LEAD SCORE CASE STUDY

Submitted by:
Arjun P
Rajat Joshi
Karishma Singh

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%.

X Education has appointed you 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%.

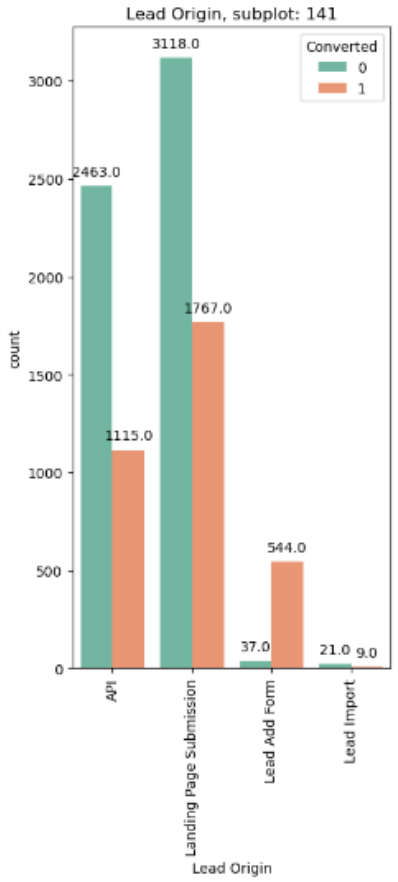
Steps to find the solution

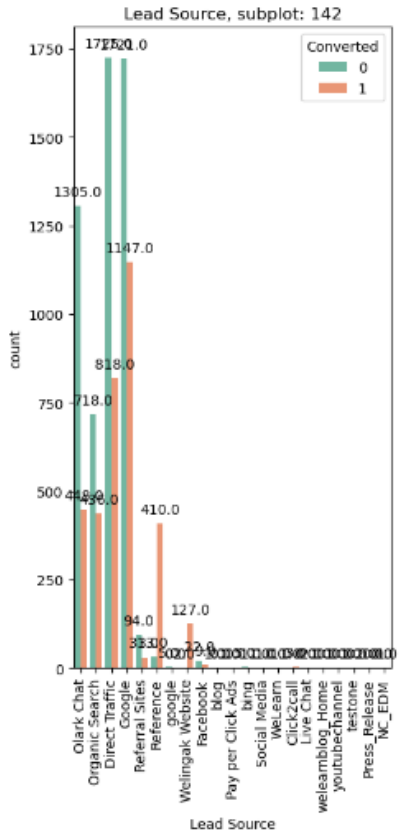
- Import & Analyze data
- Clean & prepare data for analysis
- Exploratory Data Analysis
- Train-Test splitting of data
- Scaling of data using Standard Scaler
- Feature selection using RFE
- Confusion matrix analysis
- ROC curve plotting
- Finding optimal cut-off point
- Analysis on train data set
- Prediction on test data set

