Lead Scoring Case Study Via Logistic Regression

SUBMITTED BY :

1. **Piyush Mandal**
2. **Ridheesh Bhan**
3. **Ramyaseetha**

Contents

- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

Problem Statement

- X Education, an educational organization, specializes in offering online courses tailored to professionals in various industries. Each day, a substantial number of these professionals visit the organization's website, exploring the diverse range of courses available. As part of their engagement process, the website features a form that visitors can fill out, enabling the company to categorize these individuals as leads.

- Once the leads are obtained, the dedicated sales team initiates a series of interactions, including phone calls and email correspondence, among other approaches. These efforts aim to nurture these leads, with the objective of converting them into paying customers. While this approach yields successful conversions for some leads, a majority remain unconverted.

Problem Statement

- Typically, X Education maintains an average lead conversion rate of approximately 30%. In practical terms, if they secure, for instance, 100 leads in a single day, only around 30 of them tend to progress into actual customers. In an endeavor to enhance the efficacy of this process, the organization seeks to pinpoint leads with the highest potential, often referred to as "Hot Leads."
- By successfully identifying this subset of leads, X Education anticipates a notable increase in its lead conversion rate. This optimistic outcome would be the result of the sales team concentrating their efforts on communicating more extensively with the potential leads, rather than engaging with the entire pool of acquired leads.

