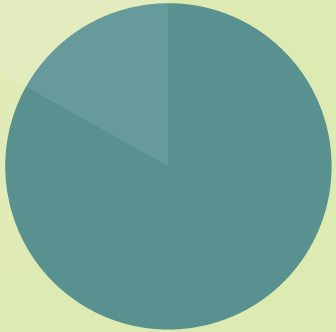
LEAD SCORE CASE STUDY



Logistic Regression Model

Problem Statement

.> **X Education is a company,Which provides online courses to industry professionals. many professionals who are interested in the courses land on their website and browse for courses like Google.**

Moreover, the X Education wants to select most promising leads that can be converted to paying customers .

Although The company generated a lots of leads but conversion rate is very poor

To make conversion process more efficient, the company wishes to identify the most potential leads,

Business Goal

The company requires a model to be built for selecting most promising leads

Company need to identify the most potential leads, also known as ‘Hot Leads’,where the conversion rate is higher.

So Company lead conversion rate will increase from 30% to 80% or more..

Strategy

Import Data

Clean and prepare the acquired data for further analysis.

Exploratory Data Analysis

Scaling Feature

Prepare the data for Model building

Build a logistic regression model

Assign a lead score for each leads

Test the model on train set

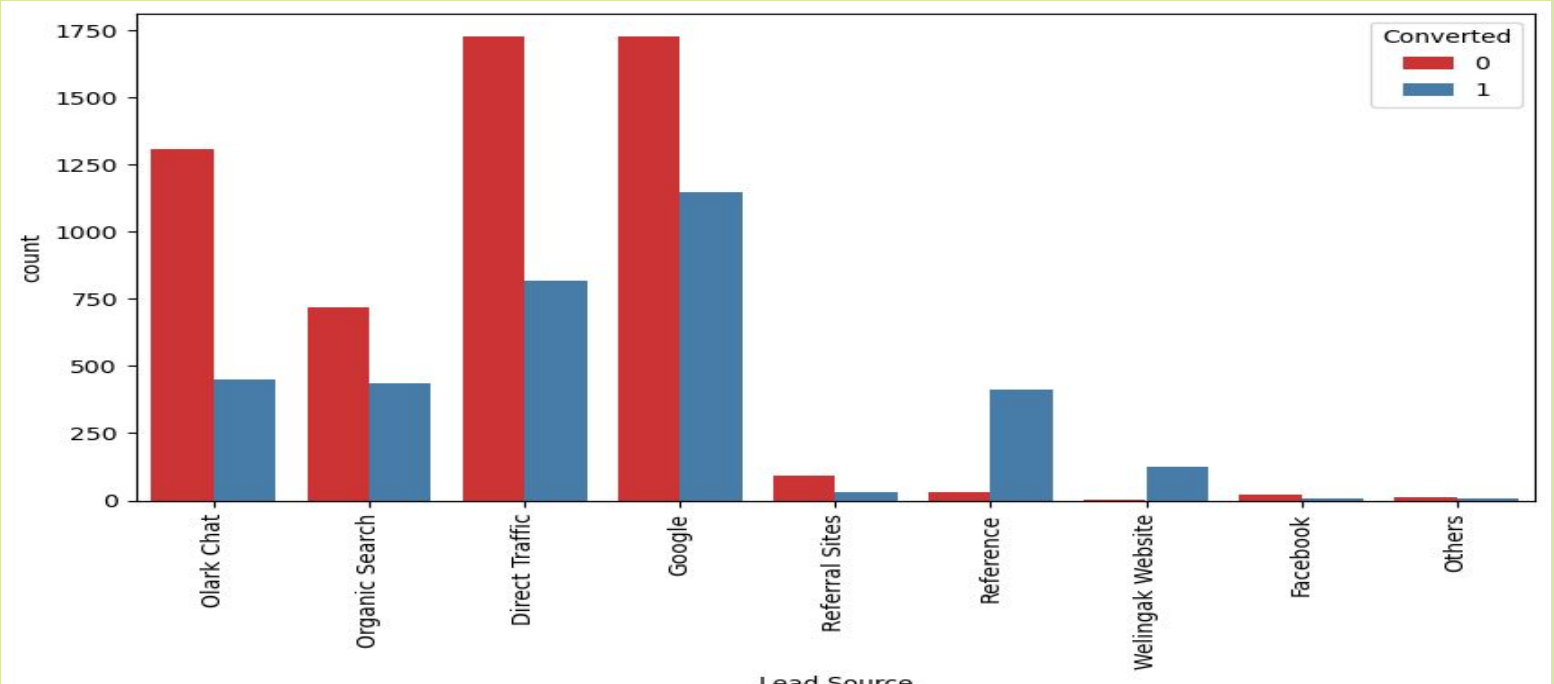
Evaluate model by different measure and metrics

