PRESENTATION ON LEAD SCORE CASE STUDY (LOGISTIC REGRESSION)

PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. 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.

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%.

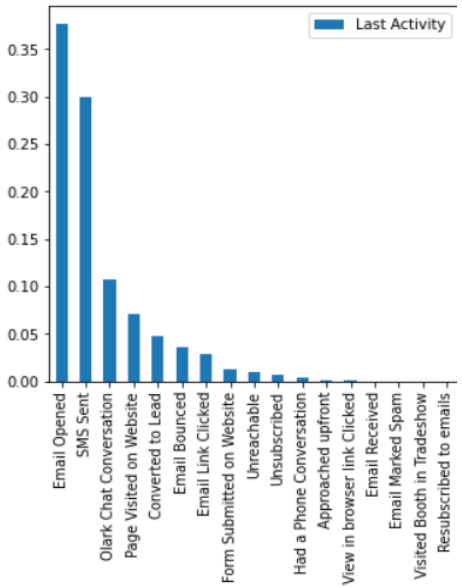
To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

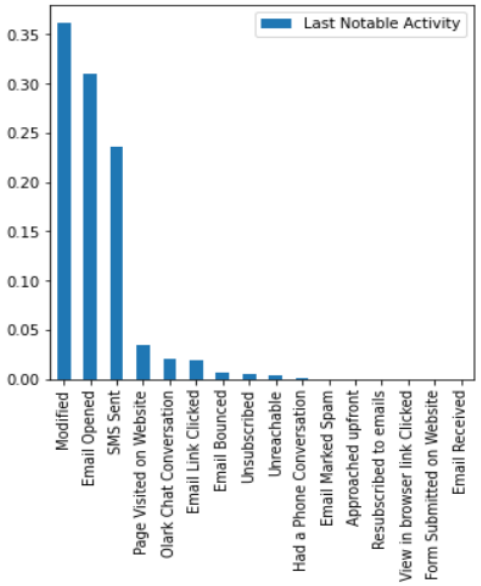
The company requires a model to be built wherein lead score has been assigned to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

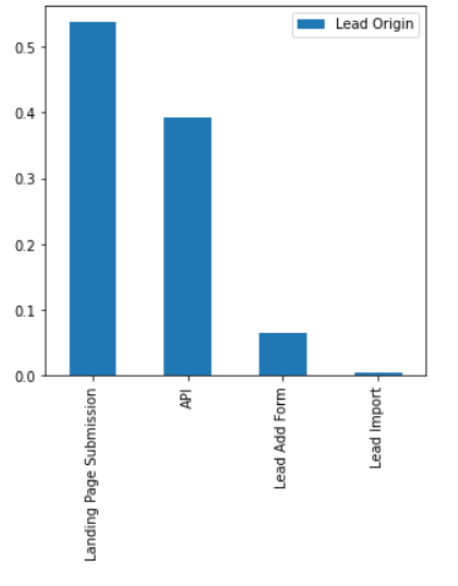
APPROACH

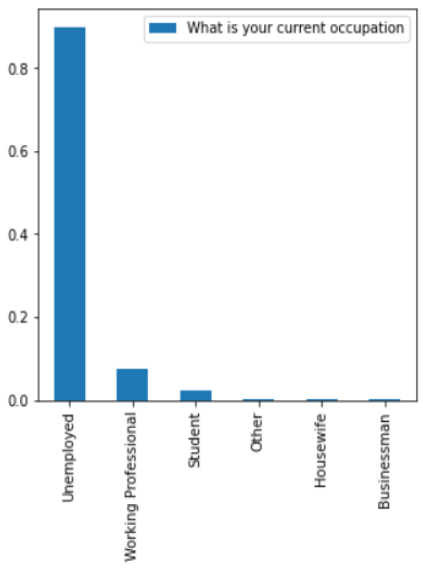
- **Data Exploration**
- **Data Cleaning**
  - **Missing values treatment**
- **Exploratory Data Analysis**
  - **Checking of Outliers**
  - **Checking of unwanted variables**
- **Data Preparation**
  - **Converting Binary variables**
  - **Creating Dummy variables**
  - **Dropping repeated variables**
- **Feature Engineering**
  - **Feature Scaling**
  - **Feature selection using RFE**
- **Model Building**
- **Model Evaluation**

