A case-study on Lead Scoring

Lucrative Learners

Enhancing the lead conversion rate for X Education

by Ashwin Hariharan Iyer

Table of Contents

1. Overview
2. Lead profiles and factors affecting conversion rate – notable ones
3. Machine learning – model building approach and results
4. Recommendations

Overview

This case study aims at enhancing the lead conversion rate for X Education, an online education company that sells professional courses to industry experts. The project focuses on identifying the most promising leads, also known as "Hot Leads," to increase the efficiency of the company's sales and marketing efforts.

To achieve this goal, the project will:

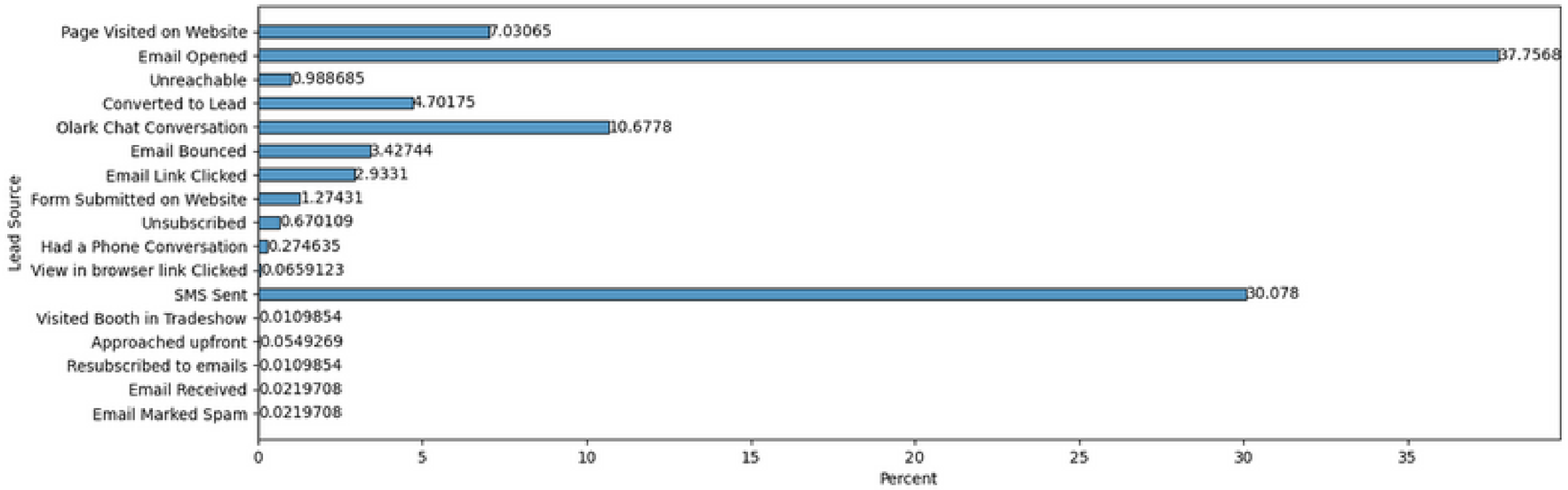
1. Develop a logistic regression model that assigns a lead score between 0 and 100 to each lead. A higher score indicates a hot lead with a high likelihood of converting, while a lower score signifies a cold lead with a low probability of conversion.
2. Ensure that the model can adapt to new requirements or changes in the company's lead evaluation strategy as specified in a separate document.
3. Provide recommendations for implementation

At the start of the analysis, the data source consisted of 9240 lead data, with almost 40 different features. After data cleaning, analysis was done for **9103** lead info using information of **13** features.

