LEADS SCORING CASE STUDY

Submitted by

Sharmila Sherin
Shashidhar M
Shivashankar S M

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:

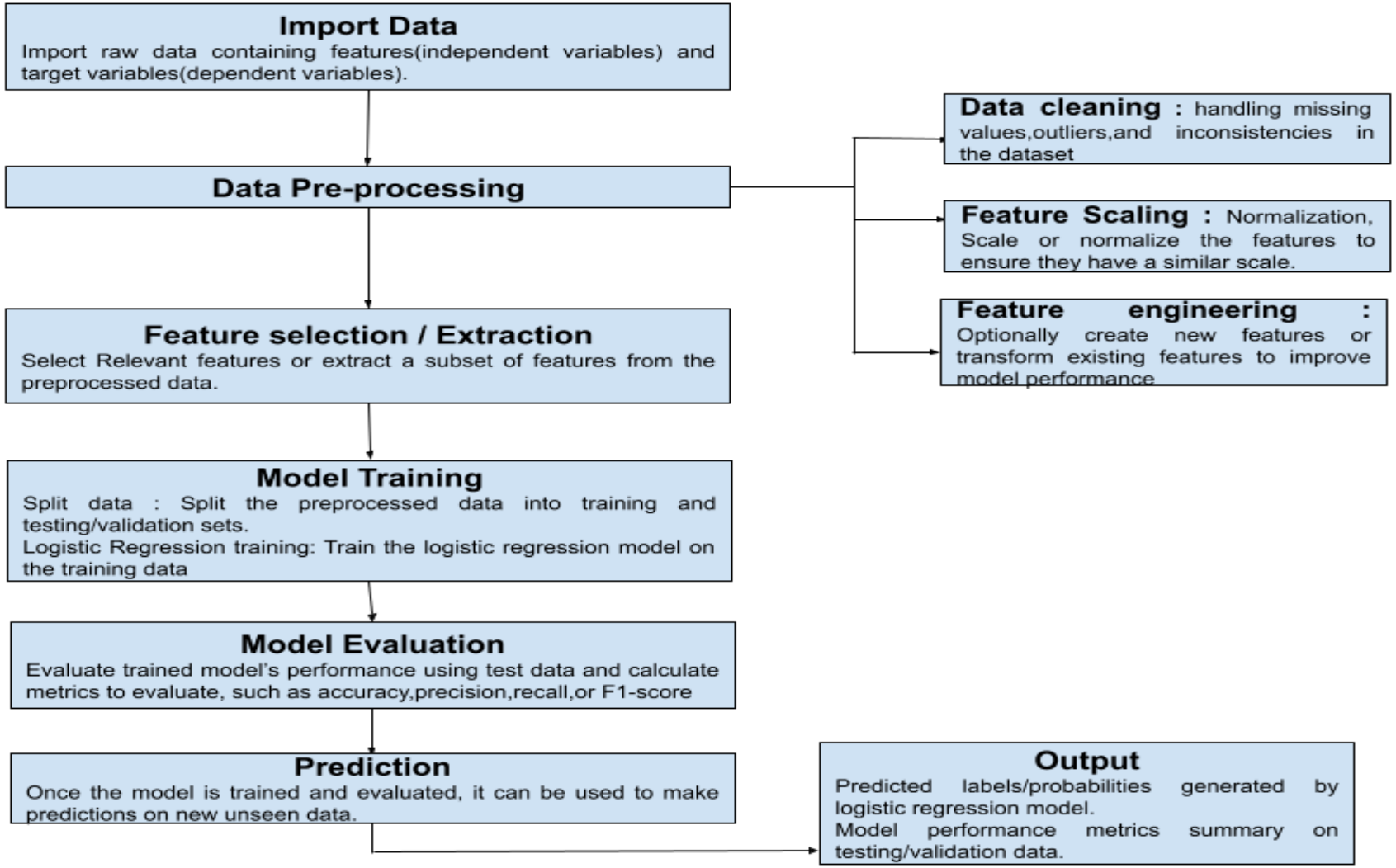
X Education needs us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

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 :



