Lead Scoring Case Study using logistic regression

Submitted By:

Ritwik

Contents

- Problem statement
- Business Objective
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclus ion

Problem Statement

- X Education, an online education company, offers courses to industry professionals. Each day, numerous professionals interested in these courses visit the website to browse. After filling out a form on the website, individuals are classified as leads by the company.

- After acquiring these leads, the sales team begins contacting them through phone calls, emails, and other methods. While some leads convert into customers, the majority do not.

- X Education's typical lead conversion rate is around 30%. This means that if they acquire 100 leads in a day, only about 30 are converted into customers. To improve efficiency, the company aims to identify the most promising leads, referred to as "Hot Leads."

- By successfully identifying this group of leads, the conversion rate is expected to increase, as the sales team will be able to concentrate their efforts on engaging with high-potential leads instead of reaching out to everyone.

Business Objective

- Lead X has tasked us with developing a model that assigns each lead a score between 0 and 100. This scoring system will help them identify Hot Leads, allowing them to focus on these high-potential prospects and improve their conversion rate.

- The CEO aims to achieve a lead conversion rate of 80%.

- They want the model to be adaptable to future challenges, such as managing actions during peak times, optimizing the use of the full workforce, and determining the best approaches after the target has been reached.

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)
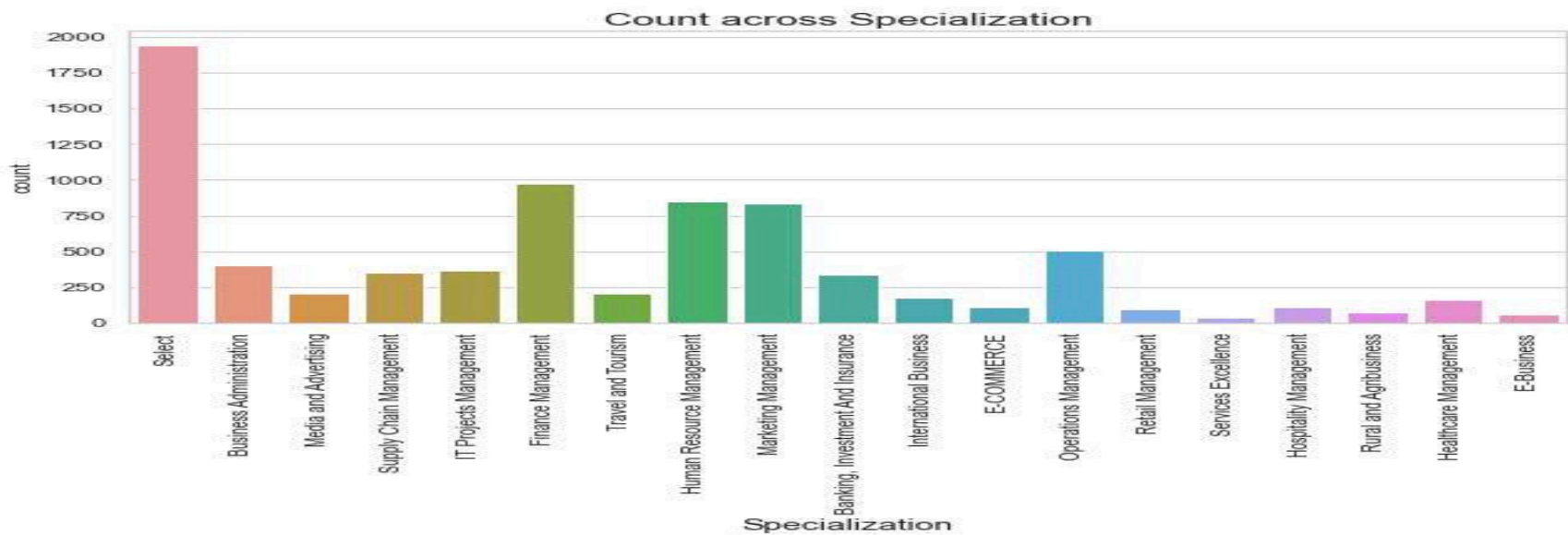- Model Evaluation
- Making predictions on test set

