Lead Scoring Case Study

By-
KAMNA SETHI
PRANAV SHETTY
SHASANK
(FROM- DS C56)

Problem Statement

- An education company named X Education sells online Courses to industry professionals. The company markets its courses on several websites like Google. Once people land on their website or fill up a form for the course. When people fill up the form by providing their email address or phone number they are classified as lead. On acquiring the leads sales team start making calls, writing emails. Through this process some of leads get converted while most do not.

- The typical lead conversion rate is around 30%. This conversion rate is very poor. To make the process more efficient, company wishes to identify the most potential leads known as 'Hot leads'. On successfully identifying this set of leads, the lead conversion rate would increase as sales team would focus more on communicating with the potential leads rather than contacting everyone.

Business Goal

- To build a Logistic Regression model to assign a lead score between 0 and 100 to each of the leads, this lead score is used by the company to target potential leads.

- A higher score would mean that the lead is hot, i.e. more likely to convert whereas a lower score would mean that lead is cold and will mostly not get converted.

- The CEO wants the target lead conversion rate to be around 80%.

Strategy

The following steps were followed in building the model –

- Reading and Understanding data
- Data Cleaning
- Visualization of data
- Data Preparation
- Train-Test split
- Feature Scaling
- Model Building
- Model Evaluation on Train set
- Evaluate model for accuracy and metrics on train set
- Model Evaluation on Test set
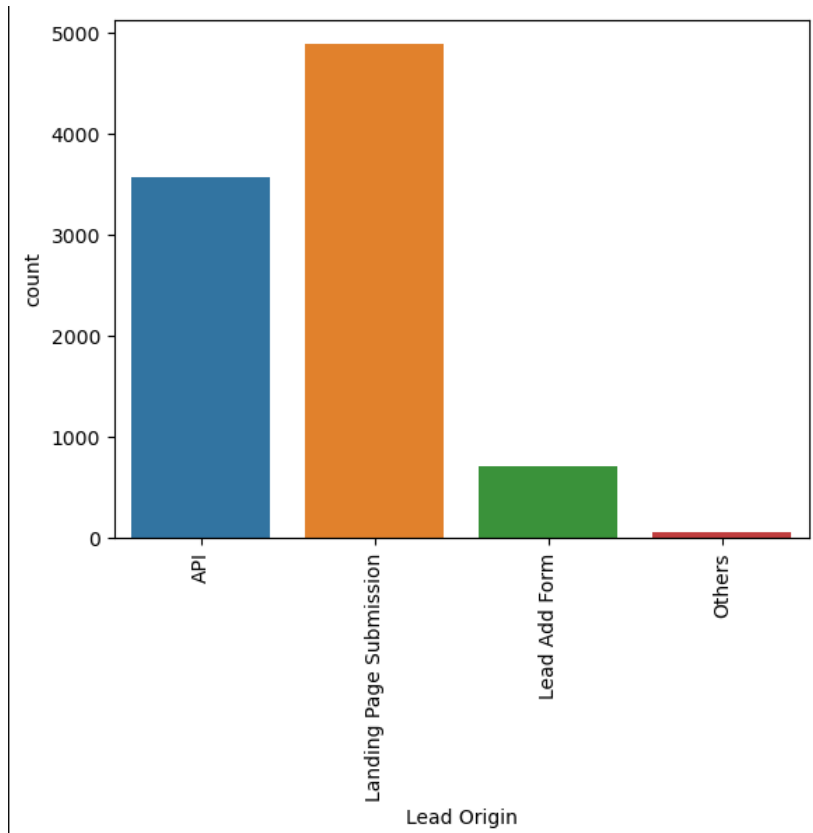- Evaluate model for accuracy and metrics on test set
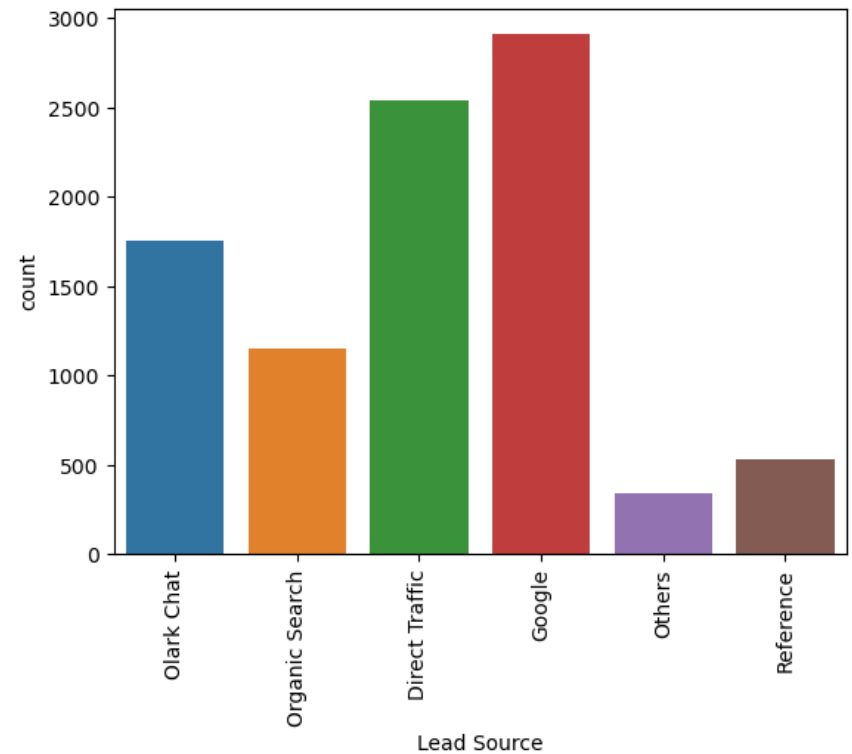
Pie chart for Converted column

