LEAD SCORING CASESTUDY –Assignment

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Business Objectives

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Two major data sheets provided for the data Analysis

1. Leads
2. Leads data Dictionary

For Programming :

1. Python
2. Notebook
3. Libraries functions

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

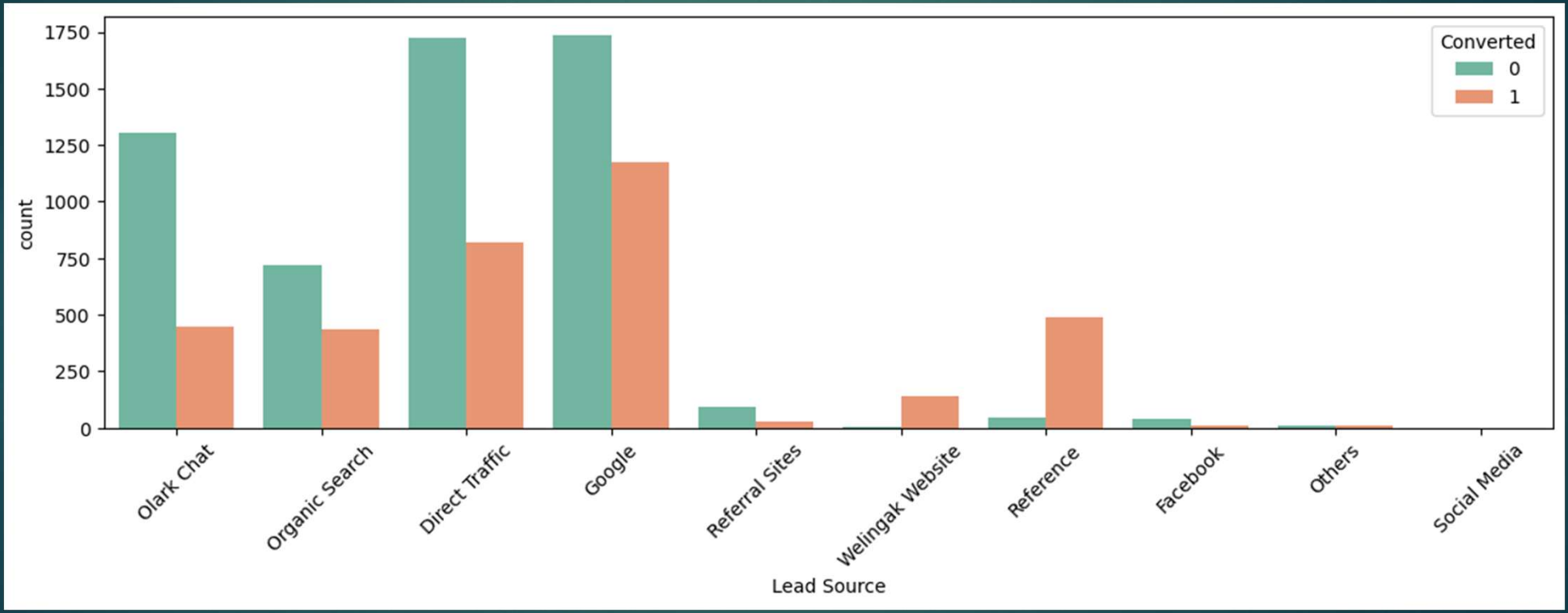
Data Cleaning & Manipulation of Data

1. Null values
2. Filtering Unwanted columns
3. Missing values
4. Sorting the data
5. Fixing the data type
6. Feature scaling
7. Model building

