

Lead Score Case Study

Mukul Dutt Bharadwaj

MS Alexander Sooraj

DSC59–Aug 2023 Batch

Prepared and Submitted

on 25 Feb 2024

Agenda

- ❖ Problem Statement
- ❖ Business Goals
- ❖ Solution Approach
- ❖ Exploratory Data Analysis
- ❖ Model Building
- ❖ Model Evaluation – TRAIN
- ❖ Model Evaluation – TEST
- ❖ Conclusion

Problem Statement

- ❖ X Education is an organization that provides online courses for Industry Professionals.
- ❖ The company markets its courses on several websites and search engines like Google.
- ❖ Even though the X Education gets many potential customers interested, only a small number end up actually buying courses. X Education wants more of these potential customers to become paying customers.
- ❖ The company converts about 30% of its leads into customers by reaching out to those who seems interested in taking the course. However, the way they attract these leads isn't very effective in getting them to become customers.

Business Goal

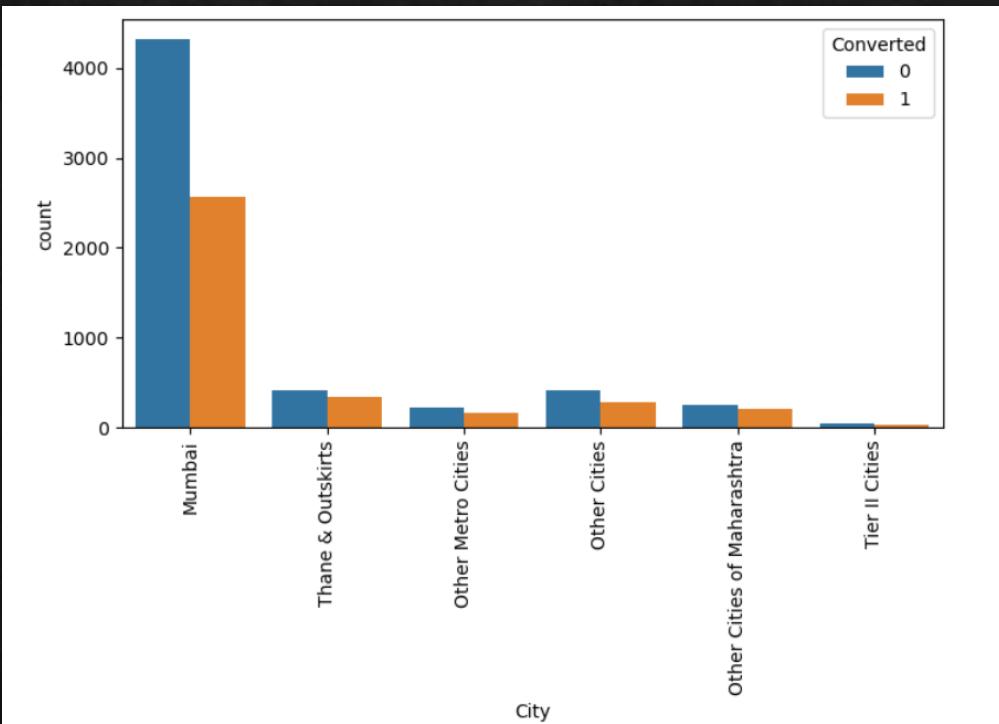
- ❖ The Company requires a model to be built for selecting most promising leads.
- ❖ Lead score to be given to each lead such that it indicates how promising the lead could be.
- ❖ The higher the lead score the more promising the lead to get converted and the lower the lead score, lesser the chances of conversion.
- ❖ The model to be built in lead conversion rate around 80%.

