Lead Scoring Case Study

Submitted by: Akshit Pratiush
apratiush@gmail.com

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

Objective

Lead 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.

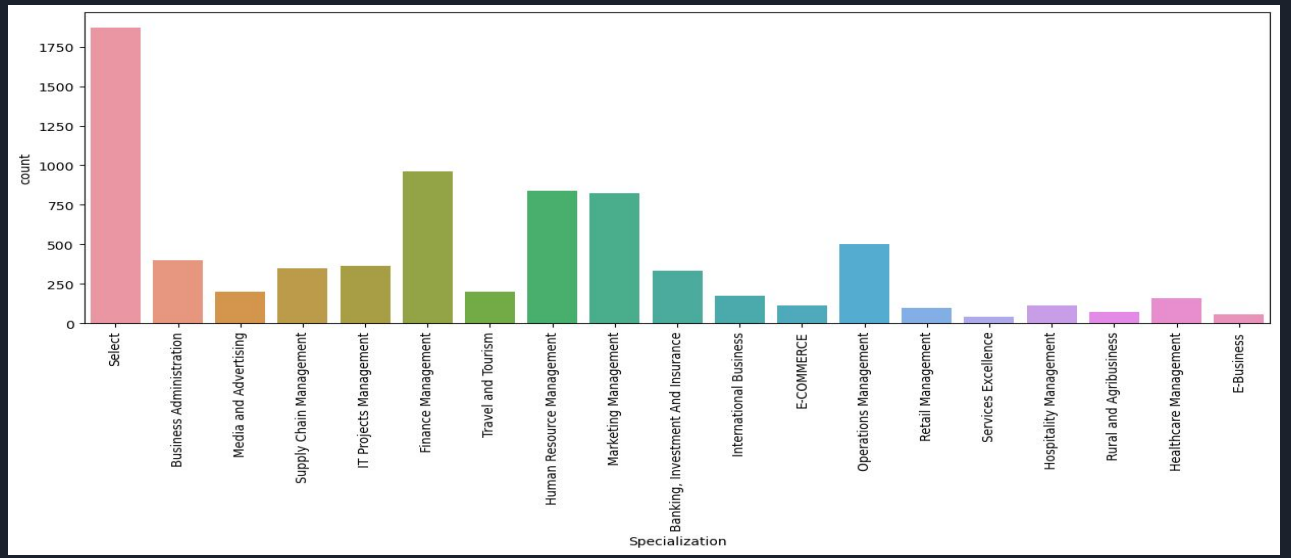
The CEO want to achieve a lead conversion rate of 80%.

They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full manpower and after achieving target what should be the approaches.

Step 1 - Data Cleaning

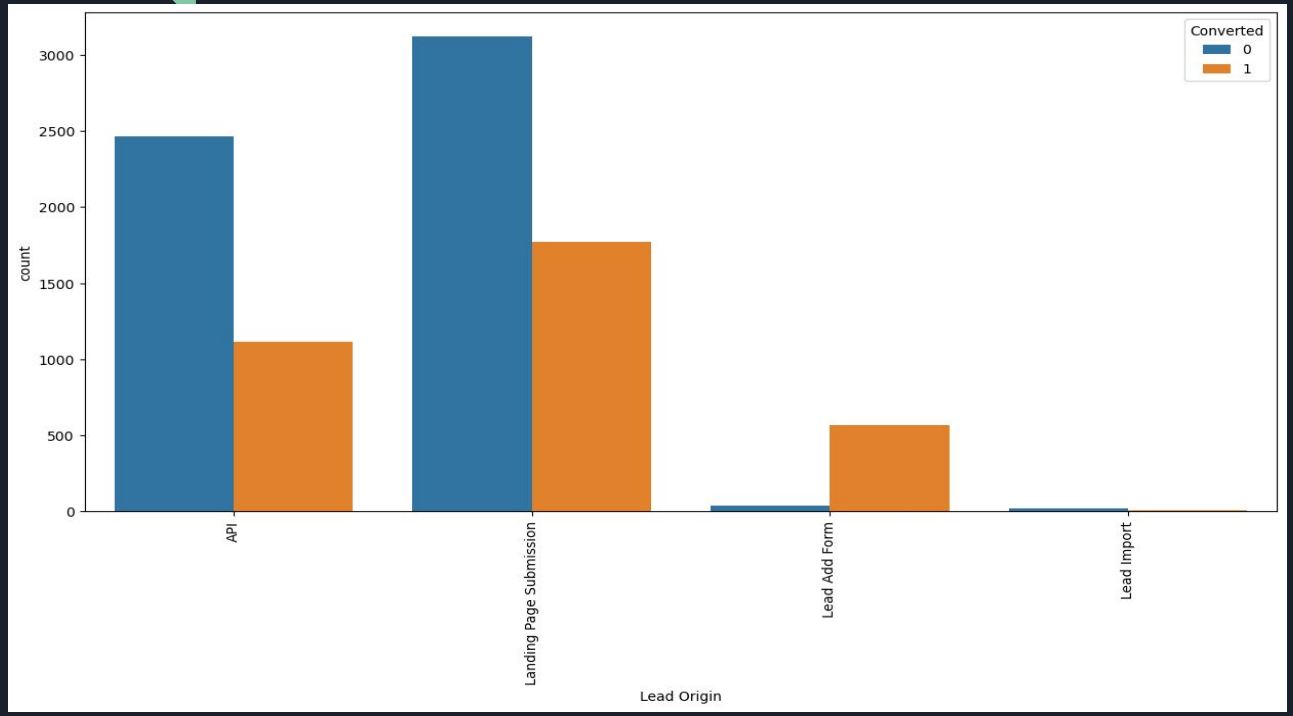
- Removed columns with more than 40% missing values
- Deleting irrelevant columns
- Columns with value “Select” is taking care of

