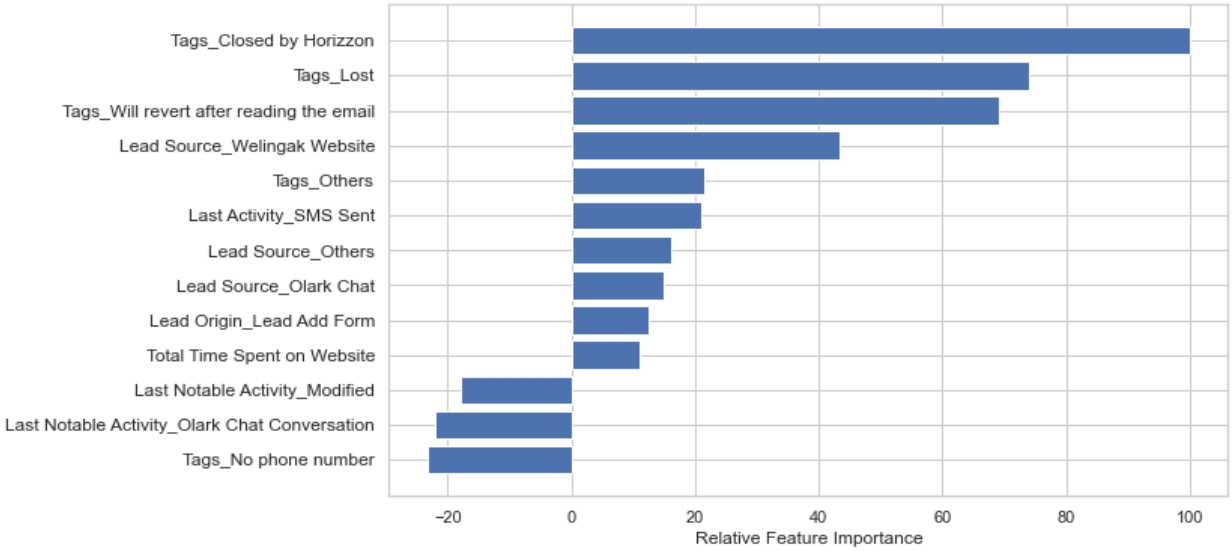
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?
   Solution:
   Following are the top three variables which contribute most towards the probability of a lead getting converted:



   From the graph, the three variables which contribute most towards the probability of a lead conversion in decreasing order of impact are:
   - **Tags_Closed by Horizzon**
   - **Tags_Lost**
   - **Tags_Will revert after reading the email**

These features are the dummy features created from the categorical variable Tags. These characteristics increase the probability that a lead will be converted. These findings suggest that the company should pay closer attention to the leads with these three tags.

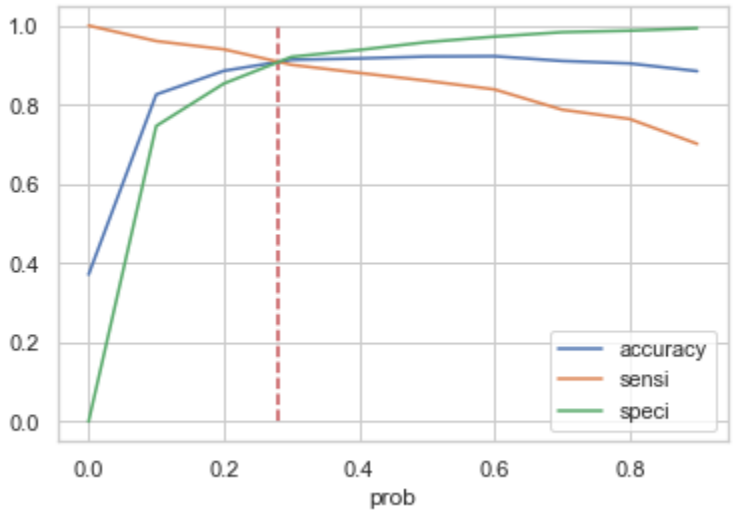
2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?
   Solution:
   From the above graph, the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are:
   - **Tags_Closed by Horizzon**
   - **Tags_Lost**
   - **Tags_Will revert after reading the email**

   The top three variables in the model are all categorical/dummy variables, hence the answers to both questions are the same.

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.
Solution:

Here, the concept of sensitivity is required.
$\boldsymbol{Sensitivity} = \boldsymbol{True\ Positives}/ (\boldsymbol{True\ Positives} + \boldsymbol{False\ Negatives})$
With respect to our model, 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 Sensitivity, Specificity and Accuracy with change in the threshold:



As the threshold rises, as we can see, sensitivity falls. We will require a high sensitivity in the current scenario since a high sensitivity will imply that our model will properly estimate most leads who are likely to convert.

Additionally, it might overstate and incorrectly label some non-conversions as conversions. However, choosing high sensitivity is a sensible strategy given that the organization has extra staff for two months and wants to increase lead conversion by calling as many potential leads as possible. We must select a low threshold value in order to attain high sensitivity.

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:

   The approach to answer this question is like the last one.
   Here, the concept of specificity is required.
   $\boldsymbol{Specificity}$ = $\boldsymbol{True\ Negatives}$/ ($\boldsymbol{True\ Negatives}$ + $\boldsymbol{False\ Positives}$)

   With respect to our model, specificity can be defined as the number of actual non-conversions predicted correctly out of total number of actual non-conversions.

   From the above graph, we can see that the specificity increases as the threshold increases. In the given situation, we'll need a high specificity because high specificity will mean that our model will correctly predict almost all leads who are not likely to convert.

   While doing so, it can mistakenly classify some conversions as non-conversions. However, since the business has already beyond its quarterly goal and doesn't want to make calls unless absolutely required, it is a good strategy to aim for high specificity.

   It will make sure that only consumers with a very high chance of converting will receive phone calls. We must select a high threshold number in order to attain high specificity.