1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
    i. **TotalVisits,**
    ii. **Total Time Spent on Website,**
    iii. **Lead Origin_Lead Add Form**

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?
    i. **Lead Origin_Lead Add Form**
    ii. **Last Activity_Had a Phone Conversation**
    iii. **Lead Source_Welingak Website**

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.

    To answer this question, we need to have the plot for sensitivity and specificity and in our context, sensitivity is defined as the number of actual conversions predicted correctly out of total number of conversions.
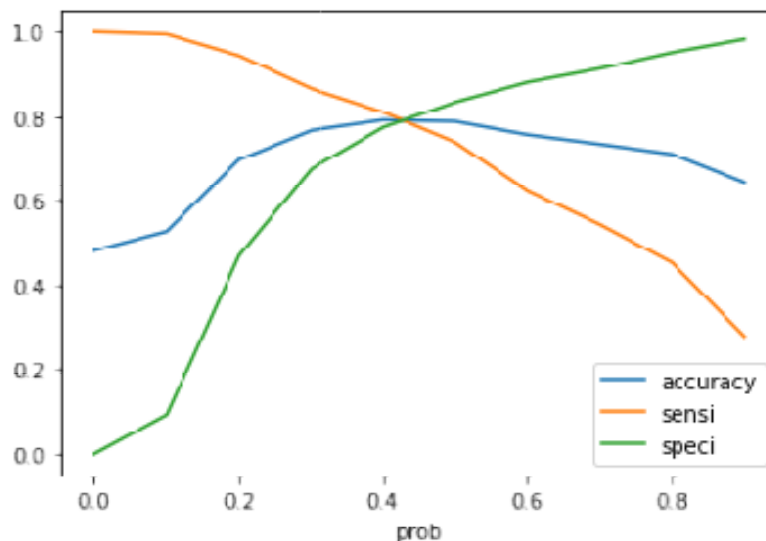
    So when the company has more interns we need have lower cutoff so that our model can predict almost all leads. The flip side to this decrease in threshold will be that we will misclassify some non-conversions as conversions but this is a

good tradeoff given we have mode manpower to deal with it.

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.

   Typically, we will be having less people to call potential customers so its good to have more accurate predictions which would be our specificity should be much more higher.

   Specificity in our context is defined as actual numbers of non-conversion predicted correctly out of total number of actual non conversion.

   This would mean form the above graph the we would have to choose a cutoff point which is much higher.