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

   Ans:

   After applying RFE and VIF checks on our logistic regression model, we have selected the following top three variables along with their coefficients:

   1. Tags_Closed by Horizzon
   2. Tags_Lost to EINS
   3. Tags_Will revert after reading the email.

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1179	0.084	-13.382	0.000	-1.282	-0.954
Total Time Spent on Website	0.8896	0.053	16.907	0.000	0.786	0.993
Lead Origin_Lead Add Form	1.6630	0.455	3.657	0.000	0.772	2.554
Lead Source_Direct Traffic	-0.8212	0.127	-6.471	0.000	-1.070	-0.572
Lead Source_Welingak Website	3.8845	1.114	3.488	0.000	1.701	6.068
Last Activity_SMS Sent	1.9981	0.113	17.718	0.000	1.777	2.219
Last Notable Activity_Modified	-1.6525	0.124	-13.279	0.000	-1.896	-1.409
Last Notable Activity_Olark Chat Conversation	-1.8023	0.491	-3.669	0.000	-2.765	-0.839
Tags_Closed by Horizzon	7.1955	1.020	7.053	0.000	5.196	9.195
Tags_Interested in other courses	-2.1318	0.406	-5.253	0.000	-2.927	-1.336
Tags_Lost to EINS	5.9177	0.611	9.689	0.000	4.721	7.115
Tags_Other Tags	-2.3737	0.206	-11.507	0.000	-2.778	-1.969
Tags_Ringing	-3.4531	0.238	-14.532	0.000	-3.919	-2.987
Tags_Will revert after reading the email	4.5070	0.188	24.002	0.000	4.139	4.875

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?

   Ans:

   After analyzing our logistic regression model, we have identified the top categorical variables that contribute the most to high probabilities:

   1. Tags_Closed by Horizzon
   2. Tags_Lost to EINS
   3. Tags_Will revert after reading the email
   4. Lead Source_Welingak Website
   5. Lead Origin_Lead Add Form

   These variables have been found to be the most significant predictors of the target variable based on their coefficients. This information can be used to gain insights into the factors that drive conversions and to inform targeted marketing campaigns.

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.

   Ans:

In our model, we have determined an optimal cut-off value of 0.3 based on the accuracy, sensitivity, and specificity. With this cut-off value, the model achieves a high positive rate of 97%, indicating that it is a good predictive model. Therefore, to maximize sales, the company should contact all leads with a conversion probability value of 1, provided they are below the cut-off of 0.3.
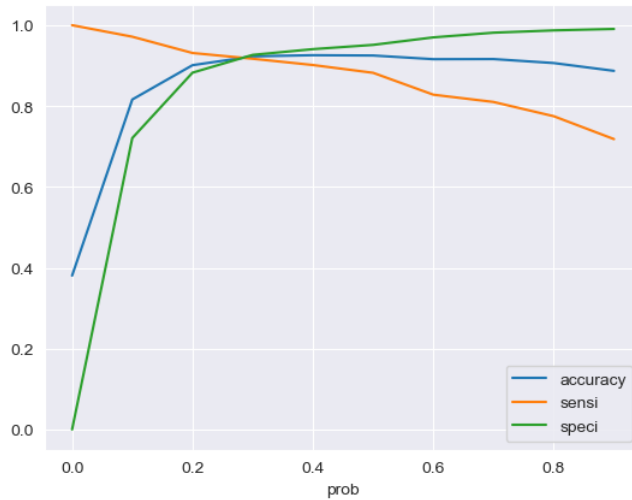
4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Ans:

To minimize the rate of useless phone calls, the company may consider contacting only those leads that have a conversion probability value of 1 under optimal cutoff 0.7. However, the flipside is that we may miss out on leads that were actually converted but were wrongly predicted as not converted. This should not be a major cause for concern since the sales target has already been achieved.

	Converted	Converted_prob	Prospect ID	Predicted	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	final_Predicted
0	1	0.283149	9196	0	1	1	1	0	0	0	0	0	0	0	0
1	0	0.031440	4696	0	1	0	0	0	0	0	0	0	0	0	0
2	0	0.576636	3274	1	1	1	1	1	1	1	0	0	0	0	1
3	0	0.006433	2164	0	1	0	0	0	0	0	0	0	0	0	0
4	1	0.989105	1667	1	1	1	1	1	1	1	1	1	1	1	1