Lead Scoring Case Study using logistic regression

SUBMITTED BY :

1. Navneet Kamaliya

2. Thanmay Das

Agenda

- ❑ Problem Statement
- ❑ Problem Approach
- ❑ Data Analysis (EDA)
- ❑ Model Evaluation
- ❑ Observations
- ❑ Conclusion

Problem Statement

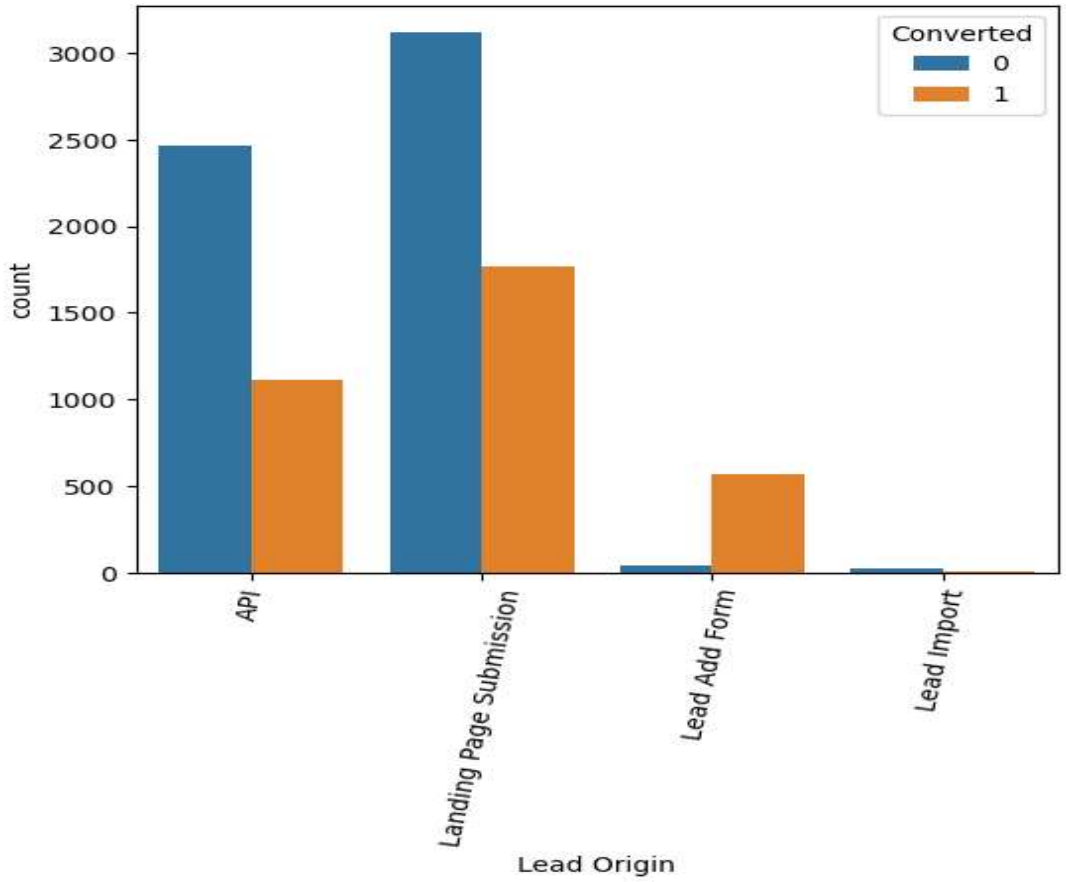
X Education, an online education company, faces a challenge with a low lead conversion rate (approximately 30%). While the company generates a substantial number of leads daily through its marketing channels, only a fraction of these leads convert into paying customers. To optimize the lead conversion process, X Education seeks to identify and prioritize "Hot Leads" — leads with the highest likelihood of converting into paying customers.

The goal is to develop a model that assigns a lead score to each lead, helping the sales team focus on the most promising prospects. By accurately identifying high-potential leads, the company aims to increase its conversion rate to around 80%, thereby improving the overall efficiency and effectiveness of its sales process.

