LEAD SCORING CASE STUDY

Submitted By: Abhay Saxena & Roopali Misra

Problem Statement

- To help X Education to select the most promising leads(Hot Leads), i.e. the leads that are most likely to convert into paying customers.

- To build a logistic regression model to assign a lead score value between 0 and 100 to each of the leads which an be used by the company to target potential leads.

This includes:

- Create a Logistic Regression model to predict the Lead Conversion probabilities for each lead.

- Decide on a probality threshold value above which a lead will be predicted as converted, whereas not converted if it is below it.

Approach

➤ 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.

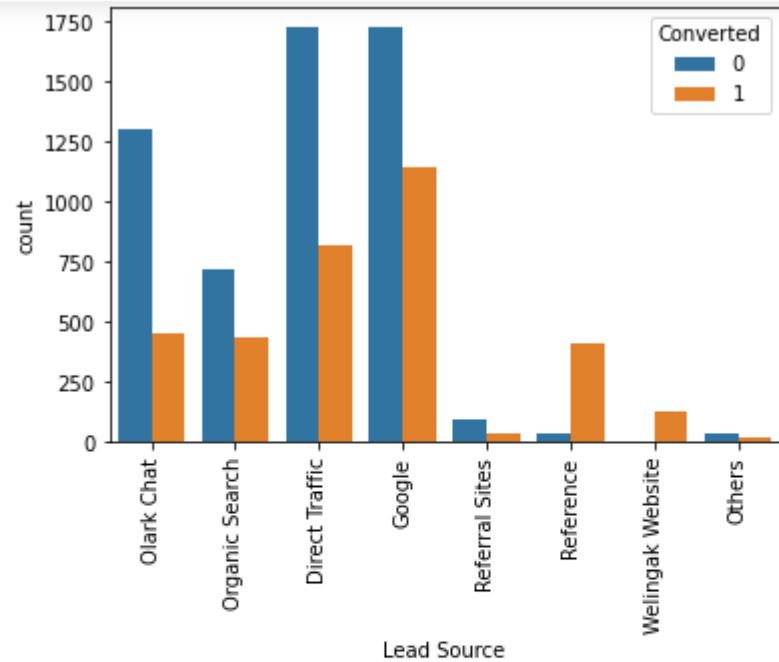➤ Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

