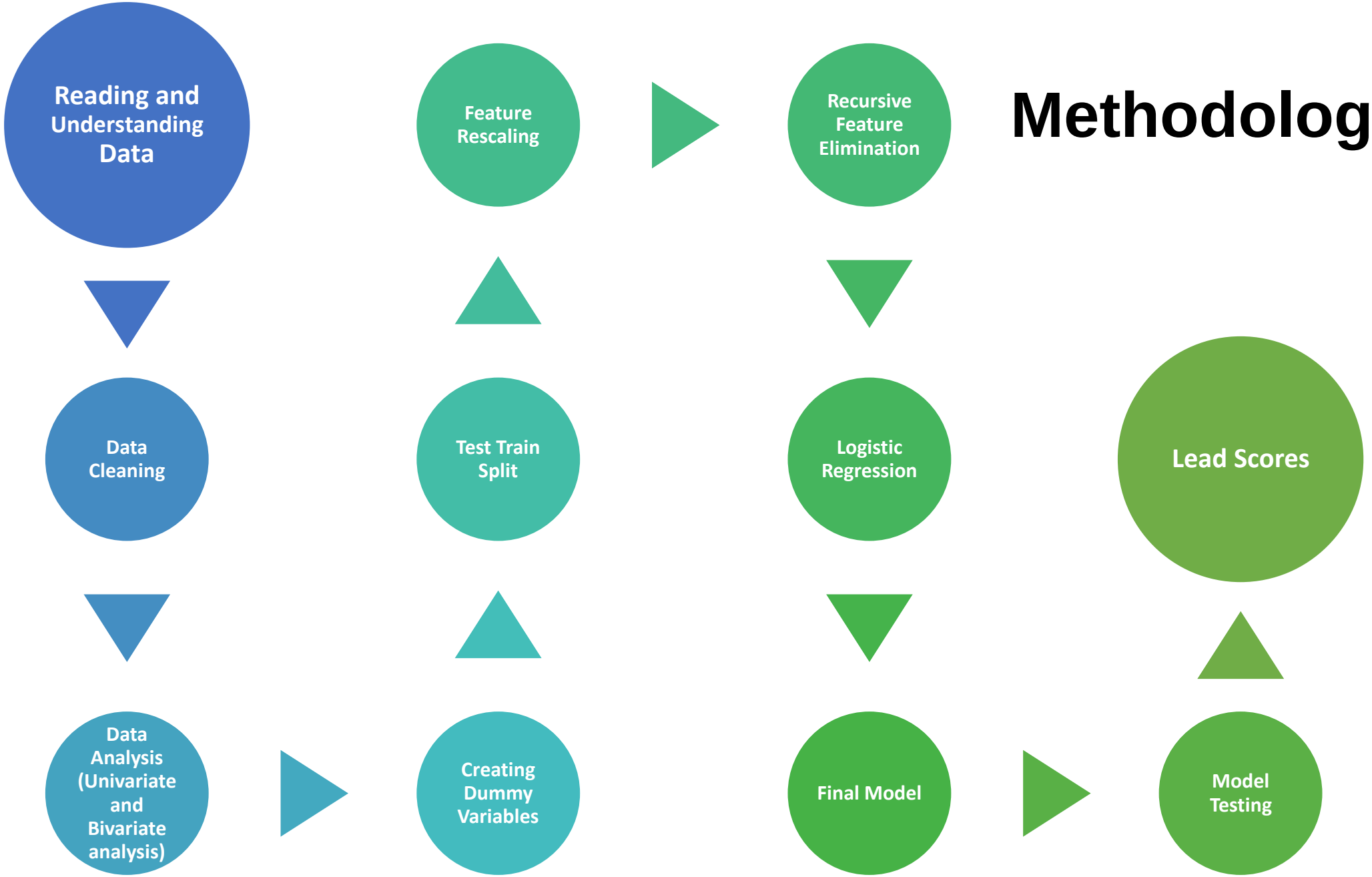
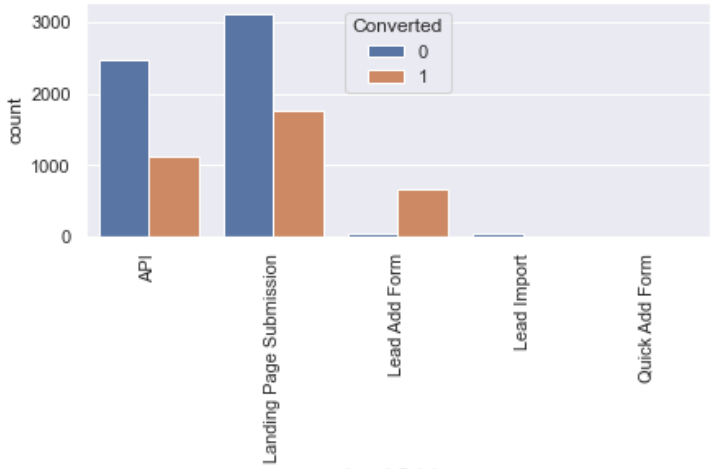
Lead Score Case Study

