Lead case study

Mahesh Jinkala Jyothi sondhi

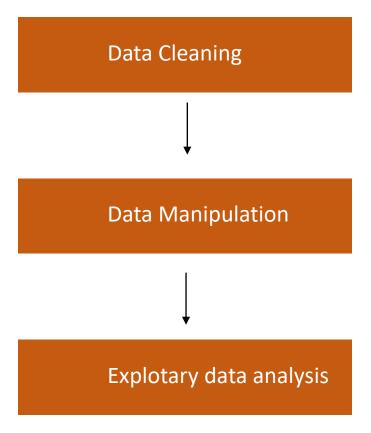
Problem Statement

- > X Educationsells online courses to industry professionals. X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also knownas 'Hot Leads'.
- ➤ If they successfully identify this set of leads, the lead conversion rate should go upas the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

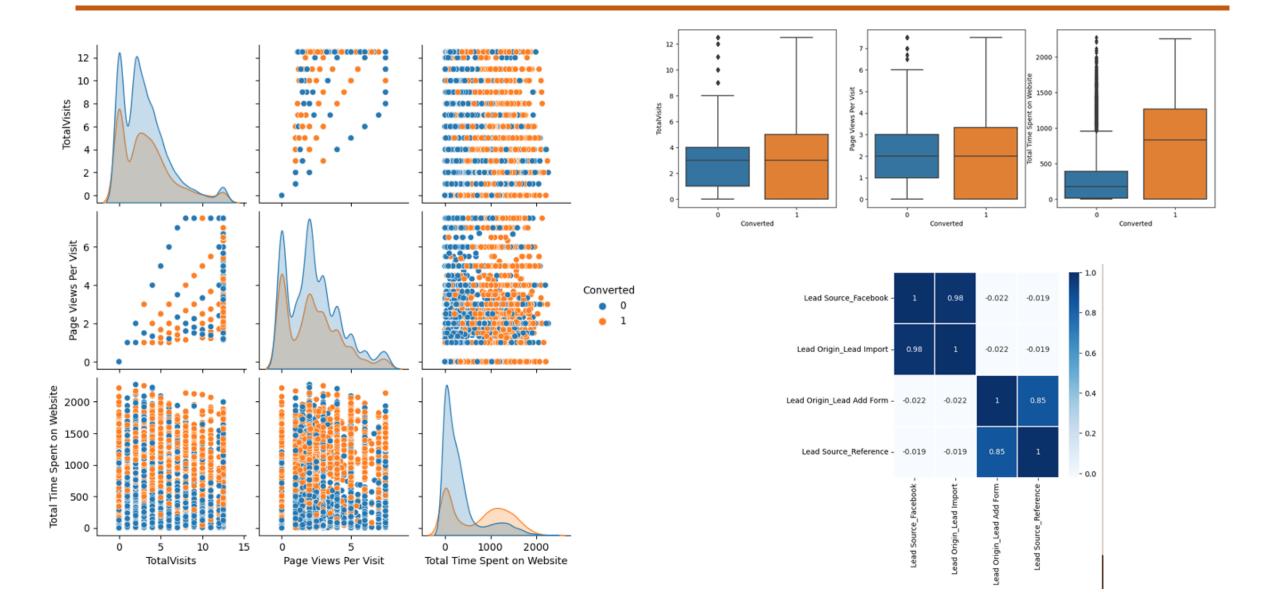
Objectives

- > X education wants to know most promising leads.
- > For that they want to build a Model which identifies the hot leads.
- > Deployment of the model for the future use.

Methodology



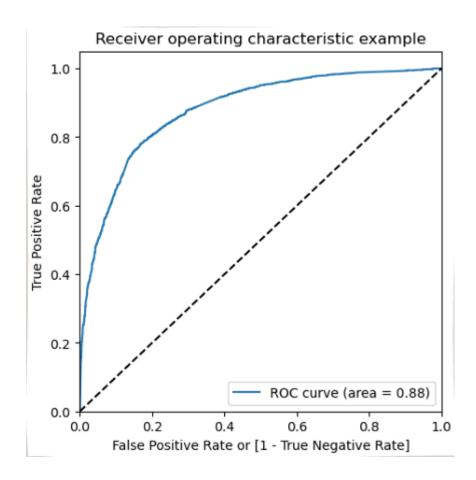
EDA insights

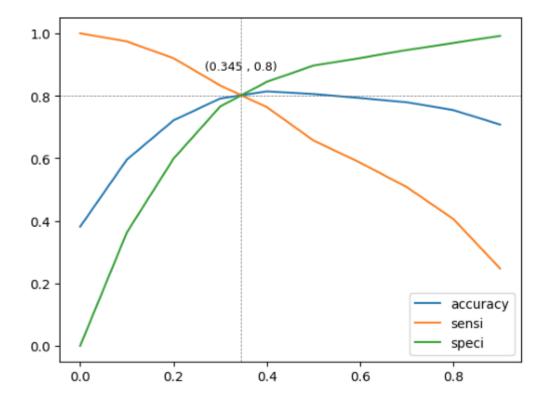


Model building

- ► Splitting the Data into Training and Testing Sets
- ▶ The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- ► Use RFE for Feature Selection
- ► Running RFE with 15 variables as output
- ▶ Building Model by removing the variable whose p-value is greater than 0.05 and vi value is greater than 5
- ▶ Predictions on test data set
- ► Overall accuracy 81%

ROC curves





Conclusion

It was found that the variables that mattered the most in the potential buyers are (In descending order):

- ▶ The total time spent on the Website.
- ► Total number of visits.
- ▶ When the lead source was:

Google Direct traffic Organic search Welingak website

- ▶ When the last activity was: SMS Olark chat conversation
- ▶ When the lead origin is Lead add format.
- ▶ When their current occupation is as a working professional. Keeping these in mind X Education can flourish as

they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.