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

    1. **Lead origin_Lead Add Form**:
       This is the largest coefficient in the model, indicating it has the strongest positive effect on the likelihood of conversion.
    2. **Occupation_Working Professional**:
       This variable also has a large positive coefficient, indicating a strong positive contribution to the probability of conversion.
    3. **Last activity_Had a Phone Conversation**:
       This variable has a significant positive effect on the likelihood of conversion, making it the third most impactful predictor.

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?

    1. **Lead origin_Lead Add Form**:
       This variable has the highest positive coefficient among all categorical variables, indicating that leads originating from a Lead Add Form have a very strong positive effect on conversion probability. Focusing on optimizing and increasing the number of leads from this source should significantly improve conversion rates.
    2. **Occupation_Working Professional**:
       This variable has a significant positive effect on conversion probability. Tailoring our marketing and sales strategies to attract working professionals can enhance the chances of conversion.
   3. **Last activity_Had a Phone Conversation**:
       This variable shows a strong positive impact on conversion probability. Ensuring that leads who have had a phone conversation are prioritized and effectively followed up with can improve conversion rates.

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.

    In the context of optimizing lead conversion during a period when X Education hires interns, focusing on **sensitivity** and **specificity** can help in efficiently using resources to maximize conversions while minimizing wasteful efforts.

Sensitivity (True Positive Rate): The proportion of actual positives (leads that will convert) that are correctly identified by the model. High sensitivity means the model identifies most leads who are likely to convert.

Specificity (True Negative Rate): The proportion of actual negatives (leads that will not convert) that are correctly identified as such by the model. High specificity means the model correctly identifies leads who are unlikely to convert.
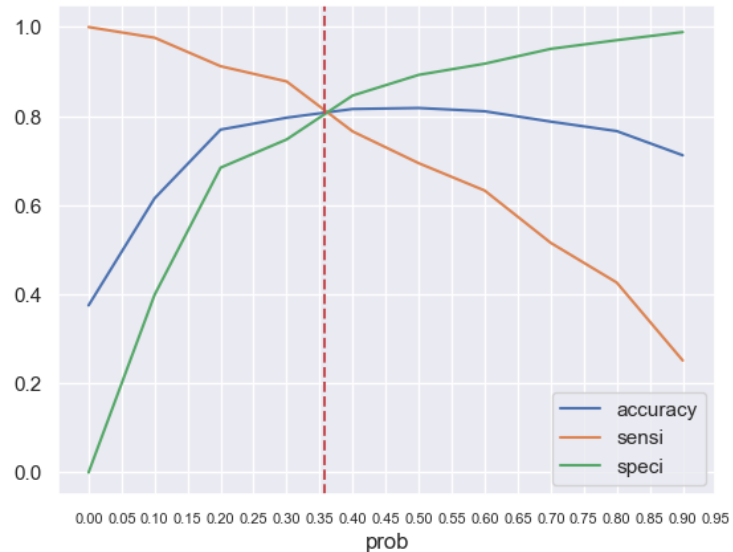
Since the goal is to convert as many high-potential leads as possible, we should prioritize maximizing **sensitivity** during the internship period. This means:

- **Adjust the Threshold for Classification**: Lower the probability threshold used to classify leads as "high likelihood to convert." Since we have used 0.358, We could reduce this value. This adjustment will increase sensitivity, allowing us to capture more leads predicted to convert, which aligns with our aggressive conversion goal.

- **Focus on High-Prediction Leads**: Direct interns to contact all leads predicted with a high probability of conversion. Even though this might include some leads with a lower probability of conversion, the primary aim is to ensure that as many potential converters are reached 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.

Specificity refers to the model's ability to correctly identify leads that are unlikely to convert. To minimize useless calls, we need to ensure that calls are made only to leads who have a high likelihood of converting. This involves:

- **Raise the Probability Threshold**: Increase the threshold used by our model to classify a lead as a high-conversion candidate. For example, the current threshold is 0.358, we can raise it to 0.5 or 0.6. This adjustment means that only leads with a very high predicted probability of conversion will be contacted, thereby reducing the number of calls to leads who are less likely to convert.

- **Filter Based on lead score**: Utilize lead score developed by the model to further filter leads. Focus on leads with the higher lead scores in their predicted probability of conversion. This reduces the risk of contacting leads who are less likely to convert.

Sensitivity measures the model's ability to identify leads who will convert. While we want to increase specificity to avoid useless calls, managing sensitivity ensures we don't miss out on potential high-conversion leads:

- **Optimize Sensitivity Balance**: Ensure that in raising specificity, we do not excessively lower sensitivity. Find a balance where we maintain a reasonable number of high-probability leads while avoiding too many false positives.