



LEAD SCORE CASE STUDY

BUSINESS UNDERSTANDING

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.

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%.



X Education gets a lot of leads, its lead conversion rate is very poor at around 30%



X Education wants to make lead conversion process more efficient by identifying the most potential leads, also known as Hot Leads



Their sales team want to know these potential set of leads, which they will be focusing more on communicating rather than making calls to everyone.

PROBLEM STATEMENT



To help X Education select the most promising leads, i.e., the leads that are most likely to convert into paying customers.



The company requires us to build a model wherein we need to assign a lead score to each of the leads



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

BUSINESS OBJECTIVE

STEPS FOR MODEL ANALYSIS



DATA CLEANING



EXPLORATORY
DATA ANALYSIS



DATA
PREPARATION
AND MODELLING



MODEL BUILDING



MODEL
EVALUATION

DATA CLEANING



Columns with more than 3000 values as null were dropped.



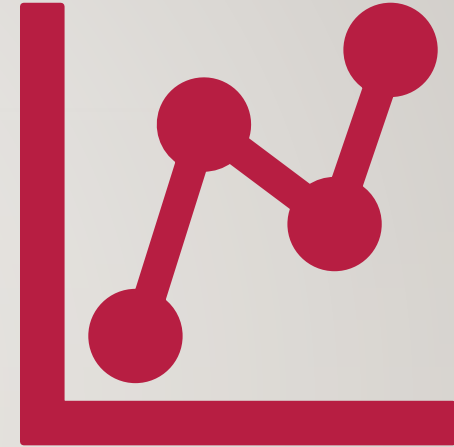
Dropping not so important columns

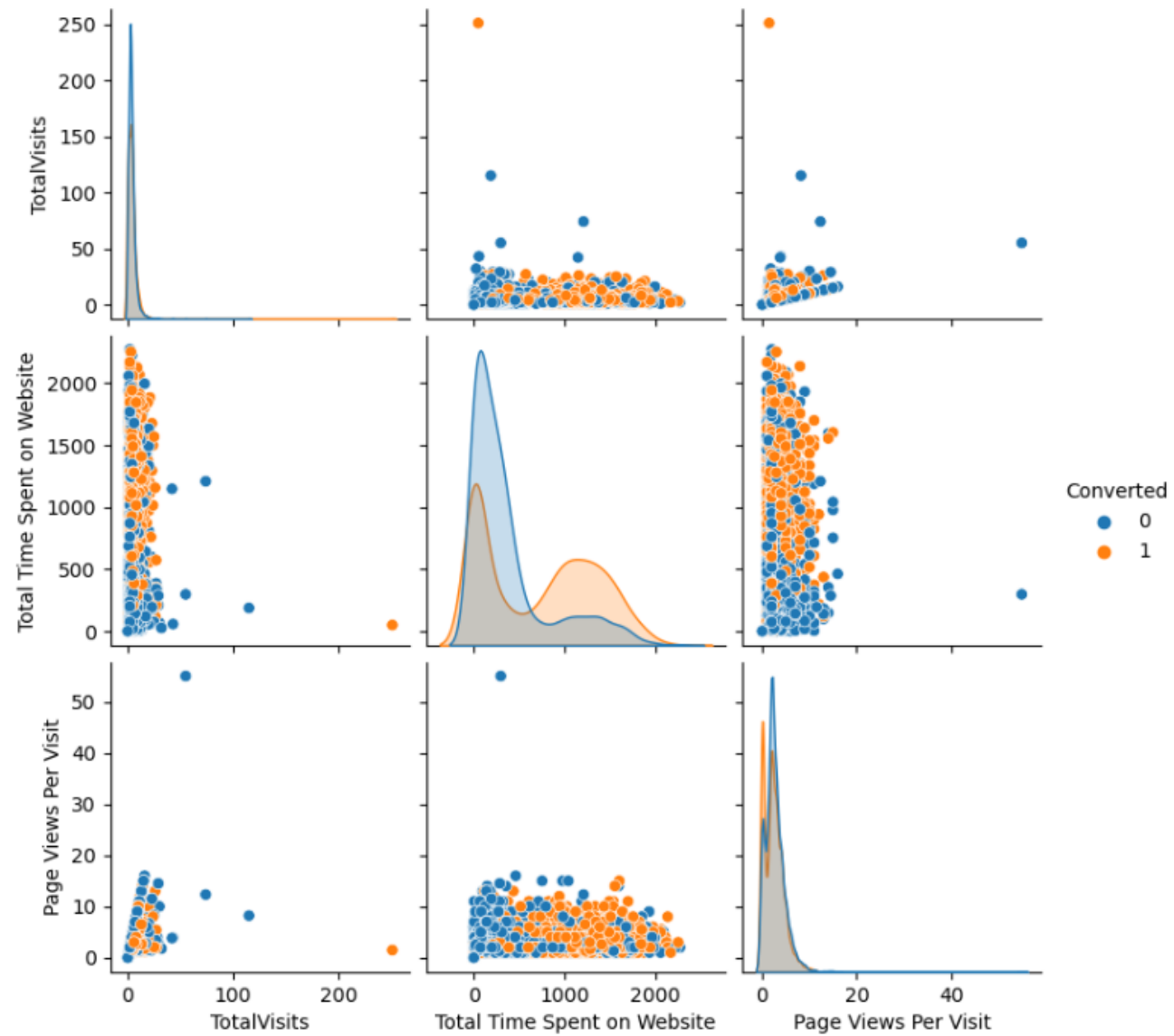


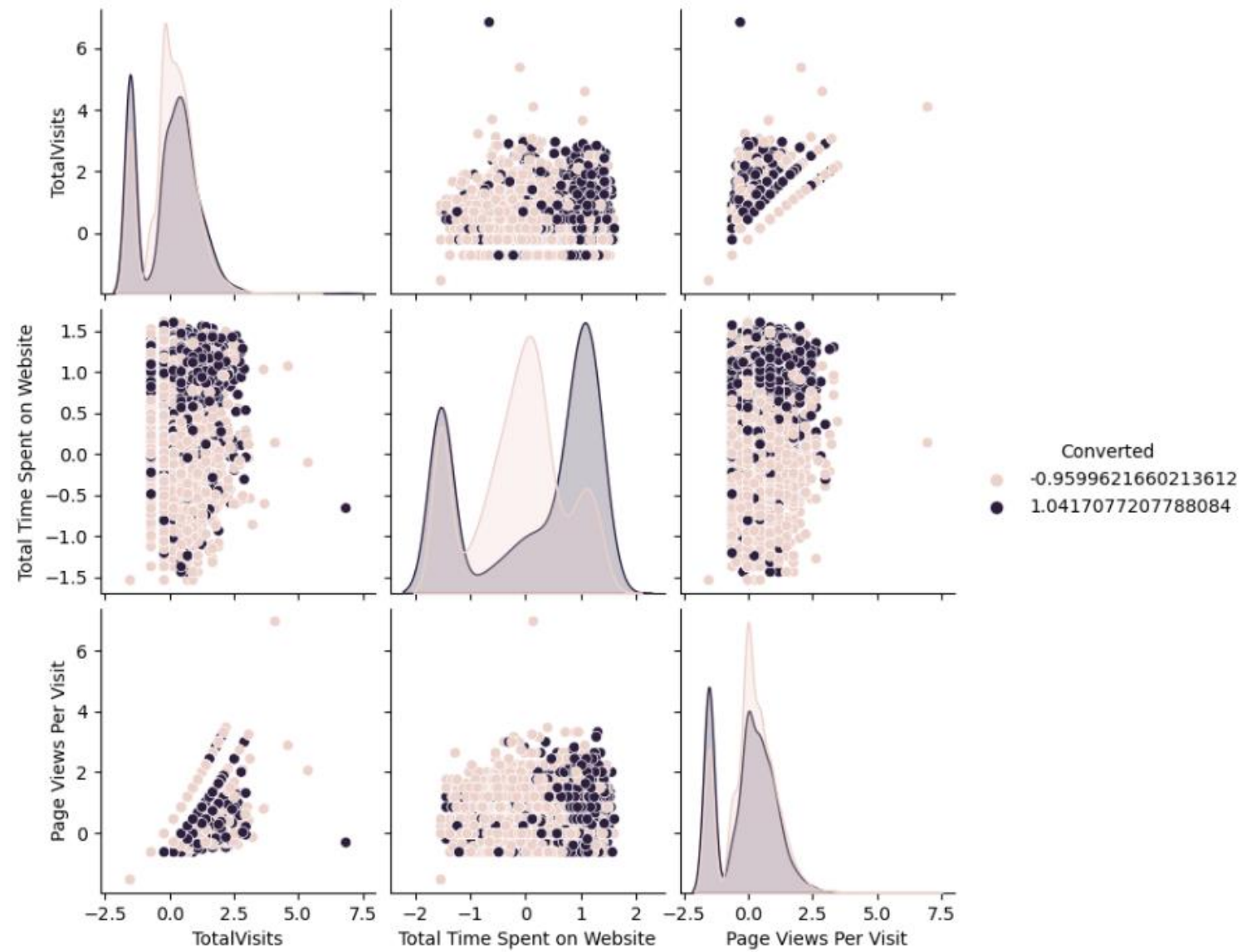
Dropping columns with high 'Select' values

