LEADS SCORING CASE STUDY

An effort by:

Muniba Naushad
Chandra Shekhar.Devanaboyina
Bhumika Nautiyal

Batch : DS C60 - 03/2023

Lead Score Case Study for X Education

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%

Business Objective:

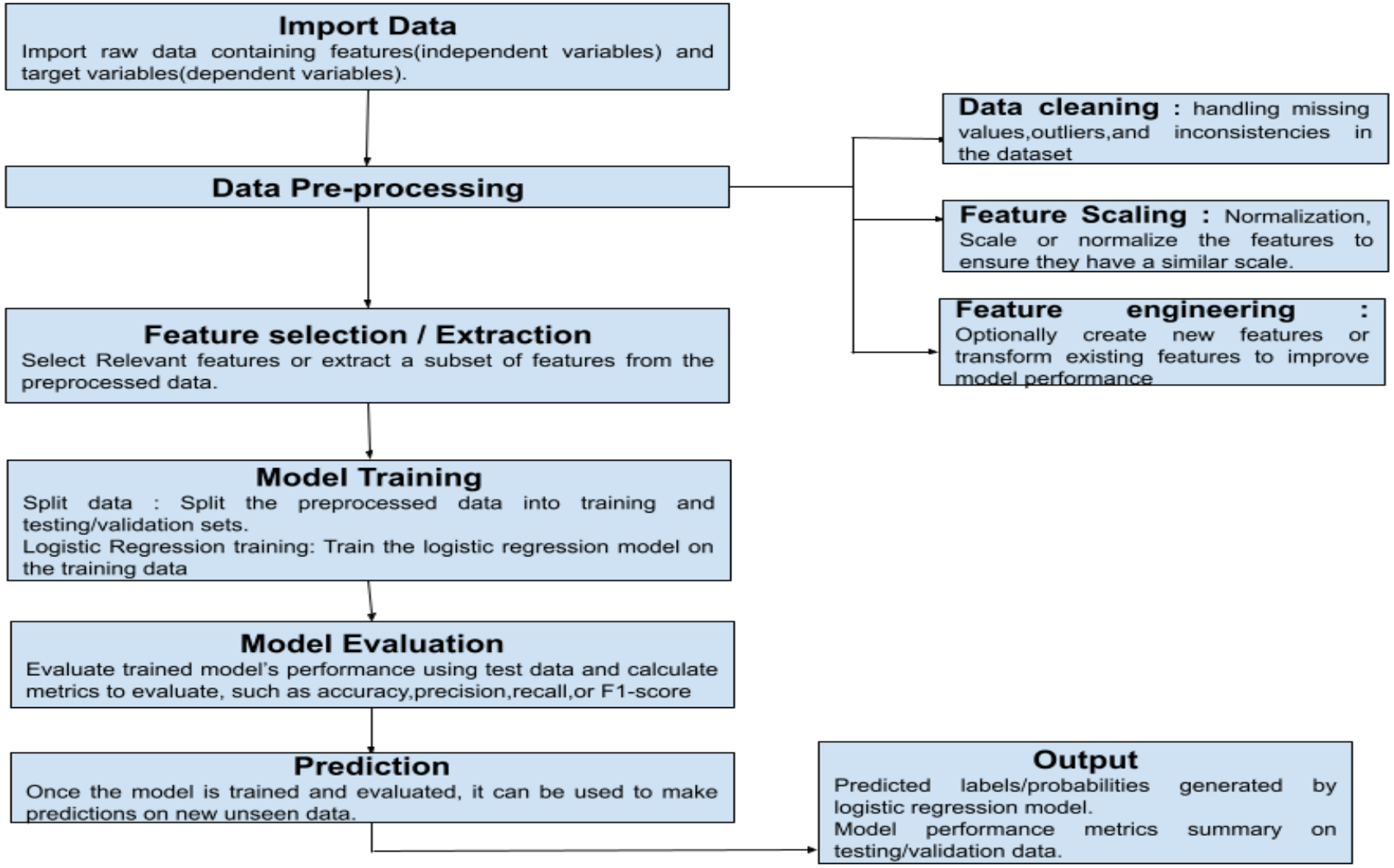
X Education needs us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a 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%.

