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

The top three variables that contribute the most towards the probability of a lead getting converted are:

1. Lead Source
2. Total Time Spent on Website
3. Current occupation
4. **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?**

The top three categorical/dummy variables that should be focused on the most to increase the probability of lead conversion are:

1. Lead Source\_Welingak Website
2. Lead Source\_Reference
3. Current occupation\_Working Professional
4. **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.**

During the 2-month intern hiring period where X Education wishes to make lead conversion more aggressive:

**a. Strategy:**

The company aims to convert as many potential leads as possible (those predicted as 1 by the model). To achieve this, they should focus on capturing a high number of true positives while allowing for a potentially higher number of false positives. The goal is to maximize sensitivity (recall) while accepting a lower precision.

**b. Implementation:**

The company should consider using a lower cutoff threshold to identify potential leads to contact. In the provided data, a cutoff value of 0.3 seems appropriate for this strategy. This means that leads with a conversion probability of 0.3 or higher will be considered for aggressive follow-up.

**Impact on Metrics:**

* Accuracy: The accuracy might decrease slightly due to an increase in false positives, but the focus is on maximizing lead conversion.
* Sensitivity: Sensitivity (recall) will increase significantly. The company will capture a higher proportion of true positives, increasing the conversion rate.
* Specificity: Specificity will likely decrease, as there might be more false positives predicted.
* Precision: Precision may decrease due to the increased number of false positives, but the focus is on conversion rate rather than precision.
* Recall: Recall will increase as the model captures more true positives.

**c. Outcome:**

Implementing this strategy would lead to a higher number of phone calls made to potential leads, resulting in a higher likelihood of converting those leads into customers. While there might be an increase in false positives (leads that are predicted as converted but don't actually convert), the primary goal is to maximize conversions during this aggressive period.

1. **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.**

During the quarter when the company wants to minimize the rate of useless phone calls:

1. **Strategy:**

The company aims to minimize the number of phone calls while still achieving their targets. They want to avoid unnecessary calls to potential leads that are unlikely to convert. The focus is on maximizing precision while accepting a potentially lower sensitivity.

1. **Implementation:**

The company X should consider using a higher cutoff threshold to identify leads to contact. In the provided data, a cutoff value of 0.7 seems suitable for this strategy. This means that only leads with a conversion probability of 0.7 or higher will be considered for follow-up.

**Impact on Metrics:**

* Accuracy: The accuracy might increase slightly due to a decrease in false positives, but there might be a decrease in true positives.
* Sensitivity: Sensitivity (recall) will decrease, as the model becomes more selective in predicting positive outcomes.
* Specificity: Specificity will likely increase, as the model becomes more cautious in predicting positive outcomes.
* Precision: Precision will increase significantly due to a reduction in false positives, resulting in higher confidence in the positive predictions.
* Recall: Recall will decrease as the model captures fewer true positives.

**c. Outcome:**

Implementing this strategy would result in fewer phone calls made, focusing on leads with a higher likelihood of converting. While there might be a decrease in true positives, the company can be more confident that the leads they do contact are more likely to convert, thus improving efficiency and reducing the rate of useless phone calls.