Lead Score Case Study

Members –

Pratik Walunj & Prajwal JM

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%.

- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Goal ...

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. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

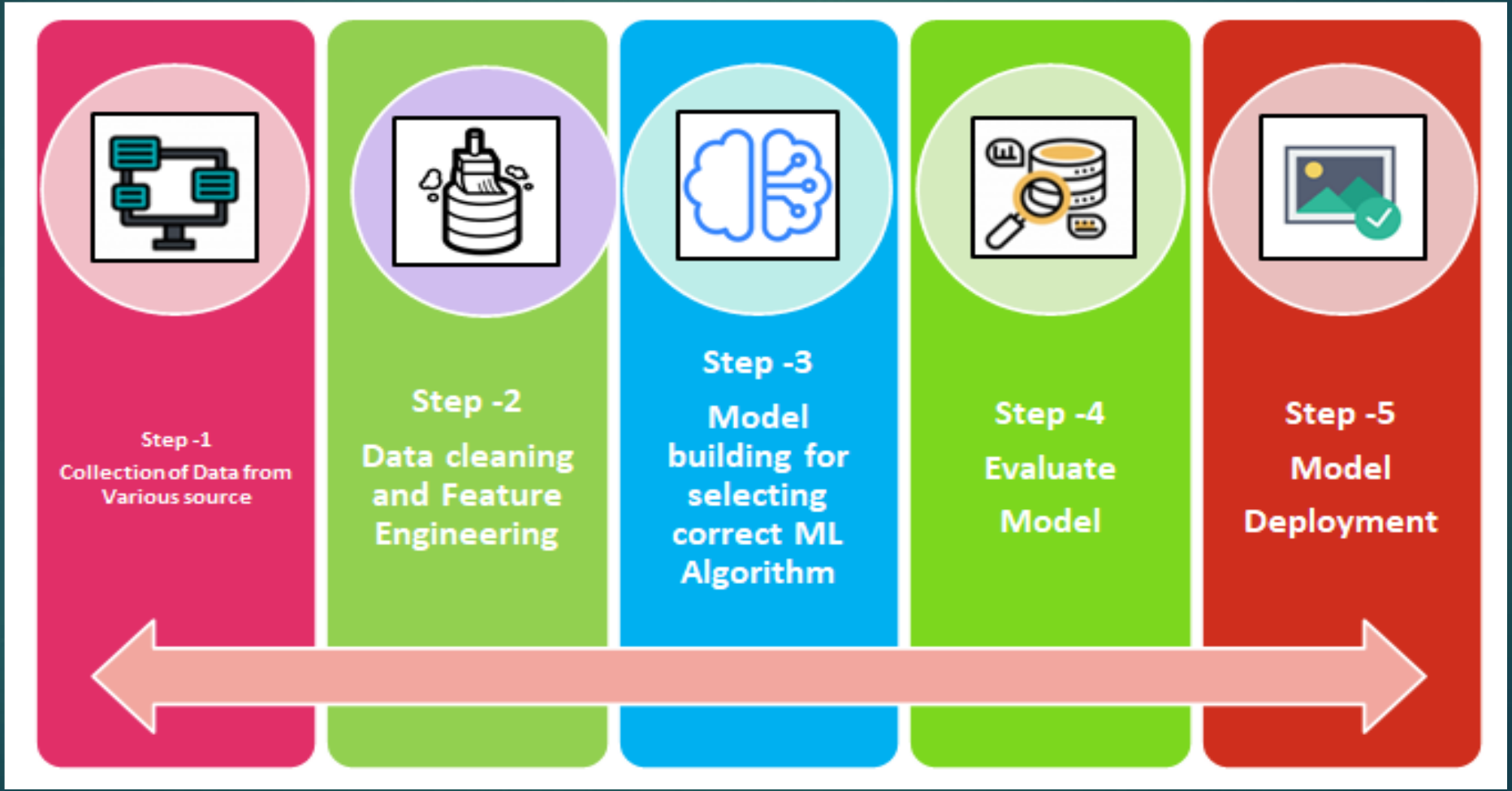
Solution Offered…

Selection of Hot Leads

Communicating with Hot Leads

Conversion of Hot Leads

Evecution Process..

