Subjective Answers

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

A. Nomenclature:
   Positive contribution - leads having this parameter in their favor are more likely to convert
   Negative contribution - leads having this parameter in their favor are less likely to convert
   The 3 variables which contribute most towards the probability of a lead getting converted (in descending order)-

   a) 'lead_source_Reference' - Denotes leads which get in through a reference (positive). The leads which come in through a reference are more likely to convert.
   b) 'lead_source_Other sources' - Denotes leads which get in through mostly Lead Add forms, along with 'Lead import' and 'Quick Add Form' in small representation (positive). Leads which come in through Lead Add Forms are more likely to convert
   c) 'do_not_email' - Leads which choose not to get e-mails from the company regarding the course (negative). Leads which choose not to receive information about the course via the email are less likely to convert.

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?**

A. Categorical/dummy variables which should be focused the most on in order to increase the probability of lead
   conversion (in descending order of significance) -
   a) 'lead_source_Reference' - As discussed before, this denotes leads which get in through a reference (positive). The leads which come in through a reference are more likely to convert.
   b) 'lead_source_Other sources' - Denotes leads which get in through mostly Lead Add forms, along with 'Lead import' and 'Quick Add Form' in small representation (positive). Leads which come in through Lead Add Forms are more likely to convert
   c) 'lead_source_Olark Chat' - denotes leads which get in through Olark Chat (positive). The leads which come in through this path are more likely to convert

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.**

A. During this phase, where the company can afford to call as many people as possible to secure leads, we have to maximize 'recall' in our model. This effectively means that out of all the leads which are likely to convert, we try to get to most of them. This requires us to lower the cutoff value of 'lead_score' (which tells us about the probability that a lead will convert) from 70 (where it is set to maximize precision) to something around 30 or lesser (going past the point of best trade-off between precision and recall on the precision-recall graph). This allows us to get to more leads which could convert despite the low probability. Doing this allows the sales team to cover more ground and get more potential customers in.

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.**

A. During this phase, since the company wants to minimize the resources spent on calling people individually, our goal is to get to the 'hot leads' i.e. the ones which have the greatest probabilities of converting. This requires us to make sure that out of all the leads we predict will convert, most of them do convert. Hence, we effectively have to maximize precision. Our model has been in such a way (with cutoff point of lead score at 70) that the precision we get is roughly 80%. We could raise the cutoff to get greater precision. Doing this would allow the sales team to focus their resources to get the most out of what they're putting in.