Lead scoring case study report

Problem statement

X education has a low lead conversion issue which is about 30%. The team at X education wants to focus only on potential high probability leads that they should go after to ensure they do not waste resources on other leads.

Analysis thought process

- Data was cleaned initially by managing null values.
- EDA was done on the data to understand the relationship between the variables.
- And eventually RFE was done to select top 15 variables.

Model details and model assessment

The optimum cut off of 0.4 probability was obtained based on the intersection point of precision and recall curve. Accuracy and precision of the model was assessed to be around ~80%.
Key features that positively impact leads are:
- Time spent on website
- Lead origin
- Lead source

Based on logistic regression the accuracy of the model is 80% and hence the team at X education and use this logic to call potential leads.

Learnings

- There is mutual tension between precision and recall and we can only maximize one of the two always. Hence it is in the best interest to discuss with the business to get more business context on choosing what is more important – precision or recall (among the 2)
- Based on the analysis there can be value addition to any business if we deploy the correct model and approach.