# LEAD SCORING CASE STUDY

## DATA ANALYSIS FOR BUSINESS IMPROVEMENT & IDENTIFYING KEY MARKETING PARAMETERS

Team
Akhilesh Mathan
Shashi Kumar Singh
Jitalaxmi Bharali

# TABLE OF CONTENT

# BUSINESS CASE

An education company named X Education operating business of online courses to industry professionals.

- Professionals who are interested in the courses land on their website.
- The company markets its courses on several websites and search engines like Google.
- These people browse the courses or fill up a form our 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%.

# PROBLEM STATEMENT & OBJECTIVE

- X Education gets leads, however conversion rate is at around 30%
- X Education wants to improve lead conversion making it more efficient by identifying the most potential leads, also known as Hot Leads
- Sales team want to know these potential set of leads, which they will be focusing more .

**Objective of the Study:**

- To increase conversion of leads into paying customers.
- To develop a model assigning a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and vice versa.
- The CEO has given a ballpark of the target lead conversion rate to be around 80%.

# SUGGESTED IDEAS FOR LEAD CONVERSION

**Hot Leads**

- Most potential leads who are most likely to get converted

**Business Communication**

- The sales team will now be focusing more on communicating with the potential leads

**Business Target**

- Concentrating on the Hot Leads to improve the current conversion rate of 30% to 80%.
- Building a model to assign lead score to each lead so that consumers with higher lead score has higher conversion rate.

# ANALYSIS APPROACH



## DATA LOADING & CLEANING

Loading Data & study. Dropping col > 35 % null values & insignificant col, imputing categorical data.

## DATA UPDATION

Treating outliers, fixing invalid data, grouping low frequency values & mapping binary categorical values.



## EXPLORATORY DATA ANALYSIS

Checking data imbalance, univariate & bivariate analysis for categorical/ numerical variables, identifying significant variables .

## DATA PREPARATION

Creating dummy features for cat variables, splitting into train and test sets, Feature scaling using standardization, dropping high correlated col.

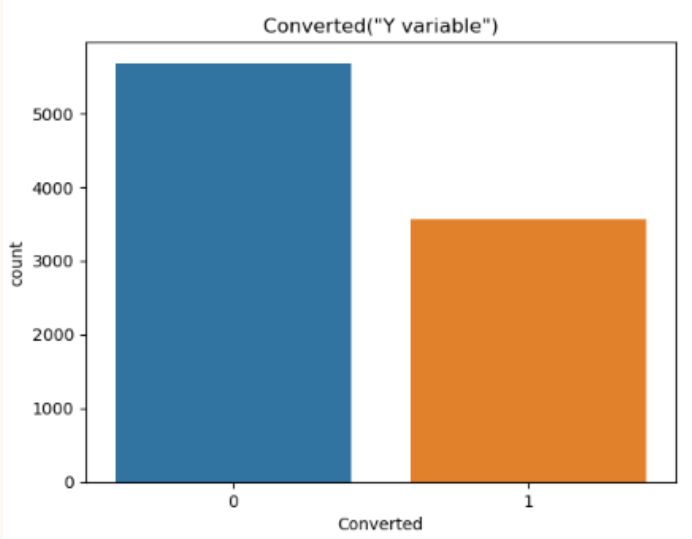


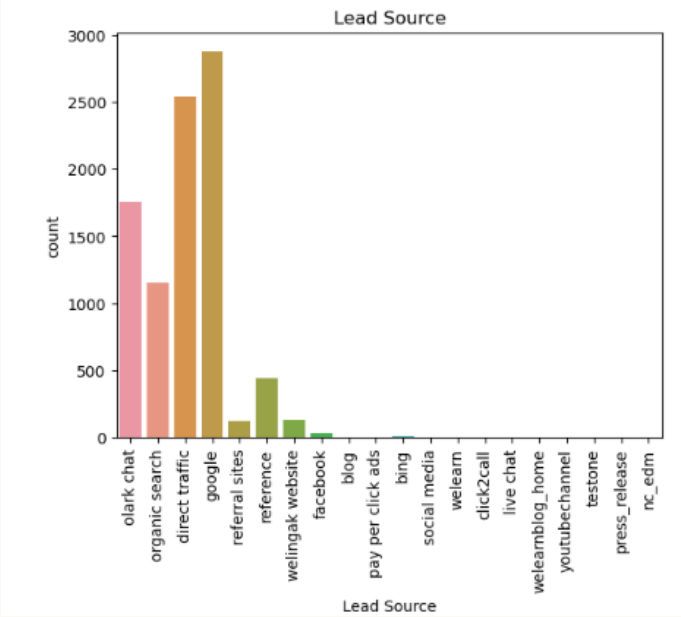
## MODEL EVALUATION

Reducing nos. of variables using RFE. Compare train vs test metrics, assign Lead Score & get top feature. Identify main features for better conv rate