Test the model on test set

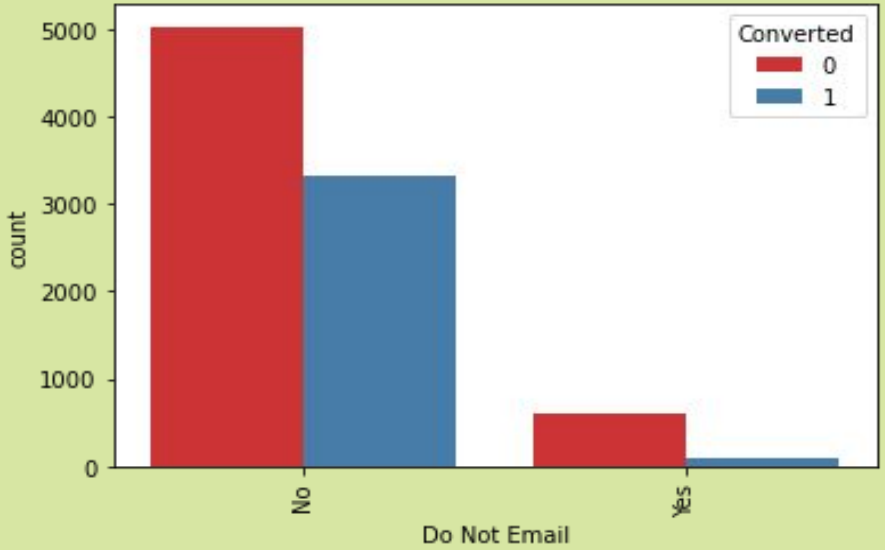
Measure the accuracy of the model

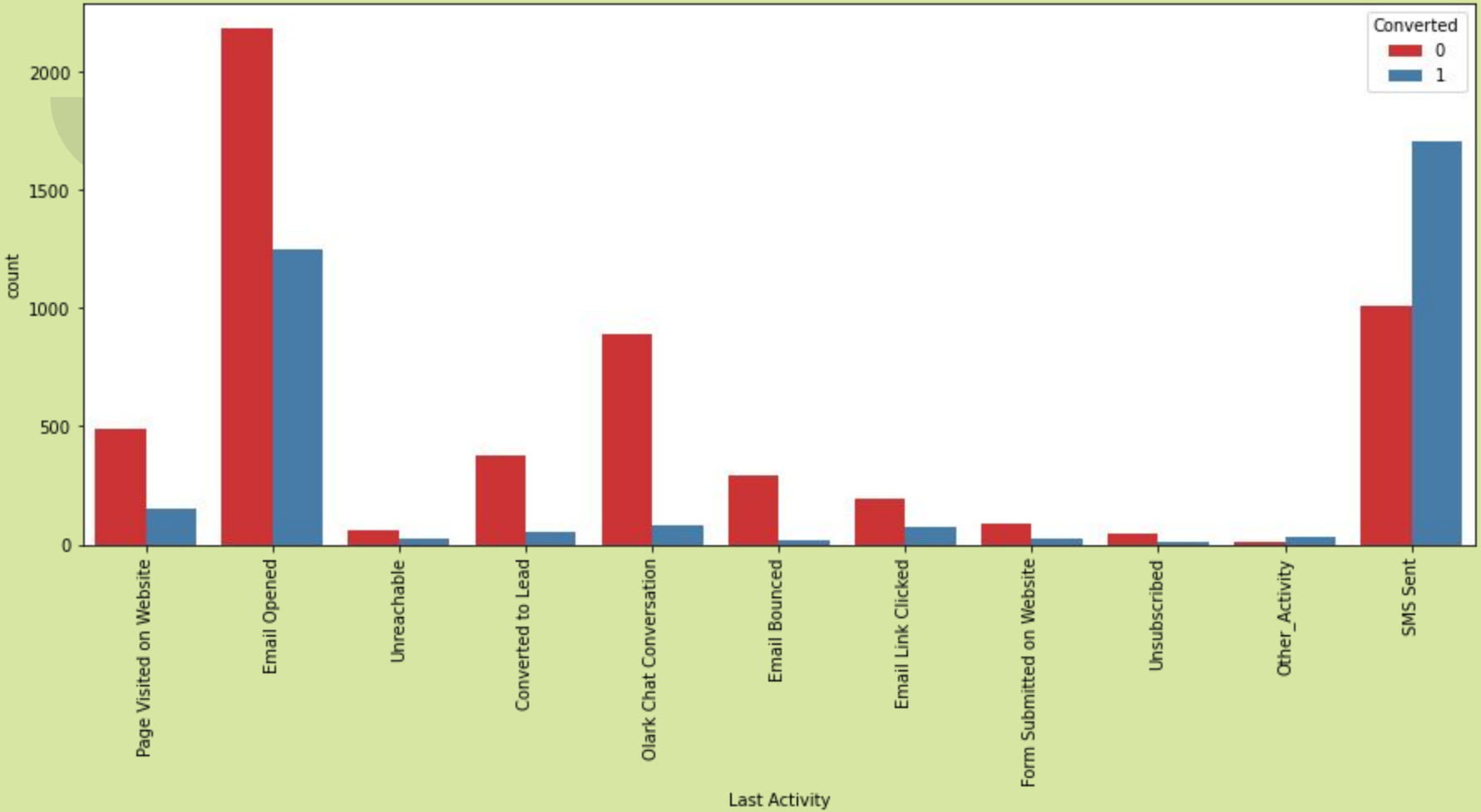
Exploratory Data Analysis