EDA – Data Cleaning

- In certain columns, there is a level labeled 'Select' that is being accounted for.
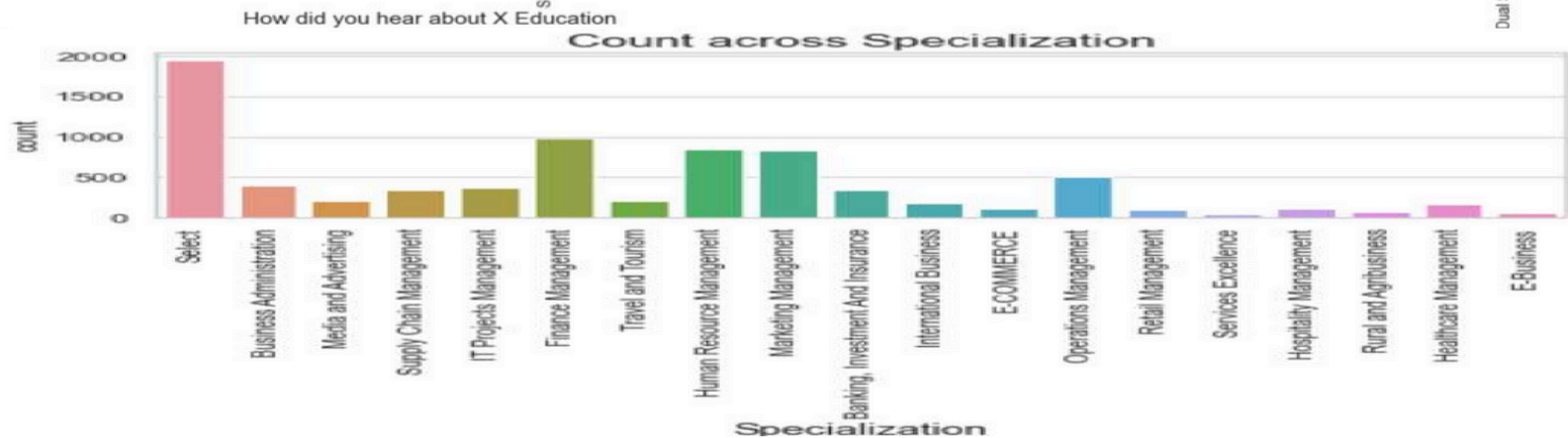