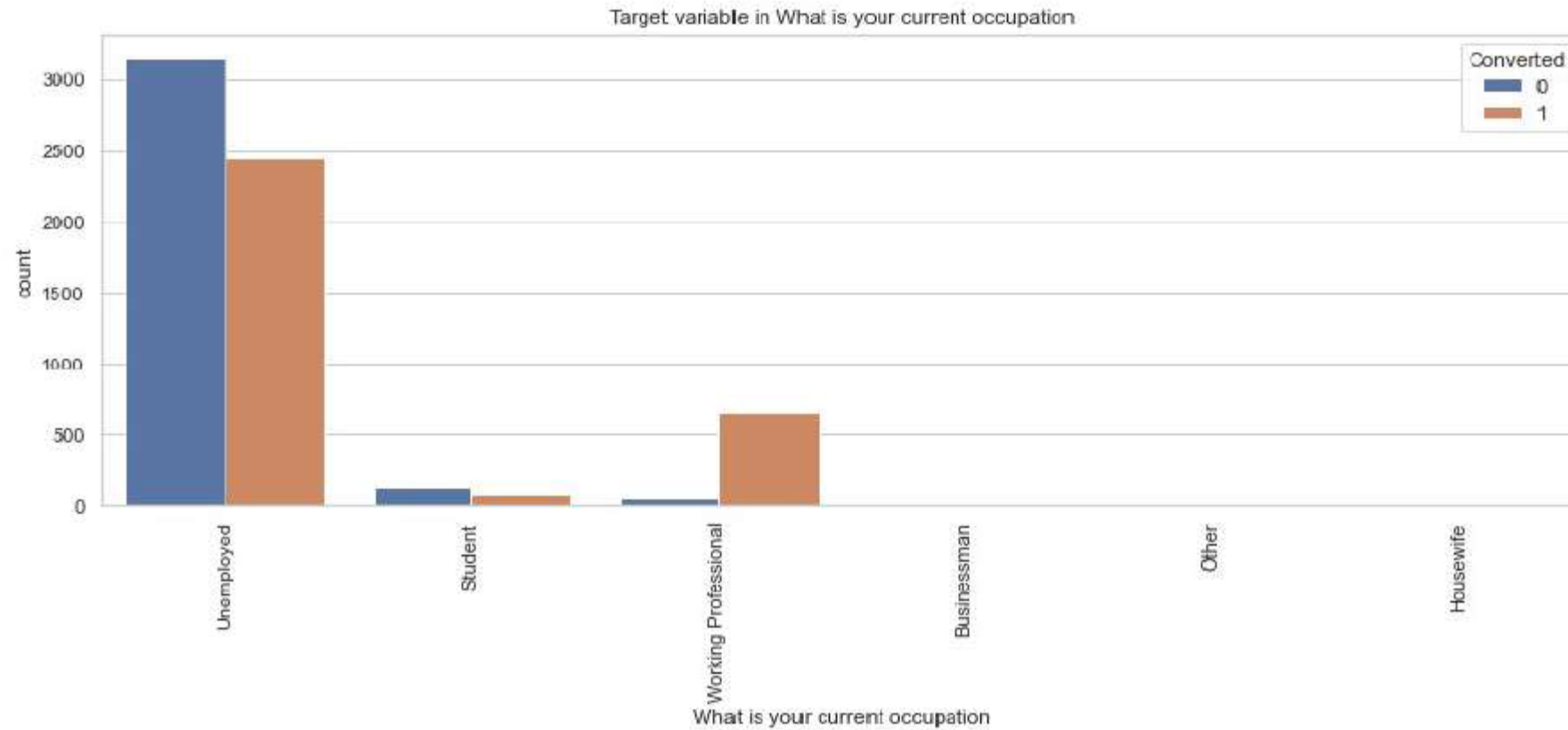
# Variables : Lead Source & Lead Origin

- In Lead source the leads through google & direct traffic are high probability to convert.
- In Lead origin the greatest number of leads are landing on Submission