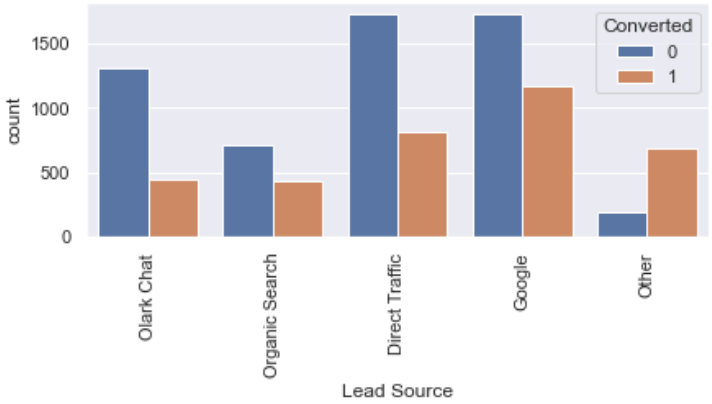
- **Problem Statement**
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When 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%.
- **Business Goal**
- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

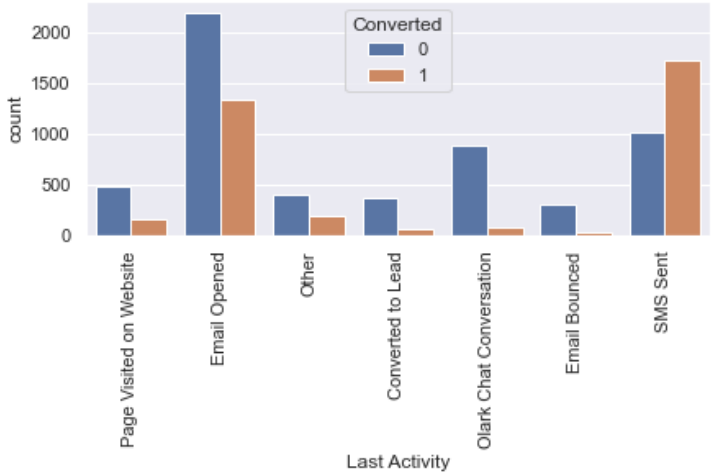
Methodology

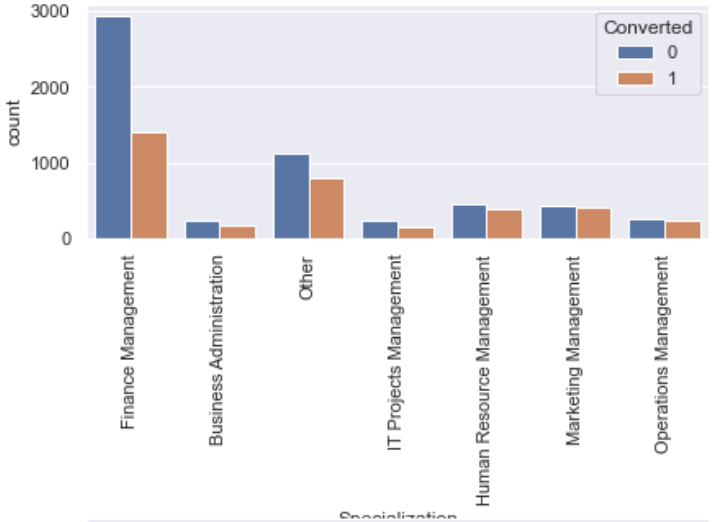


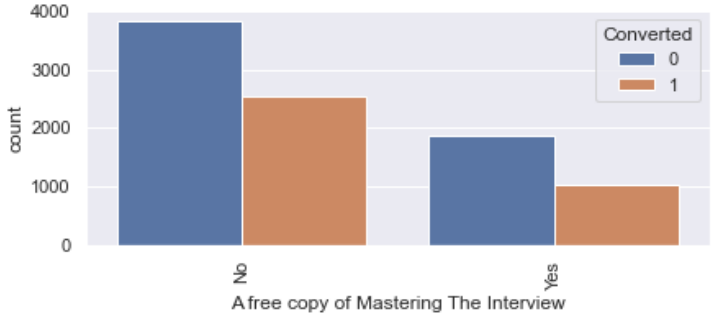
Bivariate Analysis With respect to Target Column Converted

