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 in the model contributing most to the probability of lead conversion are Total time spend on website, Last activity, and Asymmetrique Profile Square. These variables represent the features with the highest impact on conversion likelihood, guiding lead conversion strategies effectively.

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 categorical variables contributing most to conversion probability are:
   Lead Source
   Specialization
   How did you hear about X Education
   These variables have the highest average conversion rates across all categories within each variable, indicating their significance in predicting lead conversion. It's important to focus on these variables to increase the probability of lead conversion

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. During the aggressive lead conversion phase coinciding with the intern hiring period, X Education aims to maximize lead conversion by making phone calls to nearly all potential leads identified by the model. Here's how the provided information addresses the strategy for this phase:

- Threshold Adjustment: The threshold is set to 0.3, indicating a lower cutoff point for classifying leads as positive. By lowering the threshold, the model classifies more leads as positive, thereby maximizing sensitivity and capturing almost all potential leads for conversion.
- Sensitivity: The sensitivity value obtained is approximately 0.82, indicating that the model correctly identifies around 82% of actual positive cases as positive during this phase. This high sensitivity ensures that the company captures the majority of potential leads for conversion, aligning with the goal of aggressive lead conversion.
- Specificity: The specificity value obtained is around 0.61, indicating that the model correctly identifies approximately 61% of actual negative cases as negative during this phase. While specificity is lower, it is acceptable given the emphasis on maximizing sensitivity and capturing as many potential leads as possible.

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. During the phase aimed at minimizing useless phone calls after meeting quarterly targets, the company should adjust the threshold of the logistic regression model to maximize specificity. By setting a higher threshold, the model will classify only the most confident leads as positive, reducing the occurrence of false positives and unnecessary phone calls. Here's how the provided information addresses the strategy for this phase:

- Threshold Adjustment: The threshold is set to 0.7, indicating a higher cutoff point for classifying leads as positive. This ensures that only leads with the highest predicted probabilities are identified as positive by the model.
- Specificity: The specificity value obtained is approximately 0.94, indicating that the model correctly identifies around 94% of actual negative cases as negative during this phase. This high specificity ensures that the company minimizes the rate of useless phone calls by focusing only on the most confident leads.
- Sensitivity: The sensitivity value obtained is around 0.29, indicating that the model captures approximately 29% of actual positive cases during this phase. While sensitivity is lower, it is acceptable given the primary objective of minimizing useless phone calls and focusing on new work after meeting quarterly targets.