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

Answer: The top 3 variables that contribute most towards the probability of a lead getting converted are:

1. Total Time Spent on Website
2. Last Activity
   a. SMS Sent
   b. Olark chat
   c. Modified
3. Total Visit

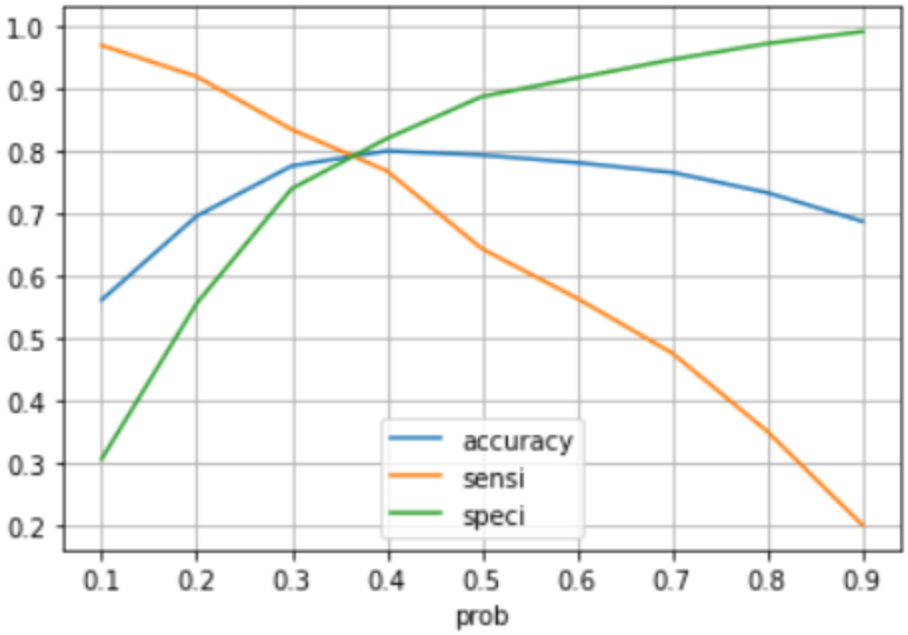
2. **What are the top 3 categorical/dummy variables in the model which should be focused the most on to increase the probability of lead conversion?**

Answer: The top 3 categorical/dummy variables that contribute the most towards the probability of a lead getting converted are:

1. Lead Origin_Landing page submission
2. Last Activity_SMS sent
3. Last Activity_Email opened

3. **X Education has a period of 2 months every year during which they hire some interns. The sales team 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 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 many of such people as possible. Suggest a good strategy they should employ at this stage.**

Answer: In general, the logistic regression model gives the probability of the output happening. The probabilities are converted to binomial results by defining a cutoff. By default, the cutoff is taken as 0.5 or 50%. But based on the problem we are solving we might tweak the cutoff to get our desired output. For this, we use the Sensitivity and Specificity tradeoff. Sensitivity is defined as the ratio of the total number of actual Conversions correctly predicted to the total no of actual Conversions and Specificity is defined as the ratio of the total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions. For a particular model, as one increases, the other decreases and vice versa. Different values of 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 Sensitivity and Specificity rating changes with a change in the threshold value:




As per the business problem mentioned above, we need to get a maximum number of leads. Even if there might be false positives, they should not miss any false negatives. The model should have high Recall or sensitivity. To do so we can reduce the cutoff keeping the accuracy and specificity also optimum. As per the analysis, we can reduce the cutoff to 0.41 to maintain high sensitivity, specificity, and accuracy.

4. **Similarly, at times, the company reaches its target 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.**

Answer: The 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. Since X Education has already reached its target for a quarter and doesn't want to make phone calls unless it's 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 that the leads having high lead scores are only converted. As a result, the agents won't have to make unnecessary phone calls and can focus on some new work. The company can also concentrate the leads on important features making them in priority. For example, Working professionals, people who have spent more time on the website etc.