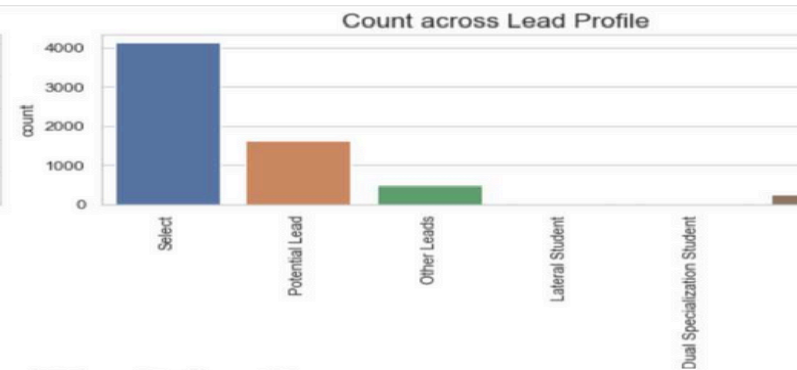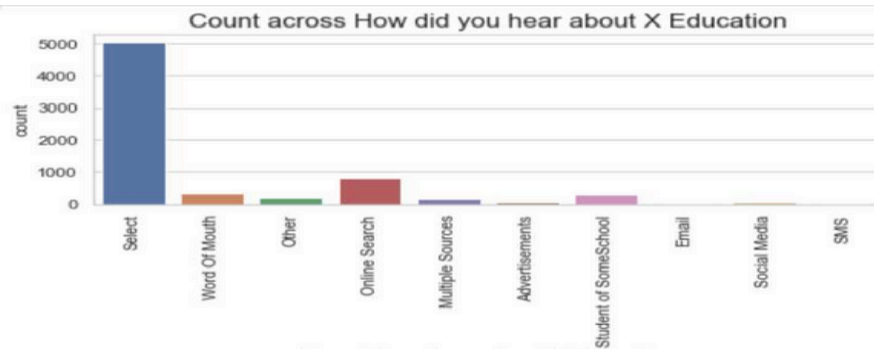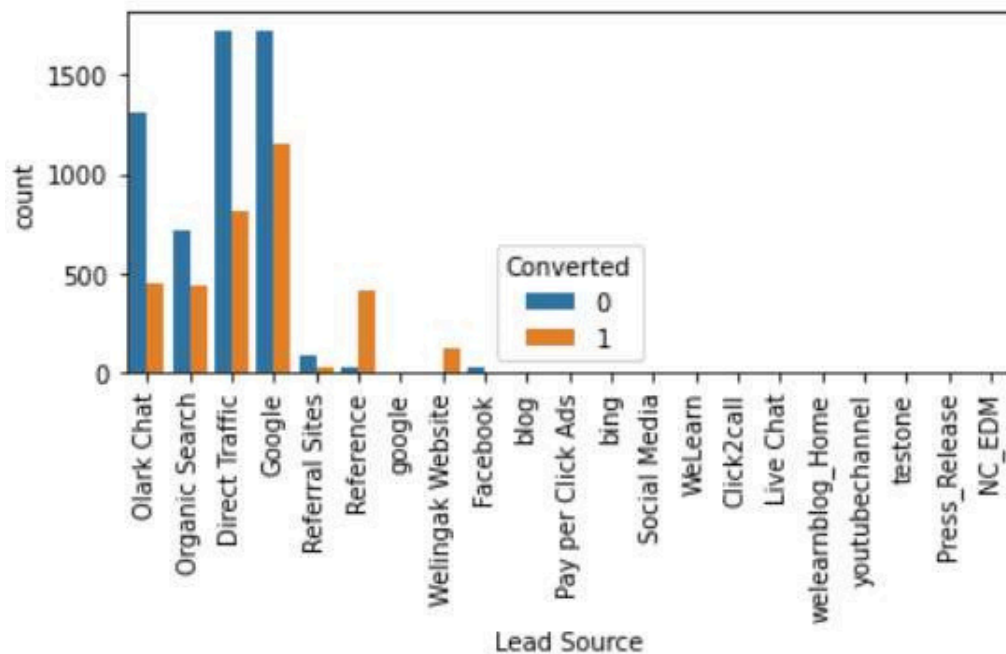
Specialization

