

Assignment Questions Subjective

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

The final model's top three features, based on their coefficients, are:

- Time on Website
- Lead Source_Reference
- Current Occupation Working Professional

	coef	std err	z	P> z	[0.025	0.975]
const	-3.4948	0.117	-29.779	0.000	-3.725	-3.265
TotalVisits	0.2110	0.181	1.163	0.245	-0.145	0.567
Time on Website	4.0452	0.157	25.817	0.000	3.738	4.352
Lead Source_Olark Chat	0.8722	0.105	8.325	0.000	0.667	1.078
Lead Source_Reference	3.3581	0.215	15.653	0.000	2.938	3.779
Last Activity_Email Bounced	-1.7839	0.316	-5.638	0.000	-2.404	-1.164
Last Activity_Email Opened	0.5938	0.080	7.410	0.000	0.437	0.751
Current Occupation_Student	1.1486	0.224	5.135	0.000	0.710	1.587
Current Occupation_Unemployed	1.1460	0.084	13.719	0.000	0.982	1.310
Current Occupation_Working Professional	3.3497	0.188	17.797	0.000	2.981	3.719
Last Notable Activity_Others	1.6558	0.262	6.322	0.000	1.142	2.169
Last Notable Activity_SMS Sent	1.9421	0.090	21.512	0.000	1.765	2.119

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?

The top three dummy variables for boosting lead conversion are:

- Lead Source_Reference
- Current Occupation_Working Professional
- Last Notable Activity_SMS Sent

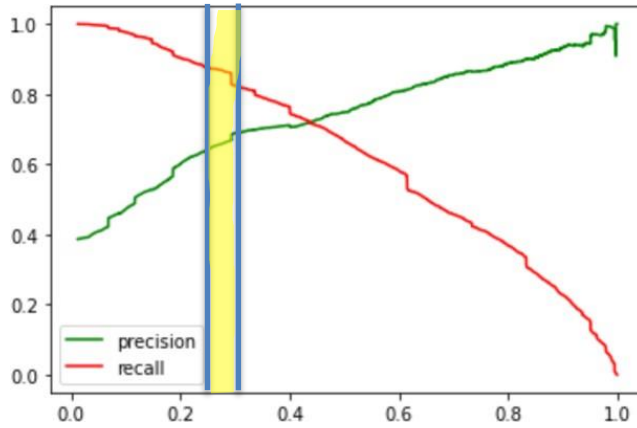
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers 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.

Recall is calculated as the proportion of actual converted leads that are correctly predicted, expressed as:

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{Total Actual Positives (TP + FN)}}$$

Our objective is to contact as many hot leads as possible to maximize the conversion rate. By lowering the probability cutoff, we can increase recall, which means we'll identify more of the actual positive leads. However, this also leads to a decrease in precision, resulting in a higher number of false positives—i.e., contacting many leads that aren't actually potential.

Given that we have sufficient resources and aim to be more aggressive in our lead conversion efforts, we can afford to accept a trade-off where we prioritize higher recall over precision.



We can set the probability cutoff within the yellow region, ranging from 0.25 to 0.3, to pursue a more aggressive lead conversion strategy.

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.

In this scenario, our goal is to focus on reaching out only to potential leads due to limited resources. Precision, in this context, is defined as the ratio of correctly predicted hot leads to the total number of leads predicted as hot. In this case, we aim to reach out to only the Potential leads as we have shortage of resources.

$$\text{Precision} = \frac{\text{True positive (TP)}}{\text{Total Predicted positives (TP+FP)}}$$

To achieve higher precision, we need to reduce the number of false positives. This means we would have to trade off some recall in order to improve precision.

In this situation, our probability cutoff should be set higher, within the range of 0.5 to 0.55, to achieve better precision.

