Lead Score Case Study (Logistics Regression)

BY - HARISH (HARIOM) BHAYAL,, HARISH KISHOR PATIL & HARSH ANAND
EMAIL - HARIOMBHAYAL86@GMAIL.COM
HARISHDESALE4629@GMAIL.COM
HARSH.ANAND3@GMAIL.COM

Problem Statement

X Education is an organization which provide online courses for industry professional. There presence on major online website like Google.

They want to select most promising leads that can be converted to paying customers. But in all those process only 30% leads convert in pay customer.

Now we will help to find out concerning point & implement make course of action for higher success rate.

Objectives

The company requires a model to be built for selecting most promising leads.

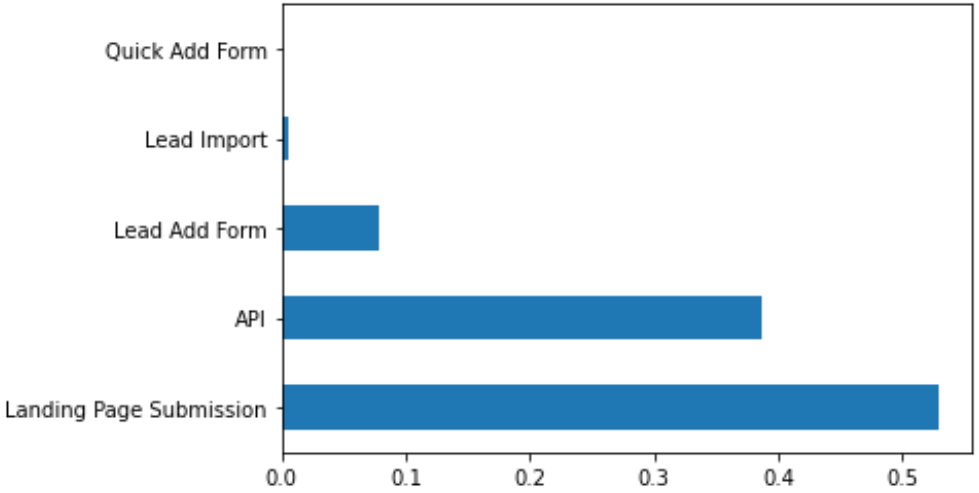
The model to be built in lead **conversion rate around 80% & more from 30%.**

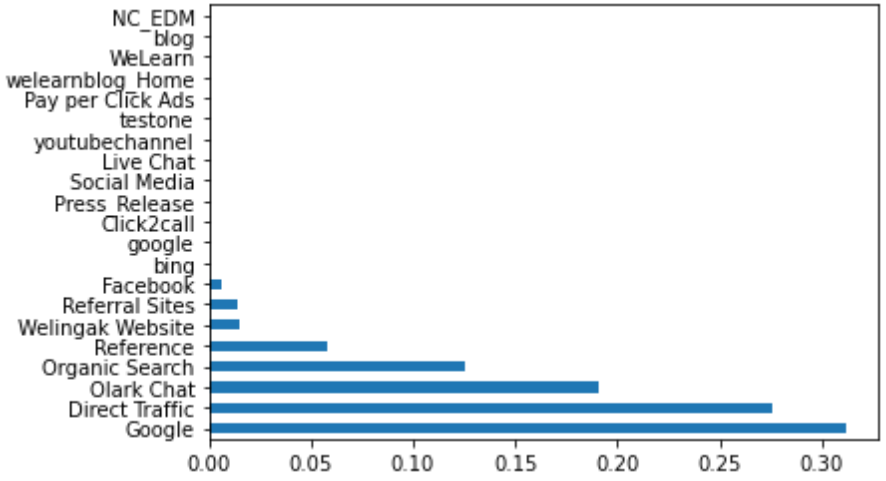
Steps

- Data Collection
- Data Mining(Cleaning & prepare for analysis)
- Apply EDA for figure out most helpful attributes for conversion
- Scaling Features
- Build Model(Logistics Regression)
- Test the model on train data set
- Evaluate Model on different required parameters
- Test the model on data set
- Measure accuracy of the model
- Conclusion

