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

Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	6395
Model:	GLM	Df Residuals:	6377
Model Family:	Binomial	Df Model:	17
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2592.8
Date:	Sun, 03 Mar 2019	Deviance:	5185.5
Time:	00:44:32	Pearson chi2:	6.73e+03
No. Iterations:	6	Covariance Type:	nonrobust

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.5360	0.158	-9.748	0.000	-1.845	-1.227
Total Time Spent on Website	1.1015	0.040	27.204	0.000	1.022	1.181
Lead Origin_Landing Page Submission	-0.9230	0.135	-6.828	0.000	-1.188	-0.658
Lead Origin_Lead Add Form	3.5867	0.218	16.422	0.000	3.159	4.015
Lead Source_Direct Traffic	-0.3228	0.089	-3.632	0.000	-0.497	-0.149
Lead Source_Olark Chat	1.1789	0.123	9.551	0.000	0.937	1.421
Last Activity_Email Bounced	-2.0357	0.414	-4.921	0.000	-2.846	-1.225
Last Activity_Email Opened	0.5346	0.092	5.794	0.000	0.354	0.715
Last Activity_Olark Chat Conversation	-0.9172	0.175	-5.256	0.000	-1.259	-0.575
Specialization_Hospitality Management	-0.7816	0.340	-2.299	0.021	-1.448	-0.115
Specialization_Unknown	-0.9700	0.126	-7.683	0.000	-1.217	-0.723
Last Notable Activity_Email Bounced	1.6392	0.605	2.710	0.007	0.454	2.825
Last Notable Activity_Had a Phone Conversation	3.4330	1.163	2.951	0.003	1.153	5.713
Last Notable Activity_SMS Sent	1.7753	0.101	17.569	0.000	1.577	1.973
Last Notable Activity_Unreachable	2.0800	0.538	3.868	0.000	1.026	3.134
What is your current occupation_Student	0.9241	0.238	3.890	0.000	0.458	1.390
What is your current occupation_Unemployed	1.0106	0.087	11.550	0.000	0.839	1.182
What is your current occupation_Working Professional	3.4090	0.205	16.623	0.000	3.007	3.811

Based on the output of the Logistic Regression Model, we can conclude that the following 3 parameters are most important (highest coefficient):

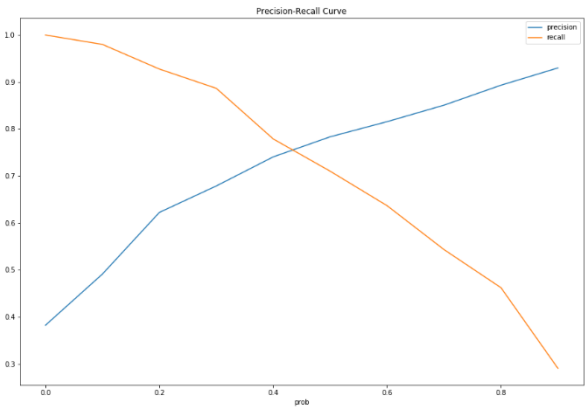
a. Lead Origin (Lead Add form)
b. Last Notable Activity (Had a phone conversation)
c. What is your current occupation (Working Professional)

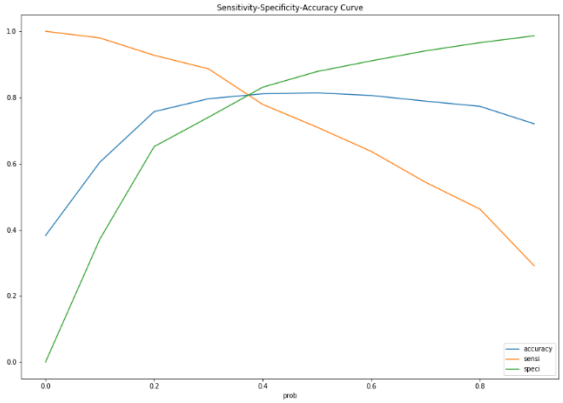
2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

   From the model summary above, we can see that the following categorical variables have the highest coefficients and contribute significantly to model conversion

   a. What is your current occupation – Working Professional
   b. Last Notable Activity – Had a phone conversation
   c. Lead Origin – Lead Add Form

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 this case scenario, we want to increase the potential leads, i.e**. reduce the cut-off** so the model predicts more leads as a potential.
Reducing the cut-off leads to higher ‘Sensitivity’ and lower ‘Precision’, as not all the leads predicted by model will eventually convert. However, with reduced cut-off we will be able to reach out to more prospective candidates.
Actual value of the cut-off will depend on the business objective

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 this scenario, we want to **increase the probability cut-off** so only the leads that are scored highest will be pursued.
   Increasing the cut-off will result in higher 'Precision', meaning the leads predicted as Yes by the model will convert.
   However, the Sensitivity will fall and we will be missing out on some border line/potential leads with slightly lower lead score
   Again, the cut-off value will depend on the business objective