

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

Below is the screenshot of final model:

Generalized Linear Model Regression Results						
=====						
Dep. Variable:	Converted	No. Observations:	6267			
Model:	GLM	Df Residuals:	6253			
Model Family:	Binomial	Df Model:	13			
Link Function:	Logit	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-1263.3			
Date:	Sun, 22 Jan 2023	Deviance:	2526.6			
Time:	14:14:19	Pearson chi2:	8.51e+03			
No. Iterations:	8	Pseudo R-squ. (CS):	0.6037			
Covariance Type:	nonrobust					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
const	-1.1179	0.084	-13.382	0.000	-1.282	-0.954
Total Time Spent on Website	0.8896	0.053	16.907	0.000	0.786	0.993
Lead Origin_Lead Add Form	1.6630	0.455	3.657	0.000	0.772	2.554
Lead Source_Direct Traffic	-0.8212	0.127	-6.471	0.000	-1.070	-0.572
Lead Source_Welingak Website	3.8845	1.114	3.488	0.000	1.701	6.068
Last Activity_SMS Sent	1.9981	0.113	17.718	0.000	1.777	2.219
Tags_Closed by Horizzon	7.1955	1.020	7.053	0.000	5.196	9.195
Tags_Interested in other courses	-2.1318	0.406	-5.253	0.000	-2.927	-1.336
Tags_Lost to EINS	5.9177	0.611	9.689	0.000	4.721	7.115
Tags_Others	-2.3737	0.206	-11.507	0.000	-2.778	-1.969
Tags_Ringing	-3.4531	0.238	-14.532	0.000	-3.919	-2.987
Tags_Will revert after reading the email	4.5070	0.188	24.002	0.000	4.139	4.875
Last Notable Activity_Modified	-1.6525	0.124	-13.279	0.000	-1.896	-1.409
Last Notable Activity_Olark Chat Conversation	-1.8023	0.491	-3.669	0.000	-2.765	-0.839
=====						

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

### 1. Tags\_Closed by Horizzon → Coefficient – 7.1955

This variable seems to have a high impact on the target variable, “Converted”. The positive value of the coefficient indicates a positive relationship, i.e., higher the number of leads Closed by Horizzon, higher is the chance of the lead getting converted.

### 2. Tags\_Lost to EINS → Coefficient – 5.9177

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 leads Lost to EINS, higher is the chance of the lead getting converted.

### 3. Tags\_Will revert after reading the email → Coefficient – 4.5070

This also contribute very well towards the probability of a lead getting converted. The positive value of the coefficient indicates a positive relationship, i.e., higher the leads having status Will revert after reading the email, higher is the chance of the lead getting converted.

- 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?

Below is the screenshot of final model:

Generalized Linear Model Regression Results						
Dep. Variable:	Converted	No. Observations:	6267			
Model:	GLM	Df Residuals:	6253			
Model Family:	Binomial	Df Model:	13			
Link Function:	Logit	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-1263.3			
Date:	Sun, 22 Jan 2023	Deviance:	2526.6			
Time:	14:14:19	Pearson chi2:	8.51e+03			
No. Iterations:	8	Pseudo R-squ. (CS):	0.6037			
Covariance Type:	nonrobust					
	coef	std err	z	P> z	[0.025	0.975]
const	-1.1179	0.084	-13.382	0.000	-1.282	-0.954
Total Time Spent on Website	0.8896	0.053	16.907	0.000	0.786	0.993
Lead Origin_Lead Add Form	1.6630	0.455	3.657	0.000	0.772	2.554
Lead Source_Direct Traffic	-0.8212	0.127	-6.471	0.000	-1.070	-0.572
Lead Source_Welingak Website	3.8845	1.114	3.488	0.000	1.701	6.068
Last Activity_SMS Sent	1.9981	0.113	17.718	0.000	1.777	2.219
Tags_Closed by Horizzon	7.1955	1.020	7.053	0.000	5.196	9.195
Tags_Interested in other courses	-2.1318	0.406	-5.253	0.000	-2.927	-1.336
Tags_Lost to EINS	5.9177	0.611	9.689	0.000	4.721	7.115
Tags_Others	-2.3737	0.206	-11.507	0.000	-2.778	-1.969
Tags_Ringing	-3.4531	0.238	-14.532	0.000	-3.919	-2.987
Tags_Will revert after reading the email	4.5070	0.188	24.002	0.000	4.139	4.875
Last Notable Activity_Modified	-1.6525	0.124	-13.279	0.000	-1.896	-1.409
Last Notable Activity_Olark Chat Conversation	-1.8023	0.491	-3.669	0.000	-2.765	-0.839

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

### 1. Tags\_Closed by Horizzon → Coefficient – 7.1955

This variable has a high positive impact on probability of lead conversion. It means that if the leads status is Closed by Horizzon, higher is the chance of the lead getting converted.

### 2. Tags\_Lost to EINS → Coefficient – 5.9177

This variable has a high positive impact on probability of lead conversion. It means that if the leads status is Lost to EINS, higher is the chance of the lead getting converted. We cannot do much here and cannot focus much in this area.

### 3. Tags\_Will revert after reading the email → Coefficient – 4.5070

This variable has a high positive impact on probability of lead conversion. It means that if the leads status is Will revert after reading the email, higher is the chance of the lead getting converted. We can focus here and reach out to customers where leads are in such status.

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 below customers:

- Focus on customers as hot leads where lead tags are closed by Horizon as they have high chances of conversion. Look for all the leads which are at status/tag just before the closed by Horizon tag. Reach out to these stakeholders to move leads to closure state for more lead conversion.
- Focus on those customers as hot leads where the tags/status is will revert after reading the email. Regularly connect with them for more lead conversion
- Focus on customers as hot lead who found about the program from Welinkak Website and who were marked as lead by adding form. There is a very high chance of these customers leads getting converted.
- Target customers who are spending more time on our website.
- Give Less priority to customers where tag/status is interested in other courses or are in ringing as these are the least probable to give us a successful lead.
- Give Less priority to people with whom we have last activity as Olark Chat conversation or if lead source is Direct Traffic.

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 following customers:

- Customers who found about the program from Welinkak Website and will respond after reading the email. We can drop email to these customers and even new customers.
- Customers who are spending more time on the website can be approached by emails and SMS.
- We can remove all such customers who are interested in other courses or tagged as ringing.