Lead source Vs Converted

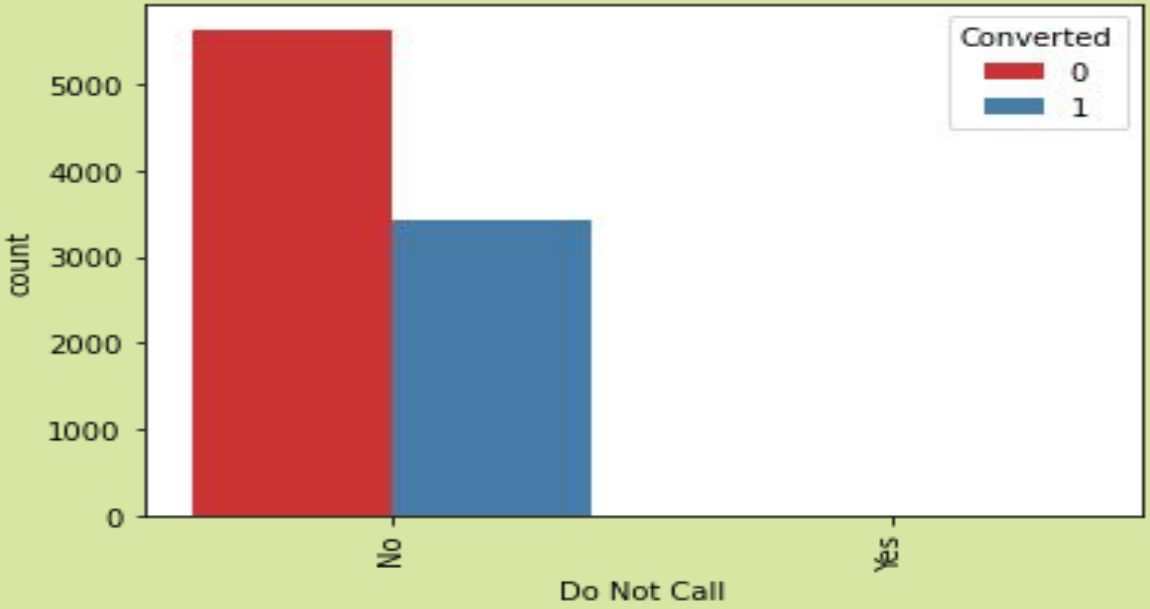


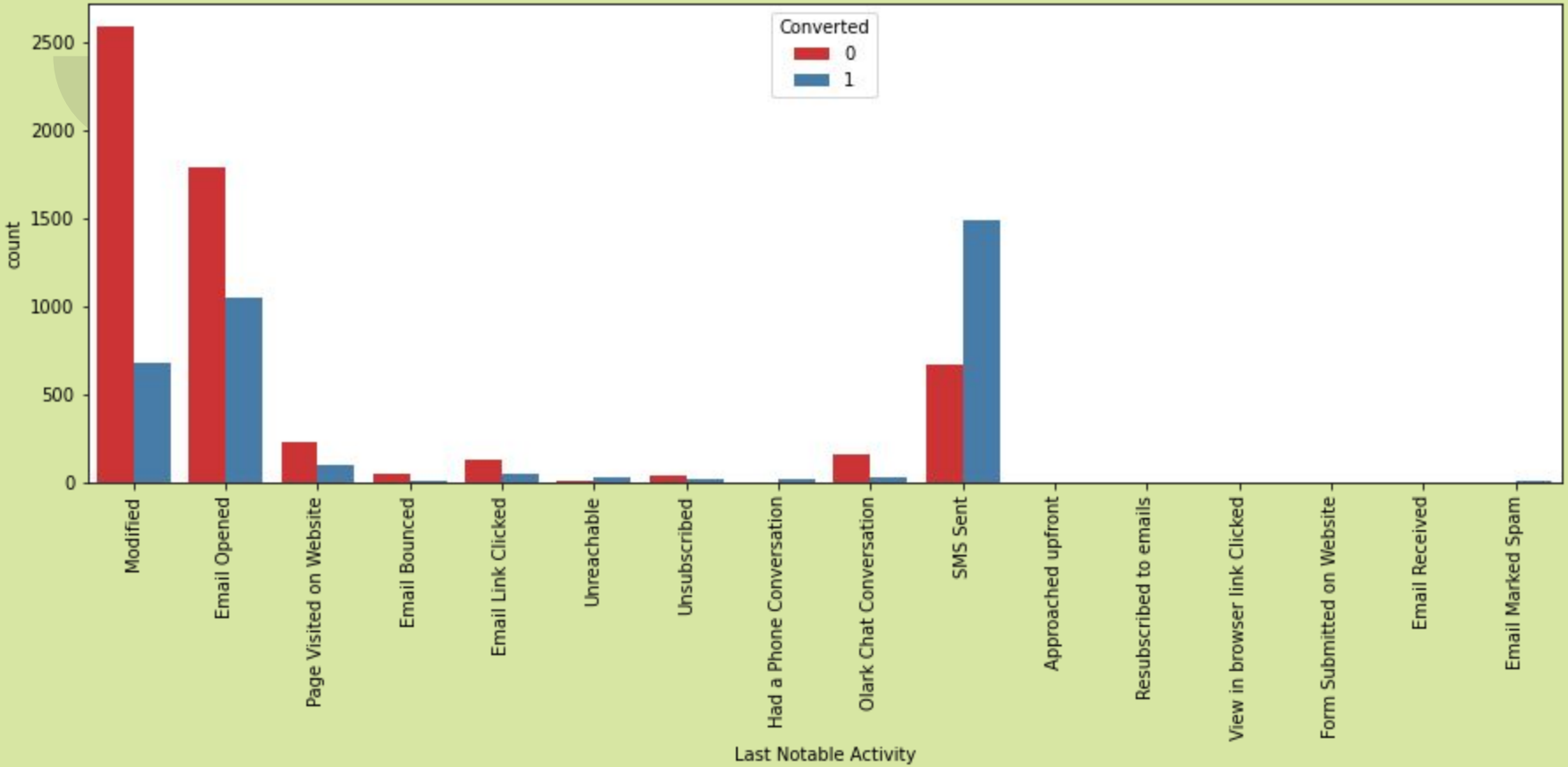
Do not Email Vs Converted

