Lead Scoring Case Study Subjective Questions

Soumyajit Mitra, Kalaivannan S and Sonal Verma

1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Ans - Three most important variables were ‘Total Time Spent on Website’,’ Last Notable Activity Having A Phone Conversation’ (variable – Last Notable Activity) and ‘Lead Origin Lead Add Form’ (variable – Lead Origin)

Dep. Variable:	Converted	No. Observations:	6463
Model:	GLM	Df Residuals:	6446
Model Family:	Binomial	Df Model:	16
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2598.7
Date:	Tue, 18 Jul 2023	Deviance:	5197.5
Time:	12:57:34	Pearson chi2:	6.68e+03
No. Iterations:	7	Pseudo R-squ. (CS):	0.4110
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.6083	0.117	-13.708	0.000	-1.838	-1.378
Do Not Email	-1.4622	0.170	-8.588	0.000	-1.796	-1.128
TotalVisits	2.4906	0.532	4.678	0.000	1.447	3.534
Total Time Spent on Website	4.5407	0.167	27.176	0.000	4.213	4.868
Page Views Per Visit	-1.4207	0.394	-3.603	0.000	-2.194	-0.648
LeadOrigin_Landing Page Submission	-0.3548	0.091	-3.896	0.000	-0.533	-0.176
LeadOrigin_Lead Add Form	3.1102	0.206	15.079	0.000	2.706	3.514
Lead_Source_Olark Chat	0.9446	0.133	7.084	0.000	0.683	1.206
Lead_Source_Welingak Website	2.6321	1.024	2.570	0.010	0.625	4.640
Last_Activity_Converted to Lead	-0.6339	0.224	-2.834	0.005	-1.072	-0.195
Last_Activity_Olark Chat Conversation	-1.0119	0.171	-5.909	0.000	-1.347	-0.676
Last_Activity_SMS Sent	1.1995	0.076	15.739	0.000	1.050	1.349
Curr_occ_Occupation NA	-1.1921	0.089	-13.469	0.000	-1.366	-1.019
Curr_occ_Working Professional	2.4654	0.182	13.523	0.000	2.108	2.823
Last_N_Activity_Had a Phone Conversation	3.2288	1.155	2.795	0.005	0.965	5.493
Last_N_Activity_Modified	-0.6478	0.085	-7.597	0.000	-0.815	-0.481
Last_N_Activity_Unreachable	1.9445	0.577	3.371	0.001	0.814	3.075

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

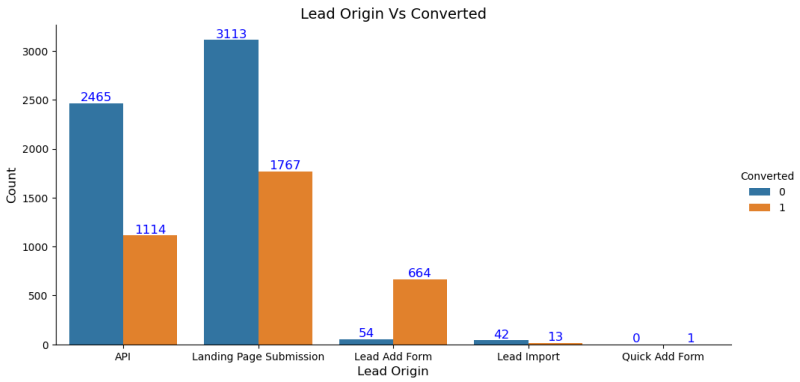
Ans - Three most important dummy variables were Last Notable Activity Having A Phone Conversation’ (variable – Last Notable Activity), ‘Lead Origin Lead Add Form’ (variable – Lead Origin) and ‘Lead Source_Welingak Website’ (variable – Lead Source)

Dep. Variable:	Converted	No. Observations:	6463
Model:	GLM	Df Residuals:	6446
Model Family:	Binomial	Df Model:	16
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2598.7
Date:	Tue, 18 Jul 2023	Deviance:	5197.5
Time:	12:57:34	Pearson chi2:	6.68e+03
No. Iterations:	7	Pseudo R-squ. (CS):	0.4110
Covariance Type:	nonrobust		

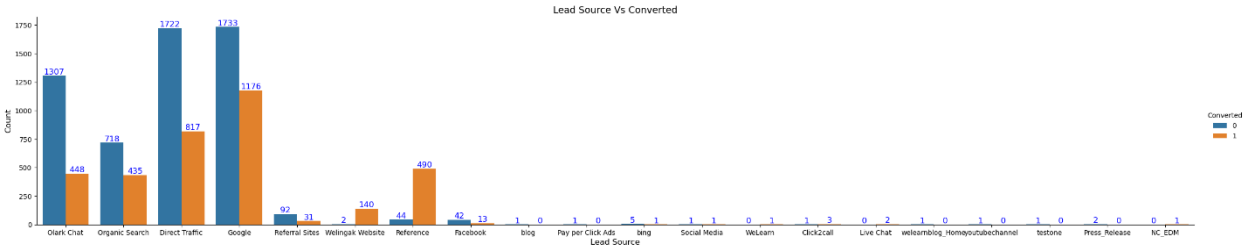
	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.6083	0.117	-13.708	0.000	-1.838	-1.378
Do Not Email	-1.4622	0.170	-8.588	0.000	-1.796	-1.128
TotalVisits	2.4906	0.532	4.678	0.000	1.447	3.534
Total Time Spent on Website	4.5407	0.167	27.176	0.000	4.213	4.868
Page Views Per Visit	-1.4207	0.394	-3.603	0.000	-2.194	-0.648
LeadOrigin_Landing Page Submission	-0.3548	0.091	-3.896	0.000	-0.533	-0.176
LeadOrigin_Lead Add Form	3.1102	0.206	15.079	0.000	2.706	3.514
Lead_Source_Olark Chat	0.9446	0.133	7.084	0.000	0.683	1.206
Lead_Source_Welingak Website	2.6321	1.024	2.570	0.010	0.625	4.640
Last_Activity_Converted to Lead	-0.6339	0.224	-2.834	0.005	-1.072	-0.195
Last_Activity_Olark Chat Conversation	-1.0119	0.171	-5.909	0.000	-1.347	-0.676
Last_Activity_SMS Sent	1.1995	0.076	15.739	0.000	1.050	1.349
Curr_occ_Occupation NA	-1.1921	0.089	-13.469	0.000	-1.366	-1.019
Curr_occ_Working Professional	2.4654	0.182	13.523	0.000	2.108	2.823
Last_N_Activity_Had a Phone Conversation	3.2288	1.155	2.795	0.005	0.965	5.493
Last_N_Activity_Modified	-0.6478	0.085	-7.597	0.000	-0.815	-0.481
Last_N_Activity_Unreachable	1.9445	0.577	3.371	0.001	0.814	3.075

3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

Ans – They should focus more on candidates belonging to categories which historically show more conversion rate. For example,



In Lead Origin, call almost all customers from Lead Add Form, Lead Import, Landing Page Submission



In lead source, call almost all customers from Reference, Welingak website, Organic Search and Referral Sites.

Like the above two examples, considering the results from the exploratory data analysis and the finalized variables post data cleaning, the categories with maximum rate of conversion as well as maximum coefficients in the logistic regression models, all candidates can be approached.

Another strategy might be to lower the lead score/probability cutoff values and call all hot leads obtained as a result, for that they will have to lower their precision value requirement

4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Ans – The simplest approach would be to increase the cut-off, giving more priority to precision and Positive Prediction Value, which will lead to employees only reaching out to candidates that have high probability of conversion