Asymmetrique Activity Index	45.65
Asymmetrique Profile Index	45.65
Asymmetrique Activity Score	45.65
Asymmetrique Profile Score	45.65



Step 2: Data Analysis

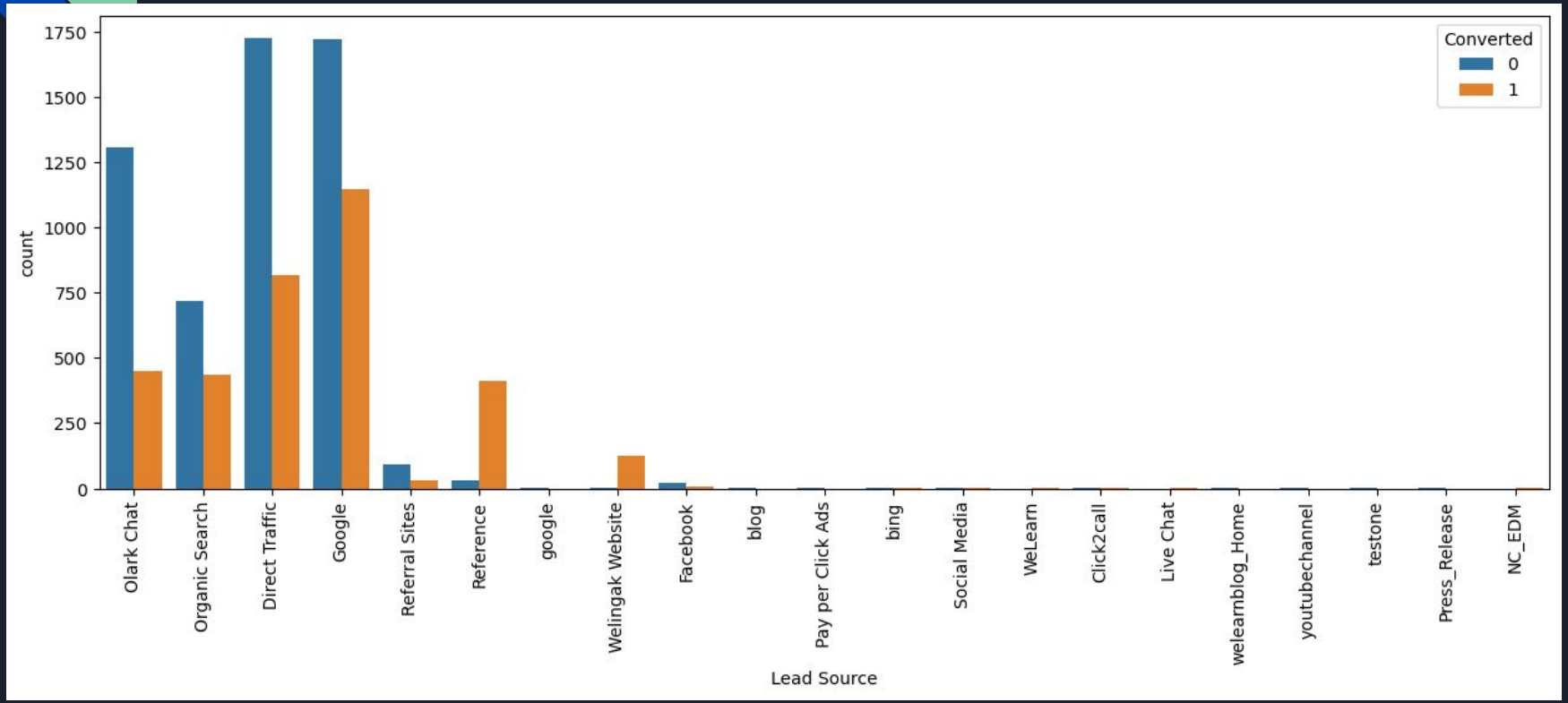
Lead Origin with “Converted”



Findings:
Conversion rate for 'API' and Landing Page Submission is above 30%
For 'Lead Add Form' number of conversion is more than unsuccessful conversion.
Count of 'Lead Import' is lesser.

