LEAD SCORE CASE - X EDUCATION

SCORING POTENTIAL LEADS TO IMPROVE CONVERSION RATE

PROBLEM STATEMENT

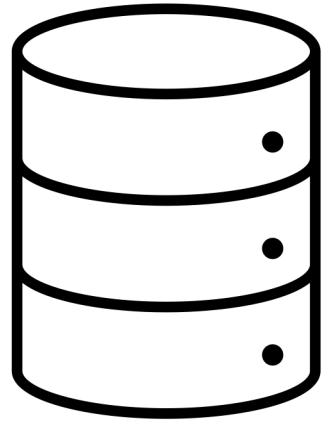
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google, social media etc. to target customers
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some 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 40%.

OBJECTIVE

- X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company needs a model wherein a lead score is assigned 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 chance.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

PROBLEM SOLVING JOURNEY

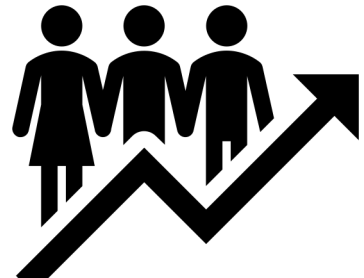
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.
- Scoring the leads based on probabilistic output of final model, and mark the hot leads

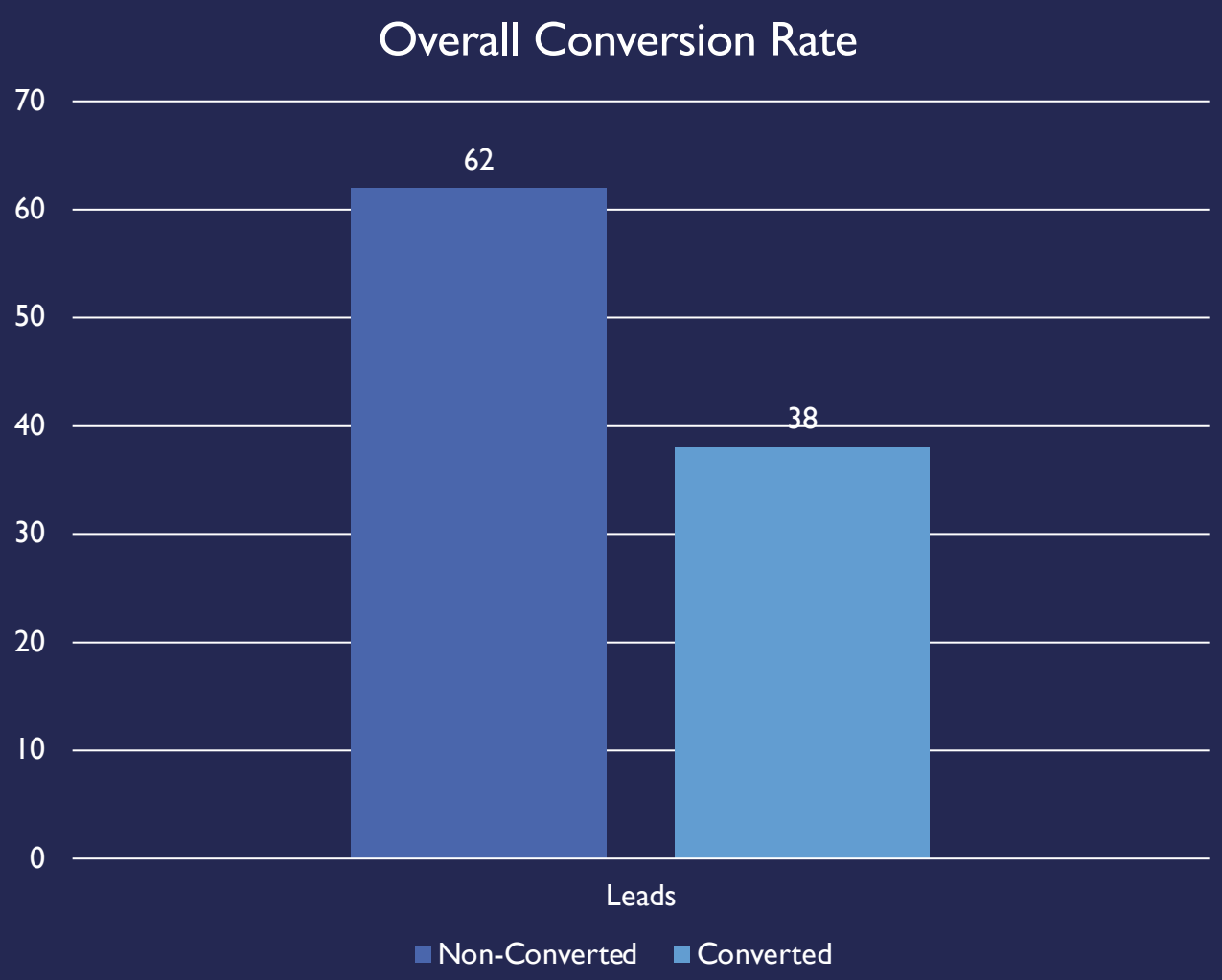


Source Raw data

Exploratory
Data
Analysis

Modelling and
Scoring of leads

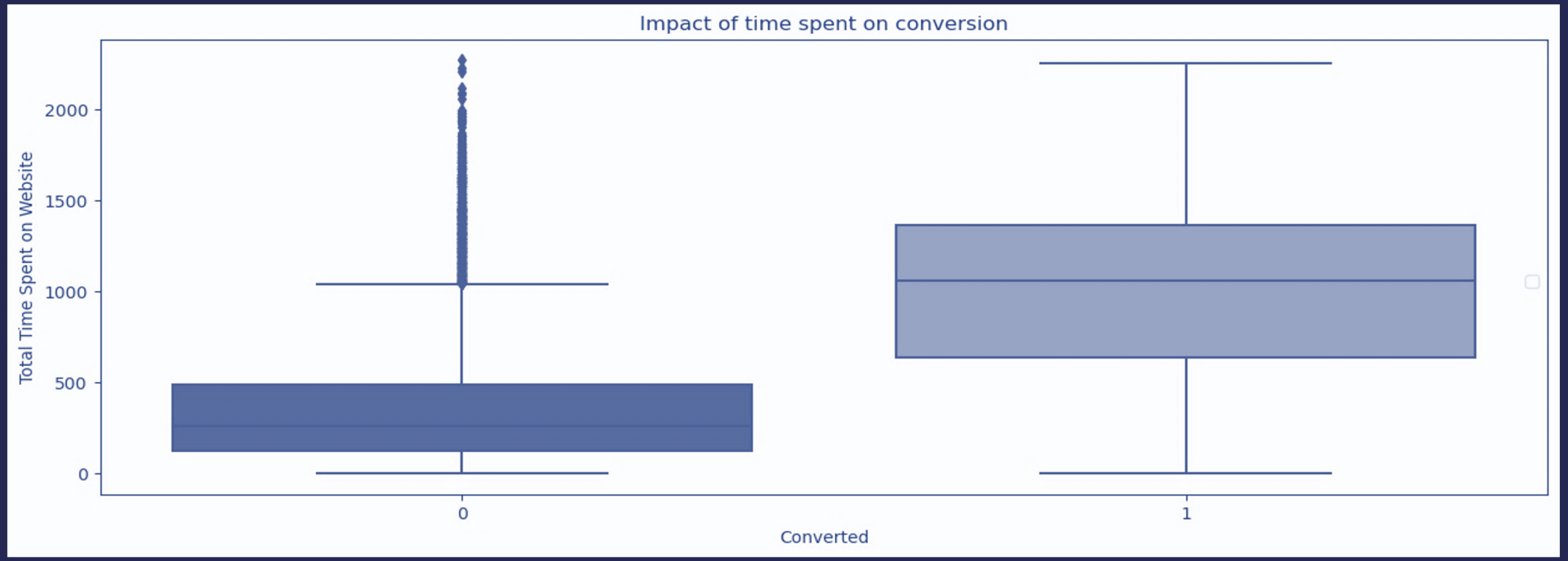




UNDERSTANDING THE BASE CONVERSION

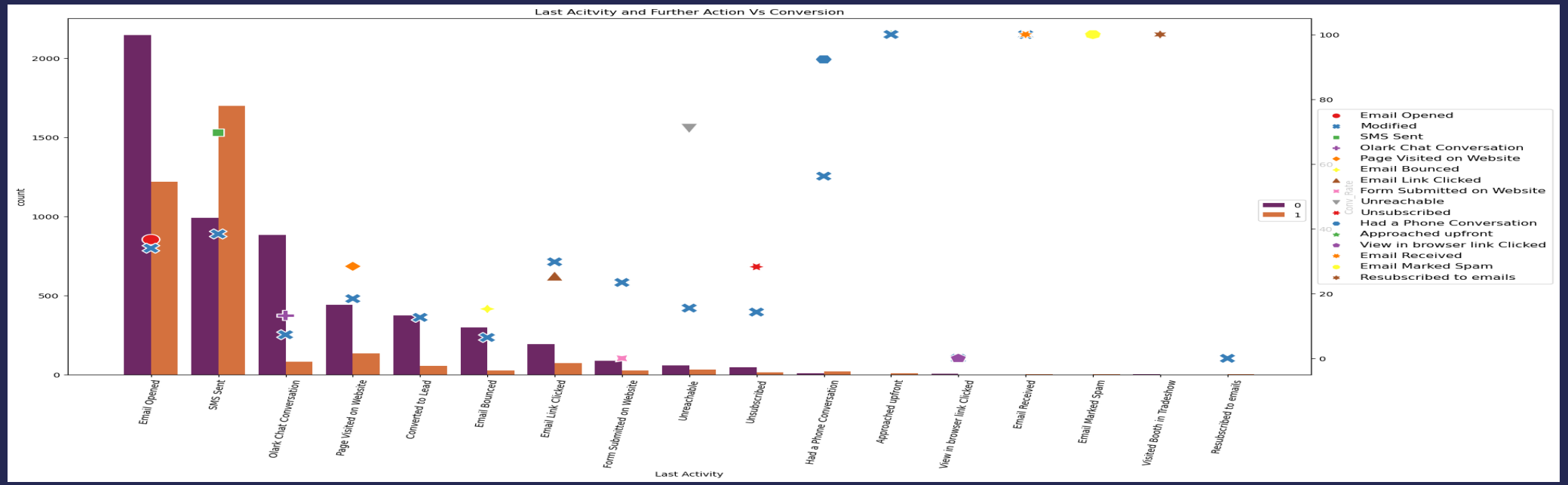
Overall data has close to 40% conversion rate for the leads, which is a maintaining the class balance and enough for building the model without bias

