Model feature results:

Generalized Linear Model Regression Results

Dep. Variable:	Converted	**No. Observations:**	6246
Model:	GLM	**Df Residuals:**	6224
Model Family:	Binomial	**Df Model:**	21
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-1223.6
Date:	Fri, 18 Oct 2024	**Deviance:**	2447.3
Time:	22:50:58	**Pearson chi2:**	1.14e+04
No. Iterations:	8	**Pseudo R-squ. (CS):**	0.6080
Covariance Type:	nonrobust		

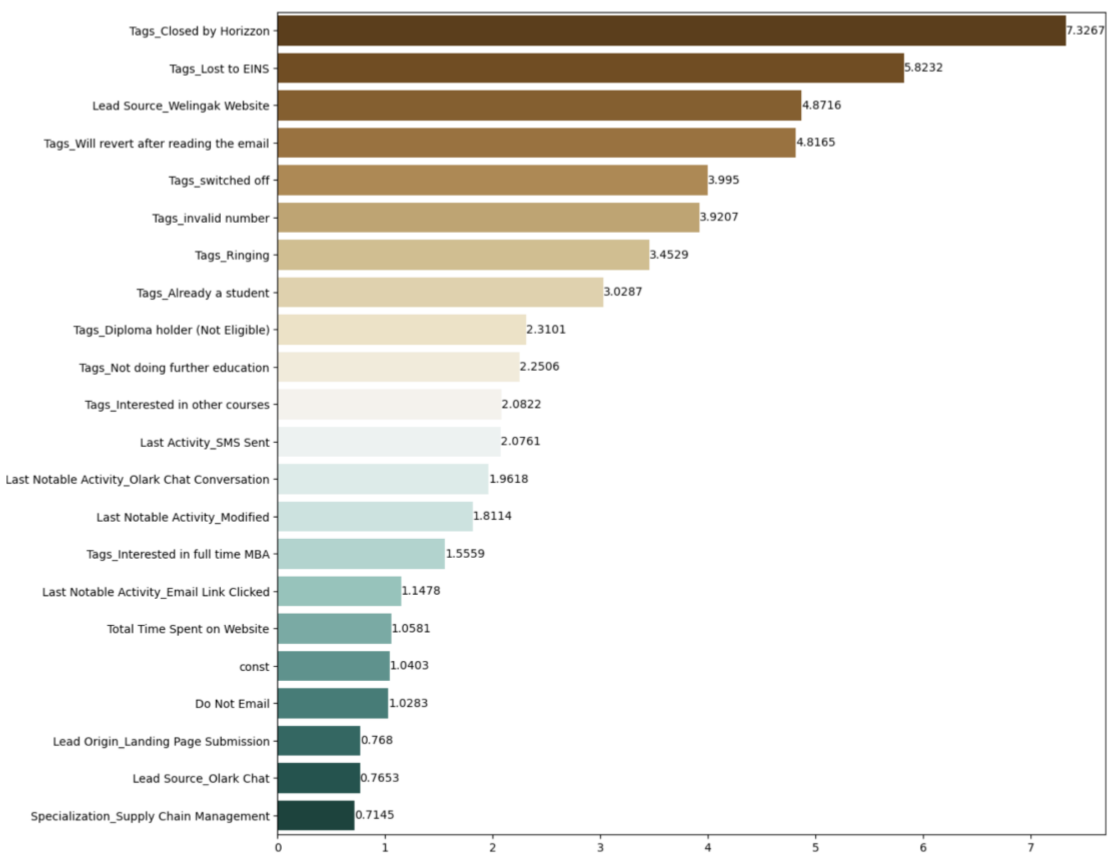
	coef	std err	z	P>|z|	[0.025	0.975]
Do Not Email	-1.0283	0.254	-4.042	0.000	-1.527	-0.530
Total Time Spent on Website	1.0581	0.061	17.451	0.000	0.939	1.177
Lead Origin_Landing Page Submission	-0.7680	0.131	-5.847	0.000	-1.025	-0.511
Lead Source_Olark Chat	0.7653	0.160	4.777	0.000	0.451	1.079
Lead Source_Welingak Website	4.8716	0.751	6.487	0.000	3.400	6.343
Last Activity_SMS Sent	2.0761	0.119	17.425	0.000	1.843	2.310
Last Notable Activity_Email Link Clicked	-1.1478	0.424	-2.709	0.007	-1.978	-0.317
Last Notable Activity_Modified	-1.8114	0.128	-14.149	0.000	-2.062	-1.560
Last Notable Activity_Olark Chat Conversation	-1.9618	0.433	-4.528	0.000	-2.811	-1.113
Tags_Already a student	-3.0287	0.607	-4.990	0.000	-4.218	-1.839
Tags_Closed by Horizzon	7.3267	0.728	10.058	0.000	5.899	8.755
Tags_Diploma holder (Not Eligible)	-2.3101	1.057	-2.185	0.029	-4.383	-0.238
Tags_Interested in full time MBA	-1.5559	0.634	-2.456	0.014	-2.798	-0.314
Tags_Interested in other courses	-2.0822	0.405	-5.145	0.000	-2.875	-1.289
Tags_Lost to EINS	5.8232	0.536	10.870	0.000	4.773	6.873
Tags_Not doing further education	-2.2506	1.023	-2.200	0.028	-4.255	-0.246
Tags_Ringing	-3.4529	0.250	-13.821	0.000	-3.943	-2.963
Tags_Will revert after reading the email	4.8165	0.191	25.163	0.000	4.441	5.192
Tags_invalid number	-3.9207	1.145	-3.425	0.001	-6.165	-1.677
Tags_switched off	-3.9950	0.743	-5.379	0.000	-5.451	-2.539
Specialization_Supply Chain Management	-0.7145	0.318	-2.248	0.025	-1.338	-0.091
const	-1.0403	0.118	-8.796	0.000	-1.272	-0.809

