Lead Score Case Study

Using Logistic Regression

Chaitanya sarath Alajangi

Problem statement

X Education sells online courses to industry professionals. 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%.

Goal

X Education needs help in selecting the most promising leads, i.e., the leads that are most likely to convert into paying customers.

The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%

Strategy

- Clean the data
- EDA (Exploratory Data Analysis)
- Feature scaling
- Dummy variable creation
- Data split to train and test
- Building the model
- Evaluate the model
- Finding out the cut off and retraining the model

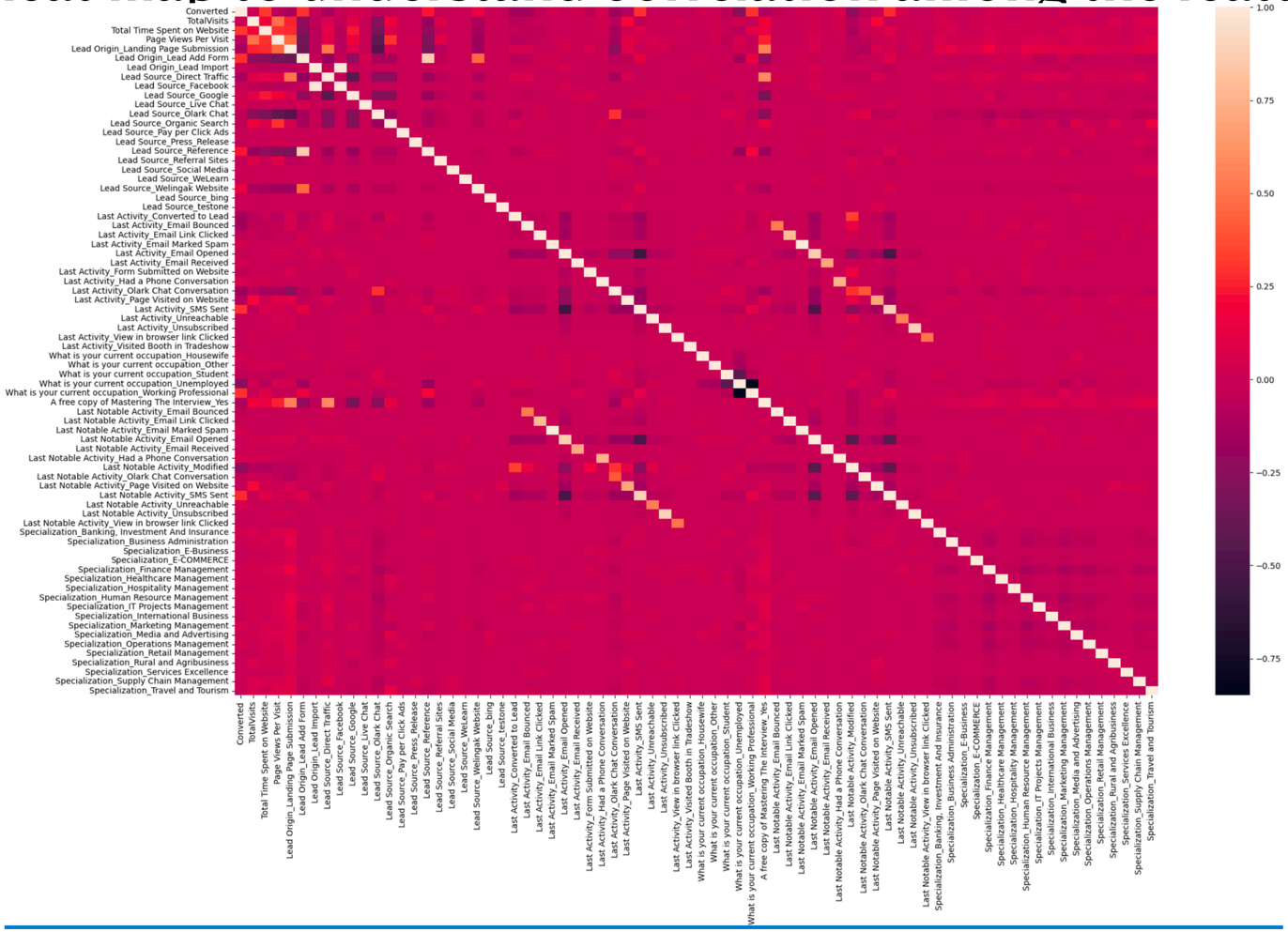
Data cleaning

We cleaned the columns below

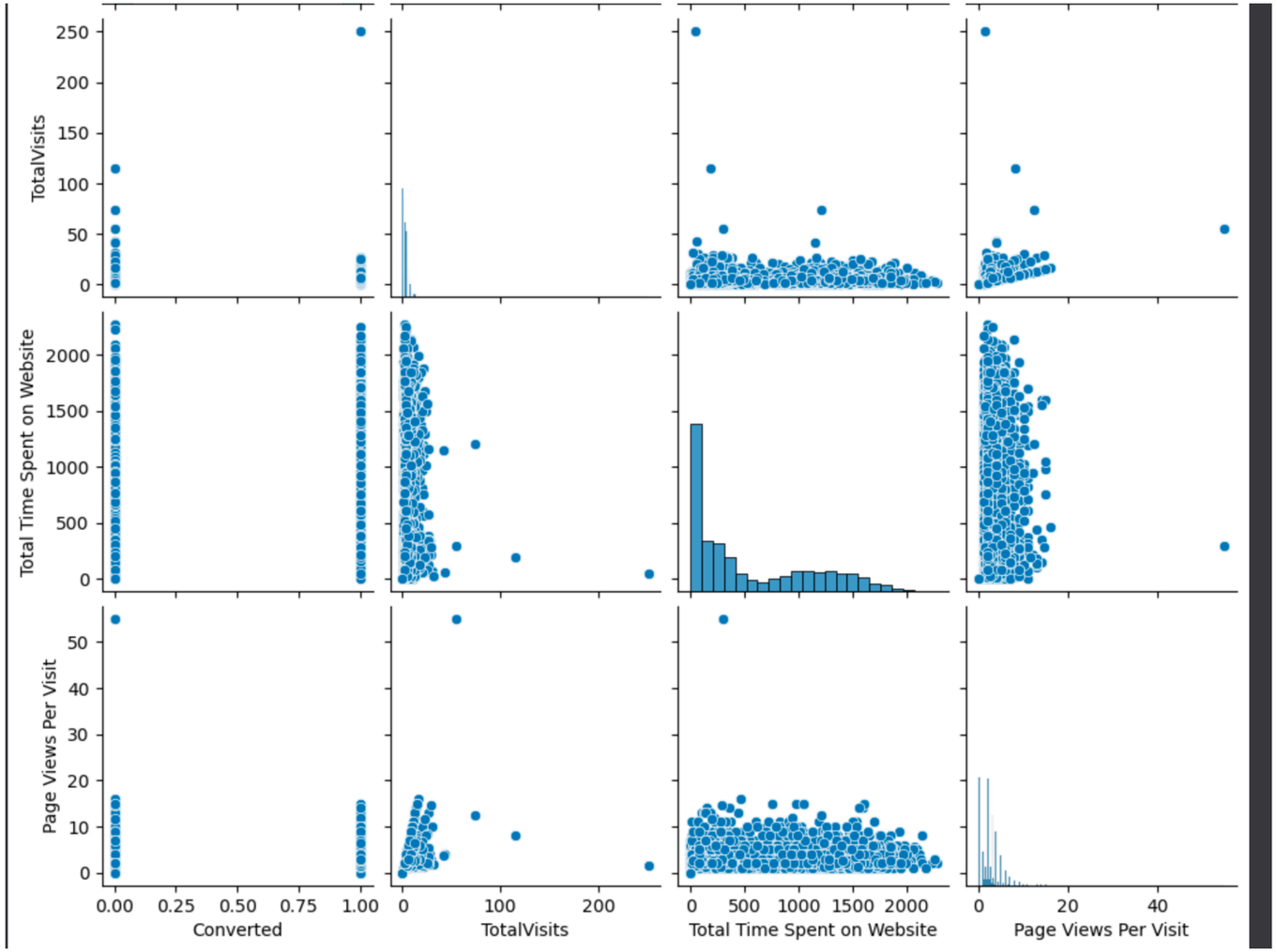
- Dropped columns with null values > 30% of data
- Dropped columns without diverse data
- Dropped columns with high VIF (correlated features)
- Dropped columns with high p-value (less significant features)

