Lead Scoring Case Study

SREVATSSA M

Problem statement

- X Education gets a lot of leads, but its lead conversion rate is very poor
- To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’
- If they successfully identify this set of leads, the lead conversion rate should go up .
- We are required to build a 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
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.

Solution

Leads Clustering

- We cluster the leads into certain categories based on their tendency or probability to convert, thus, getting a smaller section of hot leads to focus more on.

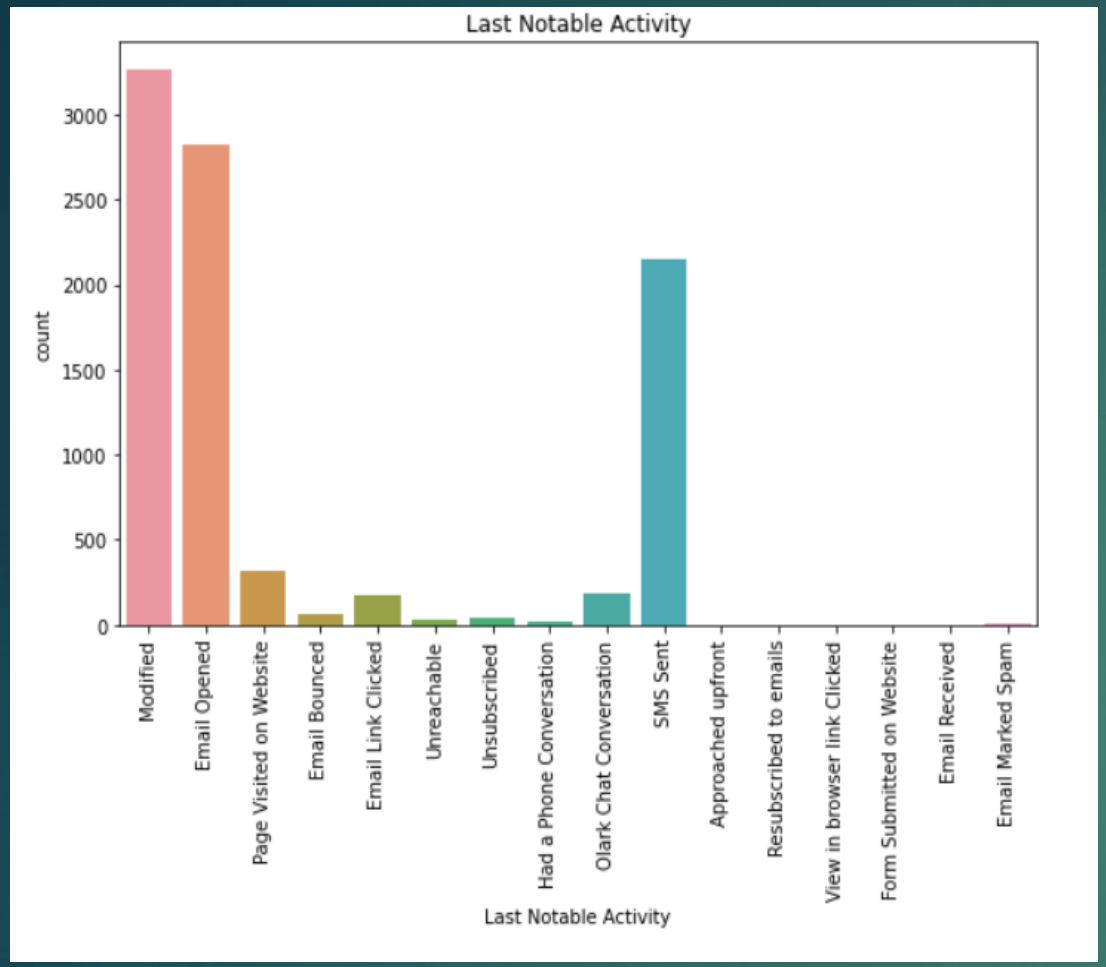
Focus Communication

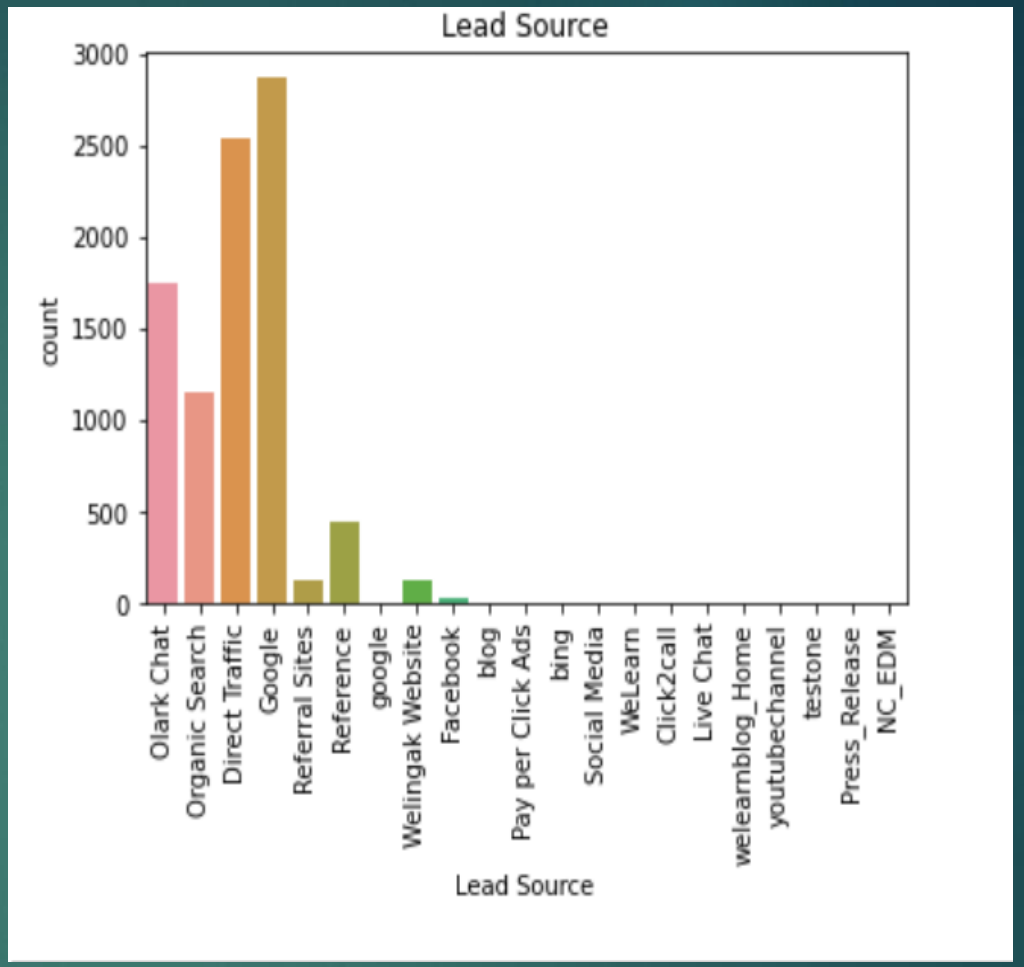
- Since we would have a smaller set of leads to have communication with, we might make more impact with effective communication .