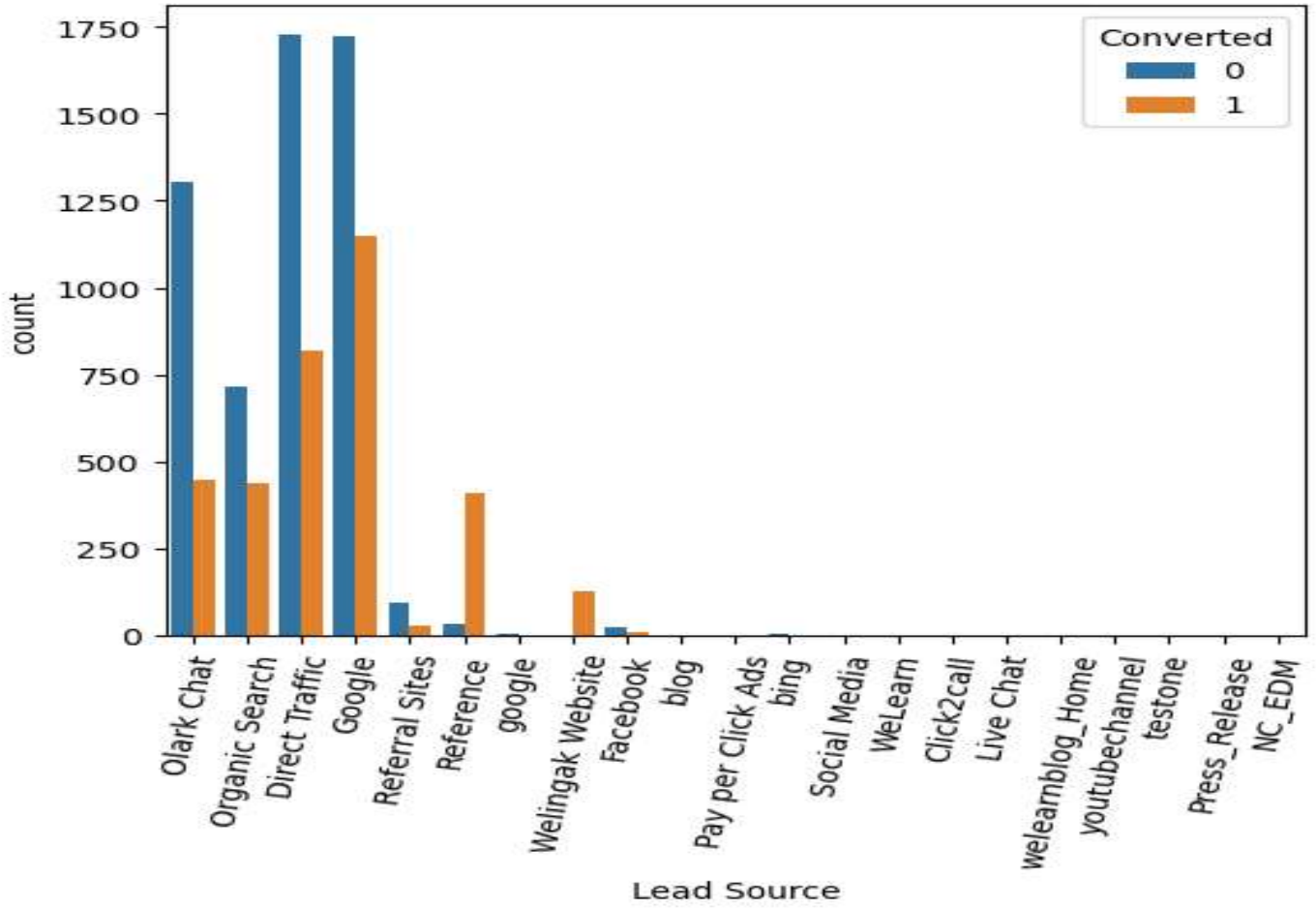
Problem Approach

- ❑ Import the dataset and understand the dataset
- ❑ Data Cleaning and Preparation
- ❑ Data Analysis
- ❑ Dummy variable Creation
- ❑ Test-Train split
- ❑ Feature scaling
- ❑ Model Building
- ❑ Model Evaluation
- ❑ Making predictions on test set