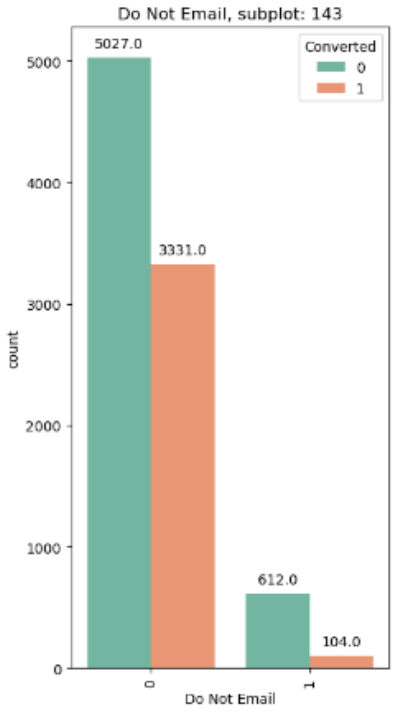
Analysis Approach

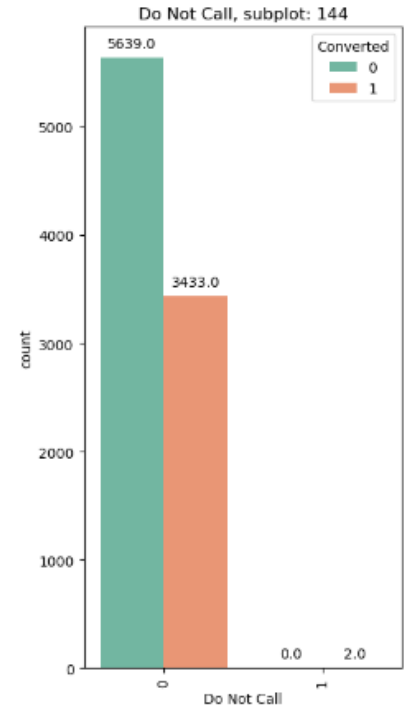
1. Data Importing, Cleaning, Preparing for EDA:
    - Reading and understanding data
    - Cleaning and imputing missing values
    - EDA
2. Train – Test Splitting, Feature scaling, RFE:
    - Splitting data into train – test data set
    - Scaling of data
    - Feature selection using RFE
3. Model Building:
    - Finding significant model using stats model
    - Calculate various metrics
4. Result:
    - Do prediction on test data set
    - Compare the values of Train – Test data set results
    - Evaluate the results.

