Lead Scoring Assignment

Anurag Anandkumar Shinde

Himanshu Thakur

Contents

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

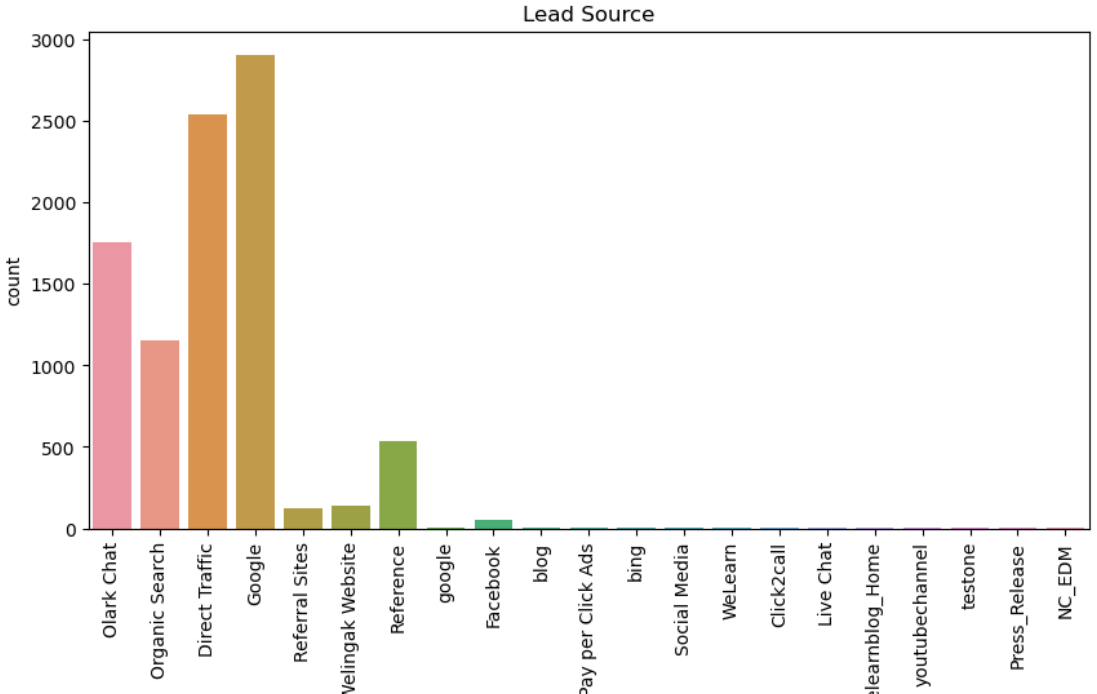
- X wants us to build a model to give every lead a lead score between 0 -100 , So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time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

Problem Approach

- Importing the data and inspecting the data frame
- 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