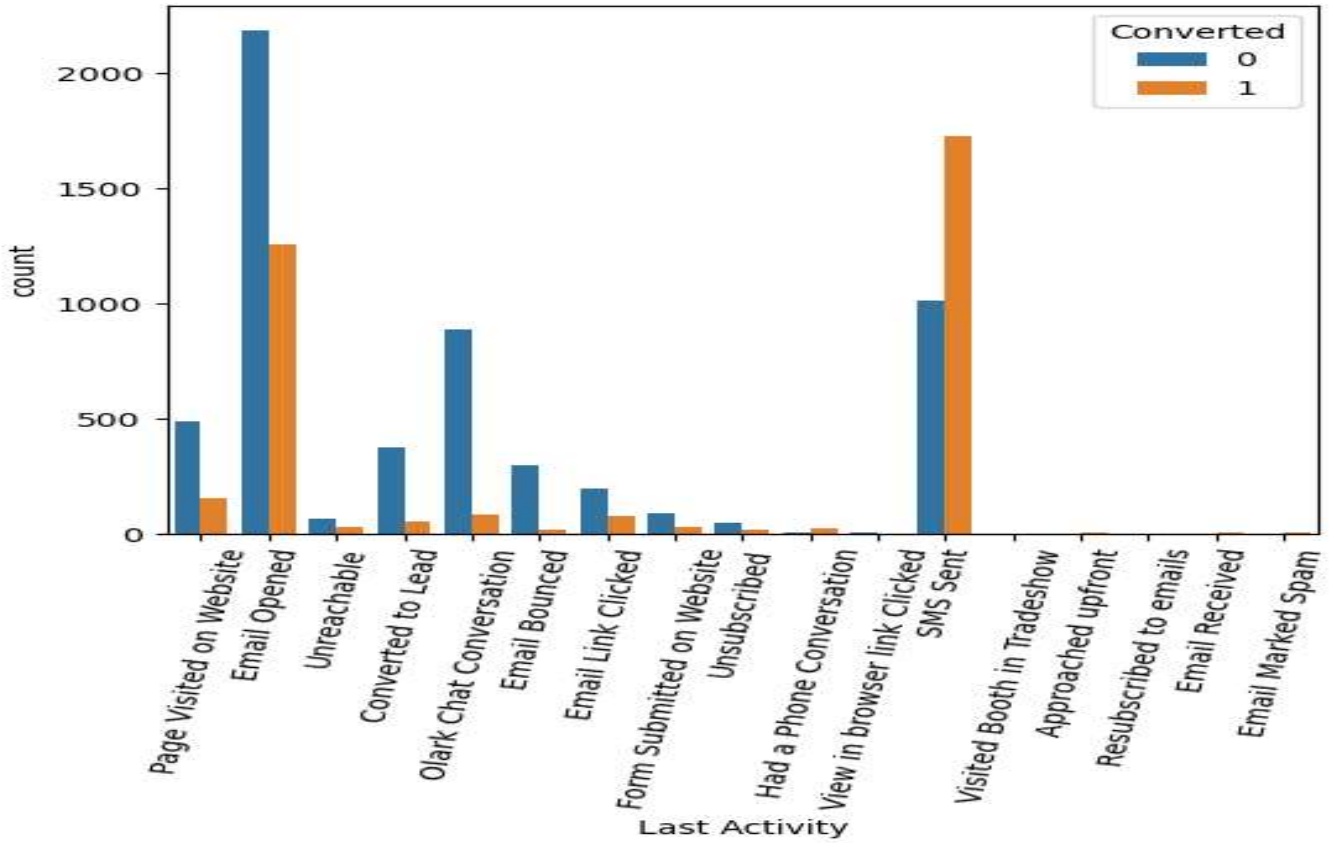
Exploratory Data Analysis



- The "Loading Page submission" generates the highest number of customer conversions, whereas the "Lead Add Form" proves to be the most effective method for converting leads.