EDA Findings

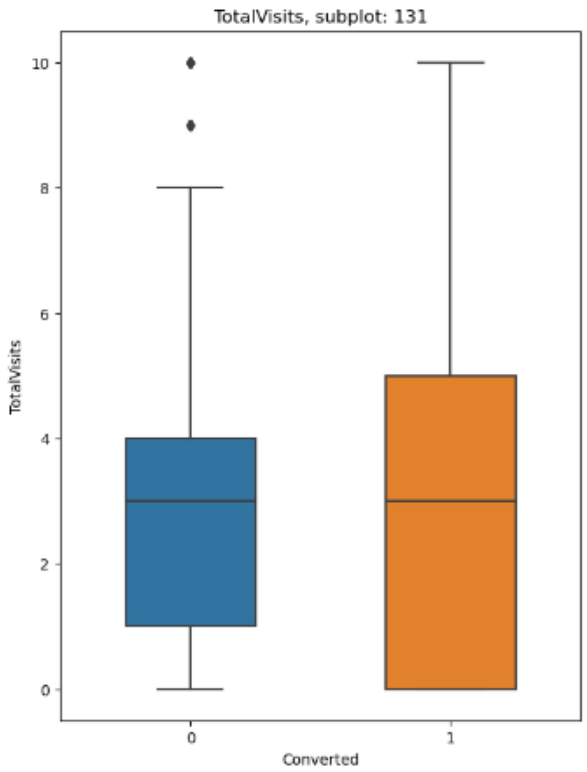


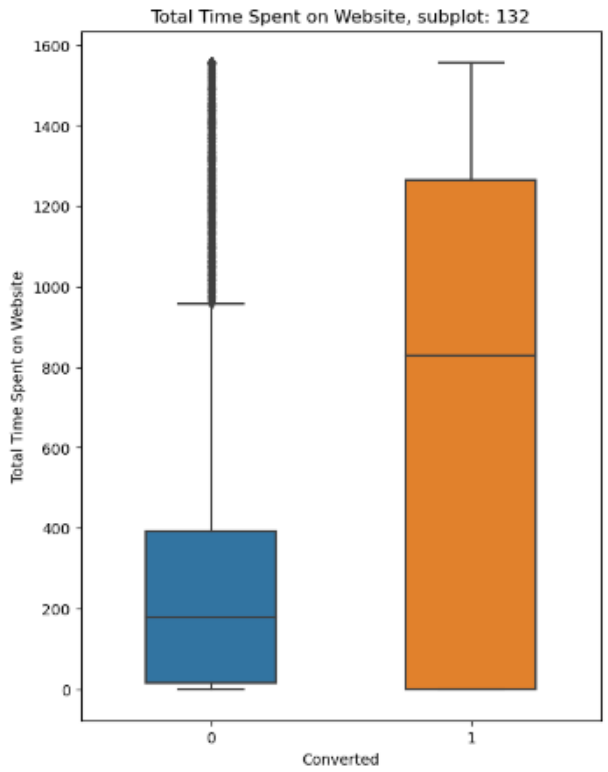


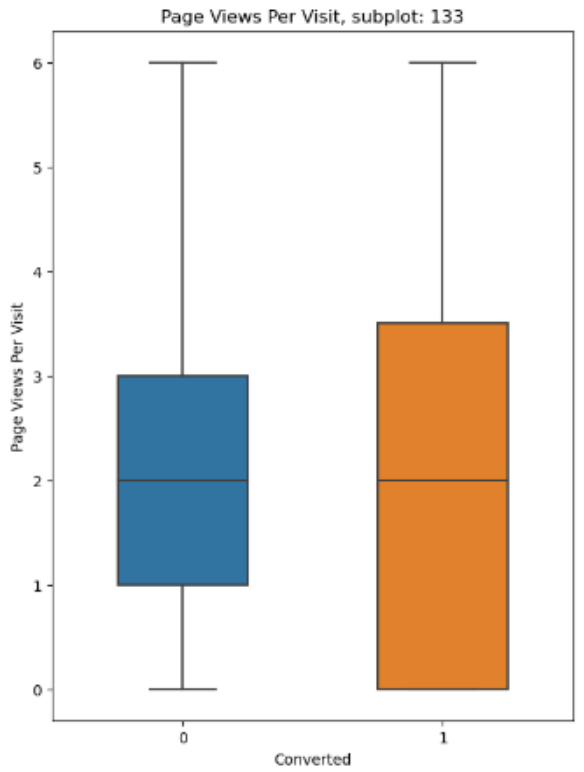




- API and Landing Page Submission have 45-55% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more conversion rate but count of lead are not very high.
- Lead Import are very less in count.
- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high compared to the number of leads generated.
- for Do Not Call & Do Not Email, most of the entries are no. nothing can be drawn from this