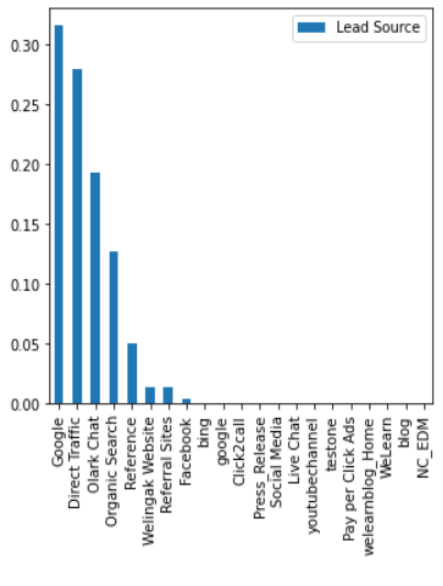
EXPLORATORY DATA ANALYSIS
UNIVARIATE ANALYSIS





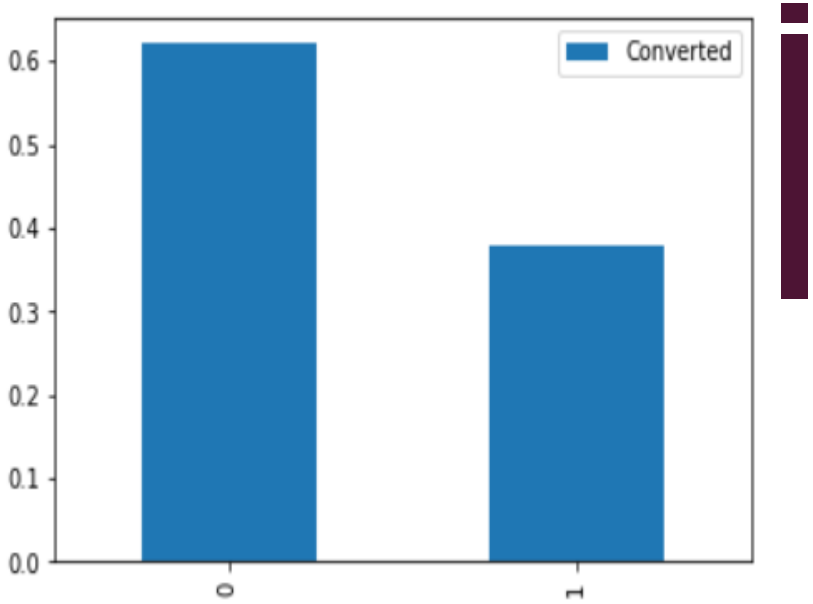


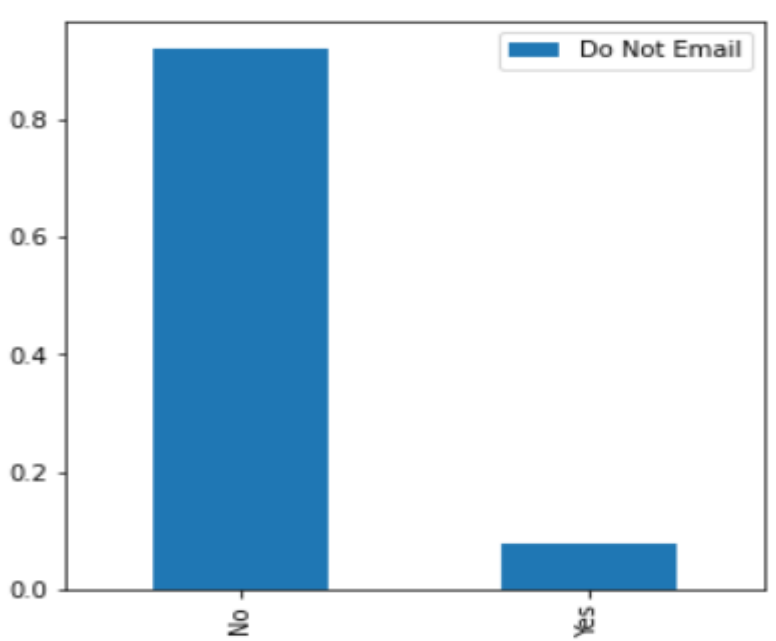


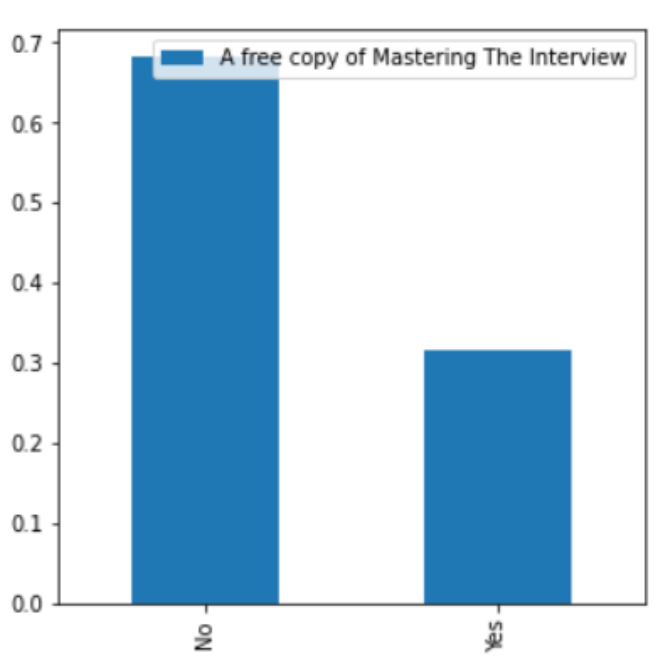


Insights:

- Last Activity: Most of the leads has Opened their email and sent SMS
- Last Notable Activity: Most of the leads has Modified, Opened their email and sent SMS
