



Presentation 2023

LEAD SCORE CASE STUDY

Submitted by Anish, Adrish, Vinita



Content

01

Problem Statement

02

Goal

03

Problem Solving Approach

04

Visualization

05

Evaluation

06

Final features



Problem Statement

X Education company that sells online courses to industry professionals markets it's courses through several websites like Google.

A person is classified as a lead when he/she fills up a form providing his/her email address or phone number

Some of these leads get converted and some do not.
The conversion rate is only 30%

Goal

- To find hot leads that are most likely going to purchase their courses
- A model that assigns a score to each individual that indicates the probability of them getting converted
- The CEO has given the conversion rate target around 80%

Approach

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut enim rhoncus vel efficitur odio odio.

1

Data Sourcing

- Loading the data
- Initial inspection

2

Data Cleaning

- Remove duplicates
- Handling missing values
- Outlier treatment

3

Data Preparation and EDA

- Standardization
- Univariate and Bivariate Analysis
- Scaling the numeric features

4

Model Building

- Splitting the data into train and test
- Feature selection
- Build an optimal model
- Evaluation (accuracy, sensitivity etc)

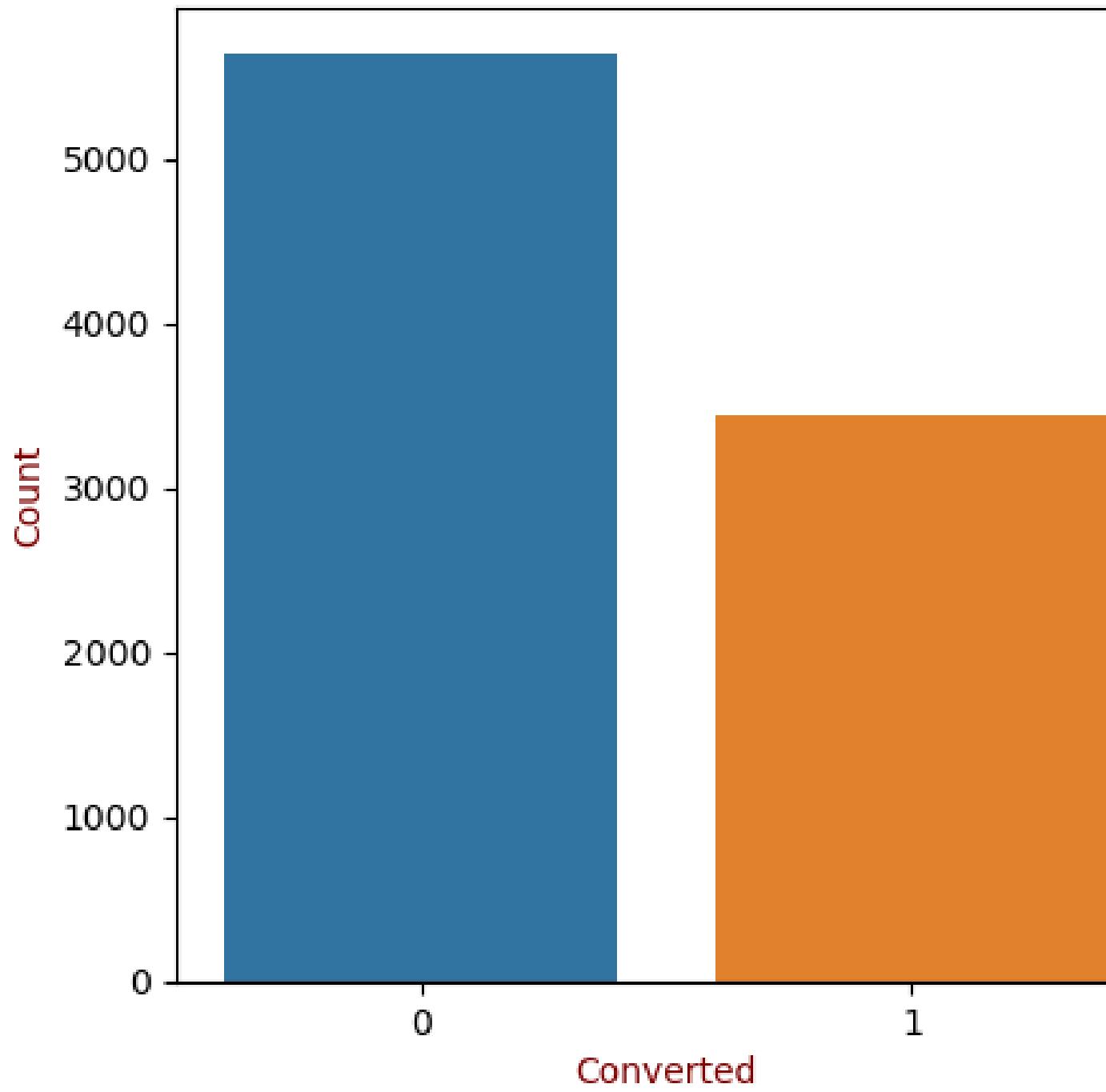
5

Result

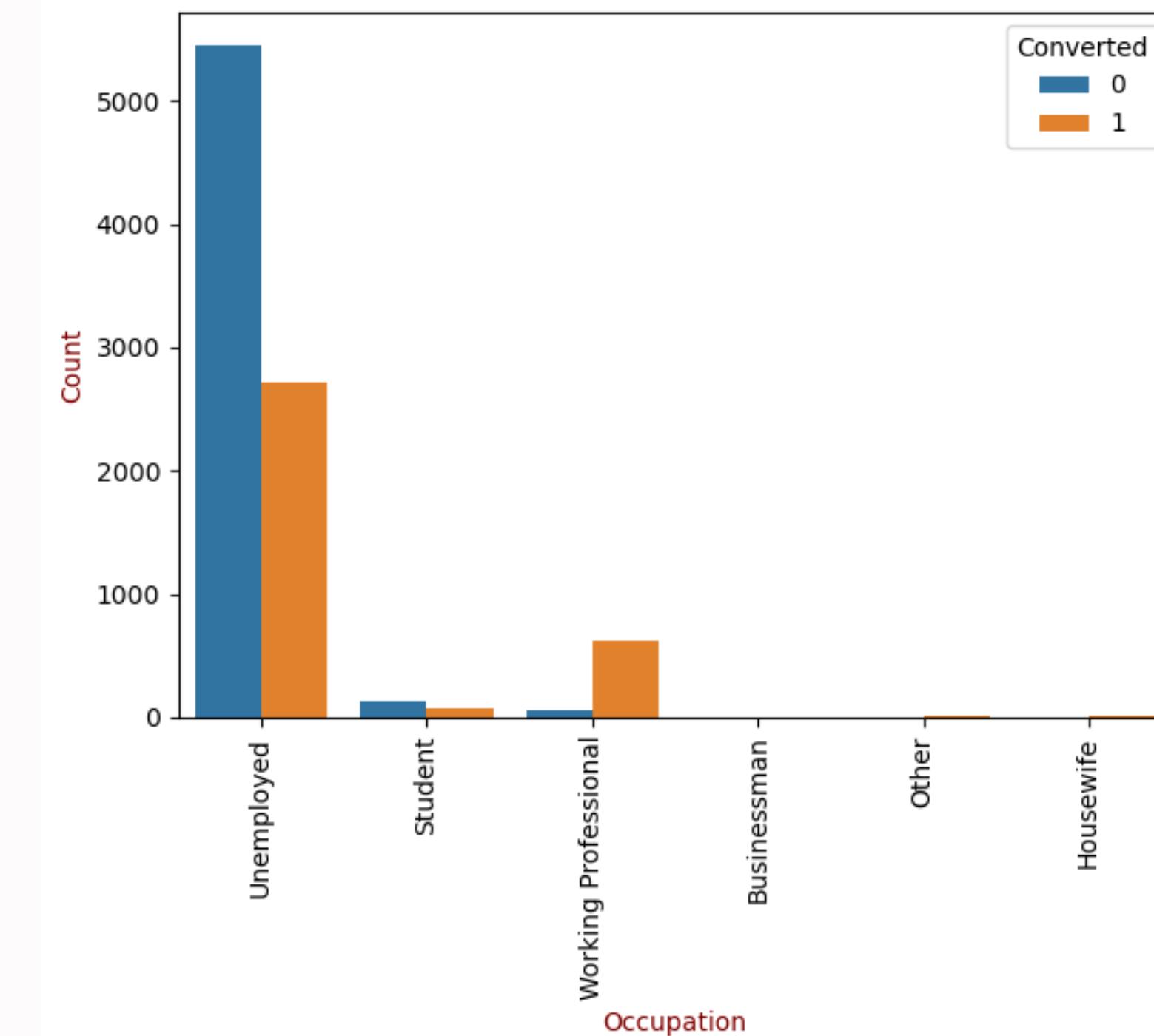
- Assign a lead score of individual customer
- Filter the data to get hot leads ($\text{lead_score} > 85$)
- Evaluation on the test dataset (Accuracy, Sensitivity, Specificity, Precision & recall)