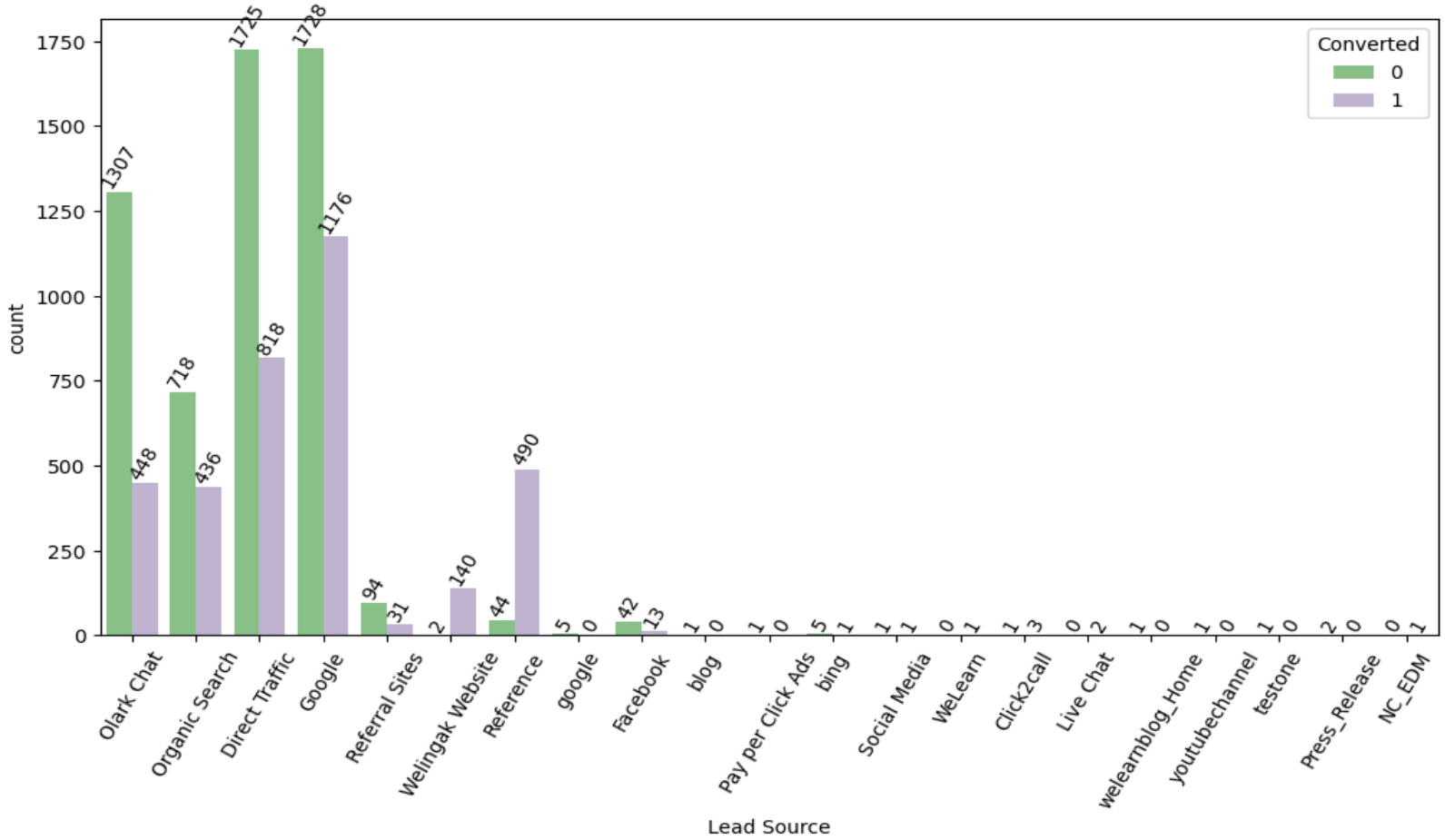
Lead Score Case Study for X Education

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 dropped the ‘City’ and ‘Country’ columns because they don’t contribute to our analysis since the platform is online based.
- We observed outliers in ‘TotalVisits’ and ‘Page Views Per Visit’ columns, so treated them using capping.
- After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: “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 :

