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

   **Ans:** Following are the top three variables in the model
   a. **TotalVisits** - The total number of visits made by the customer on the website. This positively effects conversion rate.
   b. **Lead Quality** - The quality effects the conversion rate. The value “worst” negatively affect the conversion strongly, also values not sure, might be negatively effect conversion.
   c. **Total Time Spent on Website** - The total time spent by the customer on the website. This positively effects target variable.

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: Following are the top three dummy/categorical variables:
   a. Column: “**Lead Quality**”, Value: “**Worst**”, “**Not Sure**”, “**May be**” – have high negative conversion rate, should be avoided.
   b. Column: “**Lead Source**”, Value: “**Welingak Website**” – have positive conversion rate, should follow up.
   c. Column: **Last Activity**, Value: **“Had a Phone Conversation**” – have positive conversion rate, should follow up.

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**:
   a. **Reduce the Model's Threshold**: Lower the cutoff probability threshold (cutoff_prob variable in notebook/code) threshold of the logistic regression model from the 0.4 to a lower value of 0.3 or 0.2 depending on how aggressive the campaign should be. This will classify more leads as "high potential," ensuring that a greater number of leads are flagged for follow-up.
   b. **Follow the Lead_Score**: The lead score (100 to 0) column in the model output is the indicator of how much potential is there for conversion. Higher value indicates higher probability of conversion. Sales people can start from highest lead score and work their way down.

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:

a. **Increase the Model's Threshold**: Increase the cutoff probability threshold (cutoff_prob variable in notebook/code) threshold of the logistic regression model from the 0.4 to a higher value of 0.5, 0.6, 0.7 or higher value, depending on how need of high potential of the leads. Higher the threshold, lower the number of hot leads will be, and higher will be their conversion potential.
b. **Follow the Lead_Score**: The lead score (100 to 0) column in the model output is the indicator of how much potential is there for conversion. Higher value indicates higher probability of conversion. Sales people can start from highest lead score and work their way down. Follow only the leads that are marked as hot leads in last step and start with highest lead score among them work move towards lower lead score.