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

Answer: The top three variables in my model which contribute most towards the probability of a lead getting converted are:
1.) Total Visit
2.) Total Time spent on website
3.) Lead Origin_Lead Add Form

Below screenshot supports the above results:

```
pd.DataFrame(feature_importance).reset_index().sort_values(by=0,ascending=False).head(3)
```

	index	0
0	TotalVisits	100.00
1	Total Time Spent on Website	83.08
2	Lead Origin_Lead Add Form	67.66

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?**

Answer: The top three categorical/ dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are:
1.) Total Time spent on Website
2.) Lead Origin_Lead Add Form
3.) What is your current occupation_Working Professional

Below screenshot supports the above 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.**

Answer: We can use sensitivity and specificity as metrics to evaluate classification model.
<u>Sensitivity</u>: It is defined as no. of actual conversion correctly predicted out of total no. of actual conversion. We can observe in our model although we had high accuracy our sensitivity turned out to be quite low.

Similarly,
<u>Specificity</u>: It is defined as no. of actual no. non conversions correctly predicted out of total no. of actual non conversions.
Therefore, Sensitivity and specificity is inversely proportional as one increases other decreases. We can change the Conversion Probability cut-off threshold value to get different values of the sensitivity and specificity. If we use high sensitivity in our model, it will correctly identify almost all leads who are likely to Convert. It will do that by increasing the False positive Conversion likelihood, i.e. it will wrongly classify non-Conversion cases as Conversions.
As Sensitivity will be very high, it will make sure almost all leads who are likely to Convert are identified correctly.

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.**

Answer: We have to make sure the specificity is very high, which will make all leads as getting converted. Therefore, they won't have to make unnecessary phone calls.
If we have high specificity in our model, it will correctly identify almost all leads who are not likely to convert. But it will do that at the cost of losing out some low conversion rate leads i.e. it will wrongly classify some converted cases as non-converted.
Therefore, since X Education has already reached its target for a quarter and doesn't want to make phone calls unless it's extremely necessary, it means that we want to minimize the rate of unwanted phone calls. Hence, we choose higher threshold value for conversion.