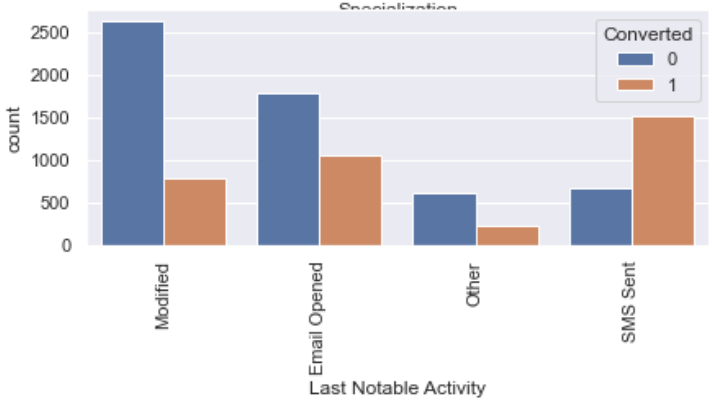


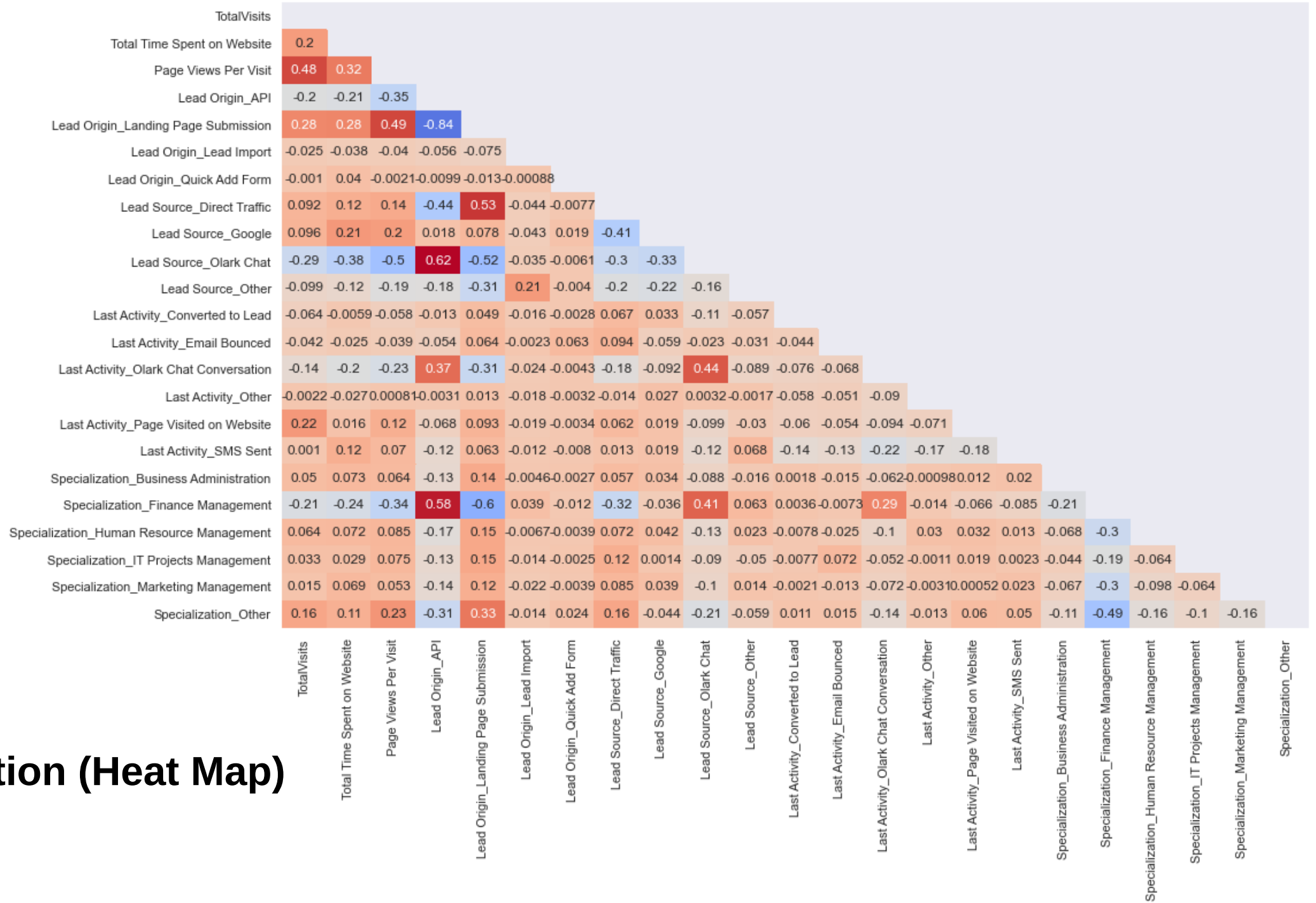












Correlation (Heat Map)

Final Model

Generalized Linear Model Regression Results

Dep. Variable:	Converted	**No. Observations:**	6468
Model:	GLM	**Df Residuals:**	6452
Model Family:	Gaussian	**Df Model:**	15
Link Function:	identity	**Scale:**	0.14832
Method:	IRLS	**Log-Likelihood:**	-2997.9
Date:	Tue, 18 Jul 2023	**Deviance:**	956.95
Time:	18:55:40	**Pearson chi2:**	957.
No. Iterations:	3		
Covariance Type:	nonrobust		