TIME SPENT ON WEBSITE HAS POSITIVE IMPACT ON CONVERSION



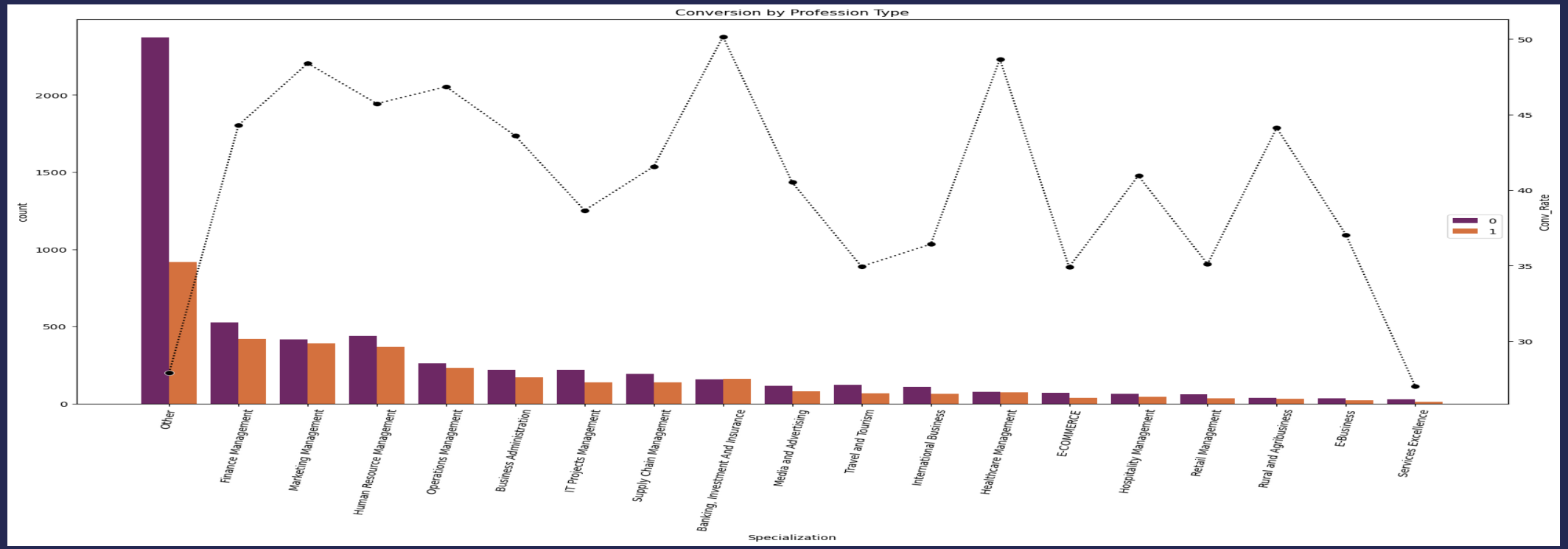
THERE IS A SIGNIFICANT DIFFERENCE IN AVERAGE TIME SPENT FOR THE ONES CONVERTED VS THE ONES WHO DID NOT, QUOTING RELEVANT CONTENT PRESENT