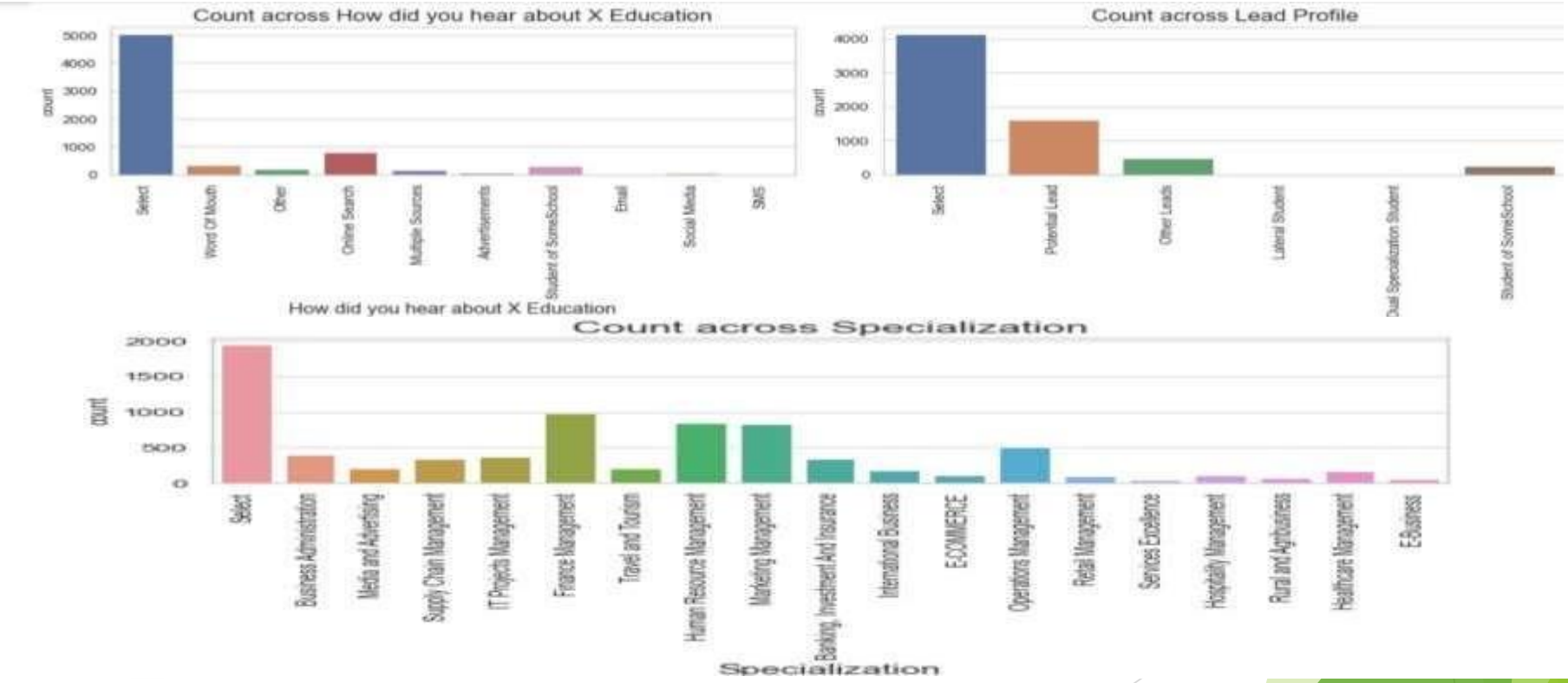
Business Objective

- Lead X has requested us to develop a model that assigns a lead score ranging from 0 to 100. This scoring system aims to assist in distinguishing the high-potential leads, facilitating an enhancement in the conversion rate.
- The CEO's objective is to attain a substantial 80% lead conversion rate
- Furthermore, the organization intends for this model to possess adaptability, enabling it to accommodate future consideration.
- This encompassed factors such as optimizing the acton., during peak activity period.

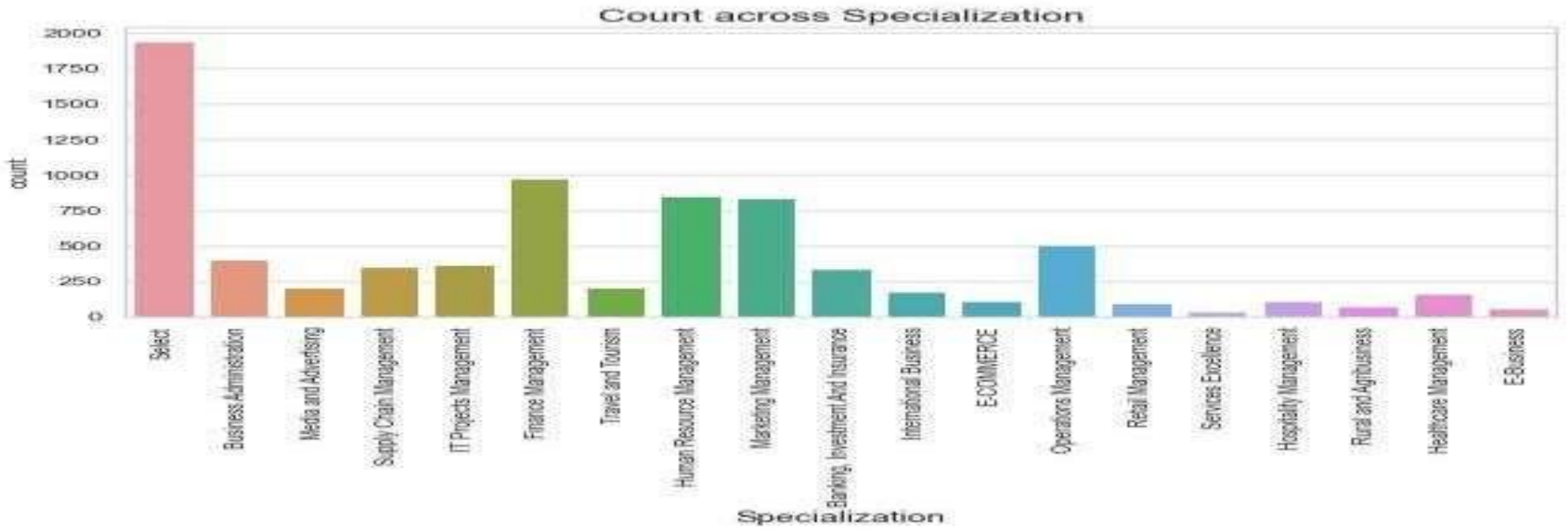
Problem Approach

- 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 p- values)
- Model Evaluation
- Making predictions on test set

