Lead Scoring Case Study

SUBMITTED BY :
1. Parimal Tapale
2. Rahul Ranjan
3. Shweta Shettigar

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 and showed some interest and filled personal data and that provides the leads.

- The current conversion rate from these leads is 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

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

- The CEO want to achieve a lead conversion rate of 80%.

- The model to be able 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.

Approach

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and p values)
- Model Evaluation
- Making predictions on test

Data cleaning and EDA

- Delete columns which is higher missing values.

- Delete columns which are irrelevant like city, country.

- Take care select level in the data set.

- Leads from HR, Finance & Marketing management specializations are high probability to convert
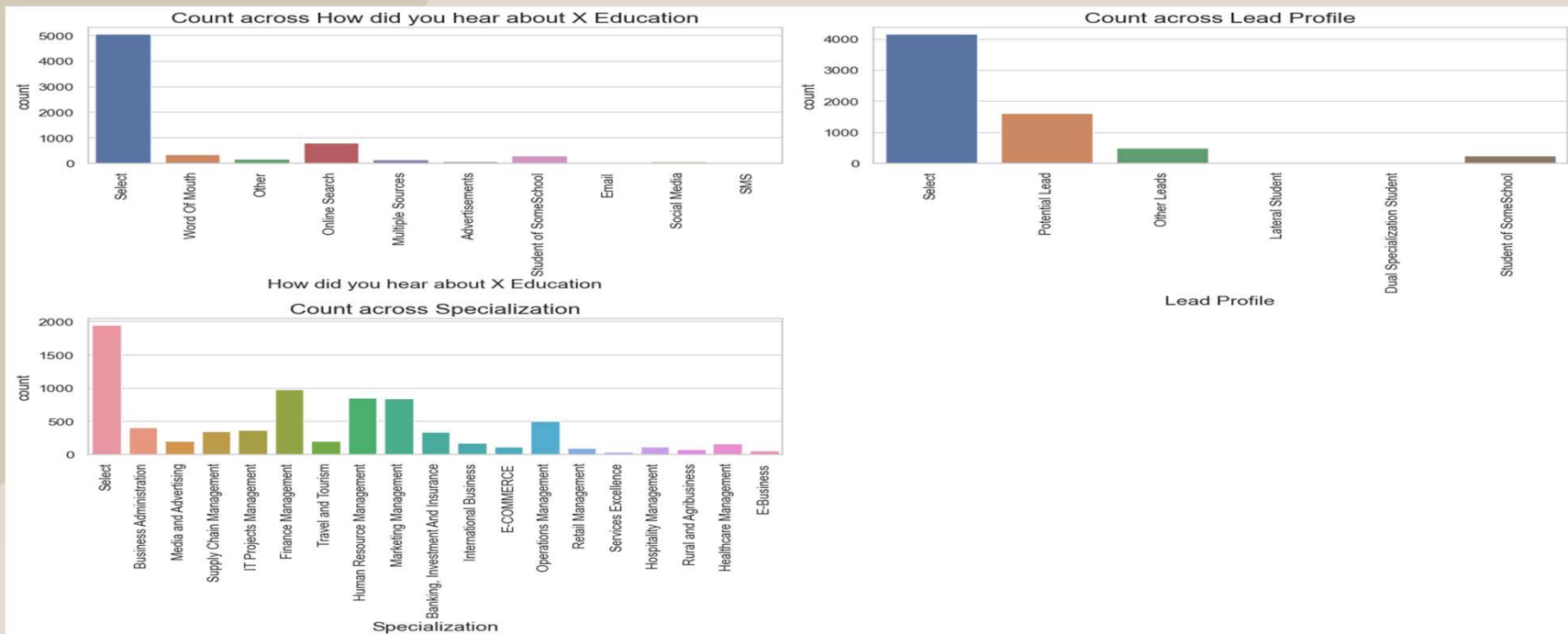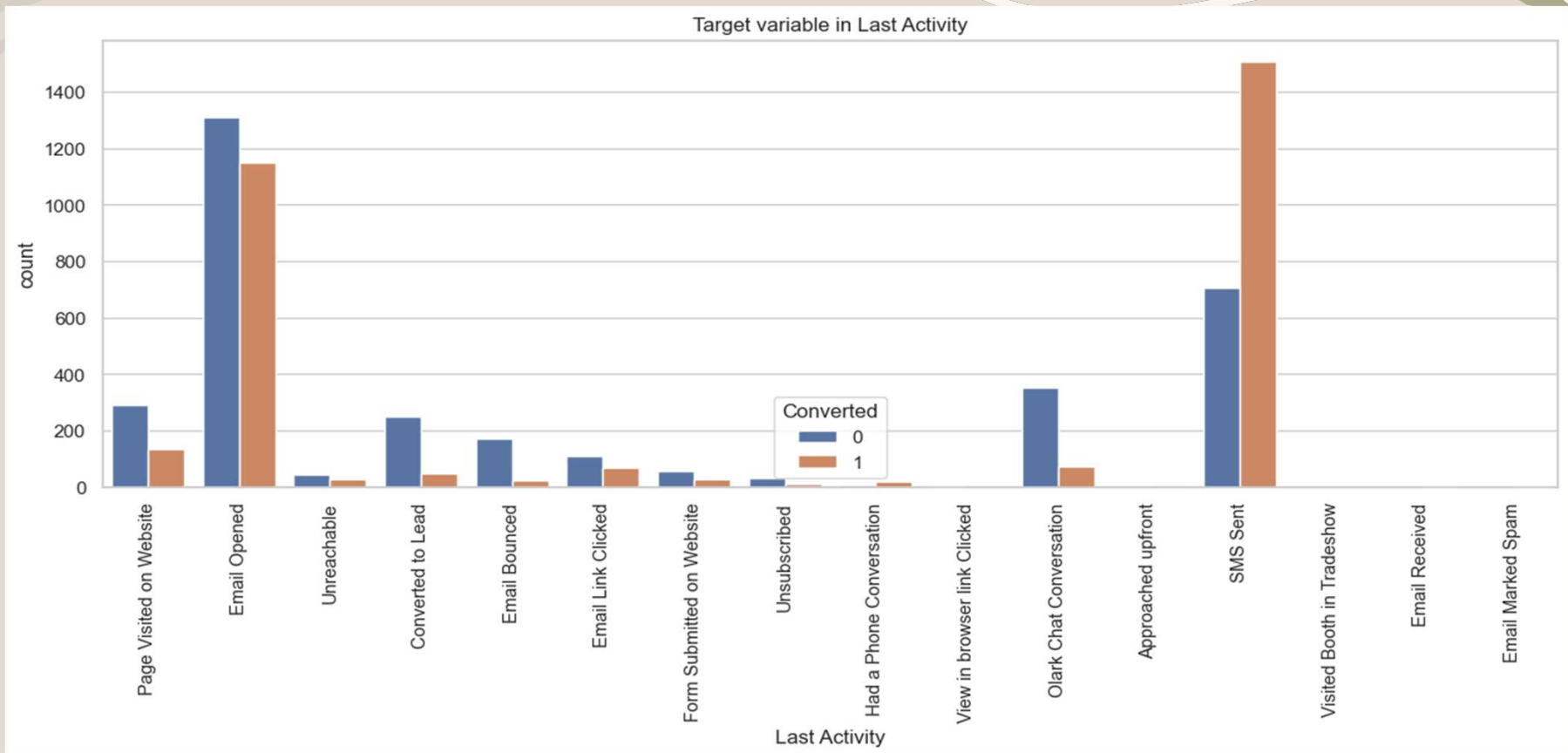
speaking impact