EDA: Based on lead origin
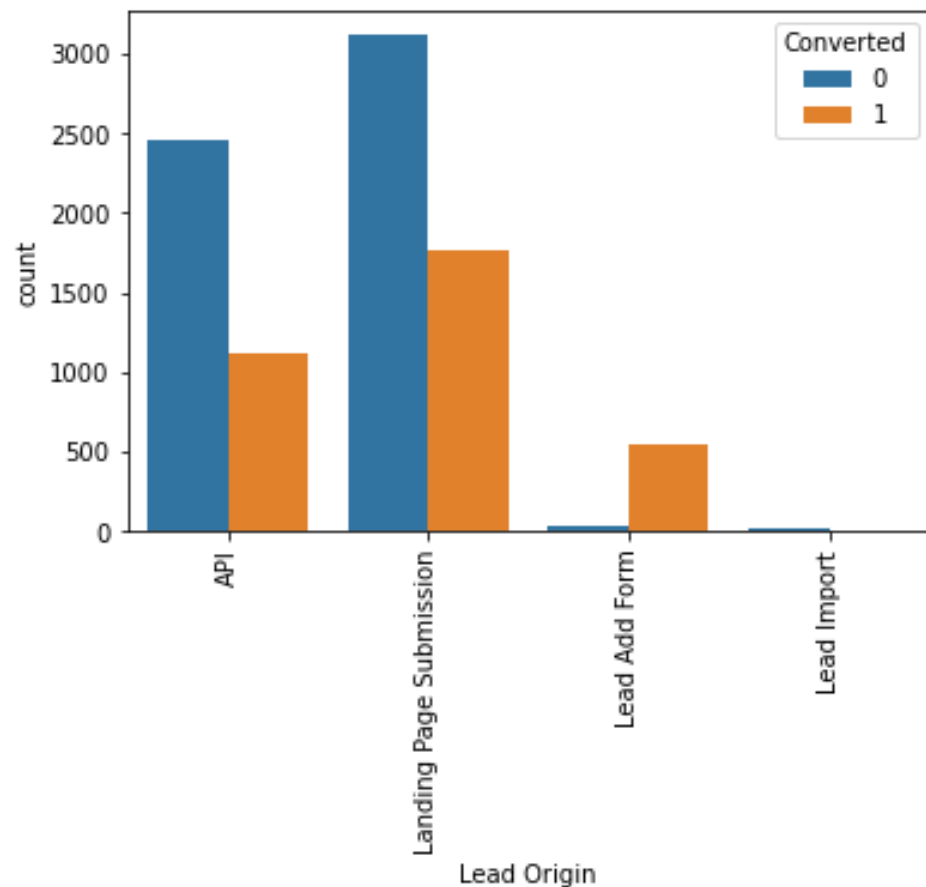
Inference

1. API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
2. Lead Add Form has more than 90% conversion rate but count of lead are not very high.

EDA: Based on lead source

Inference

1. Google and Direct traffic generates maximum number of leads.
2. Conversion Rate of reference leads and leads through welingak website is high.

