Lead Score Case Study
Using Logistic Regression

By

Kumkum Biswal

What we want to Achieve:

- 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 where a lead score is assigned to each of the leads such that the customers with higher lead score has a higher conversion chance and the customers with lower lead score has a lower conversion chance.
- The CEO has given a ballpark of the target lead conversion rate to be around 80% .

Process Steps

- 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 to find out best cut off

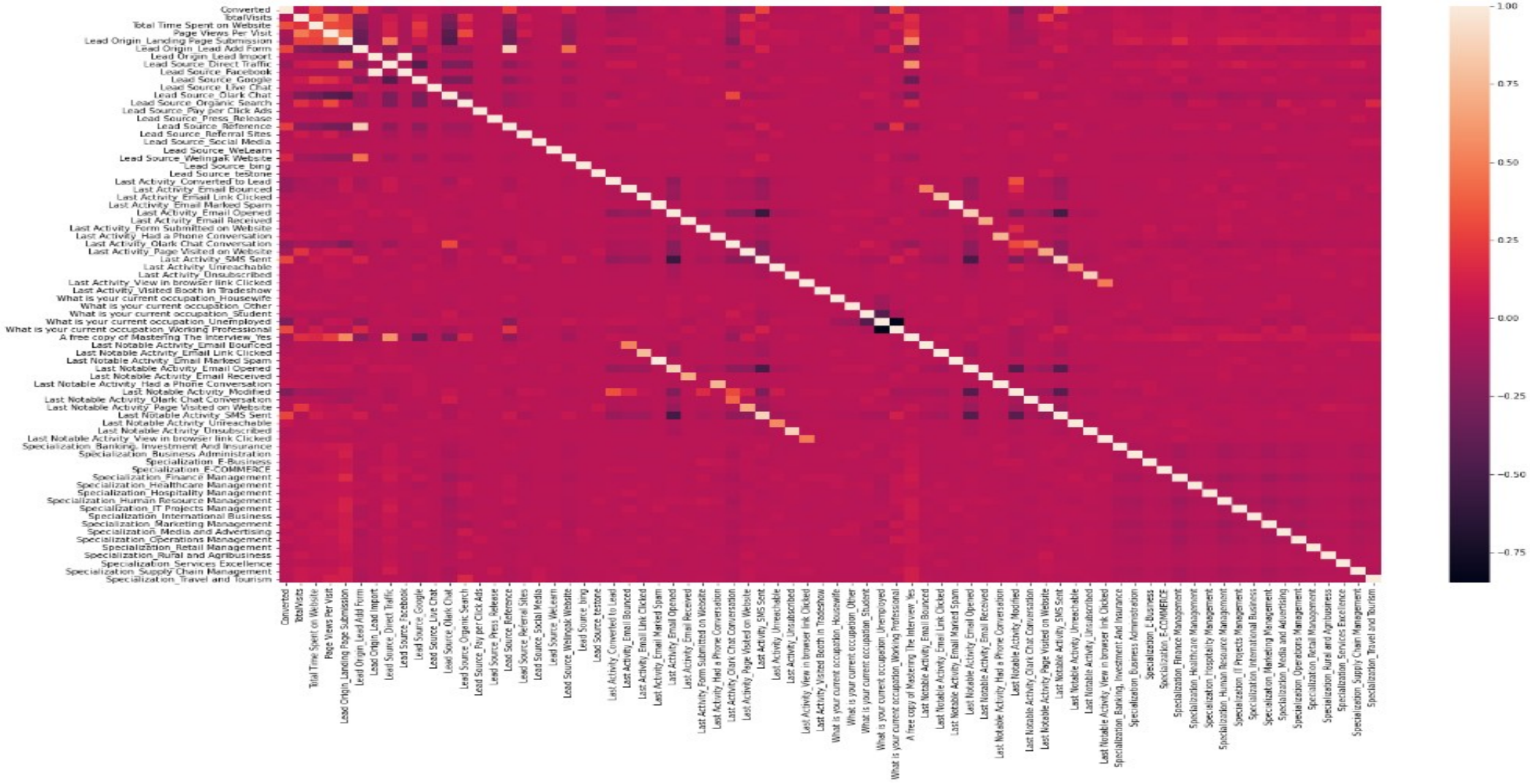
Data cleaning

Below Columns are being dropped based on its significance:

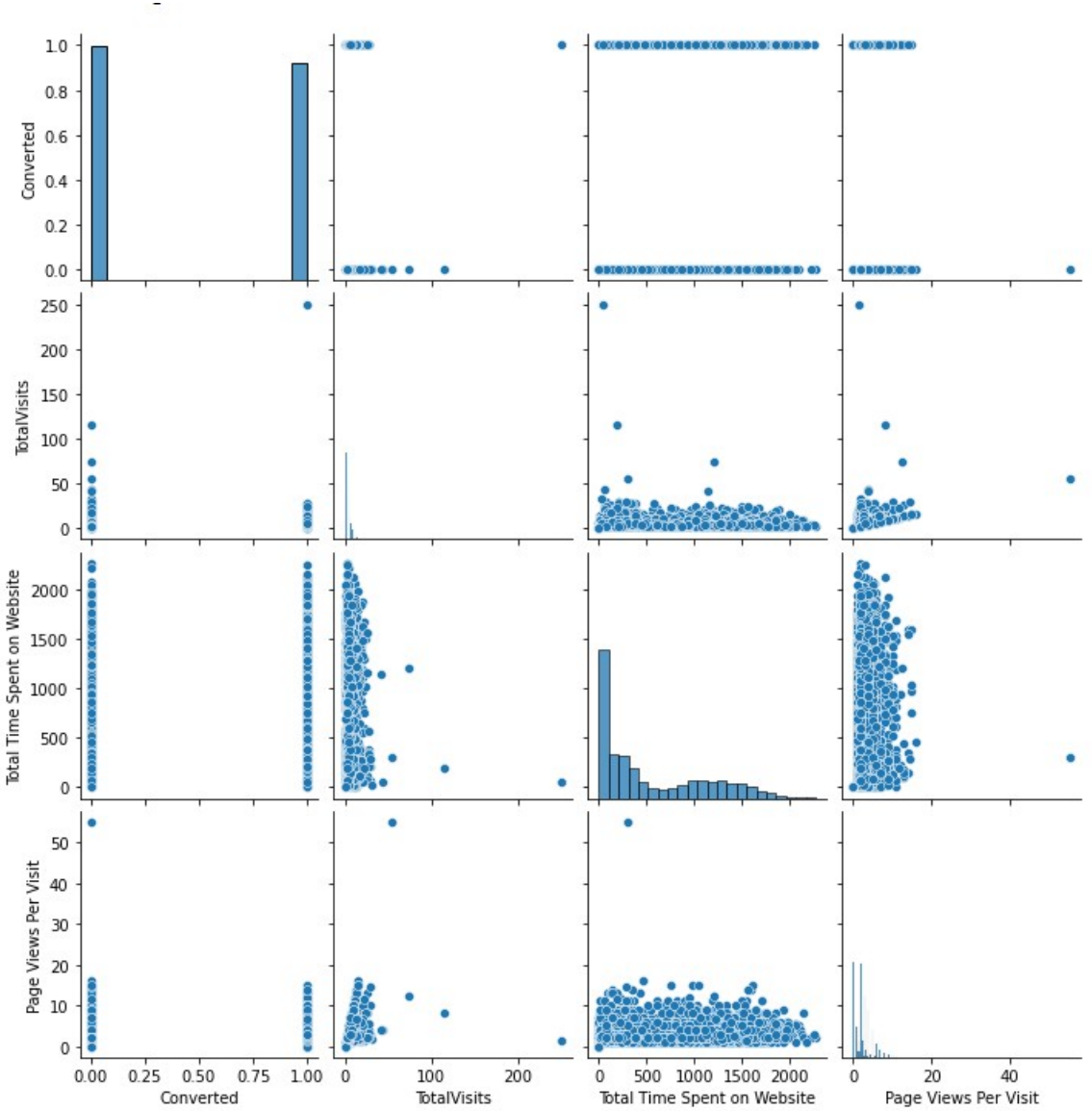
- Dropped columns with null values > 30% of data
- Dropped columns without diverse data
- Dropped columns with high VIF (correlated features)
- Dropped columns with high p-value (less significant features)