# EXPLORATORY DATA ANALYSIS ( EDA)

- Univariate Analysis & Data Imbalance:



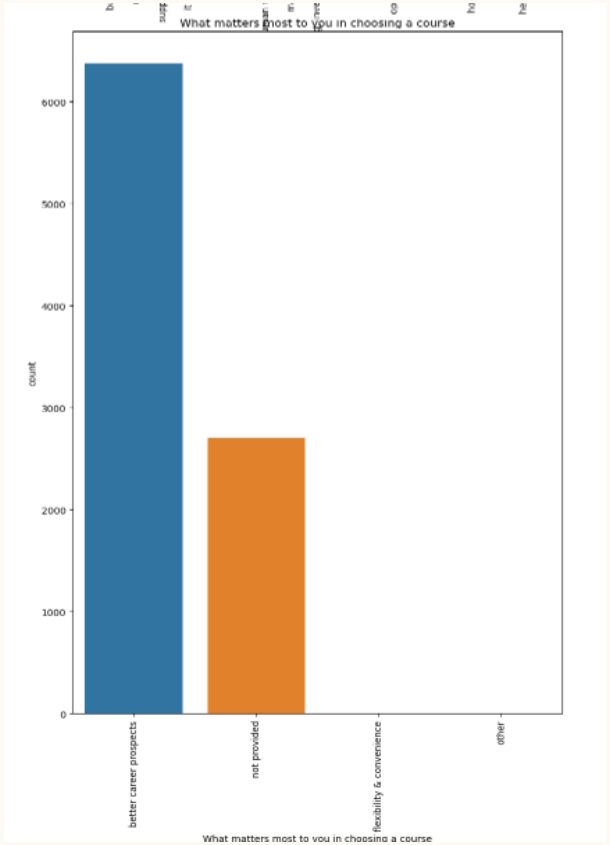

*Insights:*

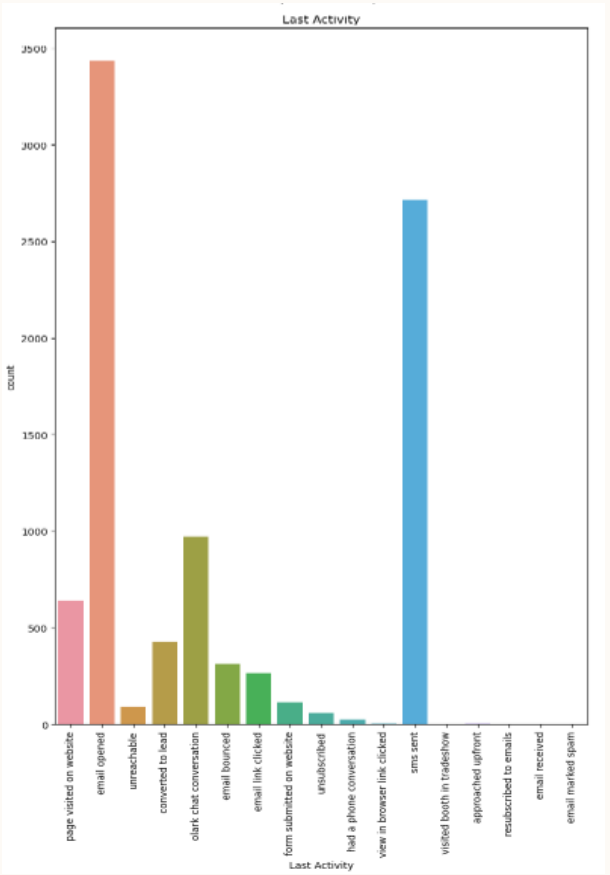
**Conversion of Leads:** The rate of conversion is lower i.e most of the people do not convert to leads.

