



# Lead Score Case Study

**SUBMITTED BY:**

DHYANI SHAH

SAVITHA D KRISHNAPPA

DURGA S

# Problem Statement:

- ▶ An education company named X education wants to sell courses to industry professionals.
- ▶ On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- ▶ Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be 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

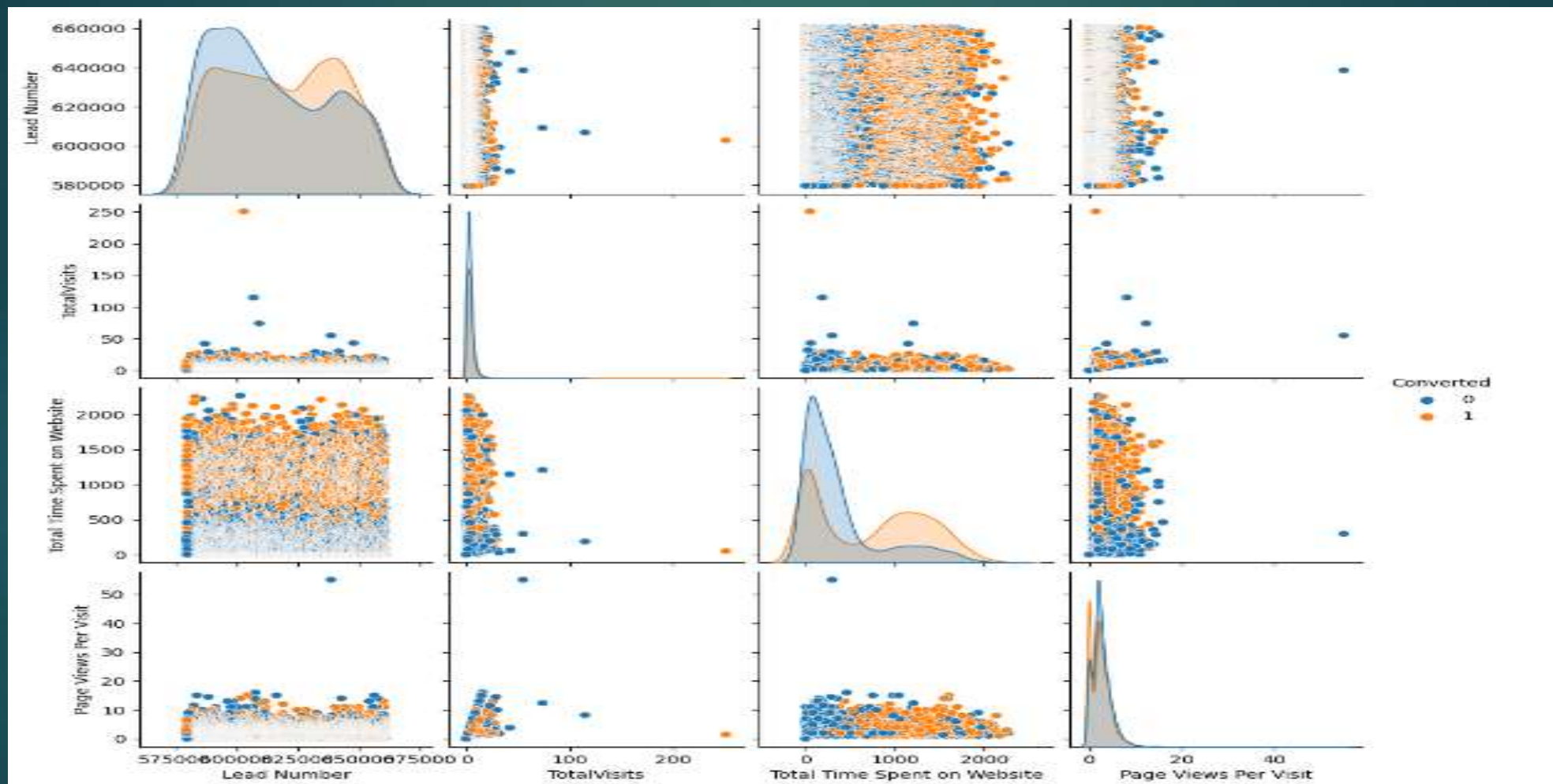
# Objective:

- ▶ Company X wants us to make a model which identifies “Hot leads” and increase there conversion as well.
- ▶ Company CEO wants to conversion rate near to 80%

