Lead Scoring Case Study

By,

Rajesh Thote

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.

- The company markets its courses on several websites and search engines like Google. 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. Moreover, the company also gets leads through past referrals. 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%.

Business Objective

- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

- The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.
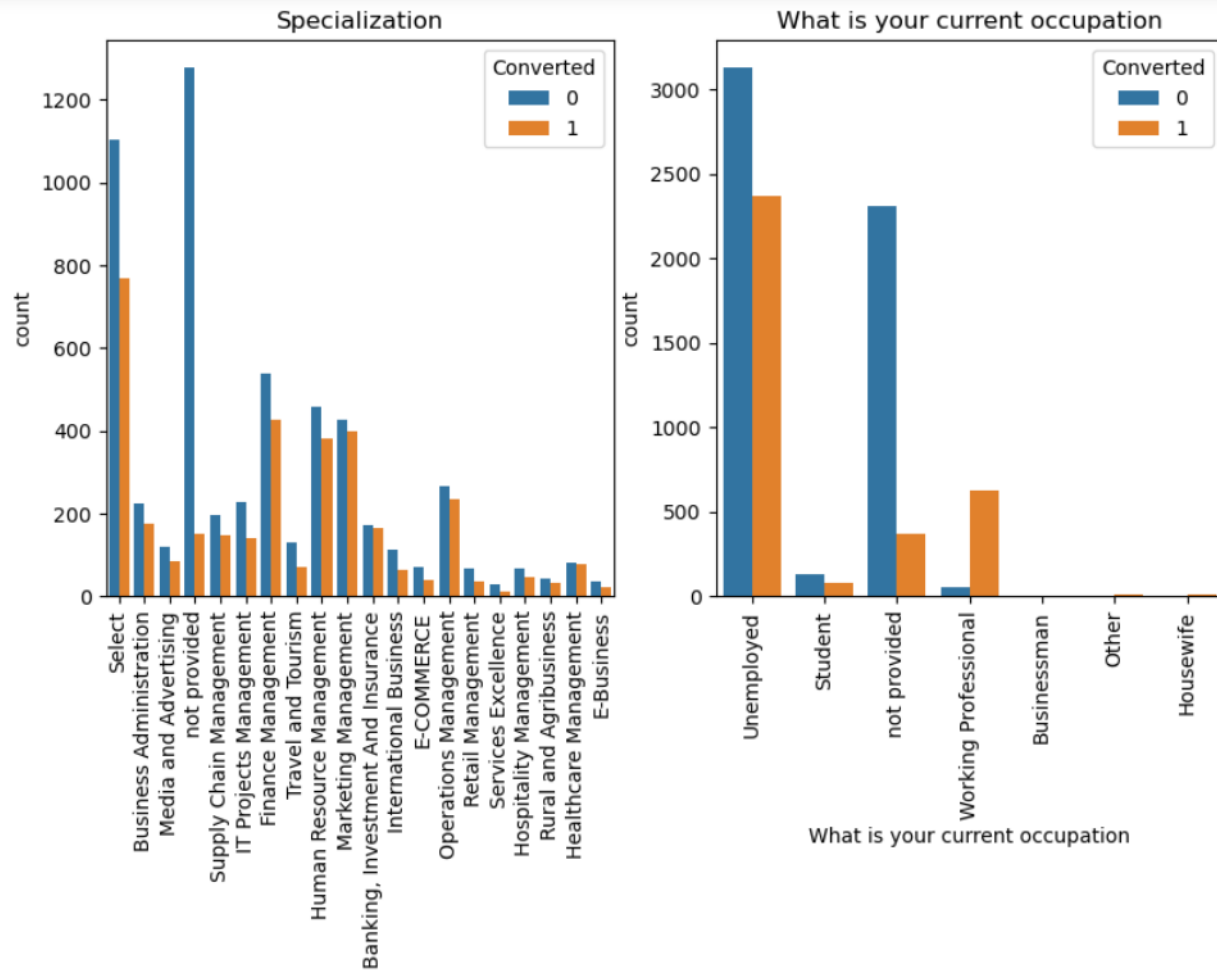
Methodology