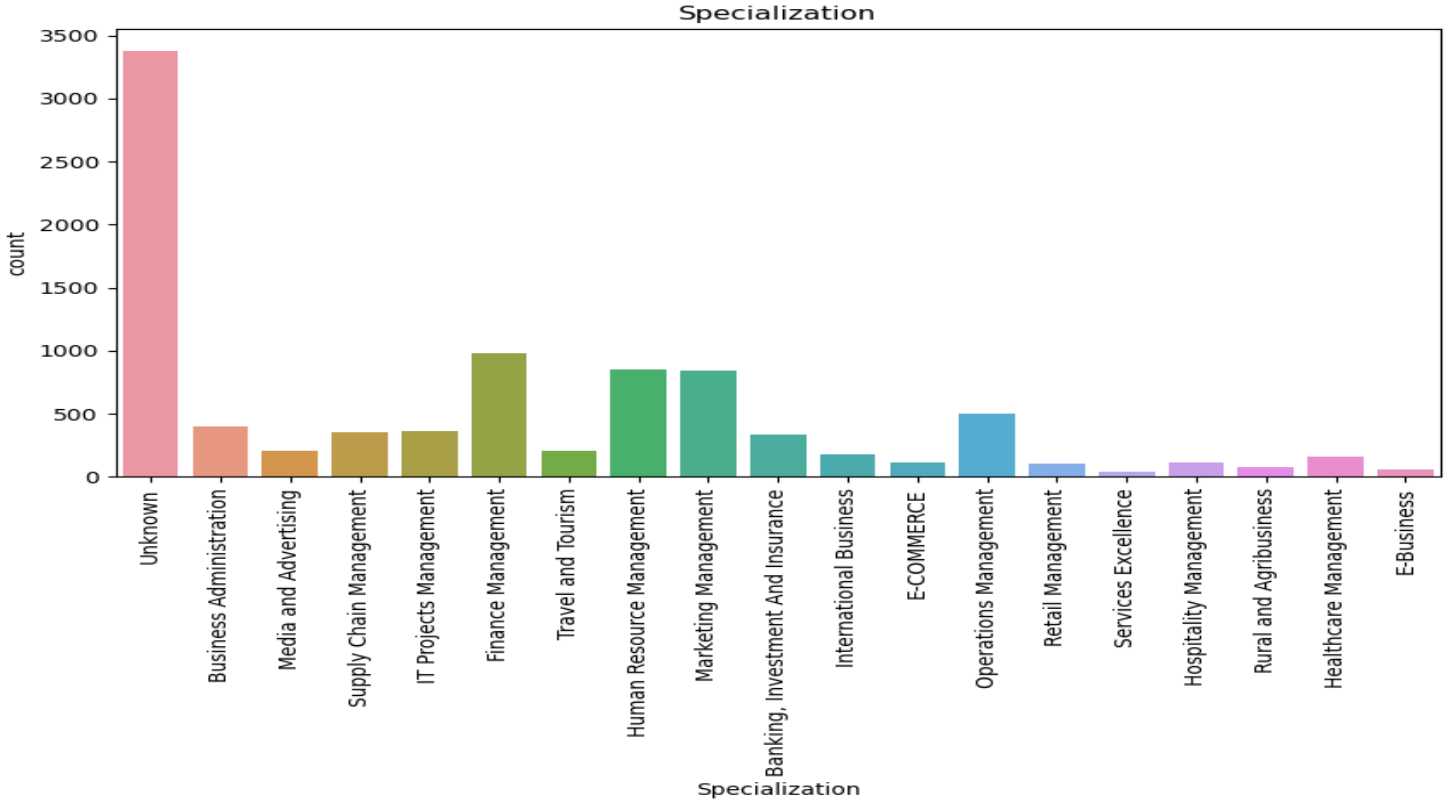
EDA

Most of the people came via Google followed by Direct Traffic source and Olark Chat.