Non-Converted

61.46

38.54

Converted
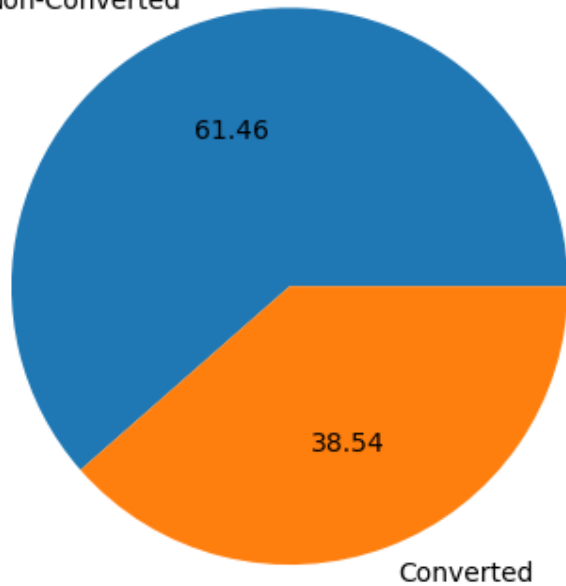
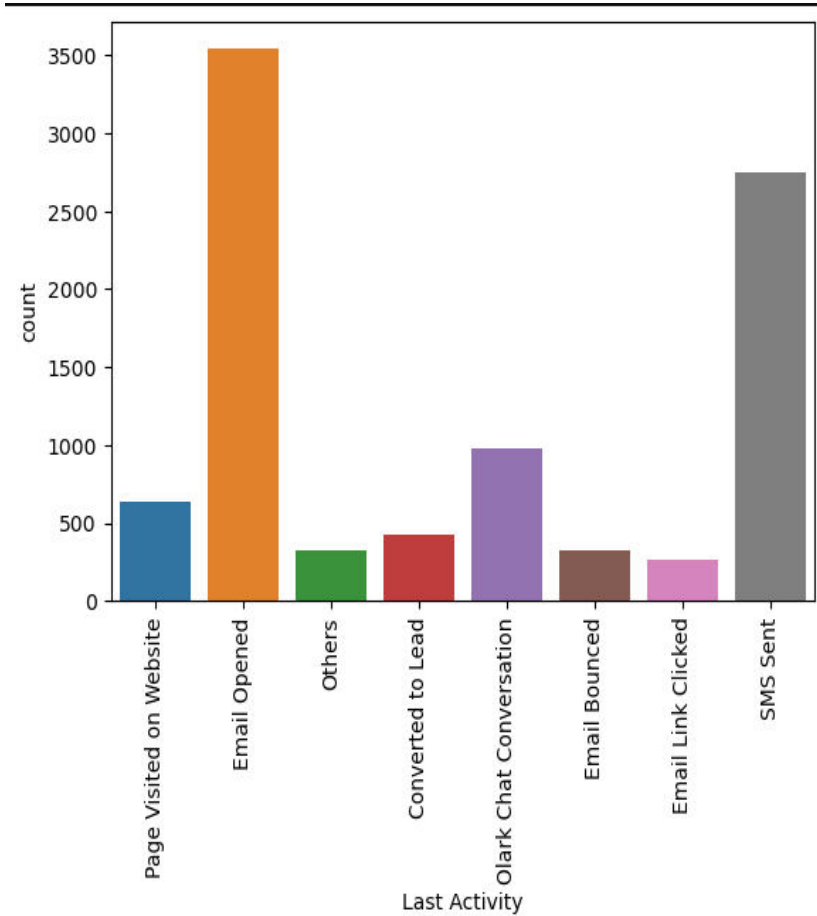Rate of conversion is quite low as 38.54%