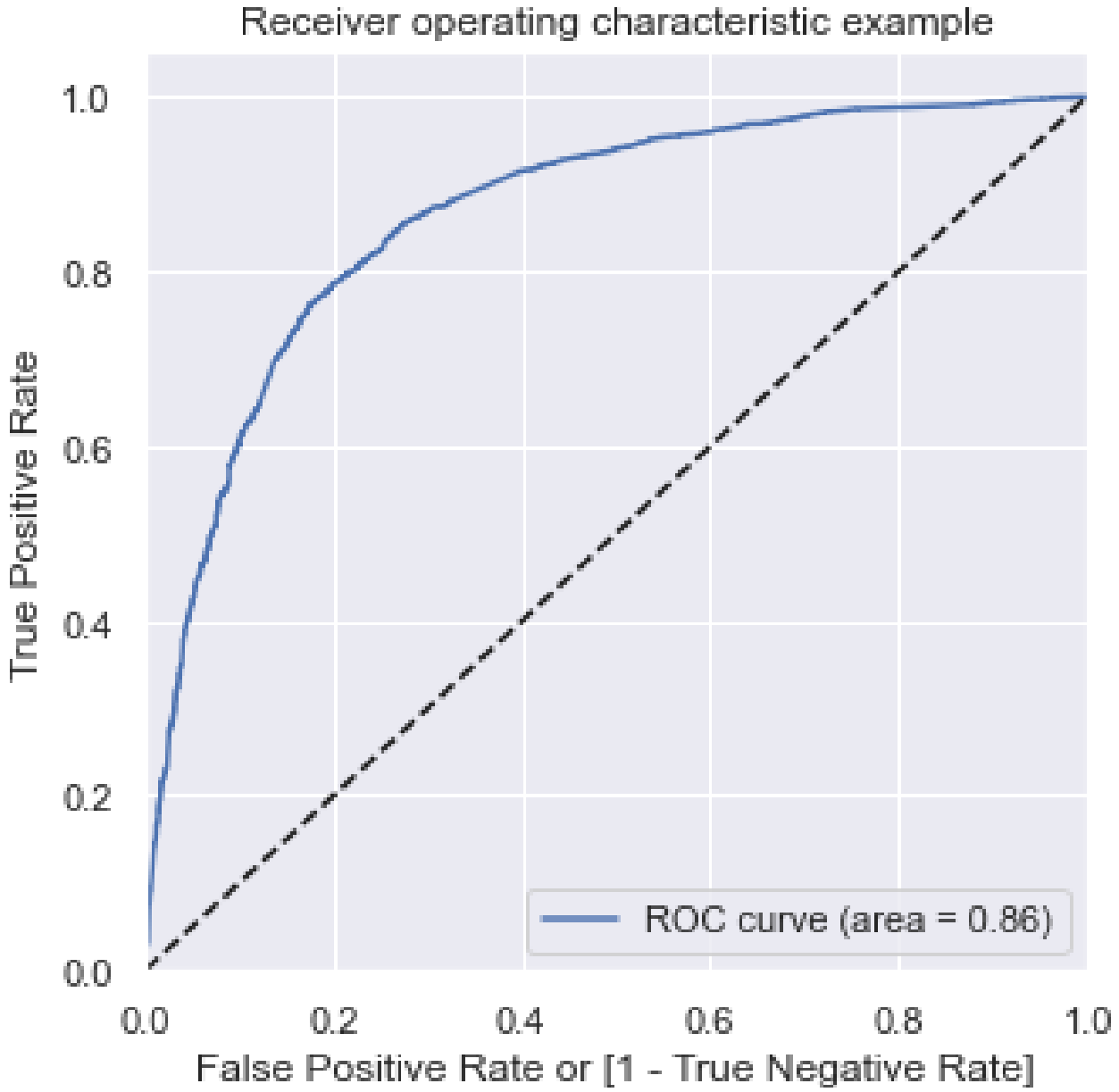
Coefficient Table of the Final Model

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.9651	0.020	47.685	0.000	0.925	1.005
TotalVisits	0.0243	0.006	4.344	0.000	0.013	0.035
Total Time Spent on Website	0.1984	0.005	36.733	0.000	0.188	0.209
Page Views Per Visit	-0.0251	0.007	-3.743	0.000	-0.038	-0.012
Lead Origin_API	-0.5913	0.022	-26.423	0.000	-0.635	-0.547
Lead Origin_Landing Page Submission	-0.6426	0.023	-28.413	0.000	-0.687	-0.598
Lead Origin_Lead Import	-0.6260	0.070	-8.895	0.000	-0.764	-0.488
Lead Source_Direct Traffic	-0.0537	0.013	-4.144	0.000	-0.079	-0.028
Lead Source_Olark Chat	0.1390	0.019	7.478	0.000	0.103	0.175
Last Activity_Converted to Lead	-0.1471	0.024	-6.095	0.000	-0.194	-0.100
Last Activity_Email Bounced	-0.2020	0.026	-7.735	0.000	-0.253	-0.151
Last Activity_Olark Chat Conversation	-0.2171	0.018	-11.820	0.000	-0.253	-0.181
Last Activity_Other	-0.0446	0.021	-2.166	0.030	-0.085	-0.004
Last Activity_Page Visited on Website	-0.1124	0.020	-5.568	0.000	-0.152	-0.073
Last Activity_SMS Sent	0.2078	0.012	17.594	0.000	0.185	0.231
Specialization_Finance Management	-0.0767	0.012	-6.211	0.000	-0.101	-0.053

