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

**Ans- Following are the top three variables in our final model;**

* **Total Time Spent on Website (Coefficient: 4.4720):**
  + **This variable has the highest positive coefficient, indicating a strong positive relationship with the dependent variable. As the total time spent on the website increases, the likelihood of the desired outcome (perhaps conversion to a lead) also increases.**
  + **The high z-score (28.132) and very low p-value (0.000) suggest that this variable is statistically significant, further supporting its importance in the model.**
* **Lead Origin\_Lead Add Form (Coefficient: 3.2381):**
  + **The positive coefficient for the "Lead Origin\_Lead Add Form" variable indicates that leads generated through the lead add form have a positive impact on the likelihood of the desired outcome.**
  + **The high z-score (16.052) and low p-value (0.000) signify the statistical significance of this variable, suggesting that it contributes significantly to the model.**
* **What is your current occupation\_Working Professional (Coefficient: 2.4117):**
  + **This variable has a positive coefficient and a statistically significant impact on the dependent variable.**
  + **The high z-score (13.127) and low p-value (0.000) indicate its importance in the model.**

1. 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?

**Ans- Following three are the top three categorical variable with significant influence over lead conversion-**

* **Lead Origin\_Lead Add Form (Coefficient: 3.2381):**
  + **The positive coefficient for the "Lead Origin\_Lead Add Form" variable indicates that leads generated through the lead add form have a positive impact on the likelihood of the desired outcome.**
  + **The high z-score (16.052) and low p-value (0.000) signify the statistical significance of this variable, suggesting that it contributes significantly to the model.**
* **What is your current occupation\_Working Professional (Coefficient: 2.4117):**
  + **This variable has a positive coefficient and a statistically significant impact on the dependent variable.**
  + **The high z-score (13.127) and low p-value (0.000) indicate its importance in the model.**
* **Lead Source\_Olark Chat (Coefficient: 0.9605):**
  + **The positive coefficient for the "Lead Source\_Olark Chat" variable indicates that leads coming from the Olark Chat source are positively associated with the desired outcome.**
  + **The high z-score (8.245) and low p-value (0.000) suggest that this variable is statistically significant and contributes significantly to the model.**

1. 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.

**Ans**- **While the sales team has additional resources to invest in reaching out to potential client, the probability threshold can be changed where the precision is higher, and recall may relatively of smaller value. As you target for higher thresholds, your precision would increase. However, someone should be mindful about the accuracy too.**

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.

**Ans- Contrary to the scenario outlined in question # 3, in this scenario going for high recall would be an appropriate decision where the team will be dealing with less false positive scenario. However, one should be mindful of the accuracy. If the accuracy drops significantly, then we will face a lot of false positive.**