Exploratory Data Analysis

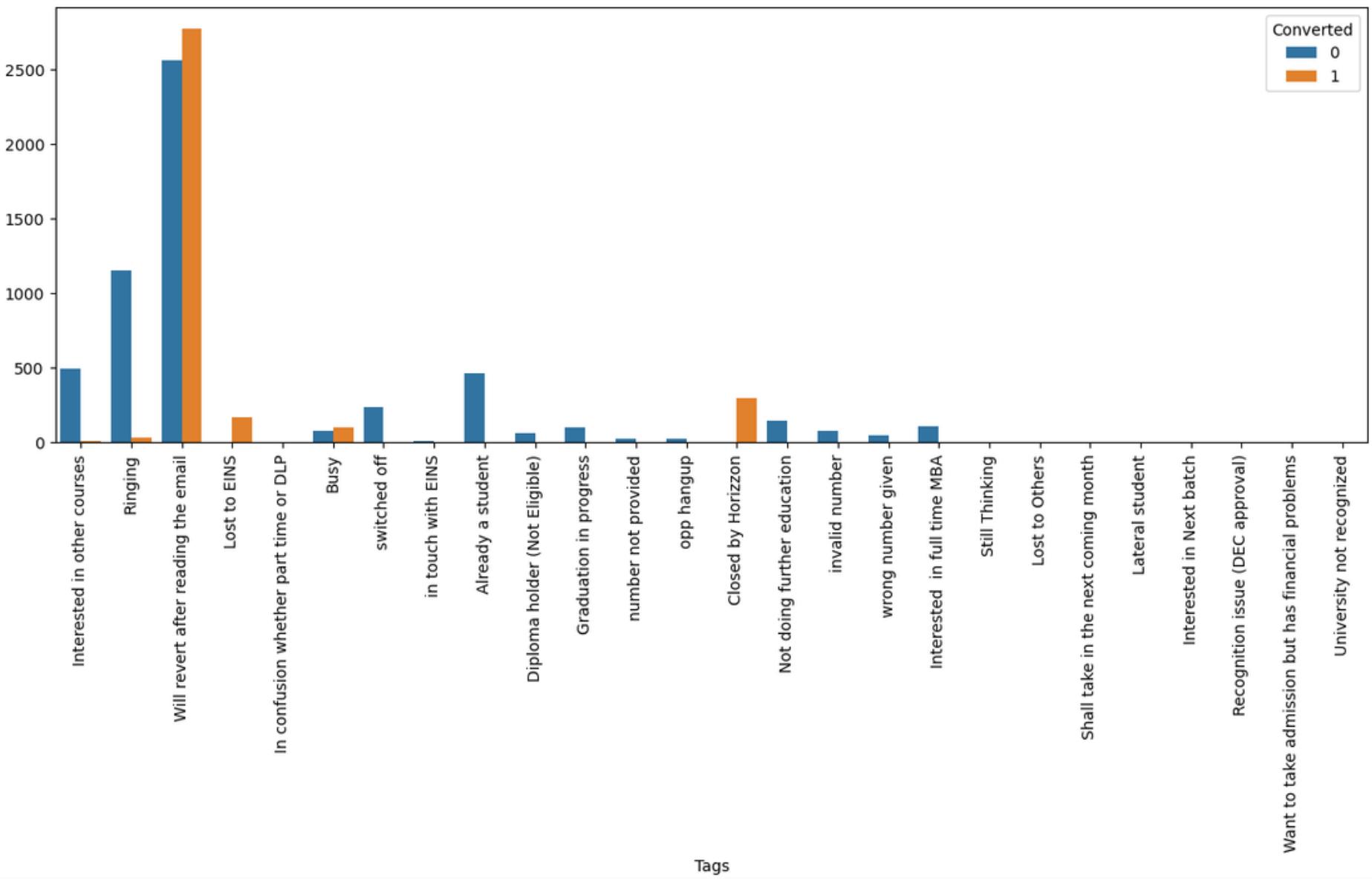
There are about 39% conversion rate