o the leads through google & direct traffic high probability to convert
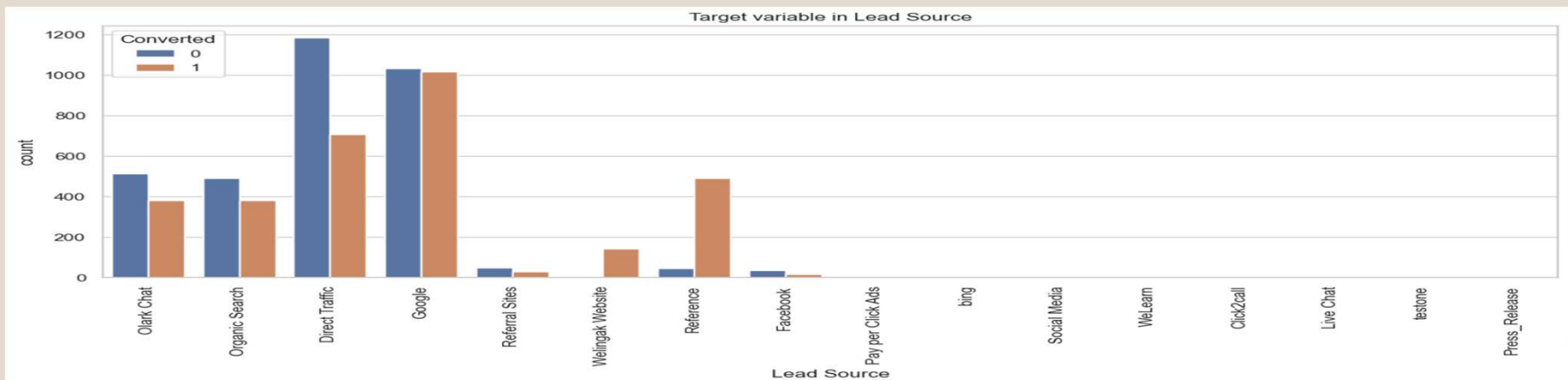