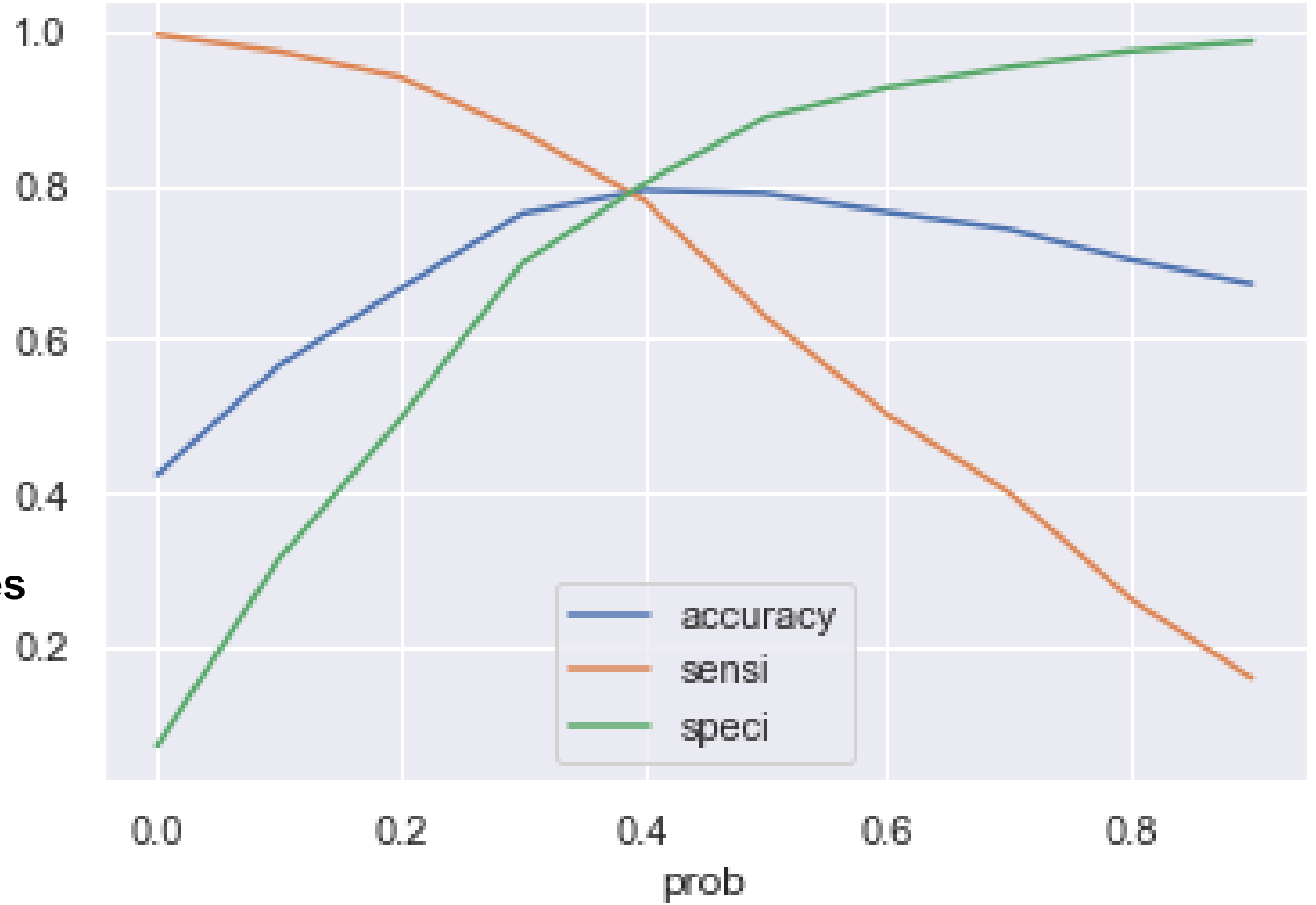
VIF Table for the Final Model

	Features	VIF
3	Lead Origin_API	4.08
4	Lead Origin_Landing Page Submission	3.40
7	Lead Source_Olark Chat	2.88
14	Specialization_Finance Management	2.86
6	Lead Source_Direct Traffic	2.02
2	Page Views Per Visit	1.86
13	Last Activity_SMS Sent	1.62
10	Last Activity_Olark Chat Conversation	1.55
0	TotalVisits	1.36
1	Total Time Spent on Website	1.26
12	Last Activity_Page Visited on Website	1.21
8	Last Activity_Converted to Lead	1.19
11	Last Activity_Other	1.15
9	Last Activity_Email Bounced	1.11
5	Lead Origin_Lead Import	1.02