- Leads from HR, Finance & Marketing management specializations are high probability to convert

Count across Specialization
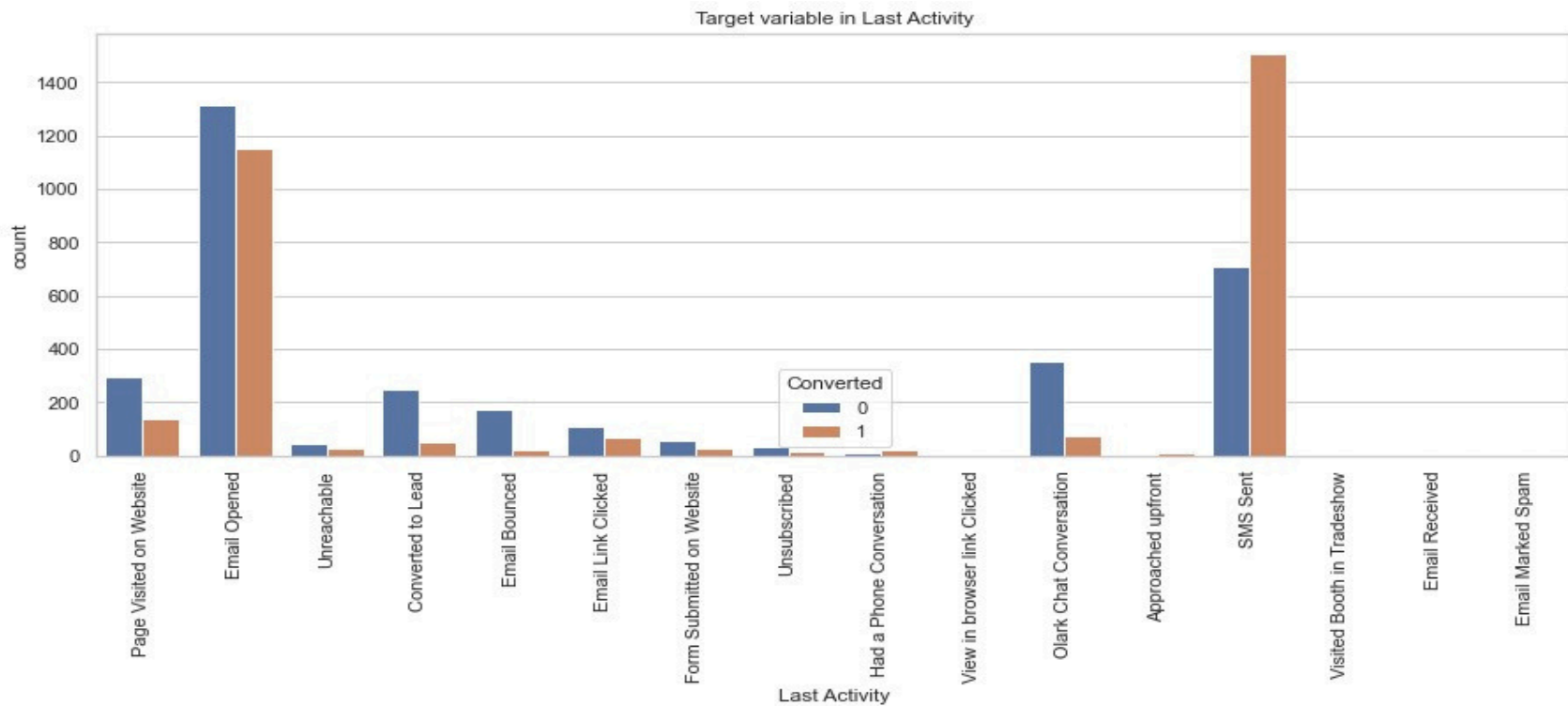
Lead Source & Lead origin