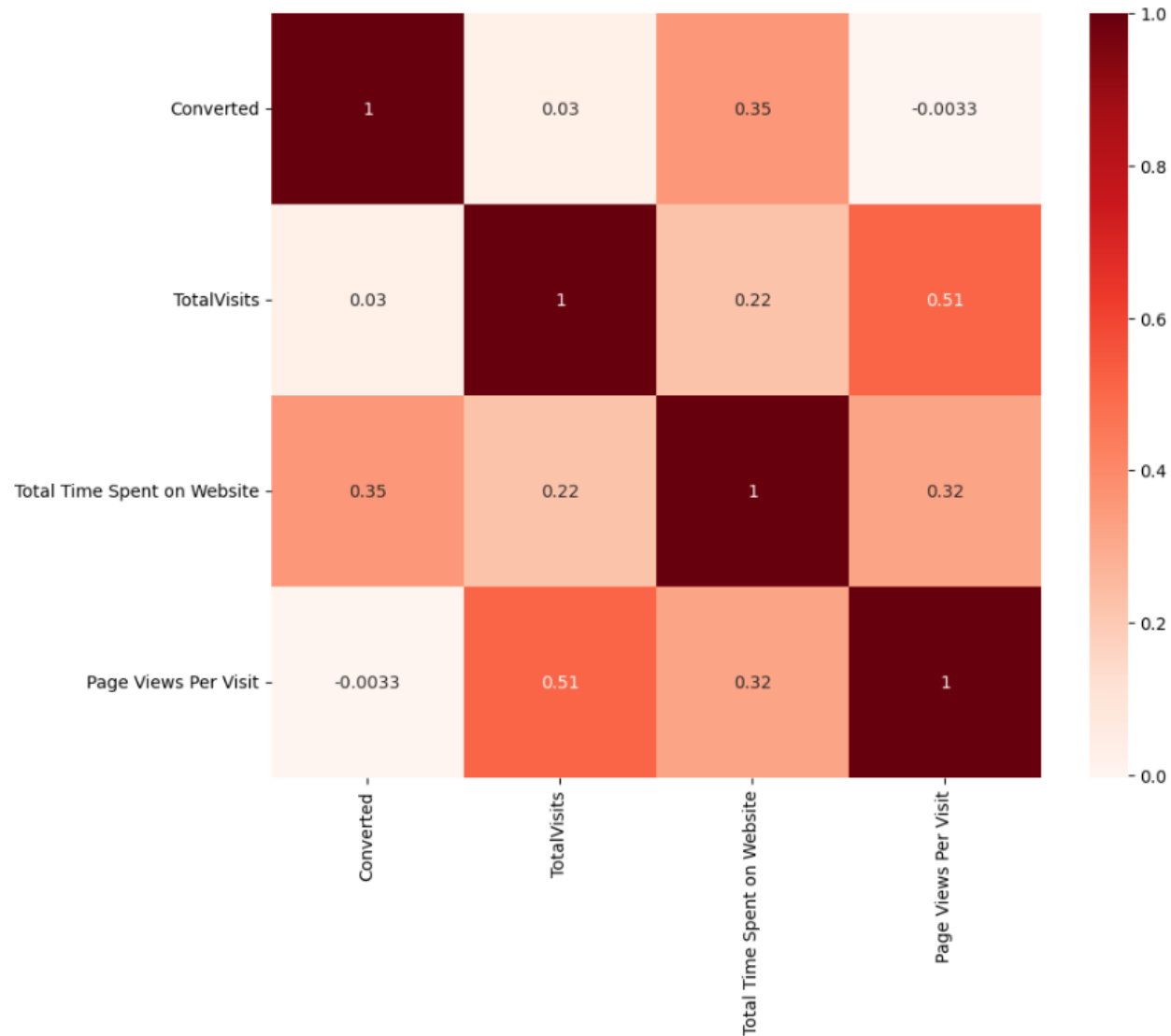
# Variable : Last What is your occupation

- Leads who are Unemployed are more interested in the course.



