Lead Scoring Case Study

Submitted BY:

- Meraj Alam
- Ganesh V
- Manish Kumar

Problem Statement

- X Education gets a lot of leads, its lead conversion rate is very poor at around 30%
- X Education wants to make lead conversion process more efficient by identifying the most potential leads, also known as Hot Leads
- Their sales team want to know these potential set of leads, which they will be focusing more on communicating rather than making calls to everyone.

Objective of the Study

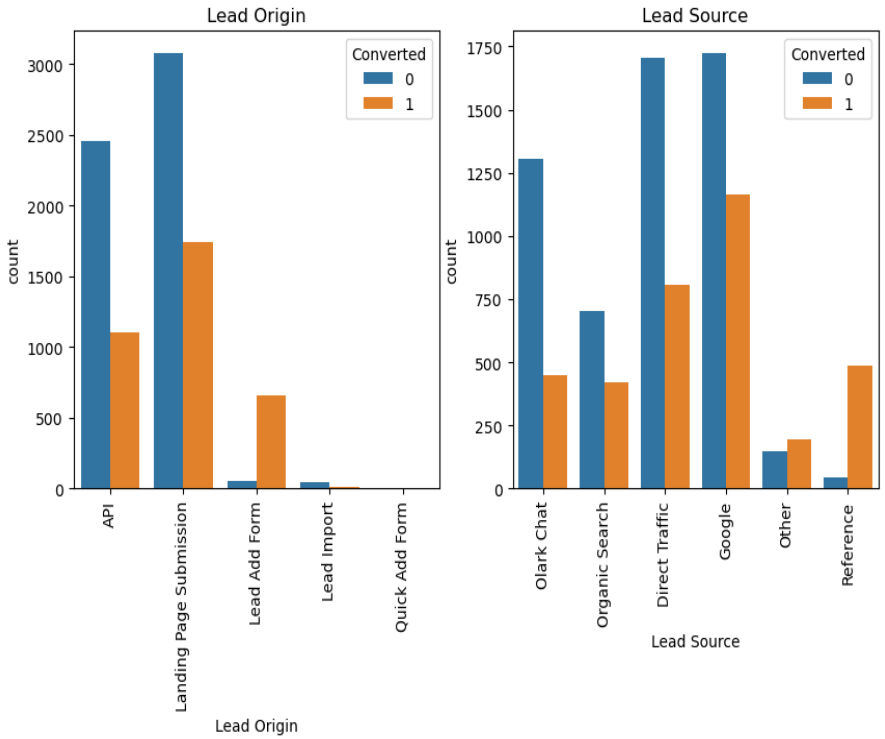
- To help X Education select the most promising leads, i.e., the leads that are most likely to convert into paying customers.
- The company requires us to build a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%.

Model Features

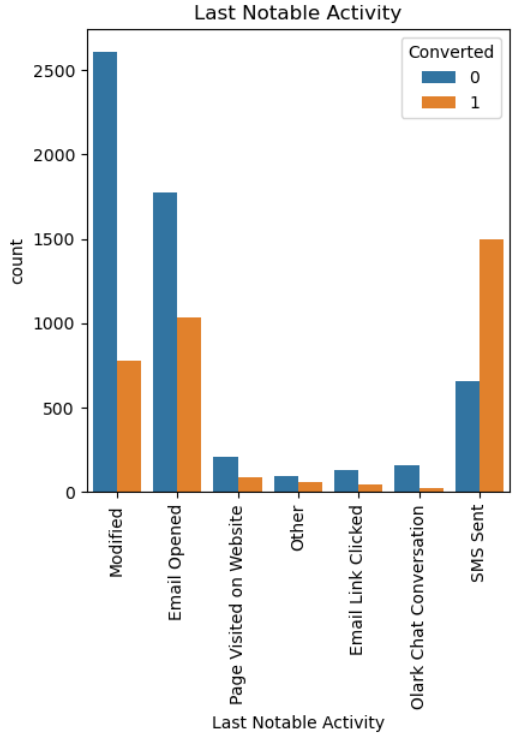
- Lead Origin_Lead Add Form
- Total Time Spent on Website
- What is your current occupation_Working Professional
- Last Notable Activity_Other
- Last Notable Activity_SMS Sent
- Lead Source_Olark Chat
- TotalVisits
- Specialization_Missing
- Lead Origin_Landing Page Submission
- What is your current occupation_Missing
- Last Activity_Olark Chat Conversation
- Lead Source_Reference
- Do Not Email