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

Base on the final model result as below picture, top three variables which contribute most towards the probability of a lead getting converted are the variables which have **highest absolute coeff**. They are:

- Tags_Closed by Horizzon (coef = 7.3267)
  - Positive contribution
  - If the lead has tag “Closed by Horizzon”, there is a high probability to convert to customer. So company should focus on such leads
- Tags_Lost to EINS (coef = 5.8232)
  - Positive contribution
  - If the lead has tag “Lost to EINS”, there is a high probability to convert to customer. So company should focus on such leads
- Lead Source_Welingak Website (coef = 4.8716)
  - Positive contribution
  - If the source of leads is “Welingak Website”, there is a high probability to convert to customer. So company should focus on such leads

Below are the feature absolute coefficients.



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 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are:

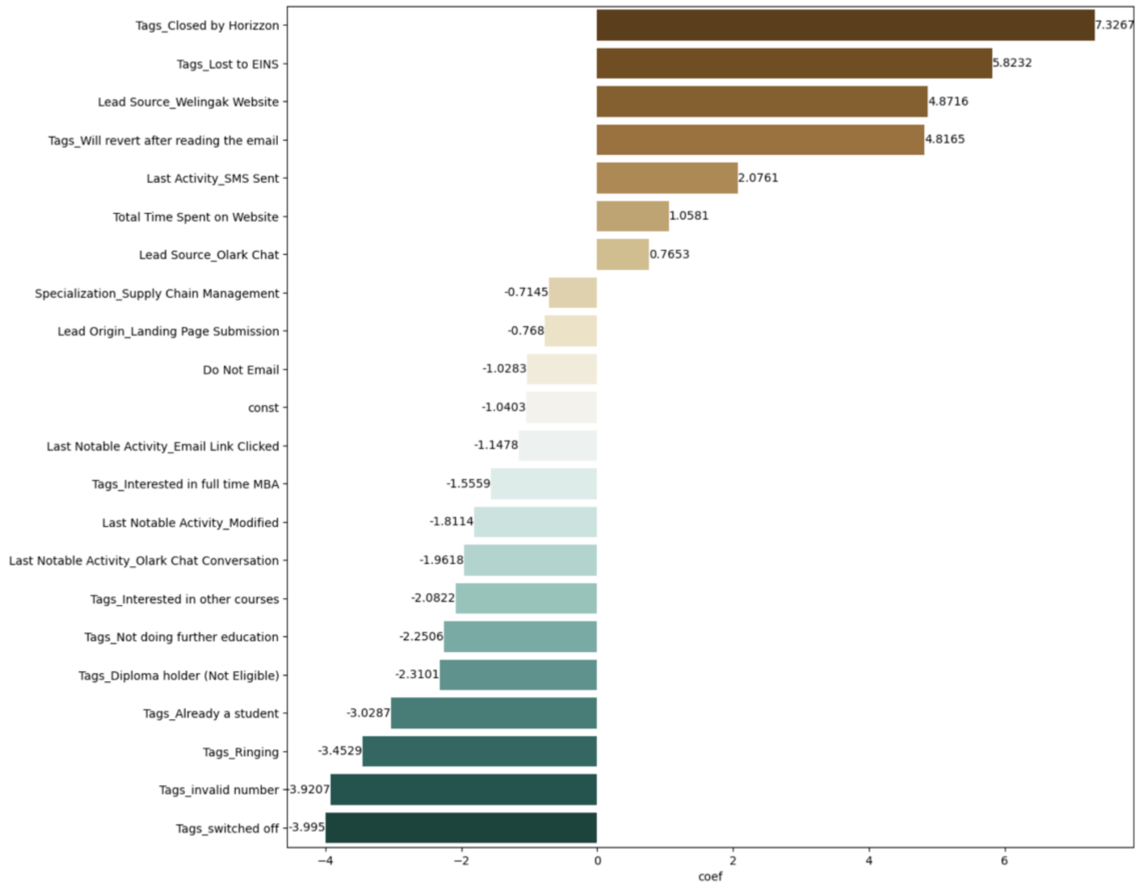
- Tags_Closed by Horizzon (coef = 7.3267)
  - Positive contribution
  - If the lead has tag “Closed by Horizzon”, there is a high probability to convert to customer. So company should focus on such leads
- Tags_Lost to EINS (coef = 5.8232)
  - Positive contribution
  - If the lead has tag “Lost to EINS”, there is a high probability to convert to customer. So company should focus on such leads
- Lead Source_Welingak Website (coef = 4.8716)
  - Positive contribution
  - If the source of leads is “Welingak Website”, there is a high probability to convert to customer. So company should focus on such leads