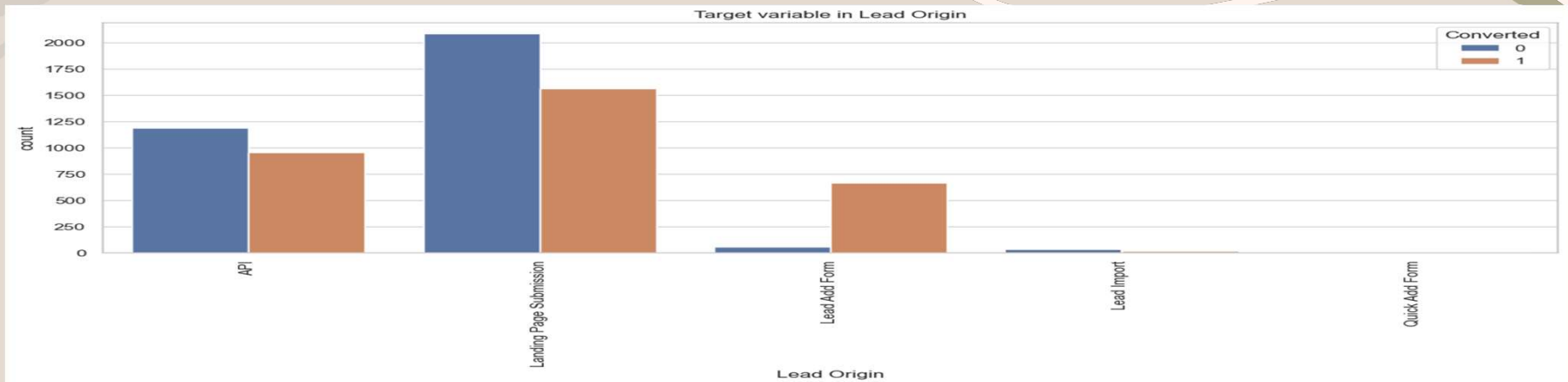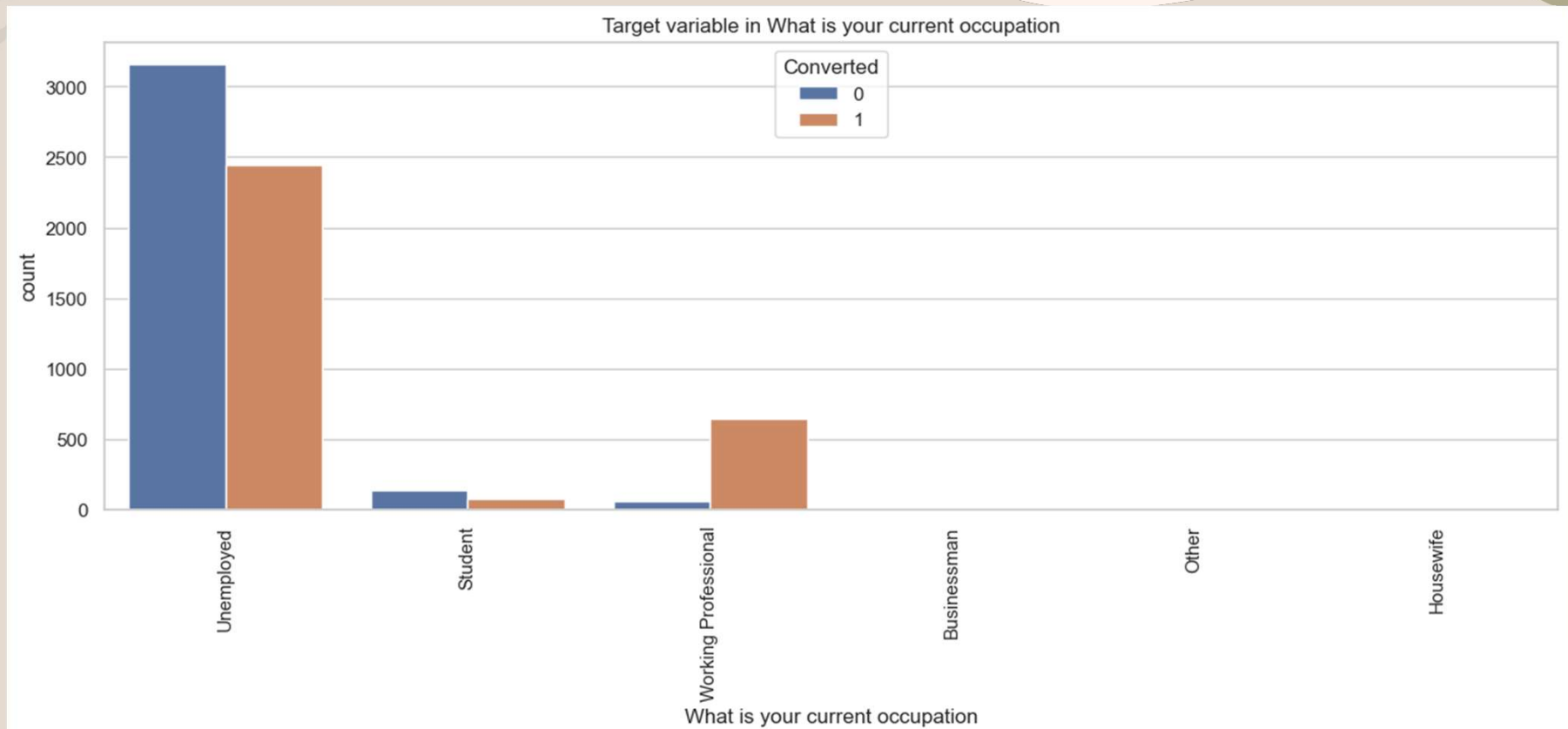
speaking impact

Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.

speaking impact

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

Model Evaluation : ROC

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.
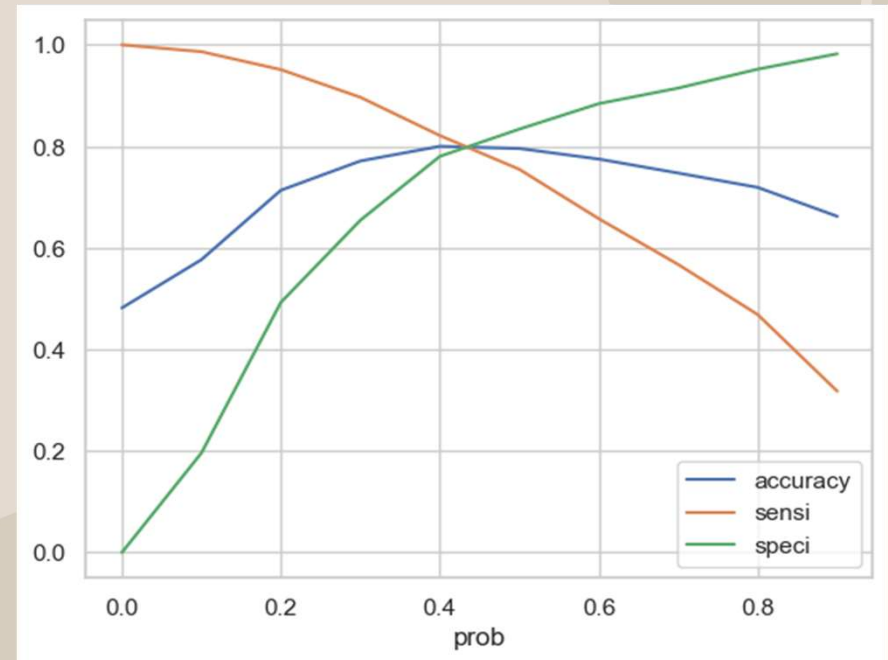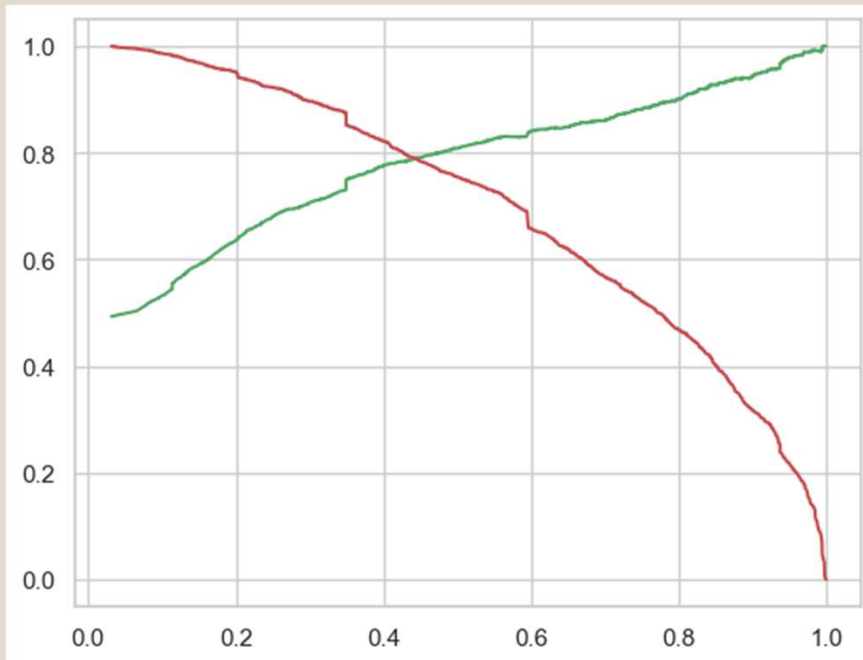
Observations

- Train Data:
- Accuracy : 80%
- Sensitivity : 77%
- Specificity : 80%

- Test Data:
- Accuracy : 80%
- Sensitivity : 77%
- Specificity : 80%

- Final Features list:
- Lead Source_Olark Chat
- Specialization_Others
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission

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.

- 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.

- We see max number of leads are generated by google. Max conversion ratio is by reference and welingak Website.