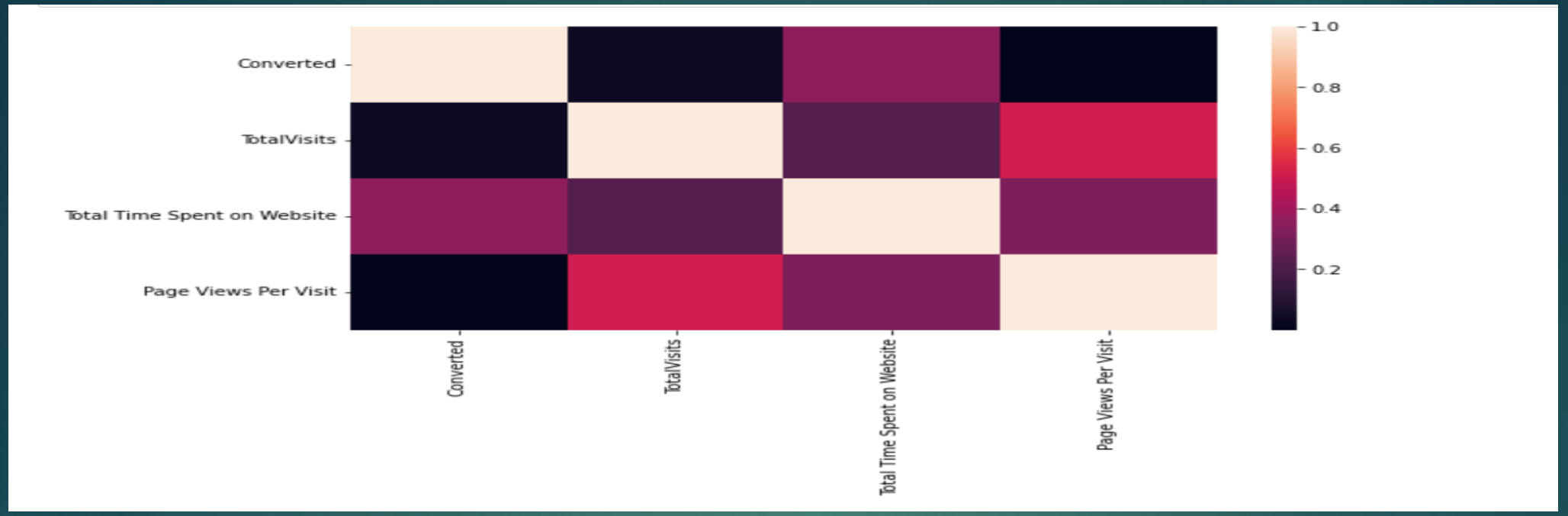
Increase conversion

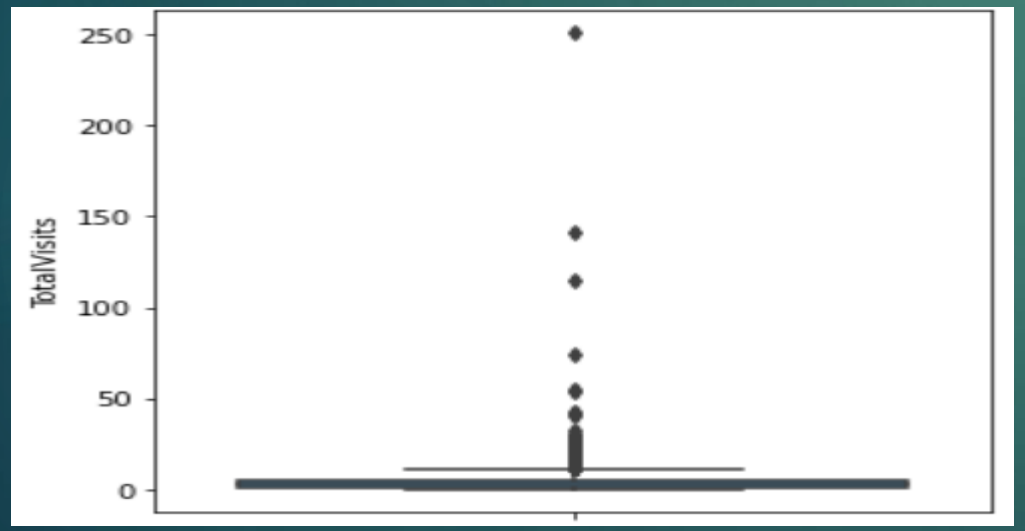
- Since we focused on hot leads, which were more probable to convert, we would have a better conversion rate, and hence we can achieve the 80% target.

