

Lead Scoring Case Study

USING LOGISTIC REGRESSION



SUBMITTED BY

Rukhsar Qureshi



Contents

- ◆ Problem statement
- ◆ Problem approach
- ◆ EDA
- ◆ Correlations
- ◆ Model Evaluation
- ◆ Observations
- ◆ Conclusion

Problem Statement

- A. 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 tags individual as a lead.
- B. 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.
- C. 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.
- D. 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

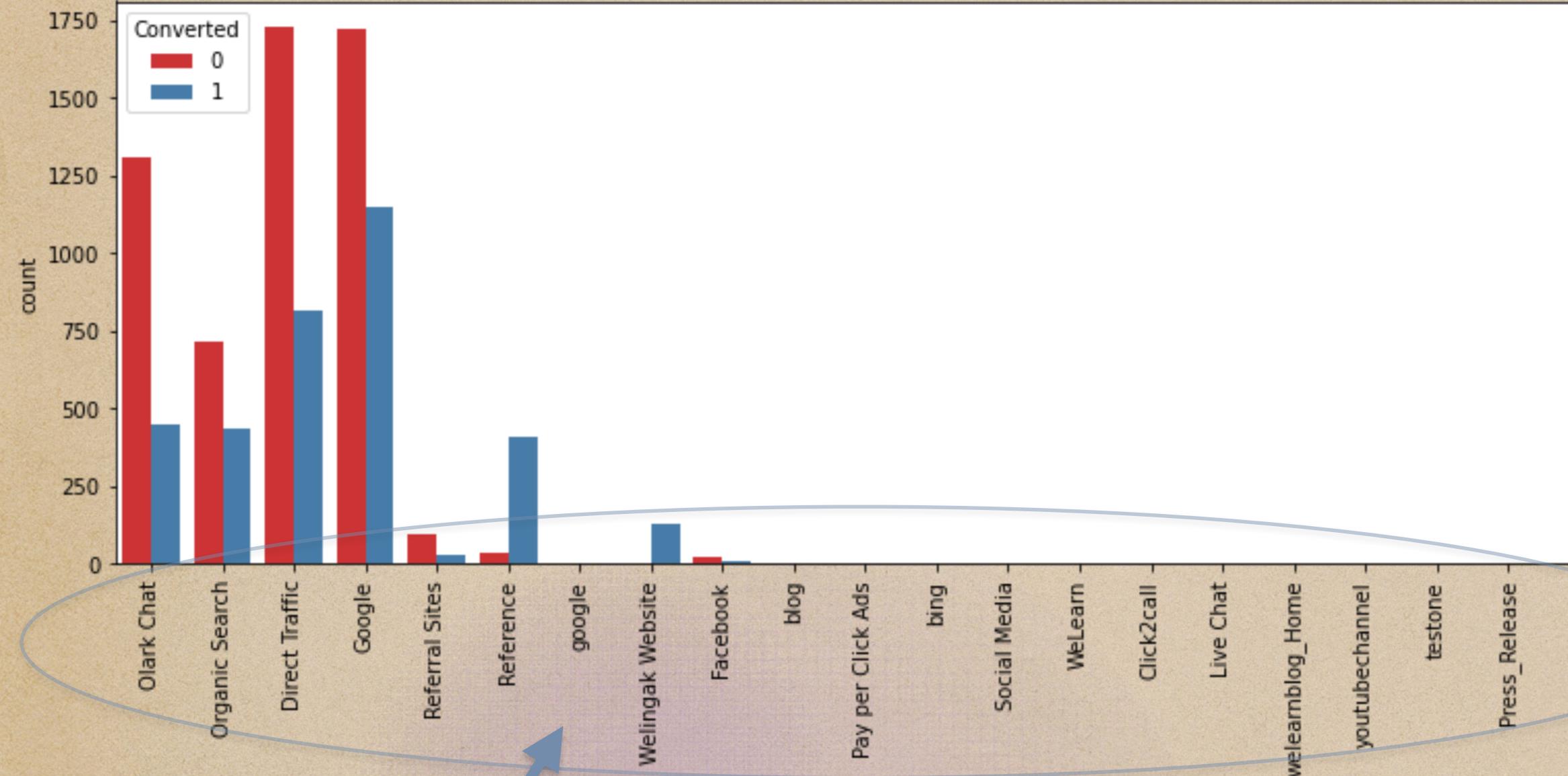
- A. The company has appointed us to help them select the most promising leads and wants us to build a model to give every lead a lead score between 0 -100. So that they can identify the Hot leads and increase their conversion rate as well. For this we have been provided with a leads dataset from the past with around 9000 data points.
- B. The CEO want to achieve a lead conversion rate of 80%.
- C. They want the model to be able to handle future constraints as well like Peak time actions required, how to utilise full man power and after achieving target what should be the approaches.

