

Lead Score Case Study

January 2023

Divya N L
Sampa Kar
Prathamesh Ingle

CONTENTS

Problem statement

Steps involved in Analysis

Conclusions &
Recommendations

Problem Statement



An education company named X Education sells online courses to industry professionals.



The company requires to build *a logistic regression model* wherein we need to assign a lead score 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.

Steps taken in Solving the Problem

Cleaning up data after treating missing values and outliers.

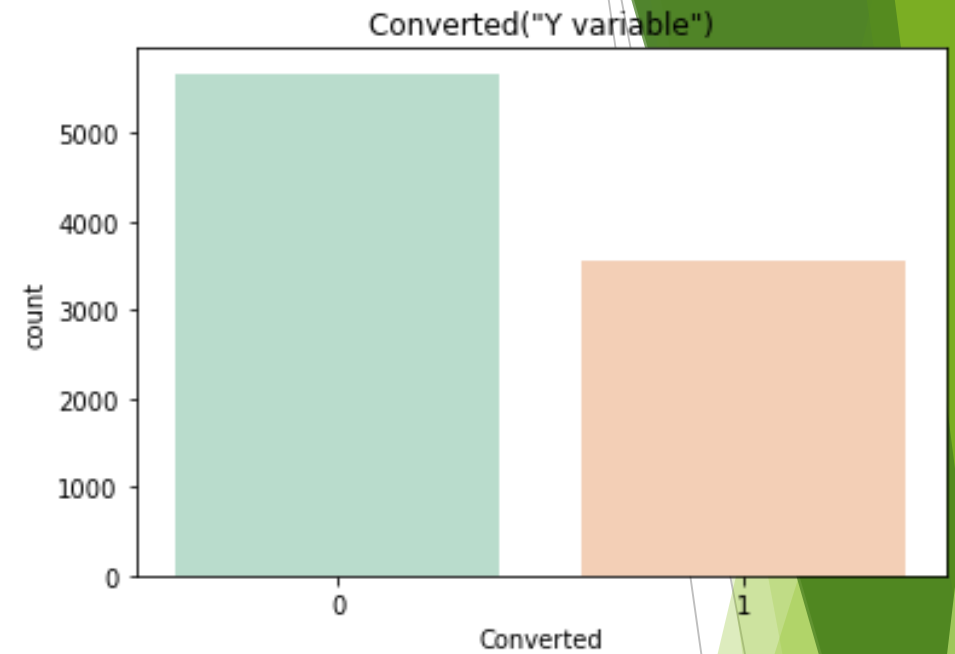
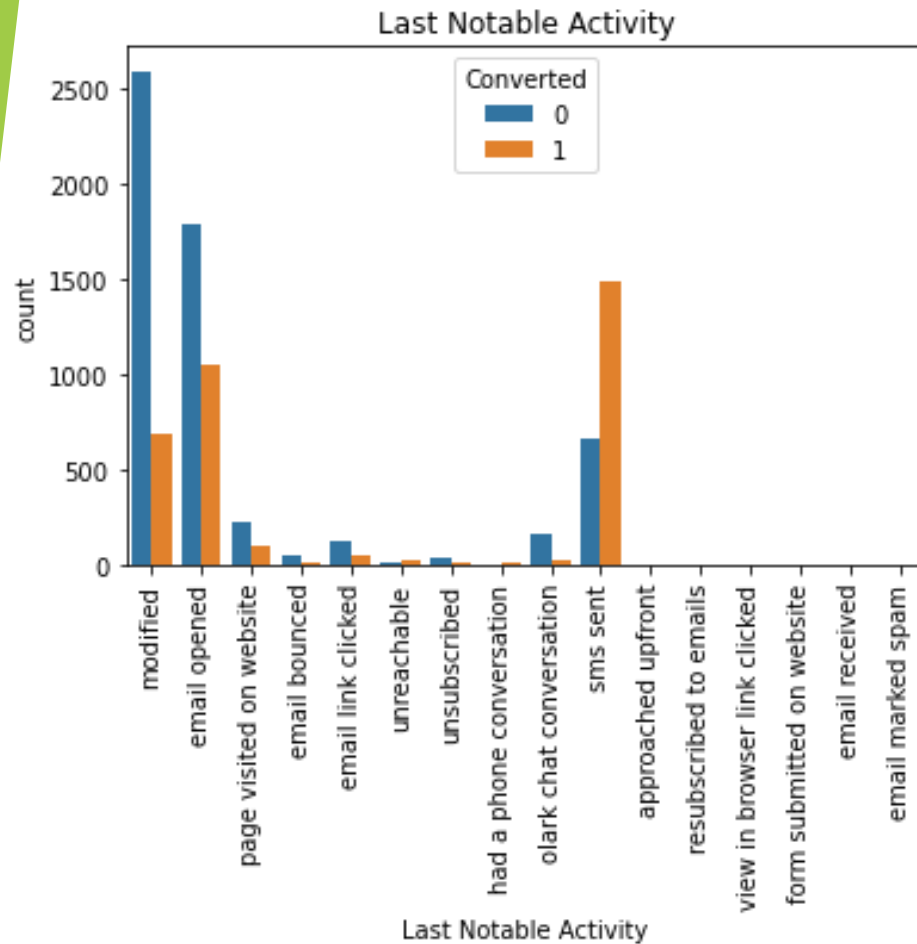
Creating train and test data by splitting the cleaned data set.

