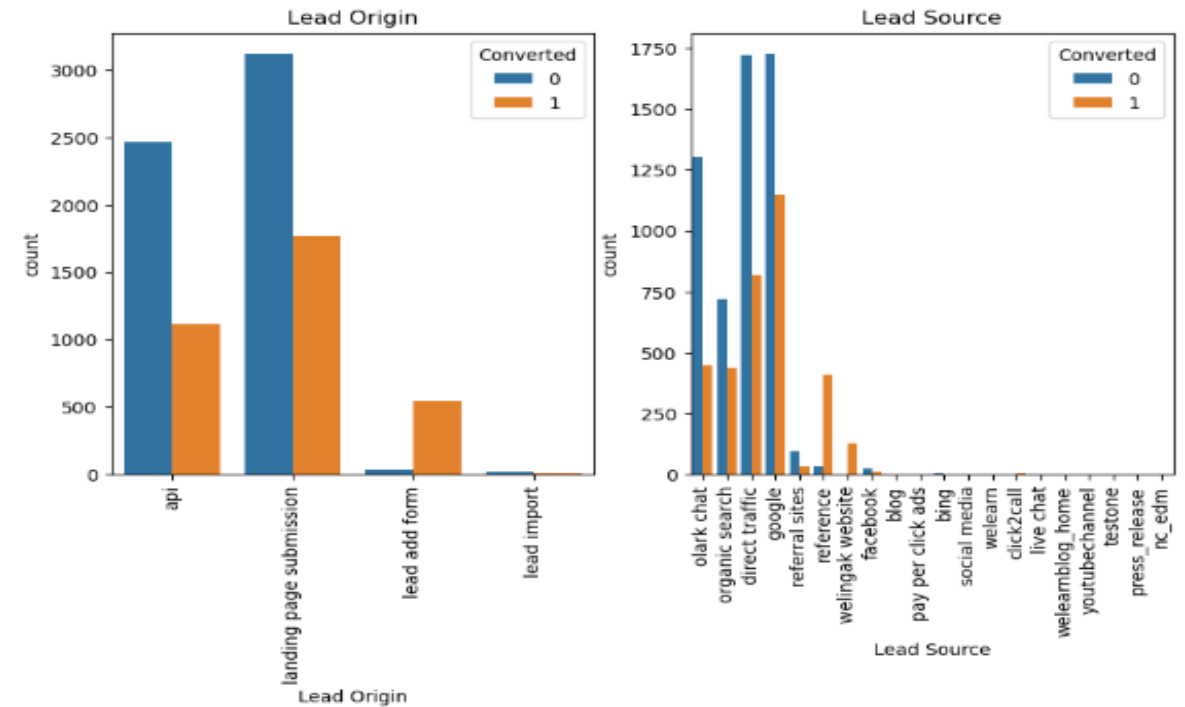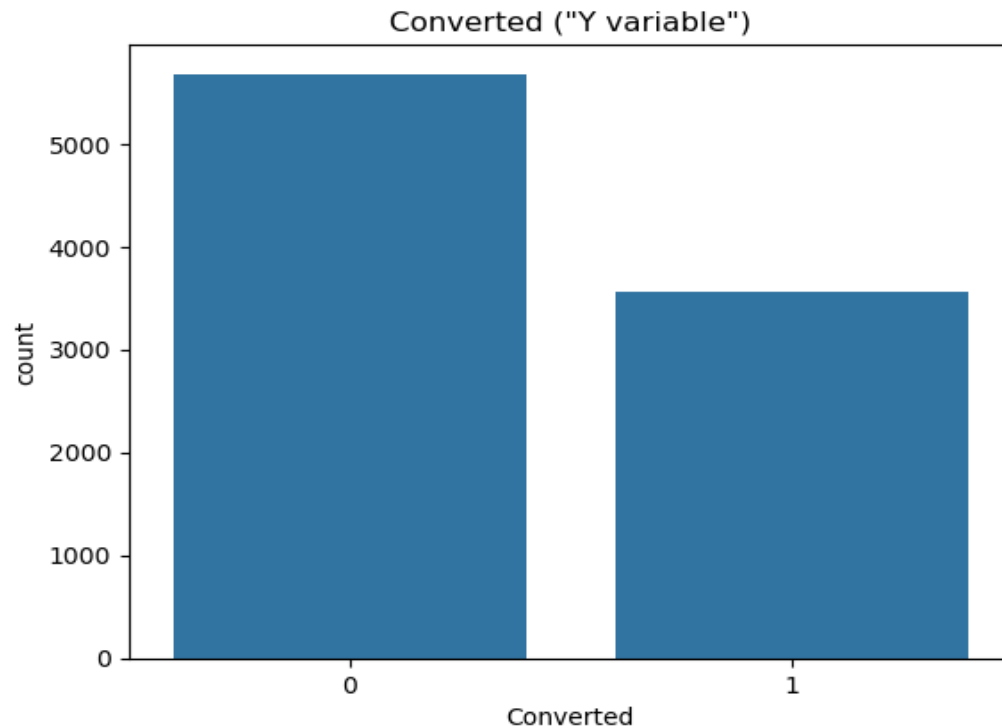# Lead Scoring Analysis

By: Rufina D Costa, Ruchir Rastogi, Robinson Kotamraju

Exploring Initial Data

- The overall conversion rate as per the available data is ~38.5%

- This conversion rate accounts all the other parameters taken into consideration

- Majority of these conversions are attributed to,

    1) The origin being Landing Page

    2) The source being Google

Feature Selection

The model-built post RFE indicates we had all the P values under 0.05 and VIF values below 2.5

