

Lead Scoring Case Study

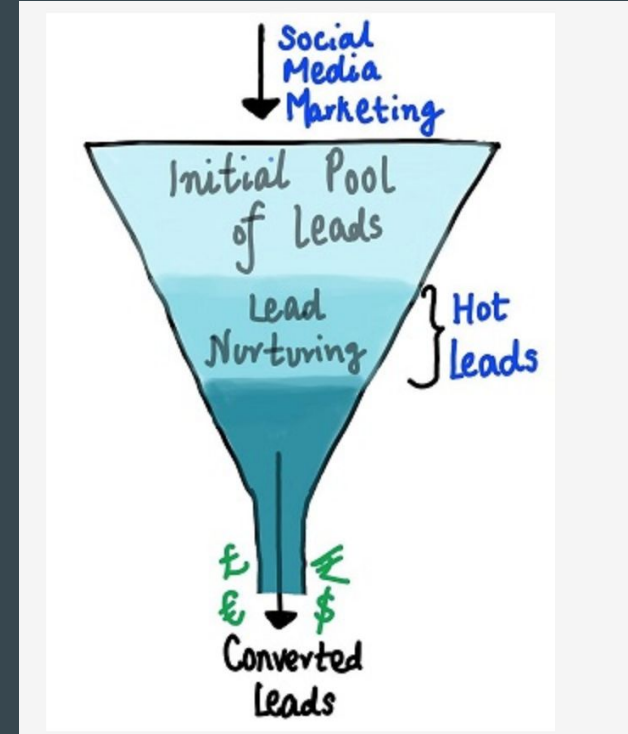


Sagir Mehmood
Vaibhav Sah

Problem at Hand

X Education provides Online Education Platform

- Currently has 30% Lead Conversion
- Increase Lead conversion rate to 80%
- Filter relevant parameters for choosing Hot Leads



Choosing an Approach

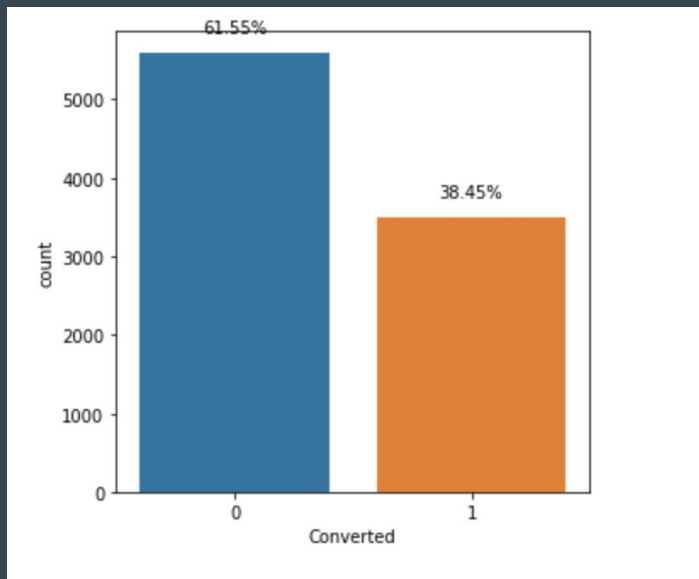
- Most of the Data is Categorical (Non - Numeric) and Binary
- We have to analyse relation b/w parameters
- Required a Predictive Analysis Model

Logistic Regression

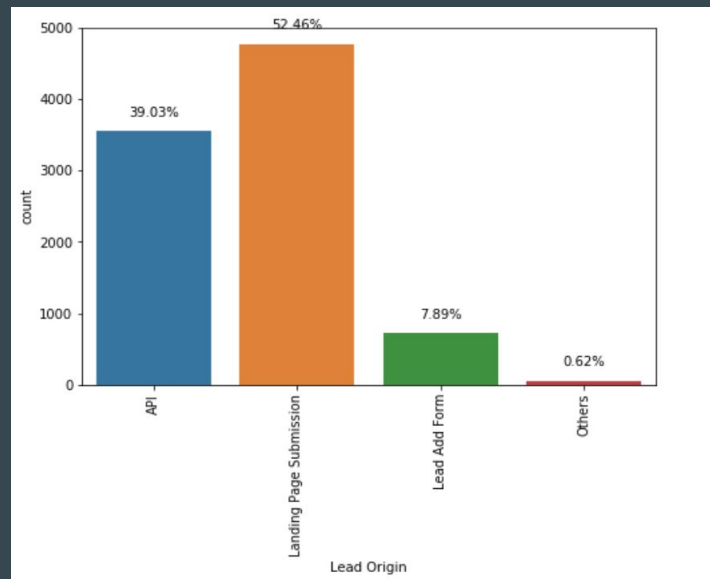
- Remove Highly Skewed Categorical Data
- Outliers Treatment
- Filtering Highly Correlated Data (Decreasing Multicollinearity)
- Scaling of Numerical Data
- Created a Model using Train Data
- Predicting Model generated on the Test Data

