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

Ans: The top 3 variables which came out to be most important are:

* Tags
* Lead Source
* Total\_Time\_Spent

1. **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: The Top 3 categorical variables in the model would be focused in order to increase the probability of lead conversion are:

* Tags\_Closed by Horizzon
* Tags\_Lost to EINS
* Tags\_Will revert after reading the email

1. **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: We can design a Logistic Regression model which will predict the probability of each instance. So, based on that probability, we can determine a score for each instance. Based on a cut-off score we can identify which leads have score more than that and hence we can determine the interns as potential lead.

1. **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: We can make a logistic regression model which can assign probability for conversion for each instance and based on that we can determine a lead score for each instance. From those scores we can select a cutoff and identify those instances which have high score than the cut-off. Thus the potential leads can be identified. Now we can see from the historical records if those potential leads are interested in phone call or not. Based on their interest level we can manipulate the number of phone calls and along with that we need to check if they are interested in getting mails. If they are interested in getting mails, then we can use this alternative for contact other than phone calls. Thus we can reduce the number of phone calls.