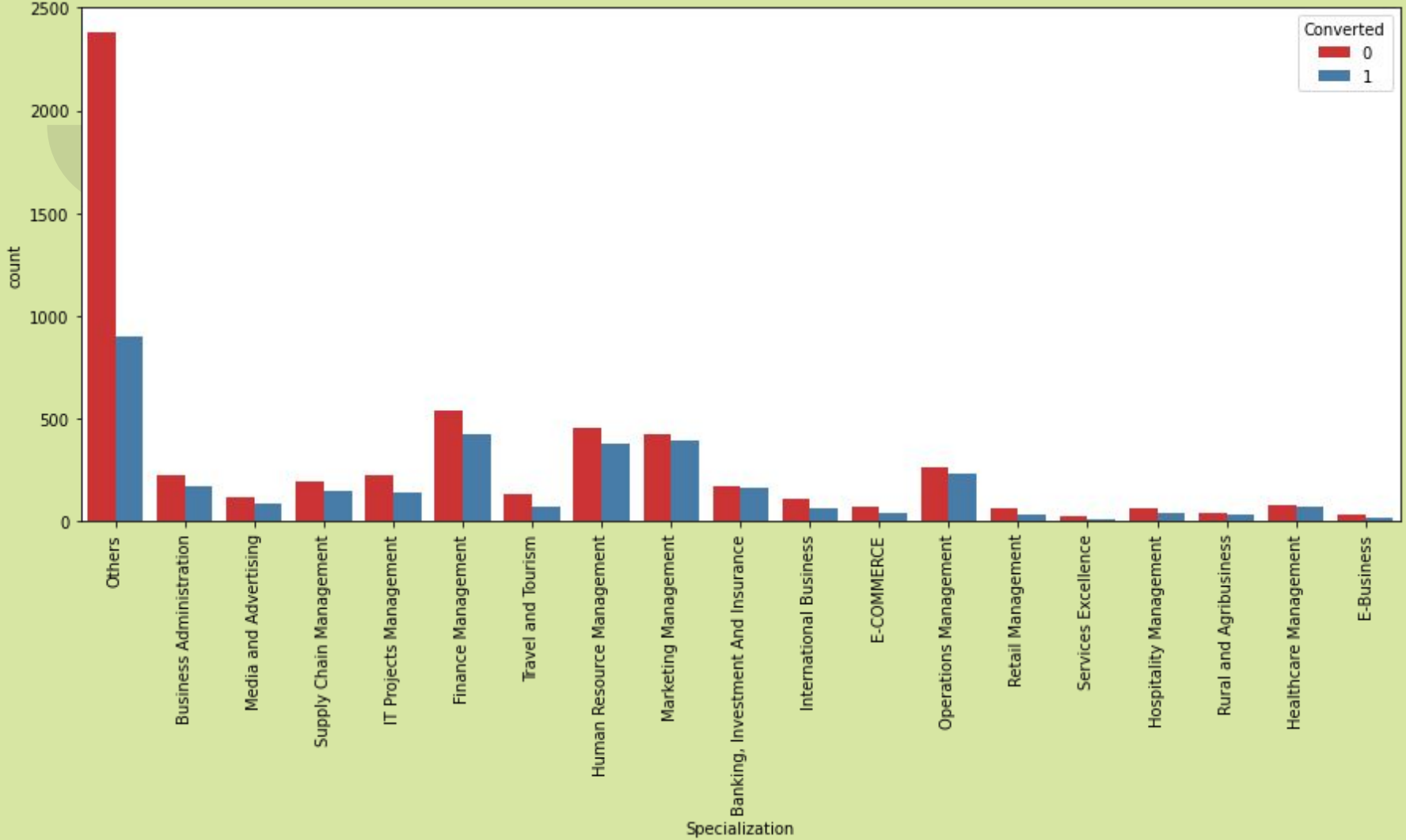



Converted
0
1
2000
1500
1000
500
0
count
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Other_Activity
SMS Sent
Last Activity