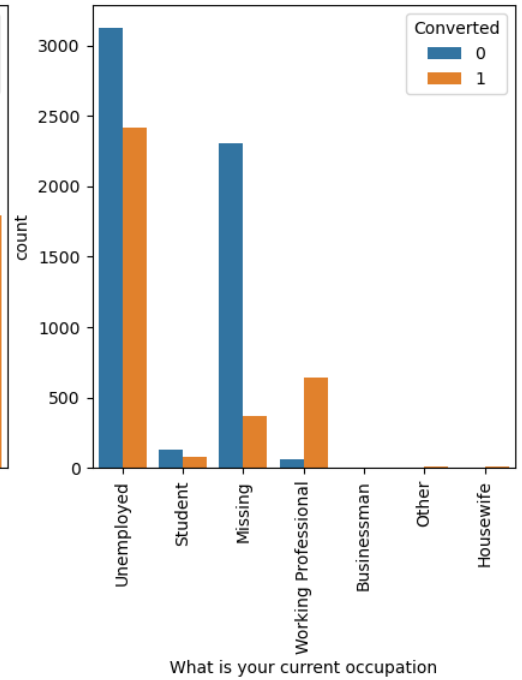
EDA

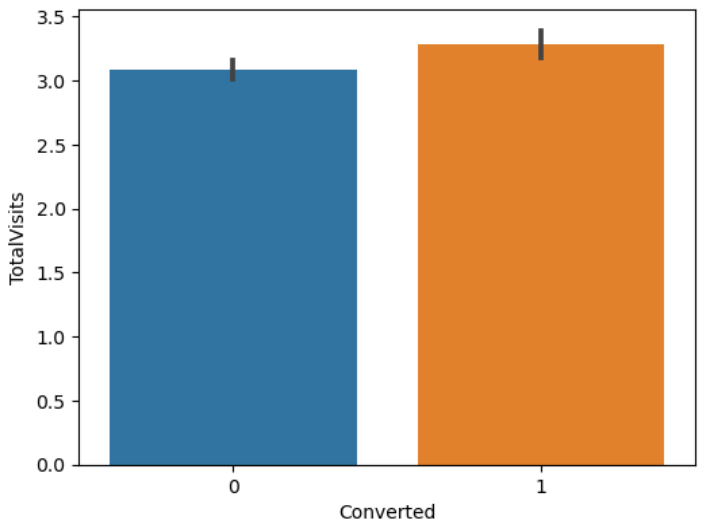
- Data imbalance checked- only 38.5% leads converted
- We have created the univariate, bivariate analysis on both categorical and numerical columns, also we have checked the plot numerical vs converted columns and categorical vs converted columns where we have understood which columns is important for further analysis even, we have 29 % of missing value.
- Maximum lead conversion happened from Landing Page Submission.
- Max. lead conversion in the lead source is from 'Google’
- Major lead conversion has happened from the emails that have been sent.
- Major lead conversion has happened from the calls they are doing.
- Maximum conversion to those customers who has opened their mails and whom SMS being sent, and those customers who has modified.
- Major lead conversion is from the Unemployed Group
- Max. lead conversion of those customer who has spent max time on the website.

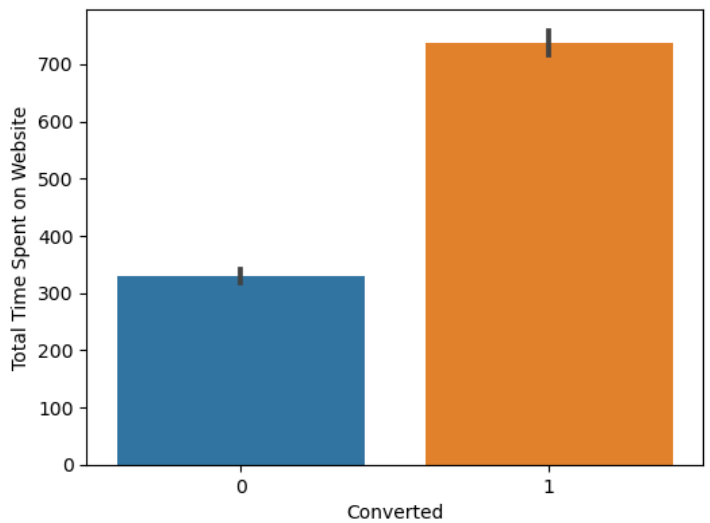

Lead Origin
Converted
0
1
count
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form
Lead Origin

