

# Leads Scoring Case Study

## Problem Statement and the Analysis Approach

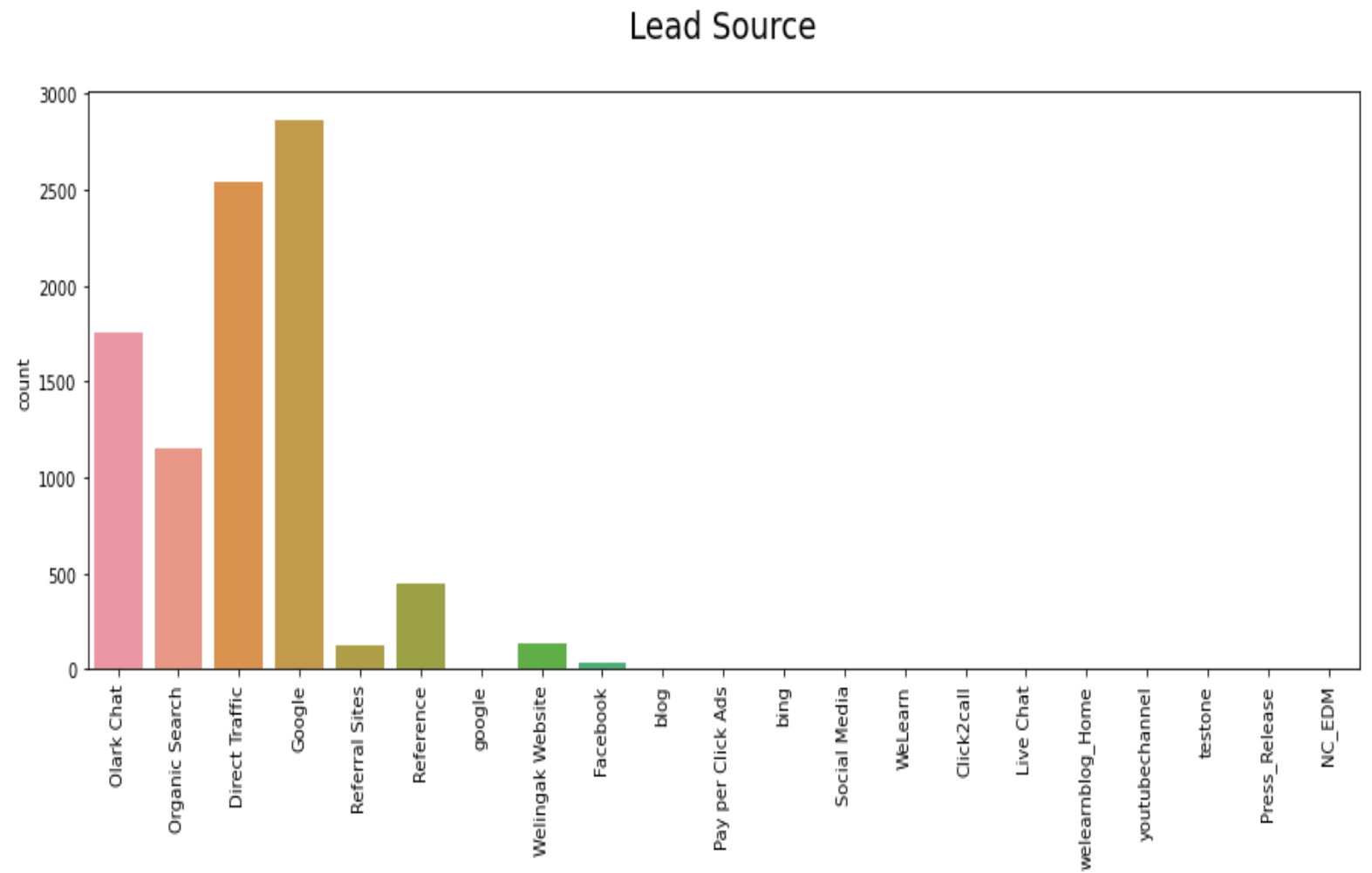
### 1. Problem Statement:

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%

### 2. Analysis Approach:

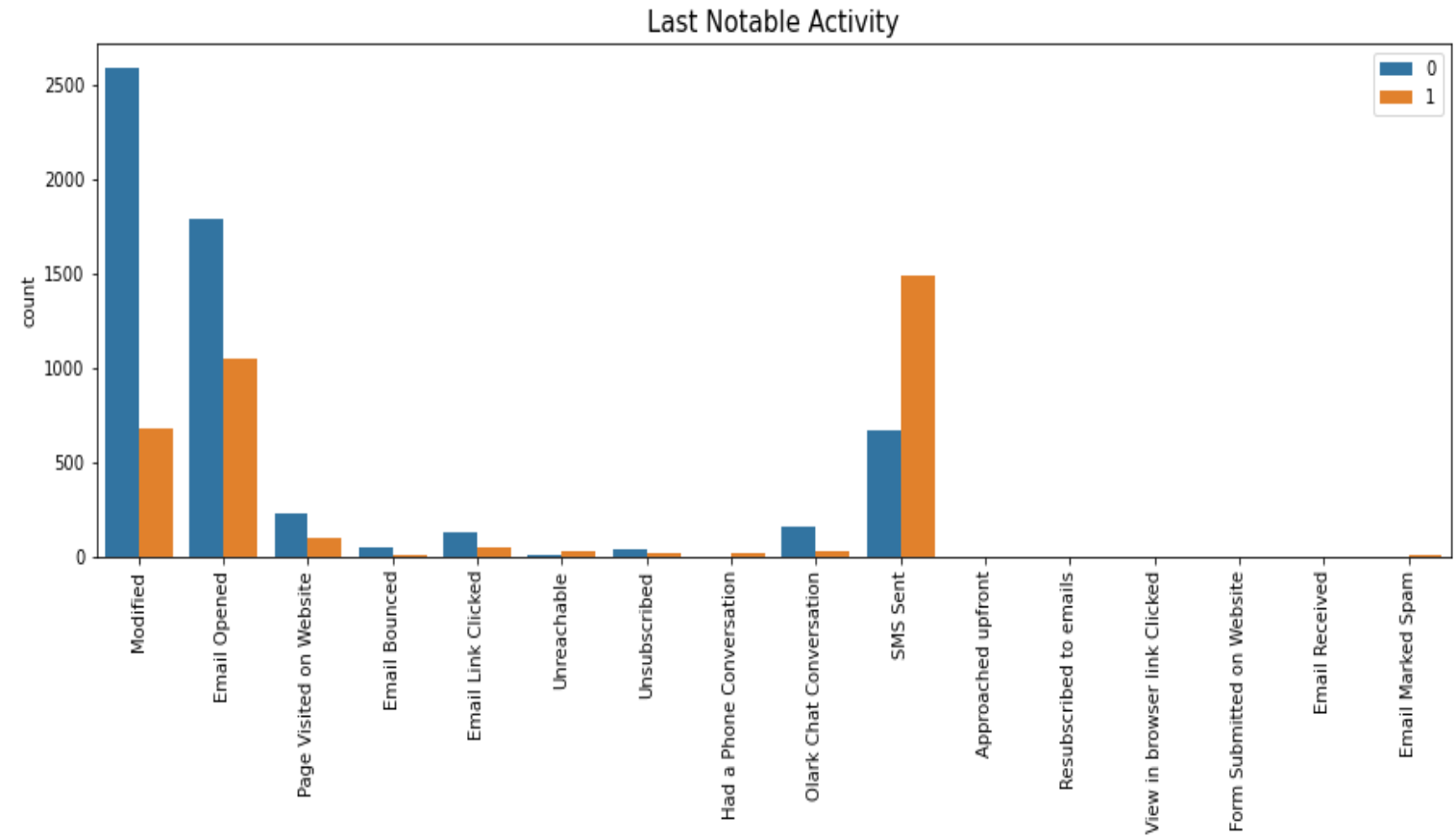
1. Reading the dataset
2. Inspecting the dataset
3. Implementing Exploratory Data Analysis
4. Doing Transformations
5. Train Test Split
6. Building ML Model using Logistic Regression
7. Evaluating the Model and generating Inferences