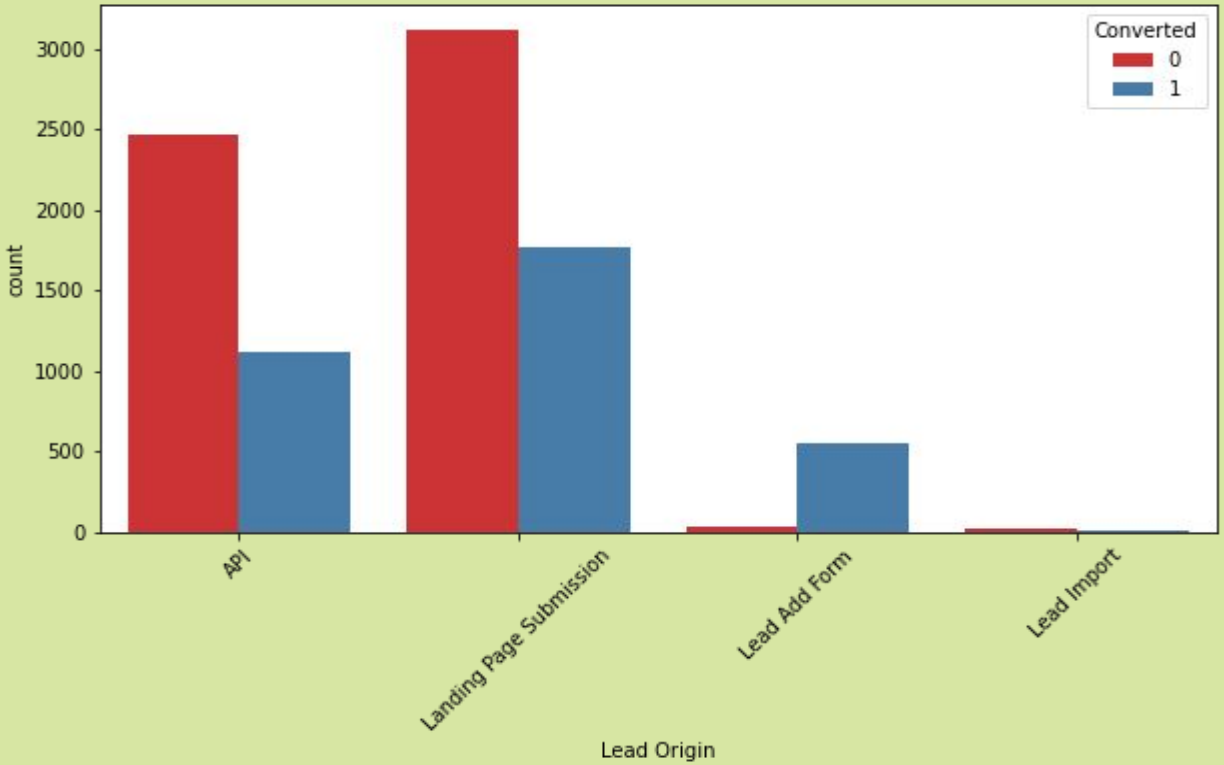
Do not call vs converted



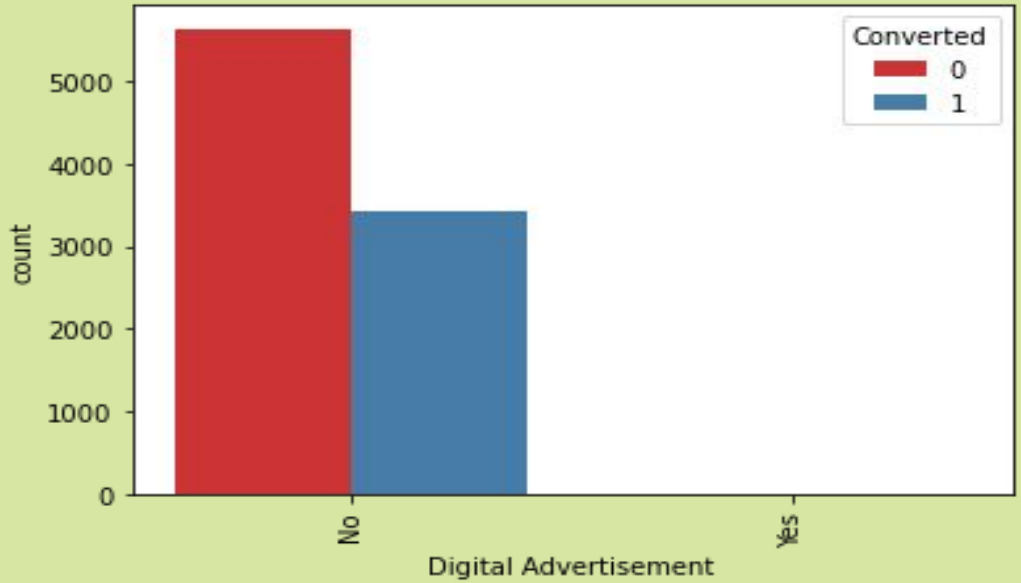

Converted
0
1
count
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
View in browser link Clicked
Form Submitted on Website
Email Received
Email Marked Spam
Last Notable Activity


Converted
0
1
count
2500
2000
1500
1000
500
0
Others
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

