

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

Top three variables which contribute most towards the conversion rate are:

- **Tags** (contains different dummy variables having coef. ranging from -4.86 to +4.89): Different tags indicated different interests, education status, contactability of a leads which can positively or negatively affect the chance of being converted.
- **Occupation_WOE** (coef. 4.52). This WOE variables indicated that different occupations can have significant impact on the chance of conversion.
- **Total_Time_Spent** (coef. 4.09): higher time spent on the website indicated a higher interest in a course, which positively affect the chance 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?

- **Occupation_WOE** (coef. 4.53): this WOE variable indicated "Working Professional" (category with highest weight of evidence) has very high chance of being converted than other occupations (refer to section 4.4 in the notebook for individual category's WOE)
- **Tags_Revert_after_email** (coef. 3.04): this dummy variable indicated that leads have very high chance of conversion after reading email from the sale team.
- **Lead_Source_WOE** (coef 1.80): this variable indicated "Welingak" and "Reference" (2 categories with highest weight of evidence) have very high chance of being converted than other sources.
- (Note: while Tags_Lost_to_EINS (4.89) and Tags_Closed_by_Horizzon (4.79) have higher coefficients, these two seems like specific stage in the sales process on a platform/portal so not much action can be recommended for these dummy variables).

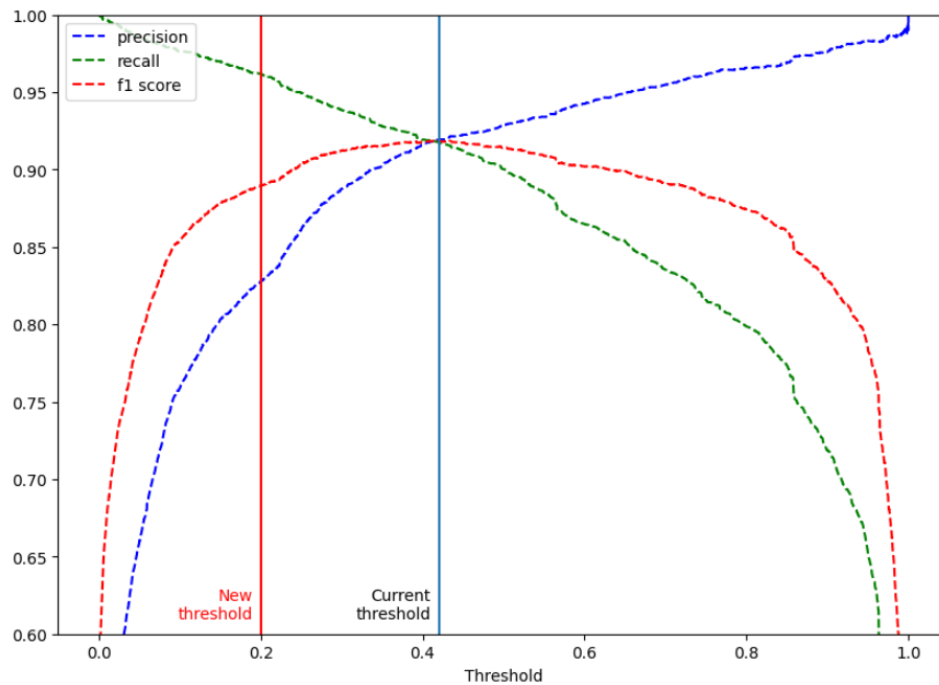
3. X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers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

The solution is to lower the threshold for 'Convert' prediction. In technical term, this will increase recall at the cost of reduced precision (lower false negative rate but higher false positive rate). In business term, there will be more leads get classified as hot leads for the interns to work on, but the chance of conversion of these leads will be lower.

The exact threshold adjustment should be made according to the capability of the interns. A suggested range can be around **0.2**. Any lower will result in a steep decline in precision without any significant gain in recall.

```
pre_rec_curve(model3, model3_top_features, threshold = 0.42, new = 0.2)
```



		Train		Test	
Actual	Not converted	3724.00	574.00	408.00	70.00
	Converted	109.00	2756.00	14.00	304.00
		Not converted		Converted	
		Predicted		Predicted	