EDA – Data Cleaning



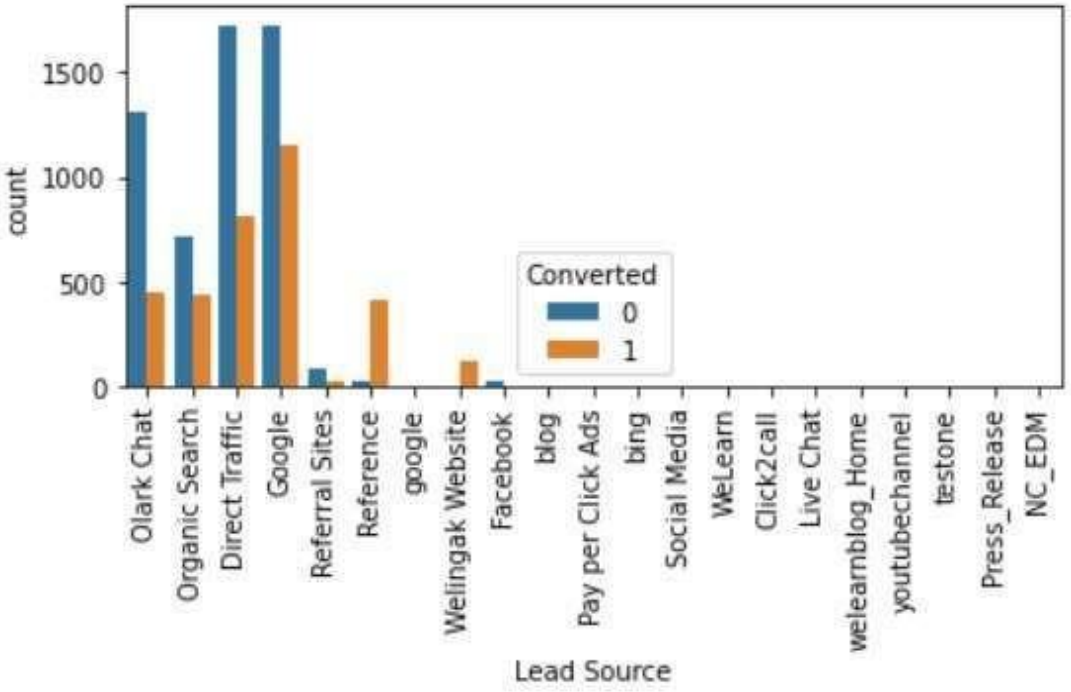
Specialization

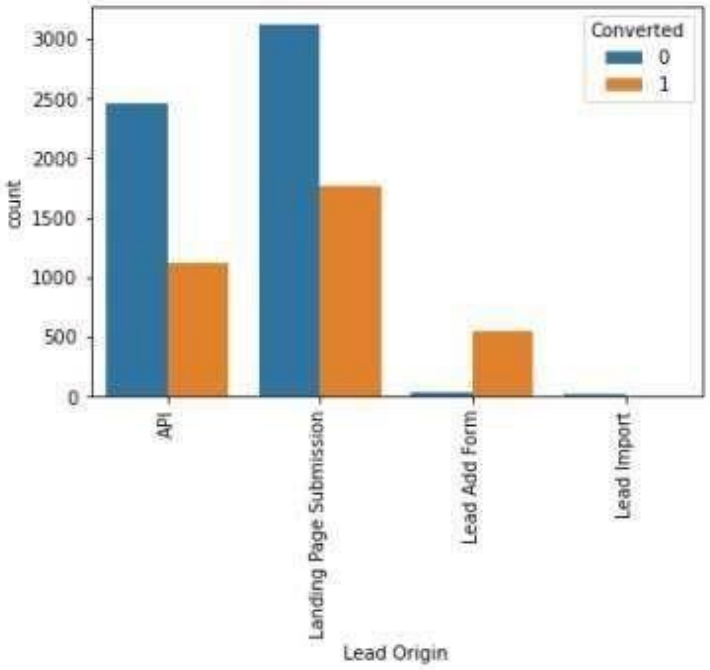


Lead Source & Lead origin

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