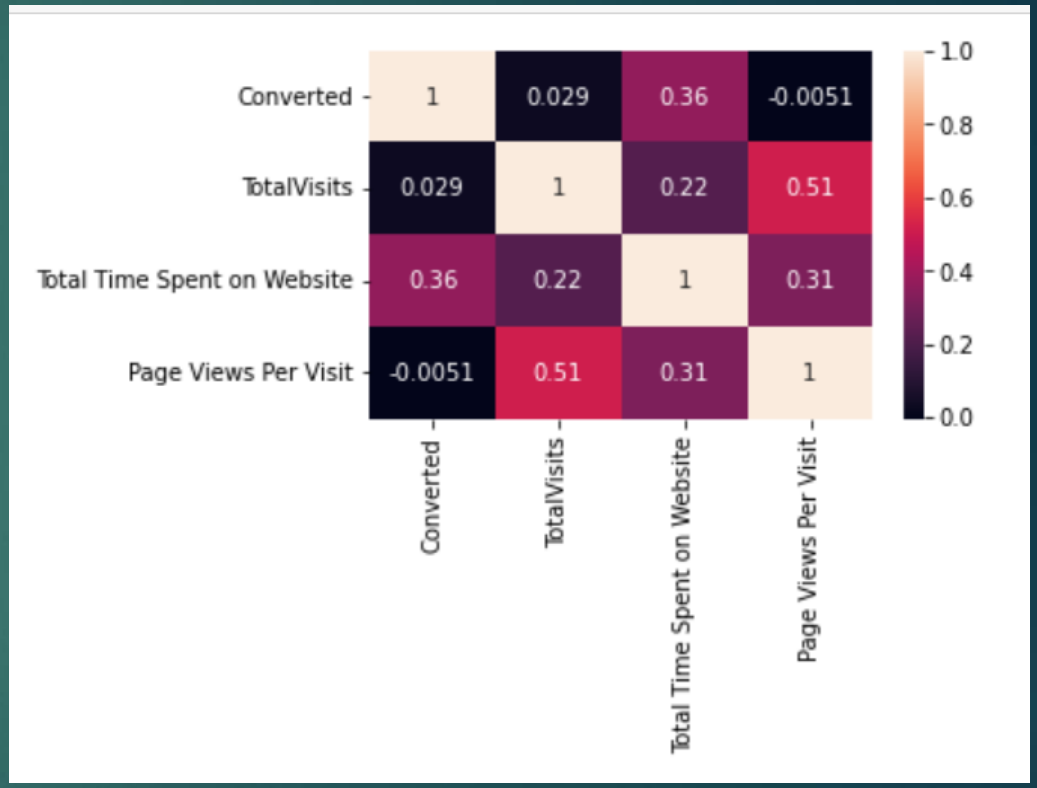


Data Visualization..

Observation:

1. **Total Visits and Page Views Per Visits have a strong association.**

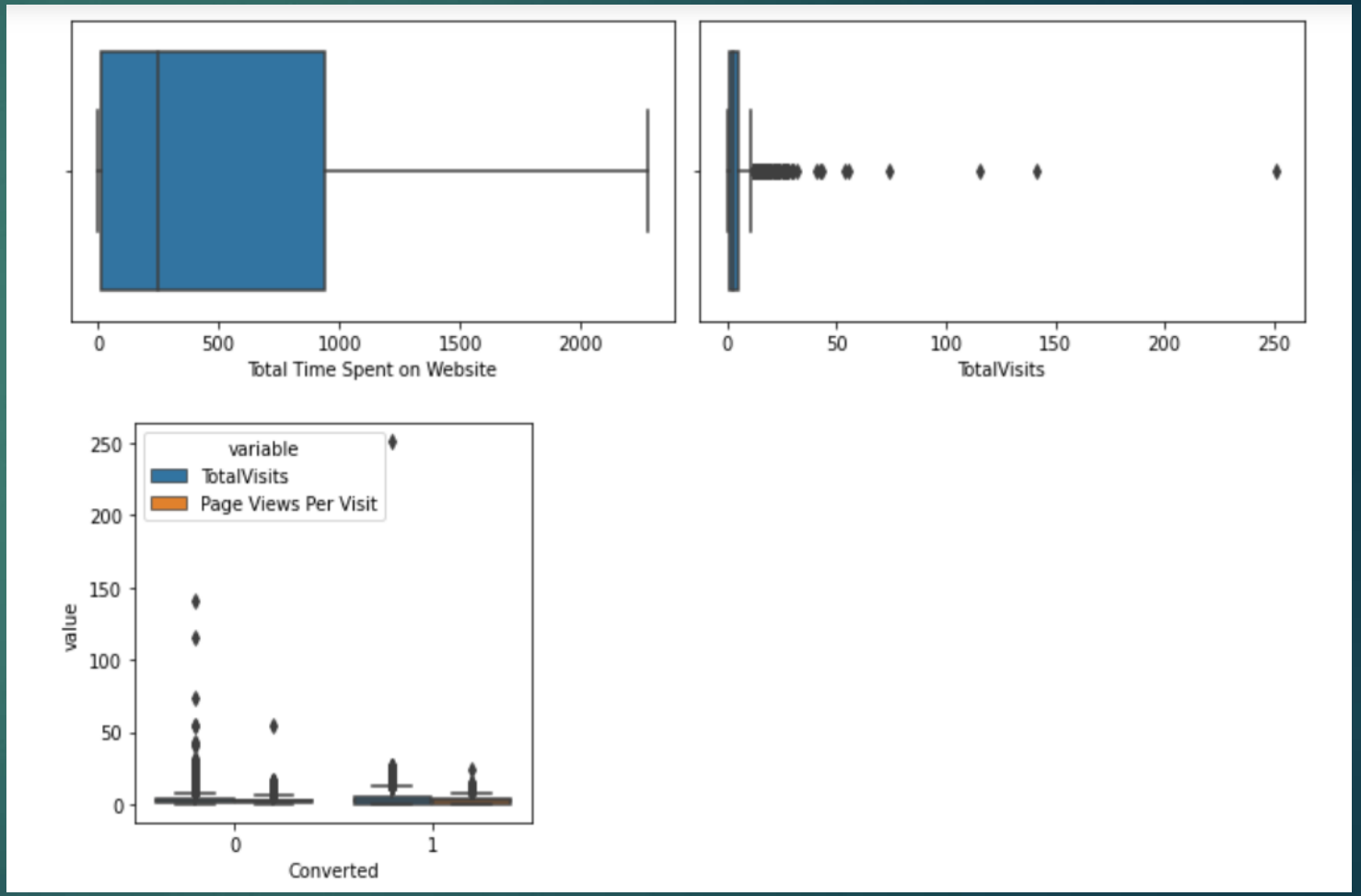
