Lead Scoring Case Study for X Education Company

The presentation on the Lead scoring
for X Education company

Logistic Regression module building to increase
lead conversion ratio.

Presentation By

Menaka Timmanpyati

Pankaj Ligade

Oct 2022

Problem statement

- X Education sells online courses to industry professionals.
- The company markets its courses on several websites and search engines like Google.
- Leads are generated from the form fill on web site and from past referrals.
- The The typical lead conversion rate at X education is around 30%.
- The has given a ballpark of the target lead conversion rate to be around 80%.
- A lead score 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 char

Lead Conversion Process - Demonstrated as a funne

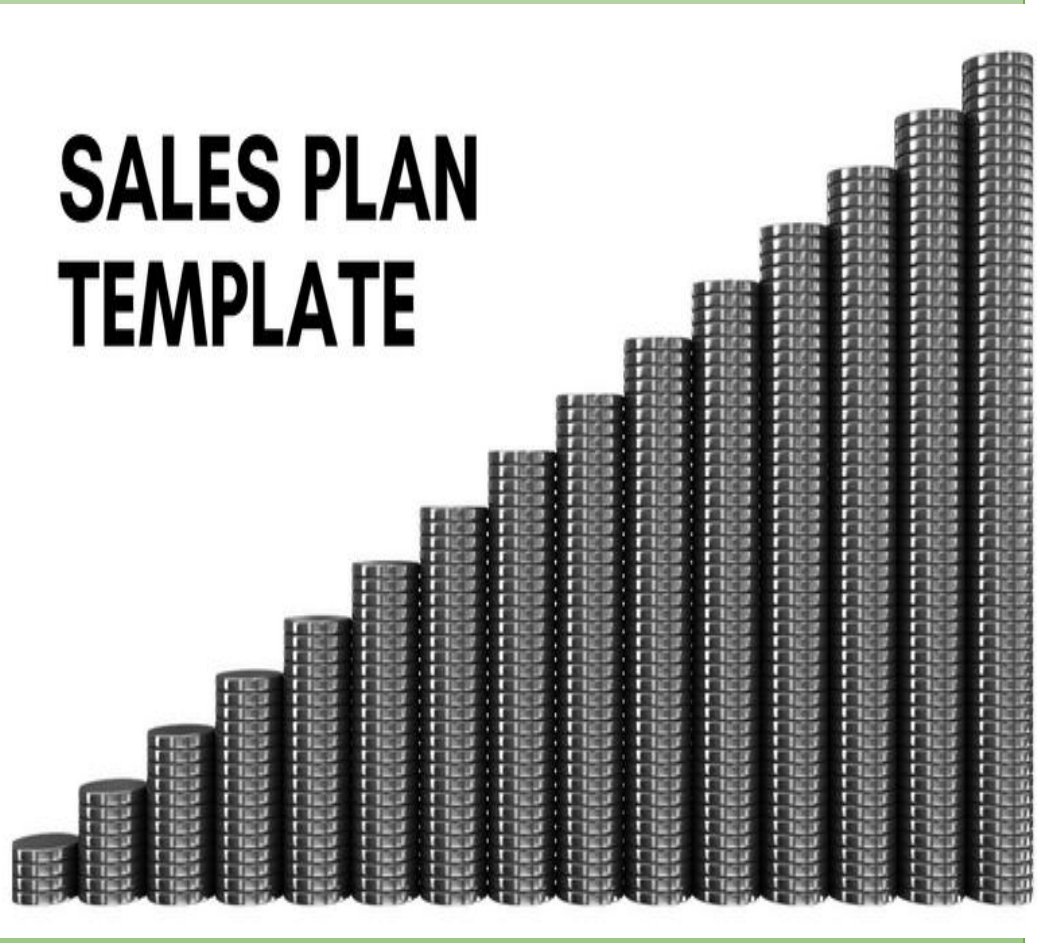
Goals of case Study

1.Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

2.There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

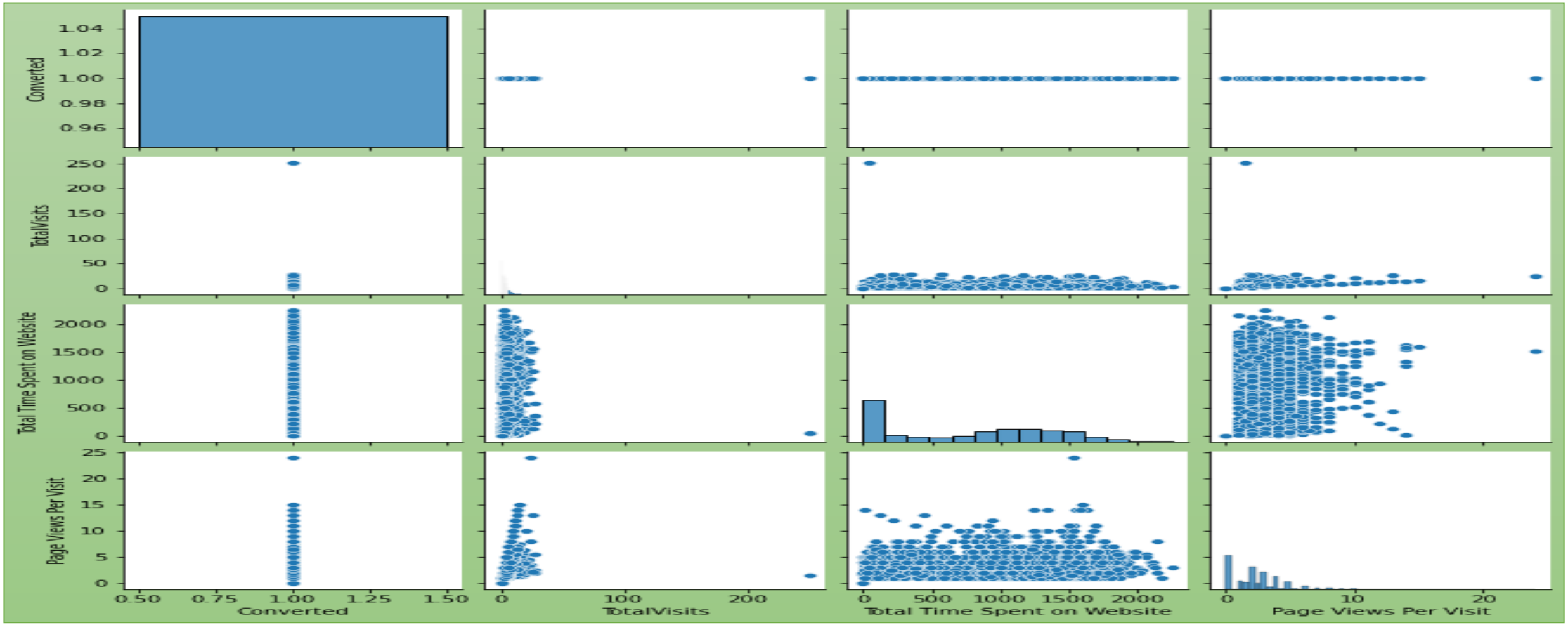
3. To adjust to if the company's requirement changes in the future so you will need to handle these as well.