EXPLORATORY DATA ANALYSIS

- Performed univariate and bivariate analysis for categorical and numerical variables







DATA PREPARATION AND MODELLING



Creating

Creating dummy variables



Splitting

Splitting the train – test data in ratio of 70:30.



Scaling

Feature Scaling

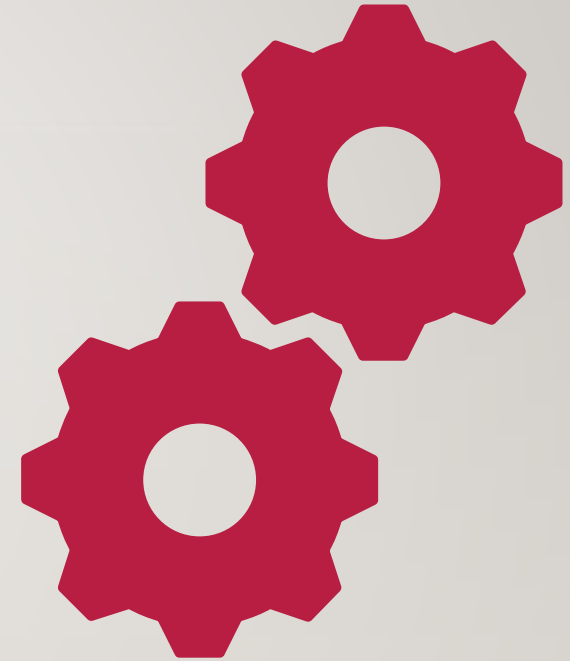


Checking

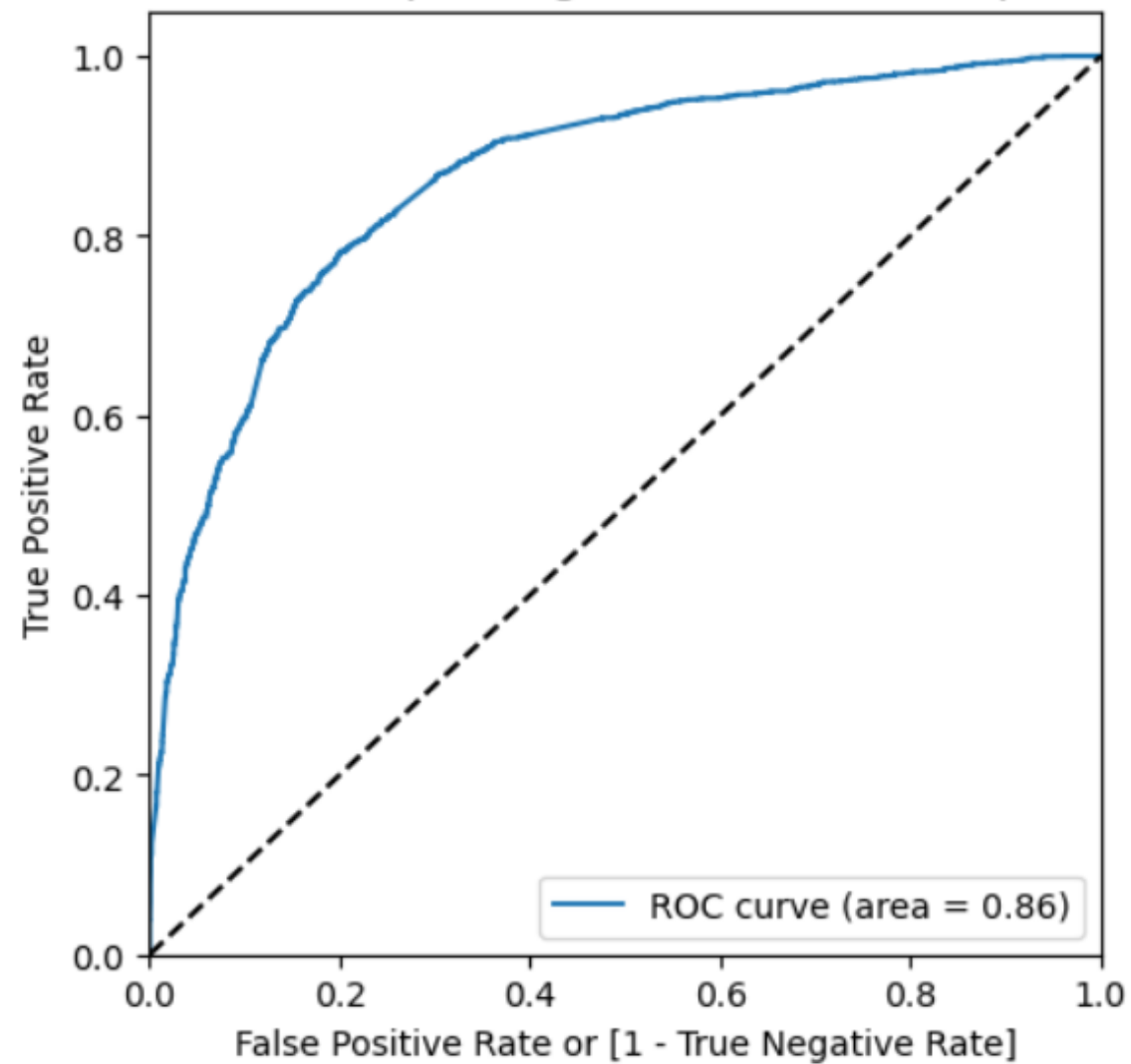
Checking the correlation.