2. **Converted has good correlation with Total time spent on Website**



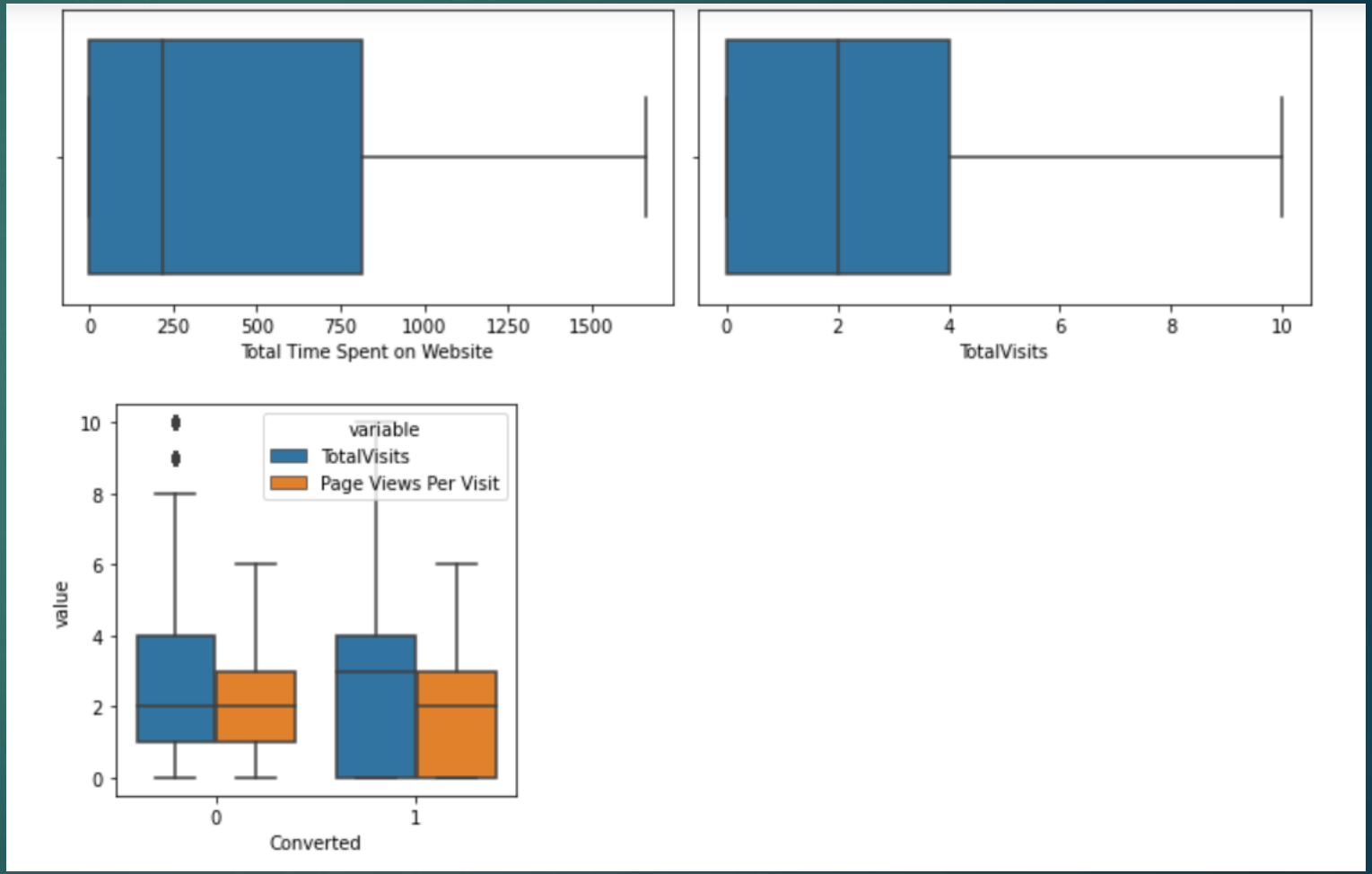
Observations:

There are upper limit outliers in both the total visits and page views per visit columns, as shown in both the box plots.

So we must remove outliers.



Observation after removing outliers…



Observations:

1.Lead Origin Graph –

a.API and Landing Page Submission have around 40-50% conversion rate

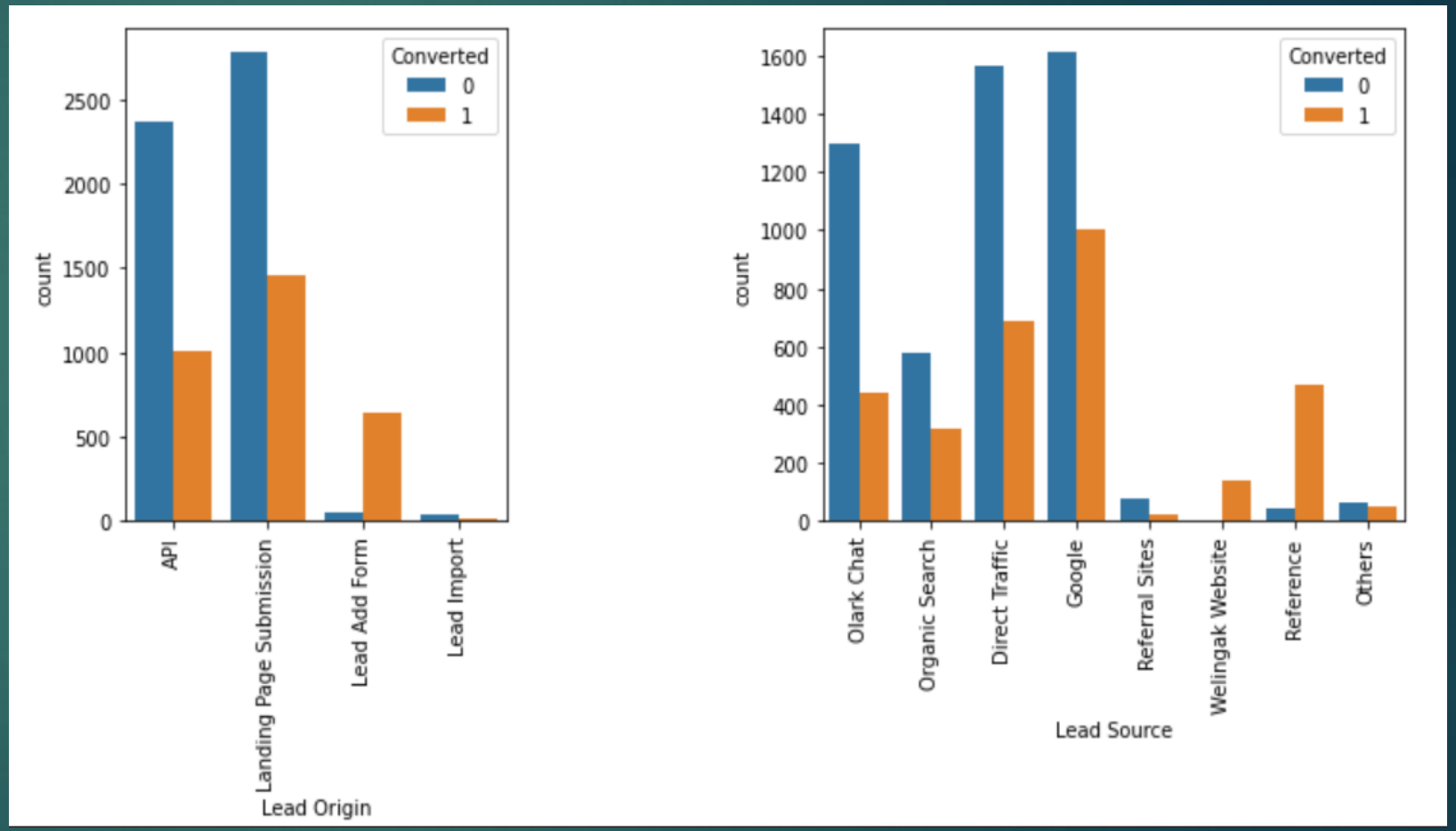
b.Lead Add Form has more than 90% conversion rate but count of lead are not very high.

