Lead Scoring Case Study

By: Manasa BR

Soumyadip Chakraborty

Satya Pratyusha

Purpose

To Identify the reasons and aspects which can be focused on to Increase the Lead Conversion Rate

Problem Statement

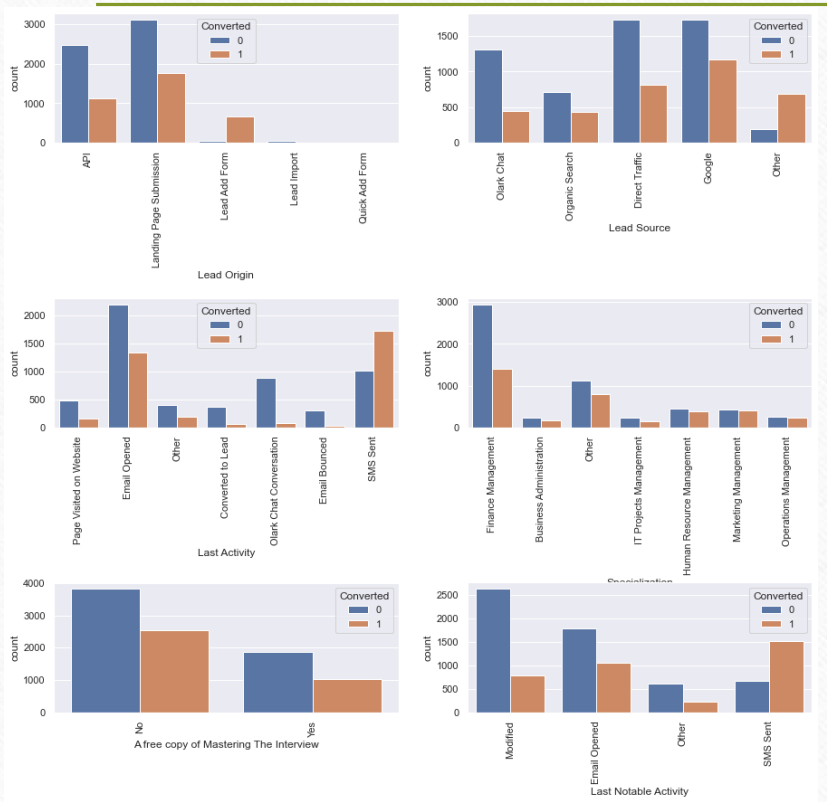
An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Approach

- Import the data for Analysis
- Read & Understand the data
- EDA
- Feature Scaling
- Split the data into train and test dataset
- Prepare the data for modelling
- Model building
- Model Evaluation
- Predictions on the Test Dataset

EDA

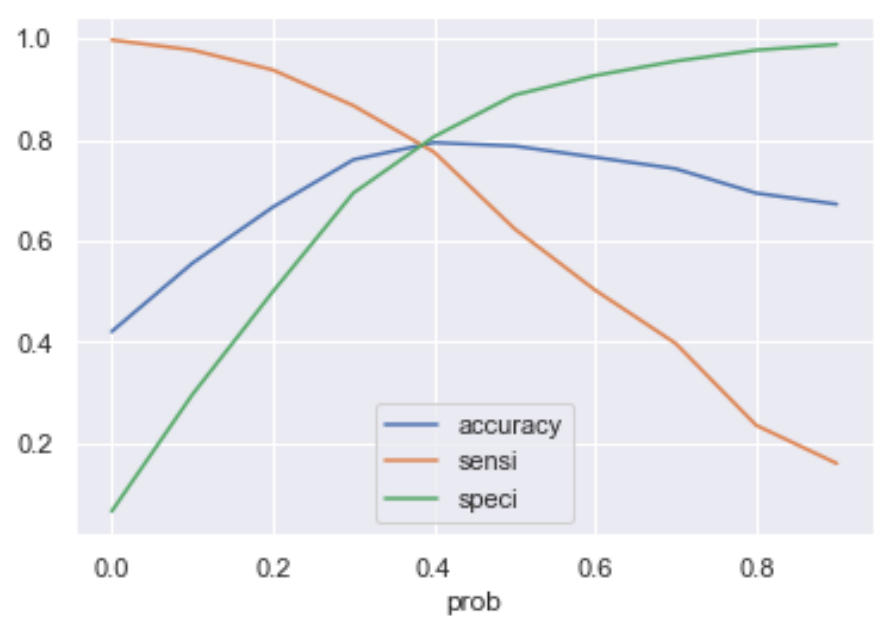


- In Lead Source The number of Hot leads is higher in Direct Traffic and Google less in Other Category
- In Last Activity the number of Hot leads is higher in SMS and in EMAIL cold leads is higher than hot leads.
- In Last Notable Activity it's mostly same as Last Activity.
- In Specialization the most of the leads are comes from Finance management but here Hot leads are lesseer than Cold leads.

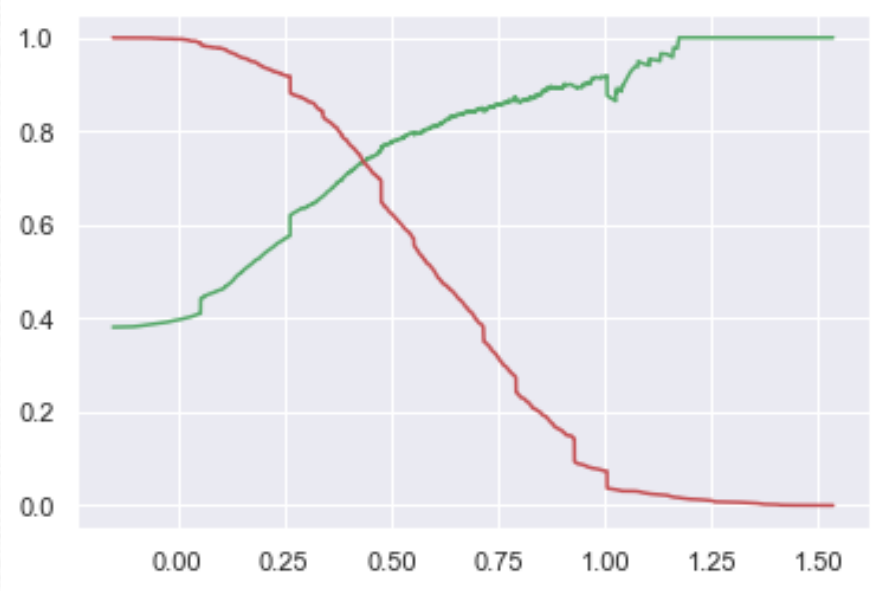
Model Building

- Splitting into Train and Test Datasets
- Scaling the features
- Find the correlation and RFE to eliminate variables
- Building Logistic Regression using Stats model
- Prediction on TEST Set

Model Evaluation



Taking value 3.7 as the Final Cutoff



we got 0.37 as the Cut-off as Precesion-Recall Threshold

Conclusion

- Total time spent on website and Total number of visits are important factors
- The Accuracy(78,77),Sensitivity(80,79) and Specifity(77,75) are approximately equal between the train and test dataset.