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	-0.0913	0.126	-0.724	0.469	-0.339	0.156
Do Not Email	-1.6724	0.190	-8.783	0.000	-2.046	-1.299
Total Time Spent on Website	1.0988	0.040	27.190	0.000	1.020	1.178
Lead Origin_Landing Page Submission	-1.1678	0.128	-9.130	0.000	-1.419	-0.917
Lead Origin_Lead Add Form	3.2825	0.235	13.959	0.000	2.822	3.743
Lead Source_Olark Chat	1.0942	0.123	8.901	0.000	0.853	1.335
Lead Source_Welingak Website	2.5816	0.756	3.414	0.001	1.100	4.063
Last Activity_Olark Chat Conversation	-0.9433	0.171	-5.501	0.000	-1.279	-0.607
Last Activity_Other_Activity	2.2026	0.463	4.761	0.000	1.296	3.109
Last Activity_SMS Sent	1.3246	0.075	17.585	0.000	1.177	1.472
Last Activity_Unsubscribed	1.3761	0.482	2.854	0.004	0.431	2.321
Specialization_Other_Specialization	-1.1841	0.126	-9.422	0.000	-1.430	-0.938
What is your current occupation_Working Professional	2.6109	0.194	13.424	0.000	2.230	2.992
Last Notable Activity_Modified	-0.8810	0.081	-10.827	0.000	-1.041	-0.722
Last Notable Activity_Unreachable	1.5512	0.474	3.274	0.001	0.623	2.480

Occupation set as Working Professional, Total Time Spent on Website and lead source are three main independent variables that affects the lead conversion.

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?

Lead Origin_Lead Add Form
Lead Source_Olark Chat
Lead Source_Welingak Website

After looking in to these the most correlated category variable is Lead Source

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.

Based on Coefficient:
Make phone call to person who spend **a lot of time in website**
Converse with working professional who have high lead possibility of conversion
Leads source who come through **Lead Add Form, Welingak Website** can be contacted for higher possibilities

Univariate and bivariate analysis shows below inference:
Lead origin:
1. Landing Page Submission have higher count of lead originated
2. Lead Add Form has high conversion rate but count of lead are less
Lead source:
1. Google have maximum number of leads.
2. reference leads and welingak website , conversion is high
Time spent:
1. Leads spending more time on the weblise are more likely to be converted.

Last Activity:
1. Email opened has highest lead
2. Conversion rate for leads SMS Sent is high

Occupation:
1. Working Professionals going for the course have high conversion
2. Unemployed leads are the high

Intern should work on conversion of lead variable having high lead numbers
He should also work on increasing the lead numbers of leads with highest 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.

   Checking negative correlation:

Specialization_Other_Specialization	-0.161448
Do Not Email	-0.140786
Last Activity Email Bounced	-0.127327

   Coefficient:

   Do Not Email -1.6724 **, Lead Origin Landing Page Submission** -1.1678

   People who are unemployed and having specialization 'Other' can be ignored for this process. Also Person with Email bounced and status of Do Not Email set also will not convert to Hot leads. So based on the correlation we can conclude this