Lead Score Case Study for X Education

Methodology :



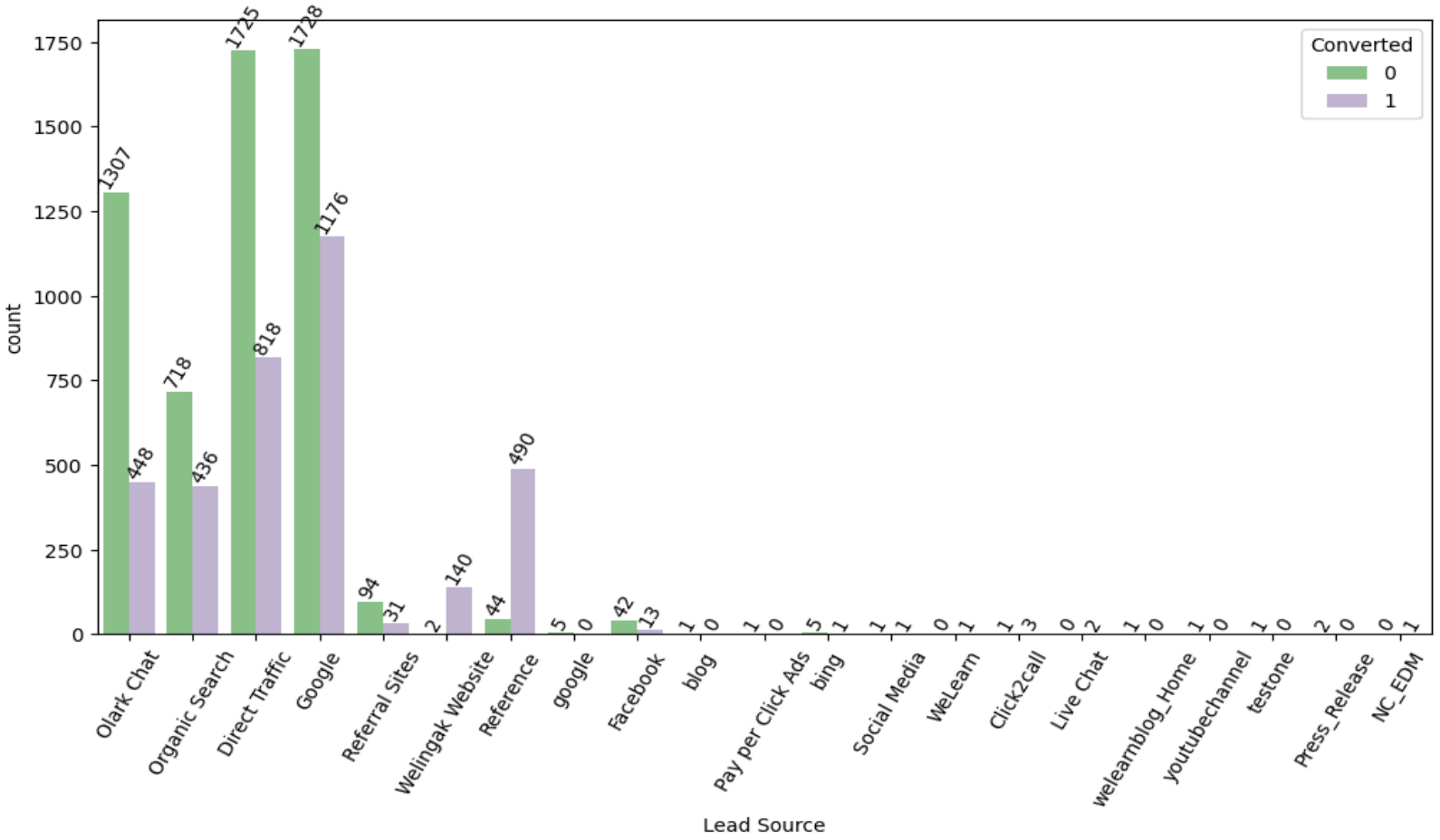
Lead Score Case Study for X Education

Data Manipulation

- Total Number of Columns=37, Total Number of Rows =9240.
- Dropped the columns that have 40% above null values that are irrelevant for analysis.
- Dropped columns of single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply” Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque”
- We dropped the columns ‘Specialization', 'Lead Profile’, and ‘How did you hear about X Education’ because they predominantly contain the value ‘select’, which is not useful for our analysis according to the problem statement.
- Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.
