SUBJECTIVE QUESTIONS - Answered

BY: *Shayak Majumder | Shreya Dutta | Shweta*

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

Screenshot of the final model for quick reference:

Generalized Linear Model Regression Results

Dep. Variable:	y	**No. Observations:**	6298
Model:	GLM	**Df Residuals:**	6284
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2815.3
Date:	Sat, 18 Jan 2025	**Deviance:**	5630.5
Time:	09:34:38	**Pearson chi2:**	6.73e+03
No. Iterations:	6	**Pseudo R-squ. (CS):**	0.3536
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-3.4097	0.135	-25.166	0.000	-3.675	-3.144
Do Not Email	-1.1678	0.154	-7.565	0.000	-1.470	-0.865
TotalVisits	0.8645	0.241	3.581	0.000	0.391	1.338
Total Time Spent on Website	3.8051	0.137	27.863	0.000	3.537	4.073
Lead Origin_Landing Page Submission	-0.3381	0.109	-3.089	0.002	-0.553	-0.124
Lead Origin_Other	3.5027	0.180	19.428	0.000	3.149	3.856
Lead Source_Google	0.2416	0.079	3.067	0.002	0.087	0.396
Lead Source_Olark Chat	1.2019	0.135	8.896	0.000	0.937	1.467
Specialization_Management	0.3793	0.099	3.846	0.000	0.186	0.573
Specialization_Other Domain	0.4641	0.121	3.847	0.000	0.228	0.701
What is your current occupation_Other	1.7291	0.483	3.580	0.000	0.783	2.676
What is your current occupation_Student	1.0709	0.231	4.645	0.000	0.619	1.523
What is your current occupation_Unemployed	1.3479	0.086	15.709	0.000	1.180	1.516
What is your current occupation_Working Professional	3.6556	0.187	19.499	0.000	3.288	4.023

Our final ML model, logmod8, shows that the top 3 variables with the highest coefficients are: **Total Time Spent on Website, What is your current occupation_Working Professional,** and **Lead Origin_Other.**

These also align with our initial observations from the EDA steps.

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?**

   Based on the coefficients, the top 3 categorical variables are: **What is your current occupation_Working Professional, Lead Origin_Other,** and **What is your current occupation_Other.**

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.**

   More interns mean more hands on deck. So, what we can do is reduce the threshold a little bit further to increase the leads predicted as 1. What this will do is increase the pool of high-priority leads for X Education.

   Also, X Education should put the interns on more efficient lead conversion. This means that they can aggressively follow up leads with the top 3 coefficients identified in Question 1.

   Matter of fact, these leads should be addressed with multiple calls in a day (twice in 24 hours), depending on the successful response on the first attempt. With more hands due to interns, this will be possible to attempt. As standard, leads who show initial interest but are hesitant at first stand more chance to be converted.

   Of course, feedback from interns will be valuable in understanding and rethinking our approach.

   Lastly, since we saw users who spend more time on the website and those who are working professionals stand more chances to be converted, X Education can start showing more offers/benefits on the website and also tailor the content to show more benefits to working professionals as well.

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.**

We can start by taking an opposite approach to Question 3, and instead of reducing the threshold, we can increase the threshold to somewhere around 0.7 or 0.8, so that the '1'

predictions that we get stand a high chance of conversion with minimal effort and lesser followups. Predictors with high coefficients, of course, will remain top priority still.

Speaking of follow-ups, X Education can also rely more on automated processes in place for follow-ups, like emails or WhatsApp/SMS messaging. This would free up manual hands in the team to focus on other work.

Of course, monitoring conversions will help in understanding which strategies are paying better dividends.

~END~