4. The CEO has given ballpark the target of the target lead conversion of 80%

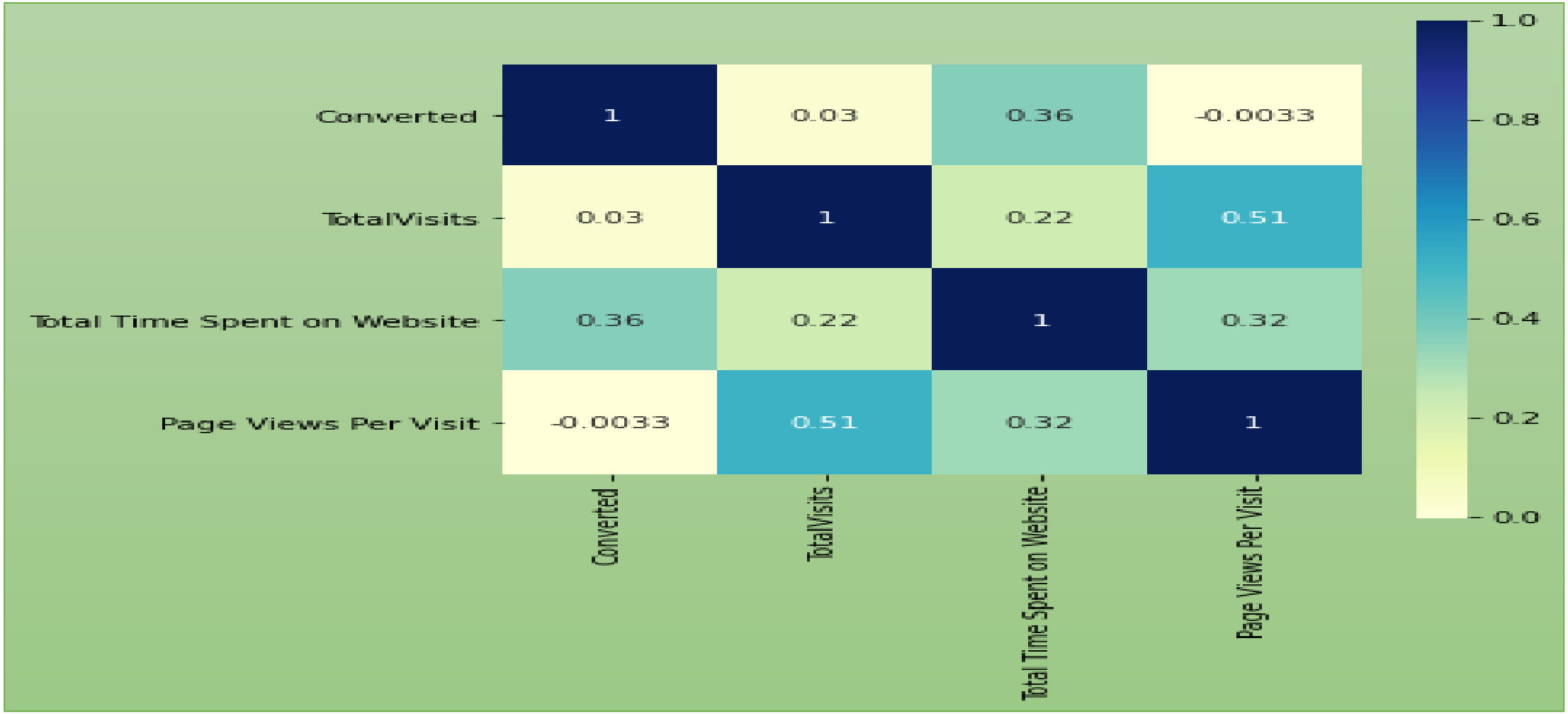


Strategy used for case study

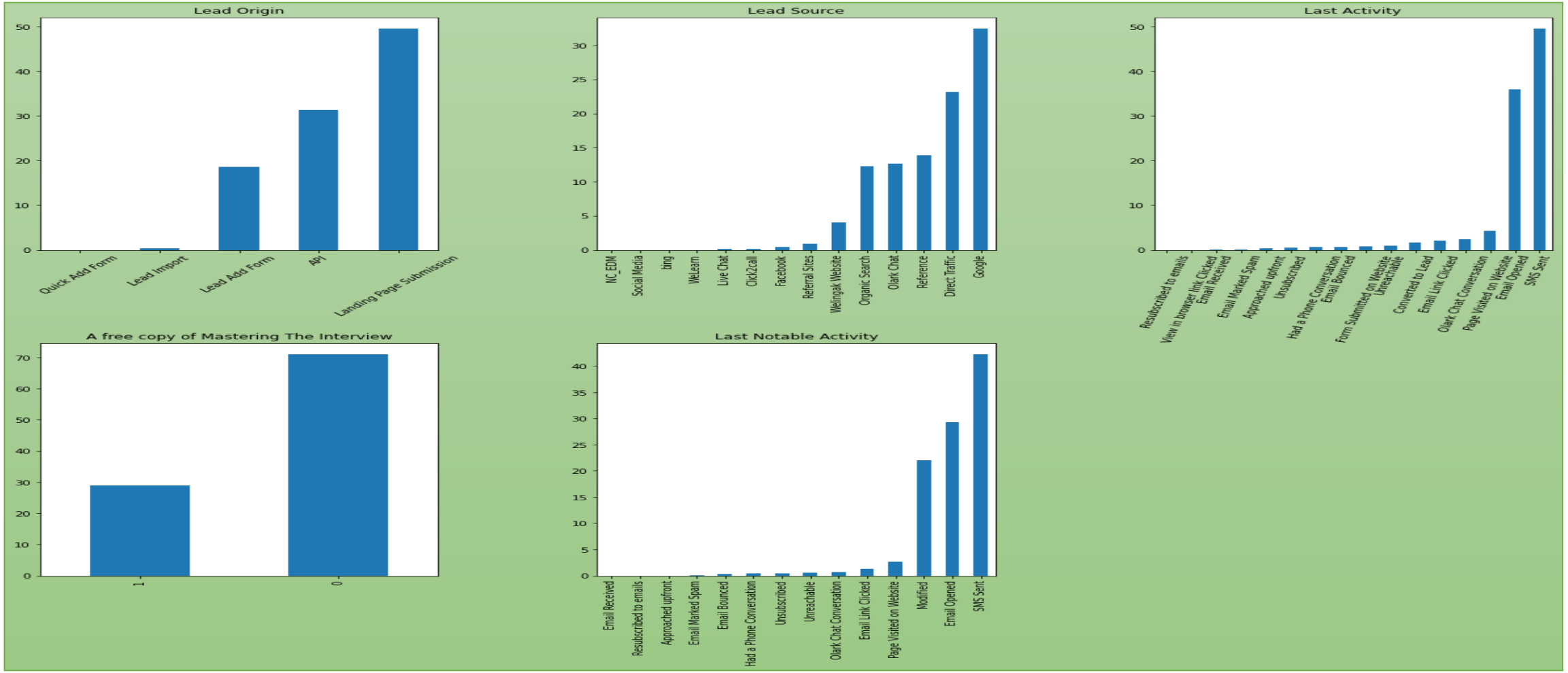
- Import the data for Analysis
- Understanding and cleaning the data
- Exploratory Data Analysis
- Categorical Analysis
- Dummy modelling
- Splitting data in to Train and Test data set
- Building Logistic Regression module
- Evaluating module by different variables
- Applying best module in Test data set by Sensitivity check