- Import the data and inspect the Data frame

- Data preparation

- EDA

- Dummy variable

- Test-Train split

- Model Building

- Creating Prediction

- Model Evaluation

- Optimise Cut off (ROC Curve)

- Prediction on Test set

- Precision-Recall

- Precision-Recall

- Conclusion

EDA-Categorical Variable Relation
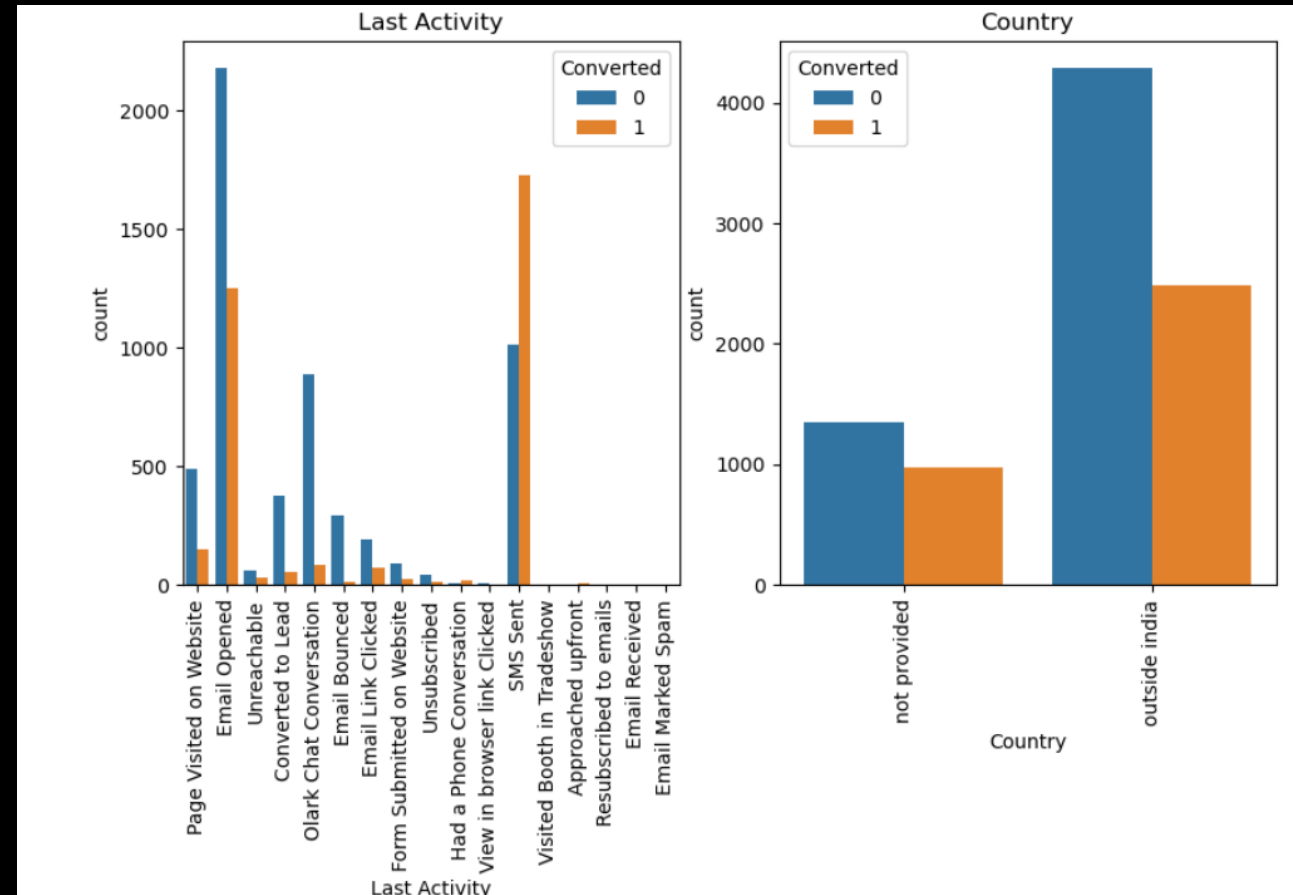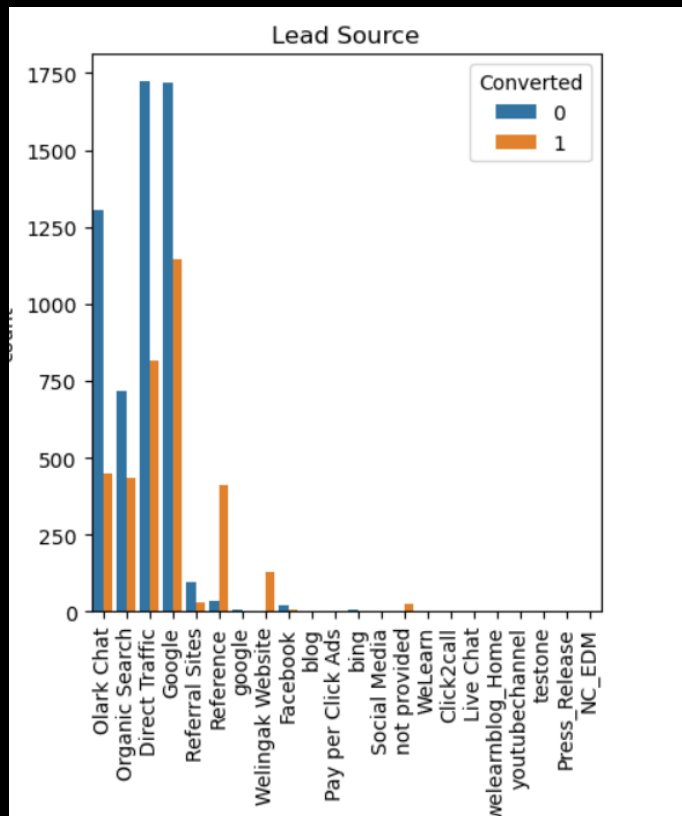
EDA-Categorical Variable Relation