Dep. Variable:	Converted	No. Observations:	6351
Model:	GLM	Df Residuals:	6338
Model Family:	Binomial	Df Model:	12
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2757.3
Date:	Thu, 14 Nov 2024	Deviance:	5514.5
Time:	04:03:09	Pearson chi2:	6.65e+03
No. Iterations:	7	Pseudo R-squ. (CS):	0.3726
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.2466	0.081	-15.398	0.000	-1.405	-1.088
TotalVisits	4.7586	1.410	3.375	0.001	1.995	7.522
Total Time Spent on Website	4.5539	0.162	28.136	0.000	4.237	4.871
Lead Origin_lead add form	2.6860	0.224	11.966	0.000	2.246	3.126
Lead Source_direct traffic	-1.4706	0.114	-12.929	0.000	-1.694	-1.248
Lead Source_google	-1.1564	0.109	-10.588	0.000	-1.370	-0.942
Lead Source_organic search	-1.2631	0.134	-9.416	0.000	-1.526	-1.000
Lead Source_welingak website	2.5923	1.033	2.509	0.012	0.567	4.617
Do Not Email_yes	-1.4186	0.168	-8.461	0.000	-1.747	-1.090
Last Activity_olark chat conversation	-1.4717	0.165	-8.909	0.000	-1.796	-1.148
Last Activity_sms sent	1.3038	0.072	18.031	0.000	1.162	1.445
What is your current occupation_working professional	2.7934	0.193	14.449	0.000	2.414	3.172

	Features	VIF
0	const	6.80
5	Lead Source_google	2.32
4	Lead Source_direct traffic	2.30
6	Lead Source_organic search	1.86
3	Lead Origin_lead add form	1.67
2	Total Time Spent on Website	1.32
1	TotalVisits	1.30
7	Lead Source_welingak website	1.29
9	Last Activity_olark chat conversation	1.25
10	Last Activity_sms sent	1.10
11	What is your current occupation_working profes...	1.09
8	Do Not Email_yes	1.03
12	Last Notable Activity_unreachable	1.00

All the VIF values are good and all the p-values are below 0.05. So we can fix model.

ROC Curve

- The AUC score is 87% which is more than fair, if not excellent and it is a testament of how good the model is and also indicates that it was worth the test
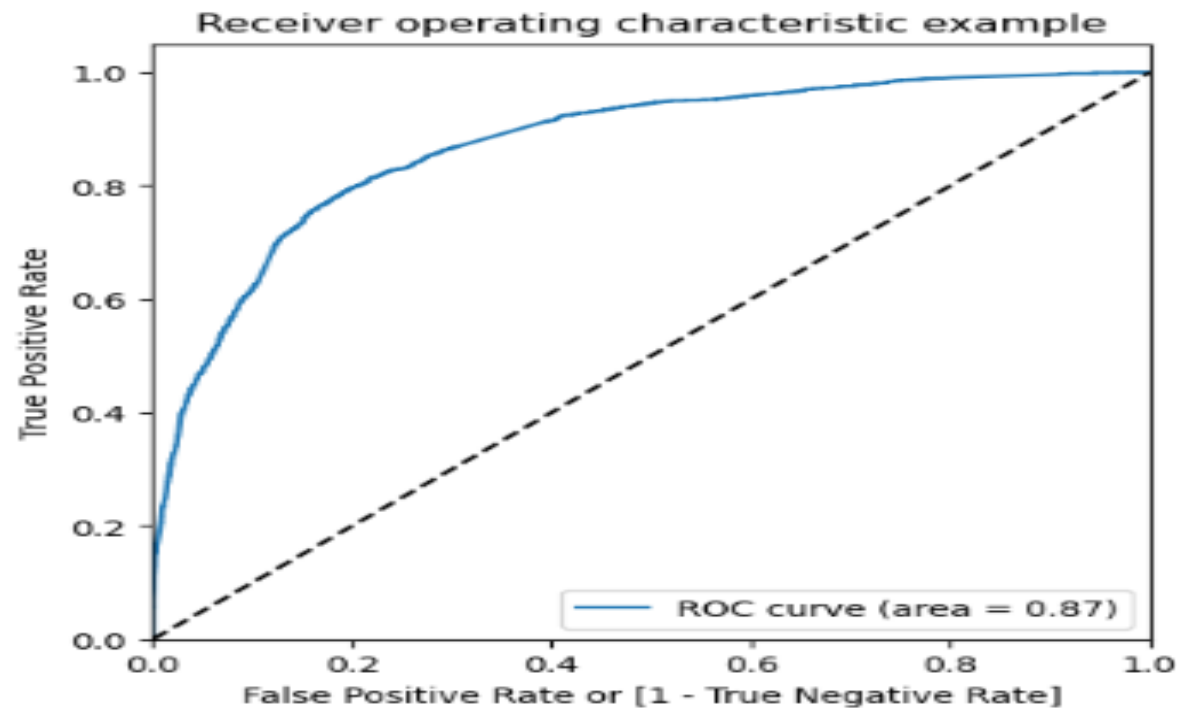
Lead Score Conversion Predicted

- The conversion score on the train data predicts a value of 80%

- The conversion score on the test data predicts a value of 75.32%

Influencing Variables

- Based on the lead score assigned for each of the leads, the few key influential variables were identified.

- **Total time spent**
- **Total number of visits**
- **Lead Source with following priorities**
  - Google
  - Direct Traffic
  - Organic Search
  - Welingak Website
  - SMS
  - Olark chat conversations
- **Lead Origin**
  - Lead add format
- **Occupation**
  - Working Professional

Thank you!