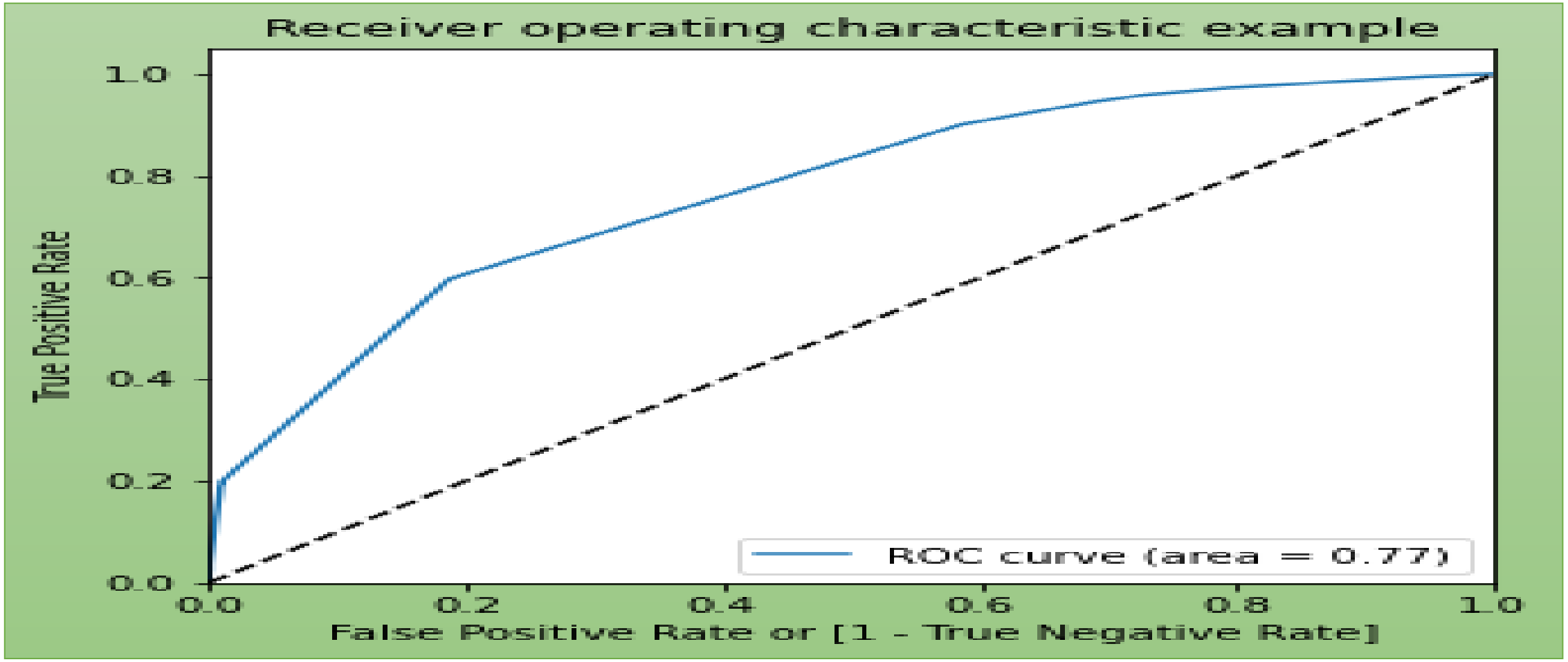
Observations:
1) Leads who get converted to clients typcially spend a lot more time on X Education's website.
2) We see that leads who get converted typcially vist lot many pages of the website.