Solution Approach

- ❖ Load / Import Dataset
- ❖ Check the Data Quality and Prepare dataset for analysis
- ❖ Exploratory Data Analysis – Univariate and Bivariate Analysis
- ❖ Feature Scaling
- ❖ Prepare the data for model building
- ❖ Build logistic regression model
- ❖ Assign a lead score to each lead
- ❖ Test the model on train set
- ❖ Evaluate model by different measure and metrics
- ❖ Test the model on Test set
- ❖ Measure the accuracy of the model and other metrics

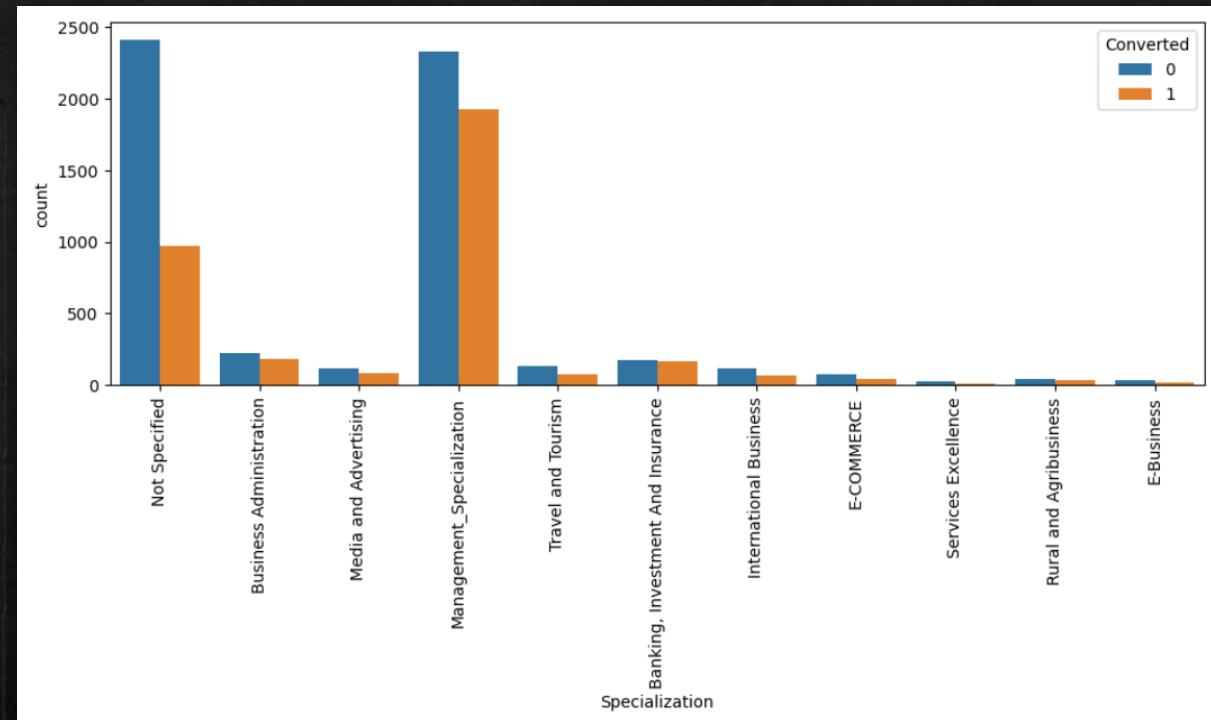
Exploratory Data Analysis

City Vs Converted



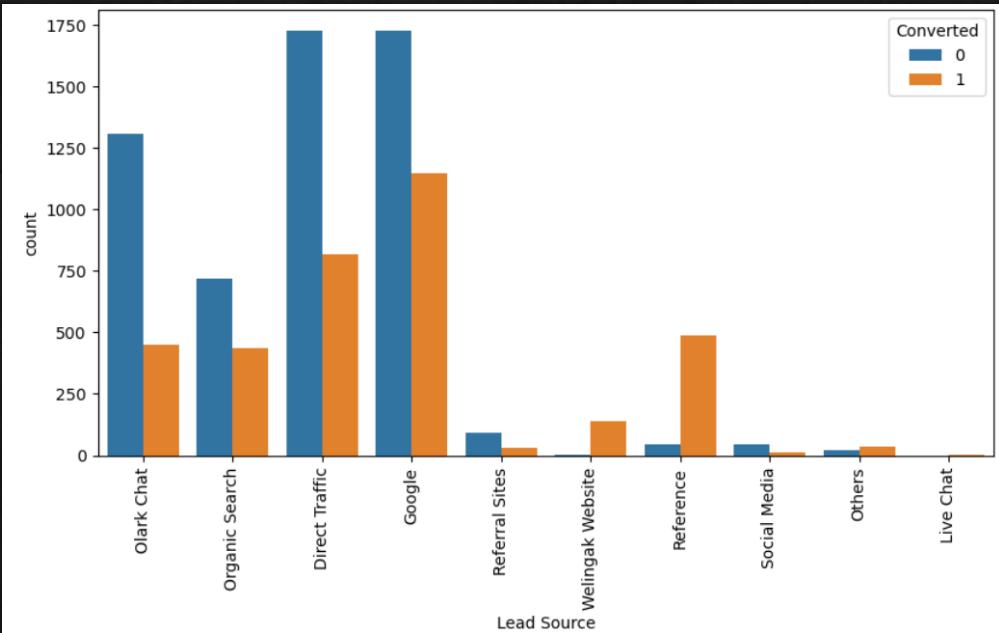
- Mumbai has highest conversion rate compared to other City

Specialization Vs Converted



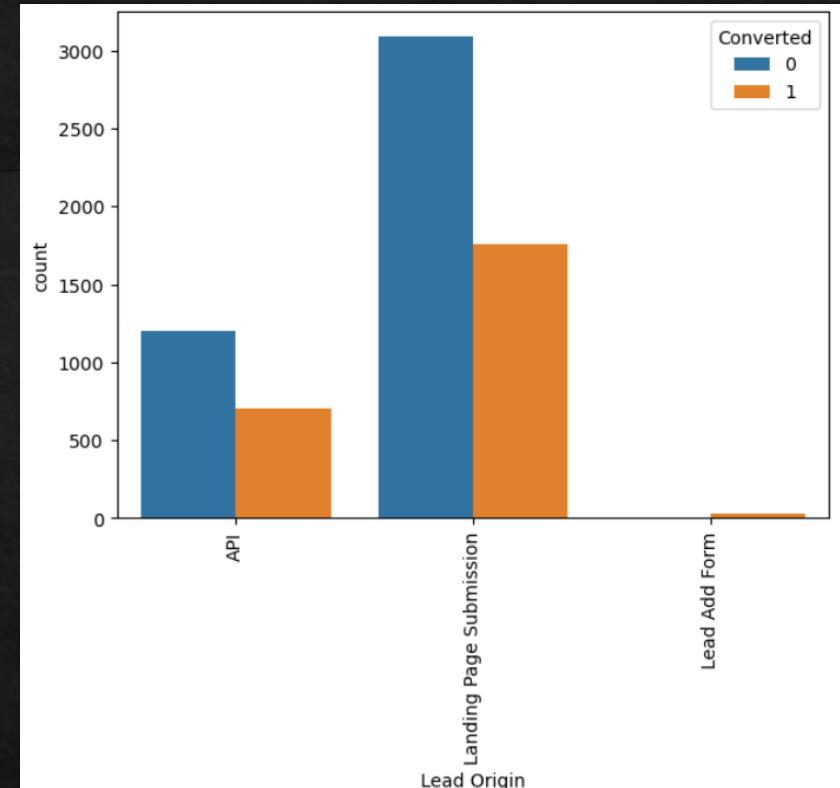
- Management Specialization has highest conversion rate compared to other Specialization.