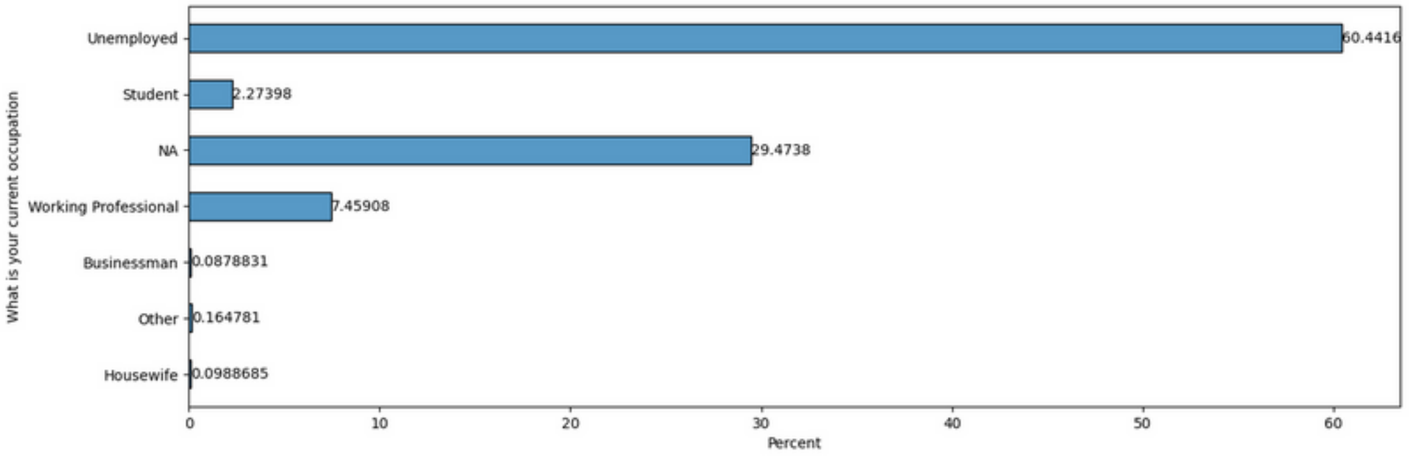
The challenge and opportunity for us is to continue expanding and improving our ability to find leads that are most likely to get converted. Read on for more insights and strategies for better conversion rates.

Lead demographics and factors affecting conversion rates

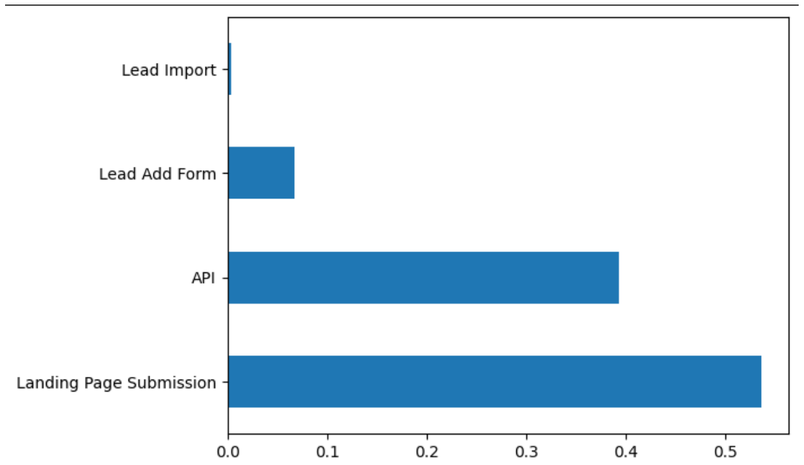
Lead origin and sources



SMSes and **emails** are the most used channels for sourcing leads

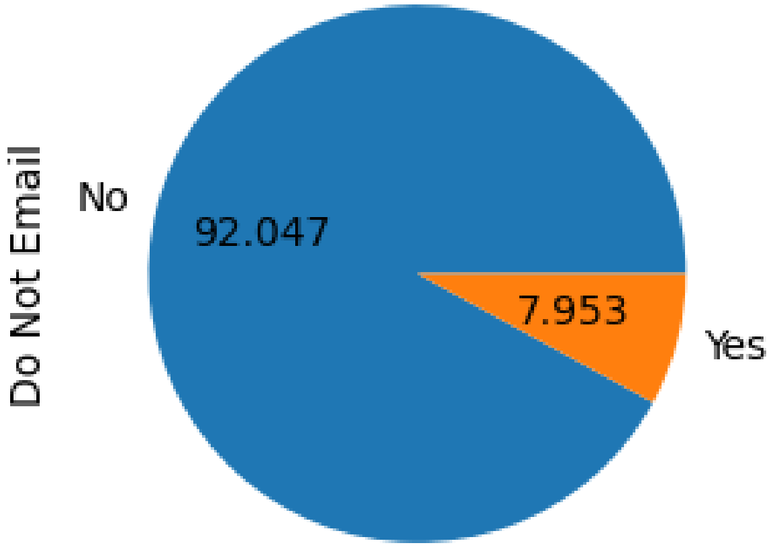


Most leads are **unemployed** or **working professionals**.

