Team 16

ML Case Study

Lead Scoring Dataset

Name: Sumit Bansod Roll no :220940325081

Name: Gauri Pandey Roll no :220940325033

Name: Keshav Yawale Roll no :220940325036

Name: Dipti Patil Roll no :220940325028

Name: Saurav Sharma Roll no :220940325063

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

```
res.summary()
```

Out[147]:

Generalized Linear Model Regression Results

Dep. Variable:	Converted	**No. Observations:**	6267
Model:	GLM	**Df Residuals:**	6253
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-1263.3
Date:	Mon, 23 Jan 2023	**Deviance:**	2526.6
Time:	21:32:34	**Pearson chi2:**	8.51e+03
No. Iterations:	8	**Pseudo R-squ. (CS):**	0.6037
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1179	0.084	-13.382	0.000	-1.282	-0.954
Total Time Spent on Website	0.8896	0.053	16.907	0.000	0.786	0.993
Lead Origin_Lead Add Form	1.6630	0.455	3.657	0.000	0.772	2.554
Lead Source_Direct Traffic	-0.8212	0.127	-6.471	0.000	-1.070	-0.572
Lead Source_Welingak Website	3.8845	1.114	3.488	0.000	1.701	6.068
Last Activity_SMS Sent	1.9981	0.113	17.718	0.000	1.777	2.219
Last Notable Activity_Modified	-1.6525	0.124	-13.279	0.000	-1.896	-1.409
Last Notable Activity_Olark Chat Conversation	-1.8023	0.491	-3.669	0.000	-2.765	-0.839
Tags_Closed by Horizzon	7.1955	1.020	7.053	0.000	5.196	9.195
Tags_Interested in other courses	-2.1318	0.406	-5.253	0.000	-2.927	-1.336
Tags_Lost to EINS	5.9177	0.611	9.689	0.000	4.721	7.115
Tags_Other_Tags	-2.3737	0.206	-11.507	0.000	-2.778	-1.969
Tags_Ringing	-3.4531	0.238	-14.532	0.000	-3.919	-2.987
Tags_Will revert after reading the email	4.5070	0.188	24.002	0.000	4.139	4.875

From the summary we can observe that the coefficients of columns **Tags, Lead Source and Last activity are high.**

Thus, these 3 categorical features contribute most towards the probability of a lead getting converted.

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?**

```
res.summary()
```

Out[147]:

Generalized Linear Model Regression Results

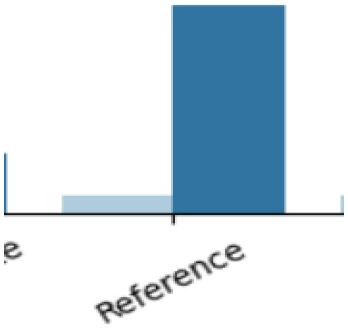
Dep. Variable:	Converted	**No. Observations:**	6267
Model:	GLM	**Df Residuals:**	6253
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-1263.3
Date:	Mon, 23 Jan 2023	**Deviance:**	2526.6
Time:	21:32:34	**Pearson chi2:**	8.51e+03
No. Iterations:	8	**Pseudo R-squ. (CS):**	0.6037
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1179	0.084	-13.382	0.000	-1.282	-0.954
Total Time Spent on Website	0.8896	0.053	16.907	0.000	0.786	0.993
Lead Origin_Lead Add Form	1.6630	0.455	3.657	0.000	0.772	2.554
Lead Source_Direct Traffic	-0.8212	0.127	-6.471	0.000	-1.070	-0.572
Lead Source_Welingak Website	3.8845	1.114	3.488	0.000	1.701	6.068
Last Activity_SMS Sent	1.9981	0.113	17.718	0.000	1.777	2.219
Last Notable Activity_Modified	-1.6525	0.124	-13.279	0.000	-1.896	-1.409
Last Notable Activity_Olark Chat Conversation	-1.8023	0.491	-3.669	0.000	-2.765	-0.839
Tags_Closed by Horizzon	7.1955	1.020	7.053	0.000	5.196	9.195
Tags_Interested in other courses	-2.1318	0.406	-5.253	0.000	-2.927	-1.336
Tags_Lost to EINS	5.9177	0.611	9.689	0.000	4.721	7.115
Tags_Other_Tags	-2.3737	0.206	-11.507	0.000	-2.778	-1.969
Tags_Ringing	-3.4531	0.238	-14.532	0.000	-3.919	-2.987
Tags_Will revert after reading the email	4.5070	0.188	24.002	0.000	4.139	4.875