**Lead Source:** More than 50% of the leads are sourced from Google and direct traffic

# EXPLORATORY DATA ANALYSIS ( EDA)

- Univariate Analysis:



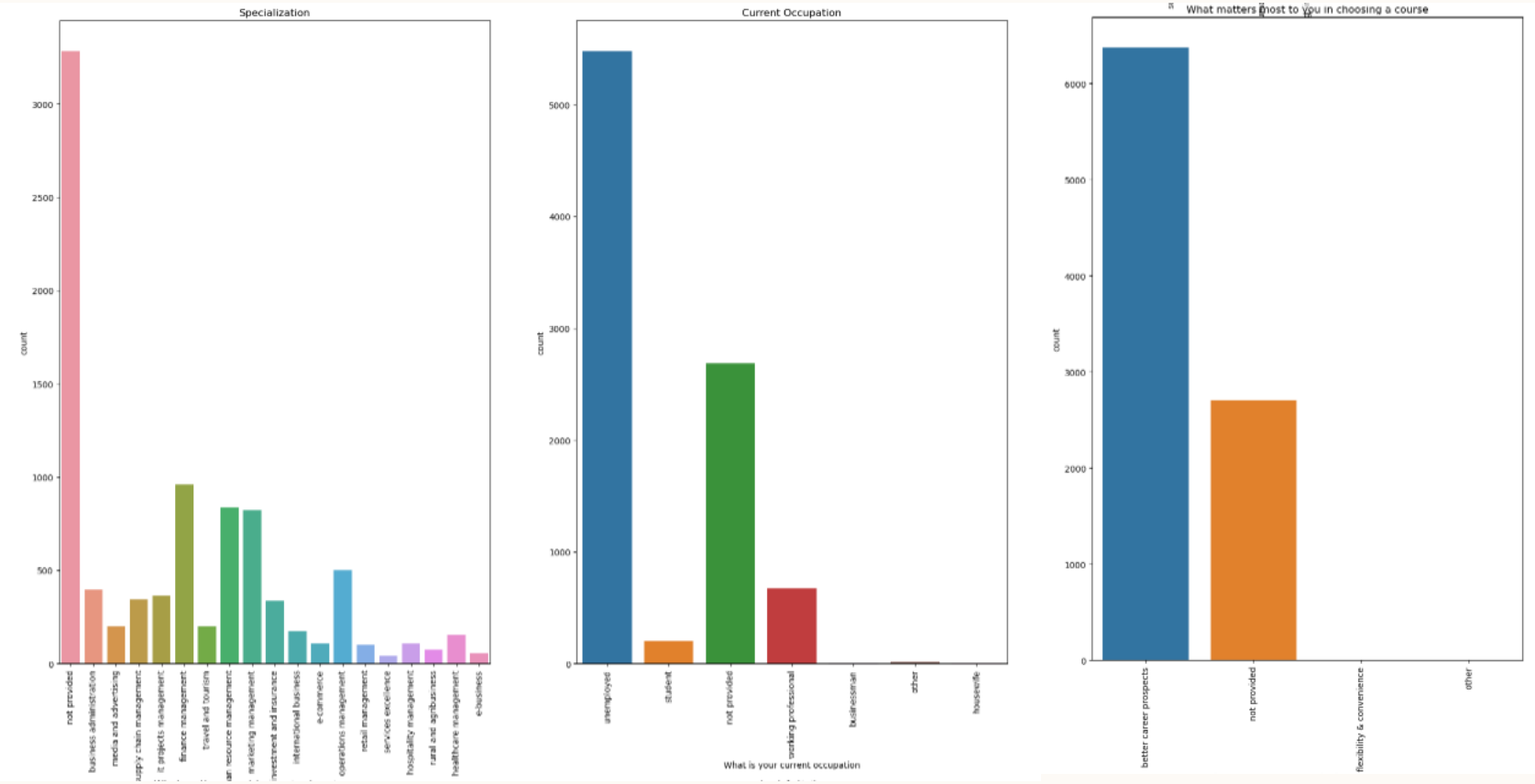

*Insights:*

**Last Activity :** Close to 70% of the customers contribution is through email open activities and SMS.

**Customer's requirement**: Around 80$ of the customers are looking at better

# EXPLORATORY DATA ANALYSIS ( EDA)

- Univariate Analysis:



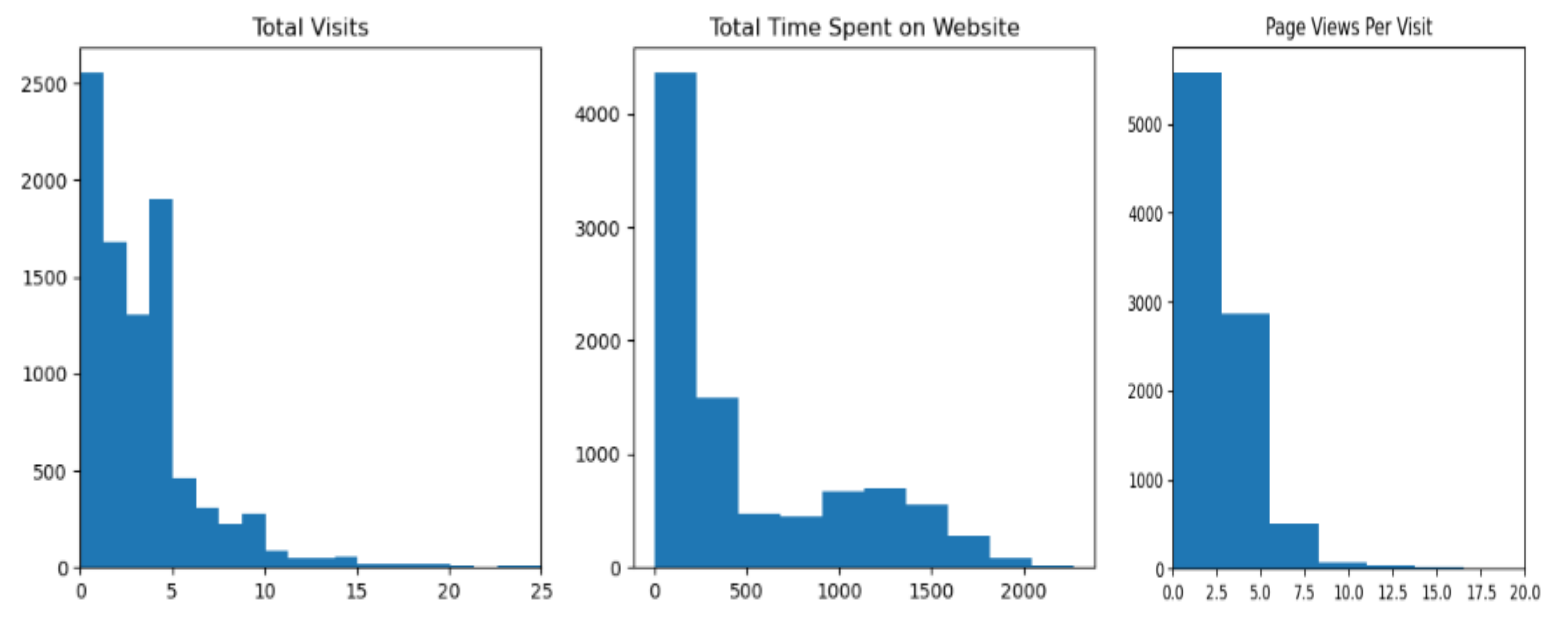

*Insights:*

Current_occupation: It has 90% of the customers as Unemployed.

Occupation: Approx 90% of the customers donot disclose their specialization.

# EXPLORATORY DATA ANALYSIS ( EDA)

- Univariate Analysis:

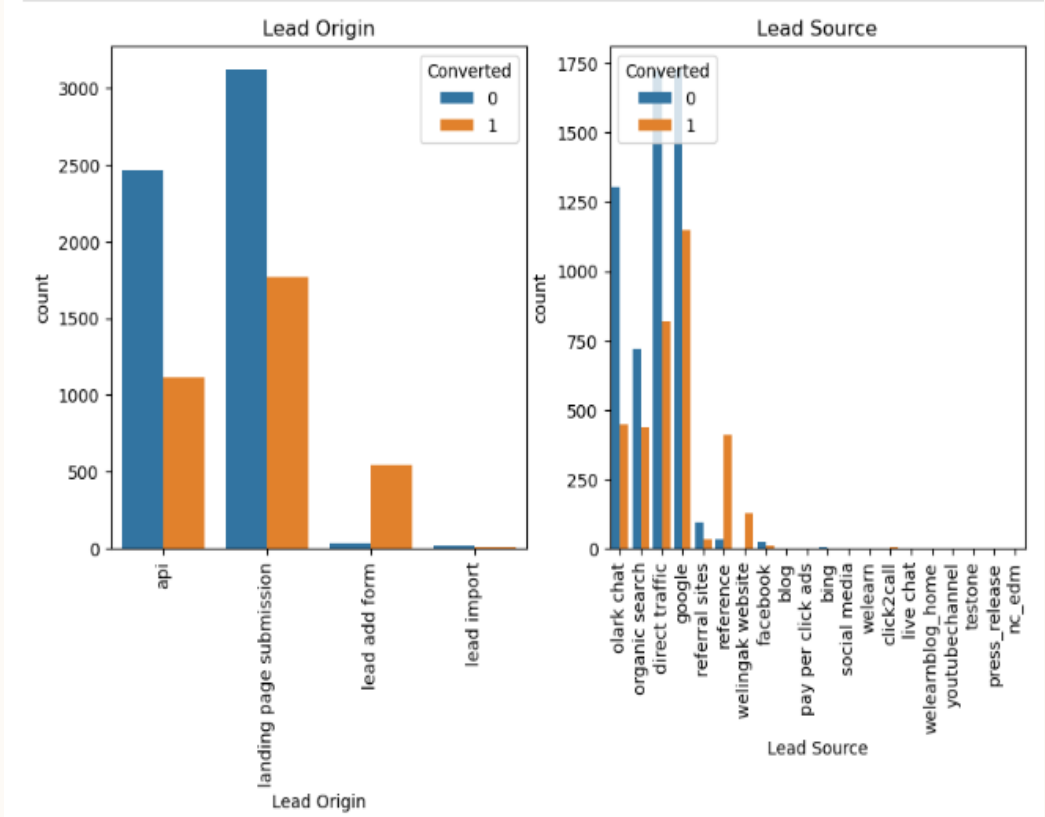



*Insights:*

Total time spent on website: Most of the people do not read thoroughly or try to understand the course material

# EXPLORATORY DATA ANALYSIS ( EDA)

- Bivariate Analysis:

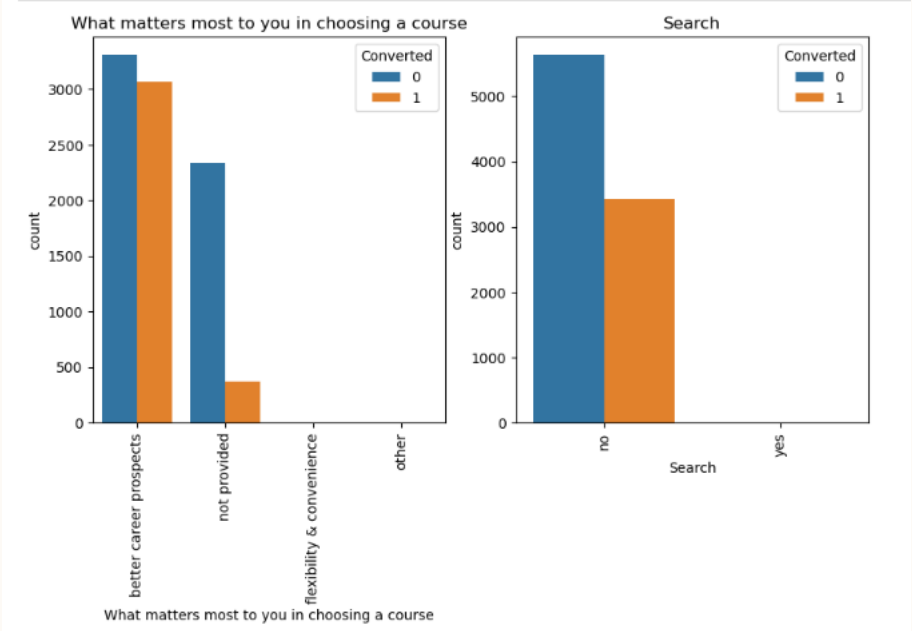



*Insights:*

**Lead Source:** Among diff sources Google and direct traffic have the highest contribution ,with substantial number of unconverted consumer.