Lead Source Vs Converted



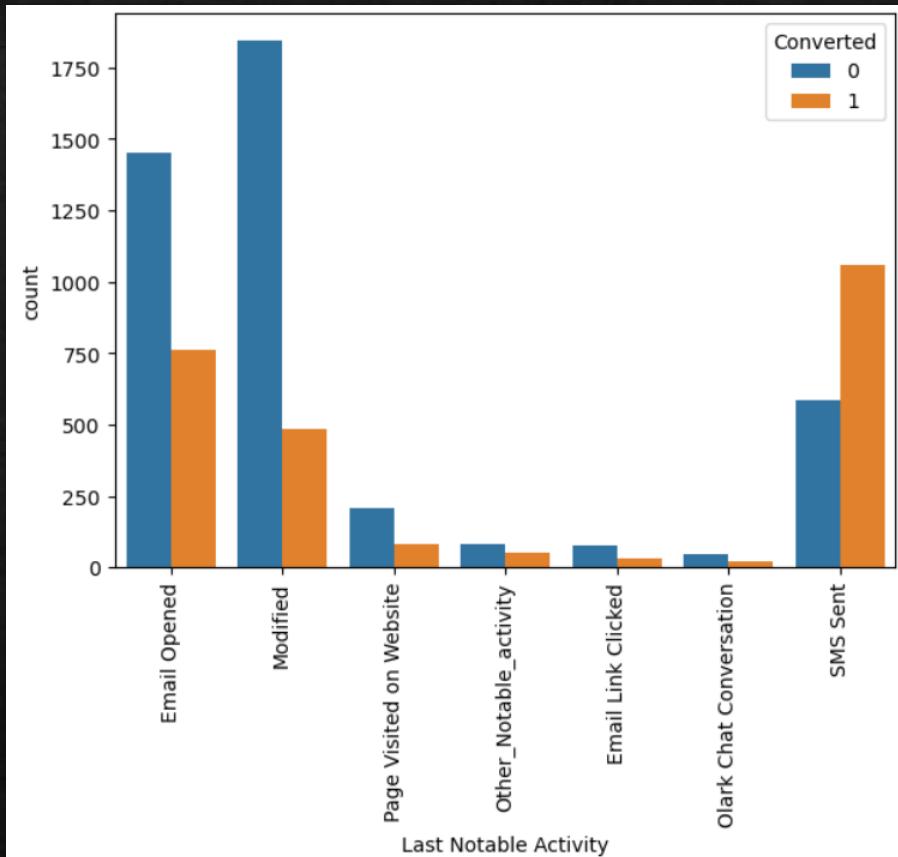
- Maximum number of leads are generated by Google and Direct traffic.
- Conversion Rate of reference leads and leads through welingak website is high.
- To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

