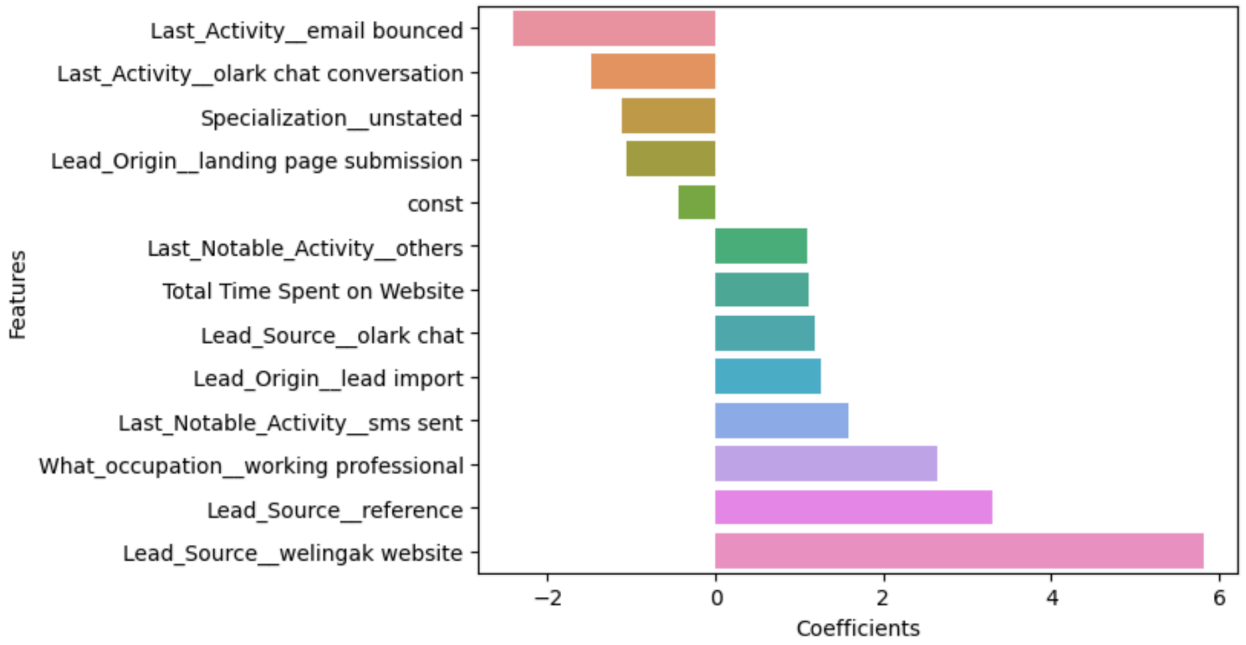
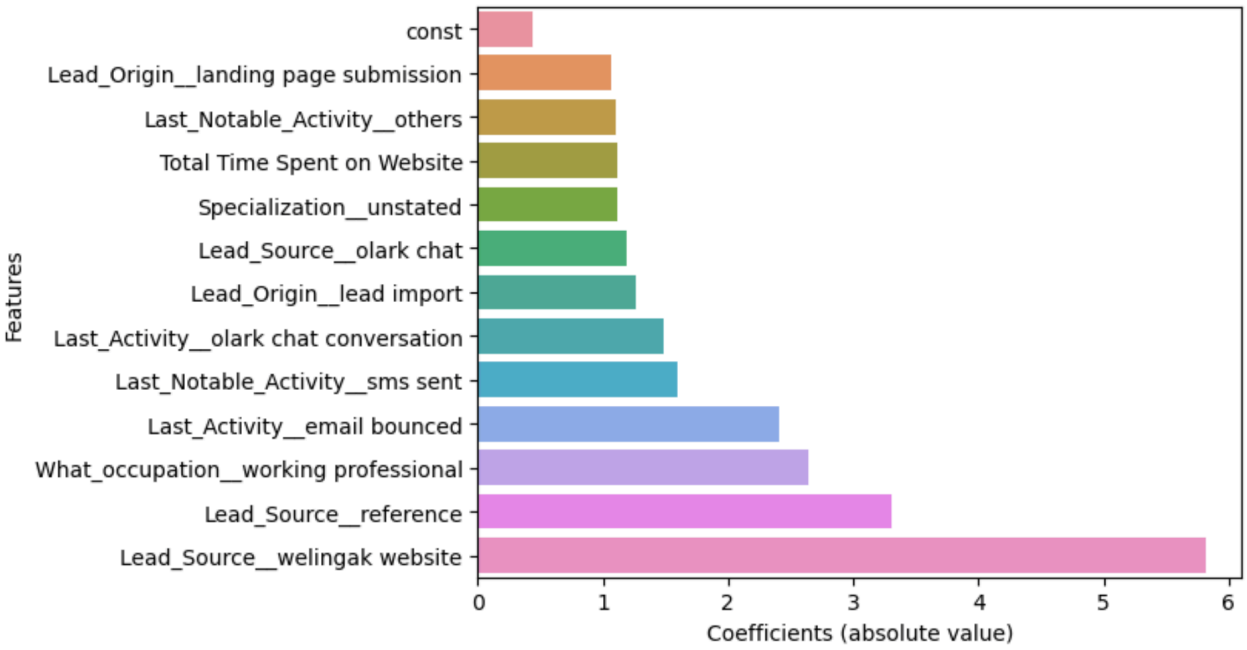
Q1. Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