- Lead Origin: Landing page submission & API has 93% than other that identified to be a lead
- Occupation: 90% of candidates are unemployed, 8% are working professional and 2% are students
- Lead source: Google, Direct Traffic, Olark Chat & Organic search has covered almost 92% of source



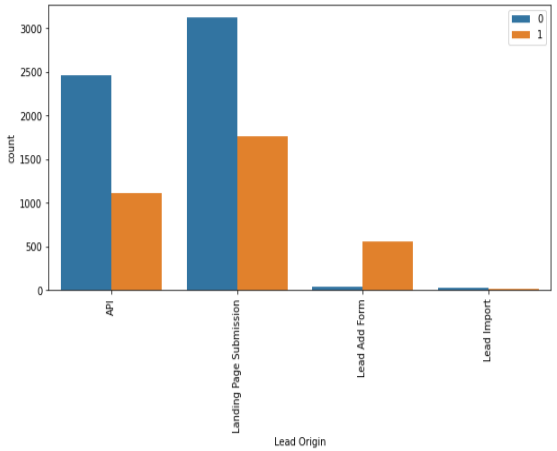


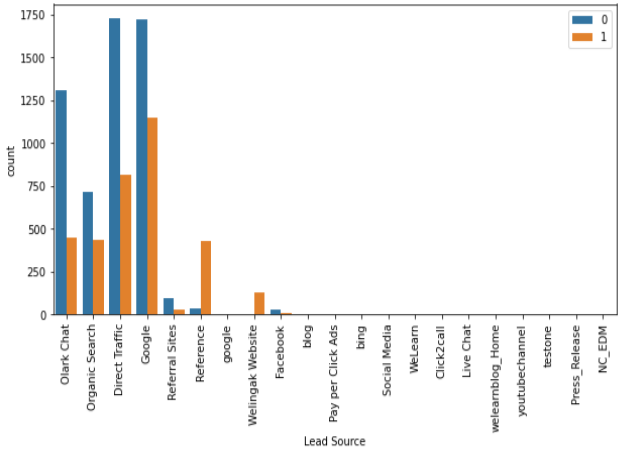


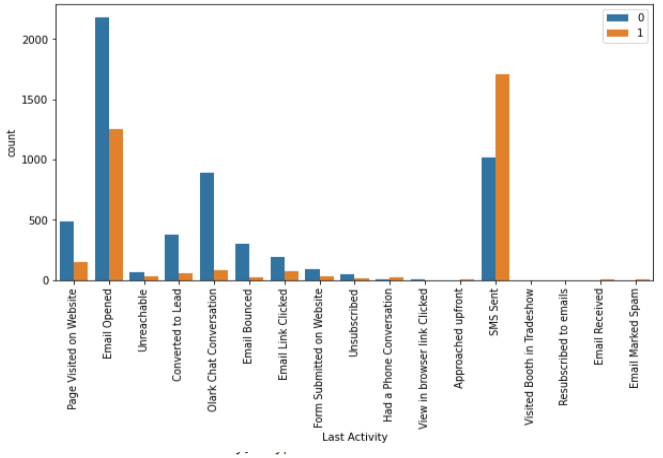
Insights:

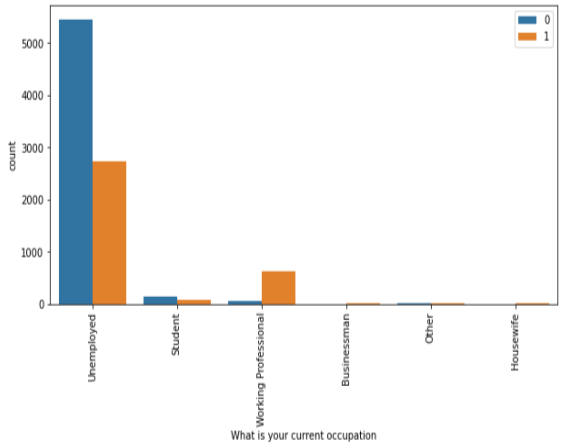
- 62% of leads are converted and 38% are not converted
- Most of the candidate chose No to do not email
- 68% candidates chose No for A free copy of mastering the Interview
- Other variables are skewed towards only one category

BIVARIATE ANALYSIS



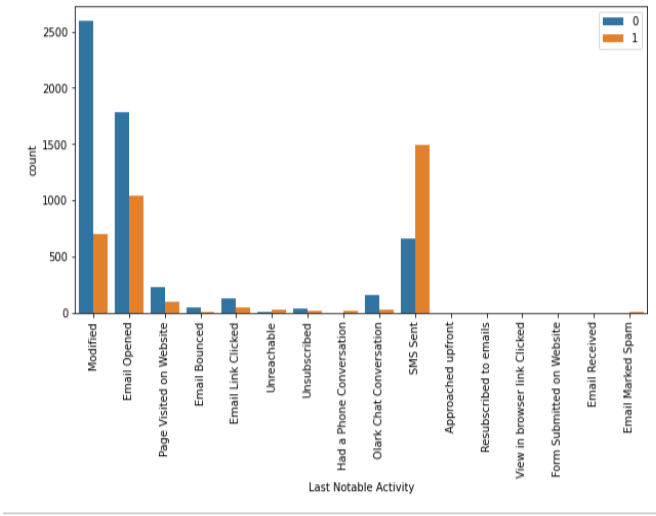


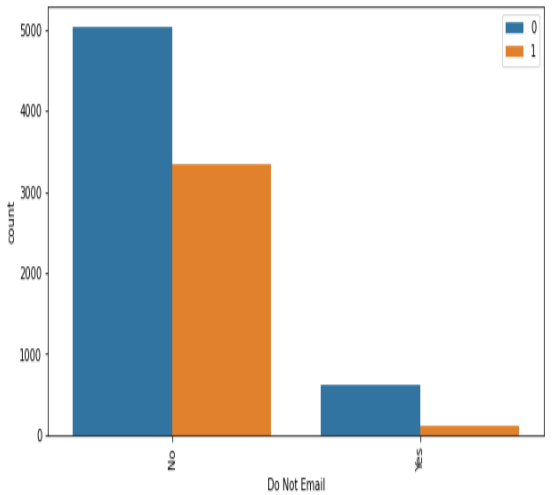