EDA- UNIVARIATE ANALYSIS

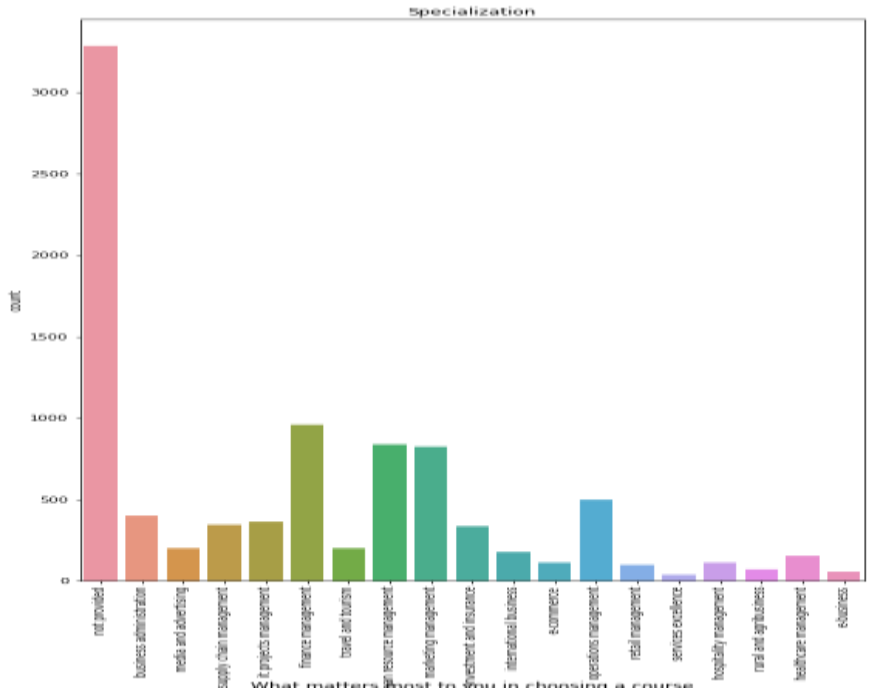
LEAD ORIGIN VS LEAD SOURCE

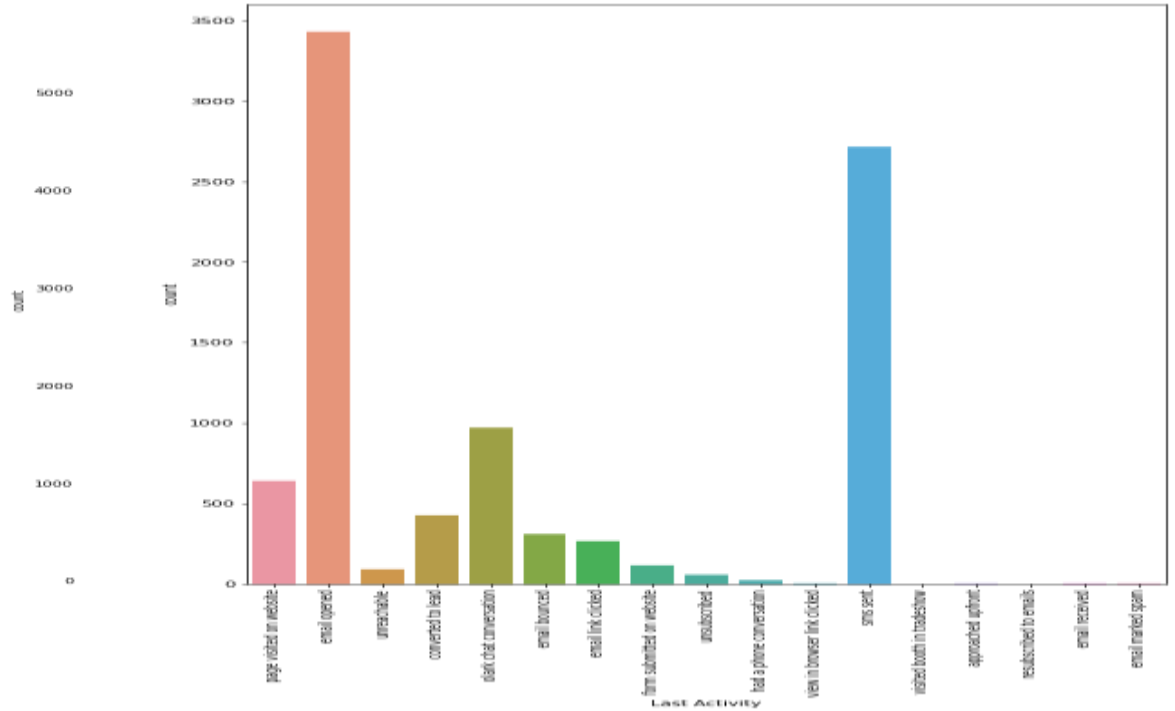




EDA- UNIVARIATE ANALYSIS

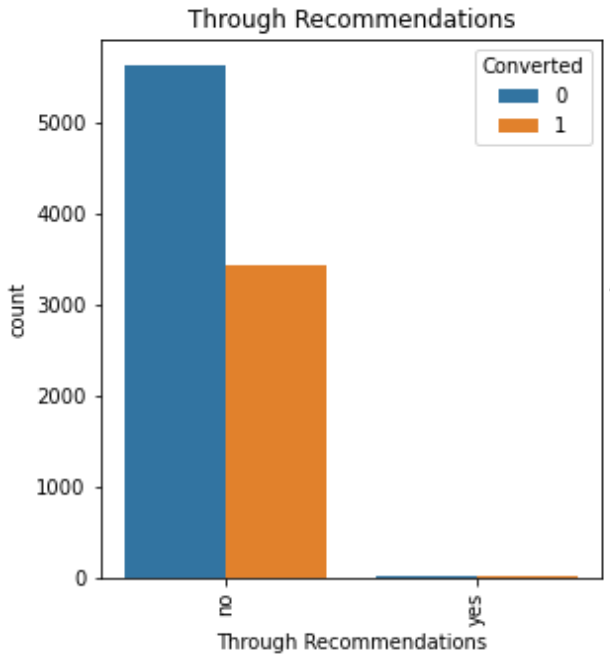
OPTING COURSE VS LAST COURSE OF ACTION

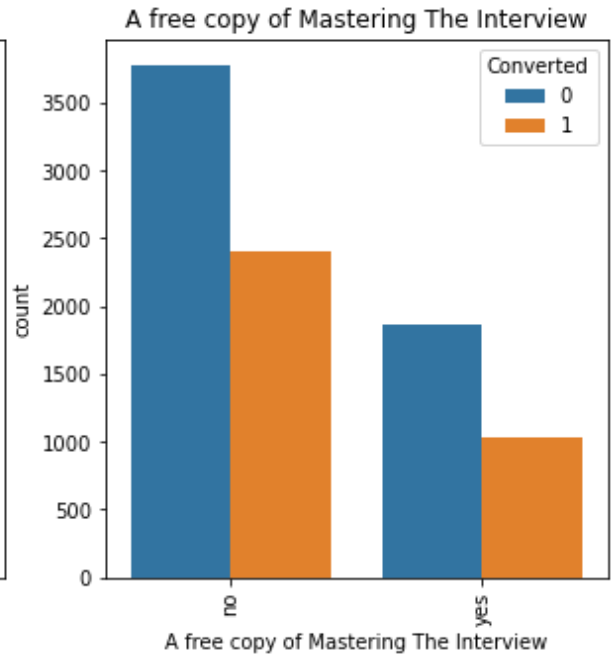


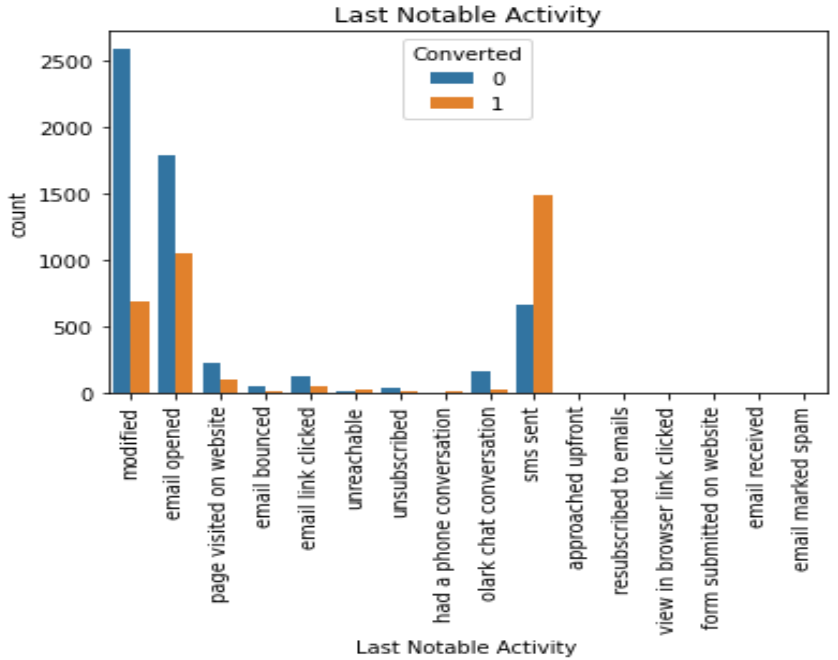


EDA- BIVARIATE ANALYSIS

CONVERSION COMPARIOSN