c.Lead Import are very less in count.

2. Lead Source Graph –

a.Google and Direct traffic generate high number of leads.

b.References and wellingak website has high conversion rate. Should be more focused on.



Observations:

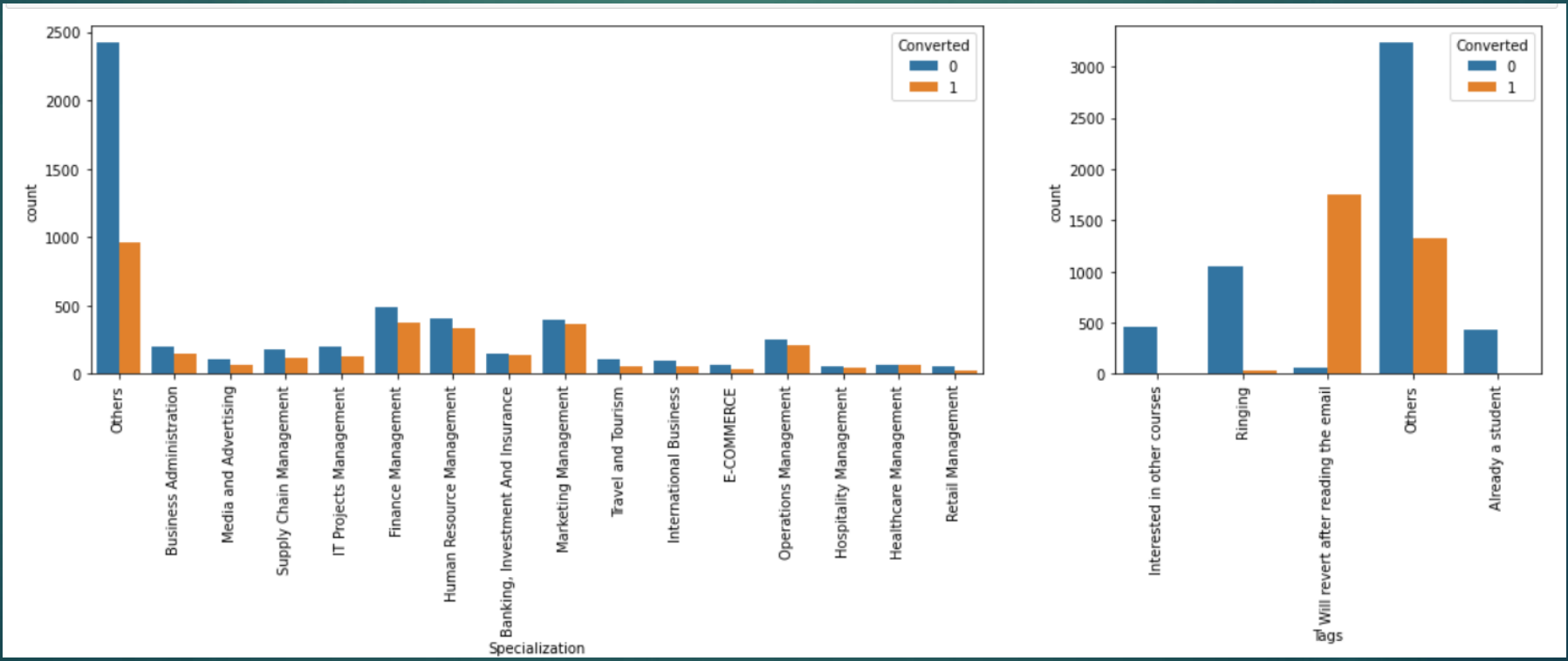
1.Specialization Graph –

a. The emphasis should be on diverse specialities with low lead generation but high conversion rates.

b. The conversion rate from other sources is below average.

2. Tag Graph –

a.The conversion rate is high for revert after reading mail .Which is justify.



Model Building…

Splitting the data into Training & Testing , we have choosen 70:30 ratio

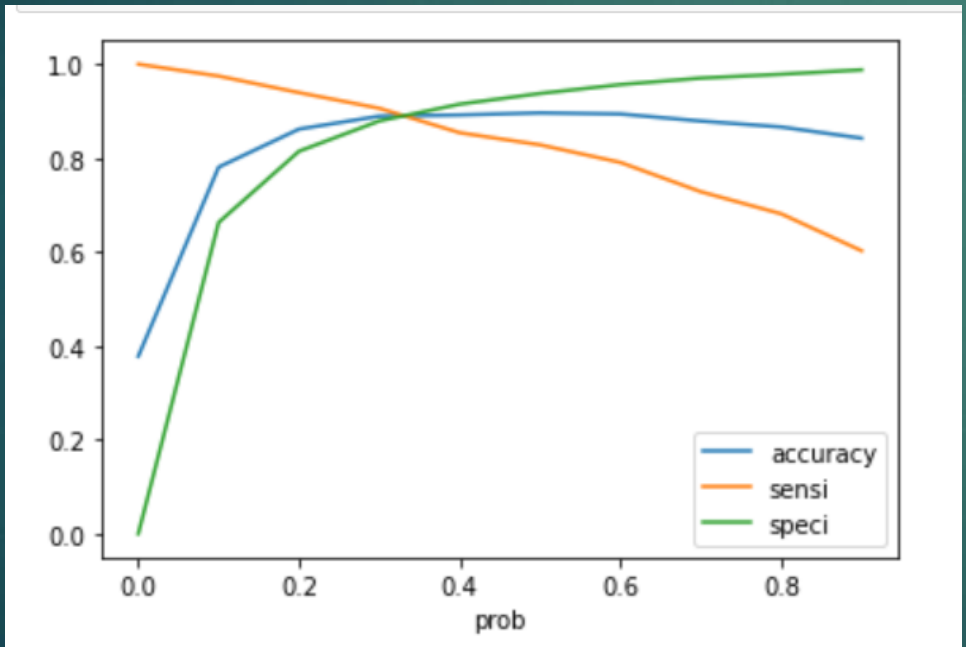
Use RFE feature for Feature Selection & run RFE with 12 variables as output..

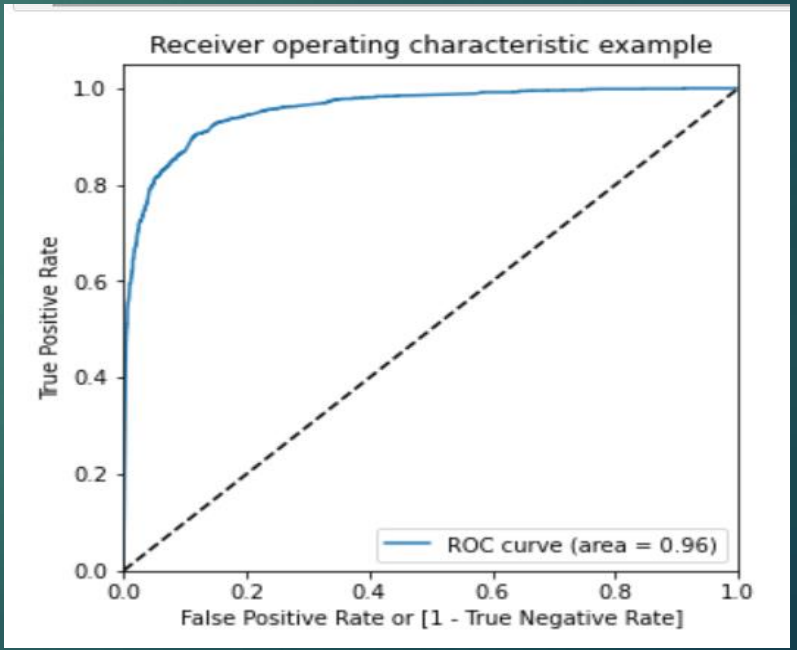
Removing the variable whose p-value is greater than 0.05 and VIF value is greater than 5