Model Building and Evaluation

- Final Model

	coef	std err	z	P>|z|	[0.025	0.975]
const	-2.2726	0.099	-22.851	0.000	-2.468	-2.078
TotalVisits	6.3855	2.176	2.934	0.003	2.120	10.650
Total Time Spent on Website	4.5897	0.165	27.821	0.000	4.266	4.913
Lead Origin_Lead Add Form	2.6538	0.228	11.662	0.000	2.208	3.100
Lead Source_Welingak Website	2.4092	1.034	2.329	0.020	0.382	4.436
Last Activity_Email Bounced	-1.6881	0.320	-5.281	0.000	-2.315	-1.062
Last Activity_SMS Sent	1.2767	0.072	17.668	0.000	1.135	1.418
Country_outside india	-1.2549	0.107	-11.731	0.000	-1.465	-1.045
What is your current occupation_Other	2.4457	0.752	3.252	0.001	0.972	3.920
What is your current occupation_Student	1.2150	0.222	5.470	0.000	0.780	1.650
What is your current occupation_Unemployed	1.2213	0.084	14.471	0.000	1.056	1.387
What is your current occupation_Working Professional	3.7615	0.200	18.819	0.000	3.370	4.153
Last Notable Activity_Olark Chat Conversation	-1.0813	0.345	-3.132	0.002	-1.758	-0.405
Last Notable Activity_Unreachable	1.9232	0.515	3.733	0.000	0.913	2.933