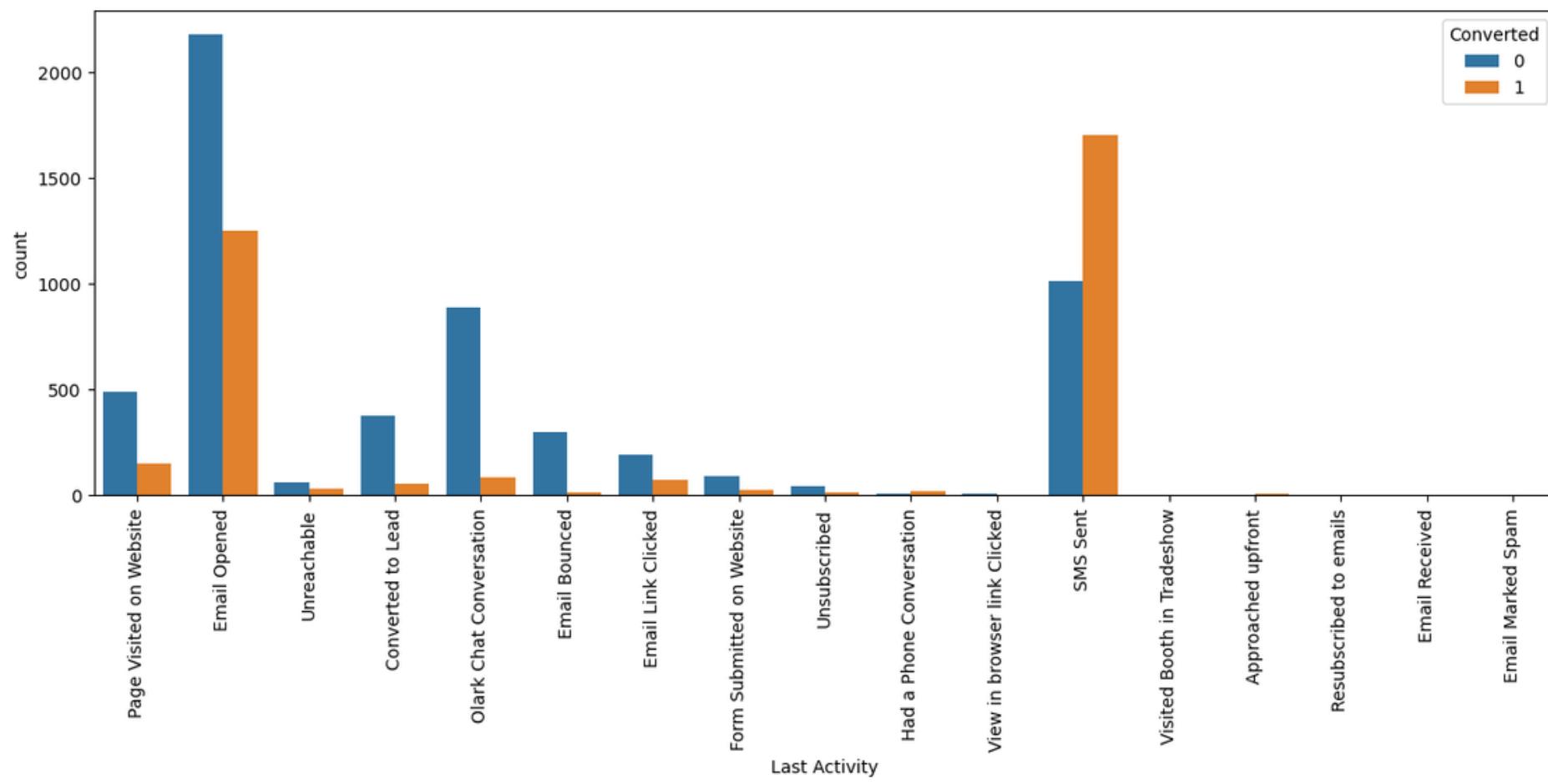
We can observe high conversion rate in the Working professions