Remove the 40% of Null values

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

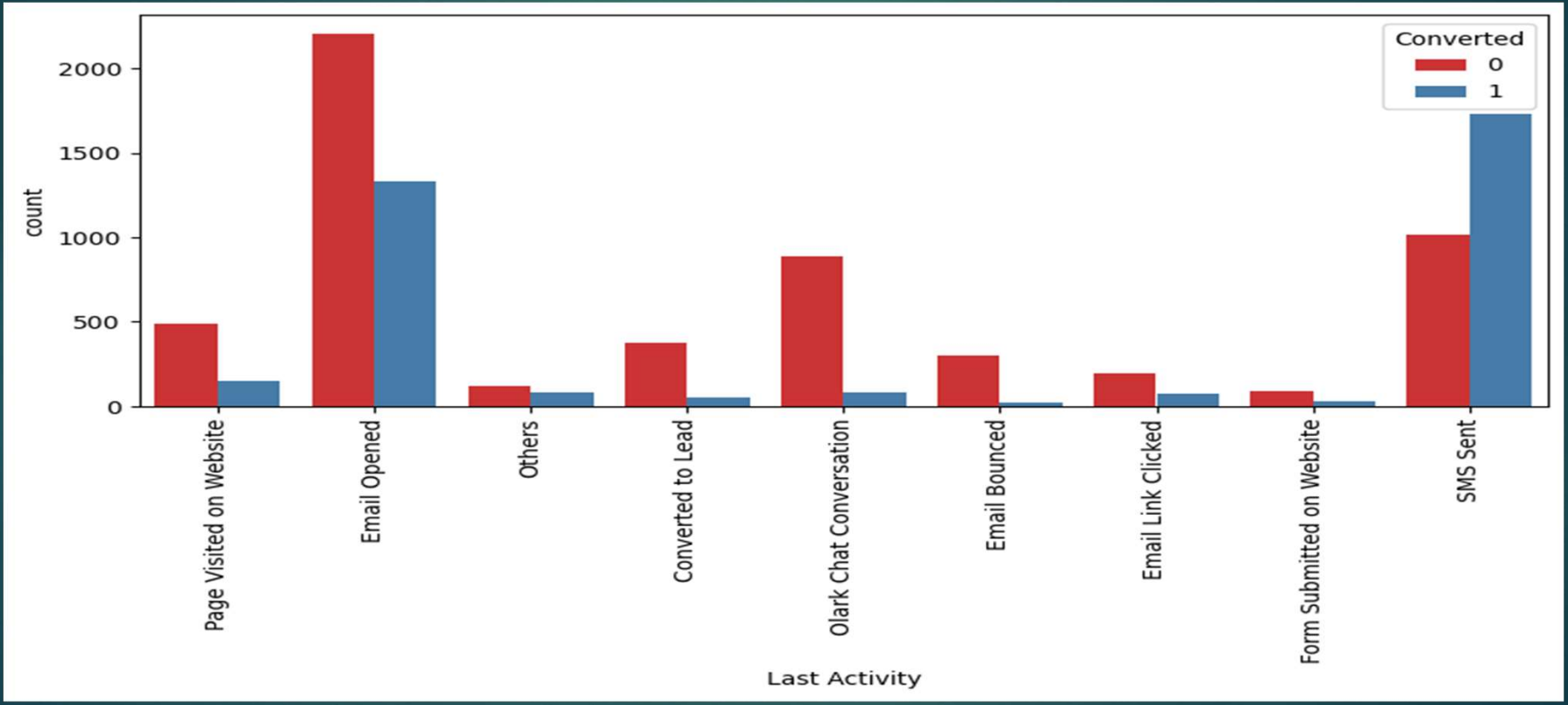
Lead Generation and rate of conversion



SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Last activity

1. Maximum leads are generated having last activity as Email opened but conversion rate is not too good.
2. SMS sent as last activity has high conversion rate