Lead Origin Vs Converted

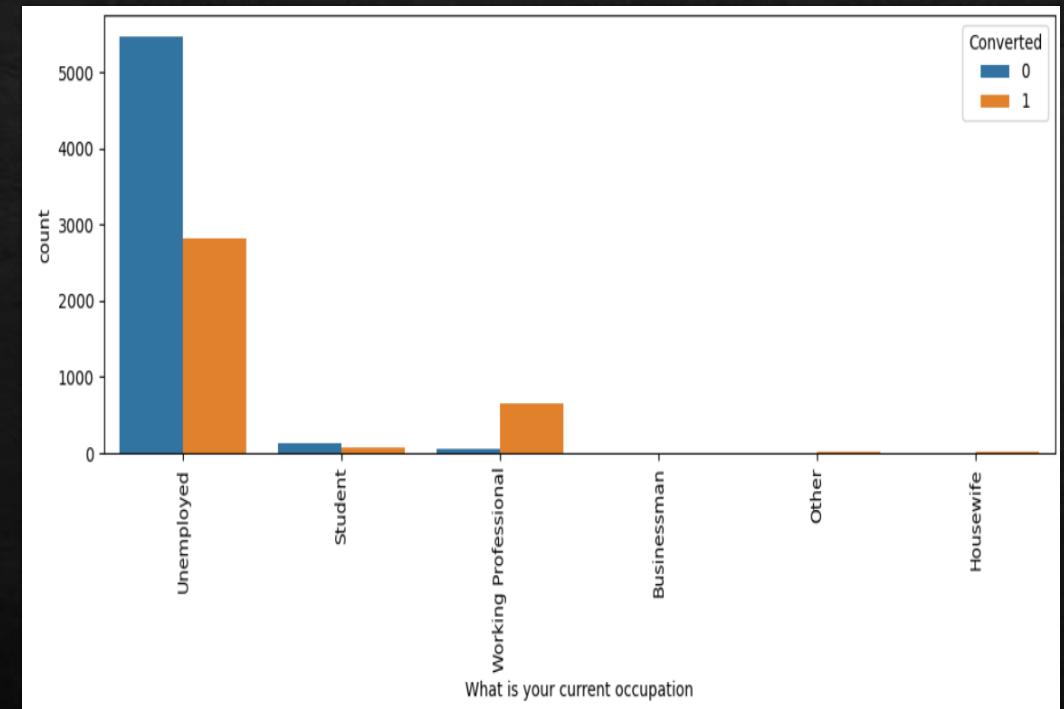


- API and Landing Page Submission bring higher number of leads as well as conversion.
- Lead Add Form has a very high conversion rate but count of leads are not very high.
- Lead Import and Quick Add Form get very few leads.
- In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