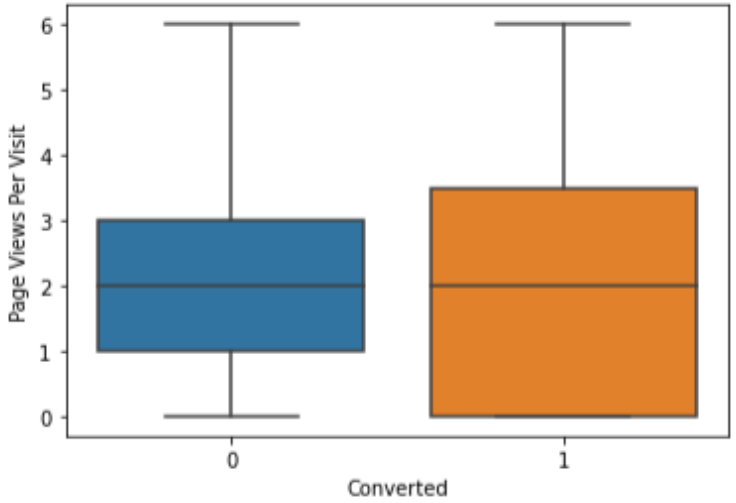


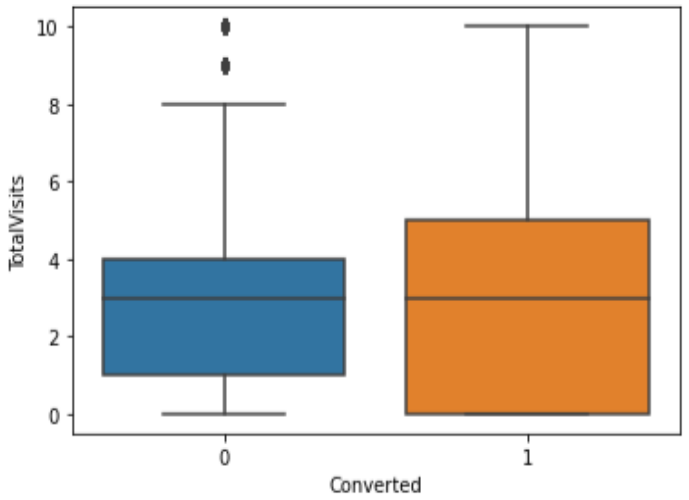
Insights:

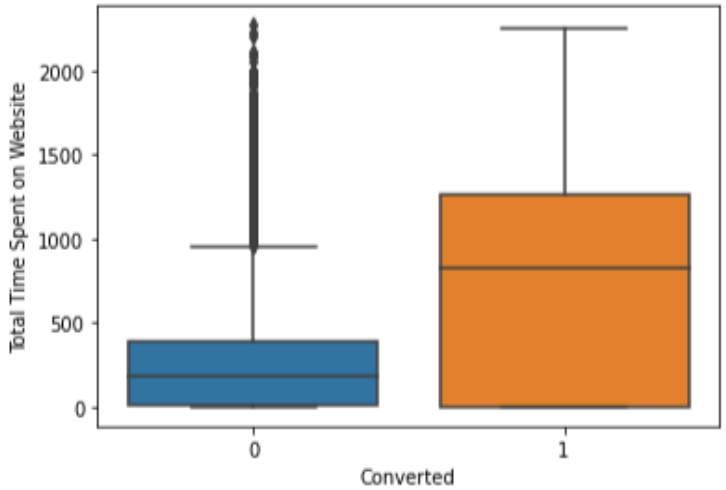
- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.
- Most of the lead have their Email opened as their last activity.
- Conversion rate for leads with last activity as SMS Sent is almost 60%.









