Assignment Subjective Questions

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

Answer

Following three variables are contributing the most towards the probability of a lead getting converted:

a. Page_views_per
b. Lead_quality_not_sure
c. Total_time_spent

Generalized Linear Model Regression Results

Dep. Variable:	converted	No. Observations:	6351
Model:	GLM	Df Residuals:	6335
Model Family:	Binomial	Df Model:	15
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2205.9
Date:	Mon, 09 Aug 2021	Deviance:	4411.8
Time:	14:27:00	Pearson chi2:	6.29e+03
No. Iterations:	7		
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7358	0.146	5.045	0.000	0.450	1.022
totalvisits	8.4965	2.504	3.393	0.001	3.588	13.405
total_time_spent	4.4432	0.183	24.282	0.000	4.085	4.802
page_views_per	-2.7754	0.610	-4.550	0.000	-3.971	-1.580
lead_source_reference	1.9921	0.248	8.023	0.000	1.505	2.479
lead_source_welingak_website	5.3765	0.732	7.343	0.000	3.941	6.812
do_not_email_yes	-1.5751	0.204	-7.704	0.000	-1.976	-1.174
country_unknown	1.0255	0.129	7.930	0.000	0.772	1.279
lead_quality_might_be	-1.5457	0.154	-10.024	0.000	-1.848	-1.243
lead_quality_not_sure	-3.5951	0.140	-25.591	0.000	-3.870	-3.320
lead_quality_worst	-5.6885	0.390	-14.568	0.000	-6.454	-4.923
asymmetrique_activity_index_low	-1.7674	0.279	-6.326	0.000	-2.315	-1.220
last_notable_activity_had_a_phone_conversation	2.4787	1.203	2.061	0.039	0.121	4.836
last_notable_activity_sms_sent	1.7102	0.088	19.544	0.000	1.539	1.882
last_notable_activity_unreachable	1.9251	0.542	3.555	0.000	0.864	2.986
last_notable_activity_unsubscribed	1.5133	0.641	2.362	0.018	0.258	2.769

	Features	VIF
2	page_views_per	3.52
8	lead_quality_not_sure	3.17
1	total_time_spent	2.03
6	country_unknown	2.01
7	lead_quality_might_be	1.90
0	totalvisits	1.87
12	last_notable_activity_sms_sent	1.56
9	lead_quality_worst	1.19
3	lead_source_reference	1.18
5	do_not_email_yes	1.17
4	lead_source_welingak_website	1.07
14	last_notable_activity_unsubscribed	1.06
10	asymmetrique_activity_index_low	1.05
11	last_notable_activity_had_a_phone_conversation	1.01
13	last_notable_activity_unreachable	1.01

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?**

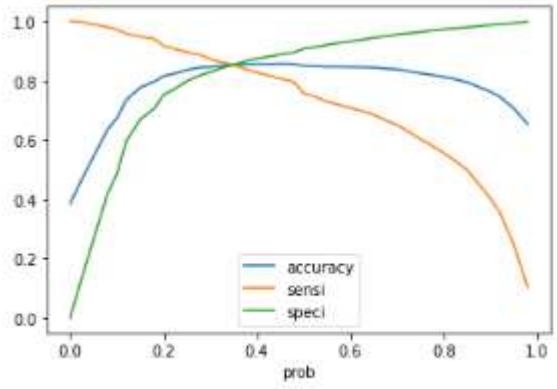
   Answer

   - Again, based on the coefficient values from the screen shot in the question above, the following are the top three categorical/dummy variables that should be focused the most in order to increase the probability of lead conversion :
     - lead_source_reference
     - lead_source_welingak_website
     - last_notable_activity_had_a_phone_conversation

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

- Sensitivity with respect to our model can be defined as the ratio of total number of actual Conversions correctly predicted to the total number of actual conversions. Similarly, Specificity can be defined as the ratio of total number of actual non-conversions correctly predicted to the total number of actual non-conversions. For a model, as one increases, the other decreases and vice versa. Different values of the sensitivity and specificity can be achieved for the same model by changing the conversion probability cutoff threshold value. For our model, the below graph shows how the Accuracy, Sensitivity and Specificity rating changes with change in the threshold value:




```
Optimum cut-off value is: 0.35
```

- When the probability thresholds are very low, the sensitivity is very high and specificity is very low. Similarly, for larger probability thresholds, the sensitivity values are very low but the specificity values are very high.
- High sensitivity implies that our model will correctly identify almost all leads who are likely to convert. It will do that by over-estimating the conversion likelihood, i.e. it will misclassify some non-conversion cases as conversions. Now, since X Education has more manpower for these 2 months and they wish to make the lead conversion more aggressive by wanting almost all the potential leads, we can choose a lower threshold value for Conversion Probability.
- This will ensure the Sensitivity rating is very high which in turn will make sure almost all leads who are likely to Convert are identified correctly and the agents can make phone calls to as much of such people as possible. The company may follow high volume low margin strategy which means the conversion rate might reduce but, the count of conversion would increase and eventually the revenue would also increase.

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

- Following the similar logic and context from the previous question, High Specificity implies that our model will correctly identify almost all leads who are not likely to convert. It will do that at the cost of losing out some low Conversion rate risky leads to the competition, i.e. it will misclassify some Conversion cases as non-Conversions.
- Therefore, since X Education has already reached its target for a quarter and doesn't want to make phone calls unless it is extremely necessary, i.e. they want to minimize the rate of useless phone calls. We can choose a higher threshold value for Conversion Probability.
- This will ensure the Specificity rating is very high, which in turn will make sure almost all leads who are on the brink of the probability of getting Converted or not are not selected. As a result, they won't have to make unnecessary phone calls and can focus on some new work. In this way the efficiency of sales team would increase as the conversion rate would be high. The sales cycle will also be reduced.