Last Notable Activity Vs Converted

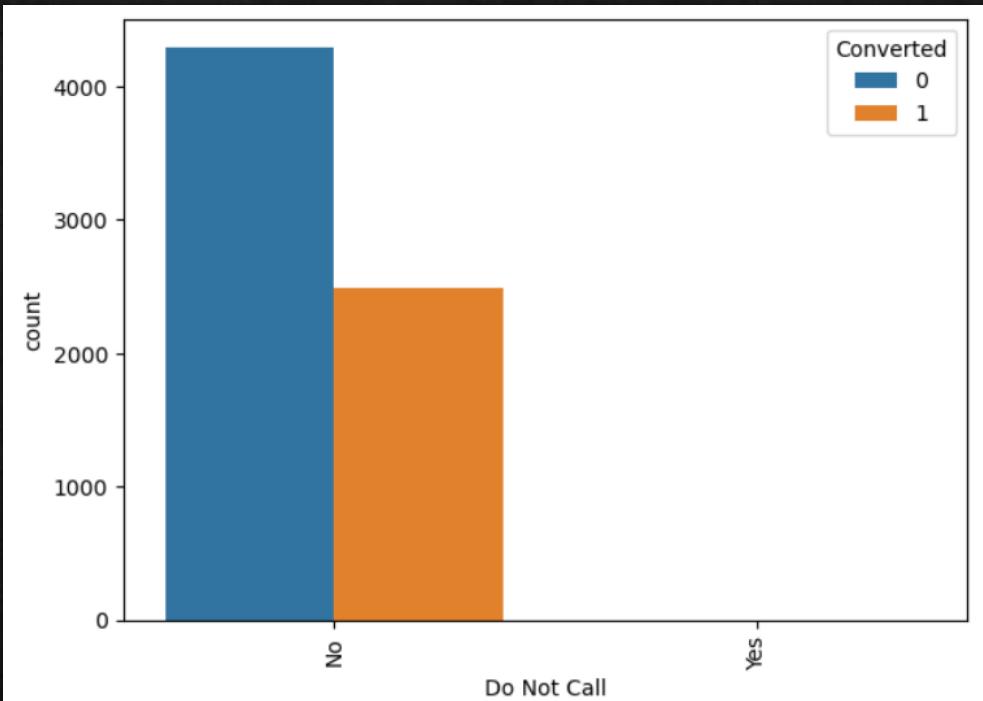


Unemployed Vs Converted

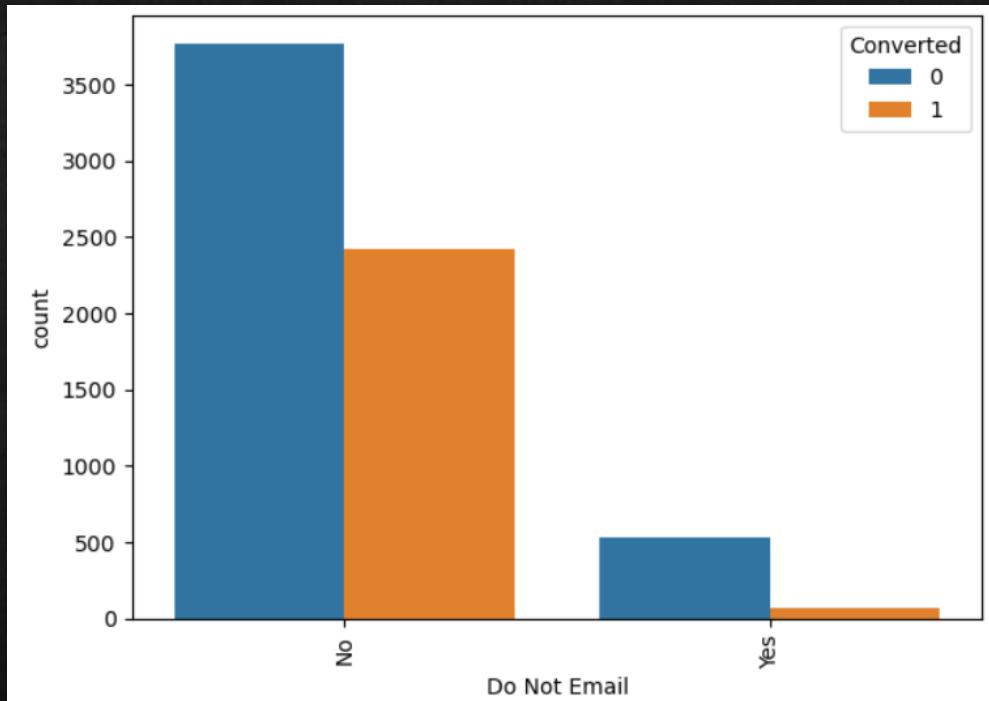


- Most leads are converted with messages
- Emails also induced leads.
- Working Professionals going for the course having the high chance of joining the course.
- Unemployed leads are the most in terms of Absolute numbers.

Do Not Call Vs Converted



Do Not Email Vs Converted

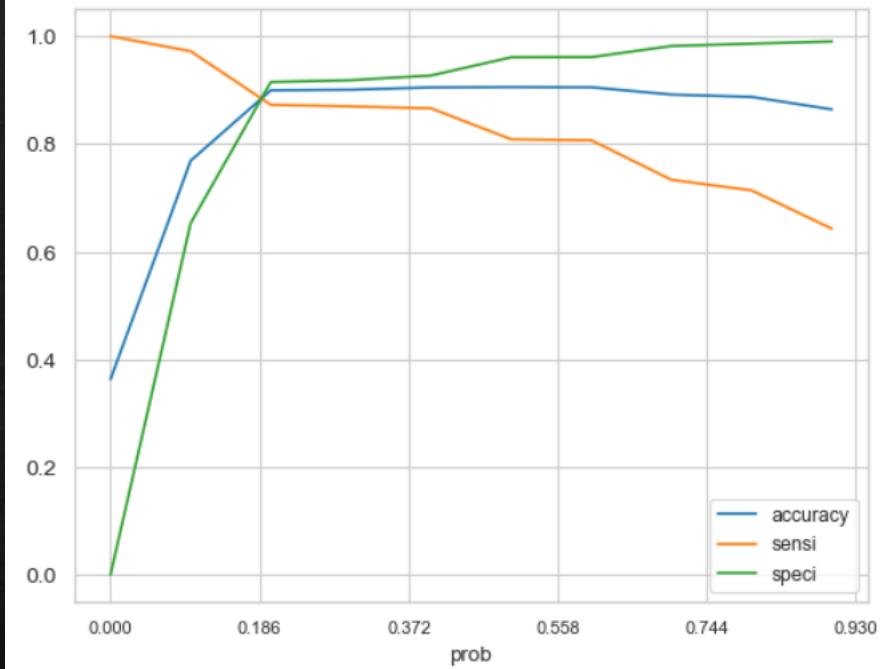


- Most leads do not prefer to inform thru phone.
- Most leads do not prefer to inform thru e-mail