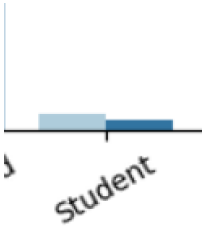
→ From the summary we can observe that the coefficients of columns **Tags_Closed by Horizzon, Tags_Lost to EINS,Tags_Will revert after reading the email are high.**

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.**

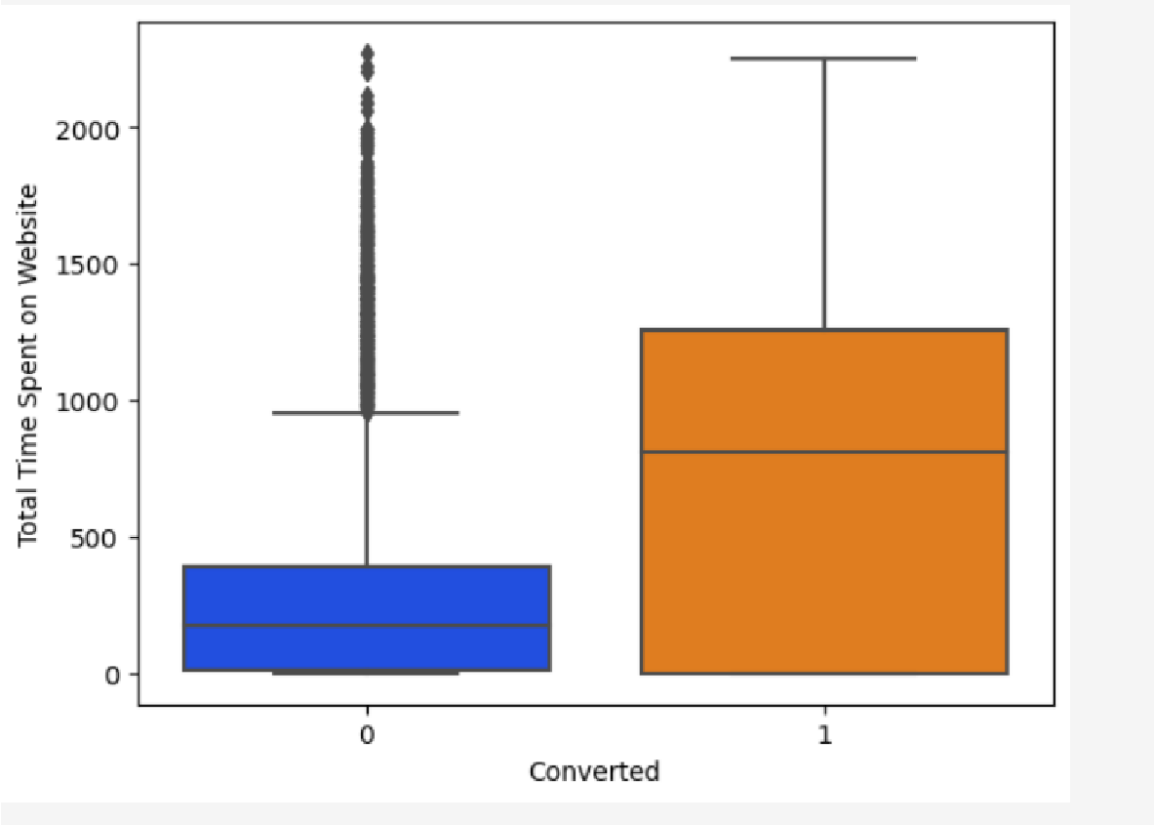
- Target leads that have come through References as they have a higher probability of converting.



- Students will have a lower probability of converting due to the course being industry based and they should be targeted after their completion of education.



- Target leads that spend a lot of time on X-Education site (Total Time Spent on Website)



- 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 which in turn will make sure almost all leads who are likely to be converted are identified correctly and the agents can make phone calls to as much of such people as possible. The company may follow a 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.

- Sales should not focus on unemployed leads as they may not have a budget to spend on the course.
- Sales should not focus on students as they are already studying and would not be willing to enroll for a professional course at the current stage.
- It is advised not to focus on unemployed leads they might not have a budget to spend on the course
- 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 the sales team would increase as the conversion rate would be high. The sales cycle will also be reduced.