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

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.6474	0.585	-1.107	0.268	-1.793	0.498
TotalVisits	4.0447	1.199	3.375	0.001	1.696	6.394
Total Time Spent on Website	4.3198	0.184	23.421	0.000	3.958	4.681
Lead Origin_Lead Add Form	3.5342	0.227	15.553	0.000	3.089	3.980
Lead Source_Olark Chat	1.5566	0.126	12.366	0.000	1.310	1.803
Lead Source_Welingak Website	2.0778	0.752	2.764	0.006	0.604	3.551
Do Not Email_Yes	-1.5573	0.193	-8.079	0.000	-1.935	-1.179
Last Activity_Converted to Lead	-1.1403	0.238	-4.795	0.000	-1.606	-0.674
Last Activity_Olark Chat Conversation	-1.3210	0.184	-7.163	0.000	-1.682	-0.960
Last Activity_SMS Sent	1.0674	0.084	12.740	0.000	0.903	1.232
What is your current occupation_Student	-1.3919	0.617	-2.255	0.024	-2.602	-0.182
What is your current occupation_Unemployed	-1.4870	0.581	-2.559	0.010	-2.626	-0.348
What is your current occupation_Working Professional	1.3025	0.613	2.125	0.034	0.101	2.504
Last Notable Activity_Unreachable	2.5712	0.814	3.158	0.002	0.975	4.167

From the co-efficients of the model built, the highest contributing factors are the following

a) Total Visits
b) Total Time Spent on Website
c) Lead Origin_Lead Add Form

Intuitively as well, these are the best indicators of a potential customer's intent to convert. Their Porbability levels are also close to 0 and we can predict with fairly high confidence in such scenarios.

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 categorical variables indicating probability of conversion include the following:

a) Lead Origin (Lead Add Form)
b) Last Notable Activity_Unreachable
c) Lead Source_Welingak Website

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.

In such a scenario, with higher capacities for outreach and the priority on increased conversions, the objective is to maximize **Recall,** which is to minimize false negatives, ensuring that the highest number of positive cases are identified as such. The model that we have built has a recall of 79% which is good. This can be further increased by lowering the threshold value for tagging a positive.

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 such a scenario, only high intent customers need to be addressed. For this we need the fhighest precision possible, which can be obtained by a higher than normal threshold value. The minor drawback here is that you could miss a few cases that can convert but by doing this we can achieve the following

- Outgoing calls can be minimized except to address urgent queries from leads on the verge of converting.
- Most queries can also be addressed via emails or portal links with answers to the queries
- An effective ticketing system with good adherence to promised TAT which is again communicated well in time to the leads will take care of such a scenario