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

Ans:

The top three variables that contribute most toward the probability of the lead getting converted are

i. Tags_Lost to EINS (coeff – 10.33)
ii. Tags_Closed by Horizzon (coeff -7.72)
iii. Tags_Will revert after reading the email (coeff – 4.94)

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

Ans:

The top 3 dummy variables in the model that should be focused on most in order to increase the lead conversion probability are

- Tags_Lost to EINS
- Tags_Closed by Horizzon
- 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:

The objective of the problem is to find a high True positive rate, that is high sensitivity.

In the statistical approach, if we fix the threshold value of probability conversion between 0.2 and 0.3, the sensitivity rate will be quite high (approx 0.94) with a decent specificity rate (approx 0.89) and accuracy(0.91).

	Prob	Accuracy	Sens	Spec
0.0	0.0	0.385136	1.000000	0.000000
0.1	0.1	0.883955	0.952576	0.840973
0.2	0.2	0.909148	0.947261	0.885275
0.3	0.3	0.914029	0.939084	0.898335
0.4	0.4	0.912140	0.934178	0.898335
0.5	0.5	0.919225	0.868765	0.950832
0.6	0.6	0.919225	0.868357	0.951088
0.7	0.7	0.904897	0.779640	0.983355
0.8	0.8	0.903322	0.773508	0.984635
0.9	0.9	0.897811	0.758790	0.984891

Keeping the conversion rate between 0.2 and 0.3 enables the interns to identify and contact the leads who tend to become hot leads at the maximum in the given limited period. That is lower the threshold value, the higher the Sensitivity.

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:

Since the company already reached its target, its prime aim is not to make unnecessary phone calls. That is, the objective of this problem is that the company should not contact any cold leads, whereas they can miss hot leads in this process as they already achieved the target.

In this problem, we have to find a high specificity rate to achieve the solution.

	Prob	Accuracy	Sens	Spec
0.0	0.0	0.385136	1.000000	0.000000
0.1	0.1	0.883955	0.952576	0.840973
0.2	0.2	0.909148	0.947261	0.885275
0.3	0.3	0.914029	0.939084	0.898335
0.4	0.4	0.912140	0.934178	0.898335
0.5	0.5	0.919225	0.868765	0.950832
0.6	0.6	0.919225	0.868357	0.951088
0.7	0.7	0.904897	0.779640	0.983355
0.8	0.8	0.903322	0.773508	0.984635
0.9	0.9	0.897811	0.758790	0.984891

As we saw the above table, it is understood that higher the conversion cutoff/threshold value the higher the Specificity. If we fix the threshold value between 0.7 and 0.8, we will get specificity rate of 98% with a good decent value for sensitivity and accuracy.