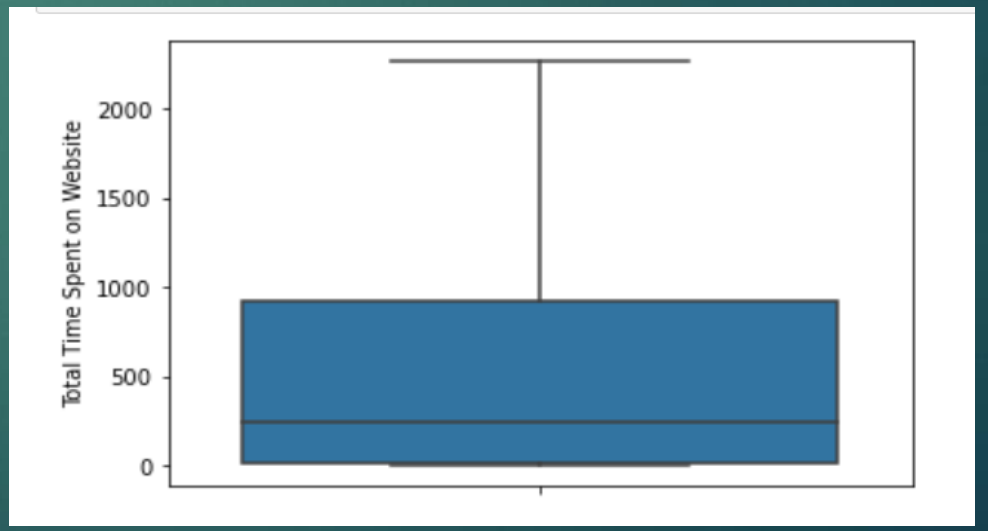
Visualisations

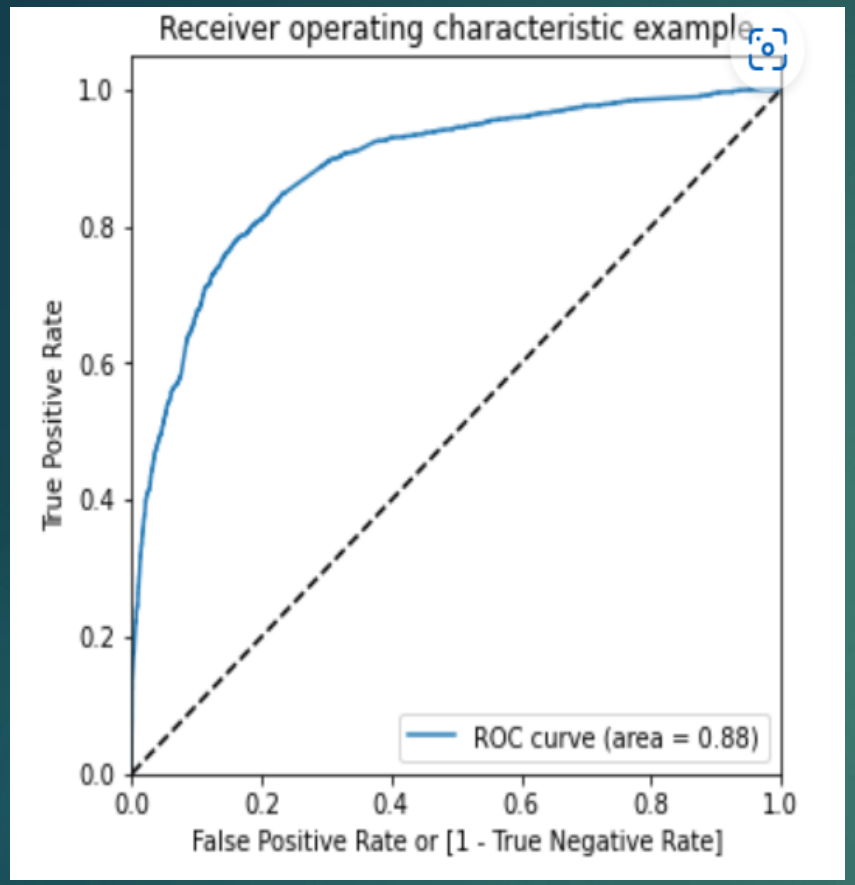

Last Notable Activity
count
0
500
1000
1500
2000
2500
3000
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
View in browser link Clicked
Form Submitted on Website
Email Received
Email Marked Spam
Last Notable Activity

