Lead Scoring – Case Study

Problem Statement

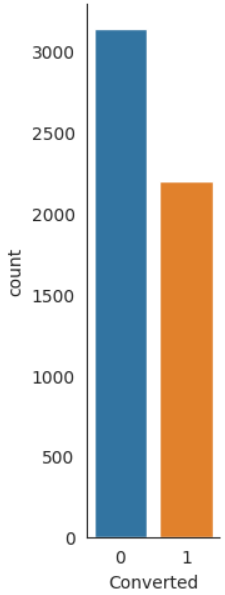
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- When these people fill up a 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

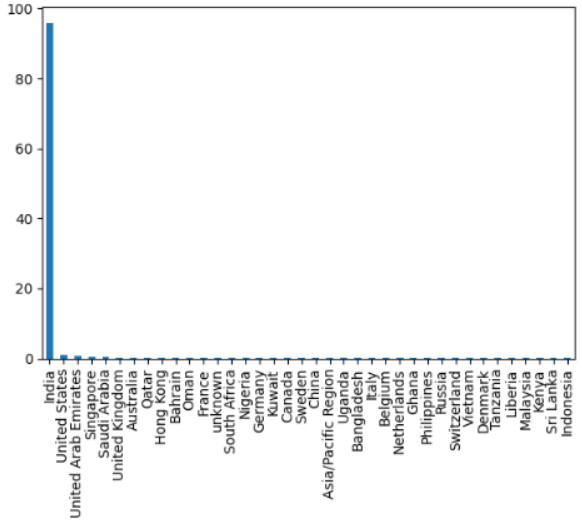
- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and 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%

Exploratory Data Analysis

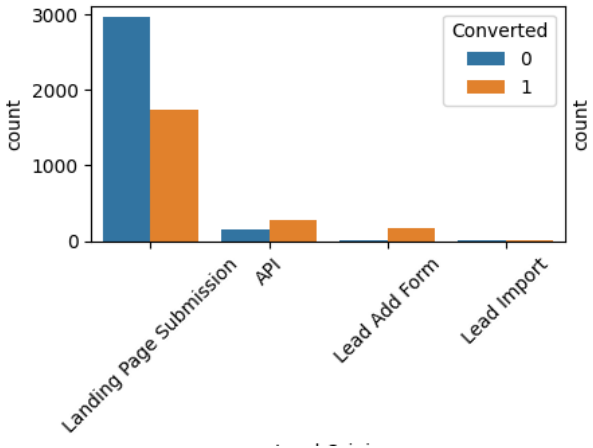
- We have around 39% Conversion rate in Total



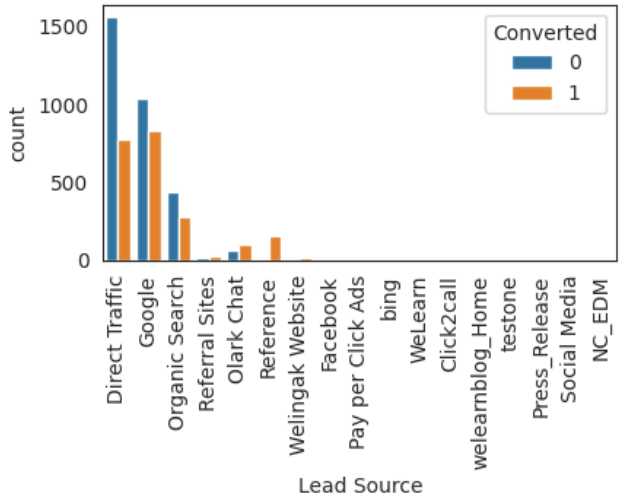
- Major traffic is from Indian customers only