# EXPLORATORY DATA ANALYSIS ( EDA)

- Bivariate Analysis for Categorical Variables:

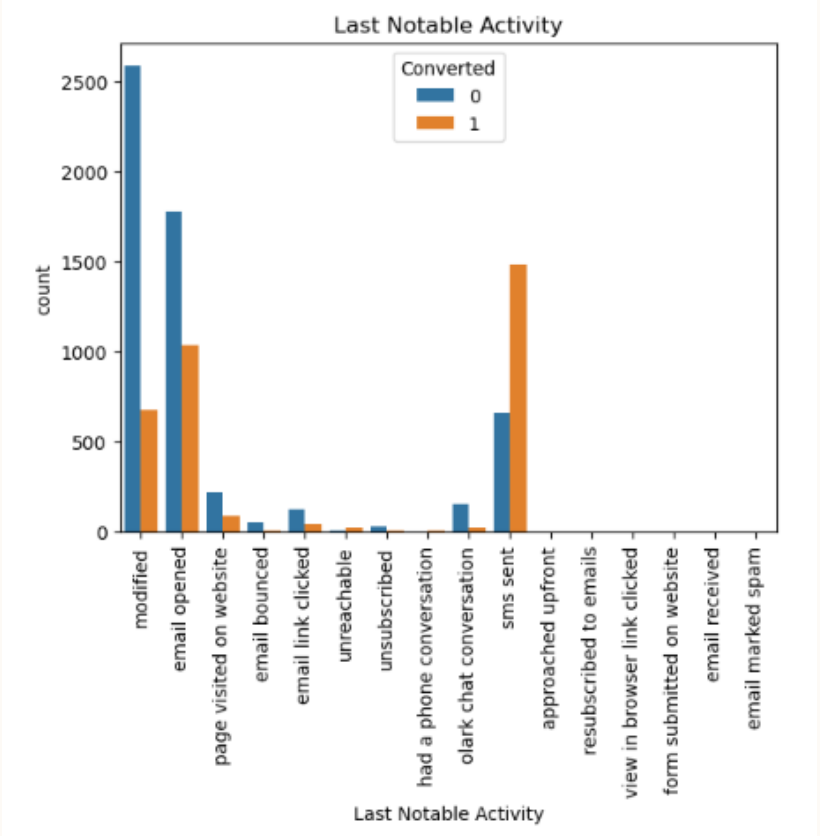



*Insights:*

Last Activity: Most of the conversion rate was from professionals looking for better career prospects.

# EXPLORATORY DATA ANALYSIS ( EDA)

- Bivariate Analysis :




*Insights:*

Last Activity: 'SMS Sent' has high lead conversion rate of > 60% with 'Email Opened' activity contributed close to 40% of last activities performed by the customers.

# DATA PREPARATION FOR MODEL DEVELOPMENT

- *Binary level categorical columns are mapped to 1 / 0 .*

- *Created dummy features for categorical variables* – Lead Origin, Lead Source, Last Activity, Specialization, Current occupation.

- *Splitting Train & Test Sets*
  - 70:30 % ratio was chosen for the split

- *Feature scaling*
  - Standardization method was used to scale the features

- *Checking the correlations*
  - Predictor variables which were highly correlated with each other were dropped (Lead Origin_Lead Import and Lead Origin_Lead Add Form).

