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

Ans: Based on the coefficient values from the below screenshot, the following are the top three variables that contribute most towards the probability of a lead getting converted :

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

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.2383	0.082	-15.040	0.000	-1.400	-1.077
Total Time Spent on Website	0.8423	0.052	16.108	0.000	0.740	0.945
Lead_Origin_Lead Add Form	1.9254	0.444	4.336	0.000	1.055	2.796
Lead Source_Welingak Website	3.8518	1.110	3.470	0.001	1.676	6.027
Do Not Email__1	-0.9424	0.238	-3.967	0.000	-1.408	-0.477
Last Activity_SMS Sent	1.9438	0.114	17.098	0.000	1.721	2.167
Specialize_Travel and Tourism	-1.2905	0.444	-2.905	0.004	-2.161	-0.420
Current Occupation_Working Professional	0.8281	0.403	2.056	0.040	0.039	1.618
Tags_Busy	0.4499	0.224	2.006	0.045	0.010	0.889
Tags_Closed by Horizzon	6.9605	1.019	6.828	0.000	4.963	8.958
Tags_Interested in other courses	-2.0624	0.399	-5.167	0.000	-2.845	-1.280
Tags_Lost to EINS	5.6820	0.609	9.323	0.000	4.488	6.876
Tags_Other_Tags	-2.4164	0.206	-11.706	0.000	-2.821	-2.012
Tags_Ringing	-3.4947	0.238	-14.672	0.000	-3.962	-3.028
Tags_Will revert after reading the email	4.3157	0.192	22.450	0.000	3.939	4.692
Last Notable Activity_Email Link Clicked	-1.3343	0.526	-2.535	0.011	-2.366	-0.303
Last Notable Activity_Modified	-1.6048	0.126	-12.763	0.000	-1.851	-1.358
Last Notable Activity_Olark Chat Conversation	-1.6023	0.482	-3.326	0.001	-2.547	-0.658

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: Again, based on the coefficient values from the screen shot in the response 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 :

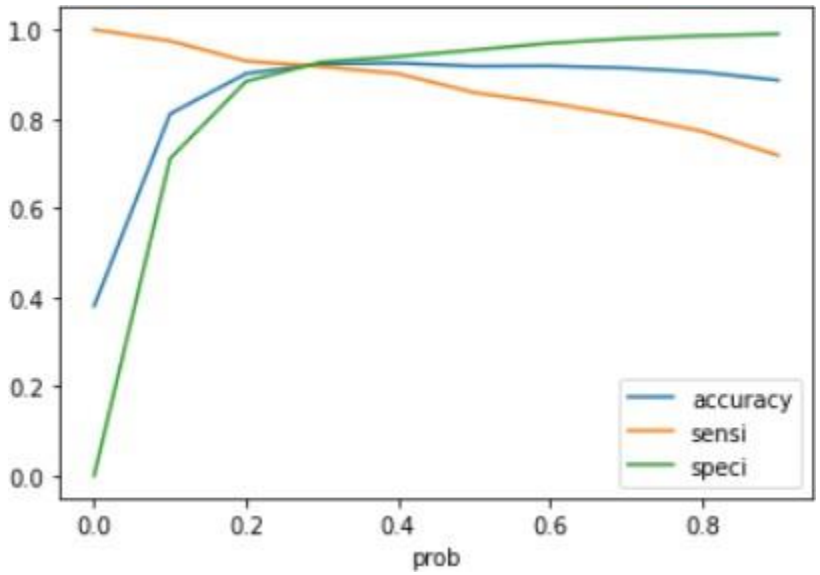
a) Tags_Closed by Horizzon
b) Tags_Lost to EINS
c) Tags_Will revert after reading the email

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.

Ans:
Sensitivity with respect to our model can be defined as the ratio of total number of actual Conversions correctly predicted to the total no of actual Conversions.
Similarly, **Specificity** can be defined as the ratio of total number of actual non-Conversions correctly predicted to the total number of actual non-Conversions.

For a particular 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. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.

High sensitivity implies that our model will correctly identify almost all leads who are likely to Convert. It will do that by over-estimating the Conversion likelihood, i.e. it will misclassify some non-Conversion cases as Conversions.

Now, since X Education has more man-power for these 2 months and they wish 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**.
This will ensure the Sensitivity rating is very high, which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible.

In the below image, the final prediction is calculated based on an optimal cut off value of 0.37.

In order to make the sales aggressive, the company may contact all the leads which have a conversion probability (value = 1) under a cut off 0.3 (column 0.3 highlighted in yellow).