To improve overall lead conversion rate, focus should be on improving lead converion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.
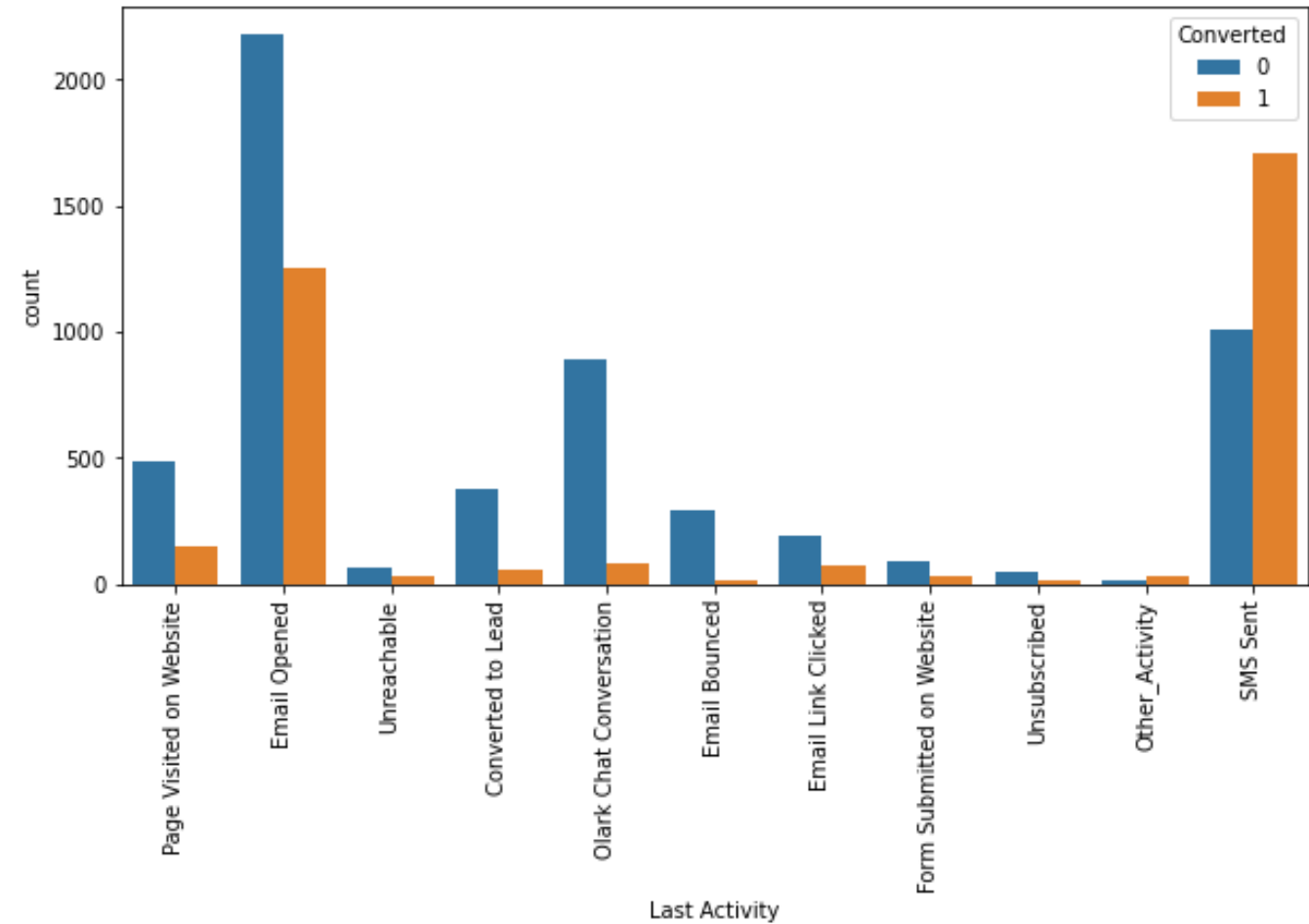
EDA: Based on time spend on website and last activity
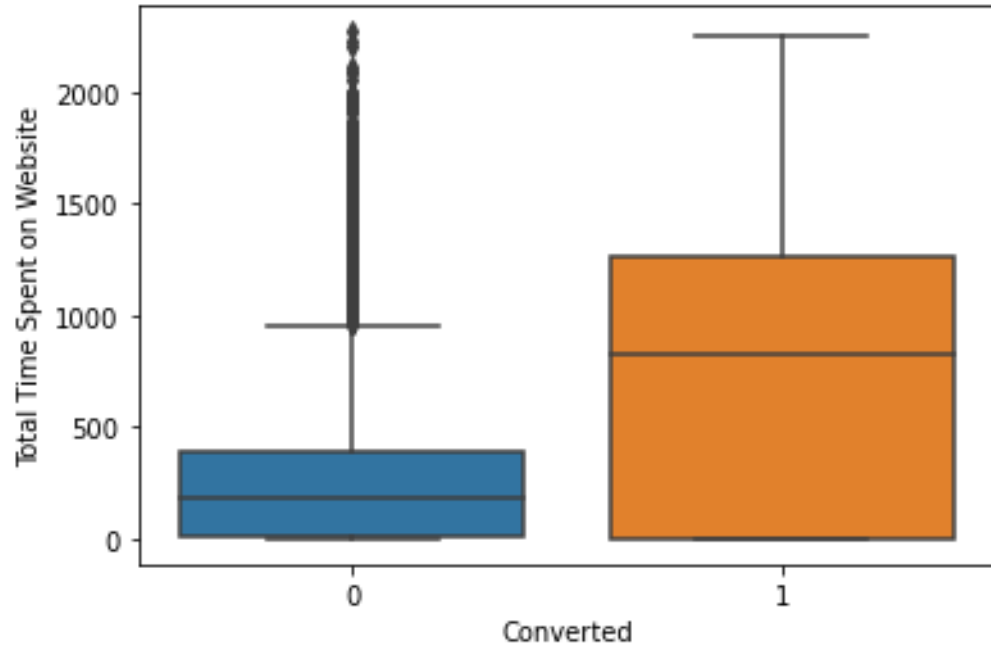
1. Leads spending more time on the weblise are more likely to be converted.