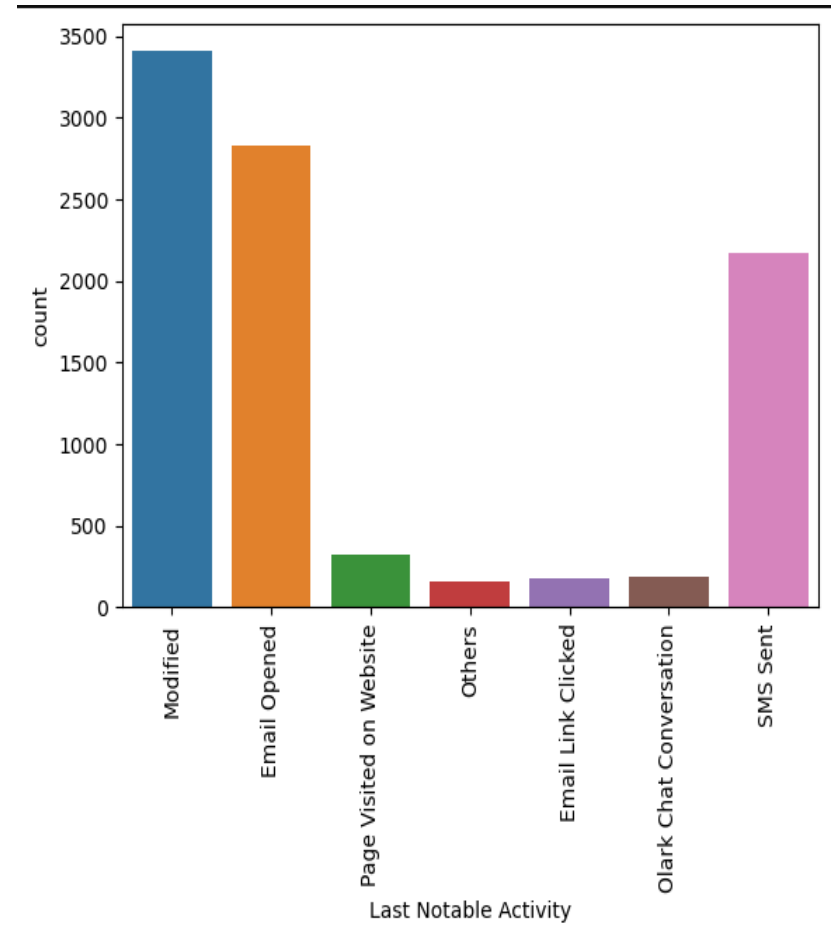**Lead Origin as Landing Page submission has the highest count**

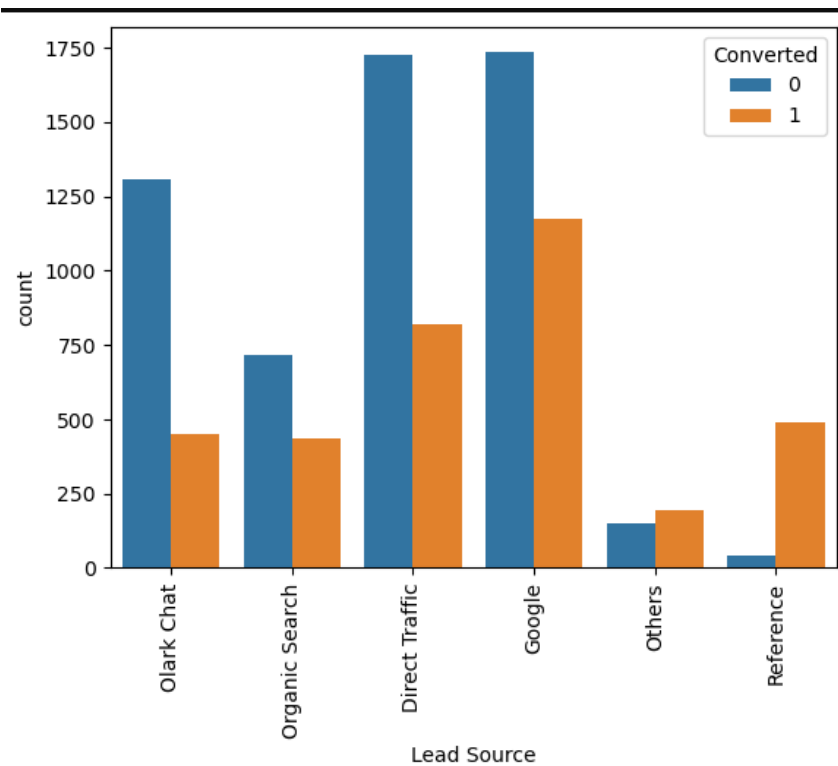**Lead source Google followed by Direct Traffic has maximum count**