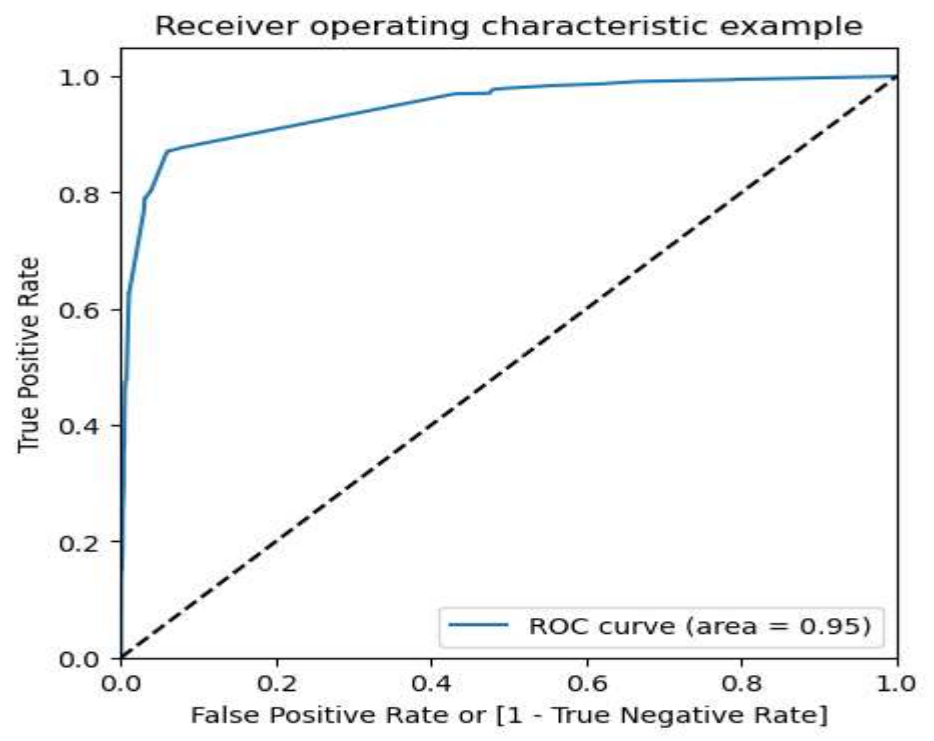


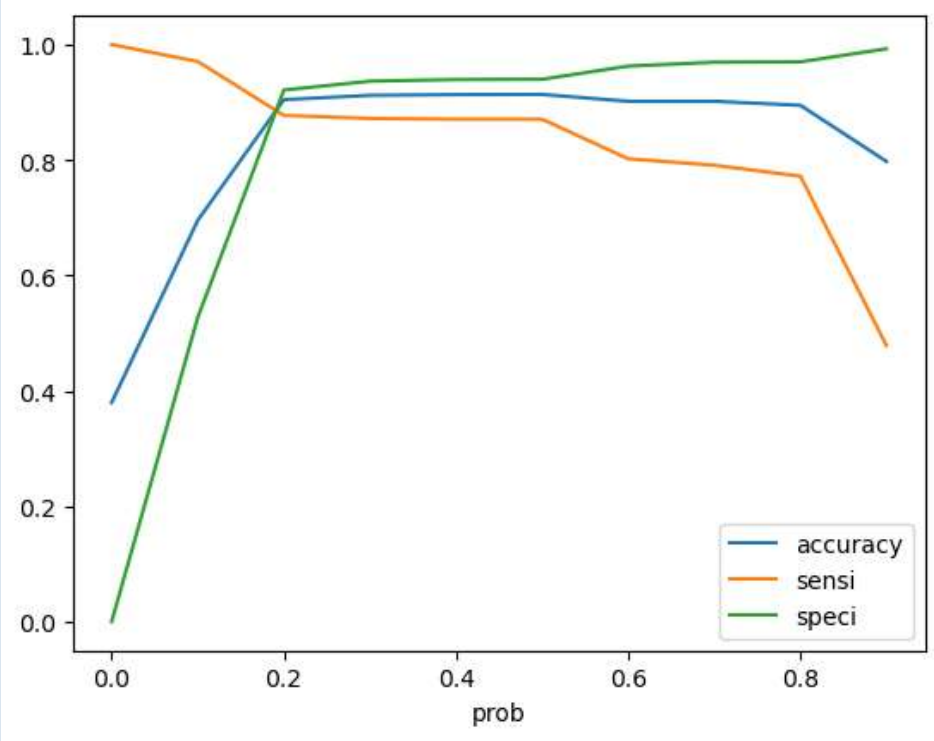
- Google stands out as the top source for lead conversions, with referrals significantly boosting trust among customers, leading to higher conversion rates.



- Customers who have had their last activity as either opening an email or receiving an SMS are more likely to convert. Among these, receiving an SMS is found to be particularly impactful in driving conversions.

Model Evaluation





- **The ROC curve has an area of 0.95, which indicates a strong model performance. Now, let's examine the tradeoff between sensitivity and**

- **As observed, the optimal values for the three metrics occur around a cutoff of `0.2`. Therefore, we will use 0.2 as our threshold.**

Observations

Train Data :

Accuracy : 91%
Sensitivity : 87%
Specificity : 93%

Test Data :

Accuracy : 90%
Sensitivity : 86%
Specificity : 93%

CONCLUSION

The conversion rate for API and landing page submissions is around 30-35%, which is close to the average. However, the conversion rates for Lead Add forms and Lead imports are significantly lower. This suggests that we should focus more on leads generated through the API and landing page submissions.

The majority of leads come from Google and direct traffic, while the highest conversion rates are seen with leads from referrals and the Welingak website.

Leads who spend more time on the website are more likely to convert. The most common last activity recorded is "email opened," while "SMS Sent" shows the highest conversion rate. Most of the leads are unemployed, and the highest conversion rates are seen among working professionals.

Thank you