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

   From the model, we obtained the equation of ln (ODDs) as

$$\begin{aligned} ln(ODDs) &= 3.574 * LastNotableActivity\_Had\ aPhone\ Conversation + 3.433 * Lead\ Origin\_LeadAddForm + 2.571 \\ &* Lead\ Source\_Welingak\ Website + 2.451 * What\ is\ your\ current\ occupation\_Working\ Professional + 1.745 \\ &* LastNotableActivity\_Unreachable - 1.663 * LastActivity\_EmailBounced - 1.481 * LastActivity\_Olark\ Chat\ Conversation + 1.254 \\ &* Lead\ Source\_Olark\ Chat - 1.218 * What\ is\ your\ current\ occupation\_Unspecified + 1.212 * LastActivity\_SMSSent + 1.141 \\ &* Total\ Time\ Spent\ on\ Website - 0.344 * Lead\ Origin\_LandingPageSubmission - 0.966 \end{aligned}$$

   As the ODDs increase when the probability increase, so from the equation given above, the top three variables which contribute the most towards the probability of lead getting converted are as below.

   - **Last Notable Activity of '*Had a Phone Conversation*'** with a weight of **3.574**.
   - **Lead Origin of '*Lead Add Form*'** with a weight of **3.433**.
   - **Lead Source of '*Welingak Website*'** with a weight of **2.571**.

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?**

   As referring from the previous question, it can be observed that 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 the same, which are as below.

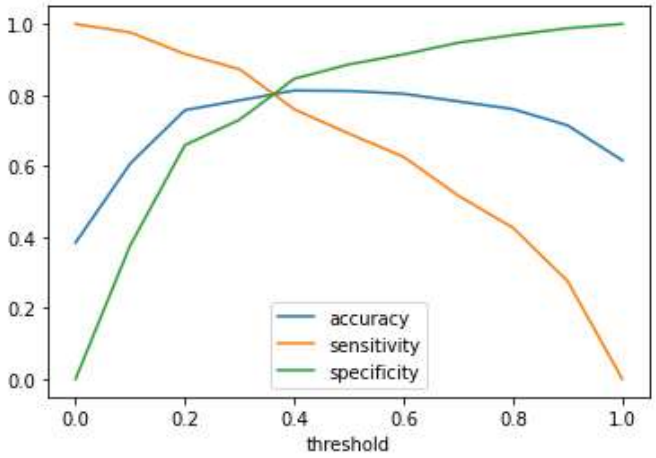
   - **Last Notable Activity of '*Had a Phone Conversation*'** with a weight of **3.574**.
   - **Lead Origin of '*Lead Add Form*'** with a weight of **3.433**.
   - **Lead Source of '*Welingak Website*'** with a weight of **2.571**.

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.**

   As the company's aim is to make the lead conversion more aggressive and have the model predicted converted leads to be actually converted, therefore, the metric which should be considered to be optimized is 'Sensitivity'.
   The Sensitivity is expressed as the ratio of True Positive values to the total actual positive values. i.e.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

   The below graph was plotted from the model, which shows the relation between accuracy, sensitivity and specificity with the threshold value.



   Hence, it can be seen that the optimum threshold value obtained is 0.33. However, as sensitivity is the metric which is to be optimized, and it can be seen that sensitivity increases as the threshold value reduces, therefore, for this scenario, **the threshold value taken should be on the lower side**. However, the other metrics also needs to be taken into consideration as per their required tolerance levels.

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.**

   Here the company's aim is to not make phone calls unless it's extremely necessary, therefore, the metric which should be considered to be optimized is 'Specificity'.

   The Specificity is expressed as the ratio of True Negative values to the total actual negative values. i.e.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

   Referring to the figure provided in the previous question, it can be seen that the optimum threshold value obtained is 0.33. However, as specificity is the metric which is to be optimized, and it can be seen that specificity increases as the threshold value increases, therefore, for this scenario, **the threshold value taken should be on the higher side**. However, the other metrics also needs to be taken into consideration as per their required tolerance levels.