Model Building

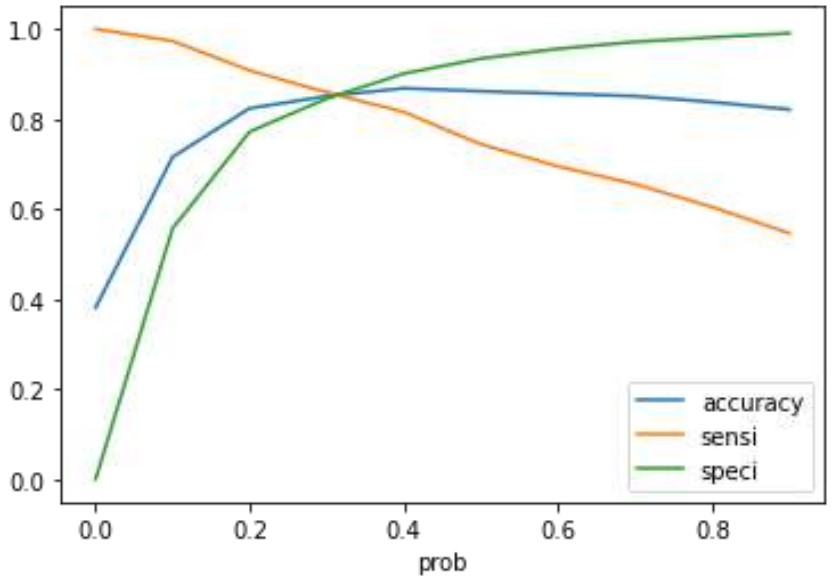
Now we starting to build Model based upon train data set. For we find VIF in different different way to minimise P-value to better outcomes.

	Features	VIF
9	What is your current occupation_unemployed	2.29
1	Total Time Spent on Website	2.06
0	TotalVisits	1.84
2	Lead Origin_lead add form	1.58
7	Last Activity_sms sent	1.53