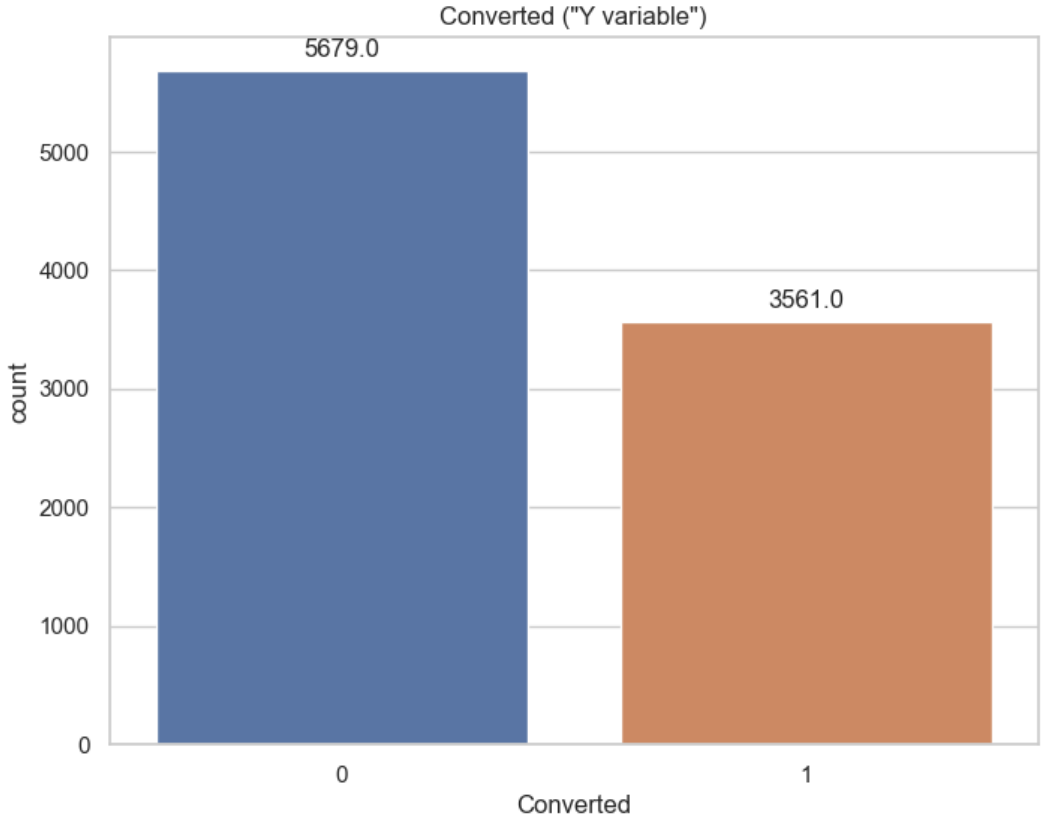


People in working in finance, human resource, marketing, operations management are more likely to be a lead.

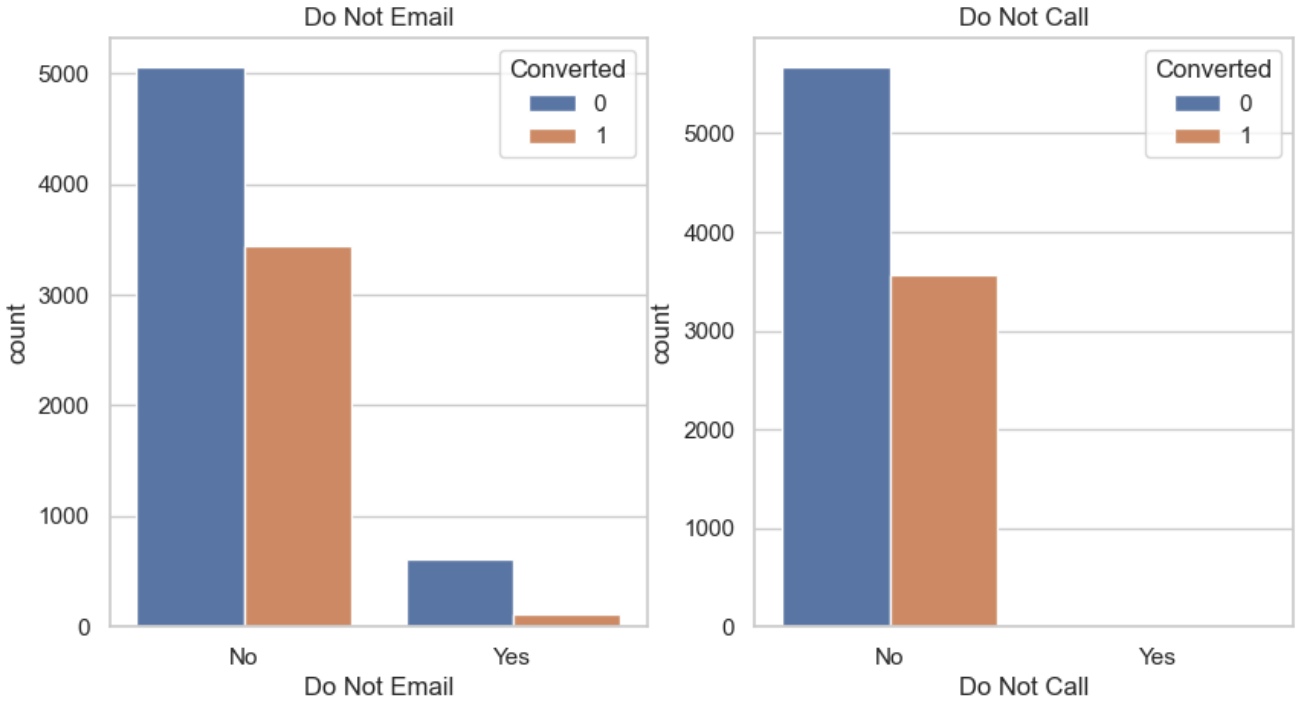
Most of the people did not mention their specialization.