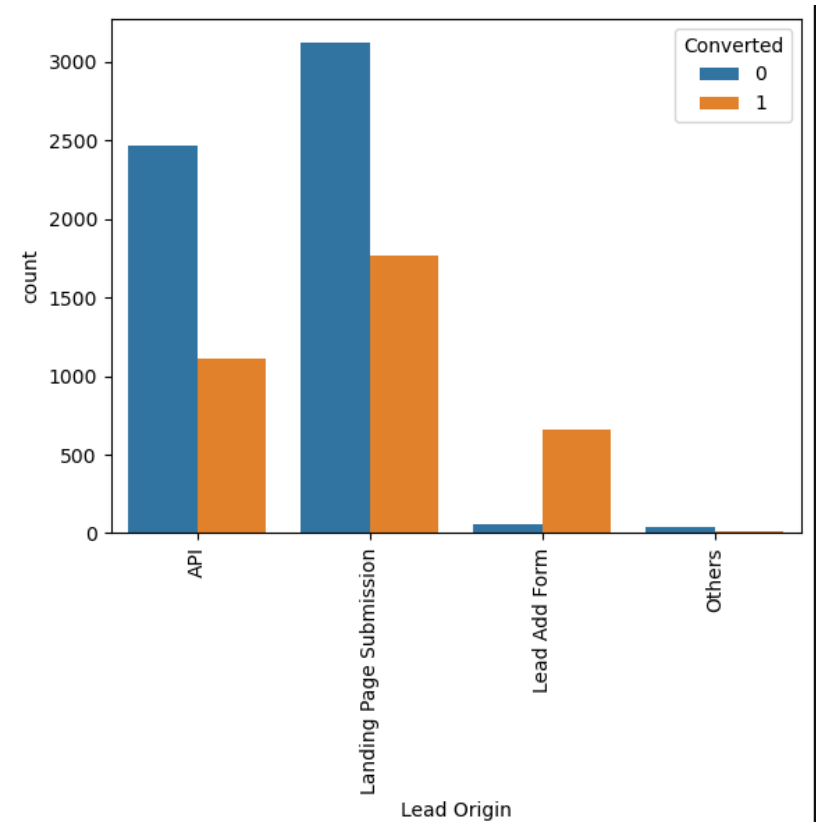**Last Activity : Email Opened followed by SMS Sent has maximum count**

Last Notable Activity : Modified followed by Email Opened has maximum count
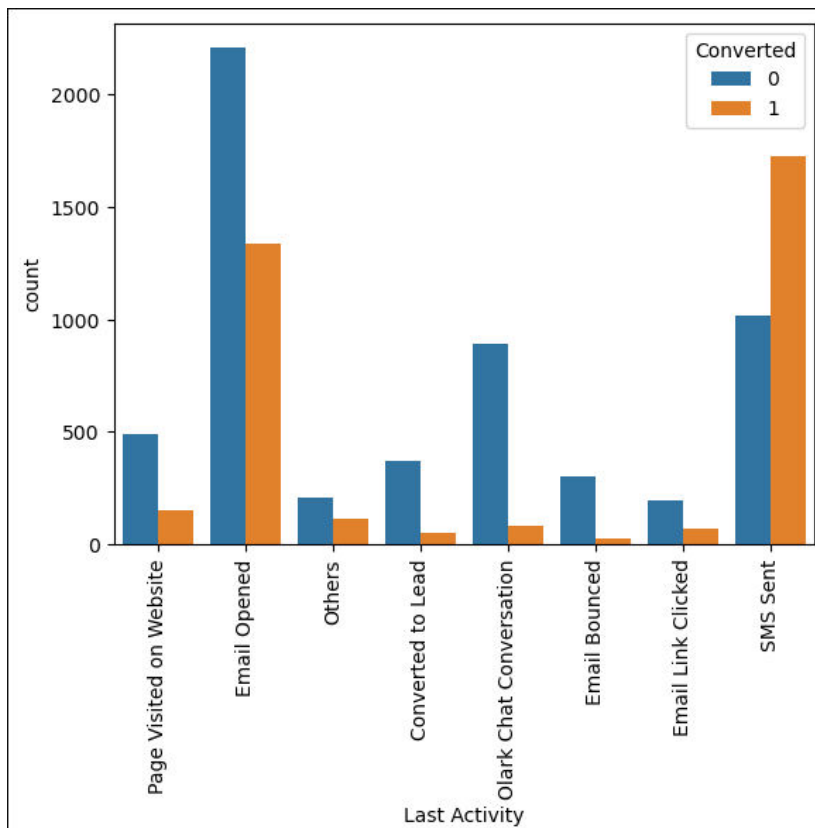
Lead Source as Google have higher conversion rate
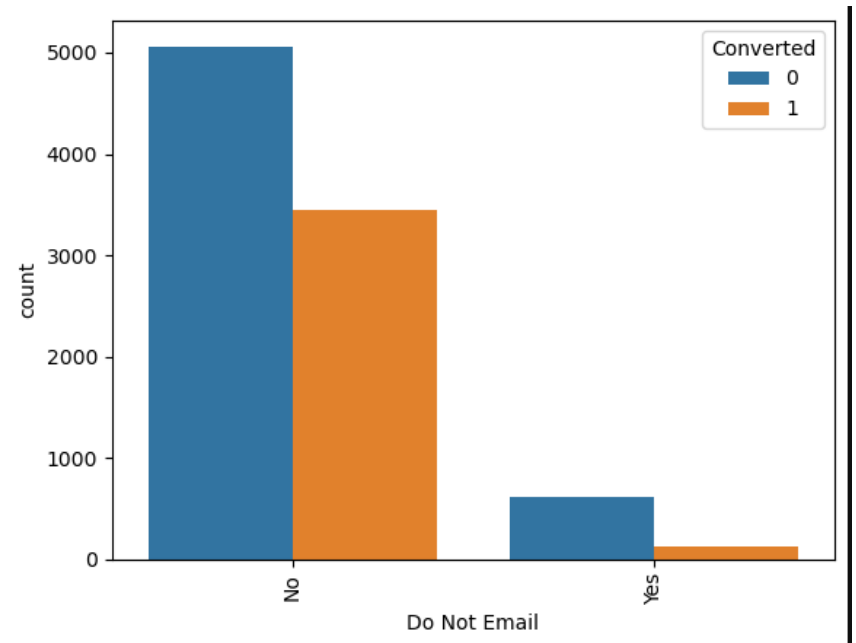