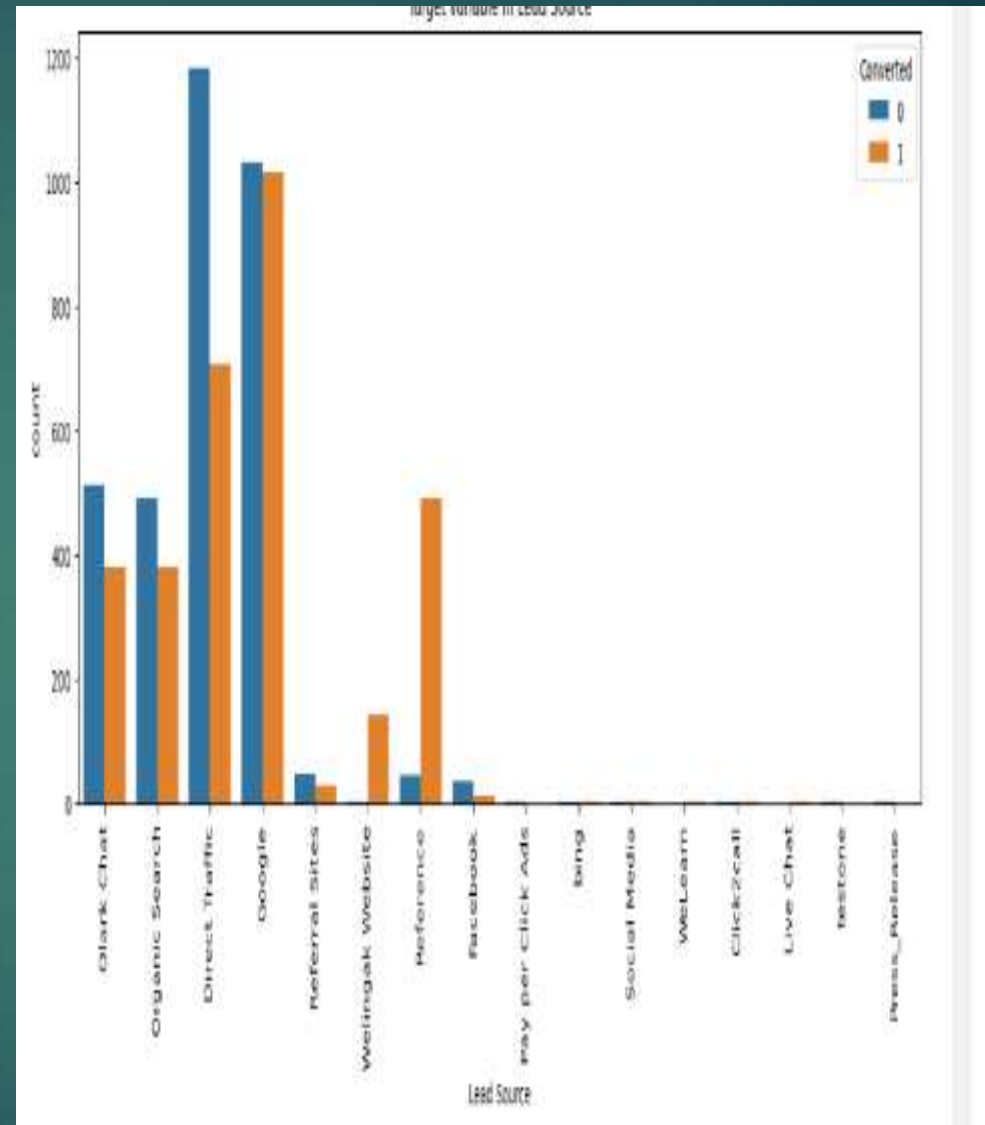