# MODEL BUILDING

*Feature Selection*

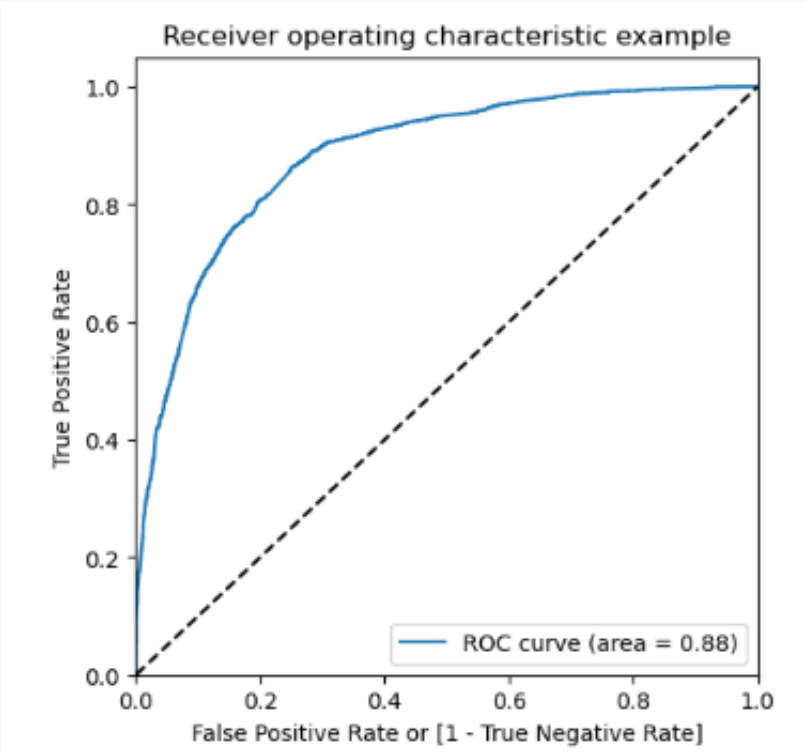
- The data set has many dimension and large number of features.
- This will reduce model performance and might take high computation time.
- Hence performing Recursive Feature Elimination (RFE) and to select only the important columns.
- The model accuracy is approx. 81%, which is as required by the project.
- RFE outcome
  - Pre RFE – 48 columns & Post RFE – 43 column

Manual Feature Reduction process was used to build models by dropping variables with p – value > 0.05.

- Model is stable after many iteration with:
  - significant p-values within the threshold (p-values < 0.05) and
  - Multicollinearity with VIFs less than 5.

# MODEL EVALUATION

Train data:

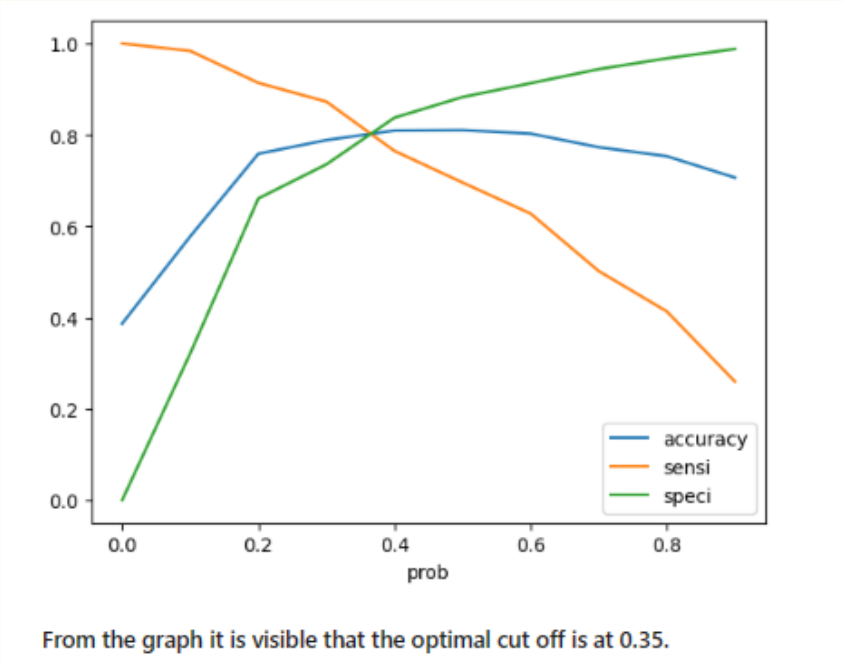



*Insights:*

The area under Receiver Operating  Characteristic curve is 0.88 , which is supports the performance of the model developed.