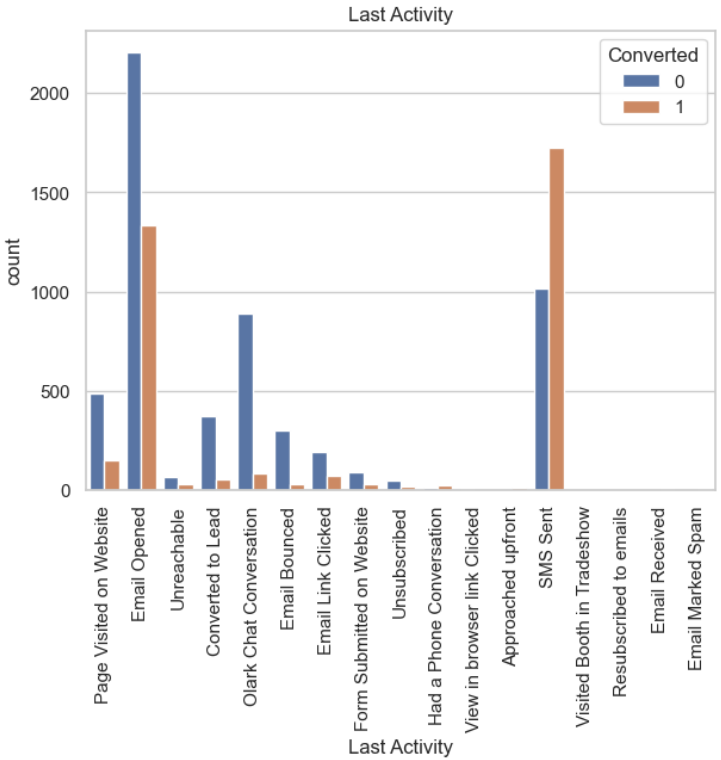
Out of the people who visited the website, 5679 haven't taken any course, while 3561 got converted to leads by joining a course from X education.

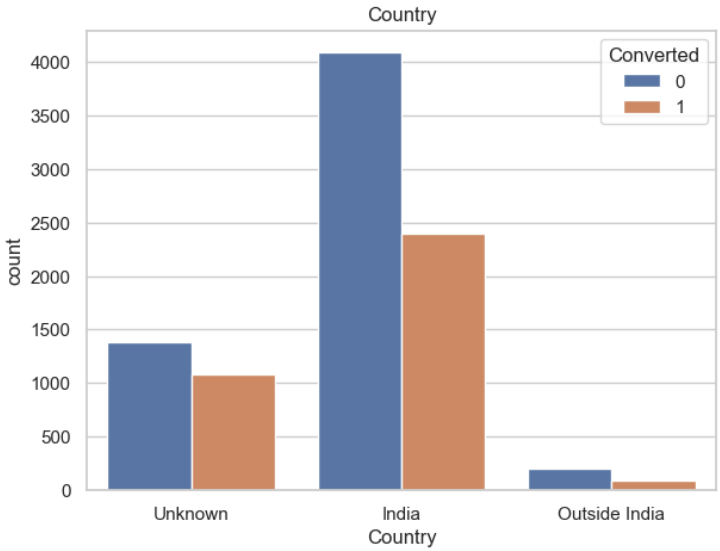


People who optioned out for no email and no call have high chances of getting converted into lead to join any course.



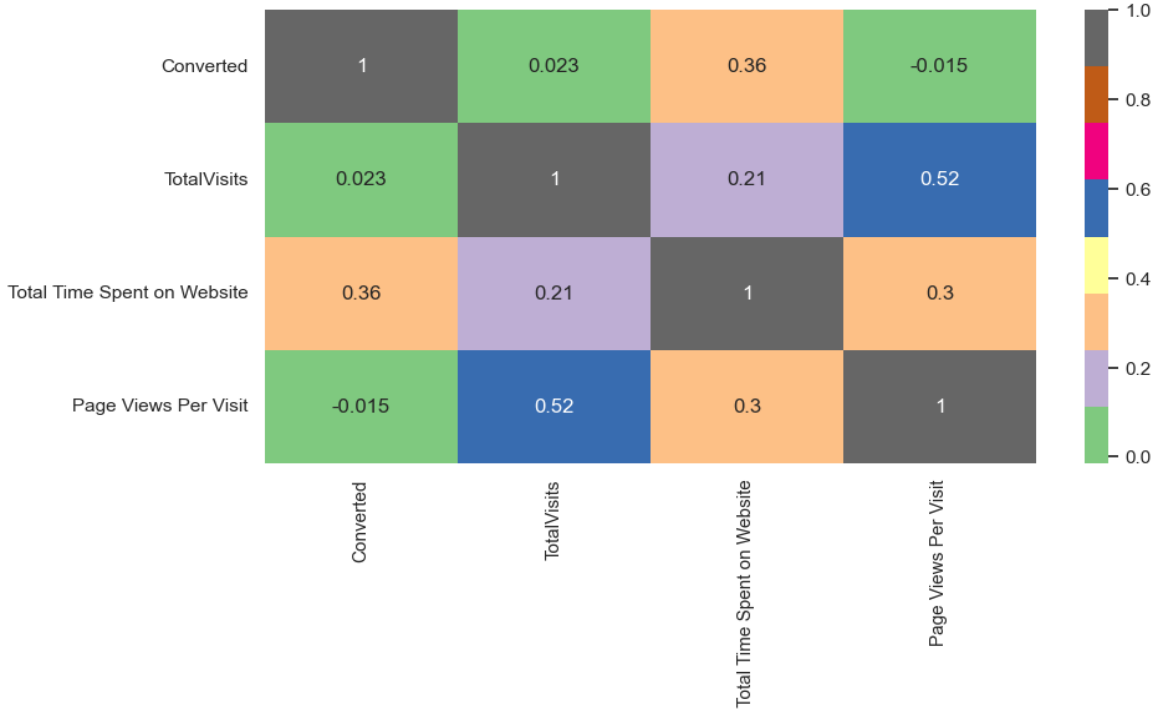
Sending SMS have very good response from people which reflects in the conversion count.
Email opened activity has less but good response from people in conversion count.
Indian people are showing positive response in conversion count compared to out of India countries.