Ans. Here are the bar plots demonstrating the final features in our model with their coefficients:





If we look at the bar plot with the absolute value of 'Coefficients', we can see that the top three features which contribute most towards the probability of a lead getting converted are:-

1) Lead_Source_welingak website
If we look at the first bar plot (the one which doesn't have absolute values of coefficients), we can see that this feature has the biggest and positive contribution to the conversion of a lead.
The coefficient value was **5.8156**

2) Lead_Source_reference
This feature had the second highest contribution (positive) to the lead conversion rate.
The coefficient value was **3.3094**

3) What_occupation_working professional
This feature had the third highest contribution (again positive) to the lead conversion rate.
The coefficient value was **2.6397**

Q2. 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. To get the top 3 categorical/dummy variables which should be focused on the most in order to increase the probability of lead conversion, we have to look at the features with the highest positive coefficients. These are features are :-

1) Lead_Source_welingak website
Coefficient Value: **+5.8156**

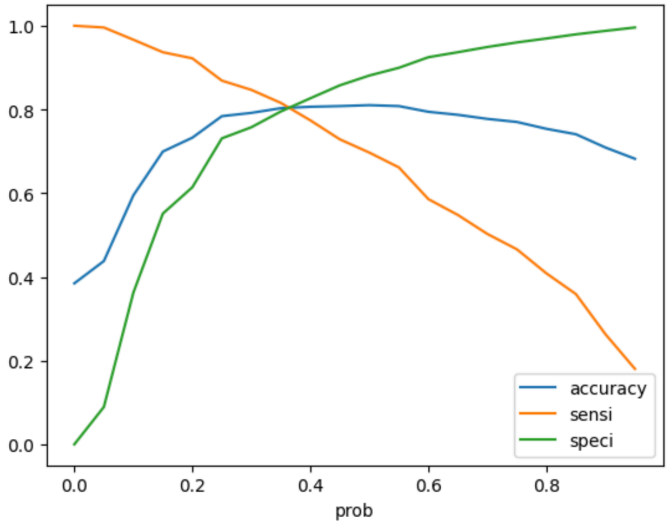
2) Lead_Source_reference
Coefficient Value: **+3.3094**

3) What_occupation_working professional
Coefficient Value: **+2.6397**

We can also see them at the bottom of the first bar plot given in the question above.

Q3. 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. In a nutshell, we have more manpower. In this case, we should reduce the threshold. We also plotted this in our jupyter notebook. As the threshold increases, the algorithm becomes more picky at choosing the right answer. It only chooses when it is pretty sure and it avoids making mistakes.
But when we have more manpower, we should reduce the threshold. This will give us almost all of the potential leads to be converted. There might be few that would be predicted wrong but that's part of the trade-off. Here is the plot of accuracy, sensitivity and specificity with different threshold values:-



Q4. 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 minimise the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Ans. In this situation, we have the opposite case. We have limited amount of time and manpower. Our goal is to maximise our output with the limited amount of time and manpower we have. So, in this case, we would do the opposite. We will increase the threshold thereby making the algorithm much more picky when it makes predictions. It will only tell us that a lead is convertible until it is highly probable that it would be converted. This will help us avoid making useless phone calls and only take action unless it is highly productive.