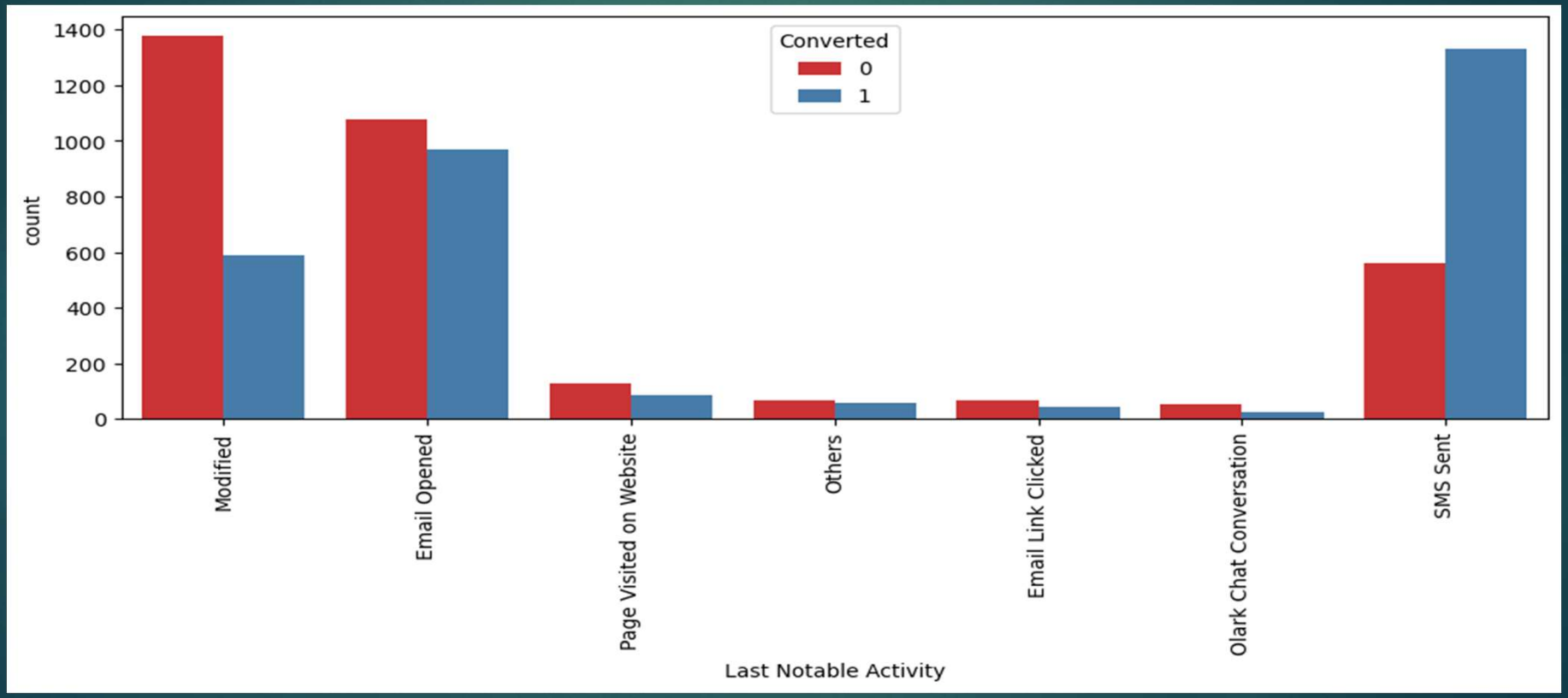


SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Last Notable activity

1.Maximum leads are generated having last activity as Email opened but conversion rate is not too good.