When we change the variable (eg: 0 to 1), it will affect to lead conversion probability the most vs. the remaining features.

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.*

Strategy to priority the leads in this case:

- Priority 1: Target leads who have the higher predict lead_score. The lead_score is in range 0..100. The higher lead_score, the higher priority to make the phone calls.
- Priority 2: Among the leads who have the same lead_score, we will priority for the leads who have most contribution features.
  - The higher coef, the higher contribution to conversion probability. The below chart sorting descending on the contribution.
  - For the categorical/dummy variables (eg: “Tags_Closed by Horizzon”), if a lead have possitive on this feature, this lead will have higher chances to convert than a lead don’t have this feature
  - For the numerical variables (such as Total Time Spent on Website…), we will prioritize for the lead have higher coef * numerical_value



○

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.*

To minimize the useless phone calls, we will prevent to call the leads who have negative coef features (eg: Tags_switched off, Tags_invalid number…).

- These features have negative coef, means that it will have negative impact on the conversions.
- For categorical variables with negative coef: the higher negative coef, the higher higher prevention to make the call these leads.
- For the numerical variables with negative coef: we consider

magnitude of a feature = coef * numeric_feature_value .

The higher "magnitude of a feature", the higher prevention to make the phone call.

- For the feature priority, from above chart, we will go from bottom to top (lowest coef to highest coef)