High conversion rate was found with tags
 'will revert after reading the email' & 'closed by horizon'

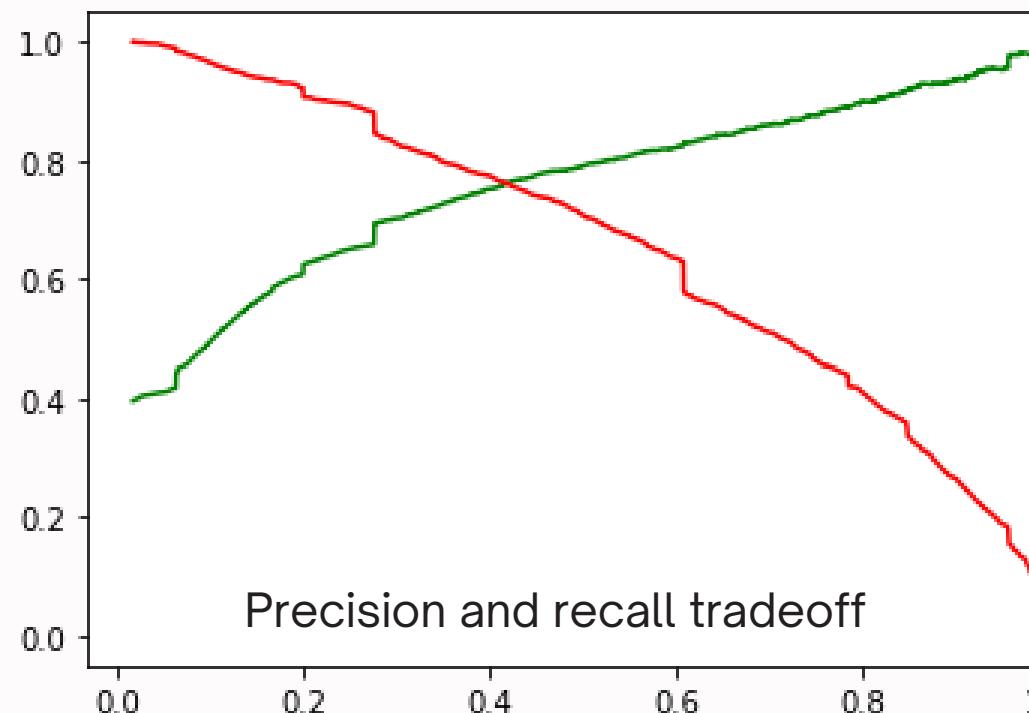
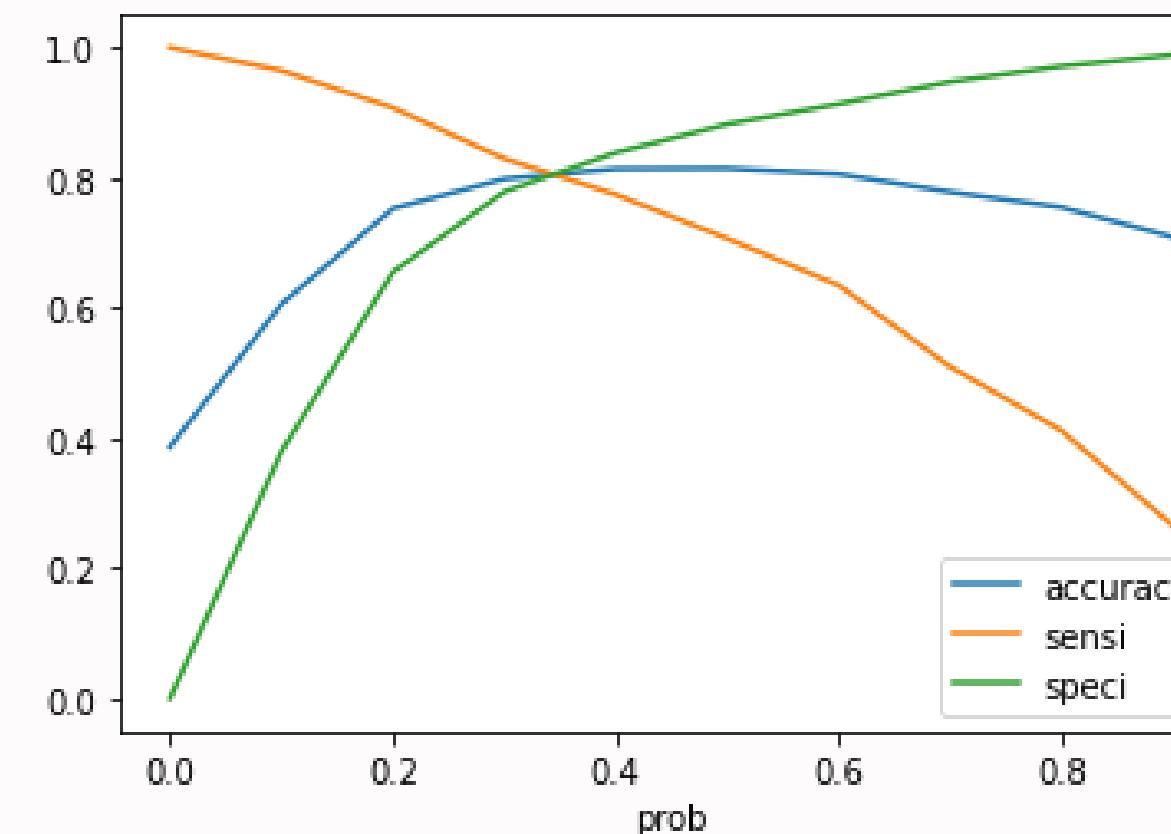


