SUBJECTIVE QUESTIONS

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

   ANSWER 1 : The top three variables that contribute the most towards the probability of a lead getting converted are as follows:

	coef
const	-1.5878
Total Time Spent on Website	3.8458
Page Views Per Visit	-1.8670
Lead Origin_Lead Add Form	2.9158
Tags_Closed by Horizzon	7.0798
Tags_Interested in other courses	-1.9577
Tags_Lost to EINS	6.0637
Tags_Other_Tags	-2.5988
Tags_Ringing	-3.8487
Tags_Will revert after reading the email	4.5059
Tags_switched off	-5.0215
Last Activity_SMS Sent	2.0162
Last Notable Activity_Modified	-1.7203
Last Notable Activity_Olark Chat Conversation	-1.6529

   Fig : 1, Variables and their coefficients

a. Tags_Closed by Horizzon
b. Tags_Lost to EINS
c. Tags_Will revert after reading the email

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?**

   ANSWER 2 : Again, based on the coefficient values from the screen shot in the question above, the following are the top three categorical/dummy variables that should be focused the most in order to increase the probability of lead conversion :

a. Tags_Closed by Horizzon
b. Tags_Lost to EINS

c. Tags_Will revert after reading the email

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.**

ANSWER 3 : To target the number of actual conversions correctly predicted by the model we would have to use the **concept of Sensitivity**.

Sensitivity = True Positives/ (True Positives + False Negatives)

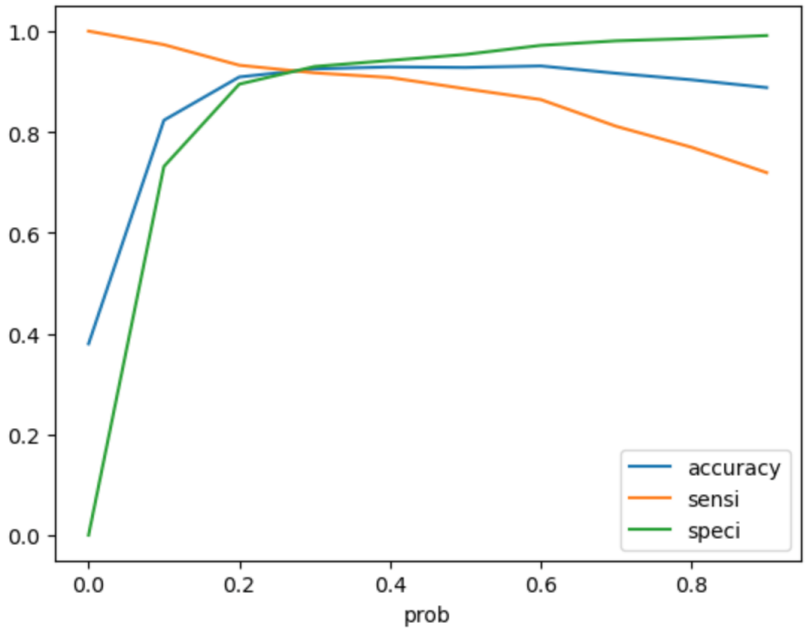

Fig : 2, Changes in Sensitivity, Specificity and Accuracy in the model.

The above figure shows the changes in Sensitivity, Specificity and Accuracy in the model with the changes in the level threshold.

In our model, **Sensitivity refers to the number of actual conversions predicted correctly out of the total number of actual conversions**. There could be different values of Sensitivity which can be achieved by varying the threshold for probability of lead conversion.

In the above figure, it is clear that **Sensitivity falls as the threshold increases**. In this given case of converting higher leads we would need a **high sensitivity,** as it would

reflect that our model would correctly predict almost all the leads that are likely to get converted. For this purpose we would need to **choose a lower threshold value**.

The idea is to contact the leads which have a conversion probability (value = 1). In the below image, the final prediction is calculated based on an optimal cut off value of 0.41.

	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	Lead_Score	final_predicted
0	1	0.998139	5647	1	1	1	1	1	1	1	1	1	1	1	1	100	1
1	0	0.004723	8449	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0.549778	1414	1	1	1	1	1	1	1	0	0	0	0	1	55	1
3	0	0.815512	8927	1	1	1	1	1	1	1	1	1	1	0	1	82	1
4	0	0.086321	3115	0	1	0	0	0	0	0	0	0	0	0	0	9	0

Fig : 3.

The only problem in this approach could be, it might **overestimate and misclassify some of the non-conversions as conversions.**

Also, according to the results of the model, phone calls could to be made to people who :

a. People who spend a lot of time on the website. The website needs to be made more insightful and interactive to get people to spend more time there, that will help in assisting lead conversion.
b. Making follow-up calls to people who are listed as Tags_Closed by Horizzon, Tags_Lost to EINS or Tags_Will revert after reading the email, as these people show a higher conversion rate according to the results of the model.
c. People whose last activity was through SMS are also good target customers as they have higher conversion potential.
d. Leads who have been obtained from Lead Origin ‘Lead Add Form’ are also potential customers since they have higher conversion rates than others.

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.**

ANSWER 4 : In this case, the concept of Specificity needs to be utilised.

Specificity = True Negatives/ (True Negatives + False Positives)

In our model, specificity can be **defined as the number of actual non-conversions predicted correctly out of the total number of actual non-conversions.** From Fig : 2, it can be seen that the specificity increases as the threshold increases.

In this given question, we would need to consider a high specificity value as **high specificity value would reflect the model correctly predicting almost all the leads who are not likely to get converted.**
Since, the company has already reached its target for the quarter, they do not want to make too many phone calls, the good strategy here would be to **opt for a high specificity value.** It will ensure that the phone calls are only made to customers who have a very high probability of conversion. To achieve **high specificity, we need to choose a high threshold value.**

Here the only problem could be, **the model may misclassify some of the conversions as non-conversions.**