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

   According to the model derived the top 3 variables contributed the probability of the lead getting converted are the Total time spent on the website, the Lead origin and the lead source.

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.9347	0.124	-7.528	0.000	-1.178	-0.691
Do Not Email	-1.4107	0.169	-8.339	0.000	-1.742	-1.079
Total Time Spent on Website	1.1455	0.040	28.726	0.000	1.067	1.224
Lead Origin_Landing Page Submission	-0.2049	0.091	-2.241	0.025	-0.384	-0.026
Lead Origin_Lead Add Form	4.3319	0.226	19.204	0.000	3.890	4.774
Lead Origin_Lead Import	1.6907	0.448	3.777	0.000	0.813	2.568
Lead Source_Google	0.3644	0.080	4.539	0.000	0.207	0.522
Lead Source_Olark Chat	1.0229	0.127	8.082	0.000	0.775	1.271
Specialization_strong demand	-0.1785	0.083	-2.140	0.032	-0.342	-0.015
Tags_lost	1.0083	0.111	9.066	0.000	0.790	1.226
Tags_unresponsive	-1.1399	0.152	-7.509	0.000	-1.438	-0.842
Last Notable Activity_Modified	-0.8424	0.086	-9.792	0.000	-1.011	-0.674
Last Notable Activity_Other activity	-0.3013	0.129	-2.344	0.019	-0.553	-0.049
Last Notable Activity_SMS Sent	1.5179	0.088	17.245	0.000	1.345	1.690

   As it can be observed,when everything else is constant, leads originated from lead add form increases the log odds of a lead getting converted by 4.33 and lead originated from lead import increases the log odds of a lead getting converted by 1.69. Total time spent on the website also increases the logg odds of lead getting converted linearly.

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?

   The top 3 dummy variables which would affect the model by increasing or decreasing the values are lead origin Lead add form where the coefficient value is 4.339 the variable that would be the Lead origin _lead import and the third variable last notable activity sms sent 1.517 there are positively correlated with the y values and if these values increases it would increase the probability of the lead conversion.

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.

Since there are new interns to make phone calls, the strategy here should be to increase sensitivity as that would enable the model to predict higher number of conversions as a percentage of actual conversions. This can be achieved by reducing the cutoff for conversion probability. Also, this would increase the specificity which means there would also be decrease in true negatives which will reduce the ability of the model to predict number of leads that will not be converted. This will result in making unnecessary calls. But since interns are hired specifically for this purpose, company can afford to lose specificity.

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.

   Since the target is already achieved, sales team need to focus on reducing the number of calls made to leads that will eventually not be converted. This is achieved by increasing specificity as the increase in specificity increases the ability of the model to predict true negatives. This can be achieved by increasing the cutoff for conversion probability. Therefore, sales team will be making calls to only those leads which have a higher probability of conversion. This will reduce sensitivity as they will be losing out on some leads which can actually be converted. But since they have already achieved the target, they do not need to go after leads with lower probability of conversion.