# MODEL EVALUATION

Train Dataset:



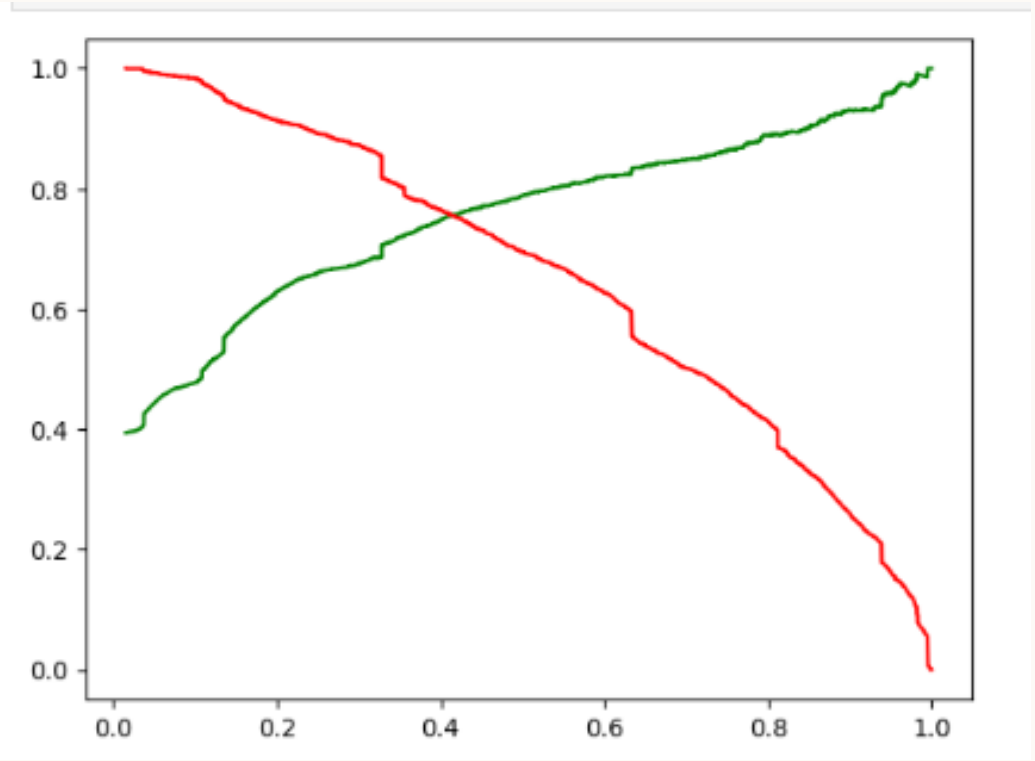

From the graph it is visible that the optimal cut off is at 0.35.

*Insights:*

Considering optimal cut off as 0.35 , the model evaluated the current cut off as 0.41 and with Precision of around 75% and Recall of around 76%

# MODEL EVALUATION

Test Dataset:




*Insights:*

The current cut off as 0.41 , the model predicts Precision of around 75% and recall of around 76%.

# HOW WE GET THERE

## BUSINESS APPROACH

Synergize communication with smaller prospective leads.
Foster holistically superior methodologies

## DATA ANALYTICS

With cut-off value of **0.35**, the model achieved a precision of 75 % in the train set & 75% in the test set.
The model evaluated Sensitivity of 80%.
Sensitivity in this case indicates nos of leads the model identify which are converting
Model achieved an accuracy of 80%, in line with the study objectives.

# AREAS OF FOCUS

## MARKET SCENARIOS

- Focus on features with positive coefficients for targeted marketing strategies.
- Develop strategies to attract high-quality leads from top-performing lead sources.
- Optimize communication channels based on lead engagement impact.
- Engage working professionals with tailored messaging.

## PROSPECTIVE OPPORTUNITIES

Lead Source

- Lead Source_Reference
- Current_occupation_Working Prof
- Last Activity_SMS Sent
- Last Activity_Others
- Total Time Spent on Website

- Specialization in Hospitality Mgmt
- Specialization in Others
- Lead Origin of Landing Page Submission

# SUMMARY

*To achieve target conversion rate of 80%,a regression model was built to identify the most significant factors that impact lead conversion. :*

- Focus on features with positive coefficients for targeted marketing strategies.
- Develop strategies to attract high-quality leads from top-performing lead sources.
- Optimize communication channels based on lead engagement impact.
- Engage working professionals with tailored messaging.
- More budget for Welingak Website in terms of advertising, etc.
- Incentives/discounts for providing reference that convert to lead, encourage providing more references. More visibility on Google.
- Working professionals to be aggressively targeted as they have high conversion rate and financial suitable to pay higher fees

*To identify areas of improvement :*

- Analyze negative coefficients in specialization offerings.
- Review landing page submission process for areas of improvement.

# THANK YOU

www.github.....