Analysis:
To improve overall lead conversion rate, we need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form

Lead Source with “Converted”



Logistic regression Model

Generalized Linear Model Regression Results

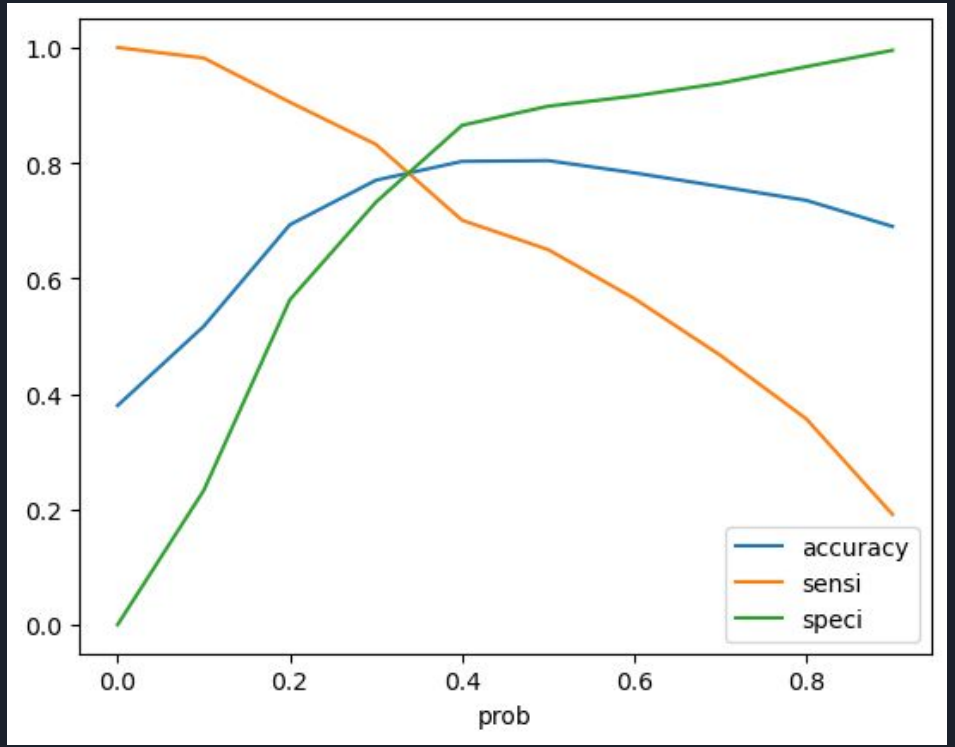
Dep. Variable:	Converted	No. Observations:	6372
Model:	GLM	Df Residuals:	6356
Model Family:	Binomial	Df Model:	15
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2859.9
Date:	Sun, 17 Dec 2023	Deviance:	5719.8
Time:	09:34:08	Pearson chi2:	6.38e+03
No. Iterations:	21	Pseudo R-squ. (CS):	0.3496
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.1812	0.084	2.166	0.030	0.017	0.345
Total Time Spent on Website	1.1070	0.039	28.690	0.000	1.031	1.183
Lead Origin_Lead Add Form	2.8832	0.225	12.838	0.000	2.443	3.323
Lead Source_Direct Traffic	-1.0860	0.108	-10.053	0.000	-1.298	-0.874
Lead Source_Google	-0.8079	0.103	-7.827	0.000	-1.010	-0.606
Lead Source_Organic Search	-0.9018	0.124	-7.256	0.000	-1.145	-0.658
Lead Source_Referral Sites	-1.3907	0.336	-4.137	0.000	-2.050	-0.732
Lead Source_Welingak Website	2.0007	0.749	2.671	0.008	0.532	3.469
Lead Source_bing	-21.6941	2.38e+04	-0.001	0.999	-4.67e+04	4.67e+04
Lead Source_google	-22.0953	2.56e+04	-0.001	0.999	-5.01e+04	5.01e+04
Do Not Email_Yes	-1.3274	0.159	-8.371	0.000	-1.638	-1.017
What is your current occupation_Housewife	22.5838	1.6e+04	0.001	0.999	-3.13e+04	3.14e+04
What is your current occupation_Other	-1.2872	0.083	-15.572	0.000	-1.449	-1.125
What is your current occupation_Working Professional	2.5064	0.178	14.093	0.000	2.158	2.855
Specialization_E-Business	0.6002	0.389	1.545	0.122	-0.161	1.362
Specialization_Services Excellence	0.8241	0.551	1.495	0.135	-0.256	1.904

Evaluation Matrix

At point 0.32 where the three parameters cut each other, we can see that there is a balance between sensitivity and specificity with a good accuracy.

So, 0.32 may be the optimal cutoff.



Business recommendation

Highly likely to be converted:

- Lead score more than 32
- Average lead Score is 73

Other Findings:

- 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