Created dummy variables and scaling the data.

Selected 15 features using Recursive Feature Elimination (RFE)

Applied Logistic Regression algorithm to build a model and more than 81% accuracy and 70% sensitivity.

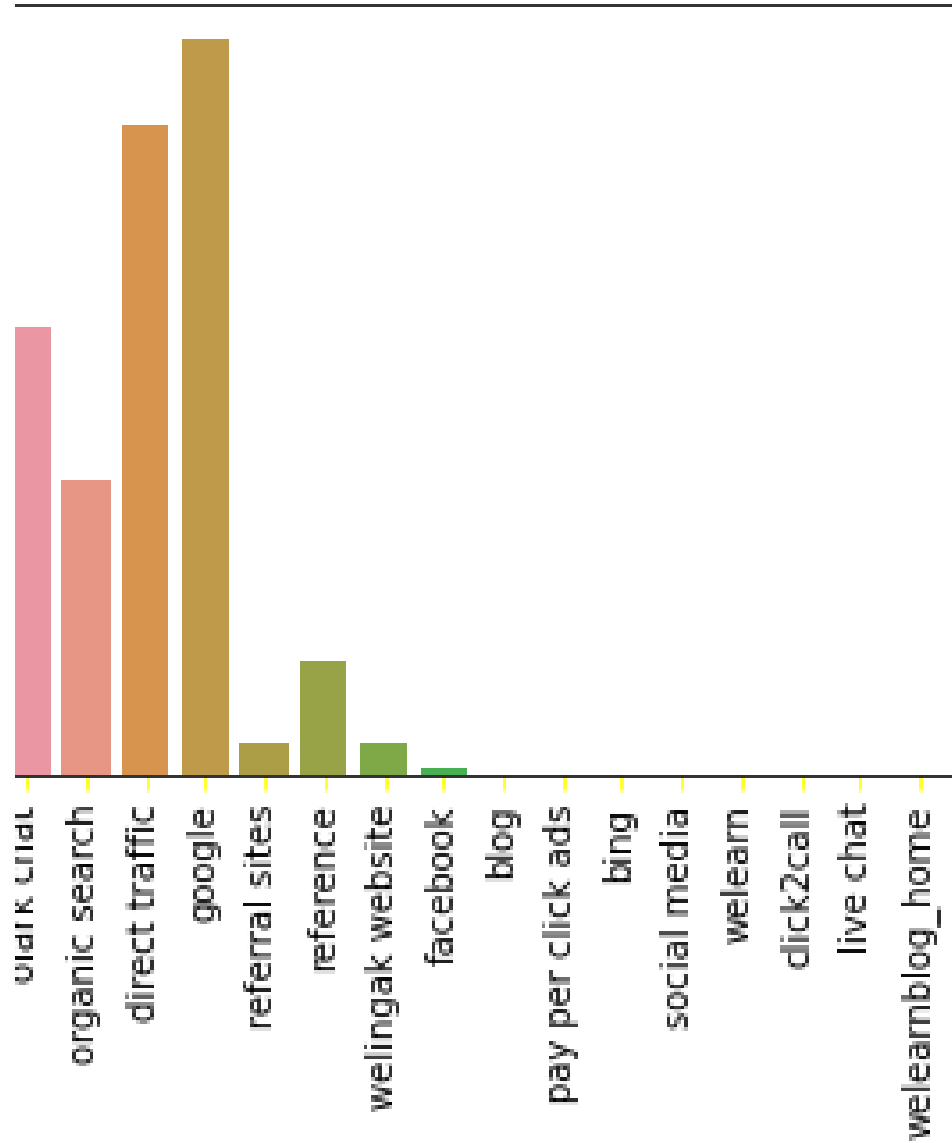
Identified the optimal probability cutoff from the accuracy, sensitivity and specificity.



