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 which contribute most towards the probability of lead getting converted are:

- Total Time Spent on Website – positive coefficient
- Lead Origin_Lead Add Form – positive coefficient
- Last Notable Activity_Unreachable – positive coefficient

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 – The top three 3 categorical/dummy variables in the model which should be focused the most on to increase the probability of lead conversion are:

- Lead Origin_Lead Add Form
- Last Notable Activity_Unreachable
- What is your current occupation_Working Professional

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 default logistic regression classification threshold is 0.4 (i.e., if the predicted probability of conversion P(y=1)>0.4, the lead is classified as 1).
However, to make the model more aggressive in identifying potential leads, the sales team can lower the threshold, for example, to 0.3. This will increase the number of leads predicted as 1, ensuring fewer potential leads are missed.

Based on the coefficients of the logistic regression model, features with larger coefficients have a stronger influence on lead conversion. Focus the outreach efforts on leads exhibiting these characteristics:

- Total Time Spent on Website (4.43): Leads spending more time on the website are significantly more likely to convert. Sales interns can prioritise contacting such leads.
- Lead Origin_Lead Add Form (3.94): Leads who originated from the "Lead Add Form" should be contacted early.
- Last Notable Activity_Unreachable (3.08): Though initially unreachable, these leads are still likely to convert upon further follow-up.

The team should also deprioritize leads with strongly negative coefficients:

- Lead Profile_Student of SomeSchool (-2.44): Students are less likely to convert and can be deprioritized.
- Last Activity_Email Bounced (-1.78): Leads with bounced emails are less promising.

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 - To minimize the number of unnecessary phone calls when the company has already achieved its quarterly target, the sales team can adopt a highly precision-focused strategy. This involves concentrating only on the most promising leads, thereby reducing the number of calls while maintaining the likelihood of successful conversions.

Increase the Classification Threshold:

- Unlike the aggressive strategy (lowering the threshold), here the threshold for classifying leads as 1 (likely to convert) should be increased (e.g., from 0.4 to 0.7 or higher).
- This ensures that only leads with a high probability of converting (e.g., $P(y=1) \geq 0.7$) are considered for phone calls, reducing false positives (leads classified as 1 that don't convert).

Utilize the logistic regression model's coefficients to identify features most strongly associated with successful conversions:

- Prioritize leads with characteristics tied to large positive coefficients, such as:
  - Total Time Spent on Website (4.43).
  - Lead Origin_Lead Add Form (3.94).
  - Last Notable Activity_Unreachable (3.08).
- Deprioritize or ignore leads with features associated with strongly negative coefficients, such as:
  - Lead Profile_Student of SomeSchool (-2.44).
  - Last Activity_Email Bounced (-1.78).