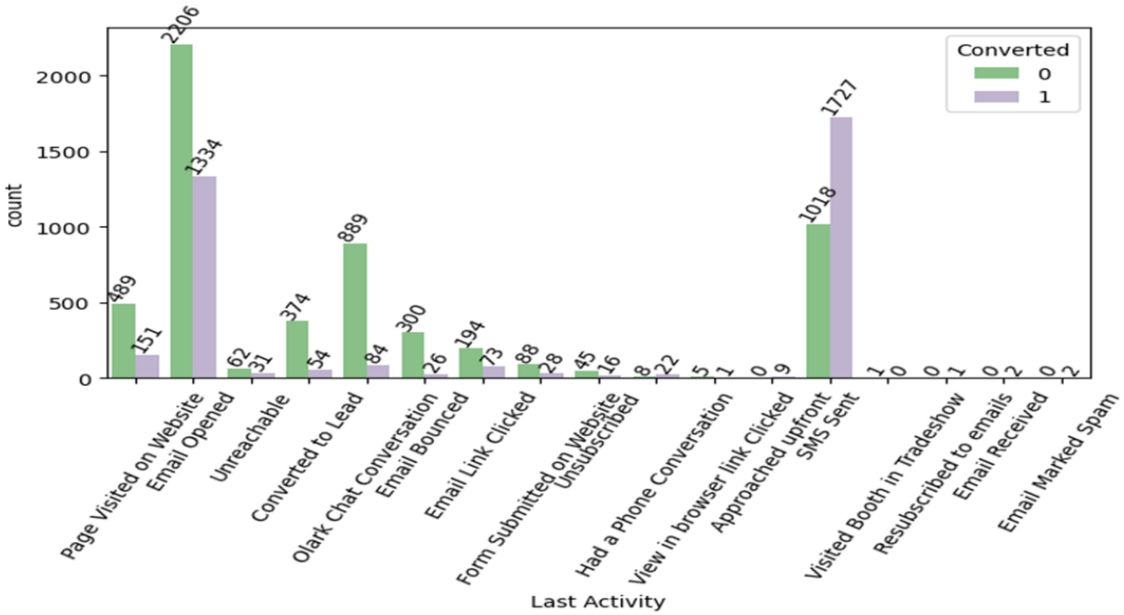


Conclusion:

- 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

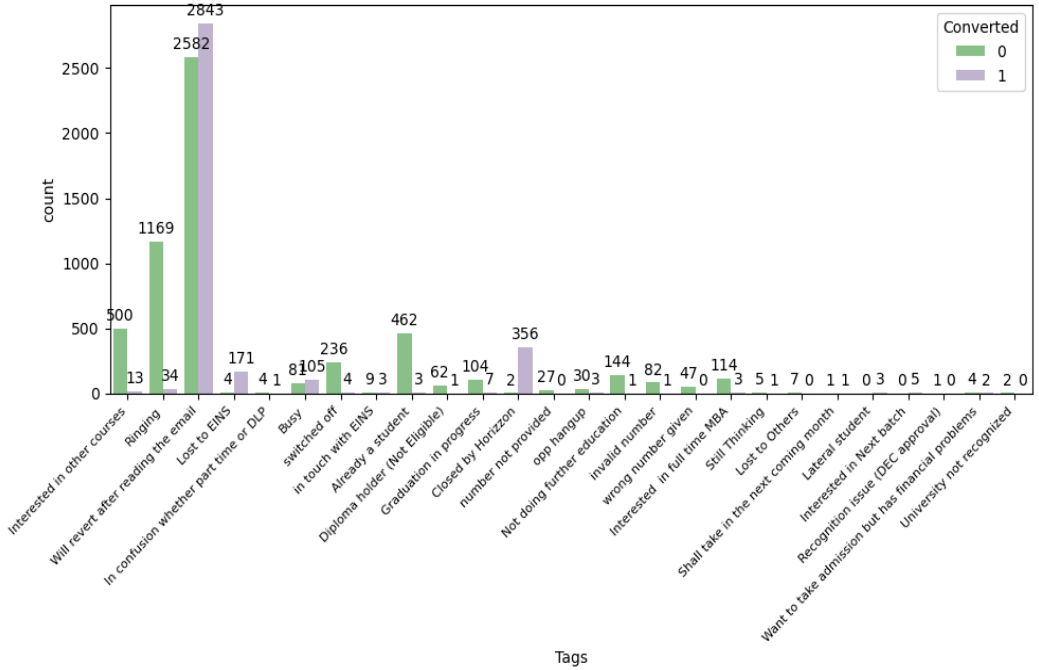


Conclusion:

- Will revert after reading the email has a very good number of converted.
- There isn't much variation in others.