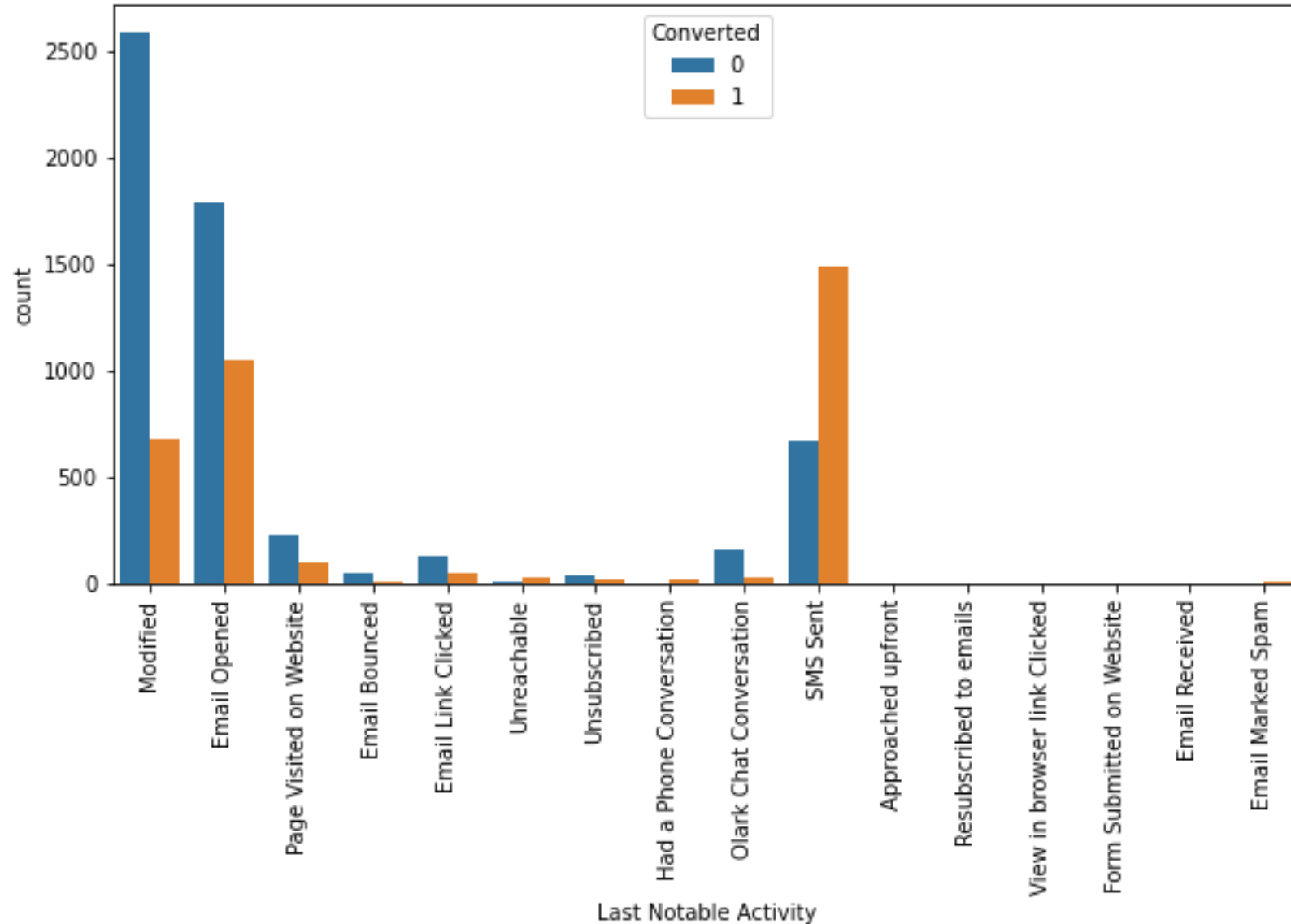**Website should be made more engaging to make leads spend more time.**

