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	-1.8488	0.111	-16.725	0.000	-2.065	-1.632
Total Time Spent on Website	0.9050	0.057	15.846	0.000	0.793	1.017
Lead_Origin_Lead Add Form	1.3654	0.461	2.963	0.003	0.462	2.269
Lead Source_Direct Traffic	-0.7038	0.136	-5.170	0.000	-0.971	-0.437
Lead Source_Welingak Website	2.5581	1.119	2.286	0.022	0.365	4.751
Do Not Email__1	-1.0457	0.255	-4.107	0.000	-1.545	-0.547
Last Activity_SMS Sent	2.0560	0.126	16.303	0.000	1.809	2.303
Specialize_Travel and Tourism	-1.0308	0.477	-2.163	0.031	-1.965	-0.097
Current Occupation_Student	2.3556	0.591	3.987	0.000	1.198	3.514
Current Occupation_Unemployed	2.4615	0.156	15.818	0.000	2.156	2.766
Current Occupation_Working Professional	3.0427	0.428	7.107	0.000	2.204	3.882
Tags_Busy	-1.3058	0.257	-5.085	0.000	-1.809	-0.802
Tags_Closed by Horizzon	5.3722	1.032	5.205	0.000	3.349	7.395
Tags_Interested in other courses	-3.9164	0.423	-9.261	0.000	-4.745	-3.088
Tags_Lost to EINS	5.0095	0.629	7.959	0.000	3.776	6.243
Tags_Other_Tags	-4.0806	0.243	-16.827	0.000	-4.556	-3.605
Tags_Ringing	-5.2198	0.273	-19.095	0.000	-5.756	-4.684
Tags_Will revert after reading the email	2.7076	0.220	12.290	0.000	2.276	3.139
Last Notable Activity_Email Link Clicked	-1.2420	0.533	-2.331	0.020	-2.286	-0.198
Last Notable Activity_Modified	-1.3937	0.132	-10.518	0.000	-1.653	-1.134
Last Notable Activity_Olark Chat Conversation	-2.1034	0.543	-3.874	0.000	-3.168	-1.039

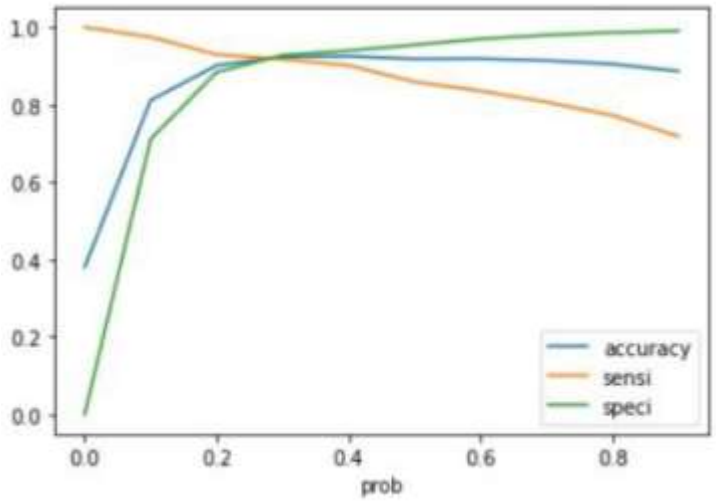
2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

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

	Converted	Converted_prob	Prospect ID	final_Predicted	Lead_Score
0	1	0.2854	9196	0	29
1	0	0.0326	4696	0	3
2	0	0.5796	3274	1	58
3	0	0.0069	2164	0	1
4	1	0.9883	1667	1	99

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.

Ans: 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
0	1	0.2854	9196	0	1	1	1	0	0	0	0	0	0	0	0
1	0	0.0326	4696	0	1	0	0	0	0	0	0	0	0	0	0
2	0	0.5796	3274	1	1	1	1	1	1	1	0	0	0	0	1
3	0	0.0069	2164	0	1	0	0	0	0	0	0	0	0	0	0
4	1	0.9883	1667	1	1	1	1	1	1	1	1	1	1	1	1
5	0	0.1368	7024	0	1	1	0	0	0	0	0	0	0	0	0
6	0	0.0260	8018	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0.2144	778	0	1	1	1	0	0	0	0	0	0	0	0
8	0	0.0032	6942	0	1	0	0	0	0	0	0	0	0	0	0
9	0	0.1020	4440	0	1	1	0	0	0	0	0	0	0	0	0
10	1	0.1368	4393	0	1	1	0	0	0	0	0	0	0	0	0
11	0	0.2292	989	0	1	1	1	0	0	0	0	0	0	0	0
12	1	0.8123	7177	1	1	1	1	1	1	1	1	1	1	0	1
13	0	0.0032	8898	0	1	0	0	0	0	0	0	0	0	0	0
14	1	0.9978	2714	1	1	1	1	1	1	1	1	1	1	1	1
15	1	0.9616	4881	1	1	1	1	1	1	1	1	1	1	1	1
16	1	0.8878	2900	1	1	1	1	1	1	1	1	1	1	0	1
17	1	0.9284	1067	1	1	1	1	1	1	1	1	1	1	1	1
18	0	0.0025	8752	0	1	0	0	0	0	0	0	0	0	0	0
19	0	0.0007	6948	0	1	0	0	0	0	0	0	0	0	0	0
20	0	0.0054	509	0	1	0	0	0	0	0	0	0	0	0	0
21	1	0.9771	2197	1	1	1	1	1	1	1	1	1	1	1	1
22	0	0.1368	3129	0	1	1	0	0	0	0	0	0	0	0	0
23	0	0.0128	7976	0	1	0	0	0	0	0	0	0	0	0	0
24	1	0.9968	6999	1	1	1	1	1	1	1	1	1	1	1	1
25	1	0.9836	1010	1	1	1	1	1	1	1	1	1	1	1	1
26	0	0.0030	5627	0	1	0	0	0	0	0	0	0	0	0	0
27	0	0.0260	7572	0	1	0	0	0	0	0	0	0	0	0	0
28	0	0.0481	6549	0	1	0	0	0	0	0	0	0	0	0	0
29	0	0.0157	3889	0	1	0	0	0	0	0	0	0	0	0	0

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.