2. SMS sent as last activity has high conversion rate

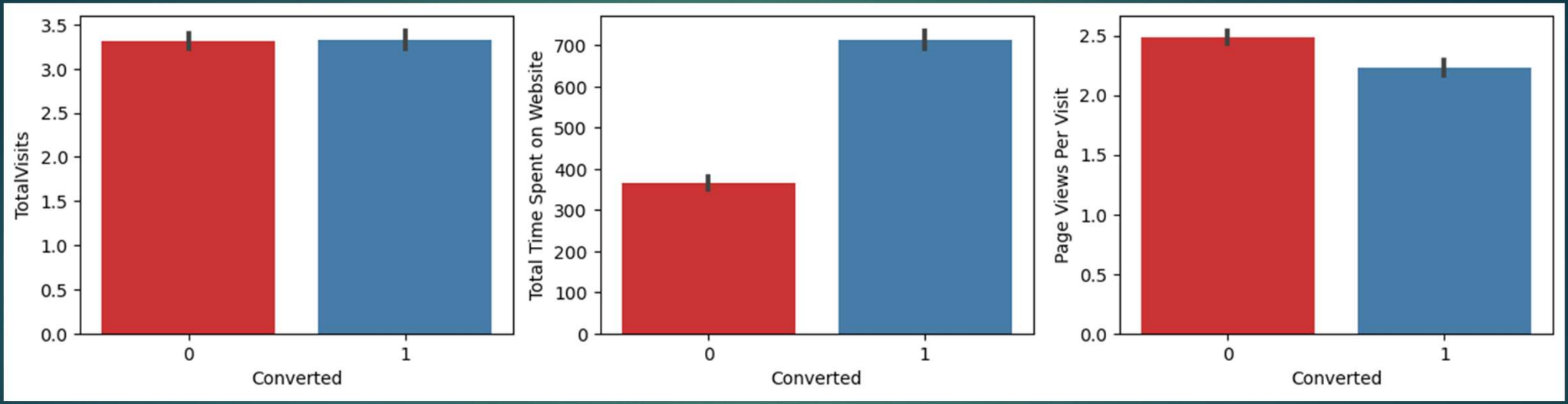


SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

The current lead conversion rate stands at 48.15%, indicating the percentage of leads that successfully converted.

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

The conversion rate is significantly higher for leads with higher Total Visits, Total Time Spent on Website, and Page Views Per Visit.



SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

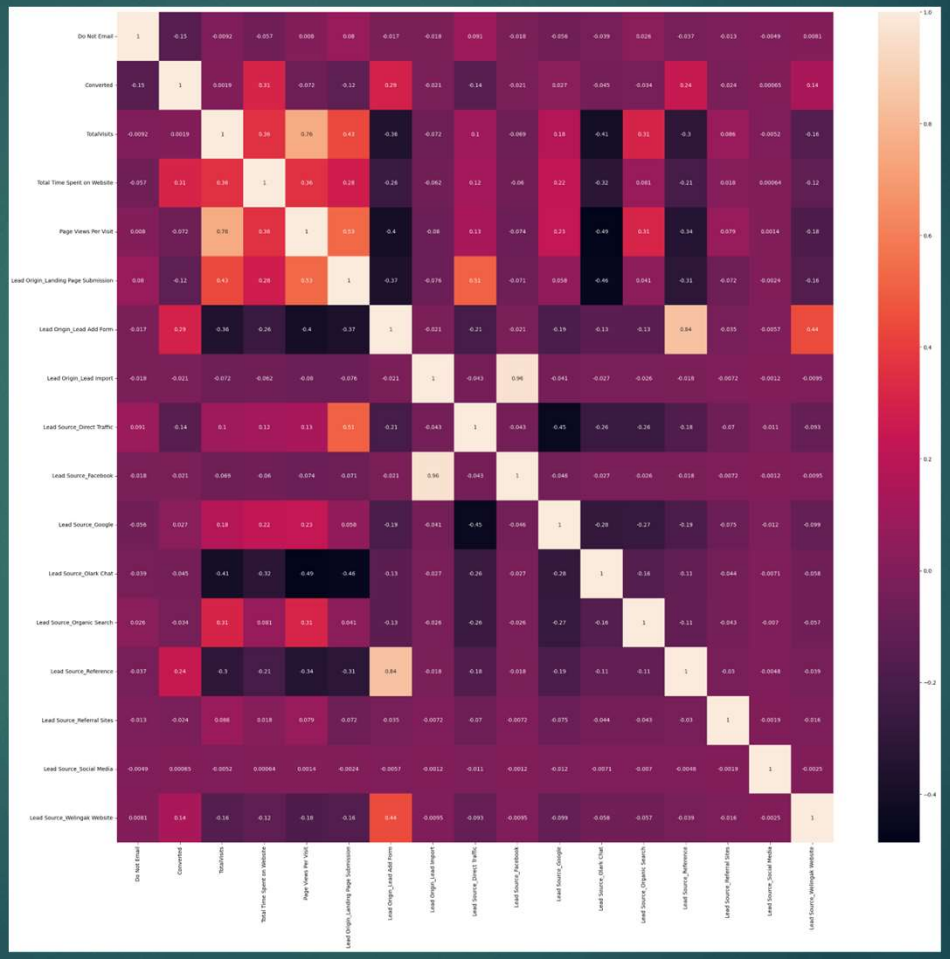
Dummy Variable Creation

	Do Not Email	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Lead Origin_Landing Page Submission	Lead Origin_Lead Add Form	Lead Origin_Lead Import	Lead Source_Direct Traffic	Lead Source_Facebook	Lead Source_Google	Lead Source_Olark Chat	S
0	0	0	0.0	0	0.0	0	0	0	0	0	0	1	
1	0	0	5.0	674	2.5	0	0	0	0	0	0	0	
2	0	1	2.0	1532	2.0	1	0	0	1	0	0	0	
3	0	0	1.0	305	1.0	1	0	0	1	0	0	0	
4	0	1	2.0	1428	1.0	1	0	0	0	0	1	0	

Dummy Variable Creation

	Do Not Email	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Lead Origin_Landing Page Submission	Lead Origin_Lead Add Form	Lead Origin_Lead Import	Lead Source_Direct Traffic	Lead Source_Facebook	Lead Source_Google	Lead Source_Olark Chat	S
0	0	0	0.0	0	0.0	0	0	0	0	0	0	1	
1	0	0	5.0	674	2.5	0	0	0	0	0	0	0	
2	0	1	2.0	1532	2.0	1	0	0	1	0	0	0	
3	0	0	1.0	305	1.0	1	0	0	1	0	0	0	
4	0	1	2.0	1428	1.0	1	0	0	0	0	1	0	

Correlated Dummy Variable Creation



SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Model Building

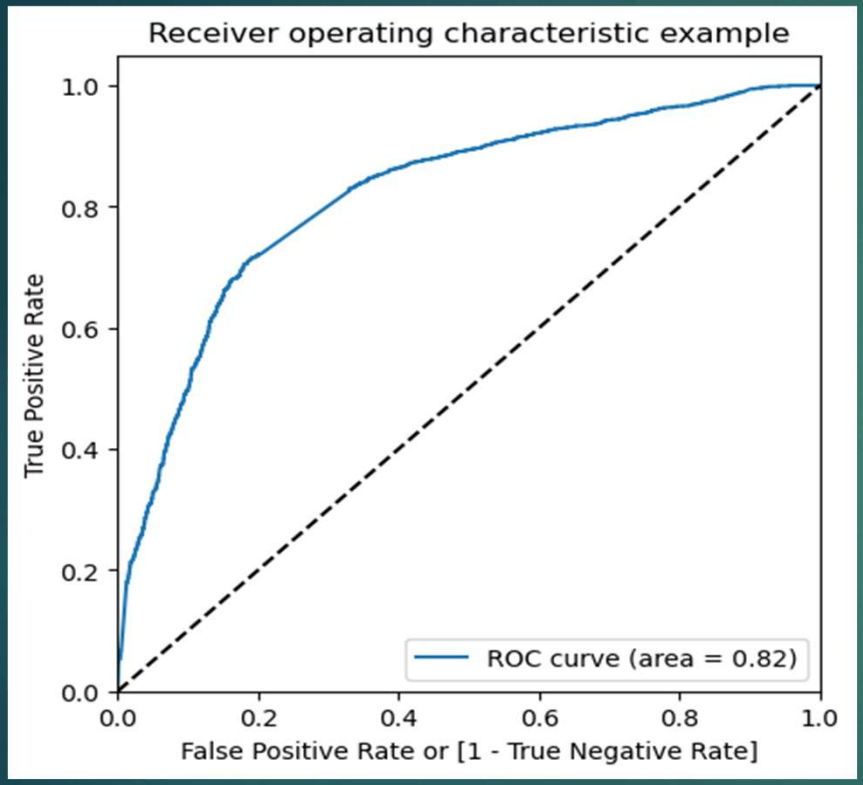
Generalized Linear Model Regression Results