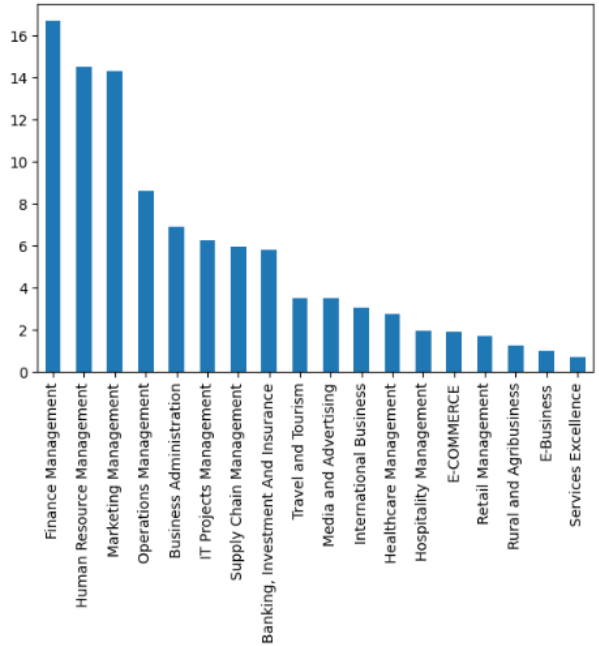
- In Lead Origin, maximum conversion happened from Landing Page Submission

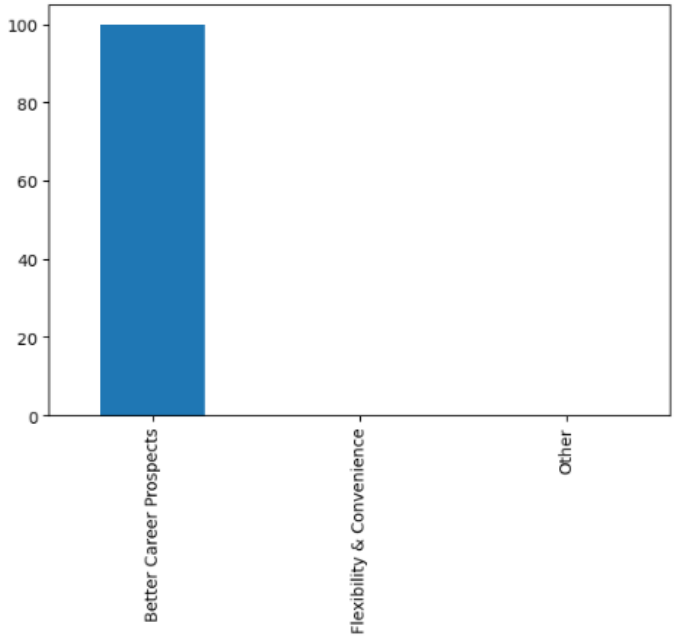


- Major conversion in the lead source is from Google



- All management related courses seems to be more interesting, and people are mainly looking at these courses for better career prospects.