Problem Approach

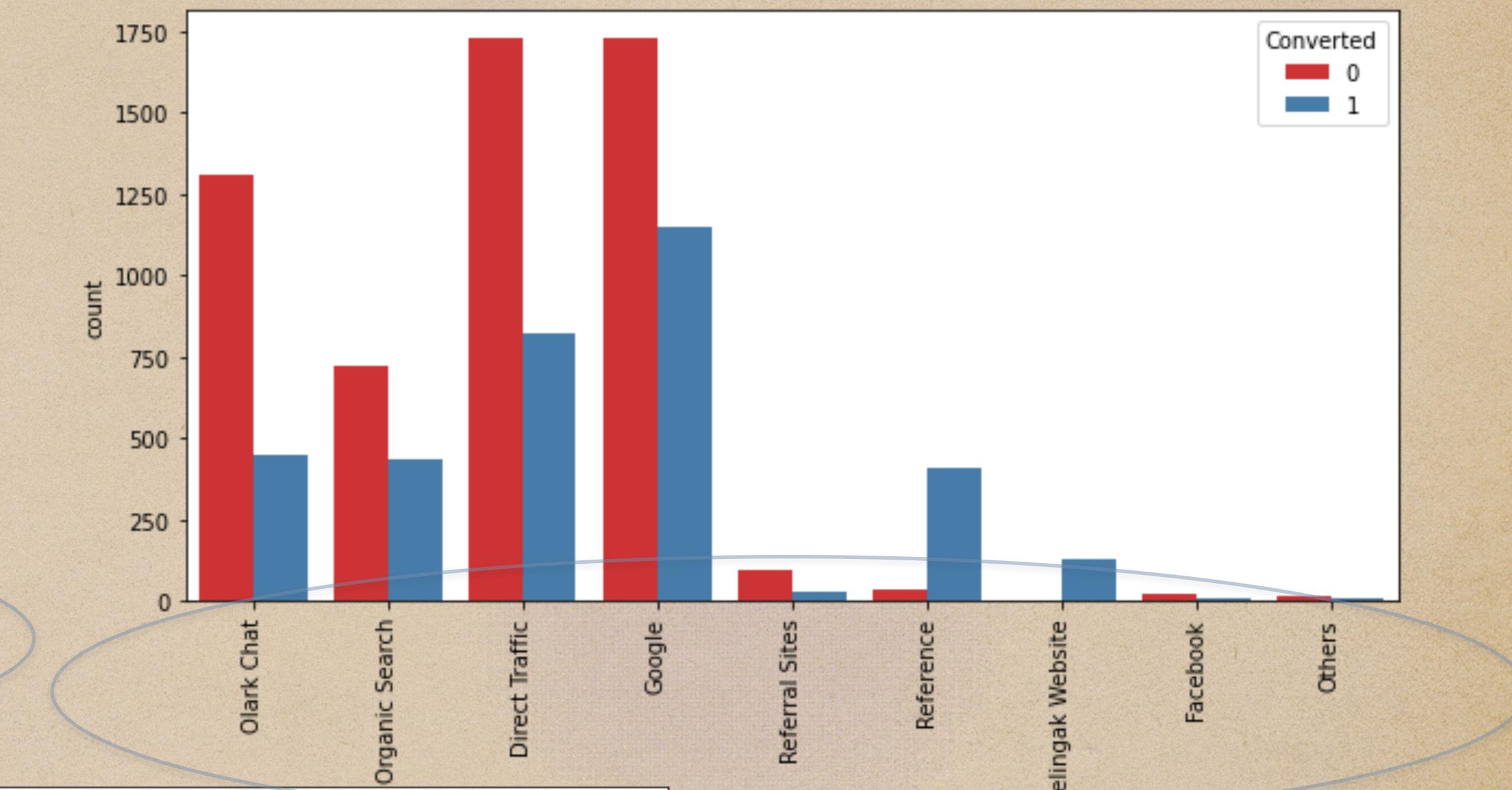
- A. Importing the data and inspecting the data frame
- B. Data preparation
- C. Dummy variable creation
- D. Test-Train split
- E. Making predictions on test set
- F. EDA
- G. Feature scaling
- H. Correlations
- I. Model Building (RFE Rsquared VIF and p- values)
- J. Model Evaluation

