Lead Scoring Case Study

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

Here is the summary of the final model:

Dep. Variable:	Converted	**No. Observations:**	4461
Model:	GLM	**Df Residuals:**	4449
Model Family:	Binomial	**Df Model:**	11
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2079.1
Date:	Mon, 02 Jan 2023	**Deviance:**	4158.1
Time:	02:21:10	**Pearson chi2:**	4.80e+03
No. Iterations:	7	**Pseudo R-squ. (CS):**	0.3642
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2040	0.196	1.043	0.297	-0.179	0.587
TotalVisits	11.1489	2.665	4.184	0.000	5.926	16.371
Total Time Spent on Website	4.4223	0.185	23.899	0.000	4.060	4.785
Lead Origin_Lead Add Form	4.2051	0.258	16.275	0.000	3.699	4.712
Lead Source_Olark Chat	1.4526	0.122	11.934	0.000	1.214	1.691
Lead Source_Welingak Website	2.1526	1.037	2.076	0.038	0.121	4.185
Do Not Email_Yes	-1.5037	0.193	-7.774	0.000	-1.883	-1.125
Last Activity_Had a Phone Conversation	2.7552	0.802	3.438	0.001	1.184	4.326
Last Activity_SMS Sent	1.1856	0.082	14.421	0.000	1.024	1.347
What is your current occupation_Student	-2.3578	0.281	-8.392	0.000	-2.908	-1.807
What is your current occupation_Unemployed	-2.5445	0.186	-13.699	0.000	-2.908	-2.180
Last Notable Activity_Unreachable	2.7846	0.807	3.449	0.001	1.202	4.367

From the summary, we can see that the top three variables which contribute most towards the probability of a lead getting converted are:

1. TotalVisits → This variable has the coefficient with highest magnitude (11.14) indicating its high impact on the target variable, “Converted”. The positive value of the coefficient indicates a positive relationship, i.e., higher the number of total visits, higher is the chance of the lead getting converted.

2. Total Time Spent on Website → has the coefficient with second highest magnitude (4.42). This will also contribute highly towards the probability of a lead getting converted. The positive value of the coefficient indicates a positive relationship, i.e., higher the Total Time Spent on Website, higher is the chance of the lead getting converted.

3. Lead Origin_Lead Add Form → also contribute very well towards the probability of a lead getting converted with coefficient of 4.2. The positive value of the coefficient indicates a positive relationship, i.e., if the customer was identified to be a lead by means of “Lead Add Form” higher is the chance of the lead getting converted.

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?**

Dep. Variable:	Converted	**No. Observations:**	4461
Model:	GLM	**Df Residuals:**	4449
Model Family:	Binomial	**Df Model:**	11
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2079.1
Date:	Mon, 02 Jan 2023	**Deviance:**	4158.1
Time:	02:21:10	**Pearson chi2:**	4.80e+03
No. Iterations:	7	**Pseudo R-squ. (CS):**	0.3642
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2040	0.196	1.043	0.297	-0.179	0.587
TotalVisits	11.1489	2.665	4.184	0.000	5.926	16.371
Total Time Spent on Website	4.4223	0.185	23.899	0.000	4.060	4.785
Lead Origin_Lead Add Form	4.2051	0.258	16.275	0.000	3.699	4.712
Lead Source_Olark Chat	1.4526	0.122	11.934	0.000	1.214	1.691
Lead Source_Welingak Website	2.1526	1.037	2.076	0.038	0.121	4.185
Do Not Email_Yes	-1.5037	0.193	-7.774	0.000	-1.883	-1.125
Last Activity_Had a Phone Conversation	2.7552	0.802	3.438	0.001	1.184	4.326
Last Activity_SMS Sent	1.1856	0.082	14.421	0.000	1.024	1.347
What is your current occupation_Student	-2.3578	0.281	-8.392	0.000	-2.908	-1.807
What is your current occupation_Unemployed	-2.5445	0.186	-13.699	0.000	-2.908	-2.180
Last Notable Activity_Unreachable	2.7846	0.807	3.449	0.001	1.202	4.367

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 from final model summary are:

1. **Lead Origin_Lead Add Form** → coefficient = 4.2051
   Lead Origin_Lead Add Form seems to have a good positive impact on probability of lead conversion. It means that if a customer is marked as lead by virtue of adding form, higher is the probability of them being a hot lead.

2. **Last Notable Activity_Unreachable** → coefficient = 2.7846
   From the model we could see that if the last notable activity is Unreachable, there is a higher chance of that customer getting converted to lead. But this couldn't be considered as a possible factor that could be practically focused.

3. **Last Activity_Had a Phone Conversation** → coefficient = 2.7552
   If that last activity had with the customer was a phone conversation, there is a higher chance of the lead getting converted. This could be taken as a positive approach and increase the use of method of contact as phone conversation rather than chat or email

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.**

   A good strategy would be to focus on those customers who had the highest number of visits to the website. Out of such a list if we could focus on those customers who has spent maximum amount of time on website. There is a very high chance of these customers getting converted.

   Focus on those customers who found about the program from Welingak Website or Olark Chat and who were marked as lead by adding form. Also, it is important to make sure that making phone calls have better impact than sending sms.

   It would be better to give lesser priority to Students and those who are unemployed, since their chance of conversion is very less. If a customer has marked Do Not Email as Yes, then it is better to not prioritize such customers as well. Focus first on employed customers who has given approval to send emails and then move on to those customers with lesser chance of conversion.

   Arranging the customers in higher probability order would help reach out to maximum prospective customers and increase the conversion rate.

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.**

   In this case when the company's aim is to not make phone calls unless it's extremely necessary, it would be better to focus only on those customers who had the highest number of visits to the website with the highest amount of time spend. We could even put a filter and consider only those customers who are employed and can completely avoid students and unemployed customers.

   We could even drop those customers who has marked "Yes" for not receiving emails. Consider those customers who found about the program from Welingak Website and who was marked as lead by adding form.