Subjective Questions and Answers

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

   Solution:

	coef
const	0.2040
totalvisits	11.1489
total_time_spent_on_website	4.4223
lead_origin_Lead Add Form	4.2051
lead_source_Olark Chat	1.4526
lead_source_Welingak Website	2.1526
do_not_email_Yes	-1.5037
last_activity_Had a Phone Conversation	2.7552
last_activity_SMS Sent	1.1856
what_is_your_current_occupation_Student	-2.3578
what_is_your_current_occupation_Unemployed	-2.5445
last_notable_activity_Unreachable	2.7846

   Hence, the top three variables contribute most towards the probability of a lead getting converted, in the model will the following.

   a) Total Visits.
   b) Total time spent on website.
   c) 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?

   Solution:

   The top 3 Categorical/dummy variable to increase probability are:

   • Lead source with direct traffic
   • Lead source with Olark Chat
   • Last activity SMS sent.

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.
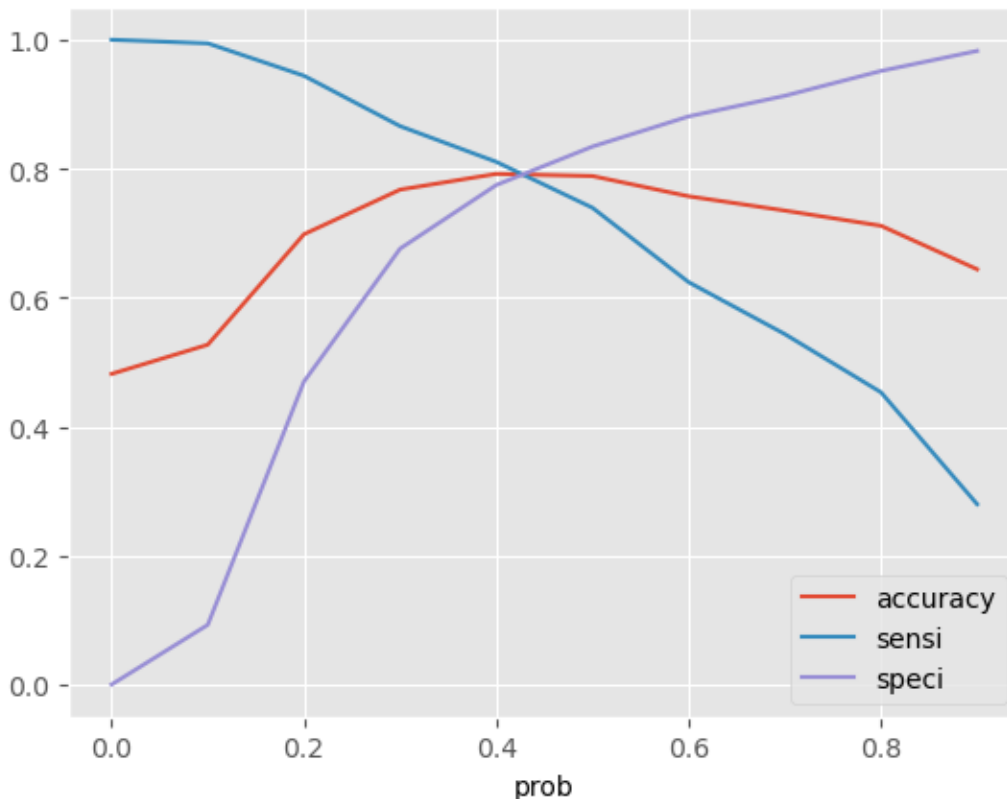
Here, the concept of sensitivity is at play,

- In the given situation, we will need a tweak of the model to increase its sensitivity because high sensitivity will mean that our model will correctly predict most all leads who are likely to convert.
- To achieve high sensitivity, we need to choose a low probability threshold value.

Below is an explanation:

- Sensitivity = True Positives / (True Positives+ False Negatives)
- With respect to our model, sensitivity can be defined as "Off all the leads that converted, how many were correctly predicted as converted by our model.
- Different values of sensitivity can be achieved for the model by changing the probability cutoff threshold for lead conversion.
- For our model the graph given in above shows changes in sensitivity, specificity and accuracy with change in the probability threshold.
  a) As we can see, sensitivity decreases with every increase in the cut-off threshold.
  b) However, we also need to be cognizant that sensitivity and precision are inversely correlated hence when sensitivity is increases the precision would decline and vice versa.
  c) In our case it would lead to our model misclassifying some of the non-converted leads are converted.
  d) But as the company has extra man-power for two months and wants to make the lead conversion more aggressively by making phone calls to as much potential leads as possible, it is a good strategy to go for high sensitivity.

Such sensitive activity can be considered for a phone call, to those people if:

- The spend time on Olark chat session direct traffic and google search under lead source.
- Considering number of Page visited on Website.
- They are unemployed.
- Their most recent action via SMS or an Olark chat session.
- Those landed on Submission page.

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.

Solution:

- ➢ Here, the concept of specificity is at play.
- ➢ In the given situation, will need a high specificity because high specificity will mean that our model will correctly predict almost all leads who are not likely to covert. To achieve high specificity, we need to choose a high threshold value.

Below is an explanation:

- ➢ *Specificity = True Negativity /(True Negatives/False Positives)*
- ➢ With respect to our model, specificity can be defined as "of all the people who did not convert, how many did our model correctly predict.
- ➢ As seen in the graph above, we can see that the specificity increases as the threshold increase.
- ➢ However, increasing the specificity may lead to misclassifying some of the converted leads as non-converted.
- ➢ As the company has already reached its target for a quarter and does not want to make phone calls unless it is extremely necessary, it is a good strategy to for high specificity.
- ➢ It will ensure that the phone calls are only made to customers who have very high probability of conversion.