



LEAD SCORING CASE STUDY ASSIGNMENT FOR UPGRAD

By Mahendrakumar Khiani



01

PROBLEM STATEMENT

PROBLEM STATEMENT

XEducation, an online education provider for professionals, experiences a daily influx of potential leads interested in their courses. The process begins when individuals visit the website, browse through the course offerings, and fill out a form, thereby becoming leads for the company.

The sales team then engages with these leads through calls and emails in an effort to convert them into customers. However, the current lead conversion rate stands at approximately 30%, meaning only 30 out of every 100 leads are successfully converted.

To optimize this process and enhance efficiency, XEducation aims to pinpoint the 'Hot Leads' - those with the highest potential for conversion. By accurately identifying and focusing efforts on these Hot Leads, the company anticipates a significant improvement in the conversion rate, as the sales team's resources will be more strategically allocated towards engaging with leads most likely to enroll in their courses.



02

BUSINESS OBJECTIVE

BUSINESS OBJECTIVE

XEducation is set on enhancing its lead conversion strategy by introducing a lead scoring model that assigns scores between 0 to 100 to each lead. This model aims to pinpoint 'Hot Leads' with the highest potential for conversion, thereby optimizing the sales team's focus and significantly improving conversion rates. The overarching goal is to elevate the lead conversion rate to an ambitious target of 80%.

To achieve this, the model will be built on logistic regression, leveraging various data points such as lead source, time spent on the website, total visits, last activities, and more. These factors are crucial in determining the likelihood of a lead converting into a paying customer. The model will also be designed to adapt to future requirements, such as managing peak times efficiently and strategizing post-target achievement



03

PROBLEM APPROACH

PROBLEM APPROACH

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and pvalues)
- Model Evaluation
- Making predictions on test set



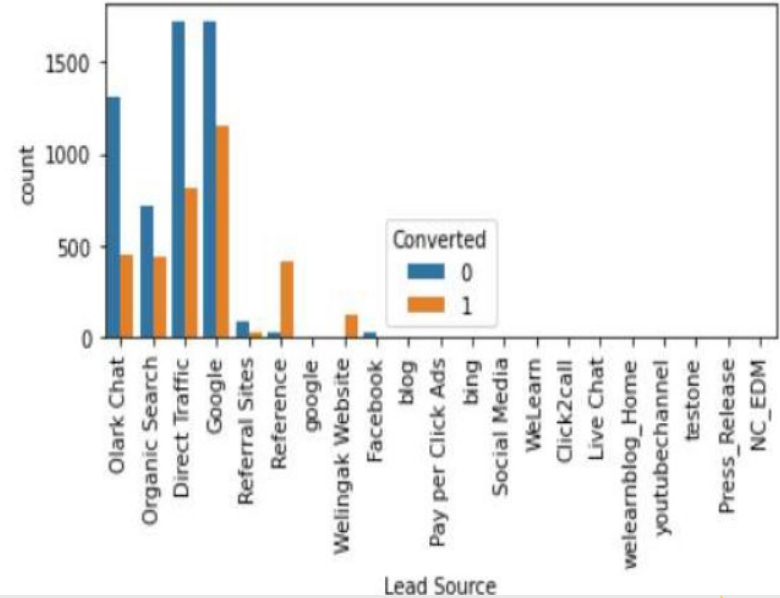
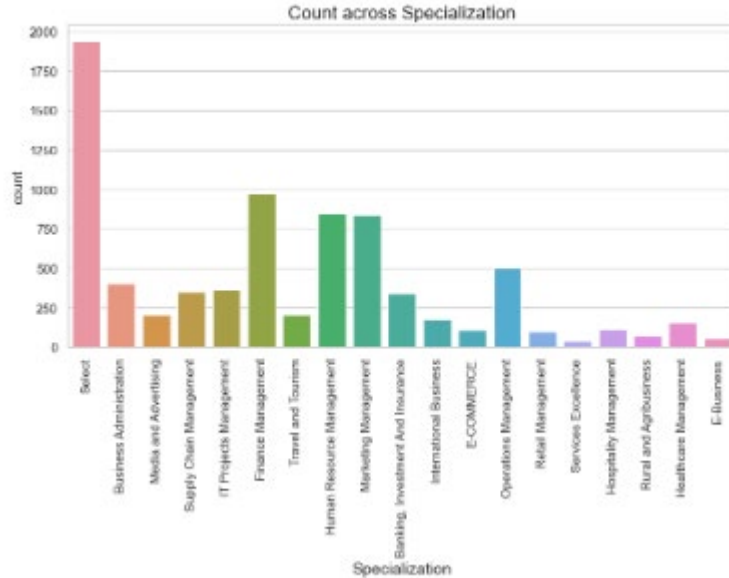
04

KEY INSIGHTS

KEY INSIGHTS (1/2)

Leads from HR, Finance & Marketing management specializations have high probability to convert.

