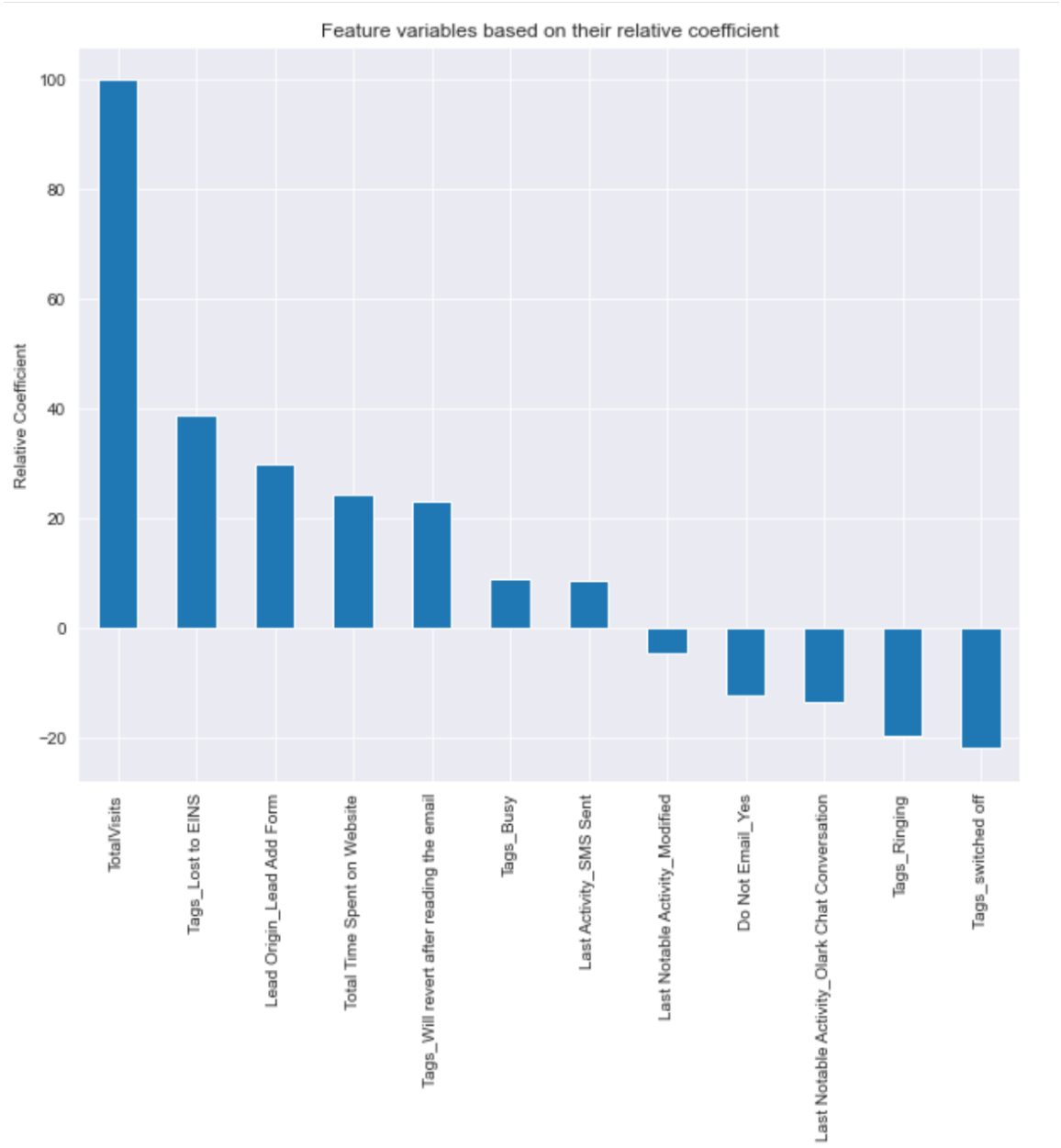
1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

   Answer:

The feature variables used to build the model have been represented below based on their importance in lead conversion as per their coefficient values.



From the above chart, the top 3 variables that are contributing to convert a lead are:

* TotalVisits
* Tags_Lost to EINS
* 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?

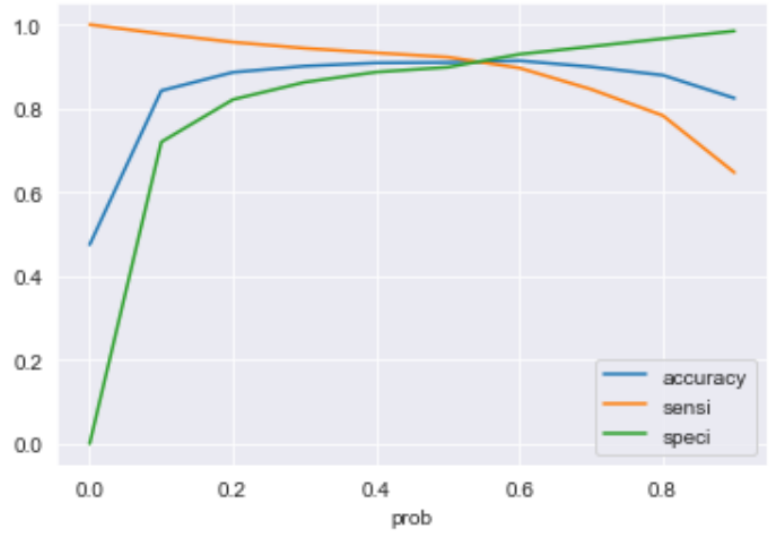
As per the chart, the top 3 categorical/ dummy variables that need improvement to convert a lead are:
* Tags_switched off
* Tags_Ringing
* Last Notable Activity_Olark Chat Conversation

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.

For our Model, Sensitivity can be defined as the ratio of total number of actual Conversions correctly predicted to the total no of actual Conversions.
And Specificity can be defined as the ratio of total no of actual non-Conversions correctly predicted to the total number of actual non-Conversions.

In case of any model, as one increases, the other decreases and vice versa.
Different values of the sensitivity and specificity can be achieved for the same model by changing the Conversion Probability cutoff threshold value.

For our model, the below graph shows how the Sensitivity and Specificity rating changes with change in the threshold value:



When the probability thresholds are very low, the sensitivity is very high and specificity is very low. And for higher probability thresholds, the sensitivity values are very low but the specificity values are very high.

High sensitivity indicates that our model will correctly identify almost all leads who are likely to be Converted. It will over-estimate the Conversion likelihood, means it will misclassify some non-Conversion cases as Conversions.

As X Education has more man-power for these 2 months and they want 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.

Making sure the Sensitivity rating is very high which in turn will ensure almost all leads who are likely to Convert are identified correctly and the customer support can call to convert as many of such people as possible.

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.

Similarly, following the logic and context from the previous question, High Specificity indicates our model will correctly identify almost all leads who are not likely to Convert. It will do that at the tradeoff of losing out some low Conversion rate risky leads to the competition, i.e. it will misclassify some Conversion cases as non-Conversions.

Thus, as 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.

This will ensure the Specificity rating is very high, which in turn will make sure almost all leads who are on the brink of the probability of getting Converted or not are not selected. As a result the agents won't have to make unnecessary phone calls and can focus on some new work instead of spending resource and time on the non-convertible leads.