EDA Data Cleaning

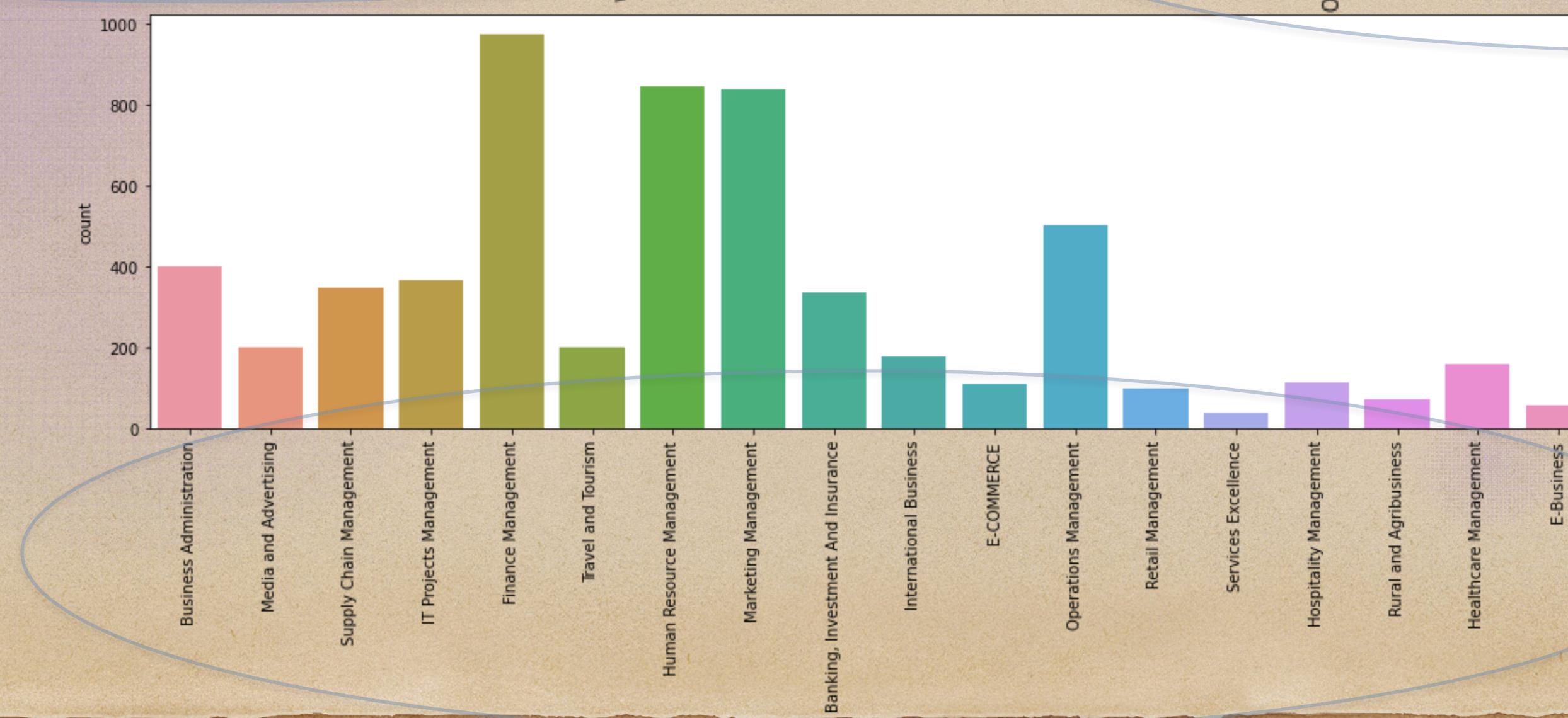
There are a few columns in which there is a level called 'Select' which needs taking care of.