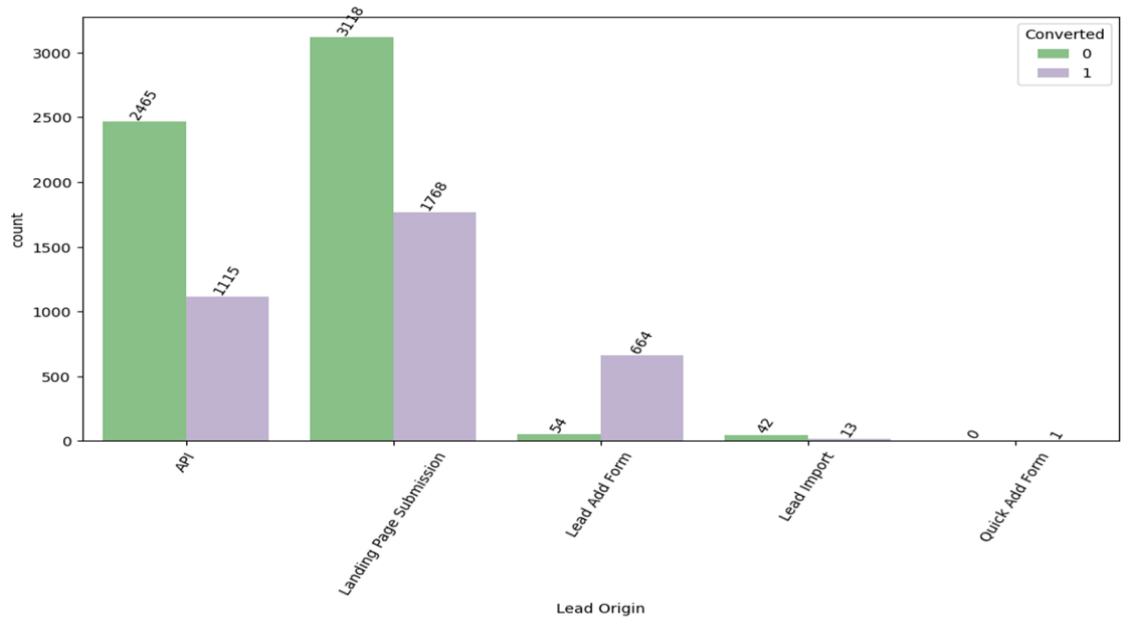
Conclusion

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



Lead Score Case Study for X Education

Inference

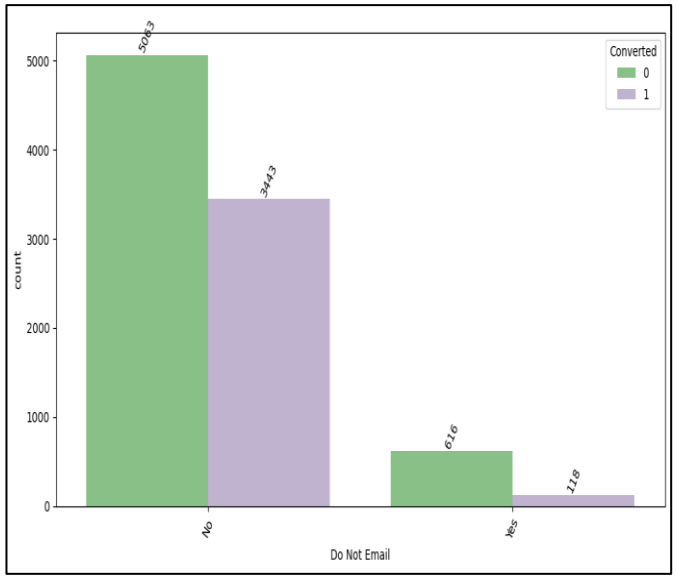


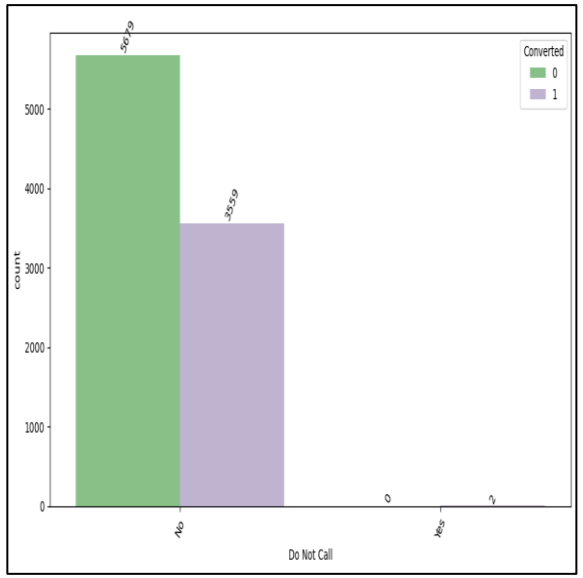
Conclusion:

- 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.