- We observed outliers in ‘TotalVisits’ and ‘Page Views Per Visit’ columns, so treated them through capping.
- After checking for the value counts for some of the object type variables, we found the features don’t have significant variance, so we decided to drop, them: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.
- We dropped the categories ‘Search', 'Newspaper’, ’Newspaper Articles' education Forums', 'Digital Advertisement’, and ‘Through Recommendations’ as they’re deemed irrelevant based on our categorical analysis.

Lead Score Case Study for X Education

Exploratory Data analysis :

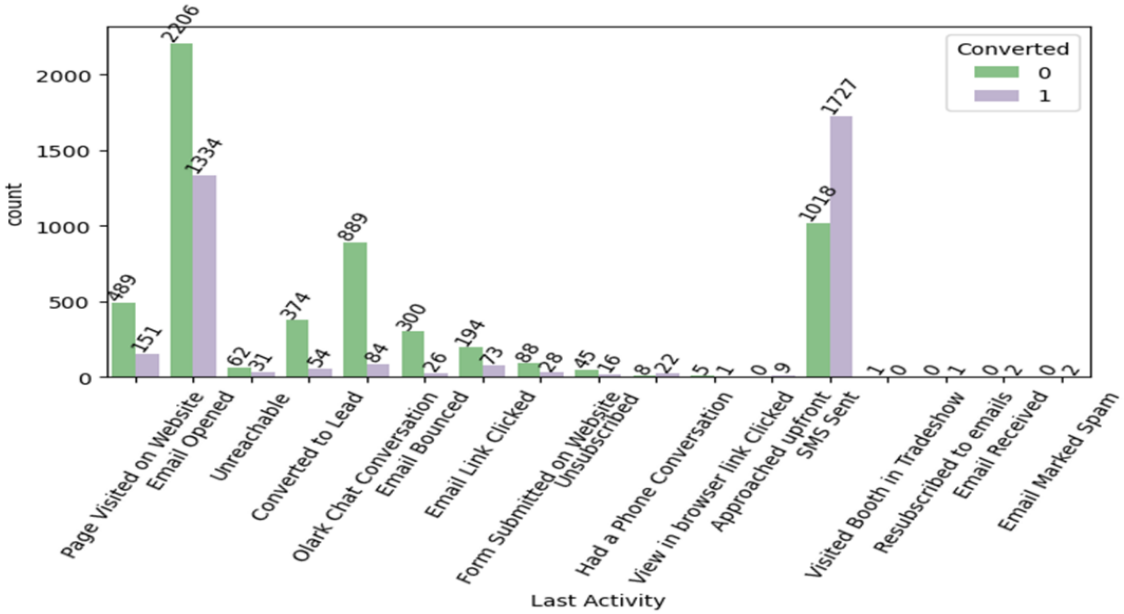


Conclusion:

- Google and direct traffic generate most leads.
- Conversion rate of reference leads and leads through welingak website is high.