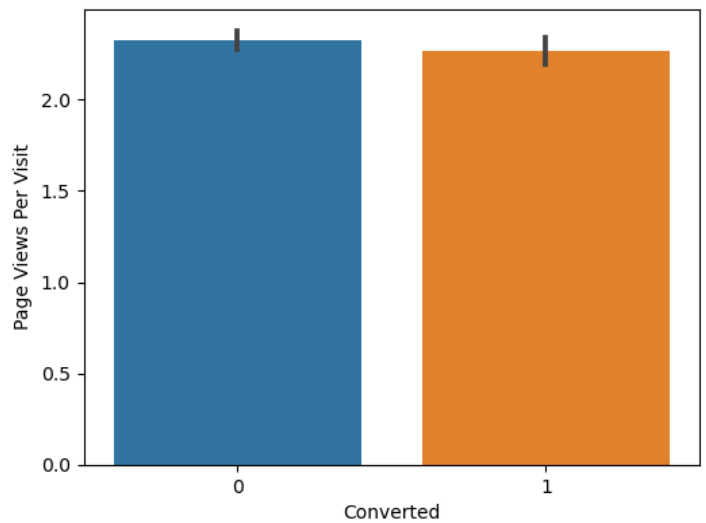
Lead Source
Converted
0
1
count
Olark Chat
Organic Search
Direct Traffic
Google
Other
Reference
Lead Source


Last Notable Activity
Converted
0
1
count
Modified
Email Opened
Page Visited on Website
Other
Email Link Clicked
Olark Chat Conversation
SMS Sent
Last Notable Activity


What is your current occupation
Converted
0
1
count
Unemployed
Student
Missing
Working Professional
Businessman
Other
Housewife
What is your current occupation