EDA

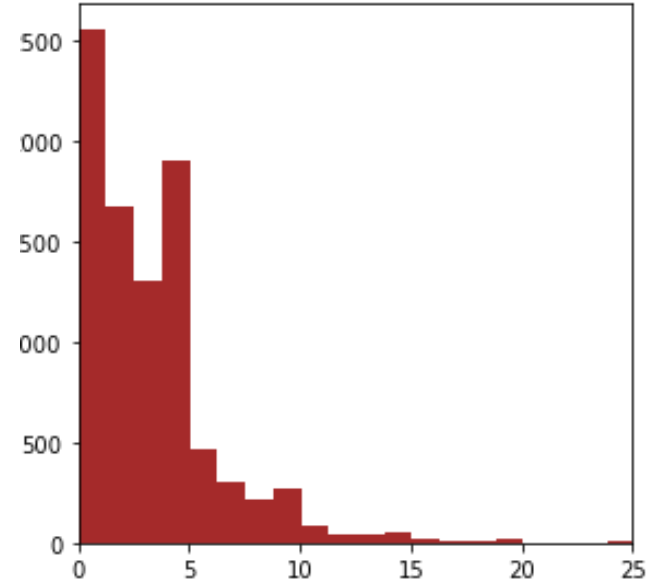
Univariate Analysis- Categorical Variables

