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

Ans ;

The top three variables in the model which contribute most towards the probability of a lead getting converted are

Tags
Lead Source
Last Activity

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.4592	0.071	-20.542	0.000	-1.598	-1.320
Lead Source_Welingak Website	4.8724	0.753	6.474	0.000	3.397	6.348
Last Activity_SMS Sent	2.1669	0.104	20.863	0.000	1.963	2.370
Tags_Closed by Horizzon	8.2211	1.008	8.154	0.000	6.245	10.197
Tags_Lost to EINS	7.2595	0.792	9.161	0.000	5.706	8.813
Tags_Ringing	-3.2303	0.220	-14.684	0.000	-3.661	-2.799
Tags_Will revert after reading the email	4.9675	0.174	28.544	0.000	4.626	5.309
Tags_invalid number	-3.4864	1.028	-3.391	0.001	-5.502	-1.471
Tags_switched off	-3.6746	0.517	-7.101	0.000	-4.689	-2.660
Lead Quality_Worst	-2.9445	0.548	-5.370	0.000	-4.019	-1.870
Last Notable Activity_Modified	-1.8106	0.114	-15.883	0.000	-2.034	-1.587
Last Notable Activity_Olark Chat Conversation	-1.1420	0.395	-2.888	0.004	-1.917	-0.367

As from the summary report generated from the logistic regression model as shown above, the *Top* most variable from the data set that contributes most towards the probability of a lead getting converted is "*Tags*" which is assigned to customer indicating the current status of the lead.

The *second* most variable from the data set that contributes most towards the probability of a lead getting converted is "*Lead Source*" which indicates the source of the lead(Includes Google, Organic Search, Olark Chat, etc).

The *Third* most variable from the data set that contributes most towards the probability of a lead getting converted is "*Last Activity*" which indicates the Last activity performed by the customer. (Includes Email Opened, Olark Chat Conversation, etc).

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:

The top 3 categorical/dummy variables which contribute most towards the probability of a lead getting converted are shown below as per the report generated from logistic regression model.

Variables	Coeff
Tags_Closed by Horizzon	8.2211
Tags_Lost to EINS	7.2595
Tags_Will revert after reading the email	4.9675

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.4592	0.071	-20.542	0.000	-1.598	-1.320
Lead Source_Welingak Website	4.8724	0.753	6.474	0.000	3.397	6.348
Last Activity_SMS Sent	2.1669	0.104	20.863	0.000	1.963	2.370
Tags_Closed by Horizzon	8.2211	1.008	8.154	0.000	6.245	10.197
Tags_Lost to EINS	7.2595	0.792	9.161	0.000	5.706	8.813
Tags_Ringing	-3.2303	0.220	-14.684	0.000	-3.661	-2.799
Tags_Will revert after reading the email	4.9675	0.174	28.544	0.000	4.626	5.309
Tags_invalid number	-3.4864	1.028	-3.391	0.001	-5.502	-1.471
Tags_switched off	-3.6746	0.517	-7.101	0.000	-4.689	-2.660
Lead Quality_Worst	-2.9445	0.548	-5.370	0.000	-4.019	-1.870
Last Notable Activity_Modified	-1.8106	0.114	-15.883	0.000	-2.034	-1.587
Last Notable Activity_Olark Chat Conversation	-1.1420	0.395	-2.888	0.004	-1.917	-0.367

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:

As we are having 10 interns and we want to wish the lead conversion more aggressive, it is better to decrease the threshold limit for the conv_probablity cutoff value so that we will have more number of potential leads identified in our model and there by all the interns will be assigned and make calls to all the potential leads and make the lead conversion more aggressive.

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 Company already reaches the target for that quarter before the dead line and there by sales team focus on new work, the company don't want to make useless call, i.e. they want to minimize the rate of useless phone calls, In this scenario the company only wants the potential leads that are having the high conversion rate, this means we will increase the threshold limit for the conversion_rate cut off and there by getting less number potential leads getting identified from the model, which sales team needs to be focused on leads who are having the highest probability of conversion.