3	Lead Source_olark chat	1.51
6	Last Activity_olark chat conversation	1.37
10	What is your current occupation_working profes...	1.32
4	Lead Source_welingak website	1.31
5	Do Not Email_yes	1.06
8	What is your current occupation_student	1.05
11	Last Notable Activity_unreachable	1.01

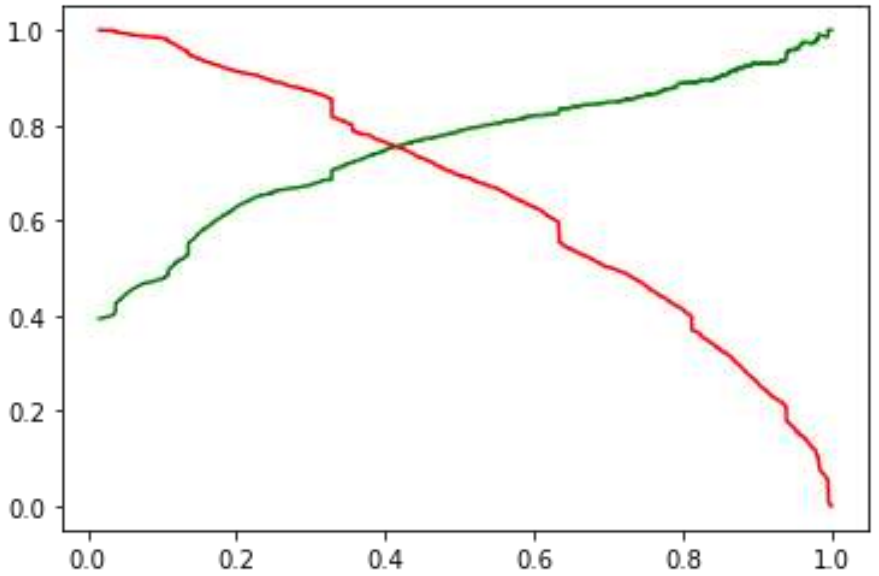
All VIF value is okay. So we can fix model.