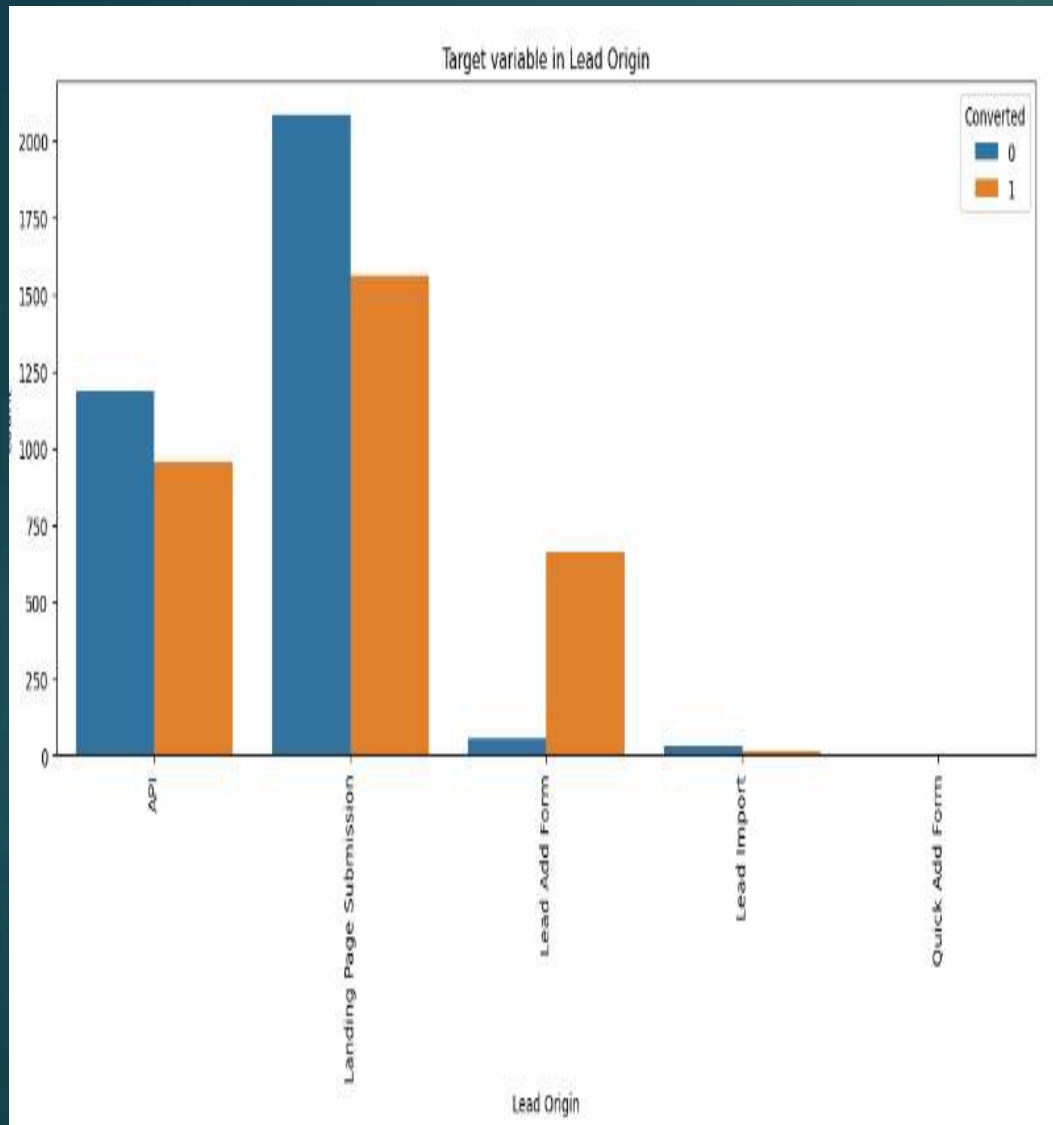
# Approach and Methodology:

- ▶ Import data and find out shape, data types, uniqueness of data
- ▶ Missing values , imbalance of data
- ▶ Dropping columns
- ▶ EDA
- ▶ Dummy Variable creation
- ▶ Train- Test Split
- ▶ Feature Scaling
- ▶ Correlation
- ▶ Model Building with help of RFE , VIF and P value
- ▶ Model Evaluation
- ▶ Making prediction on test set

# EDA:

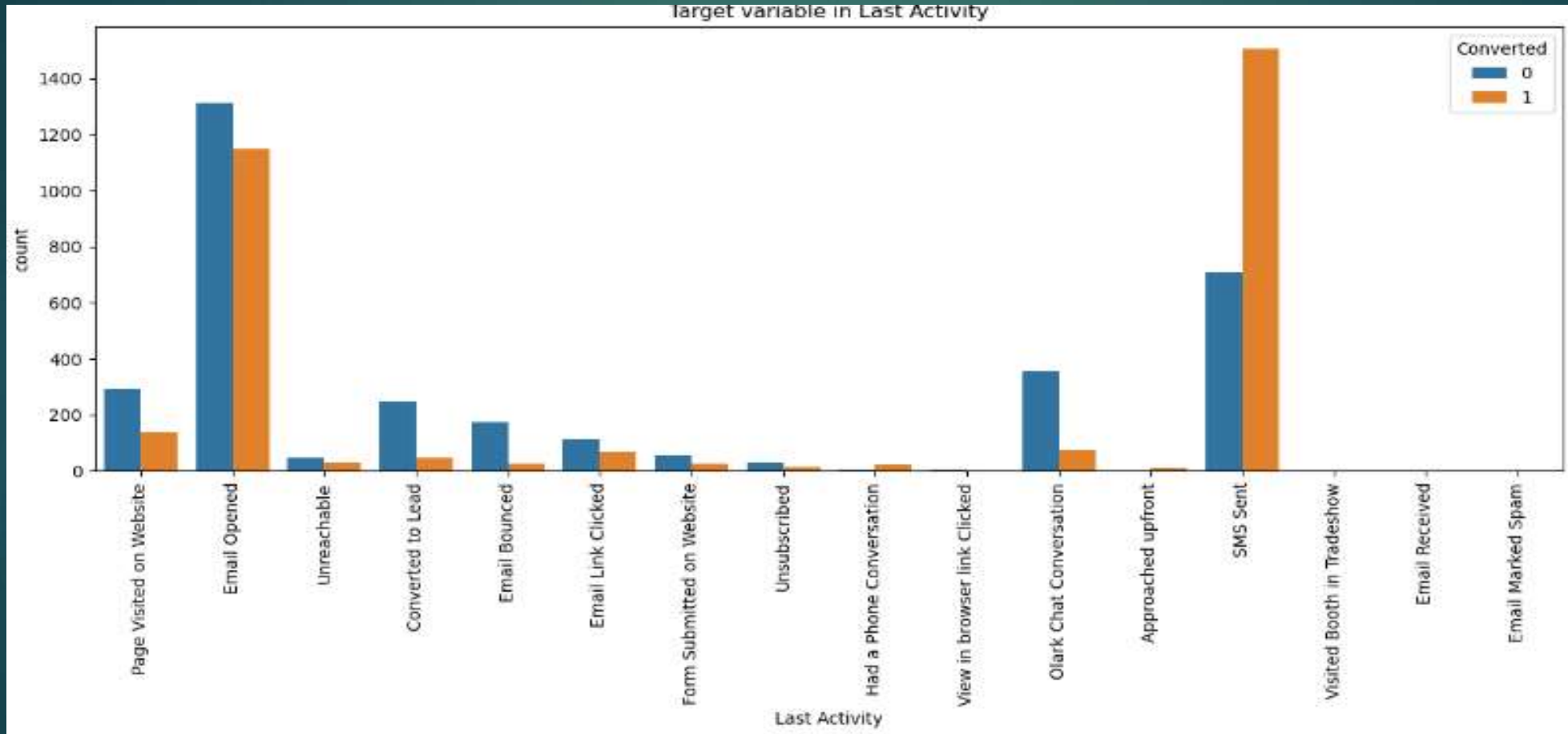


EDA on continuous variables



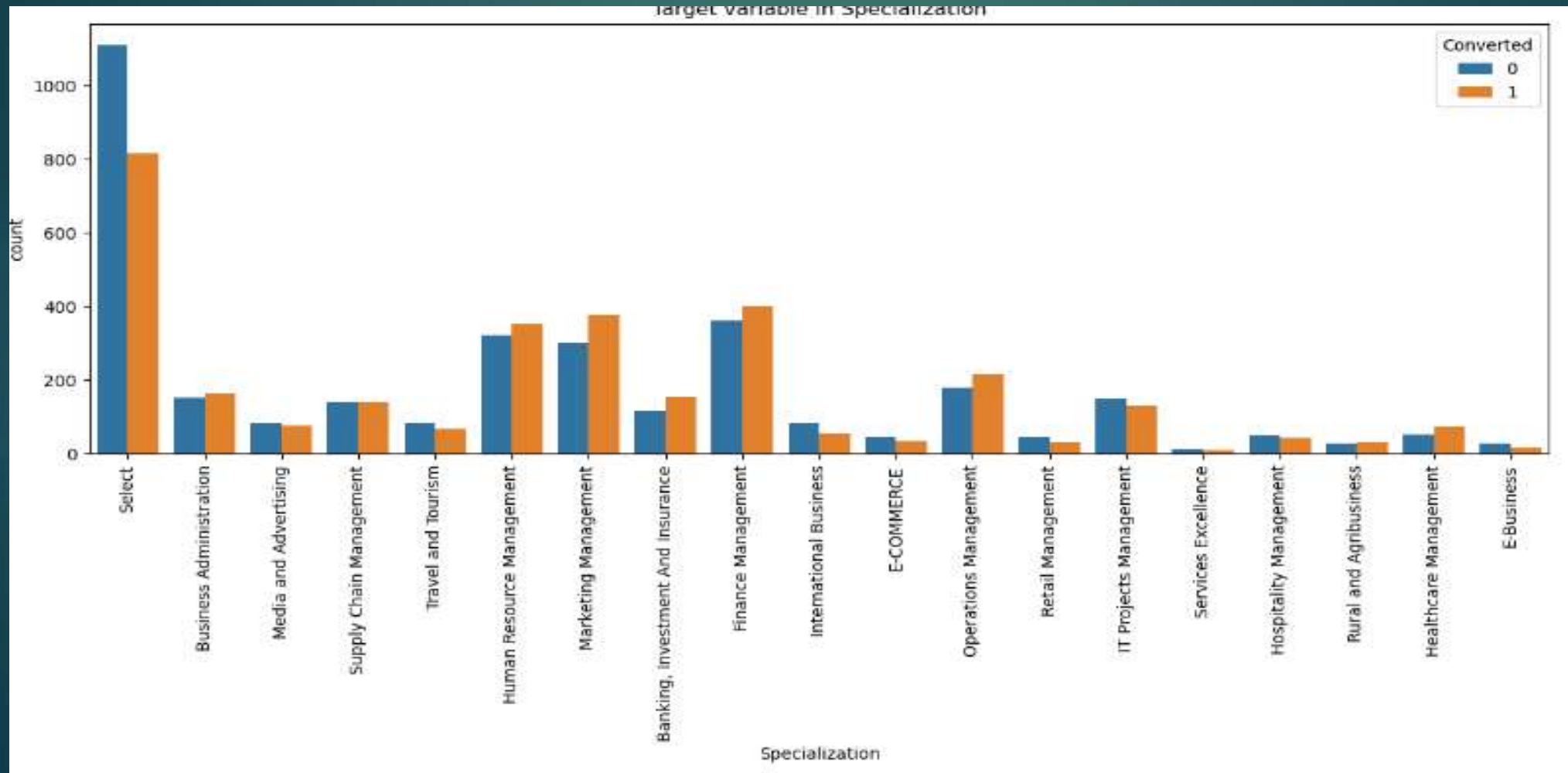
In lead origin most of the numbers of leads shows that leads are landing on submission whereas in lead source we can see high probability of convert through google and direct traffic