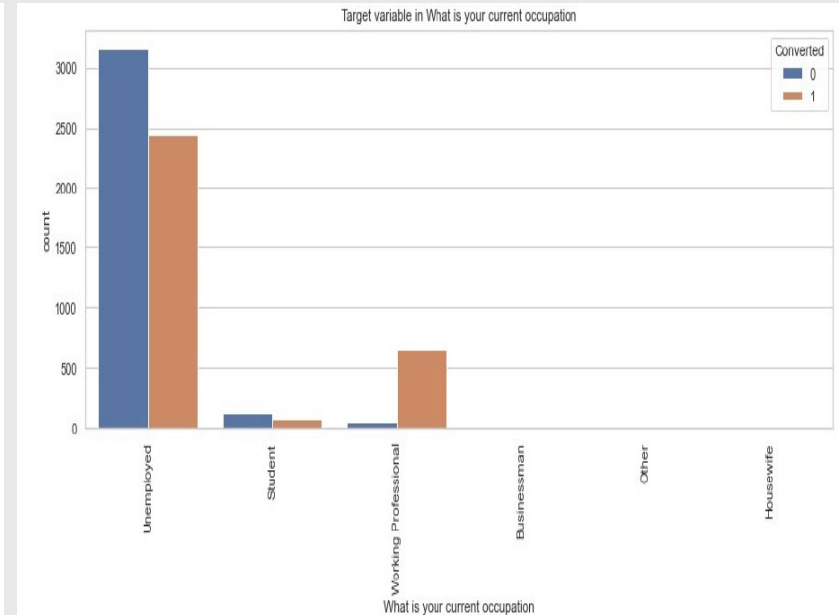
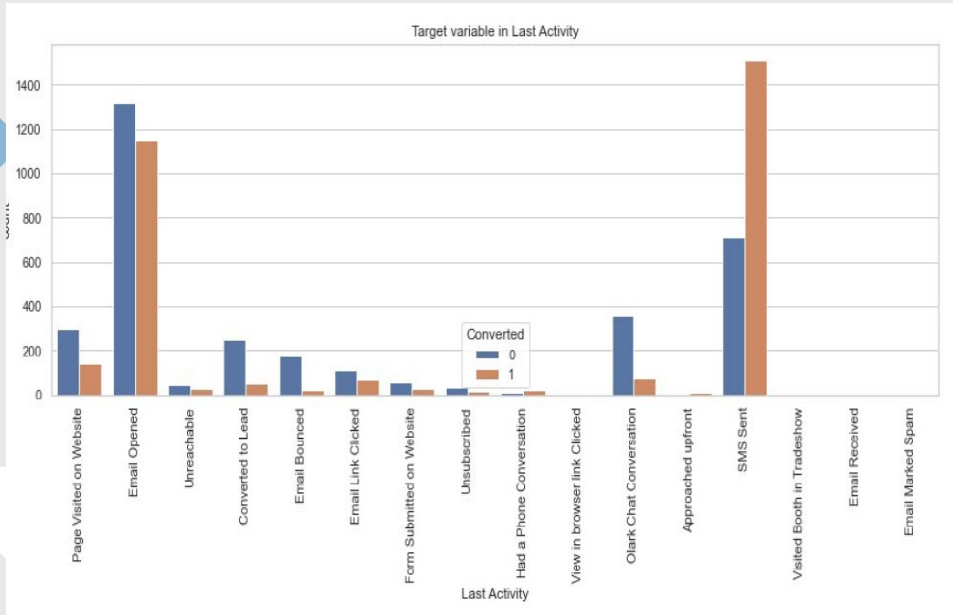
In lead source the leads through google & direct traffic high probability to convert



KEY INSIGHTS (2/2)

Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.

Leads which are Unemployed are more interested to join the course than others





05

Model Evaluation, Observation and Conclusions

Summary Performance

➤ Train Data:

- Accuracy: 80%
- Sensitivity: 77%
- Specificity: 80%

➤ Test Data:

- Accuracy: 80%
- Sensitivity: 77%
- Specificity: 80%

➤ Final Features List:

- Lead Source_Olark Chat
- Specialization_Others
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- What is your current occupation_Working Professionals
- Do Not Email

Conclusion

- a. The analysis reveals that leads originating from API and Landing Page Submission show a conversion rate of 30-35%, aligning with the average. However, conversions from Lead Add Form and Lead Import are notably lower, indicating a need for strategic focus on leads from API and Landing Page submissions to enhance overall conversion rates.
- b. Google and Direct Traffic are the primary sources for lead generation, yet the highest conversion rates are observed from references and the Welingak website. This suggests that while volume comes from broader channels, more targeted or niche sources yield better conversion outcomes.
- c. Additionally, the duration spent on the website emerges as a significant predictor of lead conversion; leads engaging longer with the site content have a higher likelihood of converting. The most frequent last activity recorded is 'Email Opened', but the 'SMS Sent' activity correlates with the highest conversion rates, underscoring the effectiveness of direct messaging.
- d. In terms of professional status, the majority of leads are unemployed, yet the highest conversion rates are seen among working professionals. This insight could guide more tailored marketing efforts towards working professionals to improve conversion rates.