# Correlations

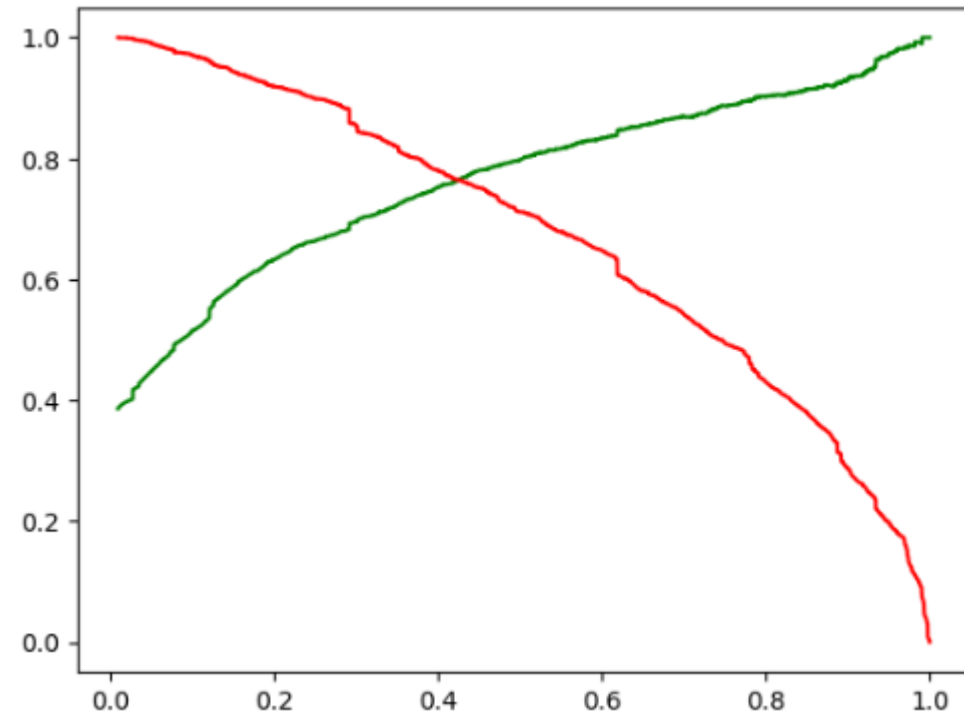
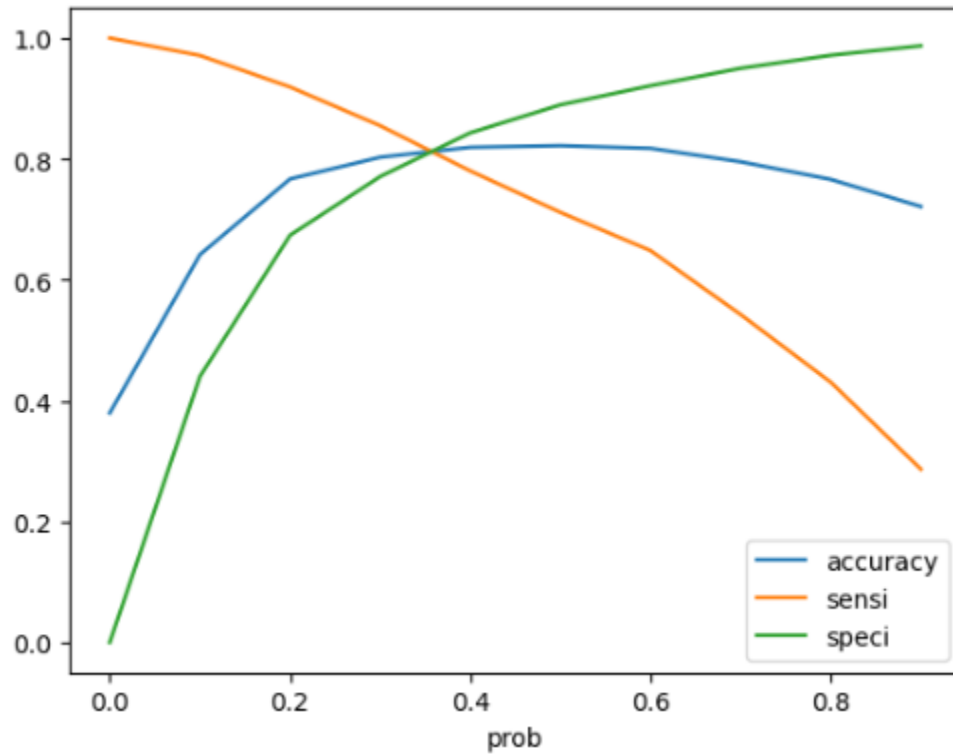
- There is no correlations between variables.





# Model Evaluation

- ROC Curve : 0.4 is the tradeoff between Precision & Recall.
- Thus, we can safely choose to consider any Prospect Lead with Conversion Probability higher than 40% to be a Hot Lead.



# Observations

---

- Train Data

- Accuracy : 80%
- Sensitivity : 85%
- Specificity : 77%

- Test Data

- Accuracy : 81%
- Sensitivity : 85%
- Specificity : 78%

## Final Features List

- Last Notable Activity\_Modified
- TotalVisits
- Last Activity\_Email Link Clicked
- Last Notable Activity\_Email Link Clicked
- What is your current occupation\_Unemployed
- Last Activity\_Page Visited on Website
- Last Notable Activity\_Page Visited on Website
- Last Activity\_Olark Chat Conversation
- Lead Source\_Olark Chat
- Specialization\_Others
- Total Time Spent on Website
- Last Activity\_Email Bounced
- Last Notable Activity\_Email Opened
- Do Not Email\_Yes
- Lead Origin\_Lead Add Form
- Last Notable Activity\_Olark Chat Conversation
- Last Activity\_Converted to Lead
- Lead Source\_Welingak Website
- What is your current occupation\_Working Profes
- Last Activity\_Form Submitted on Website
- What is your current occupation\_Student
- Lead Origin\_Lead Import
- Specialization\_Hospitality Management

# Conclusion

---

- We see that conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore, we can intervene that we need to focus more on the leads originated from API and Landing page submission
- We observed that max number of leads are generated by google /direct traffic. Max conversion ration is by reference and welingak website
- Leads who spent more time on website are more likely to convert.
- Most common last activity is email opened. Highest rate = SMS Send. Maximum are Unemployed. Max conversion with working professional.