Lead Source



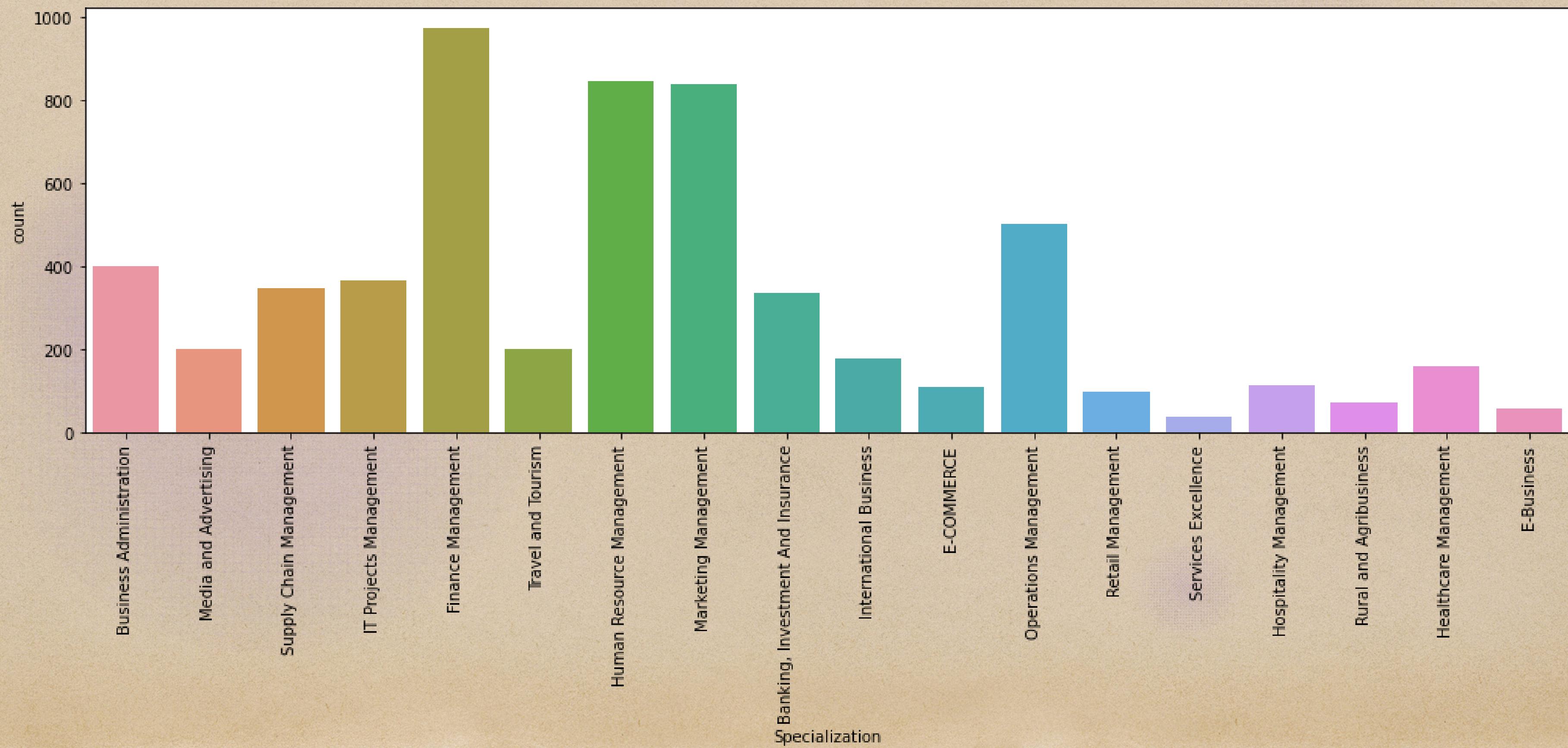
Lead Source



Specialisation