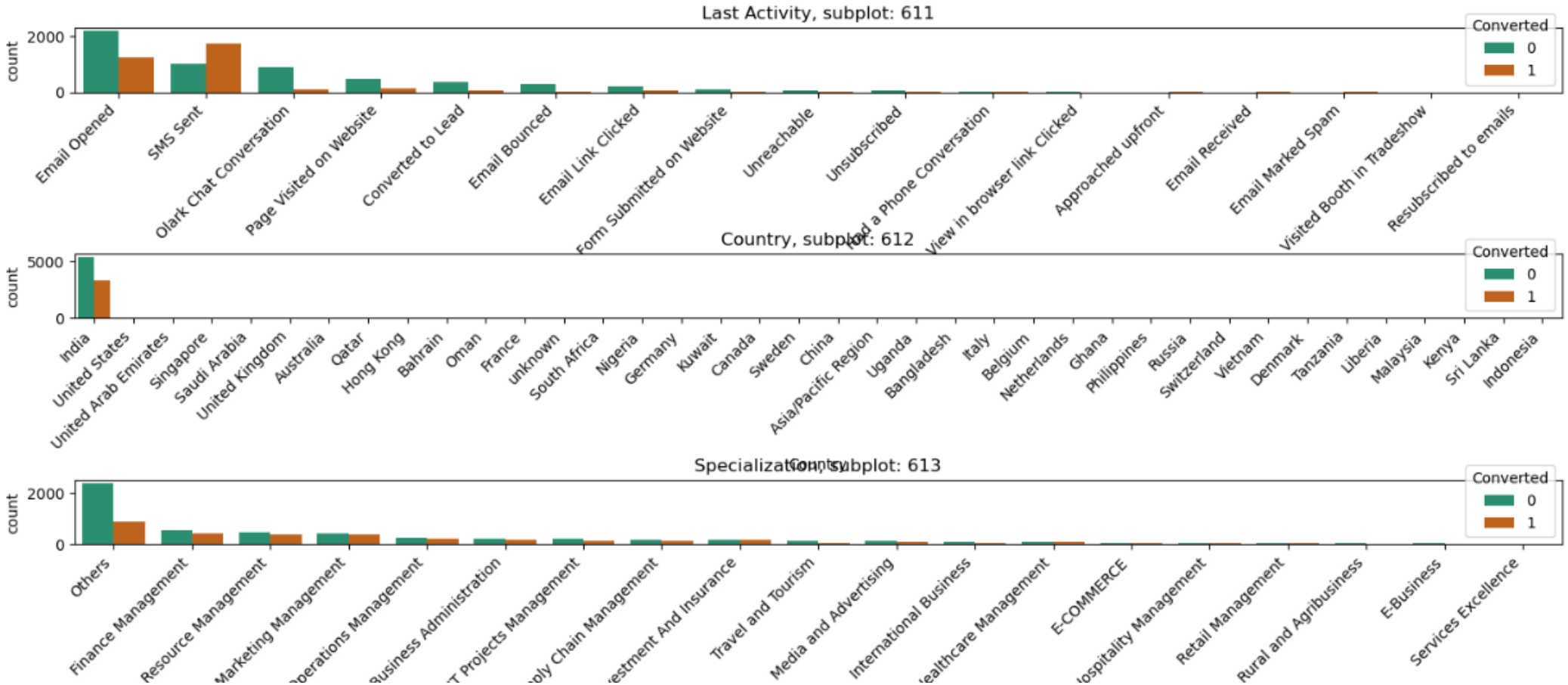


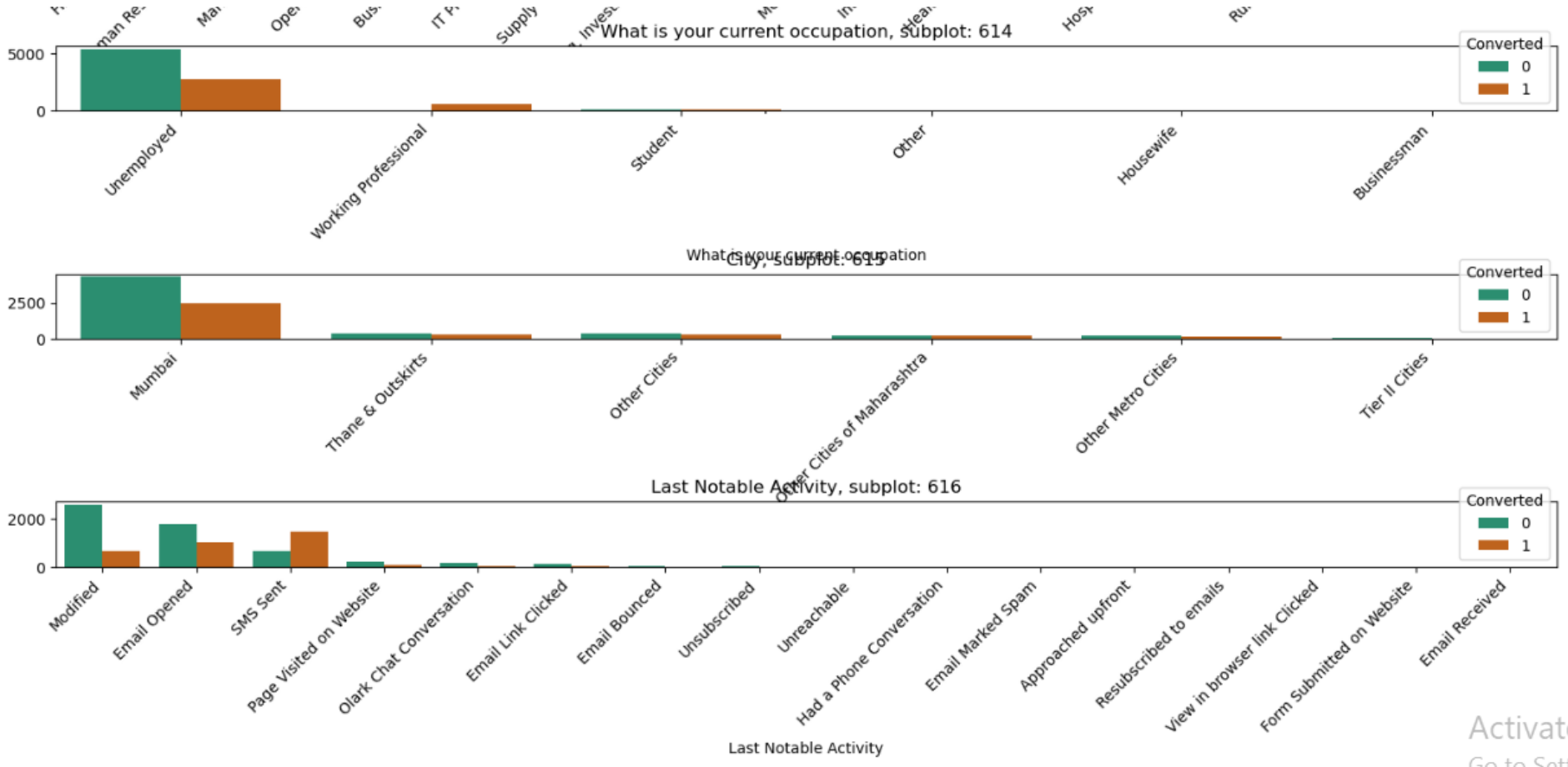




- For TotalVisits & Pages Views per visit column the Median for converted and not converted leads are the same.

