

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

```
Top features with the highest correlation with 'Converted':
Tags_dummy_Will revert after reading the email    0.644307
Lead Profile_dummy_Potential Lead                0.378061
Total Time Spent on Website                      0.362483
Last Notable Activity_dummy_SMS Sent             0.351845
Last Activity_dummy_SMS Sent                     0.325600
Lead Origin_dummy_Lead Add Form                  0.321702
What is your current occupation_dummy_Working Professional 0.313837
Lead Source_dummy_Reference                      0.270830
Tags_dummy_Closed by Horizon                    0.251229
Name: Converted, dtype: float64
```

💡 **Top 3 Features** - Highest positive coefficients: Indicate variables that strongly increase the probability of lead conversion. Based on typical scenarios:

- **Tags_dummy_Will revert after reading the email**: A specific lead source category showing a strong impact.
- **Lead Profile_dummy_Potential Lead**: Potential Lead are often **hot leads** and correlates with interest and leads to higher conversion.
- **Total Time Spent on Website**: Indicates higher engagement, which is a strong predictor 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?

```
Top features with the highest correlation with 'Converted':
Tags_dummy_Will revert after reading the email    0.644307
Lead Profile_dummy_Potential Lead                0.378061
Total Time Spent on Website                      0.362483
Last Notable Activity_dummy_SMS Sent             0.351845
Last Activity_dummy_SMS Sent                     0.325600
Lead Origin_dummy_Lead Add Form                  0.321702
What is your current occupation_dummy_Working Professional 0.313837
Lead Source_dummy_Reference                      0.270830
Tags_dummy_Closed by Horizon                    0.251229
Name: Converted, dtype: float64
```

💡 **Top 3 Categorical Features** - Highest positive coefficients: Indicate variables that strongly increase the probability of lead conversion. Based on typical scenarios:

- **Tags_dummy_Will revert after reading the email**: A specific lead source category showing a strong impact.
- **Lead Profile_dummy_Potential Lead**: Potential Lead are often **hot leads** and correlates with interest and leads to higher conversion.
- **Last Notable Activity_dummy_SMS Sent**: SMS sent is also showing strong predictor of conversion with 0.35

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.

💡 Insights from Model Confusion Matrix

- **Focus on Recall Improvement**: Adjust follow-up strategies for leads flagged as "Not Converted" to minimize missed opportunities.
- **Review False Positives**: Analyze characteristics of false positives to refine the lead scoring model further.
- **Sales Prioritization**: Assign higher priority to leads with strong probabilities of conversion (e.g., scores above 80%).

Prioritize High-Scoring Leads: Focus on leads with a predicted probability of conversion > 80% (high confidence leads). **Action:** Assign high-probability leads to experienced sales agents or interns for immediate follow-up.

Expand to Medium-Scoring Leads: Include leads with scores between 50% and 80% to increase reach. **Action:** Distribute these leads among interns for personalized phone calls.

Follow-Up Tracking: Monitor engagement metrics post-phone calls to ensure effectiveness and identify further actionable steps.

💡 **Insights from Model Confusion Matrix**

- **Focus on Recall Improvement:** Adjust follow-up strategies for leads flagged as "Not Converted" to minimize missed opportunities.
- **Review False Positives:** Analyze characteristics of false positives to refine the lead scoring model further.
- **Sales Prioritization:** Assign higher priority to leads with strong probabilities of conversion (e.g., scores above 80%).

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.

Sales Process Optimization and Efficient Use of CRM Tools: Integrate the lead scoring system into the CRM and tag leads by probability tiers to streamline follow-up.

Leverage Automated Campaigns: For medium and low-probability leads, send targeted emails or SMS instead of direct calls.