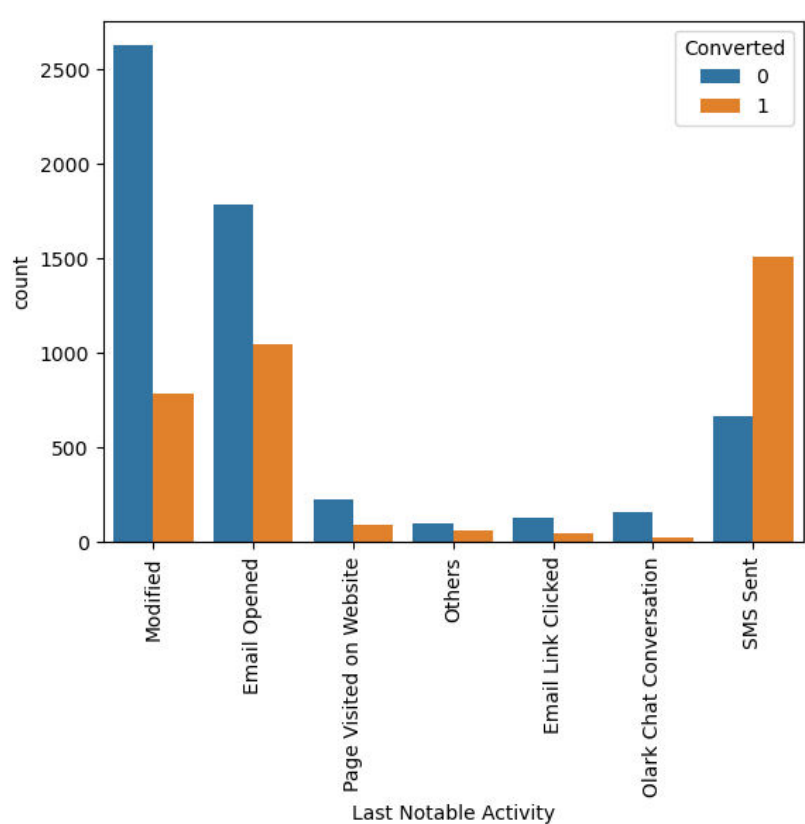# Lead origin as lead add form has higher conversion

