

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

- **Lead Origin_Lead Add Form:** Coefficient = 2.14 (strong positive influence). Leads from this origin have the highest likelihood of conversion, showing its critical importance as a lead generation channel
- **Total Time Spent on Website:** Coefficient = 1.13 (positive influence). Leads who spend more time on the website are significantly more likely to convert. The boxplot analysis also confirms that converted leads tend to spend much more time on the website than non-converted leads
- **Do Not Email_Yes:** Coefficient = -1.32 (strong negative influence). Leads marked as "Do Not Email" are much less likely to convert, highlighting the necessity of direct communication

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:** With a coefficient of 2.14, this variable has the strongest positive impact. X Education should optimize the "Lead Add Form" experience by simplifying it and reducing submission friction to increase lead quality.
- **Specialization_Select:** Coefficient = -1.31 (negative influence). Leads who leave the "Specialization" field as "Select" are less likely to convert. Encouraging users to provide detailed specialization information can improve lead quality and conversion rates.
- **What is your current occupation_Working Professional:** Coefficient = 1.47 (positive influence). Working professionals show a higher likelihood of conversion. Tailoring marketing campaigns and messaging to this segment can yield better results.

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

- **Lower the Classification Threshold:** From the threshold analysis, a threshold of 0.3 achieves high recall (90%), ensuring most potential leads are included. This will expand the pool of "hot" leads for outreach
- **Prioritize by Lead Scores:** Rank leads based on their lead scores and assign high-probability leads to experienced team members while interns handle lower-ranked leads.
 - **Hot Leads (≥ 80):** 2,165 leads for immediate manual follow-up.
 - **Warm Leads (from 50 to 79):** 1,644 leads for targeted nurturing by interns.
 - **Track Results:** Leverage interns to collect feedback during calls to refine future strategies
 - **Combine Outreach Channels:** In addition to phone calls, nurture leads with follow-up emails or SMS campaigns to maximize engagement

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.

- **Increase the Classification Threshold:** From the threshold analysis, increasing the threshold to 0.7 improves precision to 79.1%, ensuring phone calls are only made to the most promising leads. This reduces the number of non-converters contacted
- **Focus on Hot Leads Only:** Limit manual outreach to leads identified as "Hot Leads") and engage "Warm" and "Cold" leads with automated emails or SMS campaigns
- **Strategic Shift:** Redirect sales team efforts toward strategic tasks, such as refining lead generation strategies, building partnerships, or analyzing past campaign performance