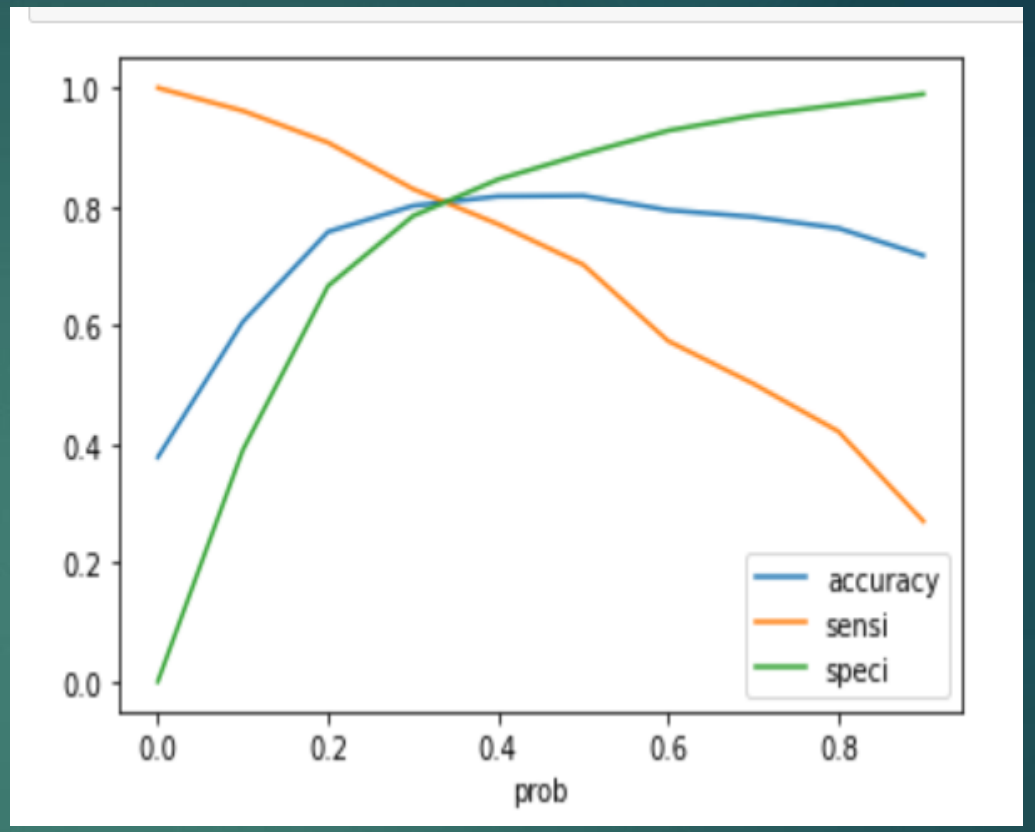

Lead Source
count
0
500
1000
1500
2000
2500
3000
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Reference
google
Welingak Website
Facebook
blog
Pay per Click Ads
bing
Social Media
WeLearn
Click2call
Live Chat
welearnblog_Home
youtubechannel
testone
Press_Release
NC_EDM
Lead Source


Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
1.0
0.8
0.6
0.4
0.2
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit


TotalVisits
250
200
150
100
50
0


Total Time Spent on Website
2000
1500
1000
500
0


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
ROC curve (area = 0.88)
0.0
0.2
0.4
0.6
0.8
1.0


accuracy
sensi
speci
prob
0.0
0.2
0.4
0.6
0.8
1.0

Analysis

- Overall accuracy on Test set: 0.786
- Sensitivity of our logistic regression model: 0.733
- Specificity of our logistic regression model: 0.823

Inferences

Top 3 variables in model, that contribute towards lead conversion are:

- Total Time Spent on Website
- Last Notable Activity_SMS Sent
- TotalVisits

Top 3 variables in my model, that should be focused are:

- Last Activity_SMS Sent (positively impacting)
- Last Activity_Olark Chat Conversation (negatively impacting)
- Lead Source_Olark Chat (negatively impacting)

Conclusion

X Education Company needs to focus on following key aspects to improve the overall conversion rate:

- Increase user engagement on their website since this helps in higher conversion .
- Increase on sending SMS notifications since this helps in higher conversion .
- Get Total visits increased by advertising etc. since this helps in higher conversion .
- Improve the Olark Chat service since this is affecting the conversion negatively .

Thank you