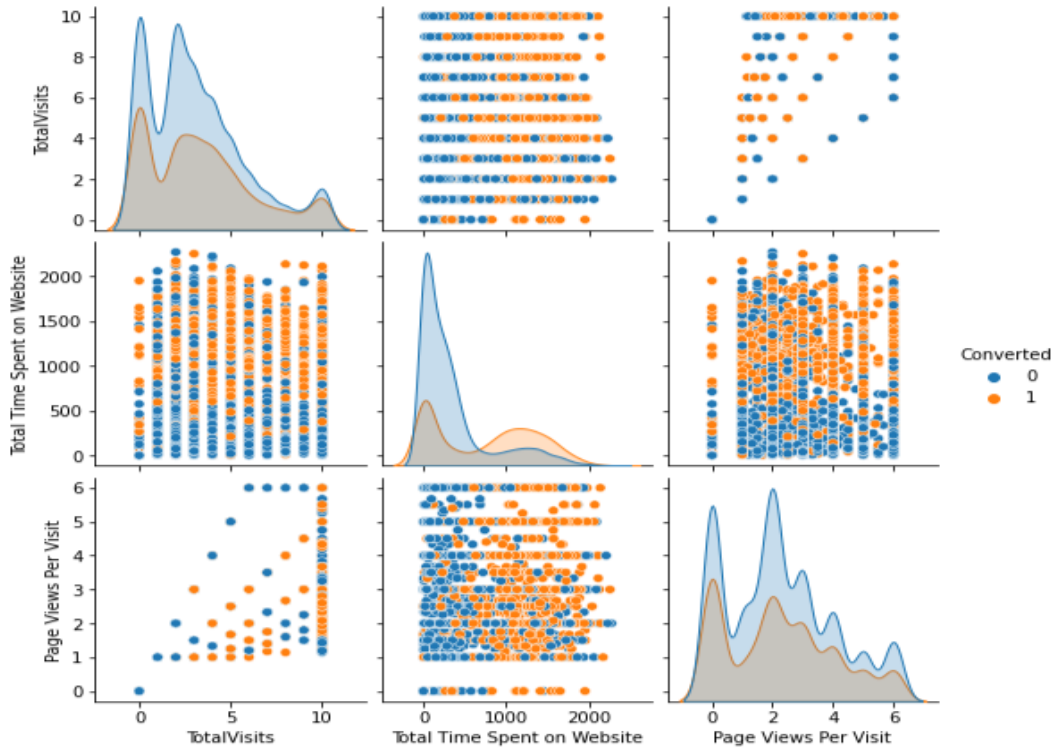


Insights:

•Median for converted and unconverted leads is the same for page views per visit

•. Median for converted and unconverted leads is the same for Total visits

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



MODEL EVALUATION

Comparing the values obtained for Train & Test data:

Train Data:

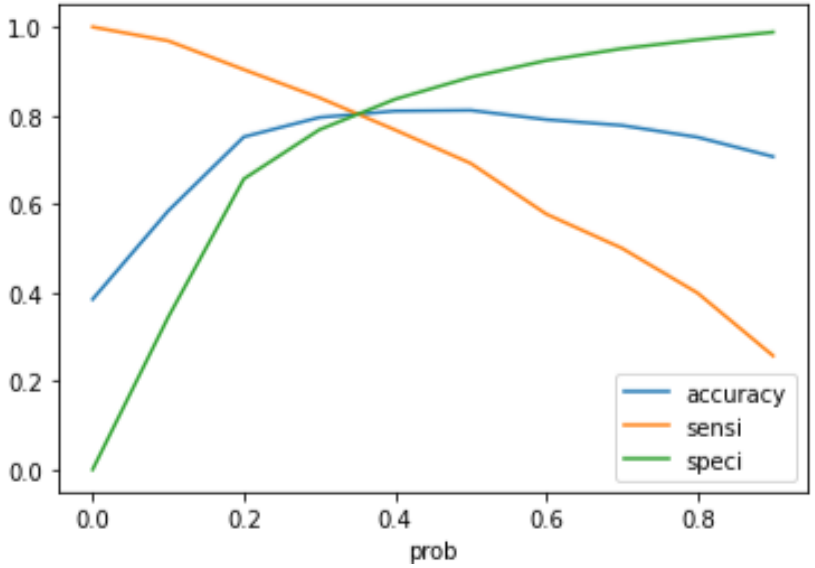
- **Accuracy : 80.46 %**
- **Sensitivity : 80.53 %**
- **Specificity : 80.41 %**

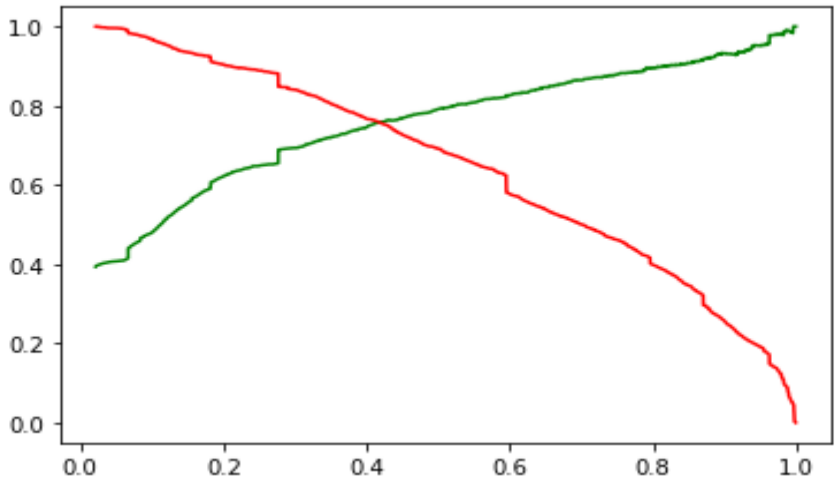
Test Data:

- **Accuracy : 80.66 %**
- **Sensitivity : 79.32 %**
- **Specificity : 81.44 %**

Precision & Recall

- **Precision: 79.21%**
- **Recall: 69.19%**





RECOMMENDATION

- 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.
- The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.
- The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.
- The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.
- The company should not make calls to the leads coming from “Do not Email” and from last notable activity “Email Bounced”, “Page visited on website” as these are not likely to get converted.