Inference

1. Most of the lead have their Email opened as their last activity.
2. Conversion rate for leads with last activity as SMS Sent is almost 60%.
3. Olark Chat also generates significant leads but conversion is poor

EDA: Based on tags
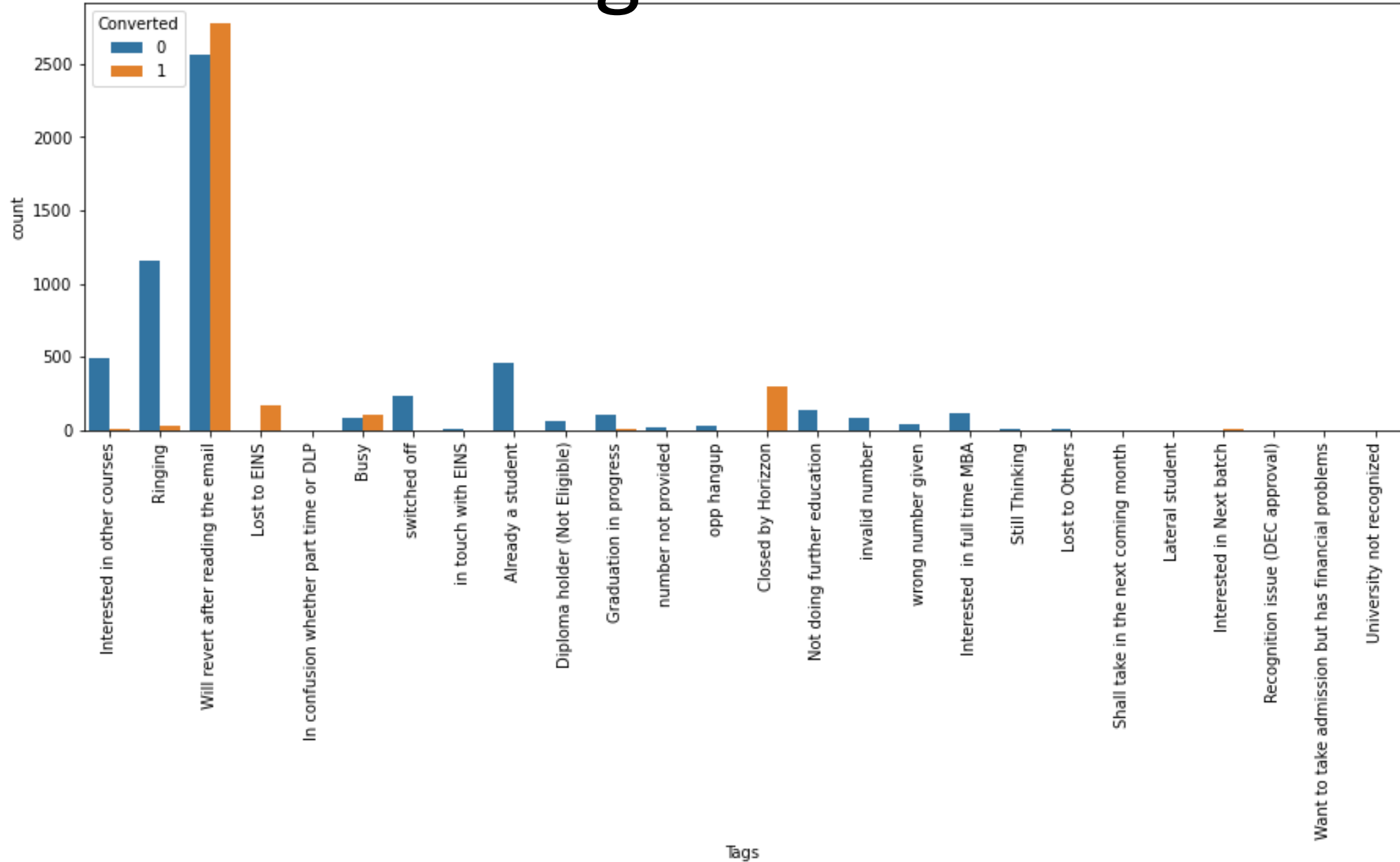
EDA: Based on Last Notable Activity

1. Most of the lead have their Email opened as their last activity.
2. Conversion rate for leads with last activity as SMS Sent is almost 60%.
3. Olark Chat also generates significant leads but conversion is poor