# Leads which are opening email and sending SMS have high probability to convert



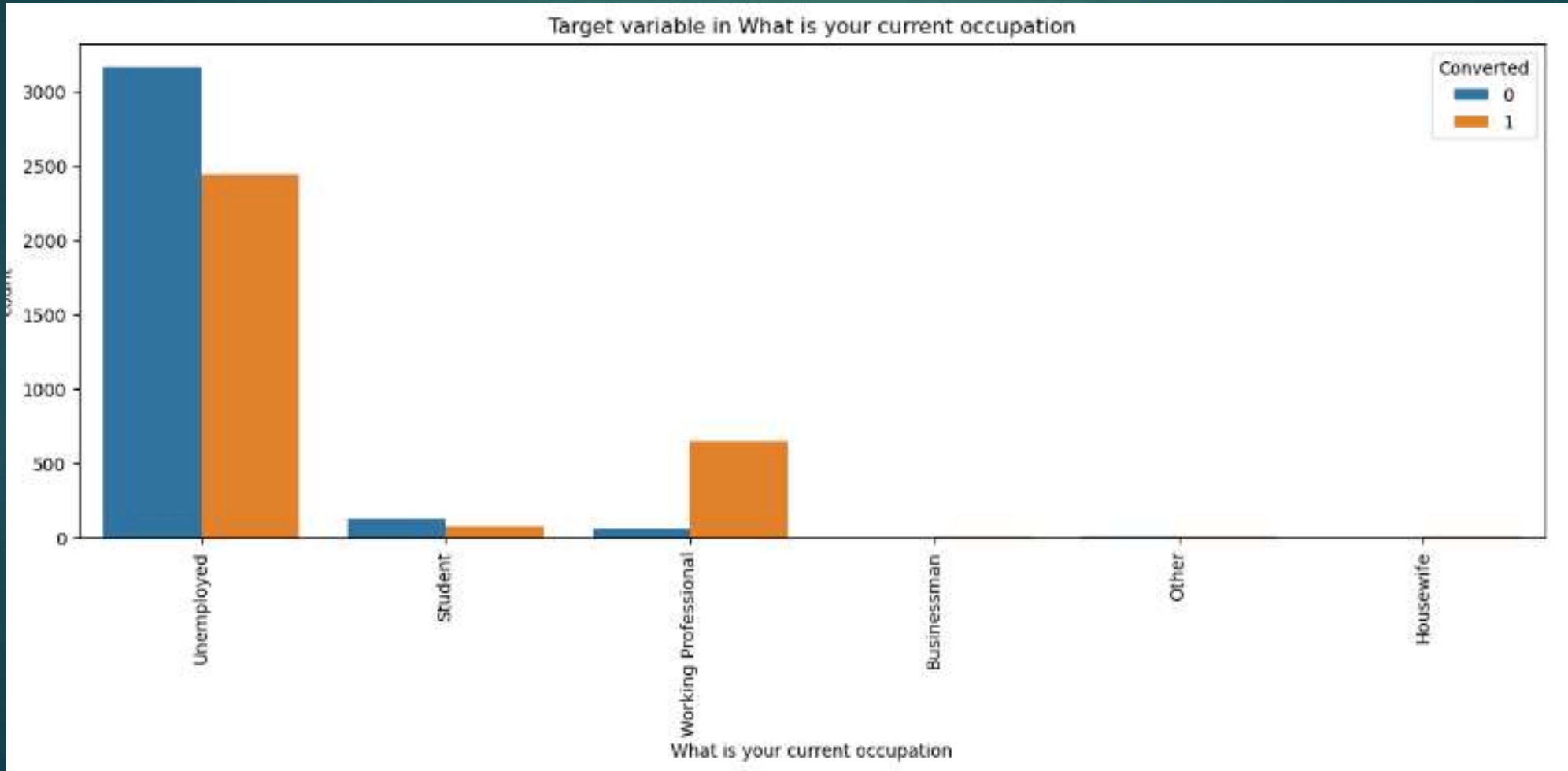


In specialization we can see leads who are in financial management ,HRM and marketing management are high likely to convert