Last Activity as SMS sent has high lead conversion

Do not email as No has high lead conversion

Last Activity as SMS Sent has high lead conversion

Current Occupation as working professional has high lead conversion
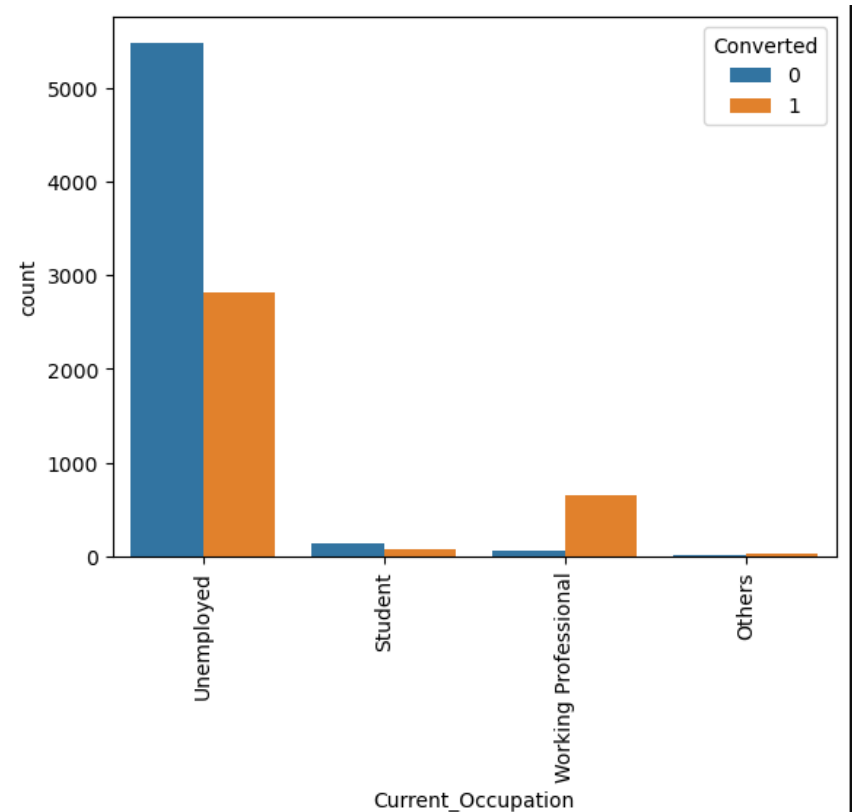
Plotting ROC Curve

• It shows the trade off between sensitivity and specificity

•The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.

•The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

Receiver operating characteristic example

Model evaluation on Train set

Accuracy = 80.61%
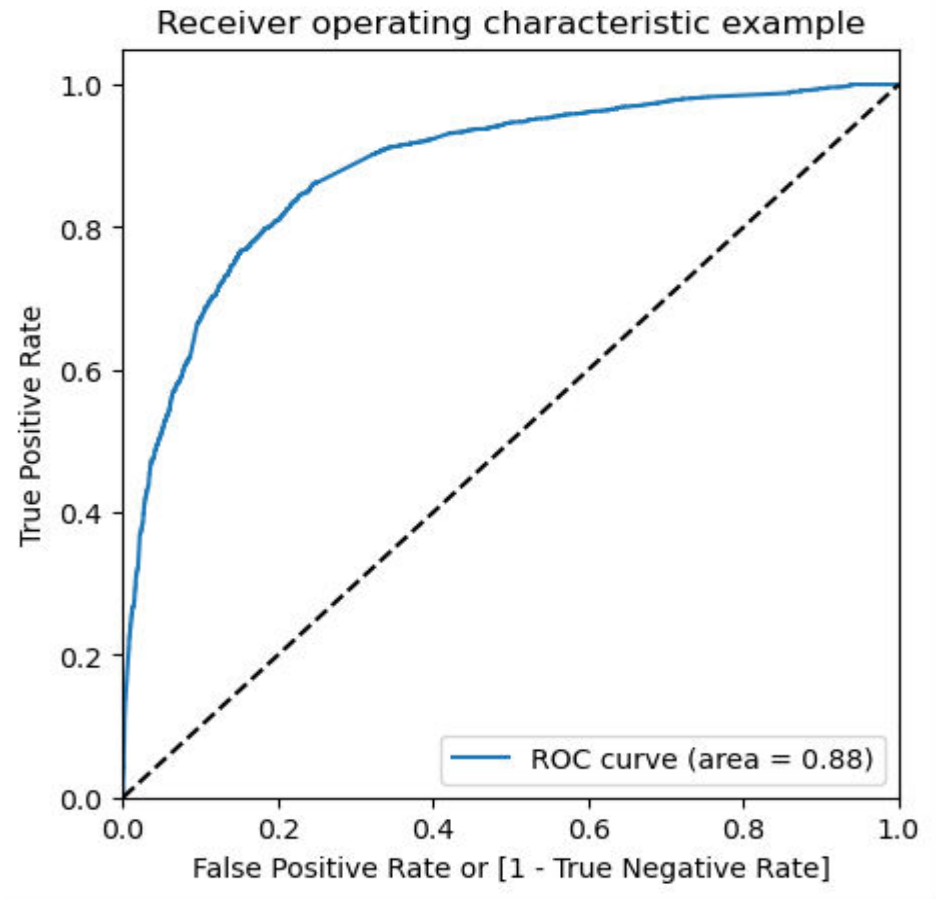
Sensitivity or Recall = 80.54%

Specificity = 80.66%

In sensitivity-specificity with cut-off threshold of 0.345 we are getting 80% for the metrices.