We can see that there is correlation converted to Total time spend on wedsite



- Lead origin has major conversion from landing page submission
- Major conversion is from Google search
- A free copy of mastering interview not having that much impact on lead conversion
- Major conversion is from SMS, email and calls made.
- We have around 39% conversion rate.
- More conversion happen with unemployed people.

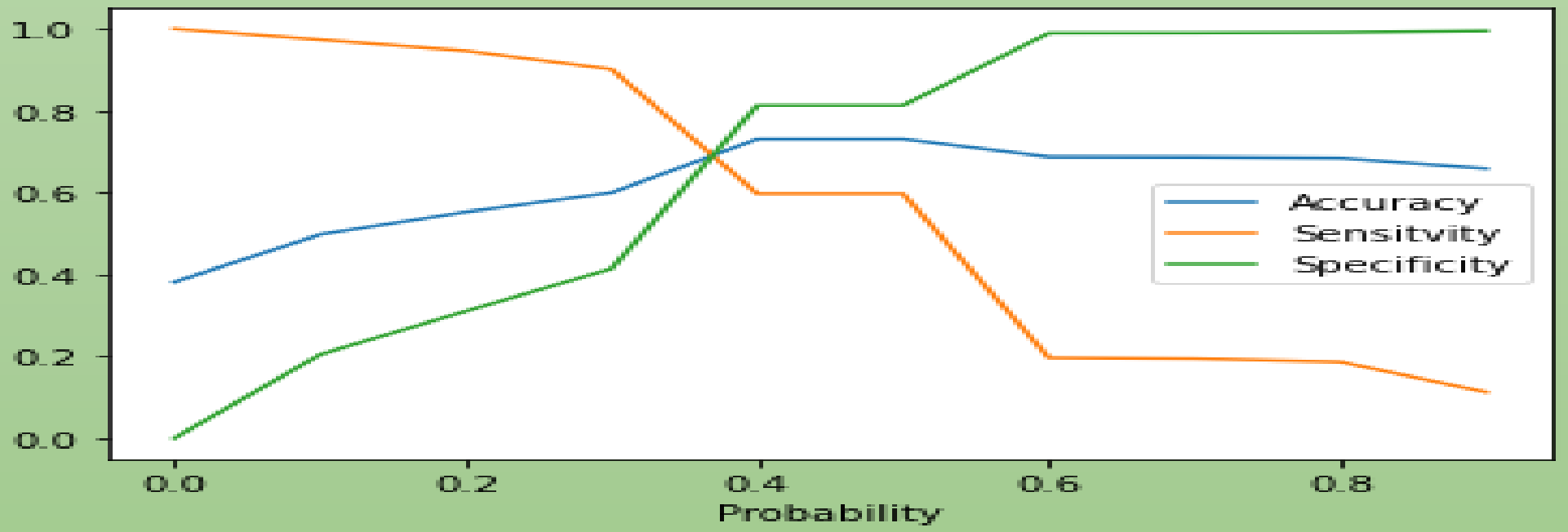


- The area of ROC curve should be close to 1. The area of the above ROC curve is 0.97

Variables Impacting the Conversion Ration

- Do not mail
- Total Visits
- Total time spend on website
- Lead page submission
- Lead form
- Olark chat
- Email bounce
- Note sure
- No Information
- Working profession
- Unreachable’

Probability cutoff point



0.375 is the optimum point for taking probability cutoff as the meeting point is slightly before from 0.375 hence final cutoff we choose is 0.375.

Module Evaluation on Test data

- Accuracy is almost 73% which very good.
- Sensitivity is 59%
- Specificity is 81%
- From confusion matrix we can say

Predicted	not converted	Converted
Actual		
not_converted	3253	994
converted	749	1472

Confusion matrix

Conclusion

1) The lead_score column can be used to identify the potential leads to focus first.

2) Higher the score, higher are the chances for the lead to convert.

3) In case, there are limited sales representatives, then the score cut-off should be higher to ensure a higher conversion probability people are contacted further to turn them into a potential customer.

4) It is the same as increasing the precision value of the model by adjusting the cut-off point to a higher value.

5) In case there are more resources available in the sales team (i.e., interns, etc.), then the score cut-off can be lowered.

6) As there are more human resources, the company can afford a higher rate of False positives as it will increase the customer outreach and, in turn, increase the potential customer who will take the online courses.