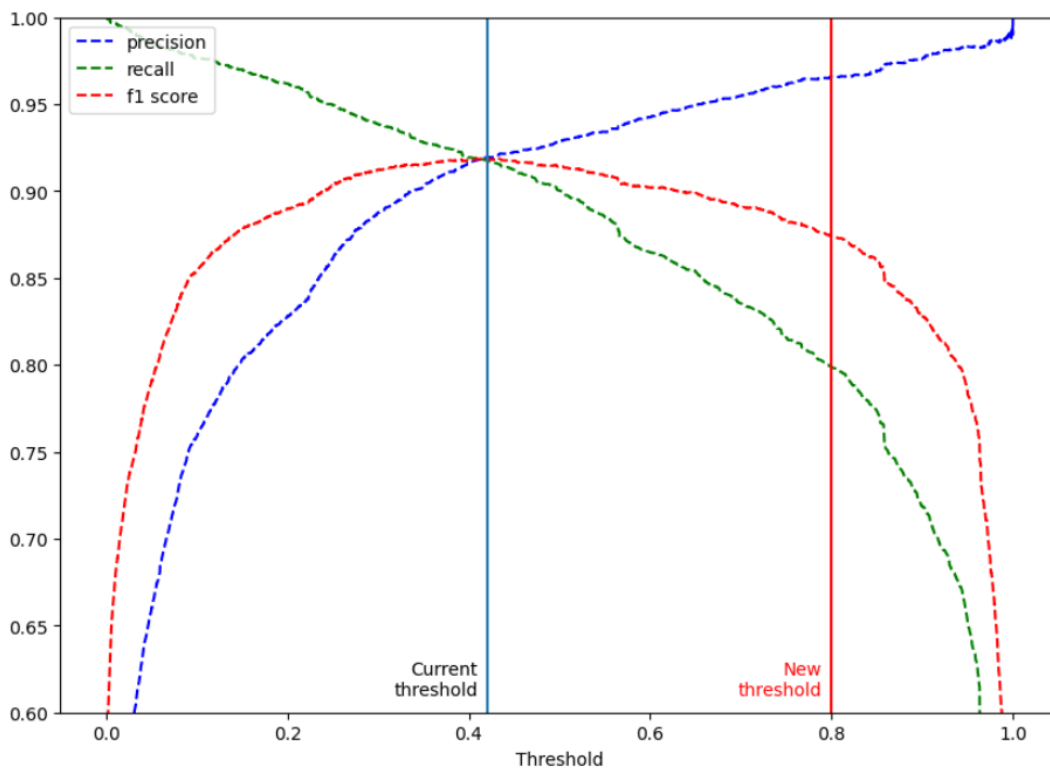
4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

Answer:

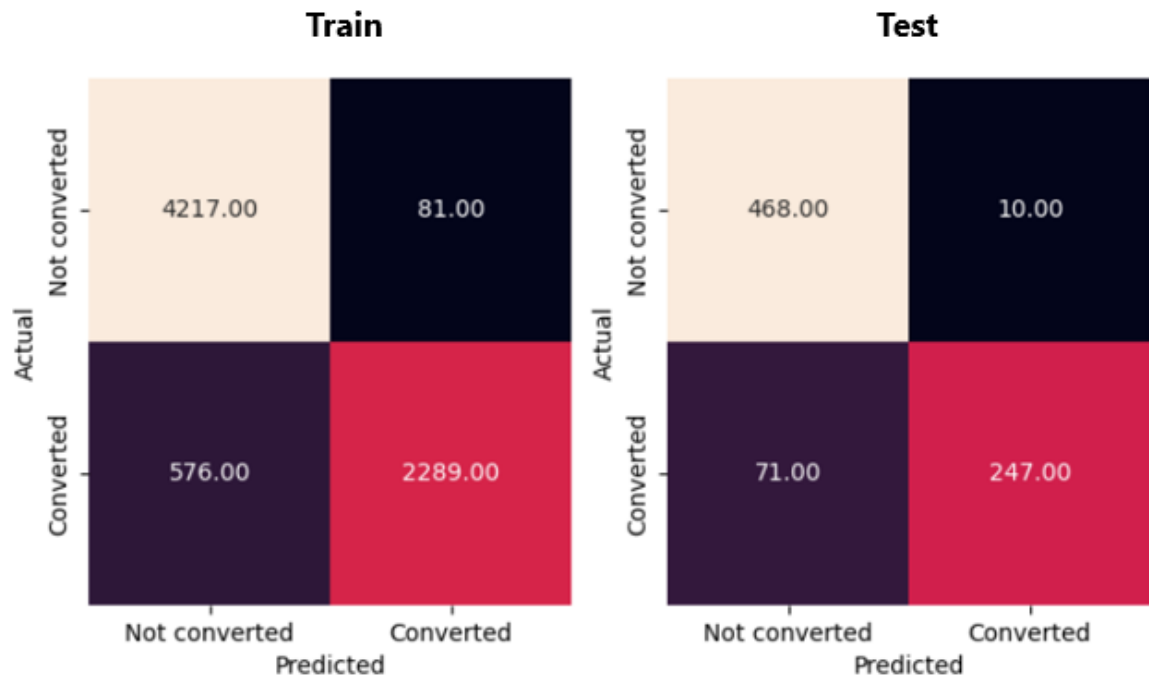
The solution is the exact opposite of the previous question, which is to increase the threshold for hot leads. Doing this will reduce the number of leads that classified as hot, but also will reduce the false negative rate. The final result is that the team will have fewer, but much more promising leads to contact.

The exact threshold increase still needs to be discussed, but around **0.7~0.8** should be a good starting point. Depends on what the team's definition of "Extremely necessary" is, the threshold can be pushed as high as **0.95~0.98**.

```
pre_rec_curve(model3, model3_top_features, threshold = 0.42, new = 0.8)
```



- Model 3 performance at adjusted threshold = **0.8**
- 20% less leads classified as hot, precision increase to 0.96 (from 0.9)



- Model 3 performance at adjusted threshold = **0.98**
- 50% less leads classified as hot, precision increase to 0.99 (from 0.9)