  Leads spending more time on the weblise are more likely to be converted.


Last Activity, subplot: 611
count
2000
0
Converted
0
1
Email Opened
SMS Sent
Olark Chat Conversation
Page Visited on Website
Converted to Lead
Email Bounced
Email Link Clicked
Form Submitted on Website
Unreachable
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
Approached upfront
Email Received
Email Marked Spam
Visited Booth in Tradeshow
Resubscribed to emails
Country, subplot: 612
count
5000
0
Converted
0
1
India
United States
United Arab Emirates
Singapore
Saudi Arabia
United Kingdom
Australia
Qatar
Hong Kong
Bahrain
Oman
France
unknown
South Africa
Nigeria
Germany
Kuwait
Canada
Sweden
China
Asia/Pacific Region
Uganda
Bangladesh
Italy
Belgium
Netherlands
Ghana
Philippines
Russia
Switzerland
Vietnam
Denmark
Tanzania
Liberia
Malaysia
Kenya
Sri Lanka
Indonesia
Specialization, subplot: 613
count
2000
0
Converted
0
1
Others
Finance Management
Resource Management
Marketing Management
Business Administration
Travel and Tourism
Media and Advertising
International Business
E-COMMERCE
Retail Management
Rural and Agribusiness
E-Business
Services Excellence

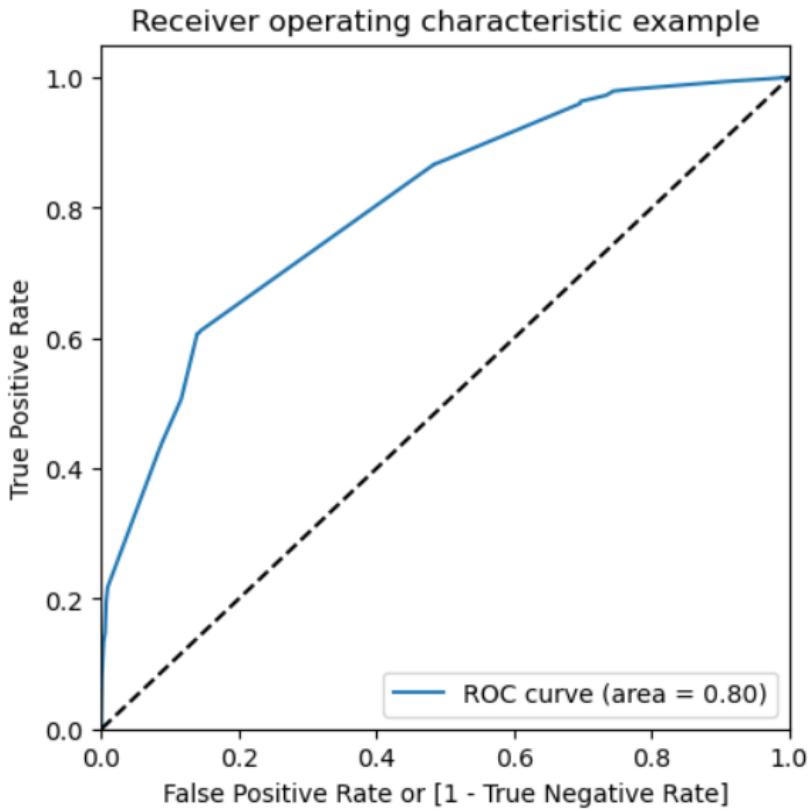

What is your current occupation, subplot: 614
5000
0
Converted
0
1
Unemployed
Working Professional
Student
Other
Housewife
Businessman
What is your current occupation
City, subplot: 615
2500
0
Converted
0
1
Mumbai
Thane & Outskirts
Other Cities
Other Cities of Maharashtra
Other Metro Cities
Tier II Cities
Last Notable Activity, subplot: 616
2000
0
Converted
0
1
Modified
Email Opened
SMS Sent
Page Visited on Website
Olark Chat Conversation
Email Link Clicked
Email Bounced
Unsubscribed
Unreachable
Had a Phone Conversation
Email Marked Spam
Approached upfront
Resubscribed to emails
View in browser link Clicked
Form Submitted on Website
Email Received
Last Notable Activity

- Most of the lead have their Email opened as their last activity.
- Conversion rate for leads with last activity as SMS Sent is more.
- Most of the leads who are joined in this course are Indians
- higest conversion rate is for people who are working
- About 50% of the unemployed people are converted
- People who sent SMS as their last activity has more conversion rate than the other activities