So, we will go with sensitivity-specificity view for optimal cut-off for final predictions.
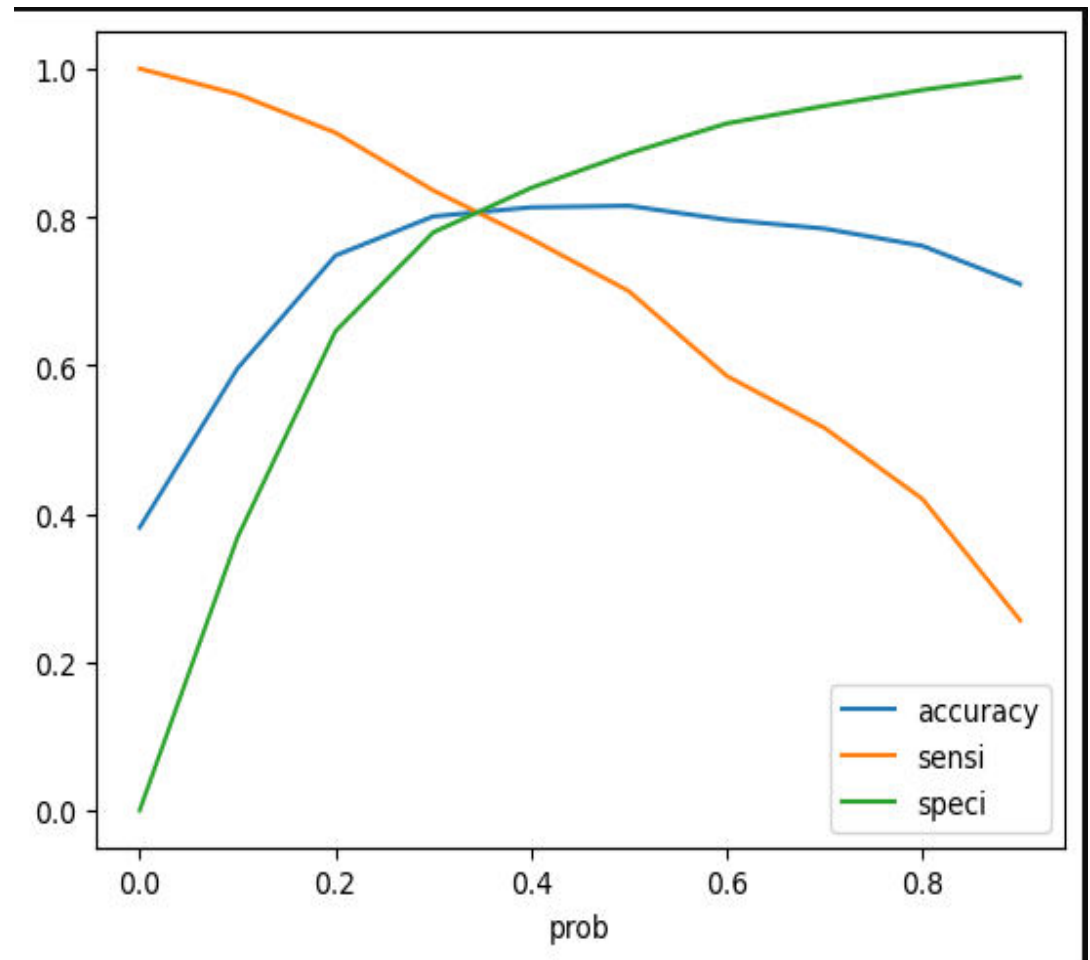
Precision and Recall Trade off

Accuracy = 81.66%

Sensitivity or Recall = 76%

Specificity = 85.15%

As we can see in above metrics, where we used precision-recall threshold cut-off of 0.42 the value of sensitivity or recall has dropped to around 76%. But we want it to be close to 80% according to Business objective.