Lead Score Case Study for X Education

Inference

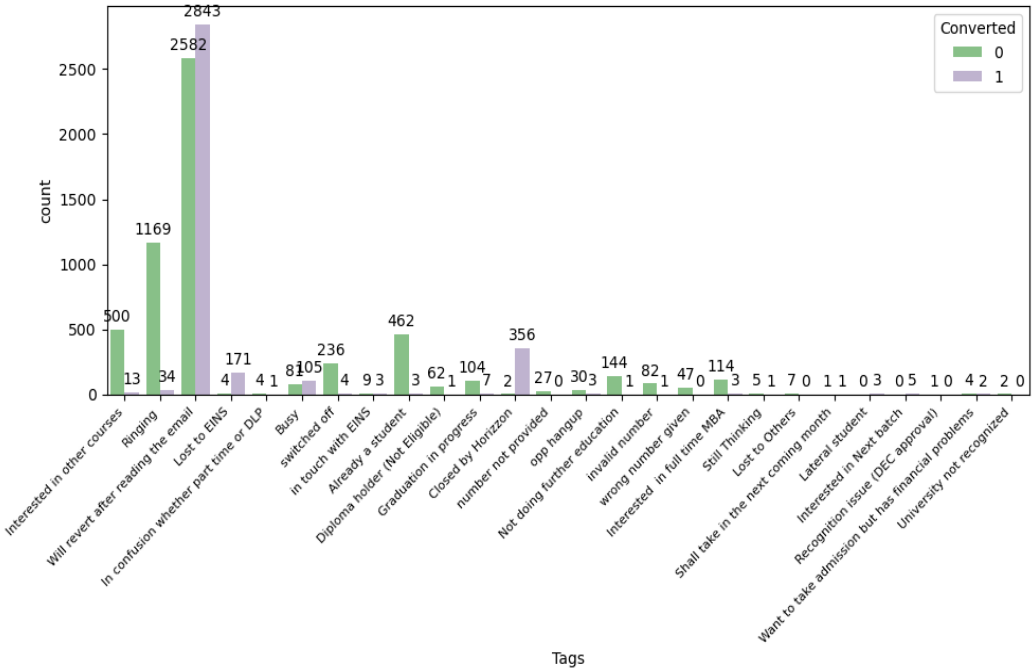


Conclusion

- Most of the lead have their Email opened as their last activity.
- SMS sent also have high lead

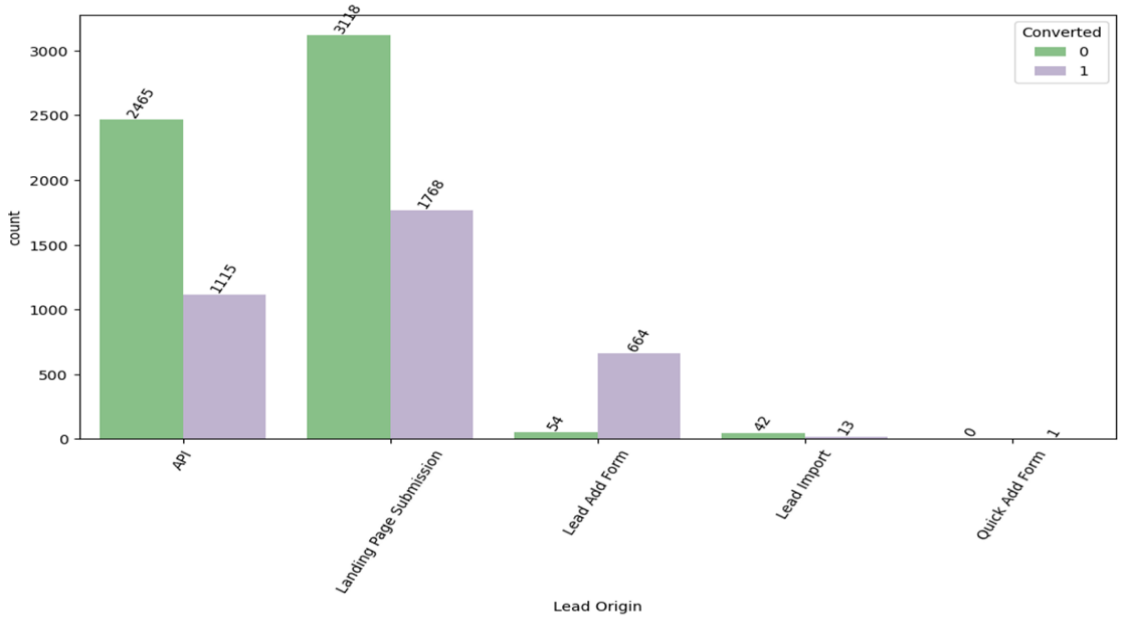
Conclusion:

- Good conversion rate of 'will revert after the email received.
- There isn't much variation in others.