Features that make an impact on Lead Conversion

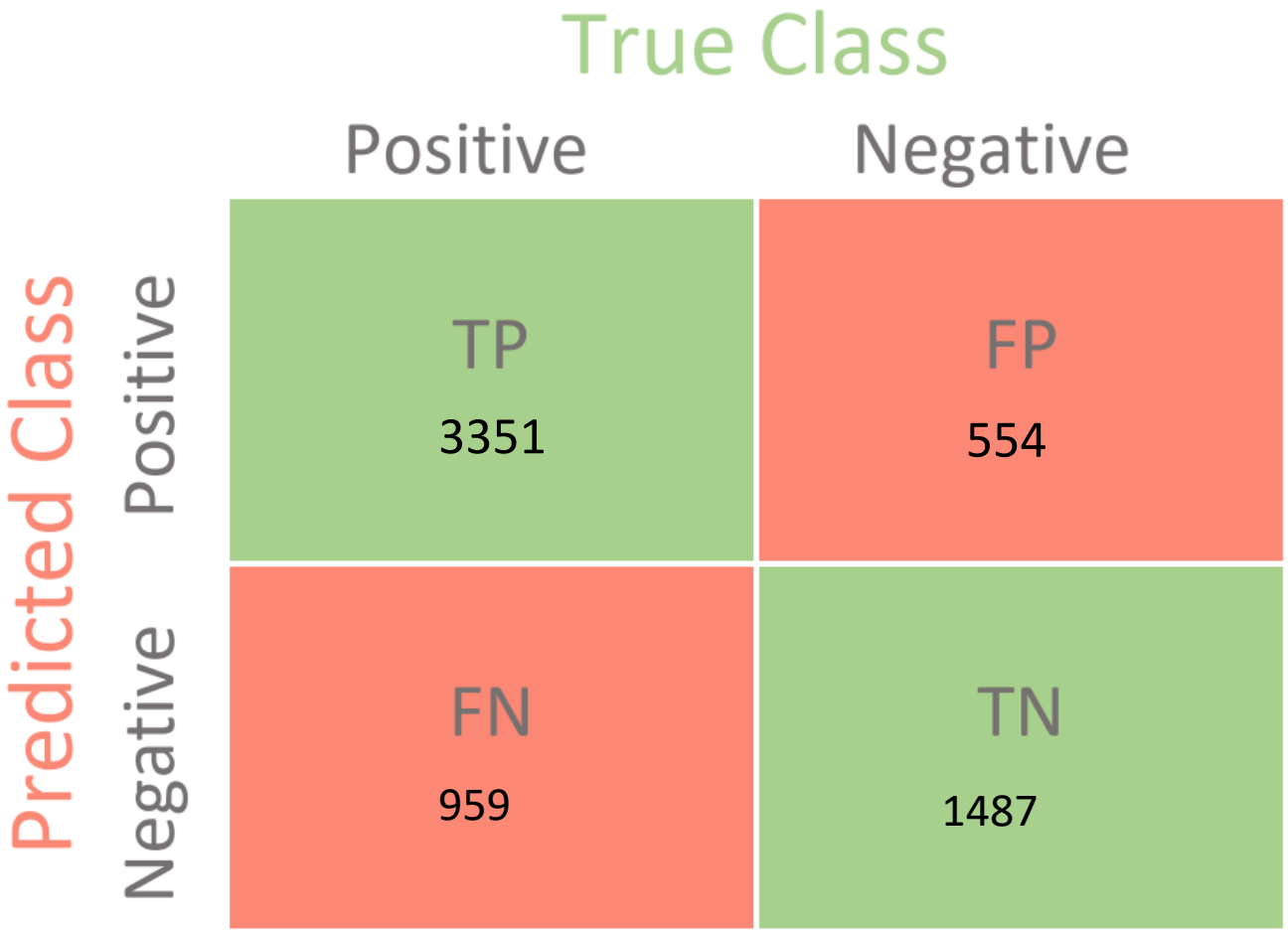
- **Do Not Email**
- **Lead Origin_Landing Page Submission**
- **Lead Origin_Lead Add Form**
- **Lead Source_Welingak Website**
- **Last Activity_Converted to Lead**
- **Last Activity_Had a Phone Conversation**
- **Last Activity_Olark Chat Conversation**
- **Specialization_Others**
- **Last Notable Activity_SMS Sent**
- **Last Notable Activity_Unreachable**
- **Last Notable Activity_Unsubscribed**