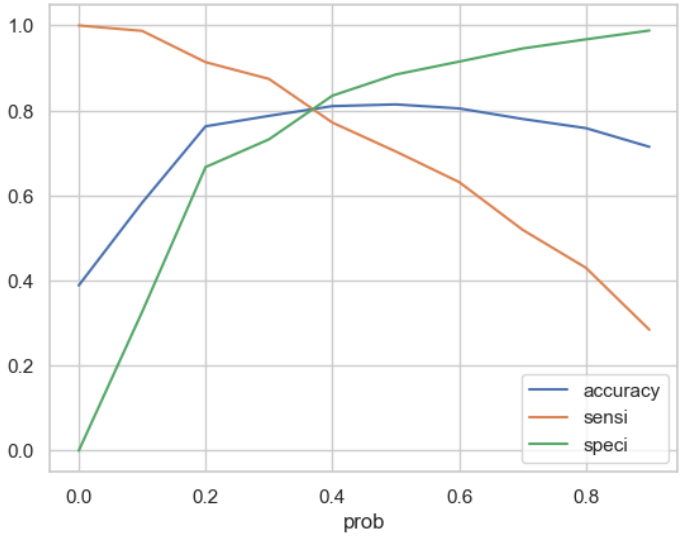
There is 0.36 correlation of "Total Time Spent on Website" with target variable "Converted".

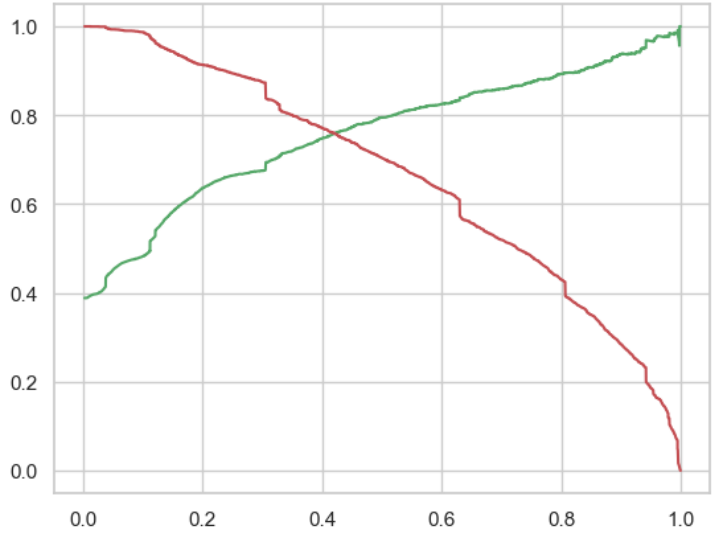
"Page Views Per Visit" have -0.015 correlation with target variable.



Model Evaluation

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





Observation

Train Set:
- Accuracy = 81.03%
- Sensitivity = 77.17%
- Specificity = 83.49%

Test Set:
- Accuracy = 81.02%
- Sensitivity = 76.31%
- Specificity = 83.88%

Conclusion

- We see that the 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 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 are unemployed. Max conversion with working professional.