Whereas in Lead origin most number of leads are landing on submission

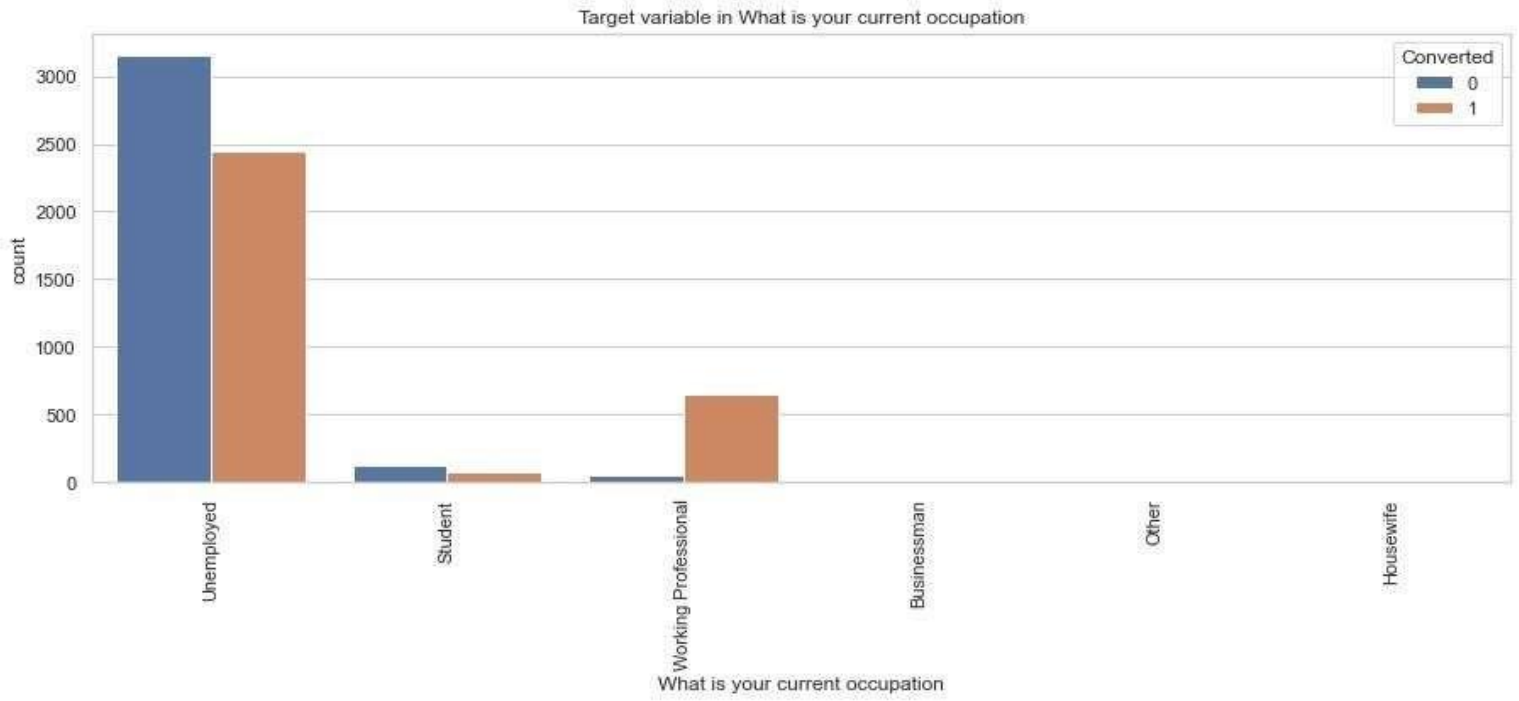




Last lead Activity

Last What is Your Occupation

hers.