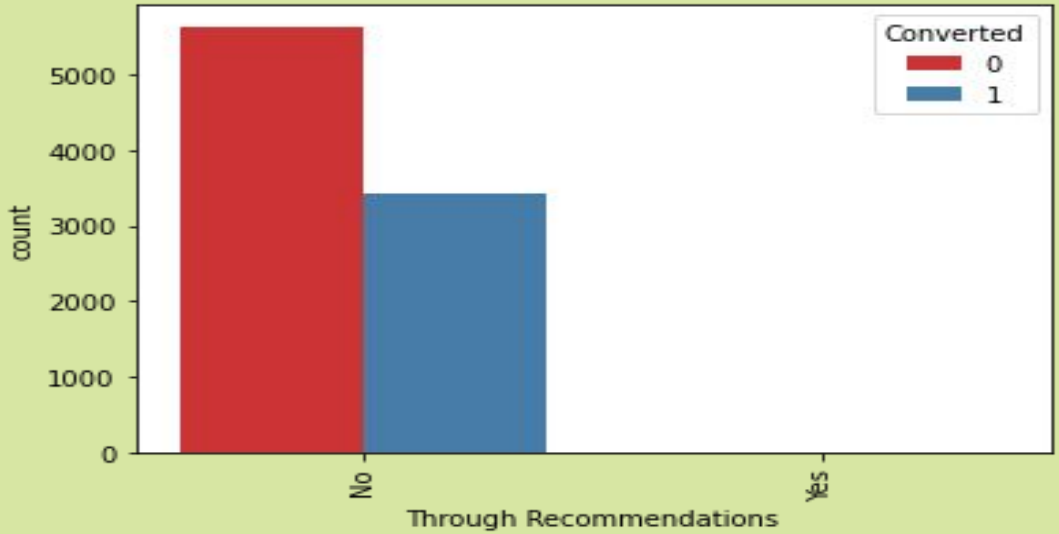
Lead origin VS Converted



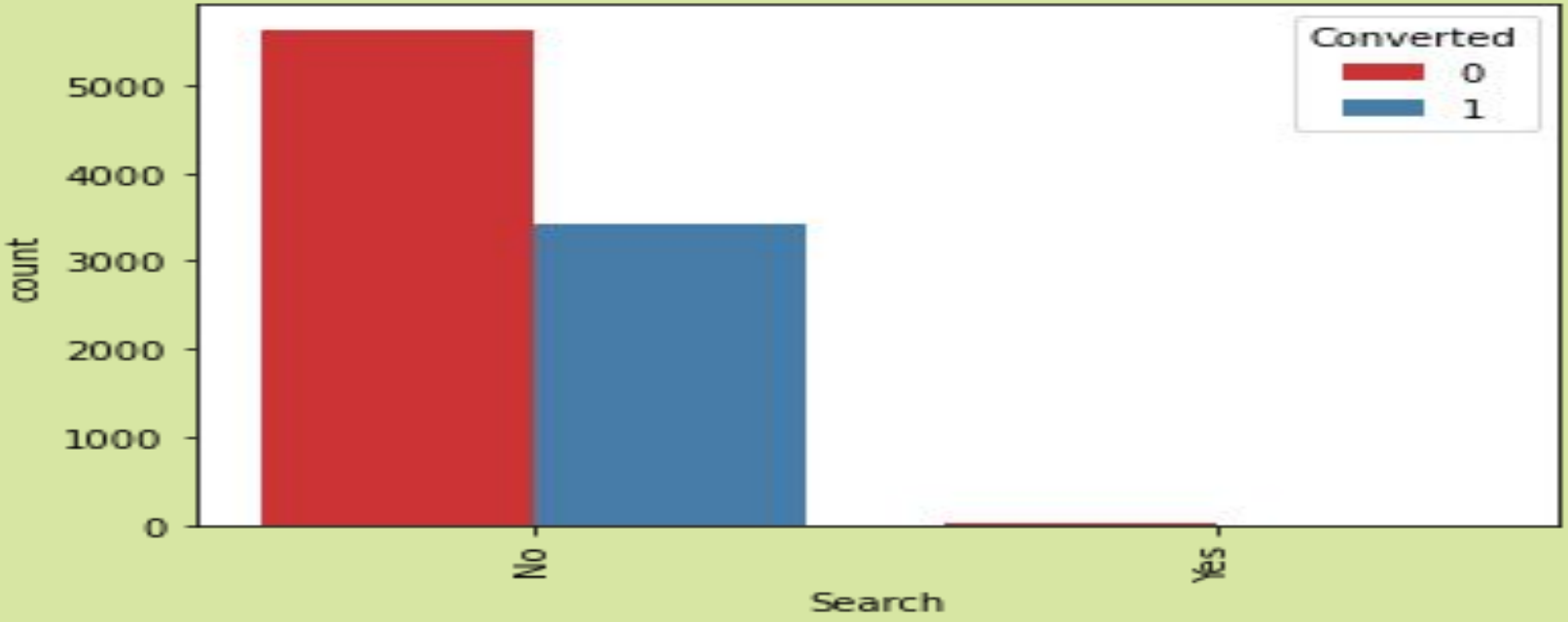
Digital Advertisement Vs Converted



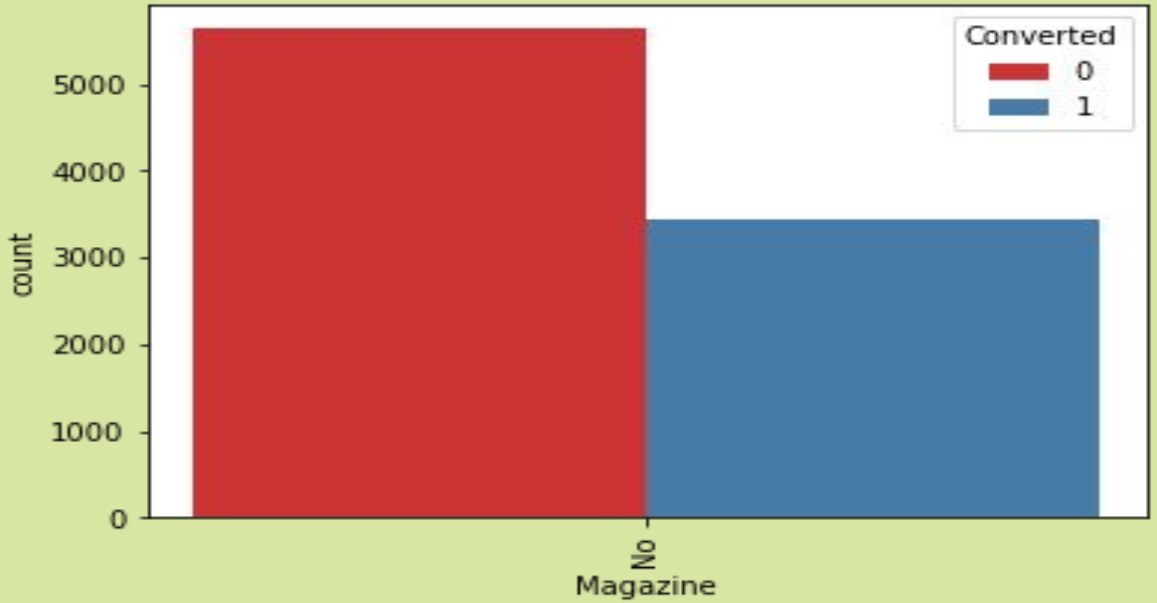
Through Recommendation Vs converted



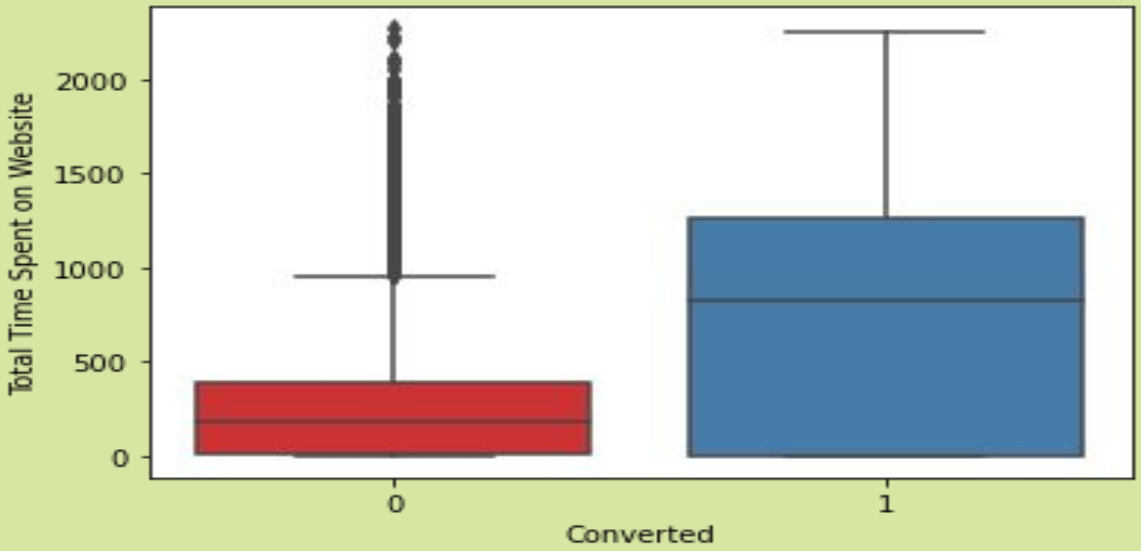
Search Vs Converted



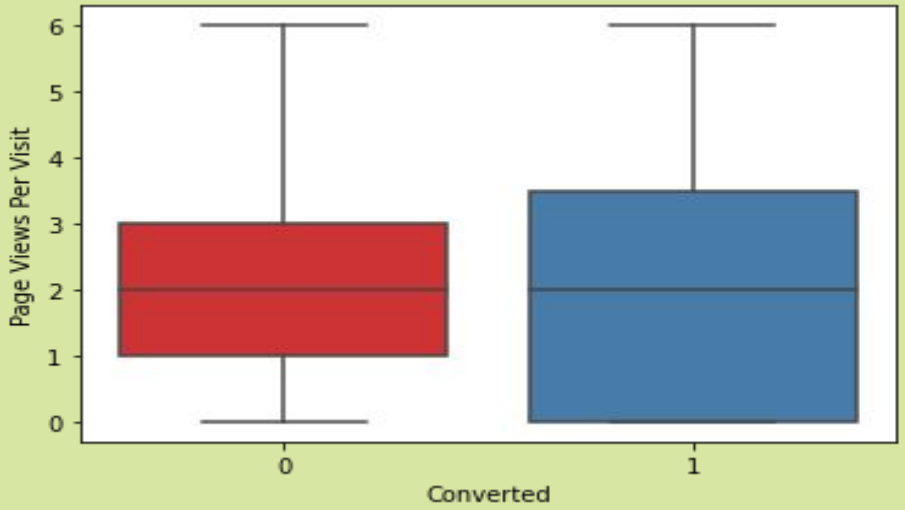
Magazine Vs Converted



Total Time spent on Website Vs converted



Total Visits Vs converted



Model Building

Splitting into train and test set