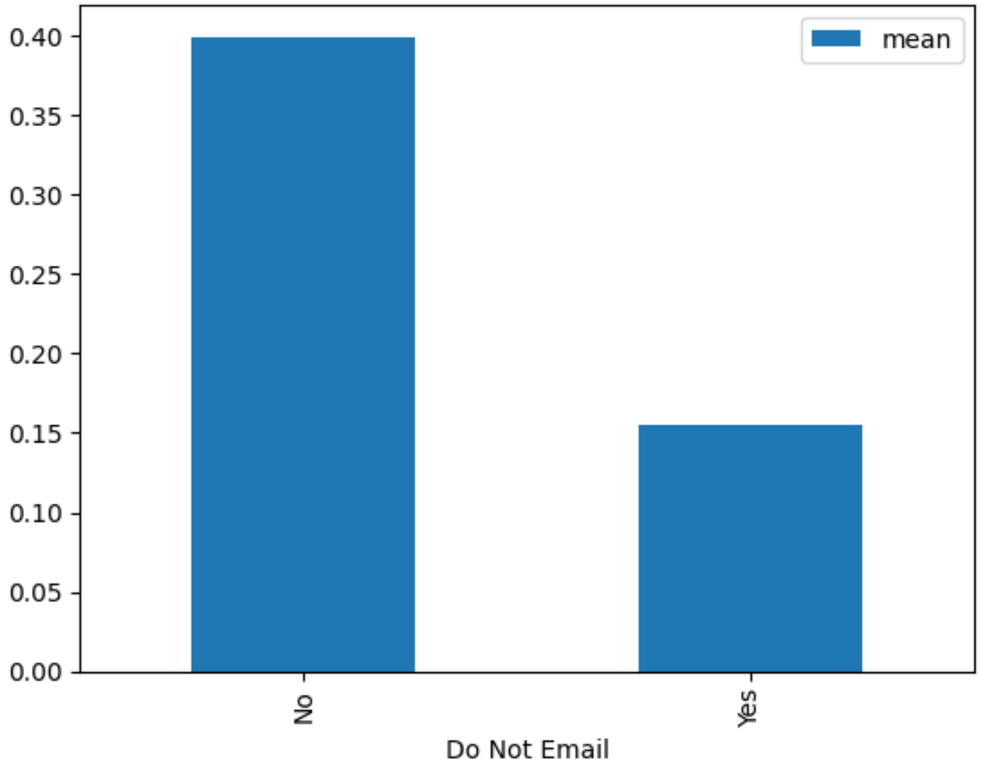


Most leads originate from **Landing Page Submissions**

Email preferences and conversion rates

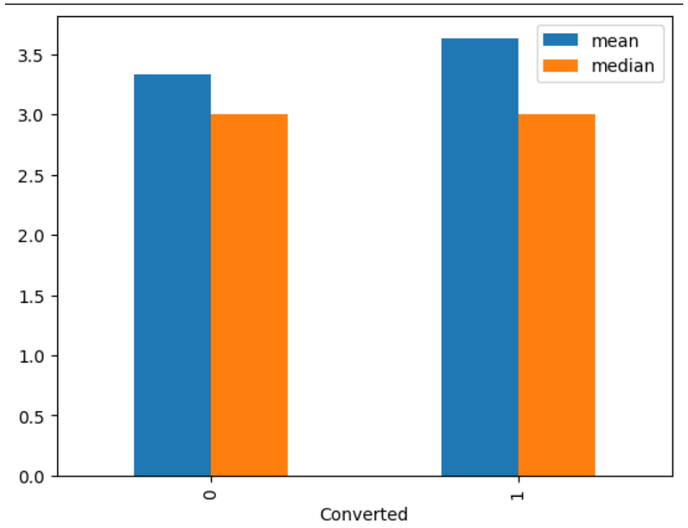


Most leads **are okay** with receiving emails



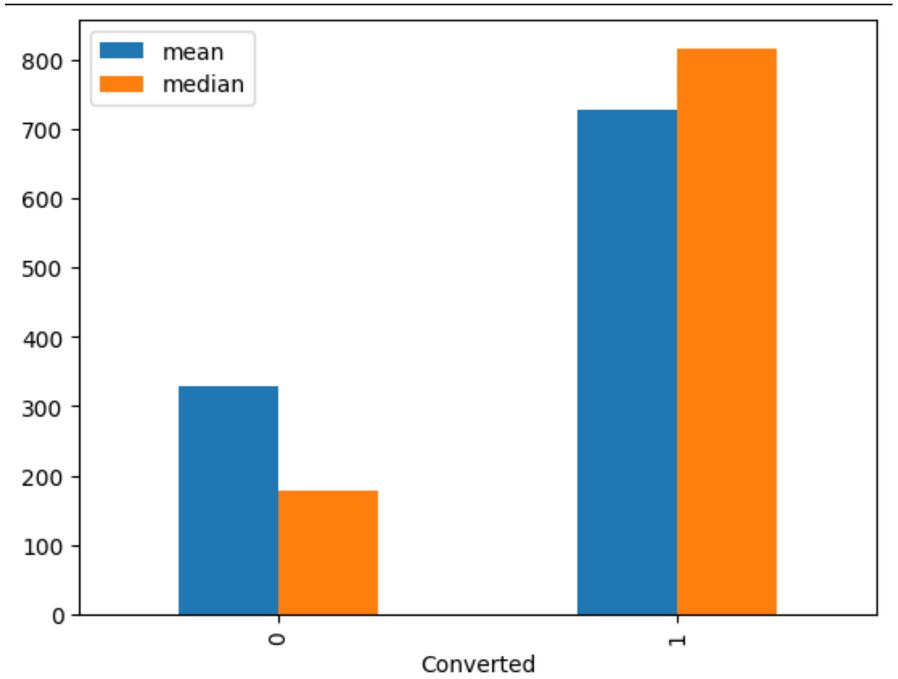
Leads that have opted out of emails have **low conversion rates**