## Technical and Business Aspects



The following Bar Graph shows the Maximum leads from each sources

## Technical and Business Aspects



The following Bar Graph shows the Last notable activities

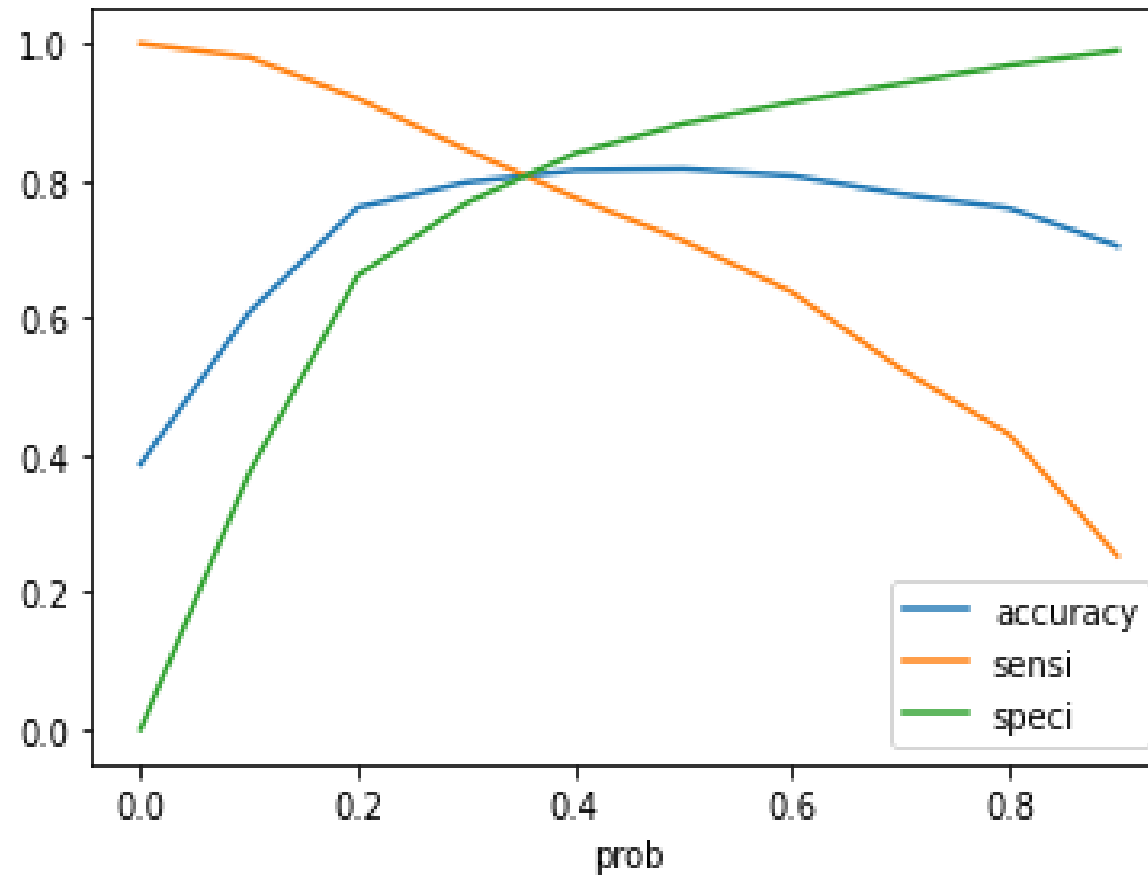
# Technical and Business Aspects

Features	coef
const	-1.2803
TotalVisits	5.7888
Total Time Spent on Website	4.6252
Lead Origin_Landing Page Submission	-1.1266
Lead Origin_Lead Add Form	1.7469
Lead Source_Welingak Website	3.1018
Do Not Email_Yes	-1.4080
Last Activity_Olark Chat Conversation	-1.3856
Last Activity_SMS Sent	1.3027
Country_Not Available	1.4710
Specialization_Not Available	-1.0591
What is your current occupation_Working Professional	2.4164
What matters most to you in choosing a course_Not Available	-1.1105
Last Notable Activity_Had a Phone Conversation	23.0318
Last Notable Activity_Unreachable	1.6143

As per the coefficients of 5.9583, 4.6086 and 3.1024 for the **TotalVisits**, **Total Time Spent on Website** and **Lead Source** (Welingak Website), respectively, these three are the top variables contributing most towards the probability of a lead getting converted.

These has positive impact on the leads to be converted.

## Technical and Business Aspects



With the cut off as 0.35 we have accuracy of 82%, sensitivity of 71% and specificity of around 88%.



## Findings and Inferences

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

- 1.The total time spend on the Website.
- 2.Total number of visits.
- 3.When the lead source was:
  - a. Google
  - b. Direct traffic
  - c. Organic search
  - d. Welingak website
- 4.When the last activity was:
  - a. SMS
  - b. Olark chat conversation
- 5.When the lead origin is Lead add format.
- 6.When their current occupation is as a working professional.



## Findings and Inferences

Target leads that spend a lot of time on X-Education Website

Target leads that have come through References as they have a higher probability of converting.