Dep. Variable:	Converted	**No. Observations:**	4494
Model:	GLM	**Df Residuals:**	4480
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2346.1
Date:	Mon, 18 Sep 2023	**Deviance:**	4692.2
Time:	01:05:39	**Pearson chi2:**	4.68e+03
No. Iterations:	19	**Pseudo R-squ. (CS):**	0.2885
Covariance Type:	nonrobust		

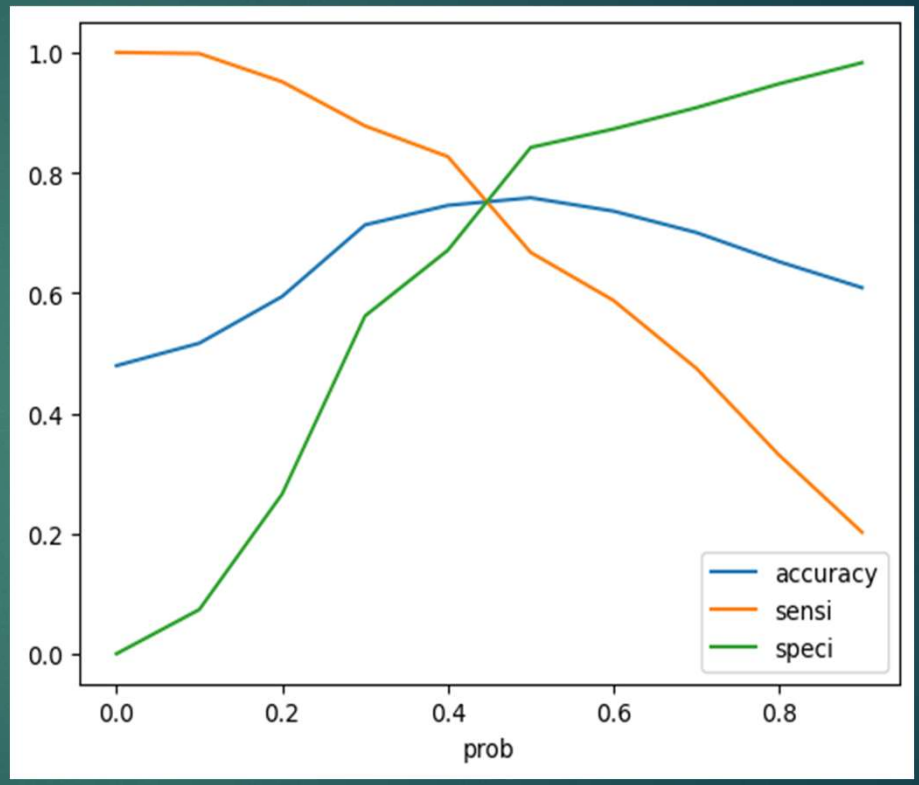
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7724	0.112	6.919	0.000	0.554	0.991
Do Not Email	-0.3699	0.046	-8.120	0.000	-0.459	-0.281
TotalVisits	0.2971	0.055	5.404	0.000	0.189	0.405
Total Time Spent on Website	1.1358	0.044	25.957	0.000	1.050	1.222
Page Views Per Visit	-0.1178	0.060	-1.971	0.049	-0.235	-0.001
Lead Origin_Lead Add Form	4.6446	0.606	7.664	0.000	3.457	5.832
Lead Origin_Lead Import	20.8650	1.77e+04	0.001	0.999	-3.47e+04	3.48e+04
Lead Source_Direct Traffic	-1.6191	0.140	-11.591	0.000	-1.893	-1.345