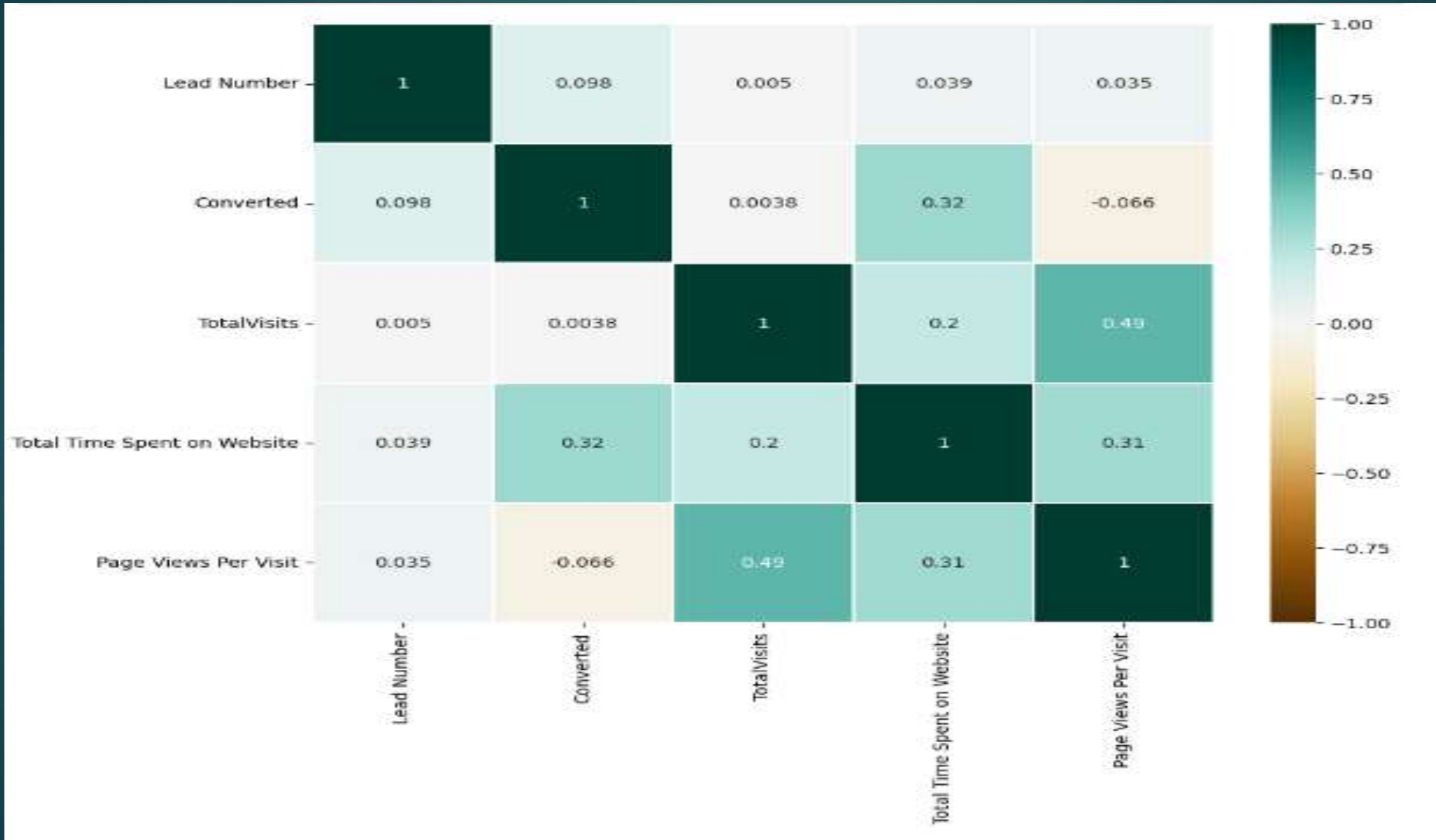




# Leads who are unemployed are more likely to convert rather than others:

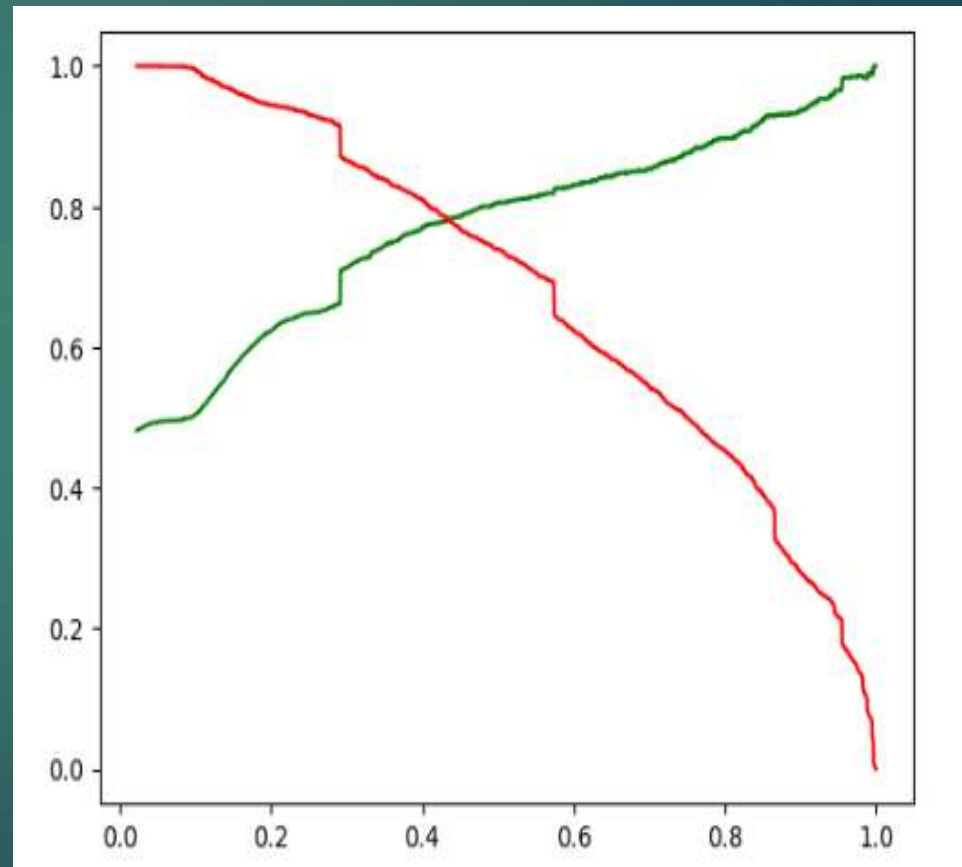
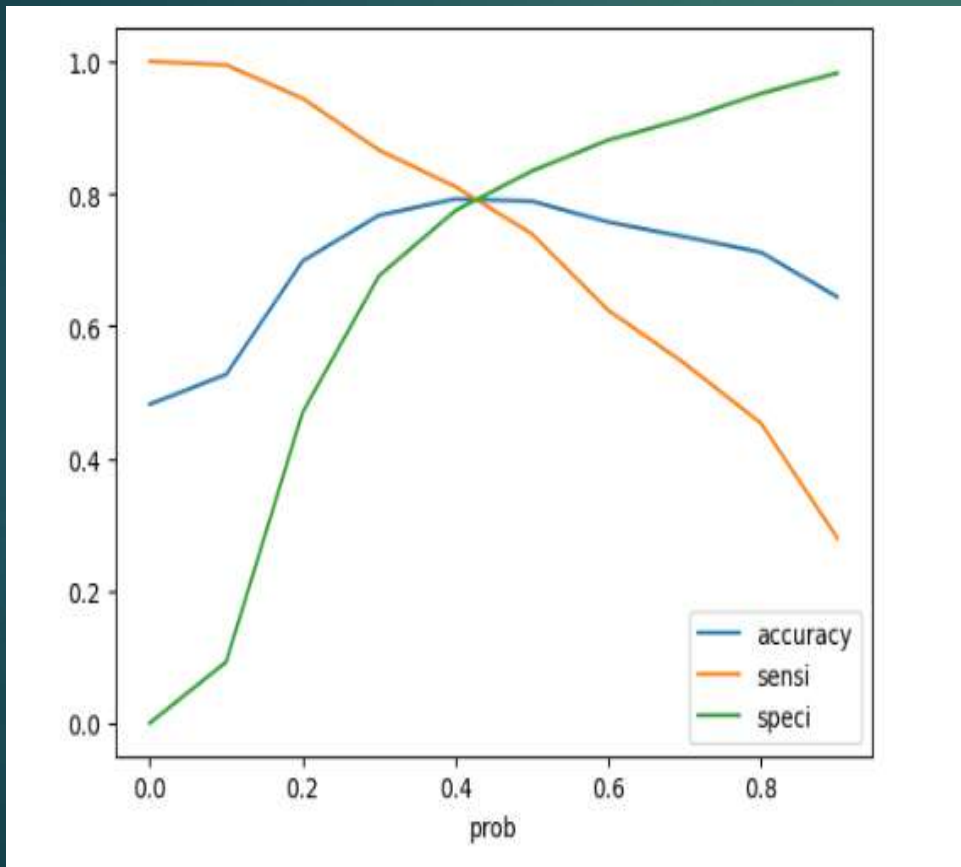


# There are no correlation between the different variables:



# Model Evaluation:

Trade off between Precision and recall comes out to be 0.42 so we can choose leads conversion probability higher than 42% as a hot lead.



# Observations:

## ▶ **Train Data:**

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

## ▶ **Test Data:**

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

# Cont...

## ▶ Final Variable List:

- ▶ Lead Source\_OlarkChat
- ▶ Specialization Others
- ▶ Lead Origin\_Lead Add Form
- ▶ Lead Source\_WelingakWebsite
- ▶ Total Time Spent on Website
- ▶ Lead Origin\_LandingPage Submission
- ▶ What is your current occupation\_WorkingProfessionals
- ▶ Do Not Emal

# Conclusion:

- ▶ We can see that maximum leads are converted by google and direct traffic.
- ▶ Maximum conversion ratios is by reference and welingak website.
- ▶ Leads who are spending more time on website most likely to convert
- ▶ Those who opened email and SMS are more likely to convert.
- ▶ Those who are more unemployed are most likely to convert
- ▶ We can also see that conversion rate is 30-35% for API and Landing page submission so we need to focus more on leads originated from the same.

*Thank you*