Exploratory Data Analysis - Univariate

Initial Lead Conversion Percentage

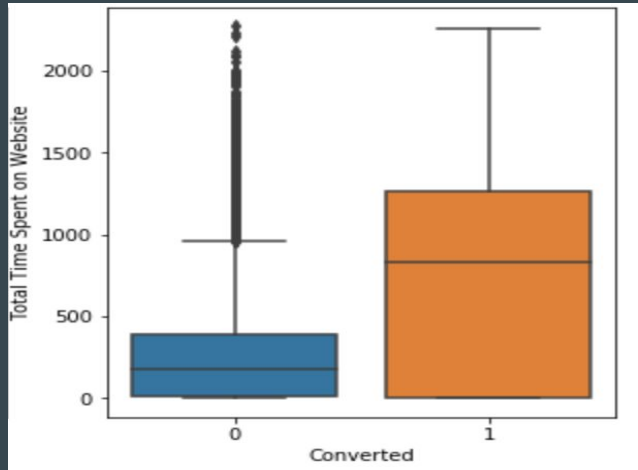


Percentage of Leads from different Origins

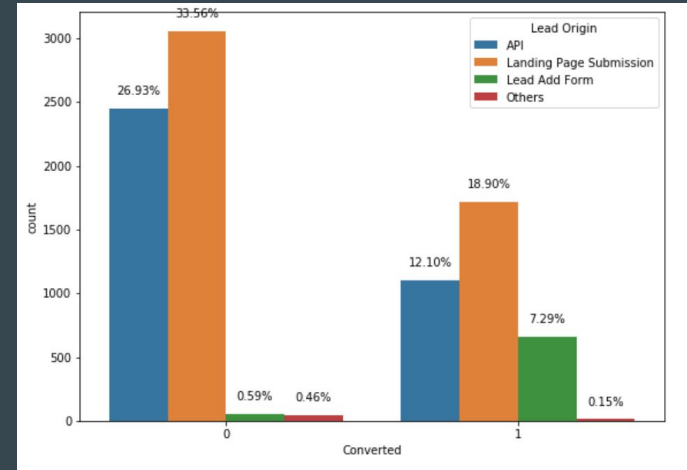


Exploratory Data Analysis - Bivariate

Average Time spent on Website



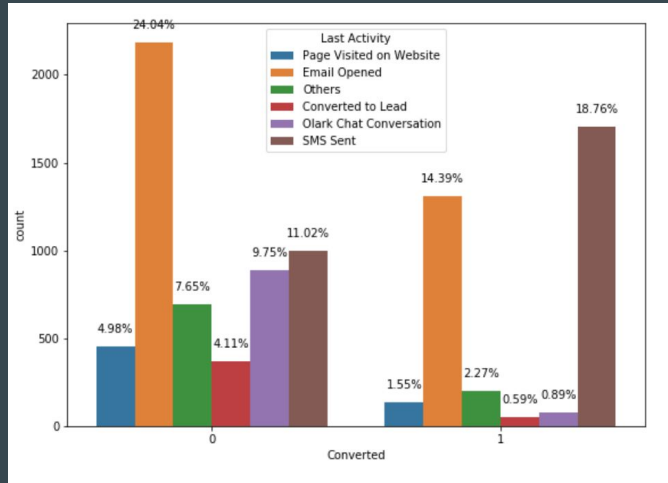
Lead Origin vs Converted



- Leads originated from Lead Add Form are more likely to Convert
- Leads spending more time on website are more probable conversions

Exploratory Data Analysis - Bivariate

Last Activity impact



1. Leads whose Last Activity is SMS Sent are more likely to convert
2. Leads who are already converted must not be called

Logistic Regression Model

Variables	Coefficient
const	-1.4575
Total Time Spent on Website	1.0201
API	0.4476
Lead Add Form	4.0795
Converted to Lead	-1.0033
Page Visited on Website	-0.362
SMS Sent	1.3736
Marketing Management	0.4415
Operations Management	0.4317

Positive Impact

- Total Time Spent on Website
- Last Activity SMS Sent
- Lead Originated from API / Lead Add Forms