EDA

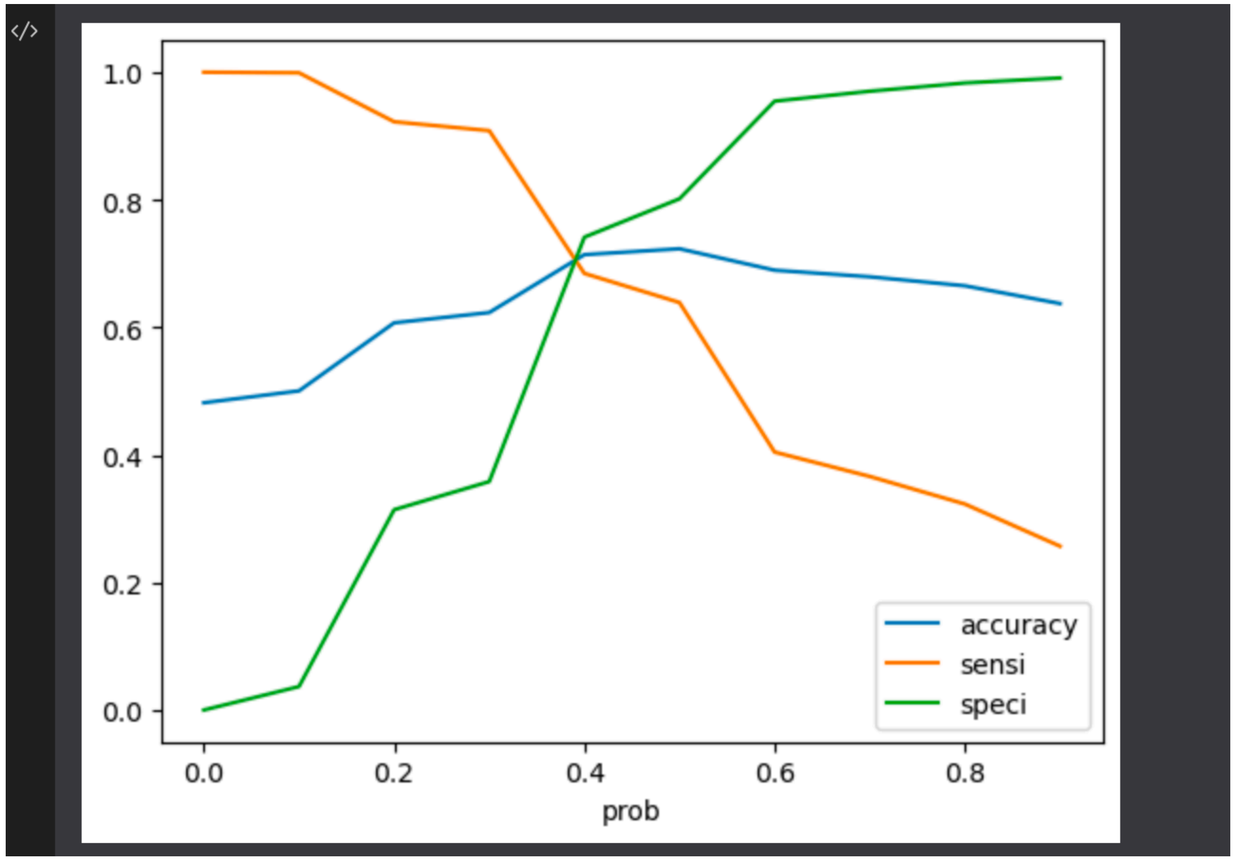
- Used heat map to understand correlation among the features
-