Conclusion

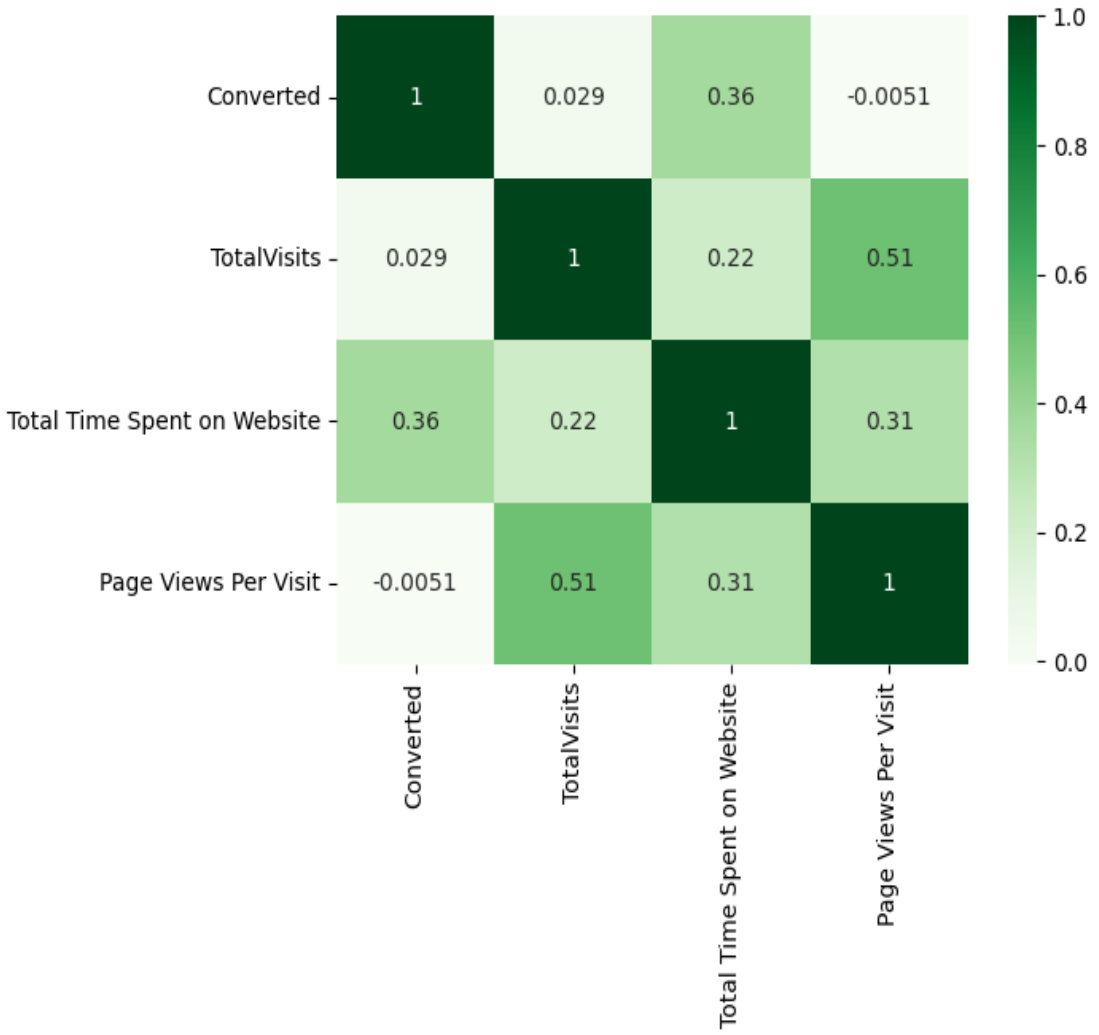
- API and Landing Page Submission have high lead.
- Lead Add form also shows high lead.