PERSONALIZING MARKETING USING SMS AND PHONE CALL IMPROVE CONVERSION



PHONE CALLS AND SMS HAS MORE THAN 50% CONVERSION RATE, ASSUMING IT HAS HIGHER PERSONALIZED FACTOR INVOLVED TO IT

PROFESSIONALS OTHER THAN STUDENTS AND BUSINESSMAN HAS HIGHER CHANCE TO GET CONVERTED

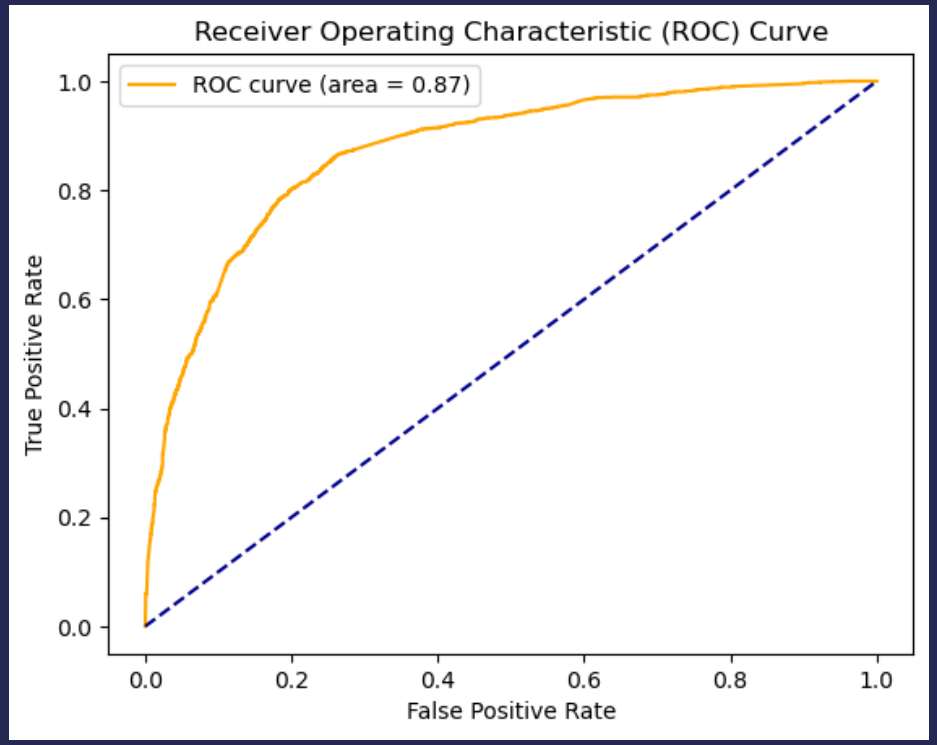


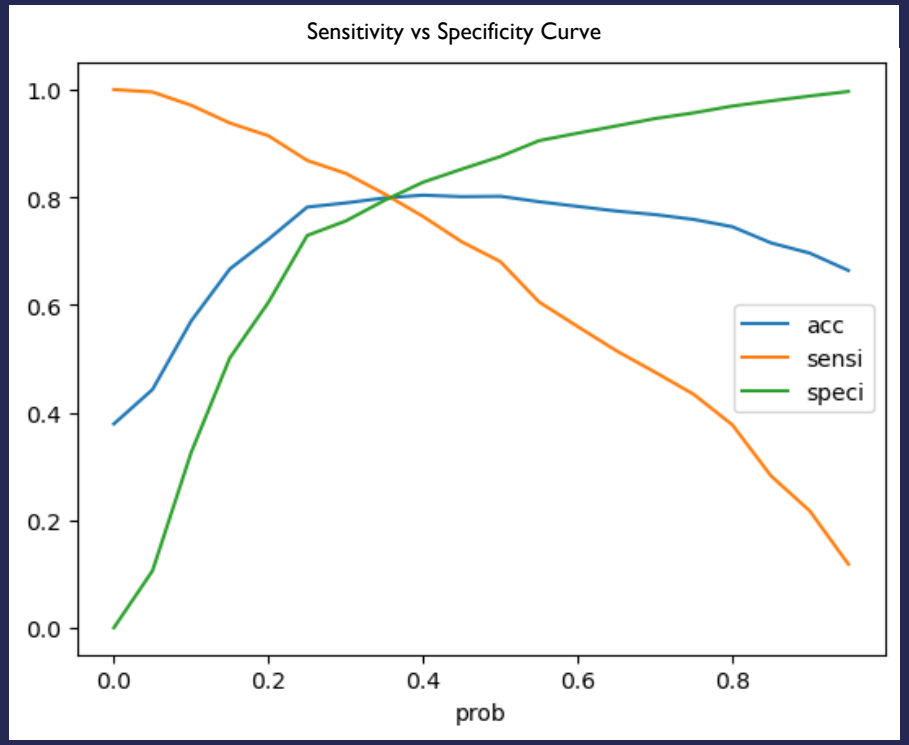
CONTENT RELATED TO FINANCE, HEALTH AND MARKETING PROFESSIONALS CAN BE FOCUSED TO IMPROVE CONVERSION RATE

FACTORS IMPACTING THE CONVERSION RATE – LOGISTIC CLASSIFICATION MODEL

- Factors Impacting the conversion Positively
  - Total Time Spent on the website
  - Leads Originating from Lead Add Forms, prominently Search, chatbots and referrals/references
  - Leads which has phone conversations
  - Leads contacted or marketed regularly through SMS
- Factors which decrease the conversion if leverage
  - Leads modifying in their last notable action, last action was a failure
  - Leads received email even when they opted no email
  - Leads originating from Landing page submission in search and chats
  - Leads interacted with Olark Chat
  - Leads which are either student or businessman
  - Leads from Saudi Arabia region