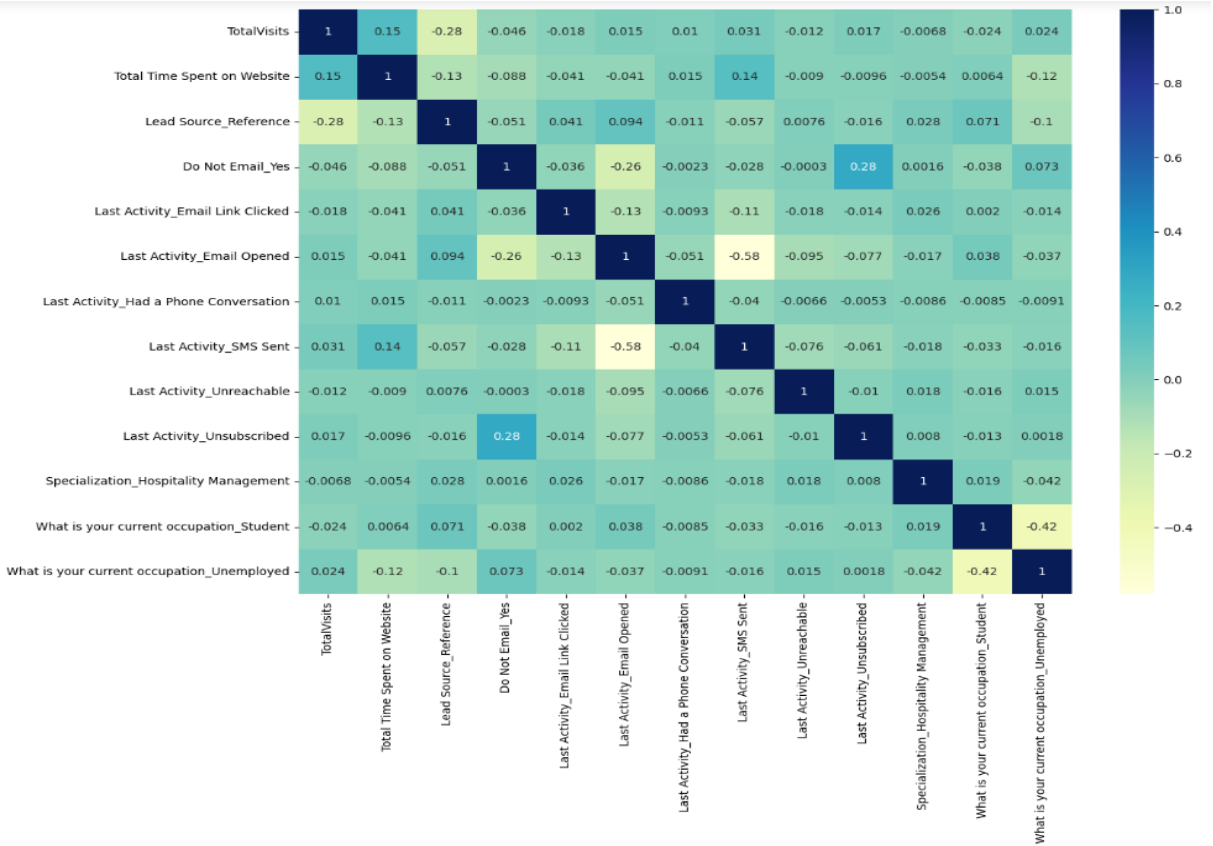


Model Evaluation

Dep. Variable:	Converted	**No. Observations:**	3728
Model:	GLM	**Df Residuals:**	3714
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-1741.0
Date:	Sun, 12 Mar 2023	**Deviance:**	3482.0
Time:	18:38:56	**Pearson chi2:**	3.95e+03
No. Iterations:	7	**Pseudo R-squ. (CS):**	0.3420
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3648	0.269	1.356	0.175	-0.162	0.892
TotalVisits	0.4571	0.159	2.882	0.004	0.146	0.768
Total Time Spent on Website	4.0353	0.183	22.093	0.000	3.677	4.393
Lead Source_Reference	5.3182	0.601	8.852	0.000	4.141	6.496
Do Not Email_Yes	-1.4666	0.227	-6.448	0.000	-1.912	-1.021
Last Activity_Email Link Clicked	0.5891	0.300	1.965	0.049	0.002	1.177
Last Activity_Email Opened	0.7175	0.128	5.608	0.000	0.467	0.968
Last Activity_Had a Phone Conversation	2.7923	0.774	3.609	0.000	1.276	4.309
Last Activity_SMS Sent	1.5496	0.130	11.906	0.000	1.295	1.805
Last Activity_Unreachable	0.6283	0.385	1.634	0.102	-0.125	1.382
Last Activity_Unsubscribed	1.8260	0.525	3.477	0.001	0.797	2.855
Specialization_Hospitality Management	-0.6046	0.331	-1.824	0.068	-1.254	0.045
What is your current occupation_Student	-2.4377	0.372	-6.556	0.000	-3.166	-1.709
What is your current occupation_Unemployed	-3.2356	0.244	-13.253	0.000	-3.714	-2.757

Correlation Matrix



Conclusion

- While we have checked both Sensitivity-Specificity , we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction at 0.4
- The overall accuracy of the prediction is around 79% calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 79% (in train set) and 78% in test set
- Below are the top 3 variables that contribute for lead getting converted in the model are
  - Total time spent on website
  - Lead Source - Reference
  - Last Notable Activity - Had a Phone Conversation

The model on an overall level seems good