Scale variable in train test

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

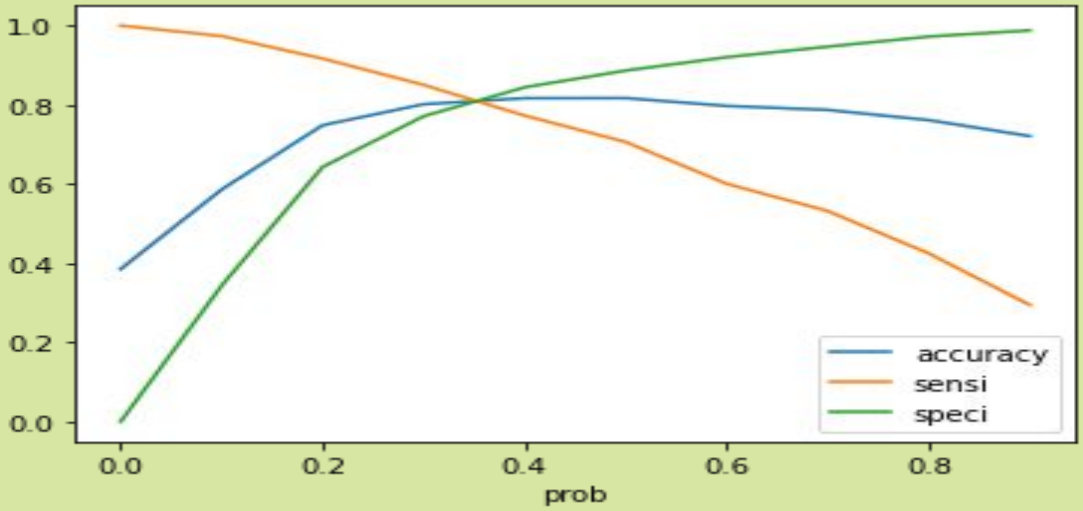
Predict using train set

Evaluate accuracy and other metric

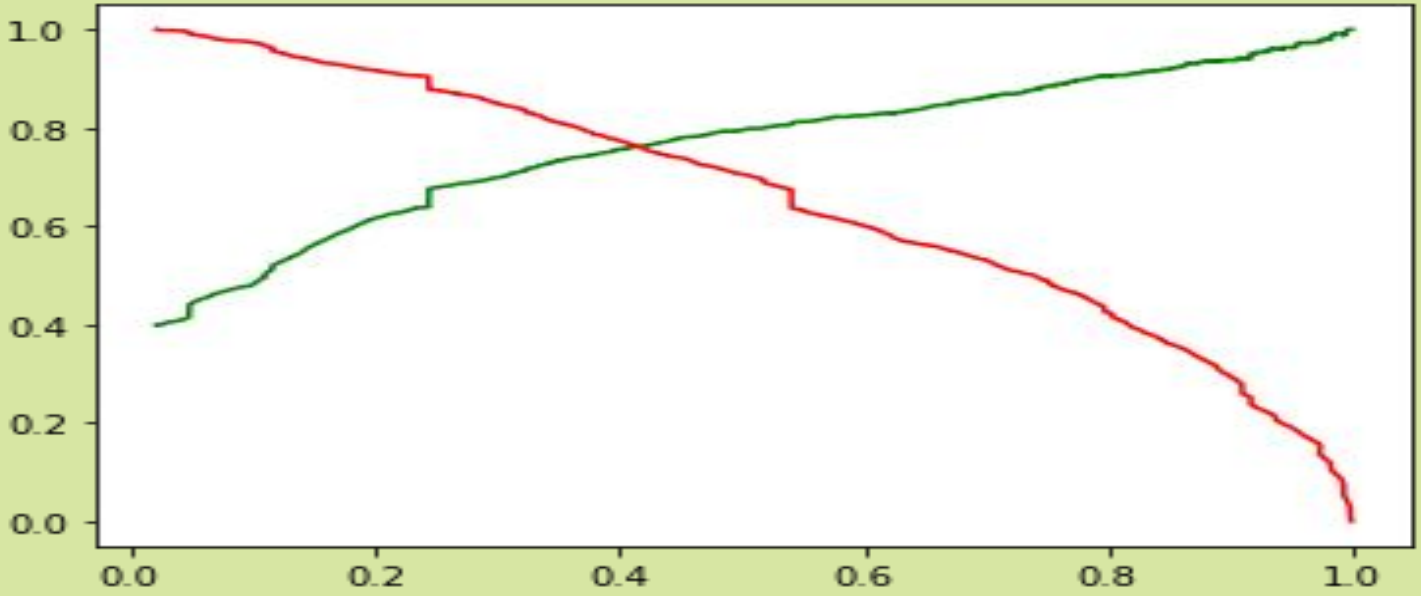
Predict using test set

Precision and Recall

Model Evaluation (TRAIN)



Precision & Recall



Conclusion

- 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 make calls** to the leads whose `last activity` was `SMS Sent` as they are more likely to get converted.
- The company **should not make calls** to the leads whose `last activity` was `"Olark Chat Conversation"` as they are not likely to get converted.
- The company **should not make calls** to the leads whose `lead origin` is `"Landing Page Submission"` as they are not likely to get converted.
- The company **should not make calls** to the leads whose `Specialization` was `"Others"` as they are not likely to get converted.