Lead Score Case Study for X Education

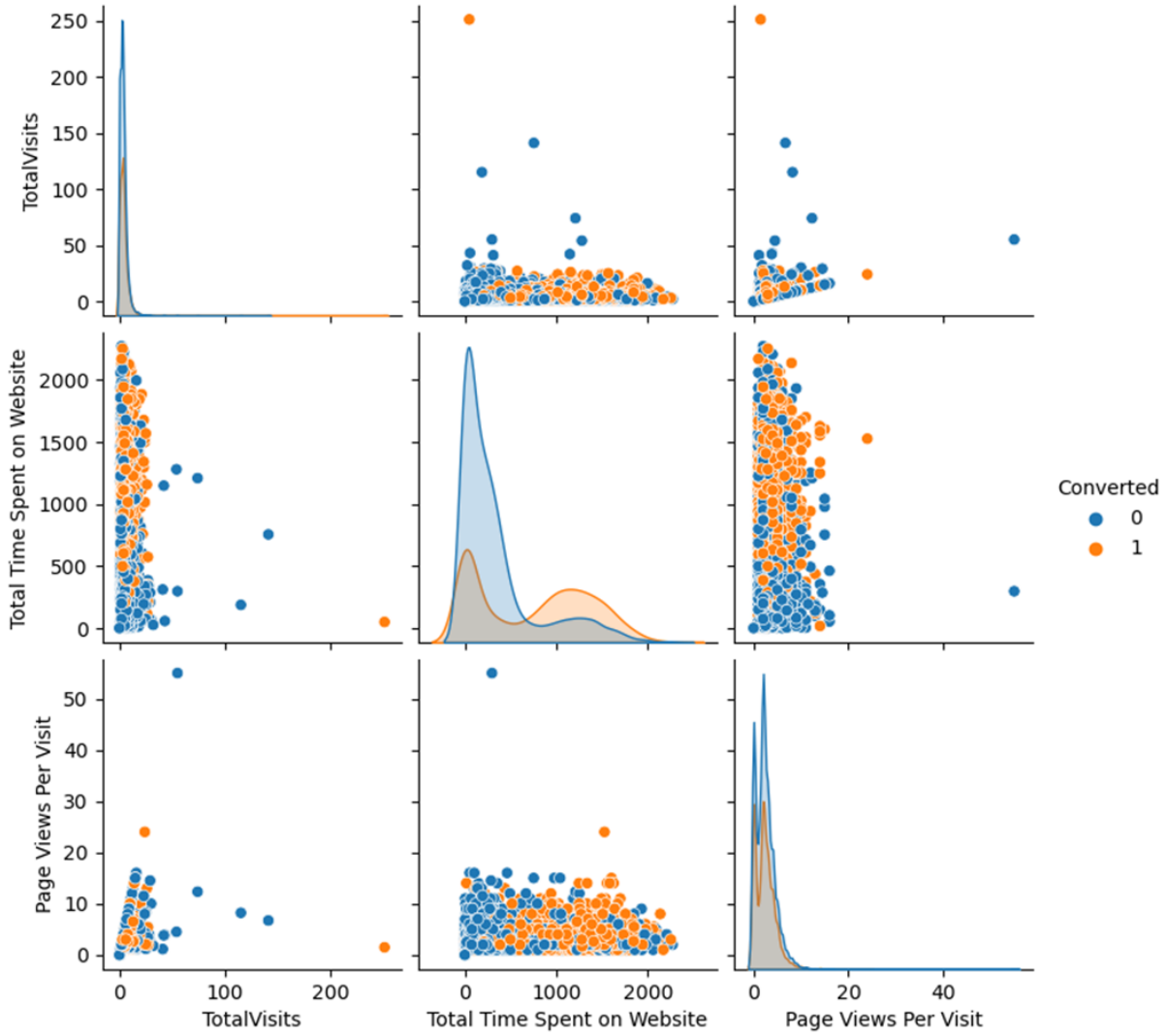
Inference



- 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



- Customers who spend more time on the website have a higher Leads, indicating that increasing the time spent on the website can lead to higher conversion rates.