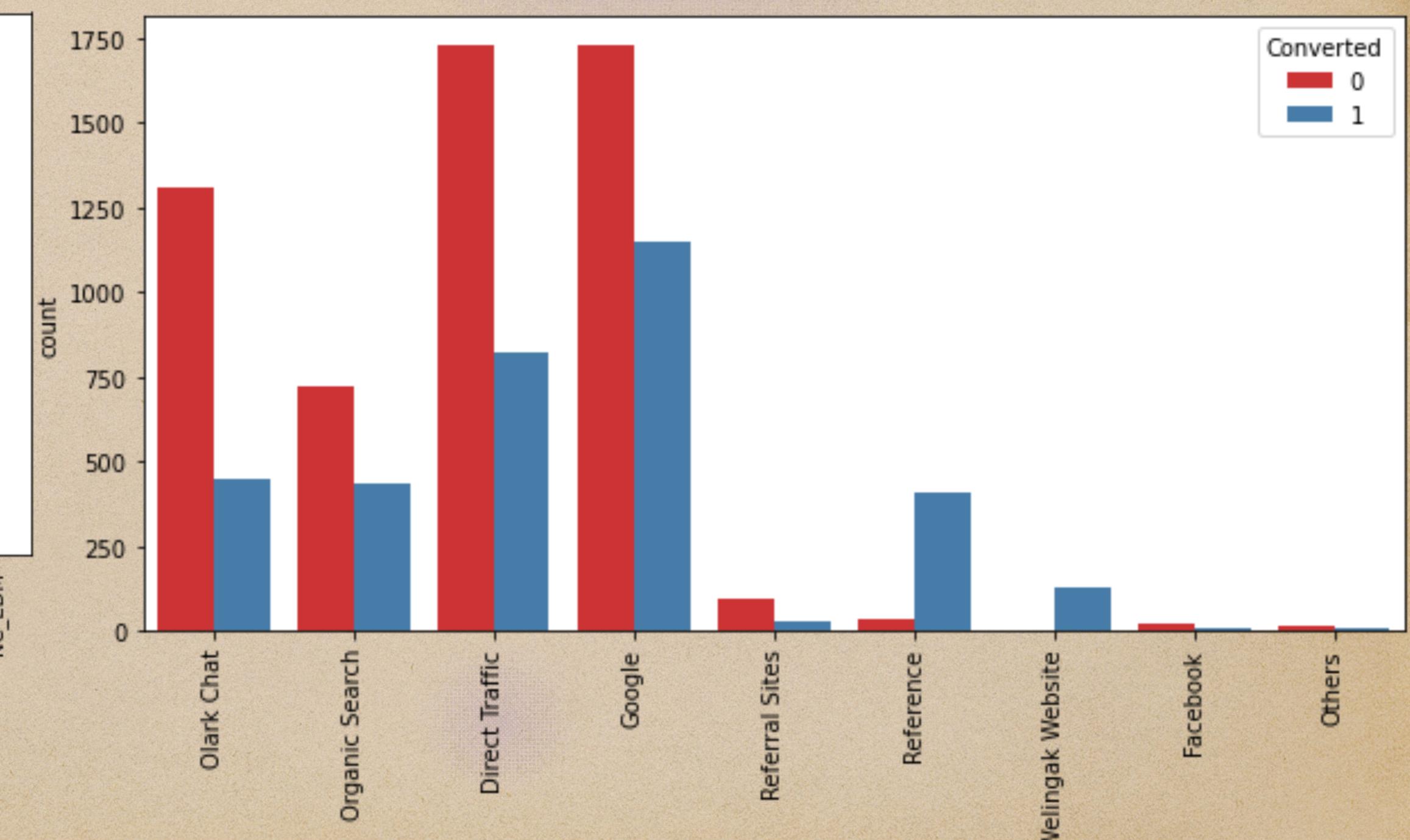
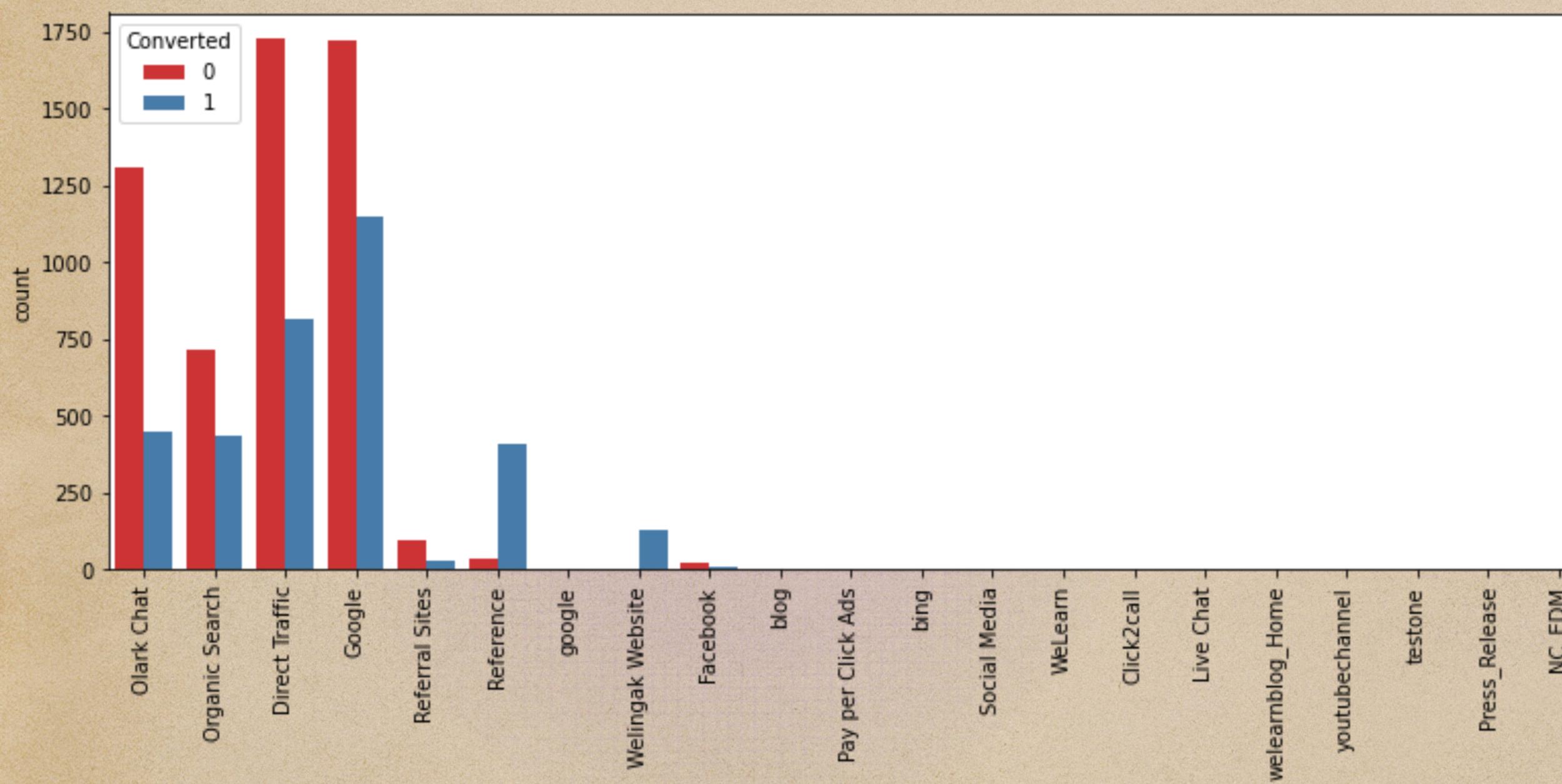
Specialization