EDA

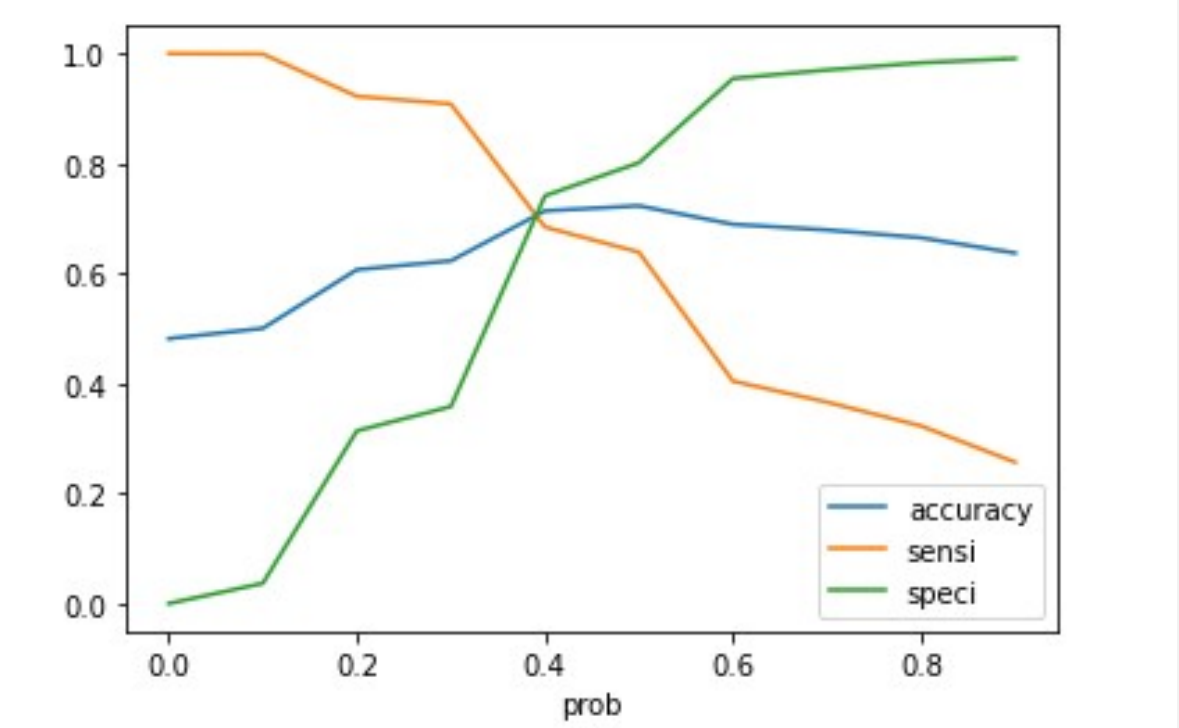
Used heat map to understand correlation among the features



Used pair plot to understand data relation among numeric columns

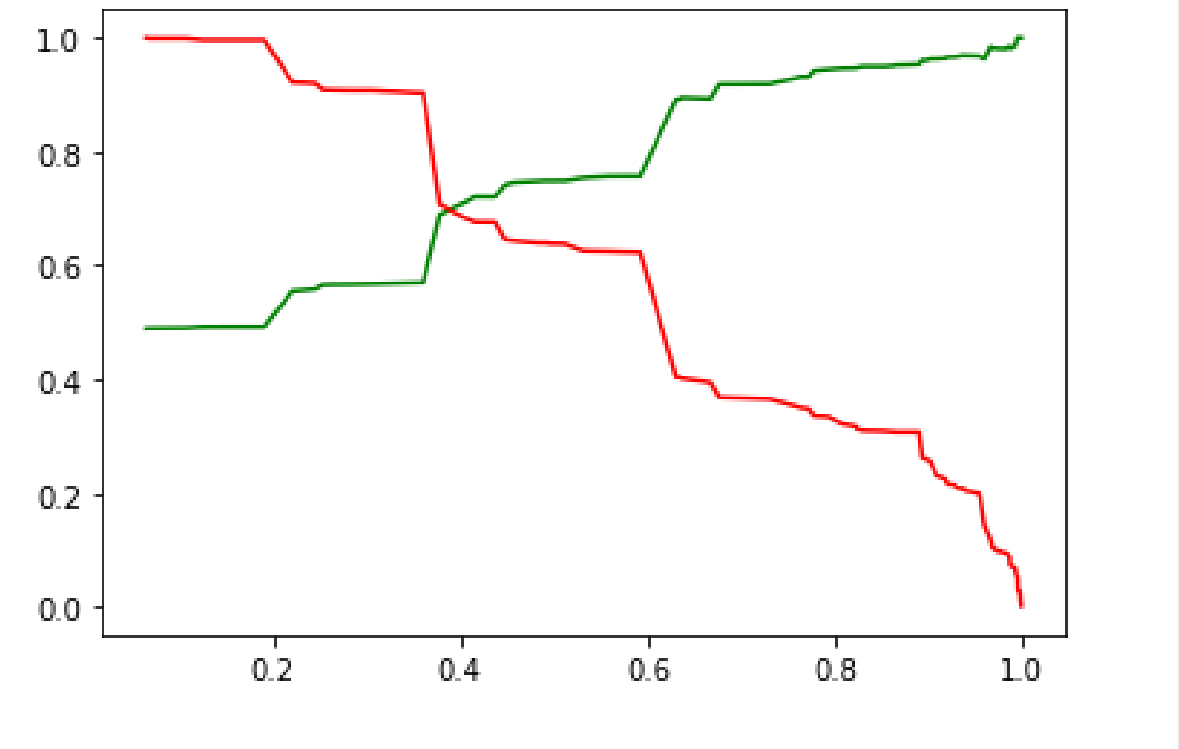


Model Building & finding optimal cut off



- 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

we use of features like below:

- 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 achieve 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