Lead Score Case Study for X Education

Data Conversion

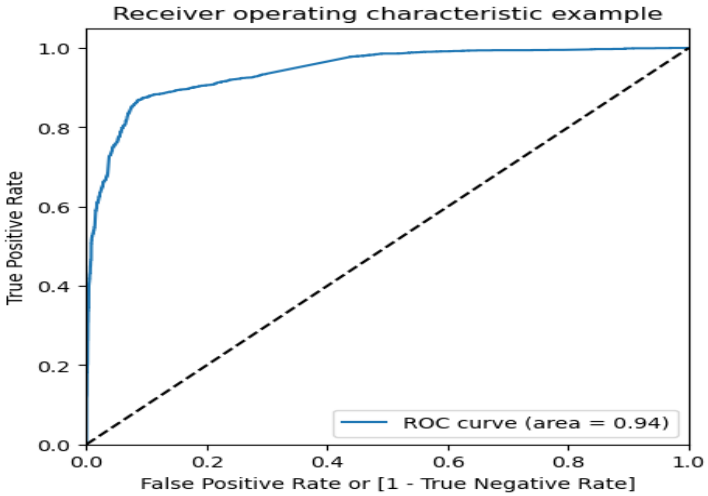
- Numerical Variables are Normalized
- Dummy Variables are created for object type variables

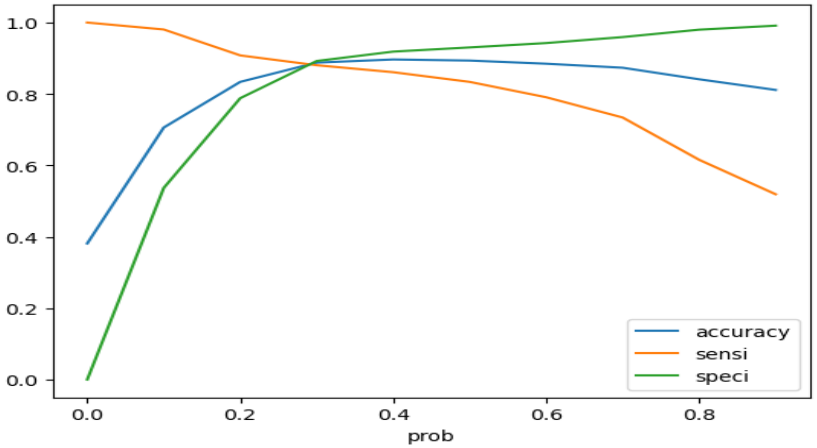
Model Building

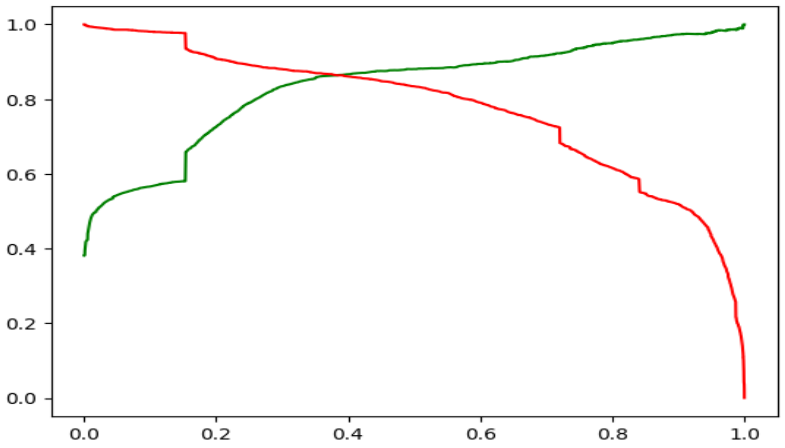
- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 20 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 89.6% on test data, and 89.6% on train data set.

Lead Score Case Study for X Education

ROC Curve







Finding Optimal Cut off Point

- Optimal cut off probability is that
- Probability where we get balanced sensitivity ,specificity and precision, recall
- The ROC curve has a value of 0.94, which is very good
- From these 3 graphs we choose our optimal cut off point as .42 where our specificity 93% sensitivity 85% precision 86% and recall 86%
- We choose specificity and sensitivity as our metric because their values are better than precision recall

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
  - Will revert after reading the email
  - in touch with EINS
  - switched off
  - Ringing
  - Closed by Horizzon
  - Busy
  - Lost to EINS
- Do Not Email
  - Yes
- Last Activity
  - Converted to Lead
- What is your current occupation_
  - Working Professional
- Lead Origin
  - Lead Add Form
- Last Notable Activity
  - SMS Sent
  - Had a Phone Conversation

LEADS SCORING CASE STUDY

THANK YOU