ROC Curve

- 0.35 is the tradeoff between Precision and Recall -

- Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than
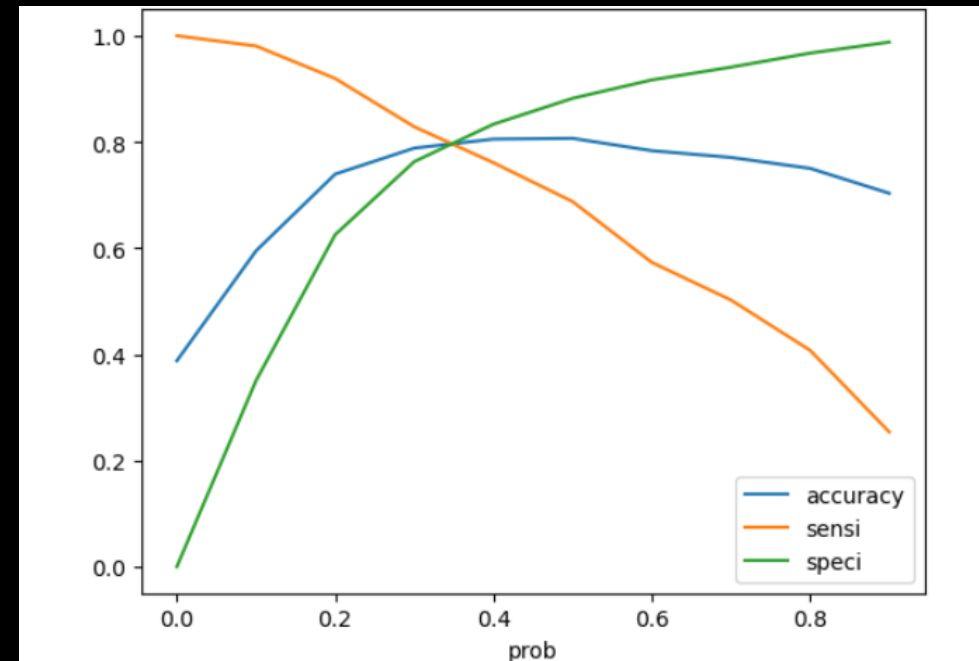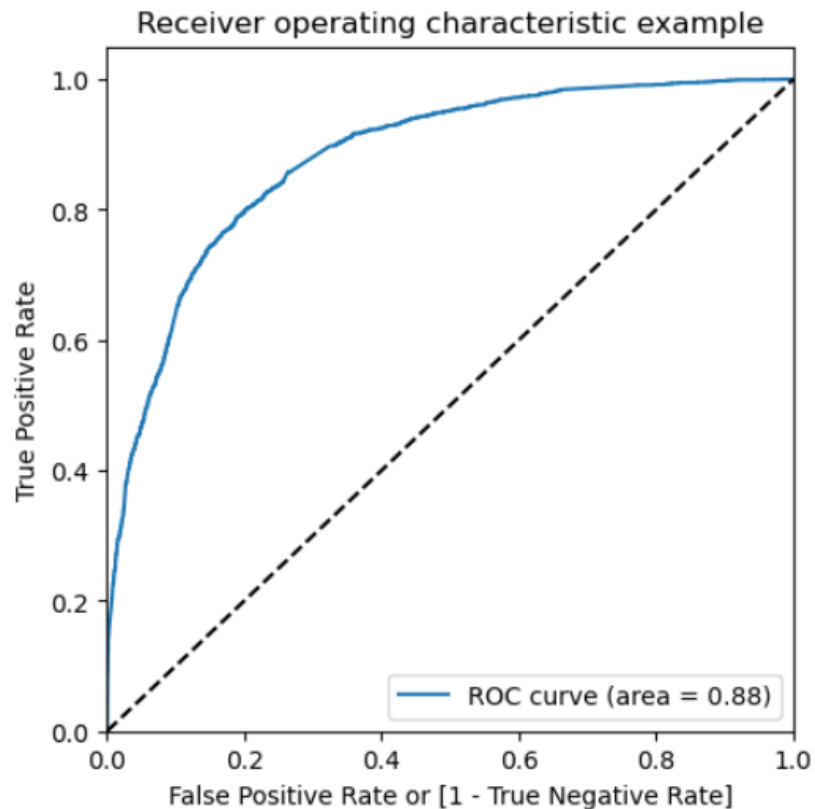
From the curve above, 0.35 is the optimum point to take it as a cutoff probability.

Observations

- Train Data:

- Accuracy : 79.8%

- Sensitivity : 79.7%

- Specificity : 79.9%

- Test Data:

- Accuracy : 81%

- Sensitivity : 79.9%

- Specificity : 81.6%

Conclusion

• Below pointers are major in converting so we can target these customers.

+ The total number of visits made by the customer on the website.

+ The total time spent by the customer on the website.

+ Source of lead as Lead Add Form

+ Major source for lead as Welingak Website

+ Occupation as Working Professional

• Below pointers are not easy to convert hence we can drop these from our conversion list

+ Those where Email bounced

+ Country other than India

+ Olark Chat conversation

+ With these variables X Education can have high chances of conversion with correct target customers.