For train data

- Accuracy = 80.61%

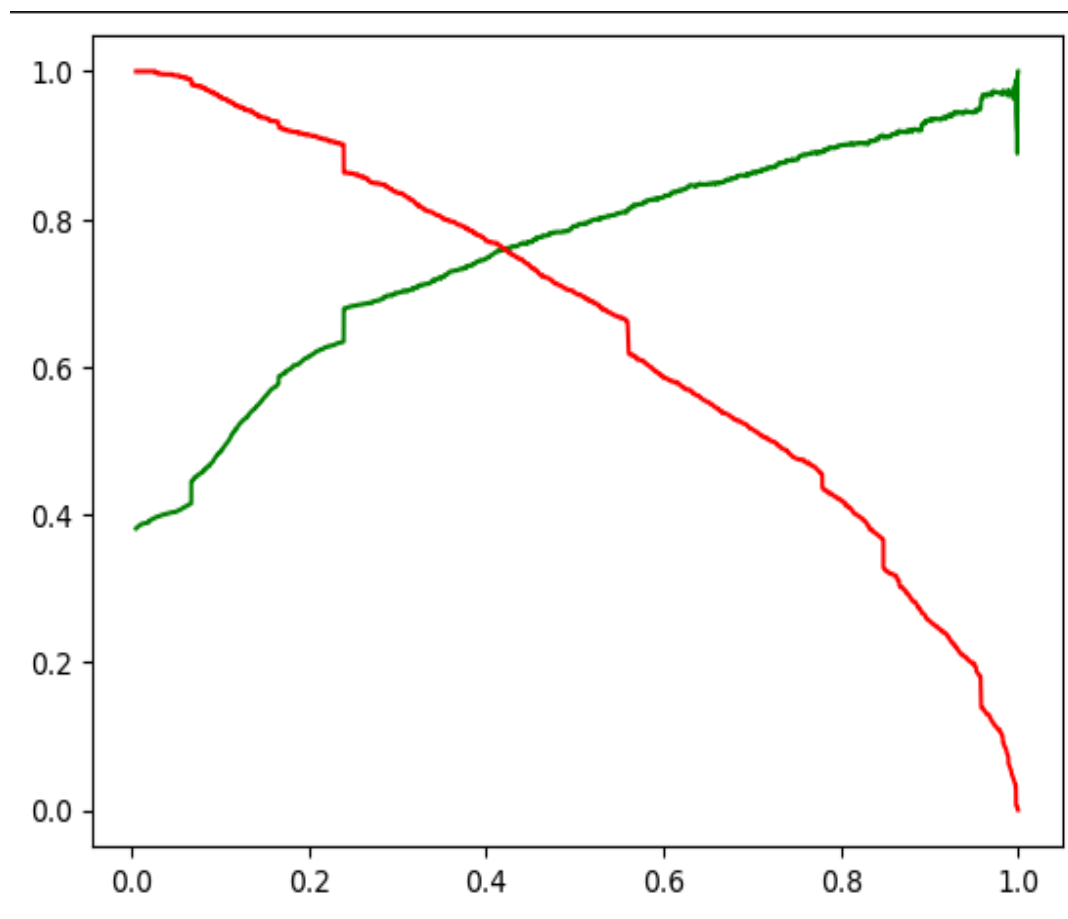- Sensitivity or recall = 80.54%

- Specificity = 80.66%

For test data

- Accuracy = 80.63%

- Sensitivity or recall = 80 %

- Specificity = 81.04%

- **The evaluation metrics are pretty close to each other so it indicates that the model is performing consistently across different evaluation metrics in both train and test dataset.**

Observations

Variables which contribute most towards the probability of a lead getting converted are –

- Total Time spent on website
- Current Occupation-Working Professional
- Lead Origin-Lead Add Form
- Last notable activity as SMS sent and others
- Lead Source – Olark Chat
- Total Visits

Recommendations

- Target lead on the basis of Total time spent on Website, as visitors who are spending more time on website are more interested in the course that's why they are exploring different options available on website about the course and its details. So sales team should focus on such leads as they are more chances of conversion into paid customer for the course.

- Sales team should contact lead whose current occupation is working professional. Since this course is specially designed for working professionals and there might be some professionals looking for such course to upgrade their skills or profile. Also this course suits best for such working professional as being an online course they can easily manage their work along with the online course.

- Target lead where lead origin is lead add form, as Lead-add form help to generate lead by letting people submit their information in a form directly in ad. Those people who are submitting their information must be interested in the course, so more chances of lead getting converted into customer. So sales team should focus on such leads

- Sales team should also try to contact visitors on the basis of Total Visits. As more visits clearly indicates that the person is interested in the course that's why visits website several times for getting complete knowledge about the course. So, Sales team can help them by providing complete insights about the course which can help them to make a decision on joining the course.

- Sales team should not try to contact those leads who have opted for Do Not Email. As it clearly indicates that are not interested in course and does not want any type of information regarding the courses.

- Sales team should not try contacting unemployed lead and students as there might be very less chances of their conversion into an online professional courses. Reason can be budget issue or already doing some other course.

Thank You