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



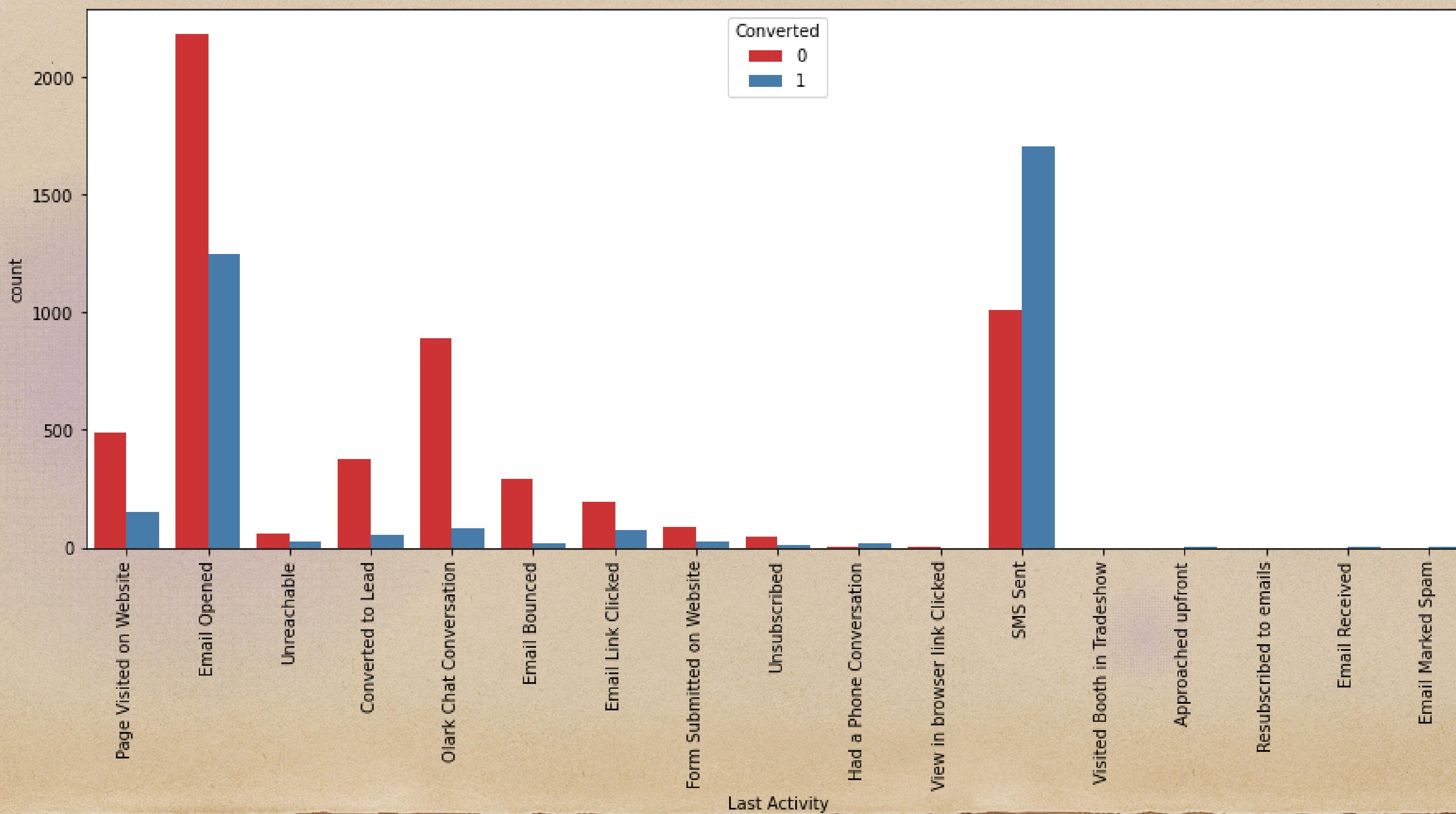
Lead Source & Lead origin

In lead source the leads through Google & direct traffic has high probability to convert, whereas in Lead origin most number of leads are on ‘Landing Page Submission’.



