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, as per the feature importance, are:

   - **What is your current occupation_Working Professionals**: This variable has the highest coefficient, indicating that leads who are working professionals are more likely to convert.
   - **Lead Origin_Lead Add Form**: The origin of the lead being from a lead add form is the second most influential feature.
   - **Lead Source_Welingak Website**: The third significant variable is the lead coming from the Welingak website, which means leads from this source have a higher likelihood of conversion.

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?

   The top categorical variables are essentially the same as the above since they are binary encoded:

   - Current occupation as Working Professionals
   - Lead Origin as Lead Add Form
   - Lead Source as Welingak Website

   Focusing on these aspects could mean targeting working professionals more aggressively, optimizing the lead add form process for higher engagement, and leveraging the Welingak website's potential to generate quality leads.

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.

   During the intern phase, since there is more manpower available, the strategy should be to:

   - Lower the threshold for predicting a potential lead to increase the pool of leads contacted.
   - Emphasize personalized outreach and follow-ups, especially for the leads classified as high potential by the model.

- Utilize the additional resources to engage in more in-depth conversations with leads, particularly focusing on the top contributing features identified by the model.

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.

   The strategy should be to:

   - Increase the threshold for lead prediction to focus only on the most likely candidates, thus reducing the number of calls.

   - Allocate resources to nurturing existing leads with the highest scores to ensure their conversion.

   - Invest in automated or less resource-intensive methods of keeping leads warm, such as targeted content delivery or automated follow-ups, reserving calls for only those above the higher threshold.

For both scenarios (3 and 4), the precision-recall curve can help determine the appropriate thresholds. A higher recall threshold is suitable for an aggressive approach, ensuring more potential leads are not missed. Conversely, a higher precision threshold is apt for a conservative approach, ensuring that the leads contacted have a higher probability of conversion, which minimizes effort and maximizes efficiency.