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

ROC

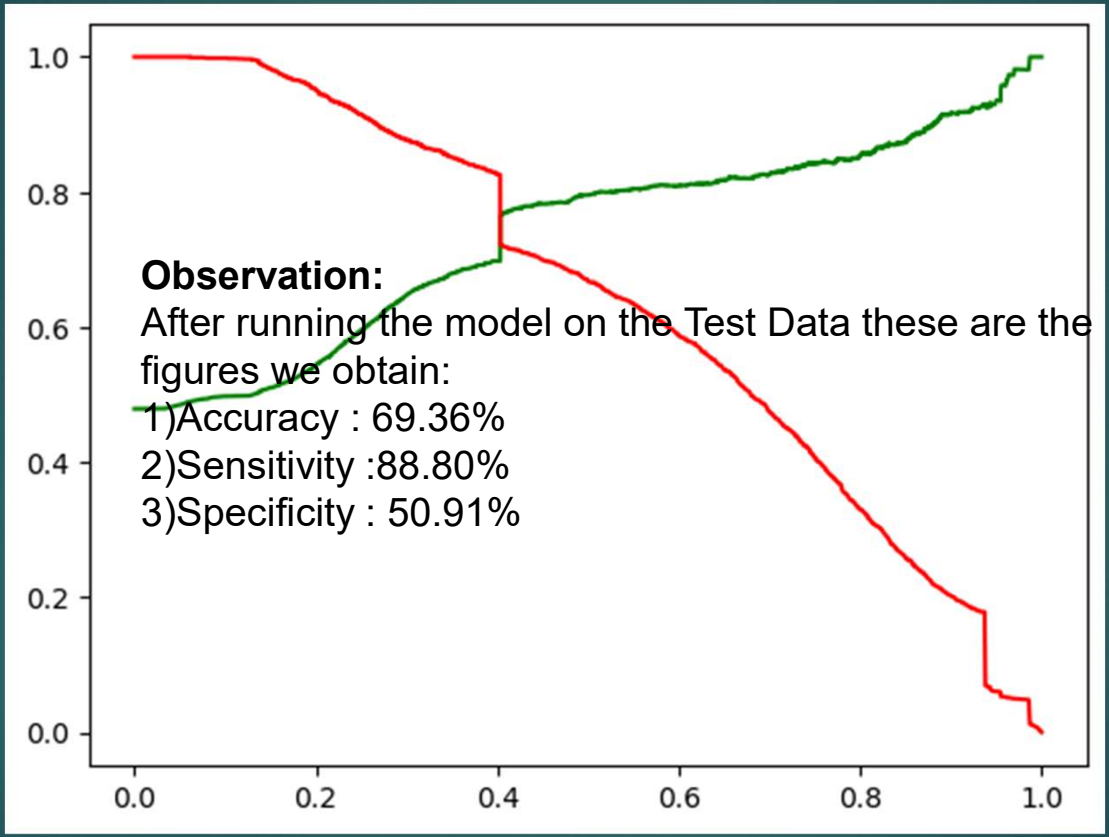


Optimal Cut-off Point



SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Precision and Recall Trade-off



Observation:
After running the model on the Test Data these are the figures we obtain:
1)Accuracy : 69.36%
2)Sensitivity :88.80%
3)Specificity : 50.91%

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

Observation:

After running the model on the Test Data these are the figures we obtain:

1)Accuracy : 69.36%
2)Sensitivity :88.80%
3)Specificity : 50.91%

SUBMISSION BY : T. NARESH KUMAR, NAMRATA SINHA,NARGIS SHEIKH– DS-C54 MARCH 2023

THANK YOU

SUBMISSION BY : T. NARESH KUMAR– DS-C54 MARCH 2023