Lead Score Case Study for X Education

Inference

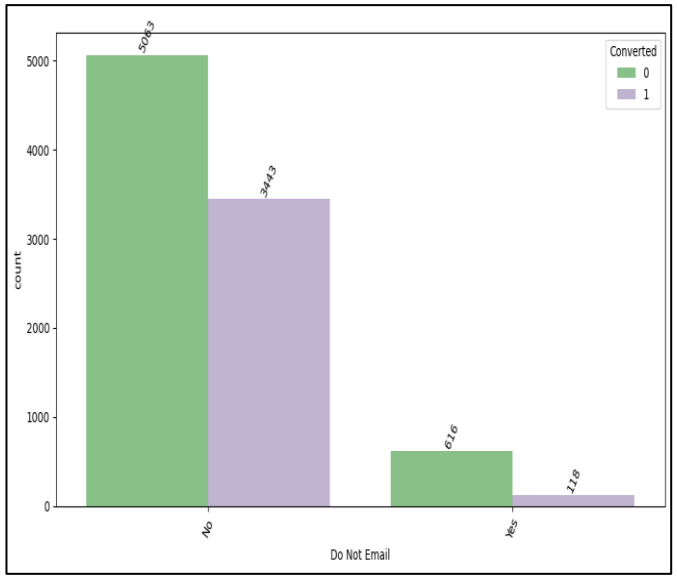


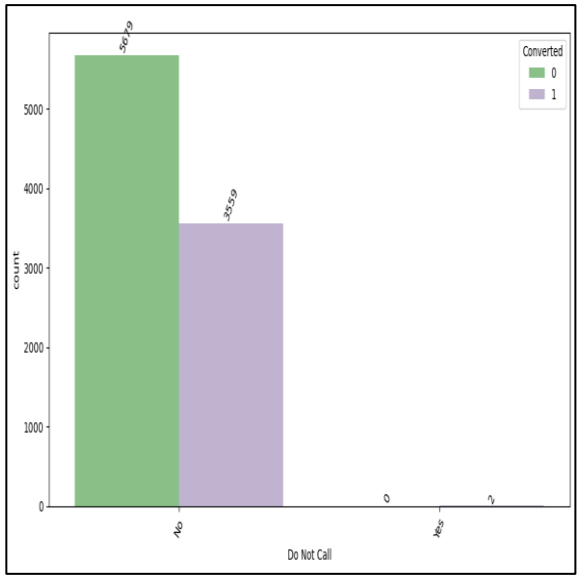
Conclusion

- API and Landing Page Submission have high lead.
- Lead Add form also shows high lead.

Conclusion:

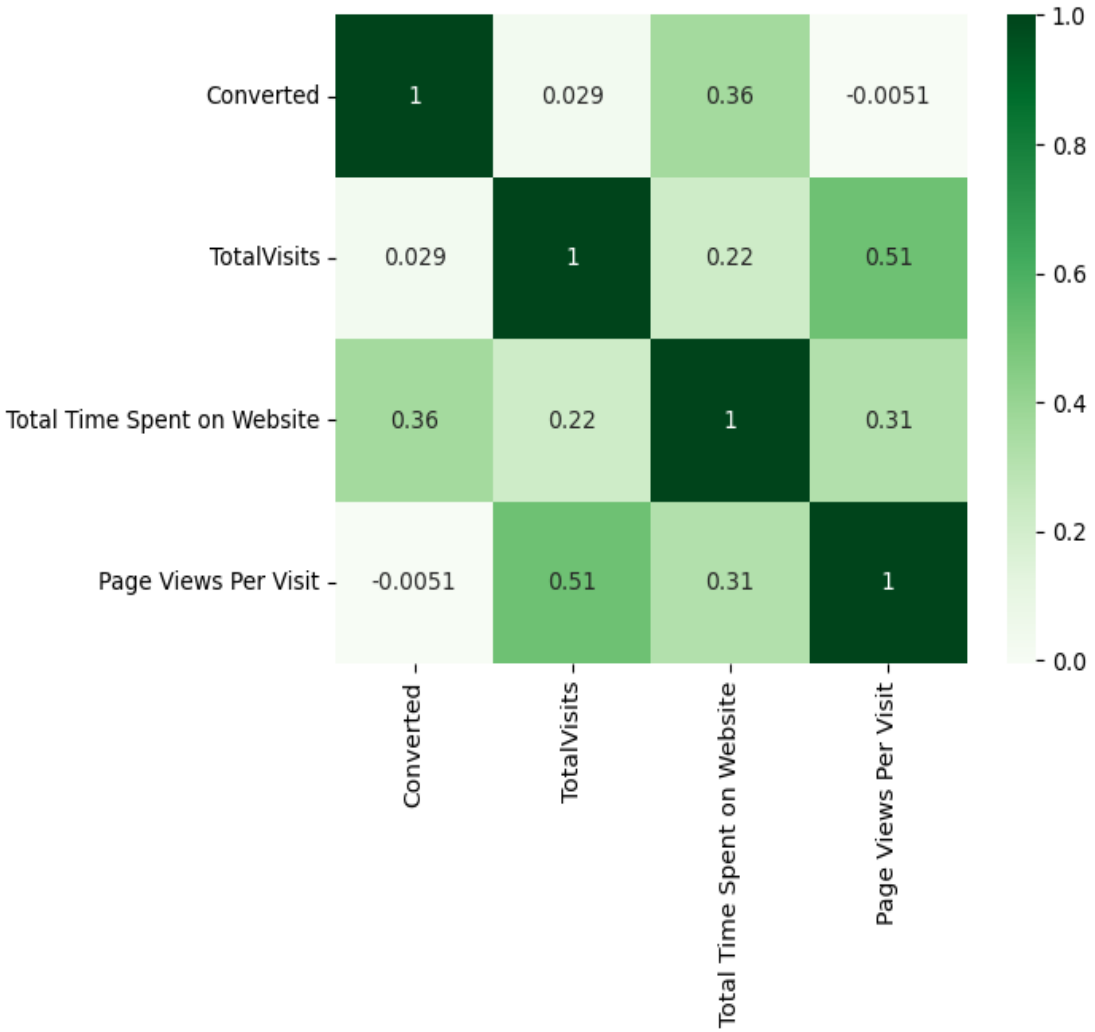
- As we can see, so many customers choose the option for Do Not Email and Do Not Call, which means most leads prefer not to informed neither through phone nor to be mailed.