MODEL BUILDING

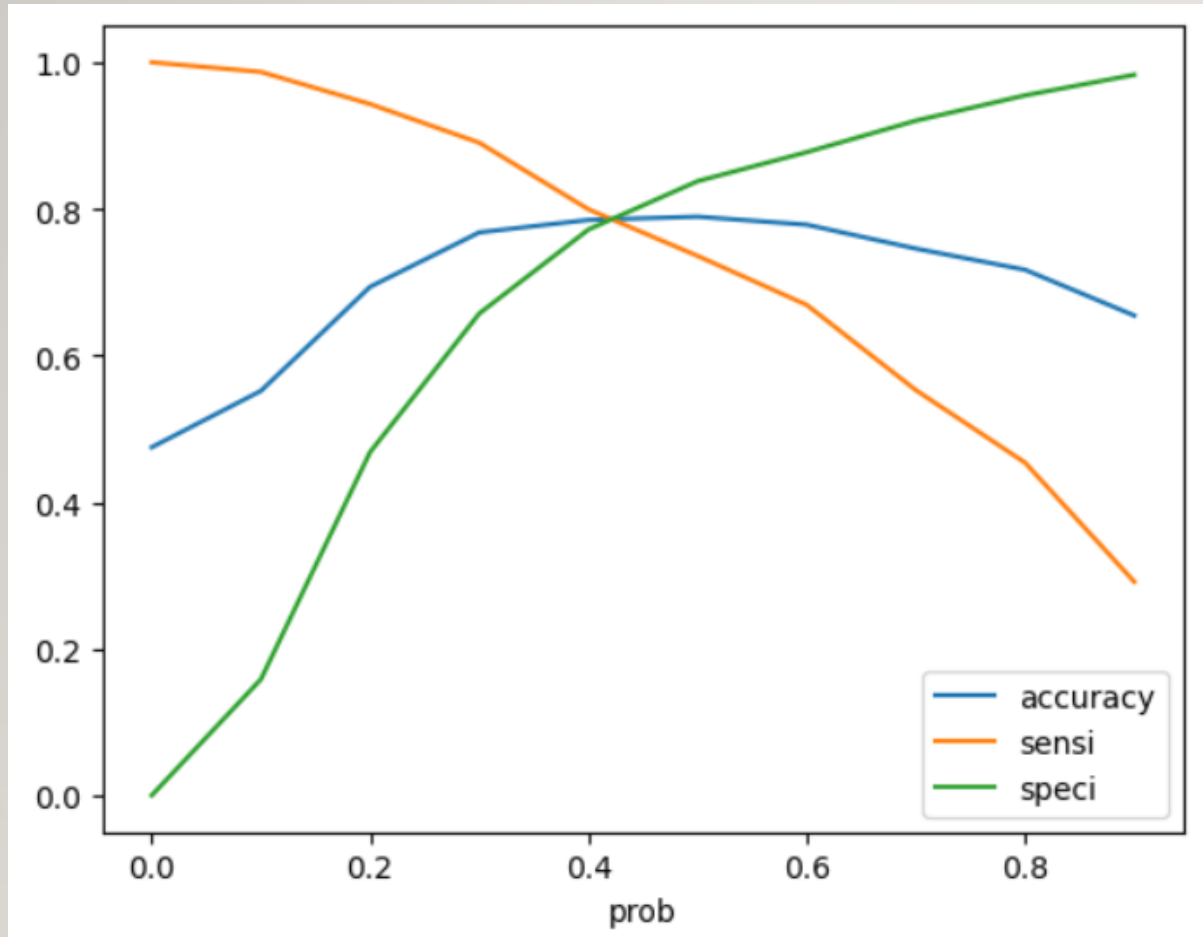
- Used RFE to reduce variables from 48 to 15 variables.
- Manual Feature Reduction process was used to build models by dropping variables with p value > 0.05
- All the variables have $VIF < 5$ and p-values are low.



Receiver operating characteristic example



-
- ROC Curve



-
- Optimal Cut-off point

MODEL EVALUATION

Train Dataset

- A. Cut-off value is 0.42
- B. Accuracy: 78.85%
- C. Sensitivity: 78.86%
- D. Specificity: 78.84%
- E. Precision: 80.40%

Test Dataset

- A. Cut-off value is 0.42
- B. Accuracy: 78.83%
- C. Sensitivity: 78.63%
- D. Specificity: 79.02%
- E. Precision: 78.31%

RECOMMENDATION



Feature responsible
for good conversion
rate are:

The total time spend
on the Website

Total number of visits

Lead Origin_Lead
Add Form

Current
occupation_Working
Professional

Last Notable
Activity_Had a Phone
Conversation



THANK YOU