Predictions on test dataset with overall accuracy above 87%

ROC & Optimal Cutoff Point –

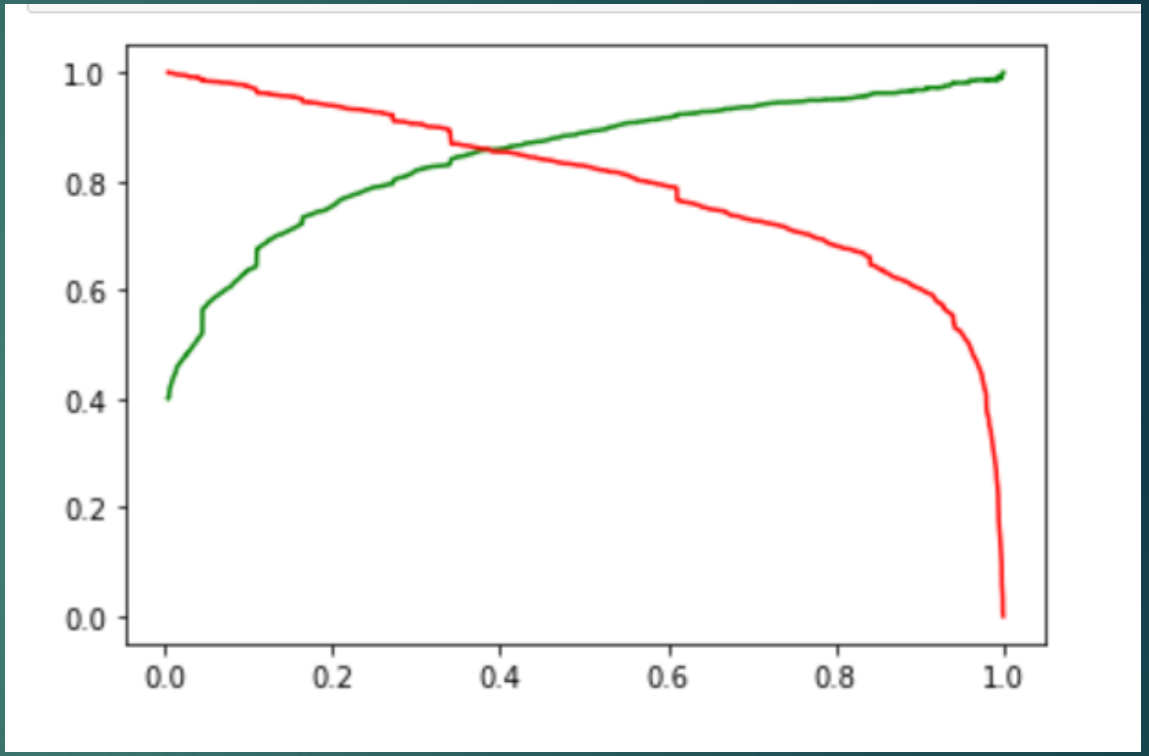
1. The 0.34 is out-cutoff
2. Accuracy doesn't change after 0.3





Precision Recall –

Cutoff is 0.39



Conclusion -

- That is, there is above 80% chance that our predicted leads will be converted
- When compared to the model we derived, our Logistic Regression Model is decent and accurate enough with approx 88% Accuracy on Test Set, 85 % Sensitivity and 90 % Specificity
- We can adjust these parameters by changing the cut-off value, and predict leads.
- Increasing total visits through advertising, etc. will result in a higher conversion rate
- In order to increase conversions, websites should increase engagement with users

Thank You…