Lead Score Case Study for X Education

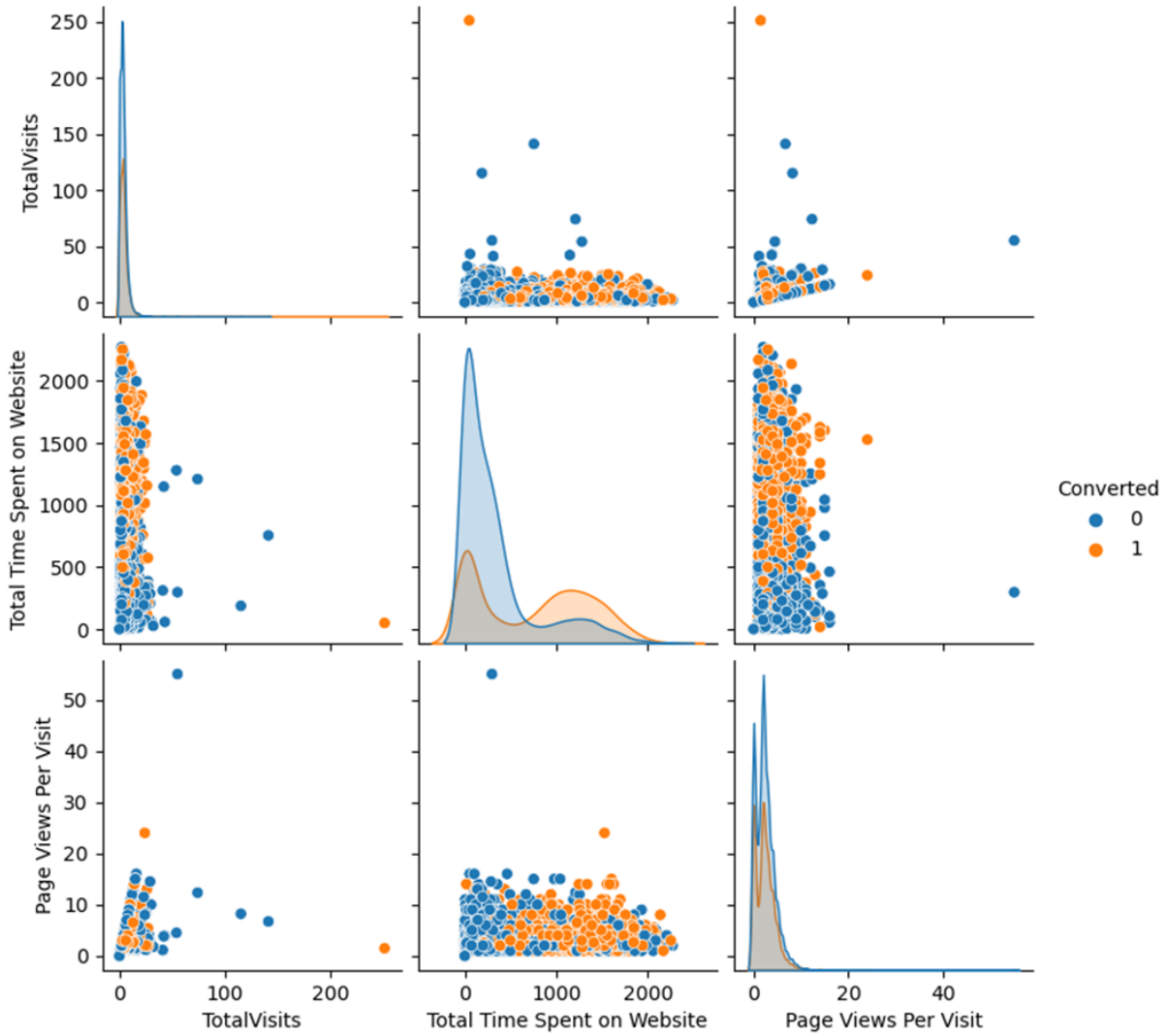
Inference



- There is positive correlation between Total time spent on website and Conversion.
- There is almost no correlation in page views per visit and total visits with conversion.

Lead Score Case Study for X Education

Inference



Total Time Spent on website

This variable has positive contribution as more time spent on the website higher chances of lead getting converted and the person actually end up taking the course.

Lead Score Case Study for X Education

Data Conversion

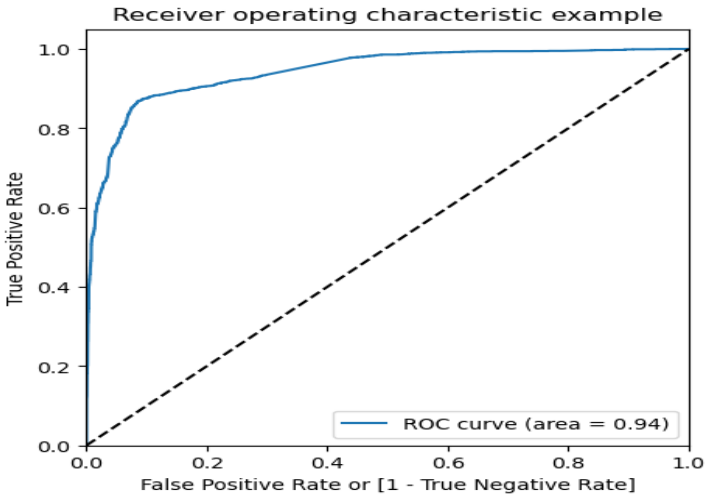
- Numerical Variables are Normalized through StandardScaler
- Dummy Variables are created for object type variables