Negative Impact

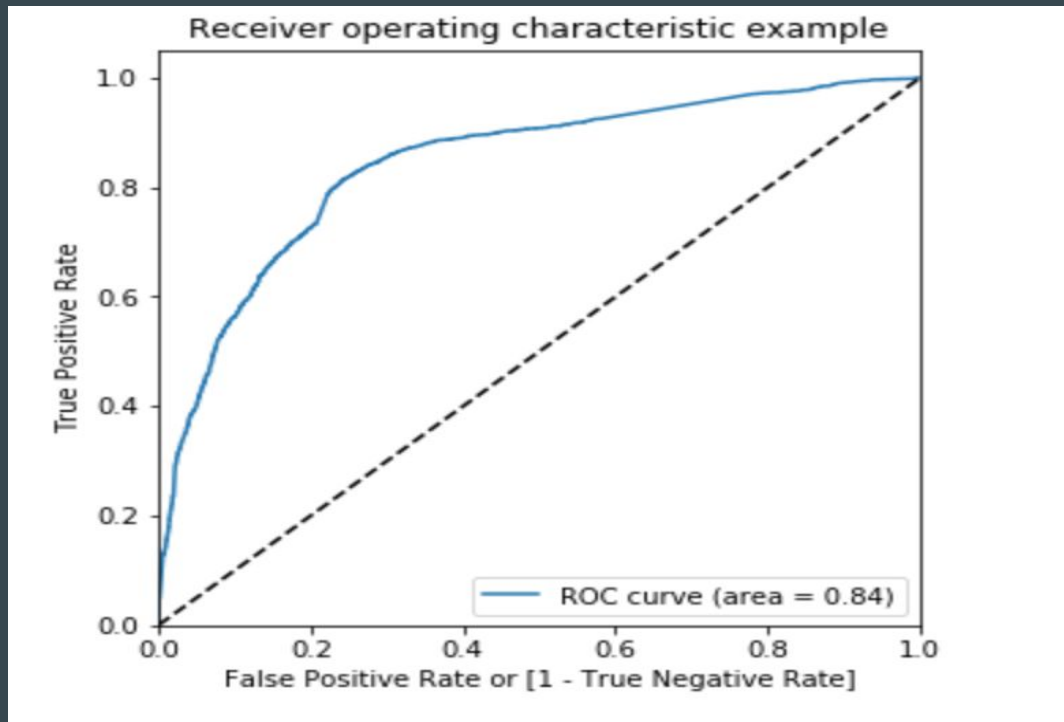
- Last Activity as Converted to Lead
- More pages visited on website

Model Evaluation - ROC Curve

GET PAID MORE:

(Model showing Good Signs)

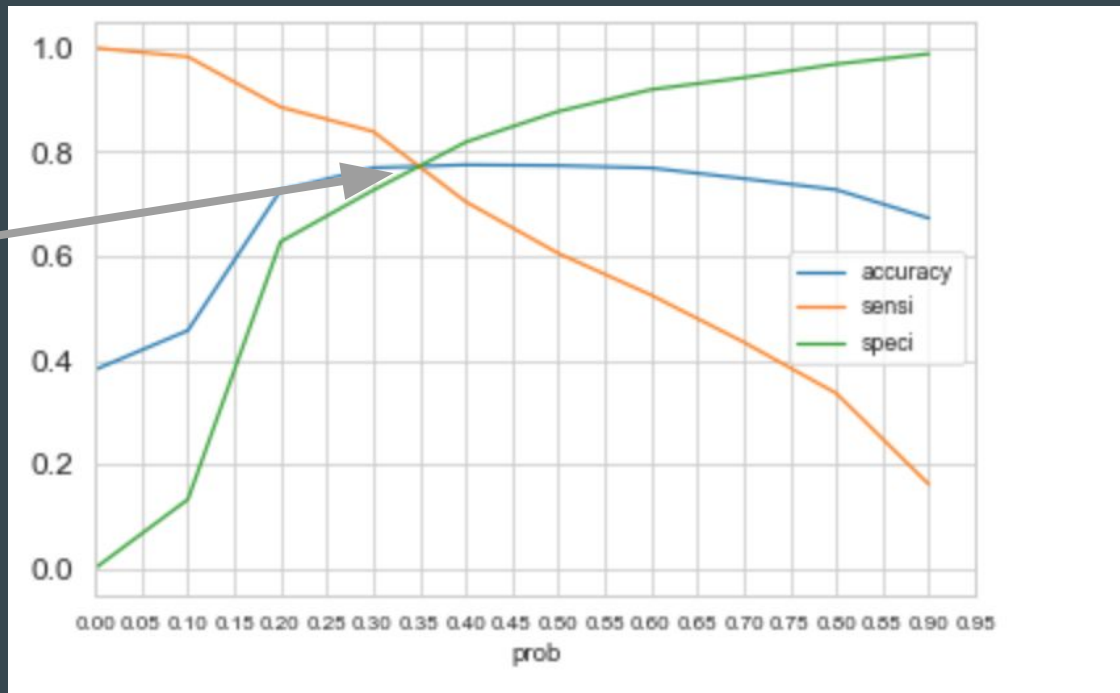
Area under the curve is 84%
of the total area.



Model Evaluation : Accuracy / Sensitivity / Specificity

Optimal cutoff probability is the probability where we get balanced sensitivity and specificity.

Optimal Cutoff Value: **0.35**



Findings

Confusion Matrix

3012	923
477	1951

Confusion Matrix

1260	400
209	858

Train Data

- Accuracy : 78 %
- Sensitivity : 80.4 %
- Specificity : 76.6 %

Test Data

- Accuracy : 78 %
- Sensitivity : 81 %
- Specificity : 76 %

Conclusion

- Overall Accuracy of Prediction is ~78%.
 - Sensitivity, i.e., how many leads are actually converted if we predict X leads. Our model has a sensitivity value of ~ 80%
 - The model has a specificity of ~76%.
-

Presented By:



Sagir Mehmood

Student,
Statistics, Honours
Calcutta University



Vaibhav Sah, Senior Engineer

Aspiring Data Scientist,
Front End Developer,
Ticketmaster