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

   Top Variables Contributing to Lead Conversion

   Lead Origin_Lead Add Form: Capturing lead-specific statuses, this variable showed the strongest contribution to conversion prediction.

   Last Notable Activity_Unreachable: Reflecting the latest interaction with the lead, it was critical in assessing engagement levels.

   Lead Source: Identified high-performing channels (e.g., referrals, Google ads) contributing to conversions.

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?

   Lead Origin_Lead Add Form: Strongly associated with successful conversions.

   Last Notable Activity_Unreachable: Important to follow up with unreachable leads via different mode of communication.

   Lead Source_Welingak Website: Highlighted leads from this website requiring proactive follow-ups.

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.

The goal is to aggressively maximize conversions by ensuring that almost all potential leads (predicted as 1) are contacted. To achieve this, **Sensitivity (Recall)** should be prioritized, even at the expense of an increased number of False Positives (non-converting leads). Here's the combined strategy:

1. Prioritize Sensitivity (Recall)
- **Sensitivity is Key:** Sensitivity measures the proportion of actual converted leads correctly identified. Missing potential converters (False Negatives) must be minimized in this phase.
- **Accept Lower Specificity:** While a decrease in Specificity will result in more False Positives, this is acceptable since the expanded team can handle the increased workload.

2. Use the ROC Curve to Adjust the Threshold
- The **ROC Curve** (Receiver Operating Characteristic) plots Sensitivity (True Positive Rate) against 1-Specificity (False Positive Rate). It helps in identifying an optimal threshold where Sensitivity is maximized.
- Lower the threshold (e.g., 0.3–0.4 instead of 0.5) to flag more leads as potential converters. This ensures that the vast majority of converting leads are identified.

3. Strategic Implementation
1. **Threshold Adjustment:** Based on the ROC analysis, adjust the threshold to a value that balances high Sensitivity with an acceptable rise in False Positives.
2. **Leverage Team Capacity:** The increased workforce (interns) can handle the rise in non-converting leads contacted due to the lower threshold. Allocate resources efficiently to follow up on the flagged leads.
3. **Track Performance Metrics:** Continuously monitor Sensitivity and False Positive Rates to ensure that

the model's predictions align with the aggressive lead conversion goal.

Outcome of This Approach
- **Maximized Conversions:** High Sensitivity ensures nearly all converting leads are identified, minimizing missed opportunities.
- **Increased Workload:** While more False Positives are flagged, the additional capacity of interns enables the sales team to manage the extra calls.
- **Optimal Lead Contact:** By leveraging the ROC curve to adjust the threshold, the balance between identifying more leads and the team's capacity is maintained.

This approach combines prioritizing Sensitivity with threshold adjustment using the ROC curve, ensuring X Education captures as many potential converters as possible during this aggressive phase. The additional effort from the sales team and interns will be well-justified by the increase in lead conversions.

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.

Strategy for Minimizing Unnecessary Calls

When X Education reaches its quarterly targets early, the goal is to minimize the rate of unnecessary calls while still focusing on high-probability leads. The company can adopt the following strategy:

1. Minimize the False Positive Rate (FPR)

- **Why FPR Matters:** The FPR represents the proportion of non-converting leads (false positives) mistakenly classified as high-probability and contacted unnecessarily. Minimizing the FPR will directly reduce wasted effort and resources.

- Adjust the **model threshold** to focus on minimizing FPR while maintaining reasonable Precision (the proportion of contacted leads who convert). This ensures the sales team only contacts leads that are genuinely likely to convert.

2. Use the Precision-Recall Curve for Threshold Optimization

- The **Precision-Recall Curve** is ideal for imbalanced datasets like this one, where the conversion rate is low (~30%).

- Optimize the threshold based on the trade-off between Precision (proportion of true conversions among predicted positive leads) and Recall (proportion of true conversions captured by the model).

- Choose a threshold that prioritizes **high Precision** with sufficient Recall to ensure that the number of unnecessary calls is minimized while capturing the most likely converters.

3. Employ Behavioral Data for Dynamic Reassessment

- Focus on **lead engagement metrics** to refine call priorities further:

- Leads who show renewed interest (e.g., open multiple emails, revisit the website, or view course details) can be reclassified for a follow-up call even if their initial probability score was lower.

- Use this dynamic engagement data to minimize calls to leads showing no activity.

4. Automate Engagement for Lower-Priority Leads

- Leads falling below the adjusted threshold can be nurtured through:

  - **Email drip campaigns** tailored to their interests.

  - Personalized SMS reminders about courses or benefits.

  - Social media remarketing efforts to maintain passive engagement.

5. Strategic Reallocation of Sales Team Efforts

During this low-call phase, the sales team can redirect their efforts towards:

- Reviewing and improving sales strategies and lead-nurturing scripts.

- Collaborating with the marketing team to identify trends and improve lead acquisition methods.

- Training and development to enhance their communication and sales effectiveness.

6. Predictive Call Assignment

- Utilize the model's **expected benefit for each lead** to prioritize calls:

  - Leads with the highest expected return based on their probability score are prioritized for calls.

  - Leads with low expected benefits are left to passive engagement methods.

7. Communication of Strategy

- Clearly communicate the updated threshold, focus on minimizing FPR, and use of engagement metrics to the sales team.
- Provide training on identifying "must-call" leads based on the adjusted strategy.