ASSIGNMENT

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

   After the application of Logistic Regression model on the Dataset, the analysis report depicts that the top 3 variables which contribute the most towards the probability of a lead getting converted are:

   - Tags
   - Country
   - Lead Origin

   These are thee variables with maximum Coefficient.

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?**

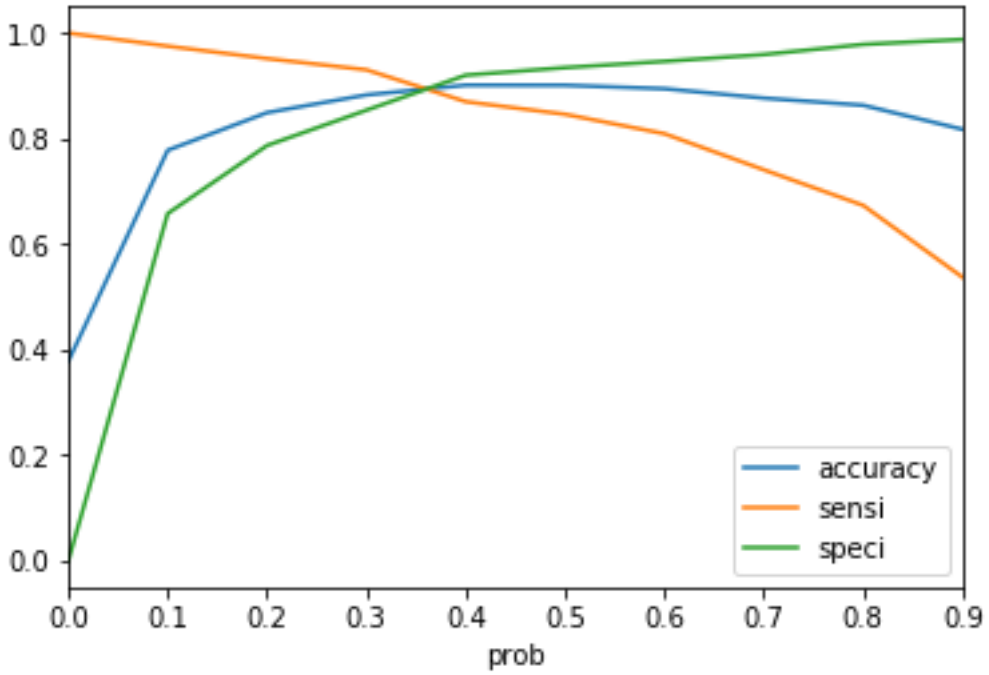
   The dummy/categorical variables which impact the most on the conversion rate are:

   - Tags_Closed by the Horizon
   - Tags_Lost to EINS
   - Tags_Will revert after reading the email

   These are the dummy variables having highest Coefficient and can be focussed on these variables to increase the probability of lead conversion.

3. **X Education has a period of 2 months every year during which they hire some interns. The sales team 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 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.**

   Company wants to make phone calls as much as possible, so in this case even if we identify some leads which are not going to convert as hot lead that won't make any difference since the company is trying to reach maximum leads as possible. So, the company wants to increase the cut-off for the models they have built for predicting the hot leads. In this case, the model needs to have less False negative count and higher True Positive Count. Ideally company should focus on Sensitivity of the model and prefer value for of Sensitivity should be high. The Sensitivity will increase with decrease in Cut-Off value.
   As shown in the plot below:



	prob	accuracy	sensi	speci
0.0	0.0	0.378084	1.000000	0.000000
0.1	0.1	0.777815	0.975144	0.657852
0.2	0.2	0.849136	0.951620	0.786832
0.3	0.3	0.882698	0.930315	0.853751
0.4	0.4	0.901158	0.869951	0.920130
0.5	0.5	0.900990	0.845983	0.934431
0.6	0.6	0.894445	0.809143	0.946303
0.7	0.7	0.876657	0.740790	0.959255
0.8	0.8	0.863064	0.673324	0.978413
0.9	0.9	0.817419	0.537062	0.987858

From above table it is seen that for the cut off value of 0.4 the Sensitivity is 0.86 and accuracy is 0.90 which is quite good. So the Company can predict based on this model and this will give them a lot of leads and can convert many of them by calling them.

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.**

   Company does not to make unnecessary calls so the model which is used to predict leads needs to have low FPR(False Positive Rate). The cut-off value should be high so that False Positives are avoided. Since, FPR = 1 – Specificity.
   The Specificity should be high for that scenario.
   The Specificity of the model increases with increase in Cut-off Value.