Model Building

- ❖ Splitting into train and test set
- ❖ Scale variable in train set
- ❖ Build the first model
- ❖ Use RFE to eliminate less relevant variables
- ❖ Build the next model
- ❖ Eliminate variables based on high p-values
- ❖ Check VIF value for all the existing columns
- ❖ Predict using train set
- ❖ Evaluate accuracy and other metric
- ❖ Predict using test set
- ❖ Precision and recall analysis on test predictions

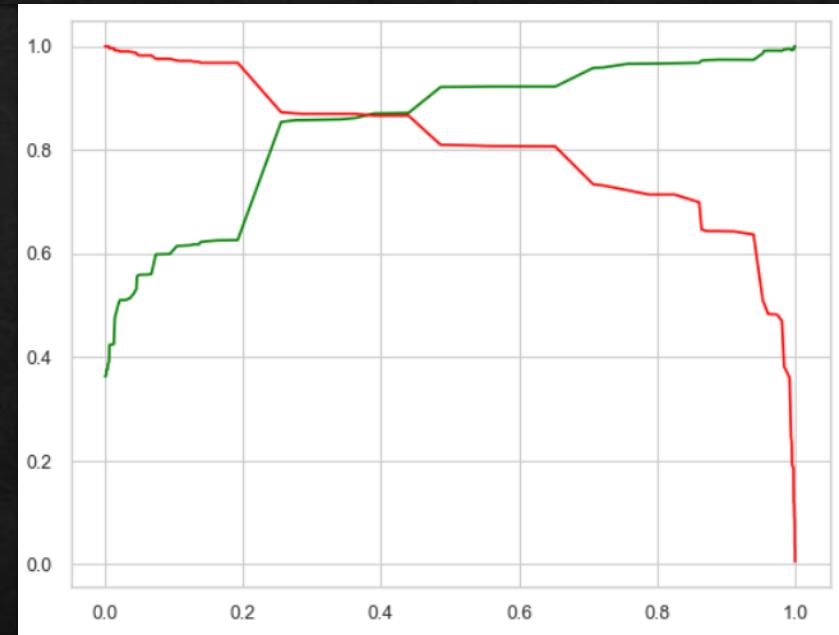
Model Evaluation (TRAIN)