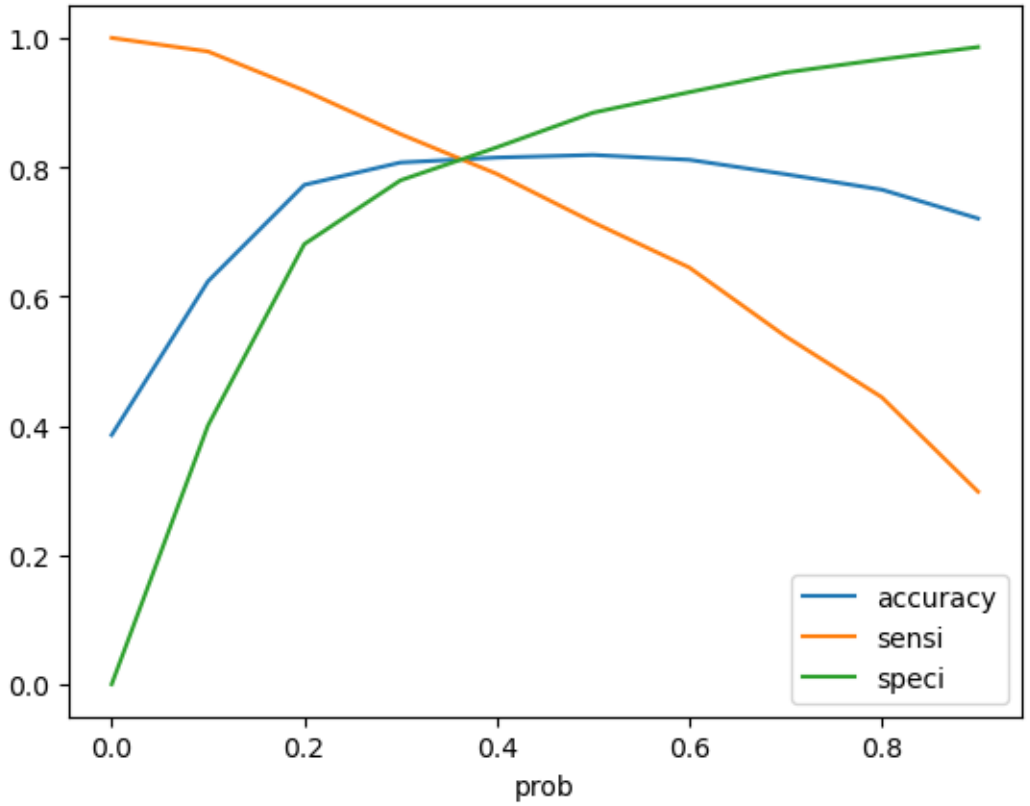

TotalVisits
Converted
0
1


Total Time Spent on Website
Converted
0
1

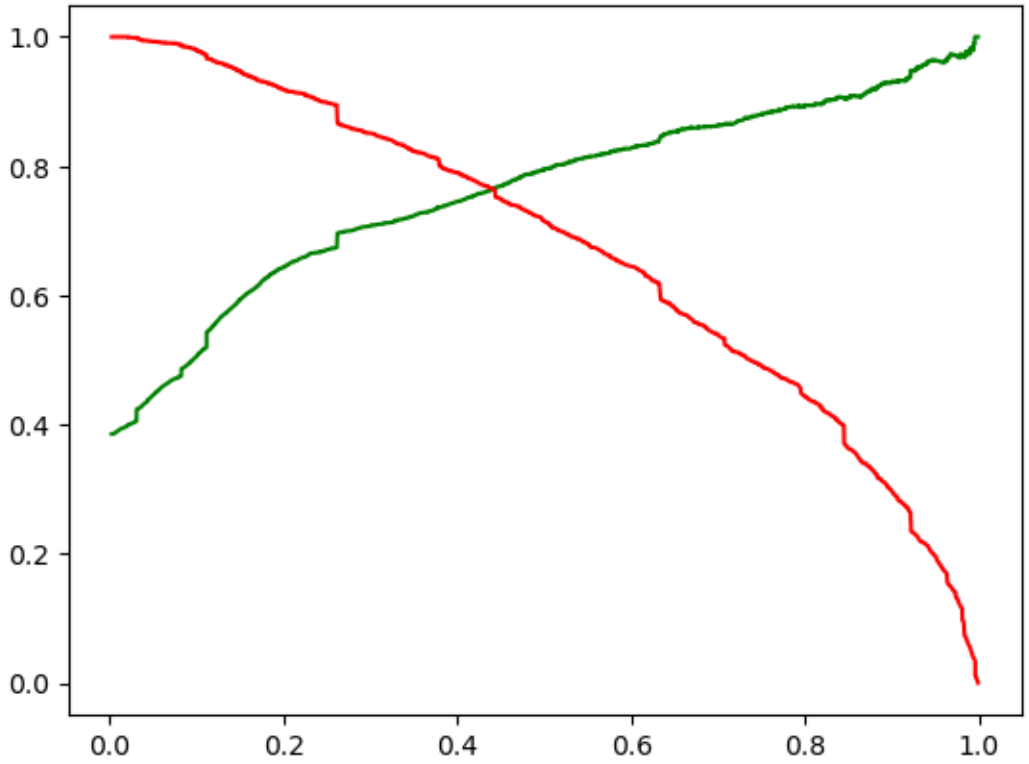

Page Views Per Visit
Converted
0
1

Model Evaluation

- The optimum cut off value was found using ROC curve.
- The area under ROC curve was 0.90.
- Training Data
- After Plotting we found that optimum cutoff was 0.35 and got
  - Accuracy 81.05%
  - Sensitivity 82.31%
  - Specificity 80.26%.



- On Training Data
- With the cutoff of 0.35 we get the Precision & Recall of 79.47% & 71.46% respectively.
- So to increase the above percentage we need to change the cut off value.
- After plotting we found the optimum cut off value of 0.42 which gave
  - Accuracy 81.66%
  - Precision 75.47%
  - Recall 77.77%
- Prediction on Test Data
  - Accuracy 81.51
  - % Precision 75.0%
  - Recall 77.7%



Conclusion

- The Sensitivity and Specificity, Accuracy, Precision and Recall score we got from test set are almost accurate.
- Important features responsible for good conversion rate or the ones which contributes more towards the probability of a lead getting converted in decreasing order:
  - Lead Origin_Lead Add Form
  - Total Time Spent on Website
  - What is your current occupation_Working Professional
  - Last Notable Activity_Other
  - Last Notable Activity_SMS Sent
  - Lead Source_Olark Chat
  - TotalVisits