Model Evaluation

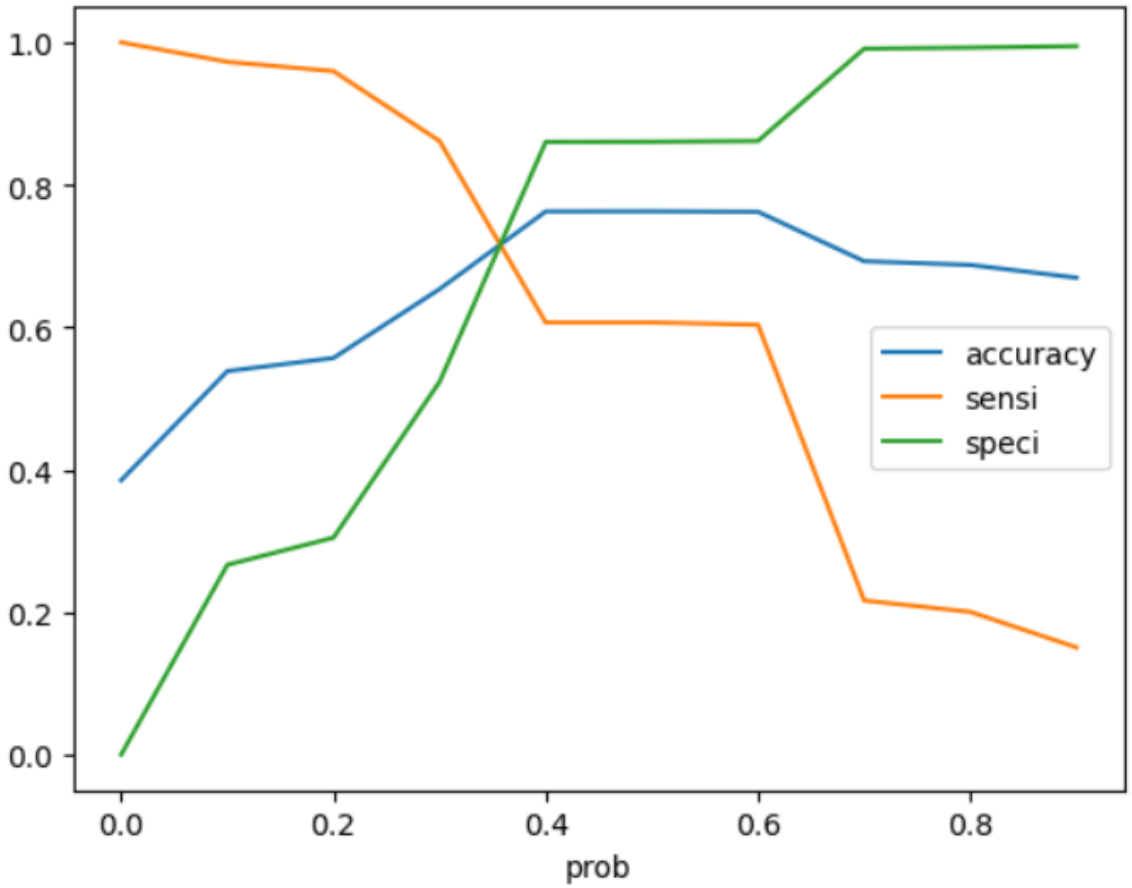


The area under the ROC Curve is higher (0.80). So the model is good one

Confusion matrix

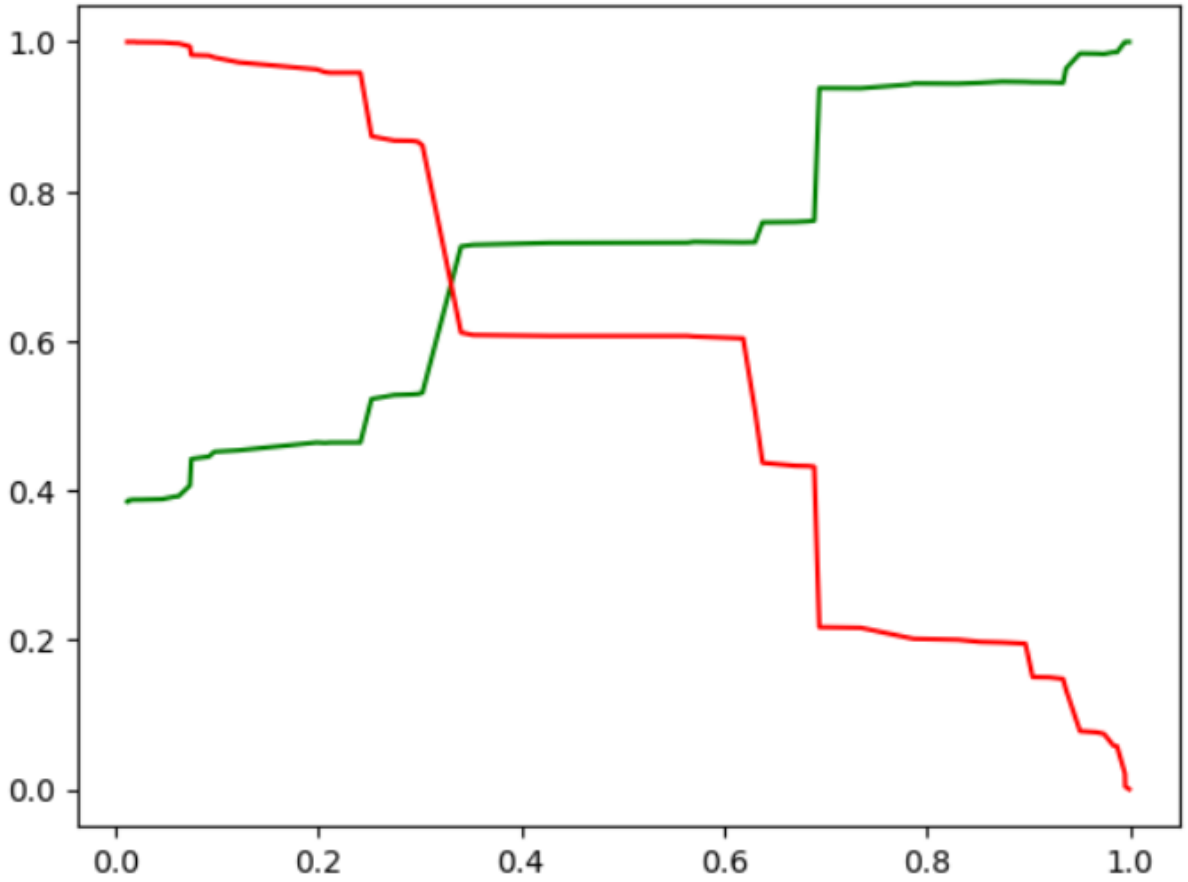


Train Data Set Values



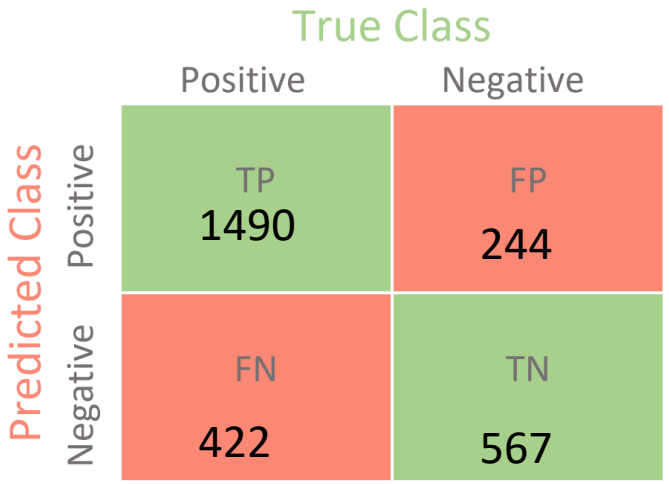
- Accuracy = 76.17%
- Sensitivity = 60.79%
- Specificity = 85.81%
- False Positive rate = 14.18%
- Positive Predictive Value = 72.85%
- Negative Predictive Value = 77.74%

Precision & Recall



- Precision = 73.14
- Recall = 60.67

Test Data Set Values



Accuracy : 75.5 %
Sensitivity : 57.3 %
Specificity : 86 %

Inference

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.
- Company should focus on working professionals
- Leads getting from Add form are more likely to converted
- The leads who had a good phone conversation with sales team are likely to join this course. The effective communication from sales team is important.
- Lead Origin_Landing Page Submission, Do Not Email, Last Activity_Converted to Lead, Specialization_Other, Last Activity_Olark Chat Conversation these leads are not likely to join this course.

THANK YOU