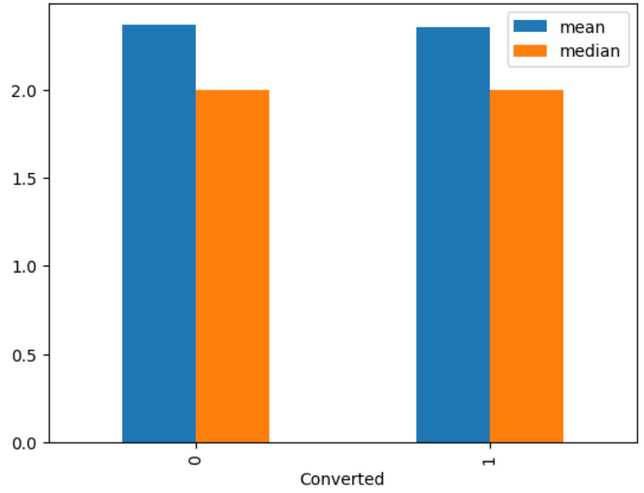
Conversions from website



Total website visits don't play significant role in conversions

Total time spent on website is correlated with a higher number of conversions





Page views per visit also have negligible impact in conversions

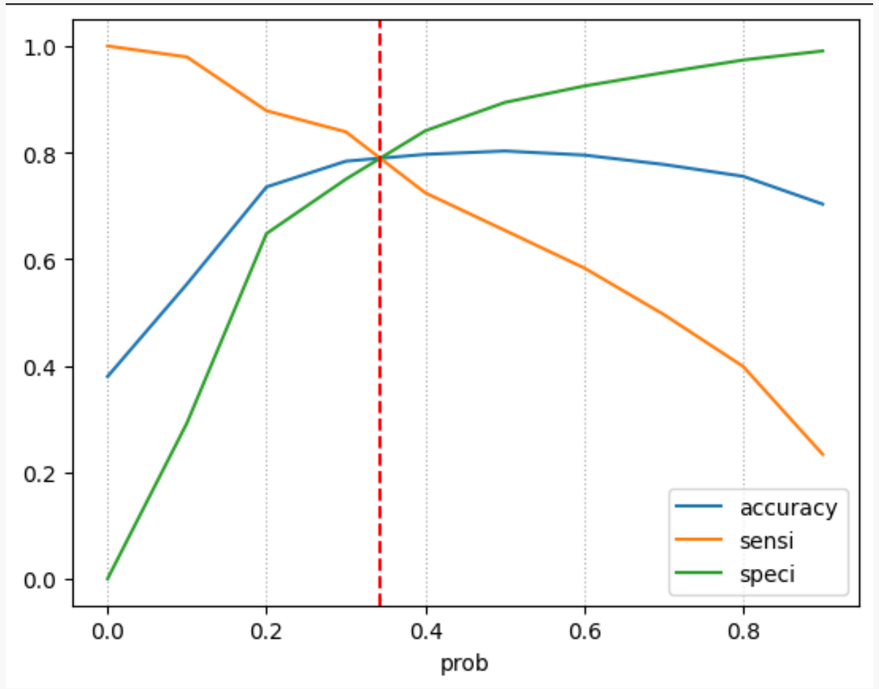
Machine learning

Model building methodology

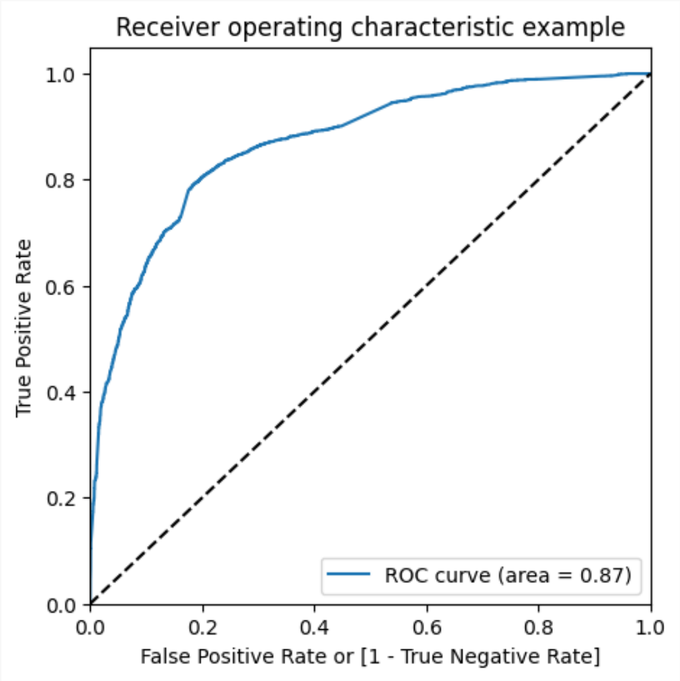
1. The first step was in data cleaning. There were a lot of columns with null values, and it was necessary to identify whether they could be removed, imputed, or dropped. Most of the columns with greater than 40% null values were **dropped**. Many columns also had just a single value, or were **highly skewed** with 99% data points with just a single value. Those were dropped, and columns with very few null values were removed.
2. I then used **one-hot encoding** to create dummy variables for the categorical columns. After that, feature scaling was done on the remaining numerical columns using **standard scaler**.
3. **RFE** was used to reduce the number of initial features to 15. Later, **manual feature selection** was done based on the VIF scores and p-values, and we arrived to a logistic regression model after 4 iterations.
4. For **model evaluation**, it was necessary to use metrics like **sensitivity** and **specificity** as the data was skewed. I used the ROC cuve to arrive at an optimal balance of both metrics.