High conversion rate was found with last activity
 'SMS' sent

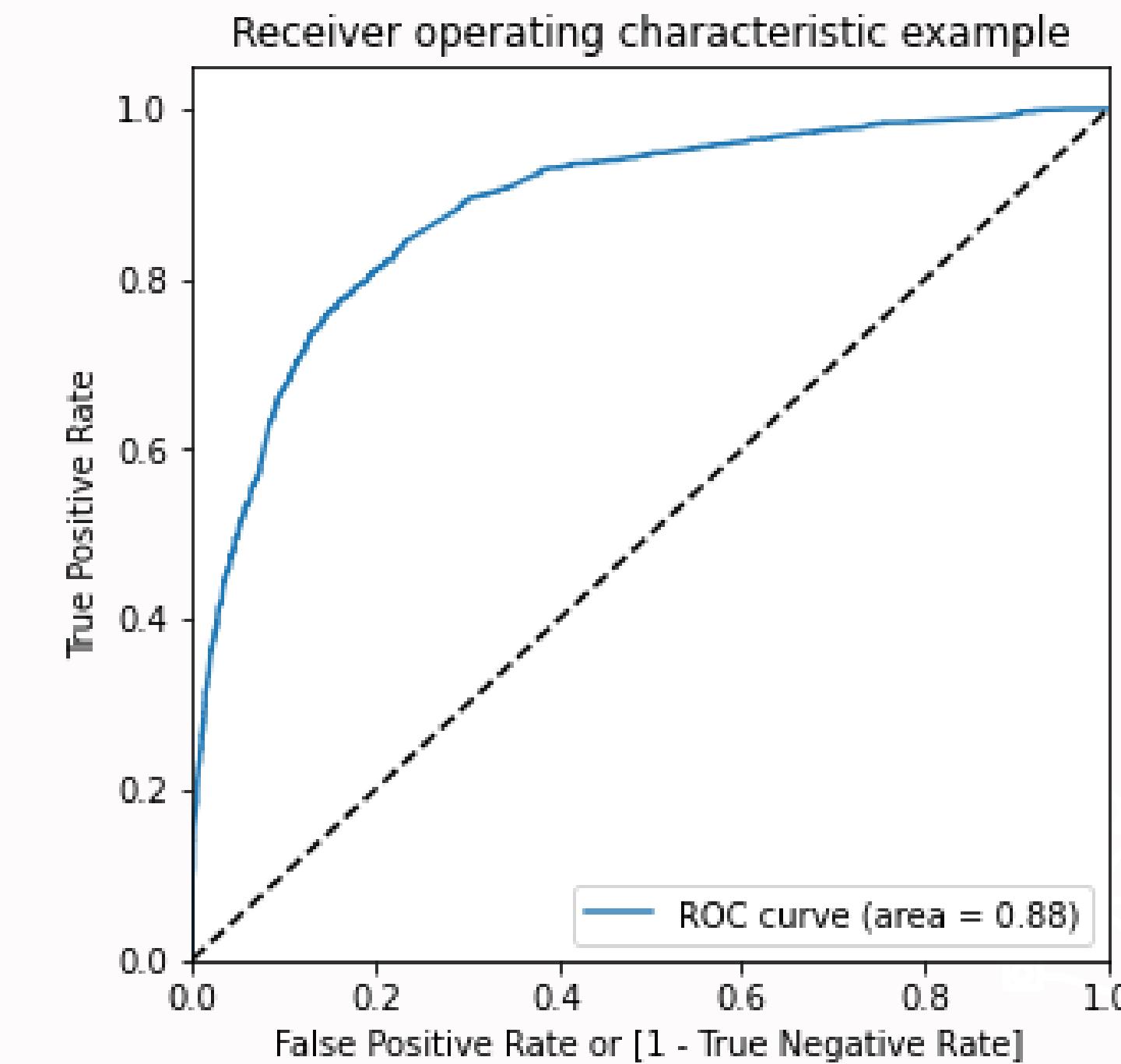


After model building visualization

Cut off value is taken as 0.35



After plotting ROC. area under the curve is 0.88



Model Evaluation

Train dataset

Accuracy: 81.24%

Sensitivity: 71.17%

Specificity: 87.59%

Precision: 74.72%

Recall: 76.91%

Test dataset

Accuracy: 80.94%

Sensitivity: 81.20%

Specificity: 79.47%

Precision: 72.11%

Recall: 76.60%



Final Set of features

- Lead Source_Google
 - Lead Source_Direct Traffic
 - Lead Source_Organic Search
 - Total Time Spent on Website
 - Last Notable Activity_Modified
 - Do Not Email
 - Last Activity_Email Bounced
 - Last Activity_Olark Chat Conversation
 - Last Notable Activity_Email Opened
 - Lead Origin_Lead Add Form
 - Last Notable Activity_Olark Chat Conversation
 - Lead Source_Welingak Website
 - Last Notable Activity_Page Visited on Website
 - What is your current occupation_Working Profesional
 - Lead Source_Referral Sites
 - Last Notable Activity_Email Link Clicked
 - Last Activity_Had a Phone Conversation
- 

THANK YOU

