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

   Solution:

   Based on the coefficient values from below screeshot, the following are the top three variables that contribute most towards the probability of a lead getting converted :
   - a) Welingak Website (from Lead Source)
   - b) Total Time spent on the Website
   - c) Reference (from Lead Source)

	coef	std err	z	P>|z|	[0.025	0.975]
const	-2.2975	0.119	-19.296	0.000	-2.531	-2.064
Do Not Email	-1.1269	0.174	-6.486	0.000	-1.467	-0.786
Total Time Spent on Website	4.4551	0.164	27.192	0.000	4.134	4.776
LeadSource_Olark Chat	1.1777	0.103	11.400	0.000	0.975	1.380
LeadSource_Reference	3.7329	0.217	17.217	0.000	3.308	4.158
LeadSource_Welingak Website	5.5265	0.723	7.641	0.000	4.109	6.944
LastActivity_Email Opened	0.4582	0.108	4.249	0.000	0.247	0.670
LastActivity_SMS Sent	1.6265	0.108	15.036	0.000	1.414	1.839
CurrentOccupation_No Information	-1.1649	0.088	-13.221	0.000	-1.338	-0.992
CurrentOccupation_Working Professional	2.5101	0.185	13.577	0.000	2.148	2.872
LastNotableActivity_Modified	-0.6494	0.089	-7.264	0.000	-0.825	-0.474
LastNotableActivity_Olark Chat Conversation	-1.1467	0.370	-3.096	0.002	-1.873	-0.421
LastNotableActivity_Unreachable	2.3843	0.592	4.030	0.000	1.225	3.544

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?

   Solution:

   The following are the top three categorical/dummy variables that should be focused the most in order to increase the probability of lead conversion :
   - a) SMS Sent (from Last Activity)
   - b) Unreachable (from Last Notable Activity)

c) Working Professional (from What is your current occupation)

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.

Solution:

In the below image, the final prediction is calculated based on a optimal cut off value of 0.38

In order to make the sales aggressive, the company may contact all the leads which have a conversion probabilty (value = 1) under a cut off 0.4

	Converted	Converted_Prob	LeadId	predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	final_predicted	lead_score
0	1	0.694944	1381	1	1	1	1	1	1	1	1	0	0	0	1	69
1	1	0.341146	9105	0	1	1	1	1	0	0	0	0	0	0	0	34
2	0	0.050494	5851	0	1	0	0	0	0	0	0	0	0	0	0	5
3	0	0.701386	4310	1	1	1	1	1	1	1	1	1	0	0	1	70
4	1	0.919767	5826	1	1	1	1	1	1	1	1	1	1	1	1	92
5	0	0.053305	3749	0	1	0	0	0	0	0	0	0	0	0	0	5
6	0	0.059801	7382	0	1	0	0	0	0	0	0	0	0	0	0	6
7	1	0.533695	686	1	1	1	1	1	1	1	0	0	0	0	1	53
8	0	0.145643	437	0	1	1	0	0	0	0	0	0	0	0	0	15
9	0	0.198589	4684	0	1	1	0	0	0	0	0	0	0	0	0	20
10	0	0.012239	3920	0	1	0	0	0	0	0	0	0	0	0	0	1
11	0	0.146980	1501	0	1	1	0	0	0	0	0	0	0	0	0	15
12	1	0.999188	818	1	1	1	1	1	1	1	1	1	1	1	1	100
13	1	0.955295	1897	1	1	1	1	1	1	1	1	1	1	1	1	96
14	0	0.020564	2792	0	1	0	0	0	0	0	0	0	0	0	0	2
15	1	0.992692	8243	1	1	1	1	1	1	1	1	1	1	1	1	99
16	0	0.138659	3110	0	1	1	0	0	0	0	0	0	0	0	0	14
17	1	0.035614	1955	0	1	0	0	0	0	0	0	0	0	0	0	4
18	0	0.050494	8315	0	1	0	0	0	0	0	0	0	0	0	0	5
19	1	0.145643	7116	0	1	1	0	0	0	0	0	0	0	0	0	15

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.

Solution:

In order to minimize the rate of useless phone calls, the company may contact all the leads which have a conversion probabilty (value = 1 highlighted in yellow color) under column 0.7. However, the flipside here would be that, we may miss out on those leads that are actually converted but then the model wrongly predicted them as not converted. (See red highlights in the image below). This should not be a major cause for concern as the target has already be achieved.

	Converted	Converted_Prob	LeadId	predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	final_predicted	lead_score
0	1	0.694944	1381	1	1	1	1	1	1	1	1	0	0	0	1	69
1	1	0.341146	9105	0	1	1	1	1	0	0	0	0	0	0	0	34
2	0	0.050494	5851	0	1	0	0	0	0	0	0	0	0	0	0	5
3	0	0.701386	4310	1	1	1	1	1	1	1	1	1	0	0	1	70
4	1	0.919767	5826	1	1	1	1	1	1	1	1	1	1	1	1	92
5	0	0.053305	3749	0	1	0	0	0	0	0	0	0	0	0	0	5
6	0	0.059801	7382	0	1	0	0	0	0	0	0	0	0	0	0	6
7	1	0.533695	686	1	1	1	1	1	1	1	0	0	0	0	1	53
8	0	0.145643	437	0	1	1	0	0	0	0	0	0	0	0	0	15
9	0	0.198589	4684	0	1	1	0	0	0	0	0	0	0	0	0	20
10	0	0.012239	3920	0	1	0	0	0	0	0	0	0	0	0	0	1
11	0	0.146980	1501	0	1	1	0	0	0	0	0	0	0	0	0	15
12	1	0.999188	818	1	1	1	1	1	1	1	1	1	1	1	1	100
13	1	0.955295	1897	1	1	1	1	1	1	1	1	1	1	1	1	96
14	0	0.020564	2792	0	1	0	0	0	0	0	0	0	0	0	0	2
15	1	0.992692	8243	1	1	1	1	1	1	1	1	1	1	1	1	99
16	0	0.138659	3110	0	1	1	0	0	0	0	0	0	0	0	0	14
17	1	0.035614	1955	0	1	0	0	0	0	0	0	0	0	0	0	4
18	0	0.050494	8315	0	1	0	0	0	0	0	0	0	0	0	0	5
19	1	0.145643	7116	0	1	1	0	0	0	0	0	0	0	0	0	15