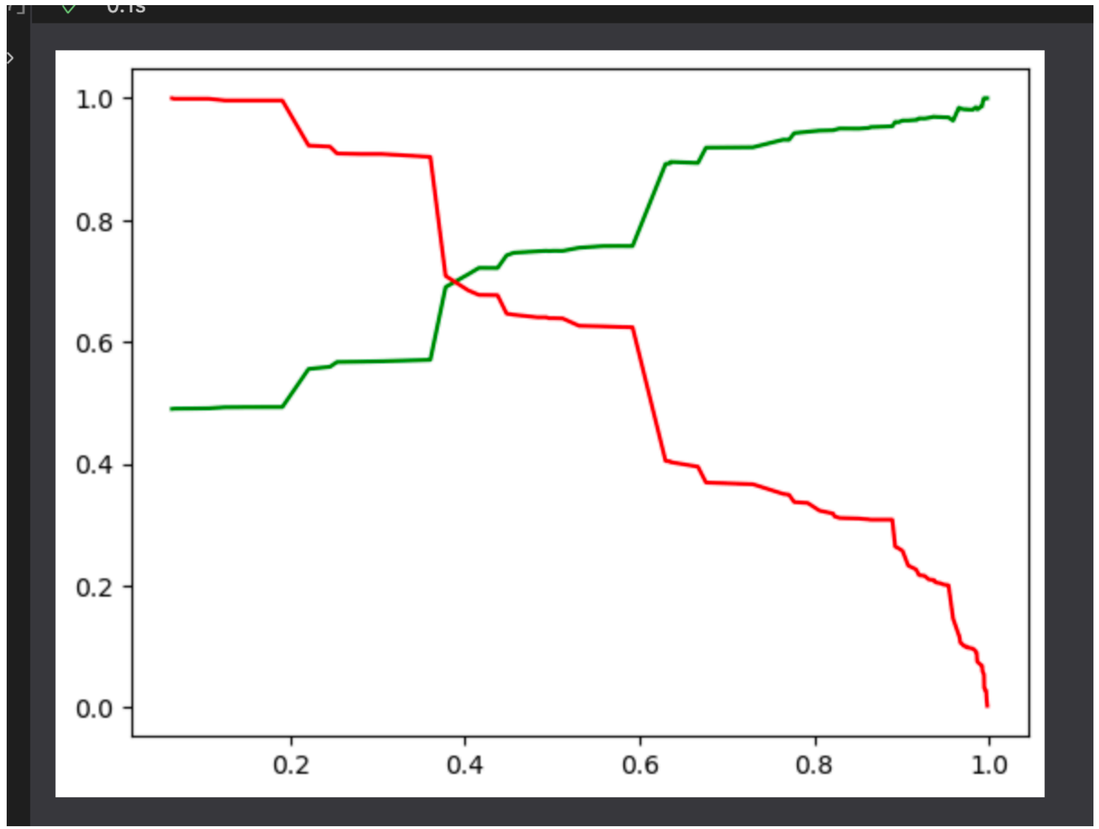
- Used pair plot to understand data relation among numeric columns



Model Building & finding optimal cut off

- Build the model with various cut offs and extract the best cut off
- 

Precision recall

- We used the precision recall to find the ideal cut off
- 

Summary

In conclusion we can ssee that we can make use of features like

- 'Lead Origin_Lead Add Form' ,
- 'Last Activity_Had a Phone Conversation ',
- 'Specialization_Banking, Investment And Insurance,Marketing Management ,Rural and Agribusiness'

can be used as hot leads , which can be quickly converted .

These leads need to be curated well to acheive better conversion , on top we have some features which would reduce our conversion , we need to pay closer attention to these leads . We can add new features to the Specializations which can boost our lead conversion