Lead Source

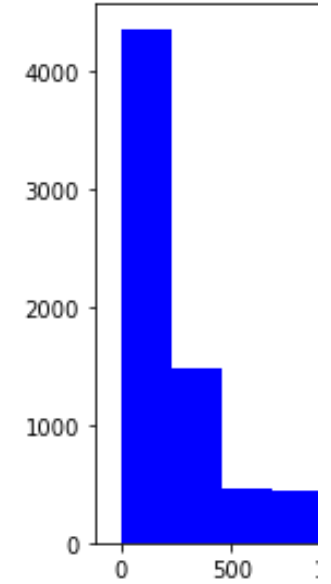


Lead Source

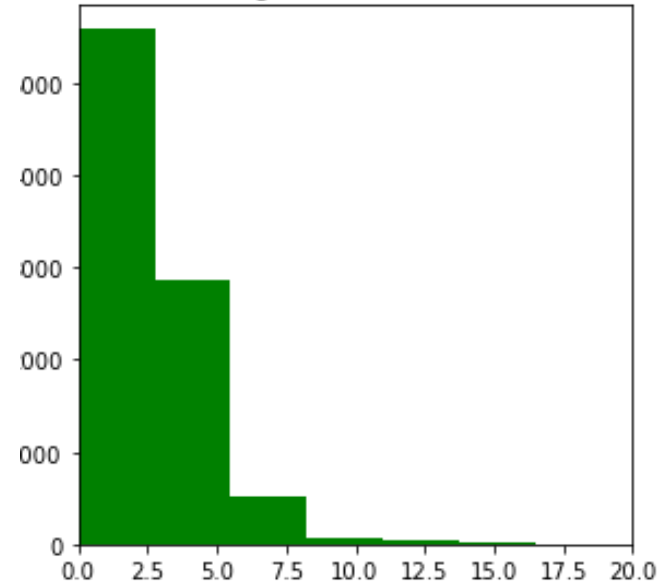
Total Visits

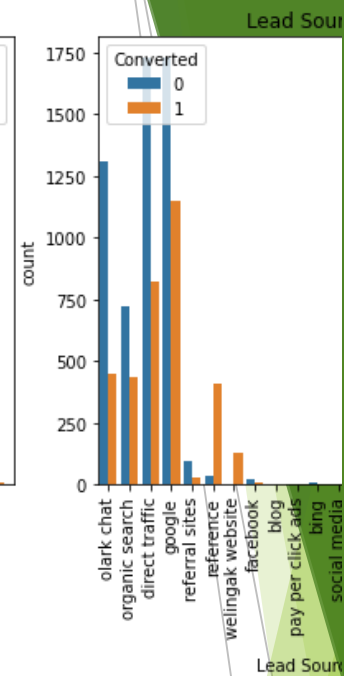
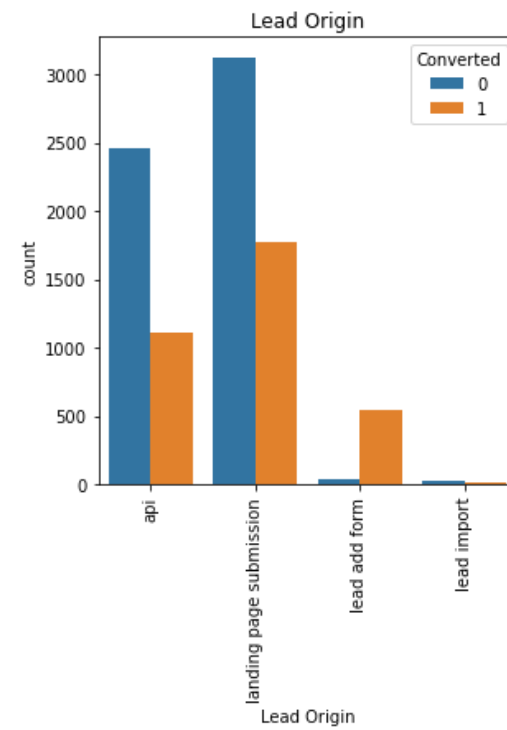


Total Time



Page Views Per Visit





Correlation among variables

Data Handling

No major Outliers



Dummy variables have been
created for Object type variable



Numerical variables have been
normalized

Model Building

Split data into train (70%) and test (30%) data

Running RFE with 15 variables as output

Building model by removing the variable where p value is > 0.05 and VIF value is > 5

Predictions on test data set

Overall Accuracy is 81%

Sensitivity is around 70%

Specificity is around 88%

Model Results

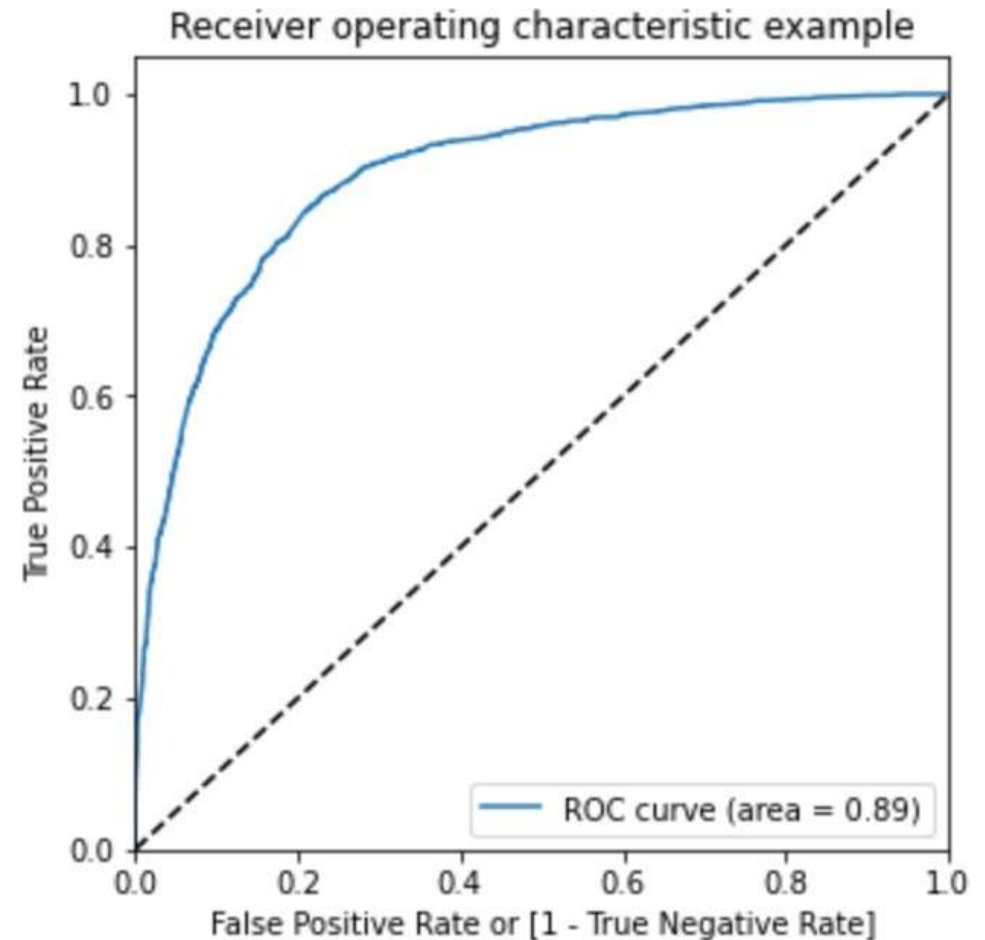
Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	6351
Model:	GLM	Df Residuals:	6337
Model Family:	Binomial	Df Model:	13
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2651.3
Date:	Mon, 02 Jan 2023	Deviance:	5302.6
Time:	16:59:37	Pearson chi2:	6.50e+03
No. Iterations:	7		
Covariance Type:	nonrobust		