MODEL EVALUATION: ROC - AUC CURVE

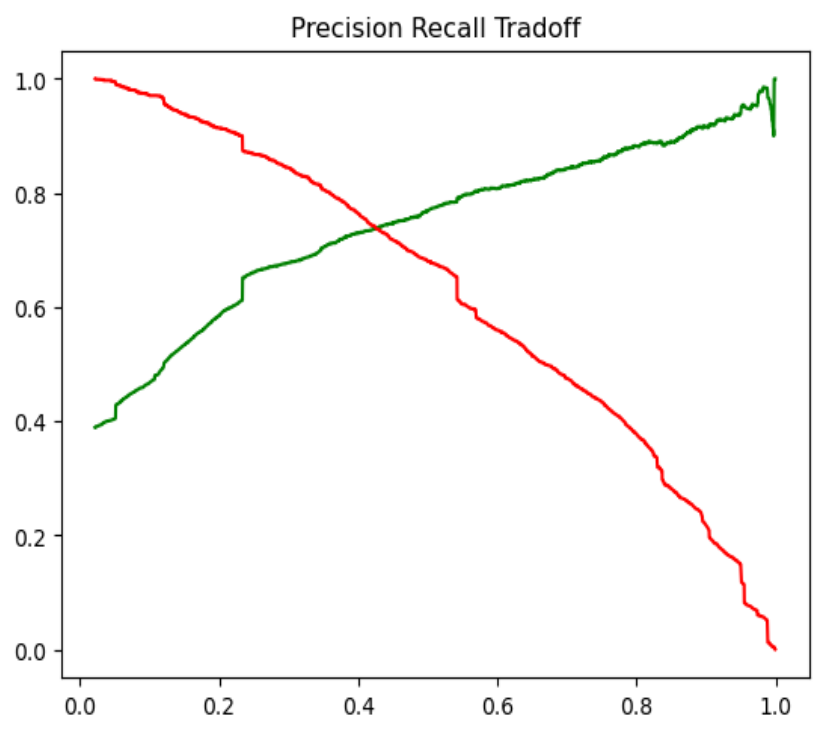




The area under ROC curve is 0.87, which suggests that current model is enough to distinguish positive and negative classes ~90% of the times

The above graph depicts an optimal cutoff of 0.35 based on Accuracy, Sensitivity and Specificity to classify the probability into 0 and 1 classes.

MODEL EVALUATION : PRECISION RECALL TRADE-OFF



- The graph depicts an optimal cutoff of 0.4 based on Precision and Recall trade-off

Performance Metrics	Train Model	Test Model
Confusion Matrix	[[3076 798] [458 1903]]	[[1324 324] [203 822]]
Accuracy	0.80	0.80
Sensitivity	0.80	0.80
Specificity	0.79	0.80
Negative PV	0.87	0.87
Precision	0.70	0.71
Recall	0.81	0.80

- Both Test and Train evaluations perform well in predicting the conversion and non-conversion 80% of the time, suggesting that model is robust and well enough to improve the business

RECOMMENDATIONS

- Focus on quality content, as consumers spending **more time on website** has higher chance to get converted. More engaging the content, better the results
- **Calls** Improve conversion, and on top of that leads coming from **chatbots** if engaged on calls get more likely to get converted.
- Making calls to working **professional, especially in 'Finance' and 'Marketing'** has more likeliness to be converted.
- **Sending SMS** to leads has positive relation towards conversion, can be used as good marketing strategy.
- Company's focus on countries other than Saudi Arabia can have better conversion rate.
- The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.
- The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.
- If company can make the interaction with the leads more streamlined using chats or SMS and **adding quick forms** to get fast subscription from the links will improve the business. **At last, website's content is the winner which will make or break the customer lead, that should be the utmost priority to keep it relevant.**