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

Ans: Top three variables which contribute most towards the probability of a lead getting converted are:
   A. Tags_Will revert after reading the email
   B. Lead Origin_Lead Add Form
   C. Lead Source_Welingak Website

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: Top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are:
   A. Tags_Will revert after reading the email
   B. Lead Origin_Lead Add Form
   C. 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.**

Ans: The good strategy to approach this problem would be to focus on sensitivity metric. Sensitivity is given by =TP/(TP+FN). Sensitivity can be defined as the number of actual conversions predicted correctly out of total number of actual conversions. Different values of sensitivity can be achieved for the model by changing the cutoff threshold for probability of lead conversion. For our model, below is the graph showing changes in accuracy, sensitivity and specificity with change in probability threshold.
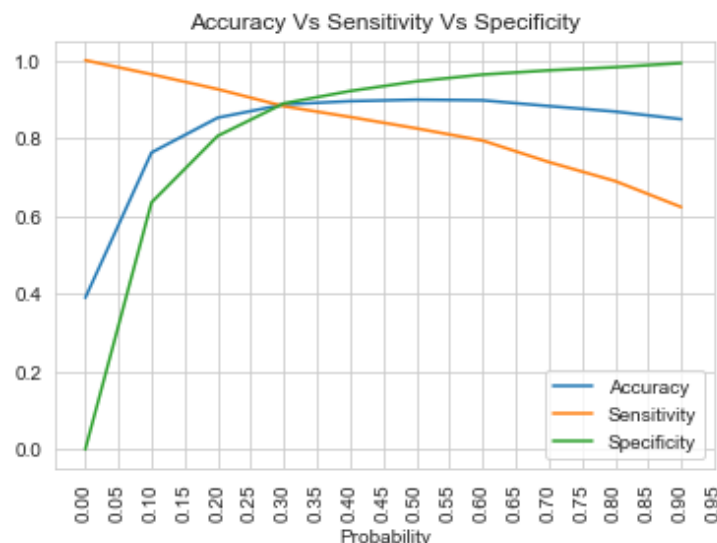
When the probability thresholds are very low, the sensitivity is very high and specificity is very low. High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e., it will misclassify some non-Conversion cases as Conversions.

Now, since X Education has more man-power for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all of the potential leads, we can choose a lower threshold value for Conversion Probability.

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: The good strategy to approach this problem would be to focus on specificity metric. It is given by $=TN/(TN+FP)$. Specificity can be defined as the ratio of total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions. For larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

From the graph used in last question, we can see specificity increases as threshold increases, so we will require high specificity, because high specificity will correctly predict almost all non-conversions. It may be a scenario where some conversions get classified as non-conversions. Therefore, 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.