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

Ans**.** Following three variables are contributing the most towards the probability of a leadgetting converted:

a. **Lead Source**
b. **Lead Origin**
c. **Lead Profile**

Reviewing coefficient of these variables indicate the importance of these variables.

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3068	0.190	1.613	0.107	-0.066	0.680
Total Time Spent on Website	1.0959	0.041	26.488	0.000	1.015	1.177
Lead Origin_Lead Add Form	3.2221	0.205	15.696	0.000	2.820	3.624
Lead Source_Olark Chat	1.2713	0.108	11.719	0.000	1.059	1.484
Lead Source_Welingak Website	3.5050	1.028	3.408	0.001	1.489	5.521
Last Activity_Email Bounced	-2.1009	0.356	-5.907	0.000	-2.798	-1.404
Last Activity_Email Opened	0.4216	0.096	4.386	0.000	0.233	0.610
Last Activity_Olark Chat Conversation	-1.0229	0.180	-5.672	0.000	-1.376	-0.669
What is your current occupation_Student	-1.7087	0.329	-5.197	0.000	-2.353	-1.064
What is your current occupation_Unemployed	-2.4014	0.181	-13.263	0.000	-2.756	-2.047
Lead Profile_Lateral Student	2.8593	1.115	2.565	0.010	0.675	5.044
Lead Profile_Potential Lead	1.8584	0.100	18.499	0.000	1.662	2.055
Lead Profile_Student of SomeSchool	-2.1485	0.455	-4.723	0.000	-3.040	-1.257
Last Notable Activity_SMS Sent	1.9094	0.104	18.368	0.000	1.706	2.113
Last Notable Activity_others	1.2984	0.298	4.363	0.000	0.715	1.882

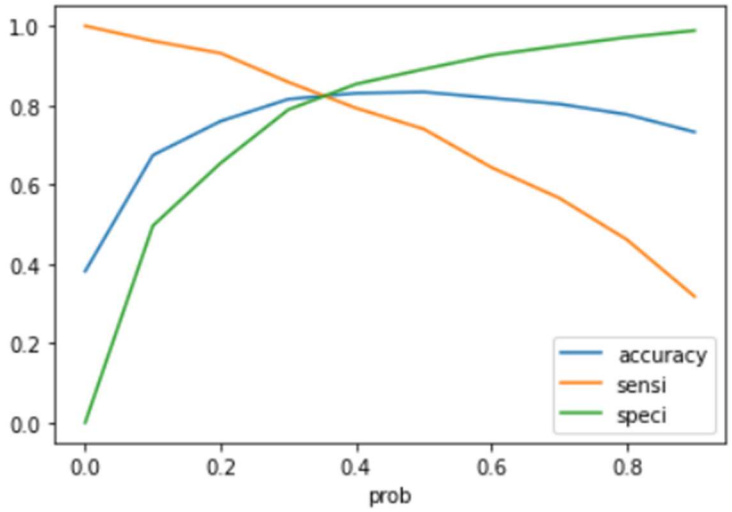
2. **What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?**

Ans. Upon reviewing the final model and as evident from EDA as well, followings are the top 3 categorical variables in the model which should be focused the most on in order to increase the probability of lead conversion:

a) **Lead Source_Welingak Website**
d. **Lead Origin Lead Add Form**
e. **Lead Profile Lateral Student**

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.**

Ans. Sensitivity with respect to our model can be defined as the ratio of total number of actual Conversions correctly predicted to the total number of actual conversions. Similarly, Specificity can be defined as the ratio of total number of actual non-conversions correctly predicted to the total number of actual non-conversions. For a model, as one increases, the other decreases and vice versa. Different values of the sensitivity and specificity can be achieved for the same model by changing the conversion probability cutoff threshold value. For our model, the below graph shows how the Accuracy, Sensitivity and Specificity rating changes with change in the threshold value:



When the probability thresholds are very low, the sensitivity is very high and specificityis very low. Similarly, for larger probability thresholds, the sensitivity values are verylow but the specificity values are very high.

High sensitivity implies that our model will correctly identify almost all leads who are likely to convert. It will do that by over-estimating the conversion likelihood, i.e. it will misclassify some non-conversion cases as conversions. Now, since X Education has moremanpower for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all the potential leads, we can choose a lower threshold value for Conversion Probability. This will ensure the

Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible. The company may follow high volume low margin strategy which means the conversion rate might reduce but, the count of conversion would increase and eventually the revenue would also increase.

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.**

Ans. Following the similar logic and context from the previous question, High Specificity implies that our model will correctly identify almost all leads who are not likely to convert. It will do that at the cost of losing out some low Conversion rate risky leads to the competition, i.e. it will misclassify some Conversion cases as non-Conversions.

Therefore, since X Education has already reached its target for a quarter and doesn't wantto make phone calls unless it is extremely necessary, i.e. they want to minimize the rate of useless phone calls. We can choose a higher threshold value for Conversion Probability. This will ensure the Specificity rating is very high, which in turn will make sure almost all leads who are on the brink of the probability of getting Converted or not are not selected. As a result, they won't have to make unnecessary phone calls and can focus on some new work. In this way the efficiency of sales team would increase as the conversion rate would be high. The sales cycle will also be reduced.