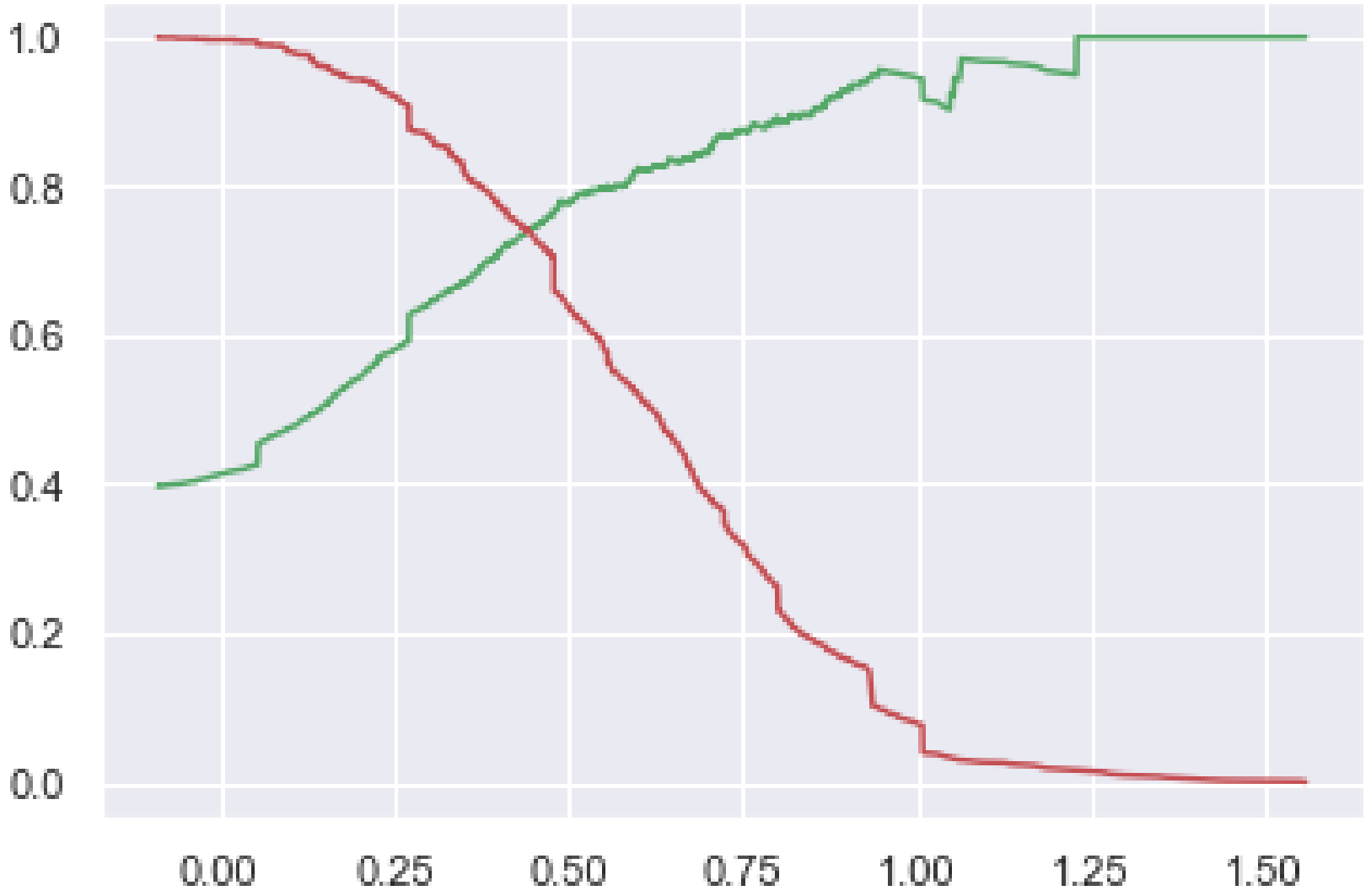
ROC Curve


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.86)

Plot for accuracy sensitivity and specificity for various probabilities



Thresholds: Precision and Recall Curve on Test Data



Confusion Matrix

	N	T
N	1275	402
T	217	878

Accuracy : 77.66%
Sensitivity : 80.18%
Specificity : 76.02%

Conclusion

- From the above analysis we can observe that the leads generated in the initial stage were high, but later only few of them become as a prospective customers.
- out of all the variables observed, 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' contribute most towards the probability of a lead getting converted.
- Focusing on converted leads. it is vital to extract the right information needed for the improving the conversion rate.
- Frequent follow-ups with the prospective customers and the leads will make the conversion rate high.