Model Evaluation -Sensitivity and Specificity on Train Data Set
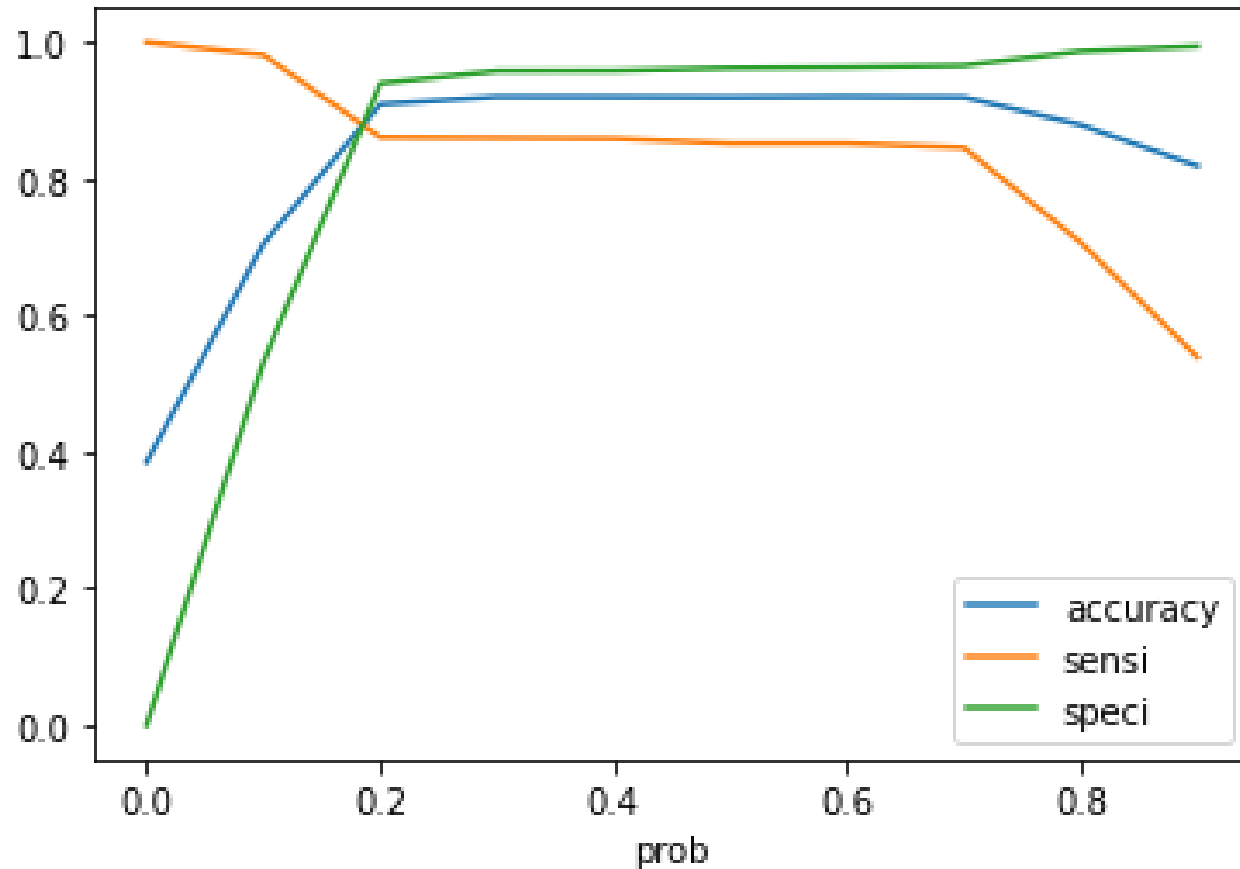
- Accuracy-90%
- Sensitivity-86%
- Specificity-93%
- FalsePositiveRate-6%
- PositivePredictiveValue-89%
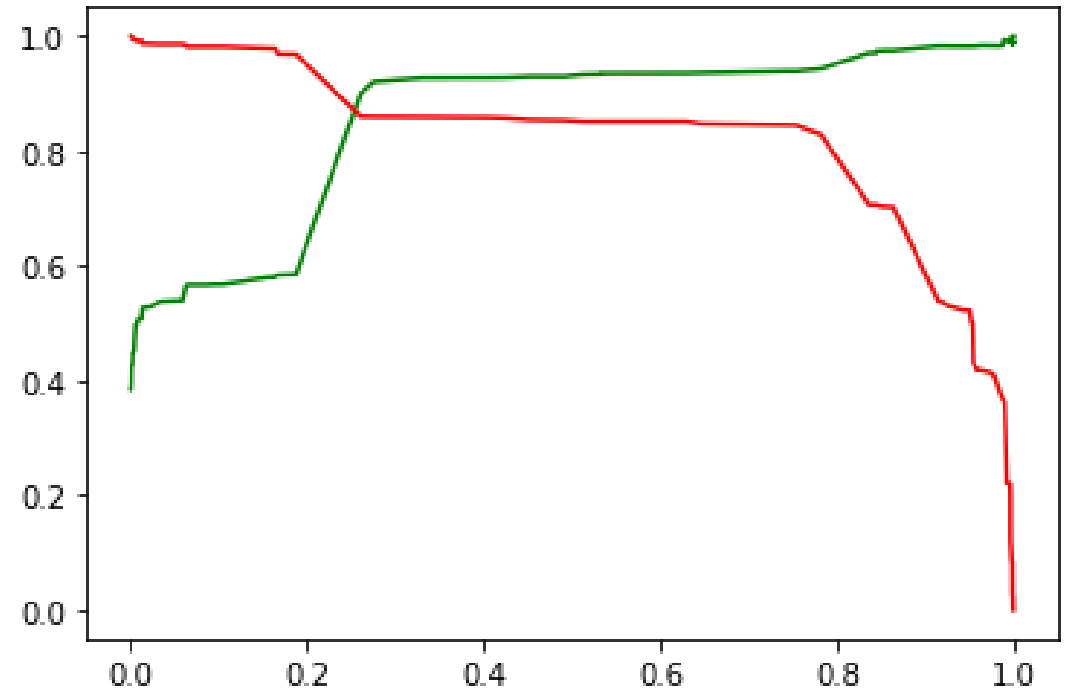- NegativePredictiveValue–91%

Model Evaluation –Sensitivity and Specificity on Test Dataset

[[3756, 149],
[363, 2083]]

Precision: 93%   Recall: 85%

[[1628, 106],
[154, 835]]

Accuracy: 90%

Sensitivity: 84%

Specificity: 93%

THANK YOU