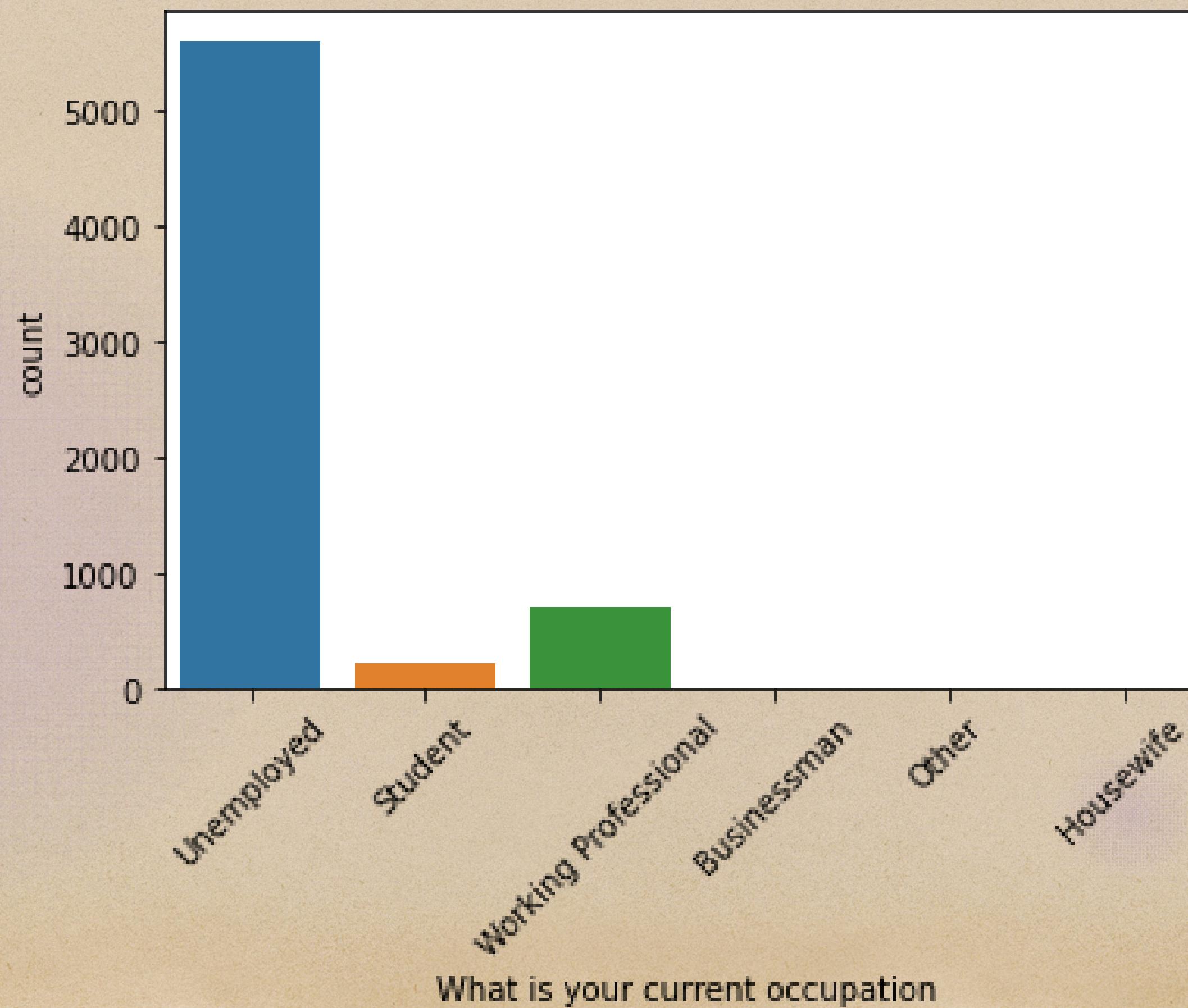
Last Lead Activity

Leads which are opening emails have higher probability to get converted, similar is the case with Sending SMSs.

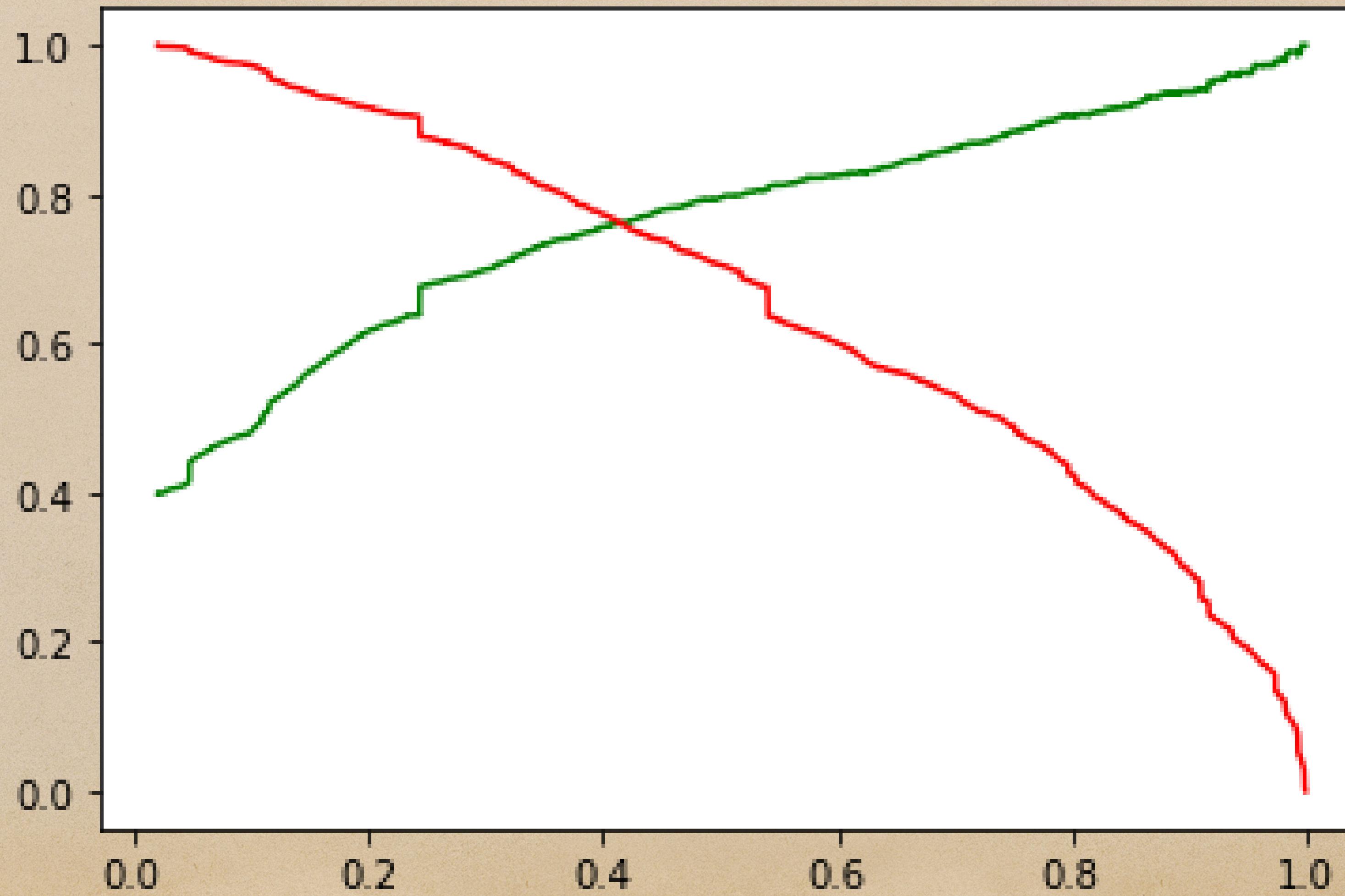


Last What is Your Occupation

Leads which are Unemployed are more interested to join the course than others.