	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
0	1	0.2257	9196	0	1	1	1	0	0	0	0	0	0	0	0	23
1	0	0.0562	4696	0	1	0	0	0	0	0	0	0	0	0	0	6
2	0	0.5410	3274	1	1	1	1	1	1	1	0	0	0	0	1	54
3	0	0.0065	2164	0	1	0	0	0	0	0	0	0	0	0	0	1
4	1	0.9843	1667	1	1	1	1	1	1	1	1	1	1	1	1	98
5	0	0.1221	7024	0	1	1	0	0	0	0	0	0	0	0	0	12
6	0	0.0272	8018	0	1	0	0	0	0	0	0	0	0	0	0	3
7	0	0.1903	778	0	1	1	0	0	0	0	0	0	0	0	0	19
8	0	0.0025	6942	0	1	0	0	0	0	0	0	0	0	0	0	0
9	0	0.1798	4440	0	1	1	0	0	0	0	0	0	0	0	0	18

From business knowledge perspective, to achieve maximum conversion, phone calls must be done to the all the people with a lead score from **40 to 100** and primarily if:

- *They spend a lot of time in the website and this can be done by making the website interesting and thus bringing them back to the site.*
- *They are seen coming back to the website repeatedly*
- *Their last activity is through SMS or through Olark chat conversation*
- *They are working professionals*

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.

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

Therefore, since 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**.
In our case, the company may contact all the leads which have a conversion probabilty (value = 1 highlighted in yellow color) under column 0.7. However, the flipside here would be that, we may miss out on those leads that are actually converted but then the model wrongly predicted them as not converted. (See red highlights in the image below). This should not be a major cause for concern as the target has already be 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	Lead_Score
0	1	0.2257	9196	0	1	1	1	0	0	0	0	0	0	0	0	23
1	0	0.0562	4696	0	1	0	0	0	0	0	0	0	0	0	0	6
2	0	0.5410	3274	1	1	1	1	1	1	1	0	0	0	0	1	54
3	0	0.0065	2164	0	1	0	0	0	0	0	0	0	0	0	0	1
4	1	0.9843	1667	1	1	1	1	1	1	1	1	1	1	1	1	98
5	0	0.1221	7024	0	1	1	0	0	0	0	0	0	0	0	0	12
6	0	0.0272	8018	0	1	0	0	0	0	0	0	0	0	0	0	3
7	0	0.1903	778	0	1	1	0	0	0	0	0	0	0	0	0	19
8	0	0.0025	6942	0	1	0	0	0	0	0	0	0	0	0	0	0
9	0	0.1798	4440	0	1	1	0	0	0	0	0	0	0	0	0	18
10	1	0.1221	4393	0	1	1	0	0	0	0	0	0	0	0	0	12
11	0	0.2023	989	0	1	1	1	0	0	0	0	0	0	0	0	20
12	1	0.7787	7177	1	1	1	1	1	1	1	1	1	0	0	1	78
13	0	0.0025	8898	0	1	0	0	0	0	0	0	0	0	0	0	0
14	1	0.9986	2714	1	1	1	1	1	1	1	1	1	1	1	1	100
15	1	0.9769	4881	1	1	1	1	1	1	1	1	1	1	1	1	98
16	1	0.8473	2900	1	1	1	1	1	1	1	1	1	1	0	1	85
17	1	0.9123	1067	1	1	1	1	1	1	1	1	1	1	1	1	91
18	0	0.0017	8752	0	1	0	0	0	0	0	0	0	0	0	0	0
19	0	0.0011	6948	0	1	0	0	0	0	0	0	0	0	0	0	0
20	0	0.0051	509	0	1	0	0	0	0	0	0	0	0	0	0	1
21	1	0.9701	2197	1	1	1	1	1	1	1	1	1	1	1	1	97
22	0	0.1221	3129	0	1	1	0	0	0	0	0	0	0	0	0	12
23	0	0.0224	7976	0	1	0	0	0	0	0	0	0	0	0	0	2
24	1	0.9968	6999	1	1	1	1	1	1	1	1	1	1	1	1	100
25	1	0.9755	1010	1	1	1	1	1	1	1	1	1	1	1	1	98

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.

From the business perspective, the company need to focus more on hot leads with probability score *higher than 90* and *other methods like automated emails and SMS*. This way manual calling won't be required unless it is an emergency. The above strategy can be used but with the customers that have a very high chance of buying the course.