Accuracy 77.83%

Sensitivity 96.79%

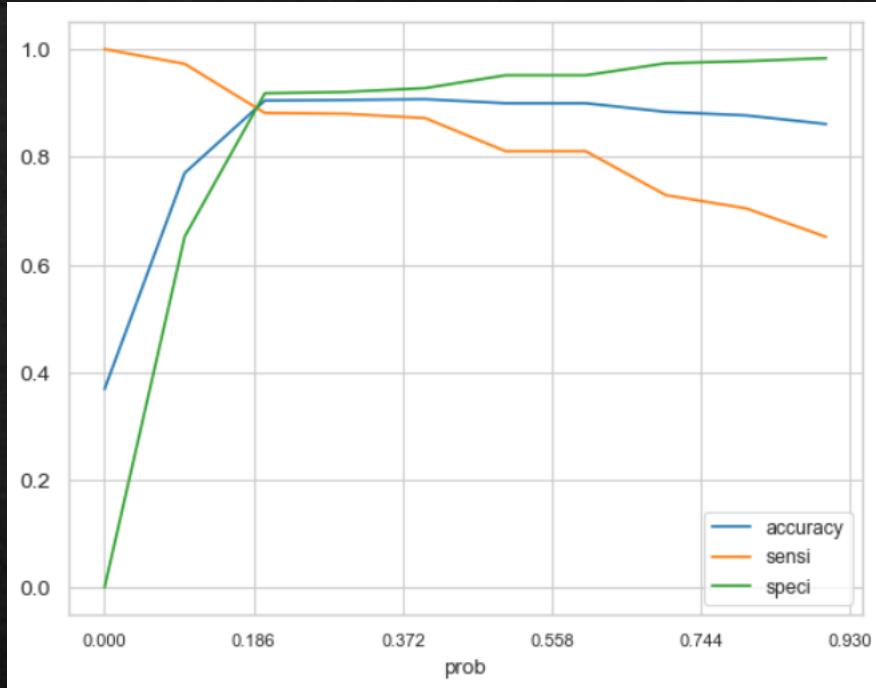
Specificity 67.03%



Precision 62.59%

Recall 96.79%

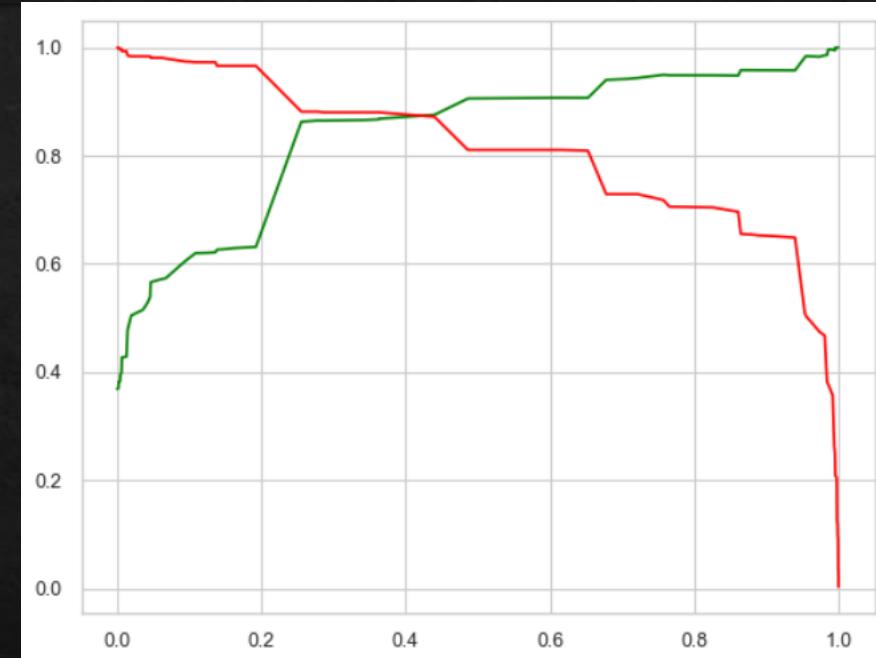
Model Evaluation (TEST)



Accuracy 77.96%

Sensitivity 96.59%

Specificity 67.09%



Precision 63.13%

Recall 96.59%

Conclusion-1

❖ EDA

- ❖ People spending higher than average time are promising leads, So, targeting them and approaching them can be helpful in conversion
- ❖ SMS messages can have a high impact on lead conversion
- ❖ Landing page submissions can help find out more leads
- ❖ Management resources has high conversion rates. People from these specializations can be promising leads
- ❖ References and offers for referring lead can be a good source for higher conversions
- ❖ An alert messages or information has seen to have high lead conversion rate

Conclusion-2

❖ Regression Model

- ❖ The model shows high close to 78% accuracy
- ❖ The threshold has been selected from accuracy, sensitivity, specificity measures and precision, recall curves
- ❖ The model shows close to 96% sensitivity and close to 67% specificity
- ❖ The model finds correct promising leads and leads that have less chances of getting converted
- ❖ Overall this model proves to be accurate.

Thank you,