Evaluation Of Model(Test)

Accuracy, Sensitivity & Specificity

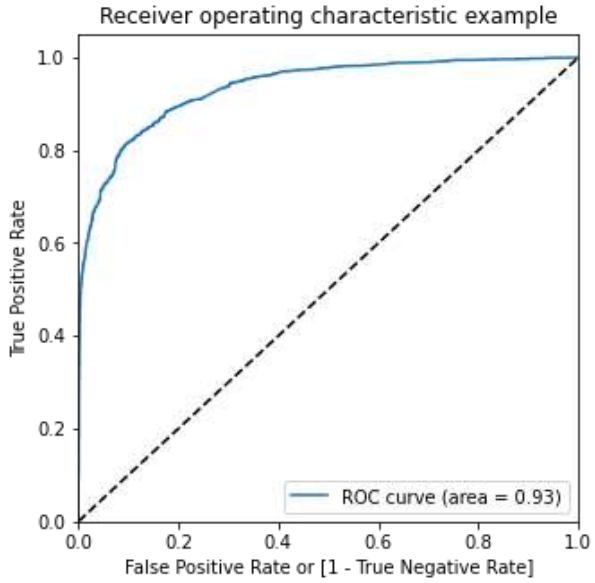


Precision & Recall

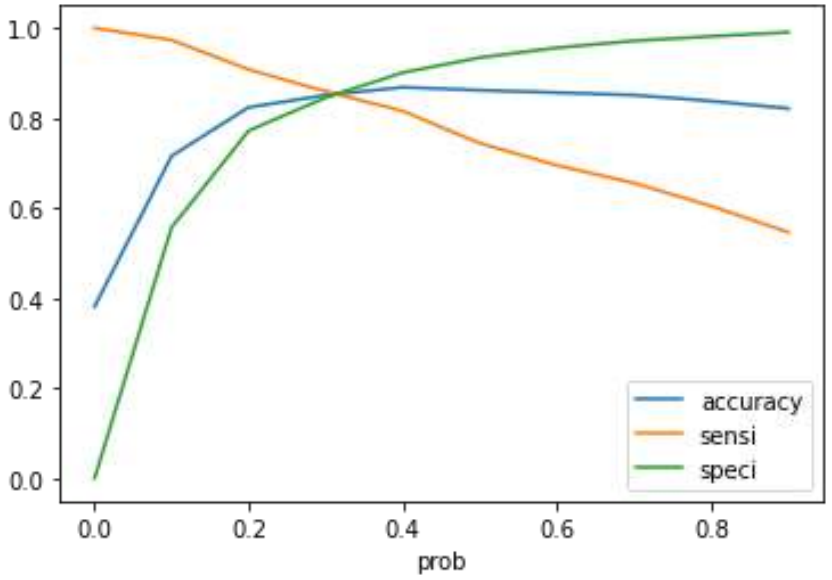


Evaluation Of Model(Train)

ROC



Accuracy, Sensitivity & Specificity



Observations

- As per data set- Accuracy comes almost 81%, Sensitivity is 70% & specificity is around 87% @ cut off 0.42.
- Precision comes around 79% & recall is 70%.
- ROC curve comes .93.
- It was found that the variables that mattered the most in the potential buyers are (In descending order) : -

I. The total time spend on the Website.
II. Total number of visits.
III. When the lead source was: Google , Direct Traffic , Organic Search, Welingak Website
IV. Last activity was- SMS, Olark chat conversation.

Thank You