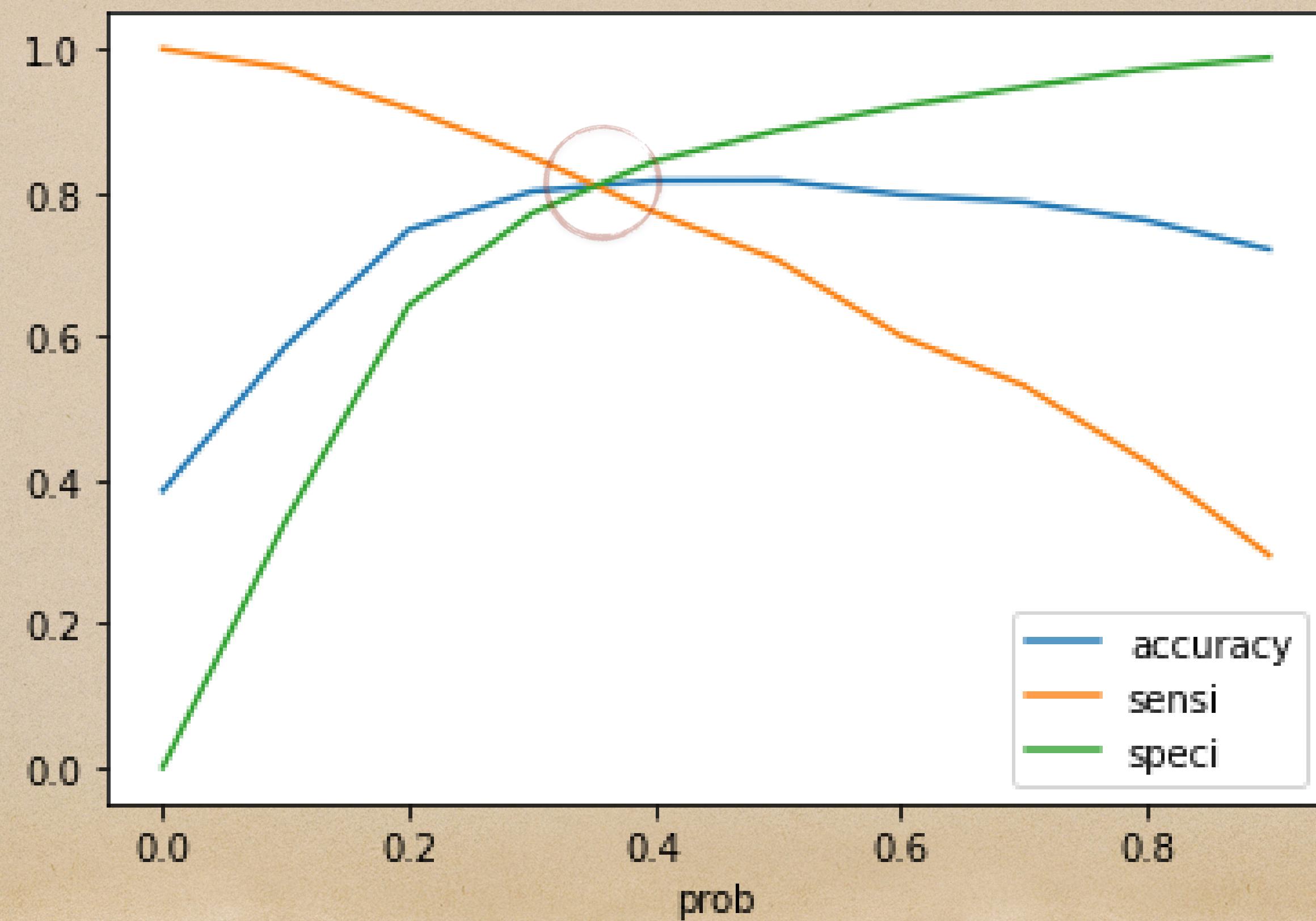
Model Evaluation



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

| | Test Data | Train Data |
|-------------|-----------|------------|
| Accuracy | 80% | 80% |
| Sensitivity | 77% | 77% |
| Specificity | 80% | 80% |

Final Feature-list

Lead Source: Olark Chat

Specialisation: Others

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

Conclusion

We see that the conversion rate is 30-35% (close to the 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 see max number of leads are generated by Google/Direct Traffic.

Max conversion ratio is by reference and Welingak website.

Leads who spent more time on website, more likely to convert.

Most common last activity is "email opened".

Highest rate = SMS Sent.

Max leads are "unemployed".

Max conversion with working professional.



THANKS