5. Finally, I computed the lead scores using the conversion rate probability predictions to identify and rank the leads in order of "hottest leads" first.

Final model performance

- Our model achieved an accuracy score of **80%**.
- Sensitivity turned out to be **81.3%**, while specificity was **79%**.



0.343 was the optimum point for a cutoff probability.



Area under ROC curve was **0.87**

Final list of features impacting lead conversions

Do Not Email	-1.591941
Total Time Spent on Website	0.997720
Lead Origin_Landing Page Submission	-0.589005
Lead Origin_Lead Add Form	3.601165
Lead Source_Converted to Lead	-1.422688
Lead Source_Had a Phone Conversation	1.684930
Lead Source_Olark Chat Conversation	-1.284270
Lead Source_Unsubscribed	1.067001
What is your current occupation_Working Professional	2.812666
Last Activity_SMS Sent	1.247845

Recommendations for lead conversion

In general, the following variables are strong indicators of lead conversion success or failure:

- Email preference
- Total time spent on website
- Leads originating from landing page form submissions
- Leads that are working professionals

X Education should therefore invest their resources into the following:

- Developing strategies for engaging leads on the **website**
- Training staff for better engagement over **phone** and chat
- Sales should also target **working professionals**.
- Marketing can look into long-term strategies to make **email campaigns** more effective.