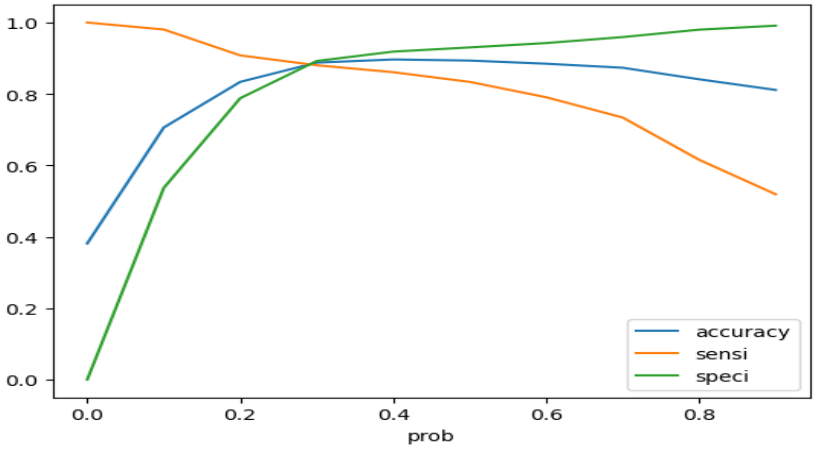
Model Building

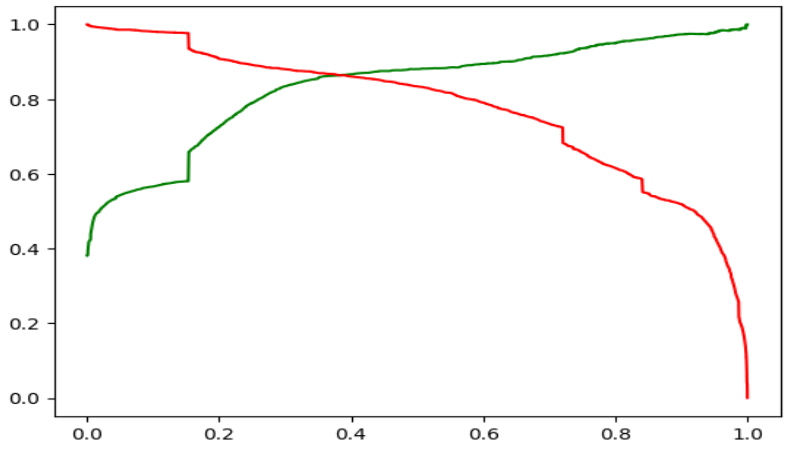
- Splitting the Data into Train and Test Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Using RFE for Feature Selection
- Running RFE with 30 variables
- Building Model by removing the variable whose p-value is greater than 0.05 and VIF value is greater than 5
- Predictions on test data set
- Overall accuracy 87.56% on train data, and 87.91% on test data set.

Lead Score Case Study for X Education

ROC Curve







Finding Optimal Cut off Point

- Optimal cut off probability is the probability where we get balanced sensitivity ,specificity, precision and recall
- The ROC curve has a value of 0.93, which is very good
- From these 3 graphs we choose our optimal tradeoff point as 0.42
- We choose Precision and Recall as our evaluation metric because their values are better than specificity and sensitivity

Lead Score Case Study for X Education

Conclusion

It was found that the variables that mattered the most in the lead converted are

- Total Time Spent on Website
- Tags
  - Lost to EINS
  - switched off
  - Ringing
  - Closed by Horizzon
  - Graduation in process
  - Invalid number
- Do Not Email
  - Yes
- Last Activity
  - Converted to Lead
  - Email Bounced
  - Olark Chat Conversation

- Cont.to next slide

Lead Score Case Study for X Education

It was found that the variables that mattered the most in the lead converted are:

- What is your current occupation
  - Working Professional
- Lead Origin
  - Lead Add Form
  - Lead Source Welingak Website
- Last Notable Activity
  - SMS Sent
  - Had a Phone Conversation
  - Unreachable

LEADS SCORING CASE STUDY

UpGrad

THANK YOU

UpGrad