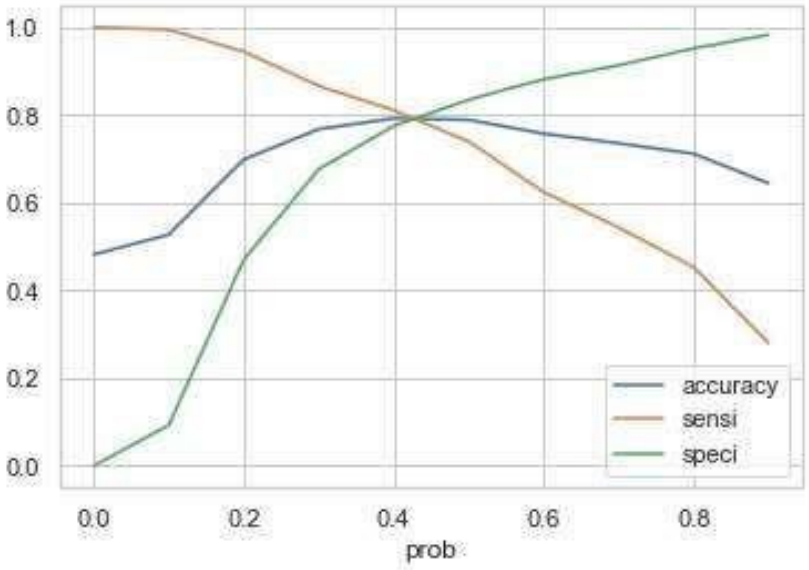
Correlation

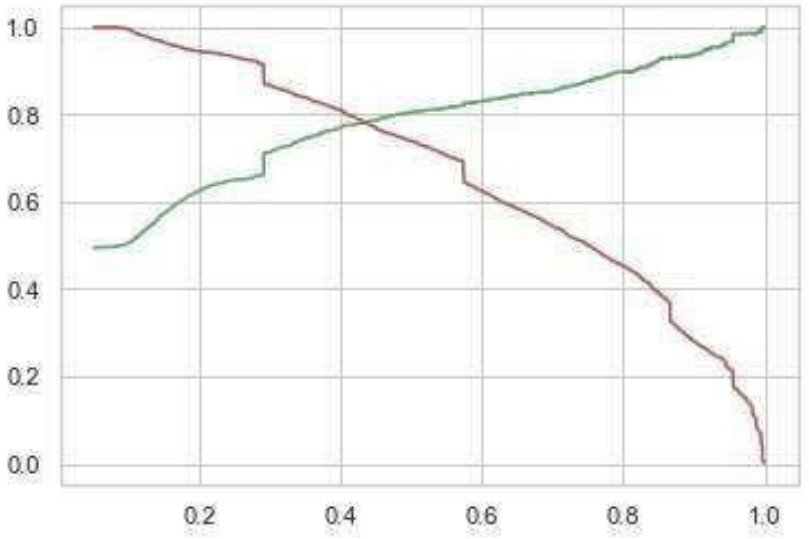
There is no correlation between the variables

Model Evaluation

The tradeoff between Precision and Recall is 0.42%

As a result any Prospect Lead with Conversion Probability Higher than 42% is a hot lead.





Observations

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

- The conversion rate for API and Landing page submissions is approximately 30-35%, which aligns closely with the average rate. However, the conversion rate is notably lower for Lead Add forms and Lead imports. This observation leads us to conclude that there is a need to allocate more attention to leads originating from API and Landing page submissions.

- An analysis of lead generation sources reveals that the majority of leads are generated through sources like Google and direct traffic. Remarkably, the highest conversion ratio is associated with leads referred by existing clients and those coming from the Welingak website.

- Furthermore, leads that spend more time engaging with the website exhibit a higher likelihood of conversion. Among the various tracked last activities, email openings stand out as the most common. Interestingly, the highest conversion rate corresponds to SMS Sent.

- When categorizing leads, it's evident that the largest segment comprises unemployed individuals. However, the maximum conversion rate is achieved with working professionals