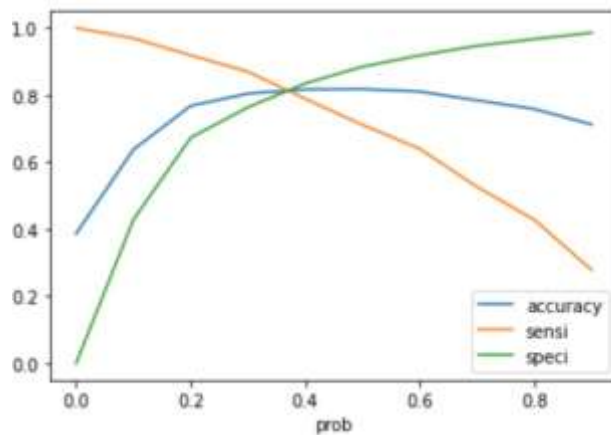
	coef	std err	z	P> z	[0.025	0.975]
const	-3.4533	0.113	-30.579	0.000	-3.675	-3.232
TotalVisits	5.5427	1.444	3.838	0.000	2.712	8.373
Total Time Spent on Website	4.6048	0.166	27.690	0.000	4.279	4.931
Lead Origin_lead add form	3.7501	0.225	16.651	0.000	3.309	4.192
Lead Source_olark chat	1.5802	0.111	14.187	0.000	1.362	1.798
Lead Source_welingak website	2.5821	1.033	2.500	0.012	0.558	4.607
Do Not Email_yes	-1.4360	0.170	-8.437	0.000	-1.770	-1.102
Last Activity_olark chat conversation	-1.3974	0.167	-8.348	0.000	-1.725	-1.069
Last Activity_sms sent	1.2672	0.074	17.164	0.000	1.123	1.412
What is your current occupation_other	2.1567	0.755	2.857	0.004	0.677	3.636
What is your current occupation_student	1.2456	0.226	5.502	0.000	0.802	1.689
What is your current occupation_unemployed	1.1632	0.086	13.582	0.000	0.995	1.331
What is your current occupation_working professional	3.6797	0.204	18.008	0.000	3.279	4.080
Last Notable Activity_unreachable	1.8153	0.601	3.022	0.003	0.638	2.993

ROC Curve

- ▶ The ROC curve shows that the 88% of the curve is under the curve
- ▶ The classification probability of lead conversion is very high by the model



Optimal probability Cutoff



- ▶ Optimal probability cutoff is identified as 0.35 for better accuracy of the classification of lead conversion.
- ▶ With the current cut off as 0.35 we have accuracy, sensitivity and specificity of around 80%.

Actual/Predicted	Not Converted	Converted
Not Converted	3445	450
Converted	710	1746

Confusion Matrix

Observations And Recommendations

► The variables that are significant in identifying 'Hot Leads' are : (in descending order)

- Total number of time spent on websites
- Total number of visits
- When the lead source was:
 - a. Google
 - b. Direct traffic
 - c. Organic search
 - d. Welingak website
- When the last activity was:
 - a. SMS
 - b. Olark chat conversation
- When the lead origin is Lead add format.
- When their current occupation is as a working professional.

► Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.