- 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.
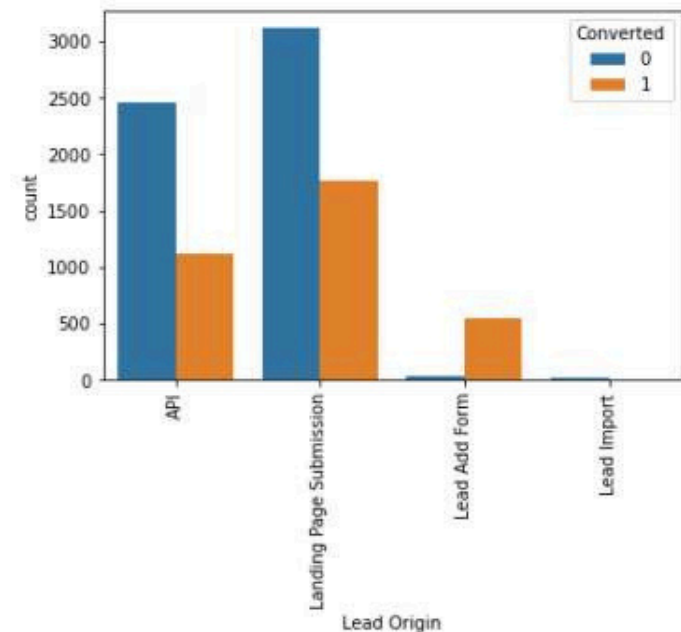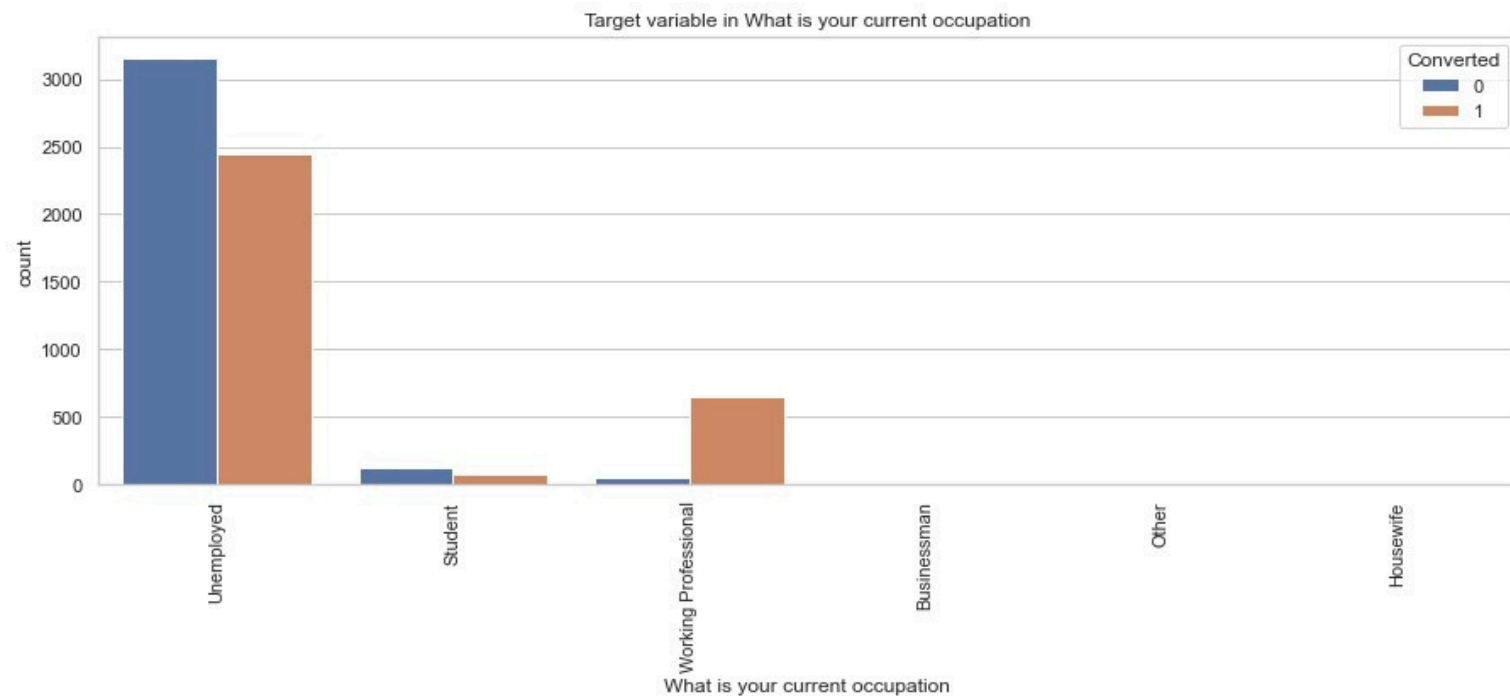
Last lead Activity

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

Last What is Your Occupation

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

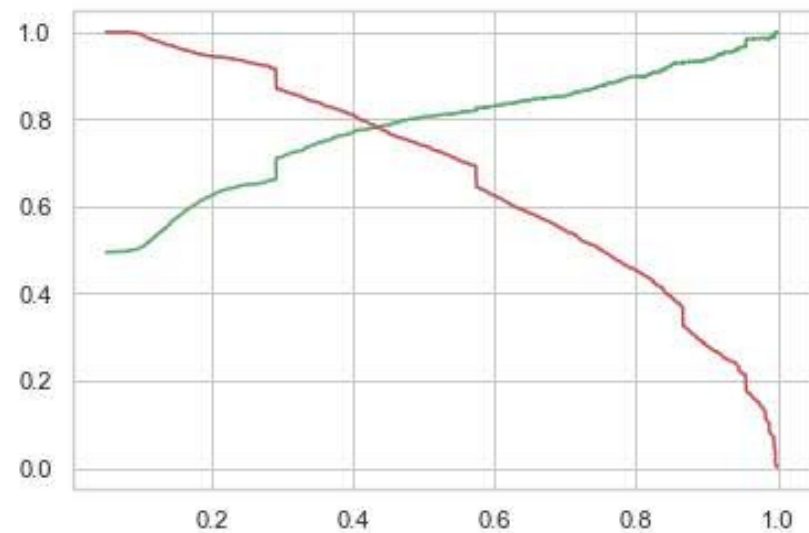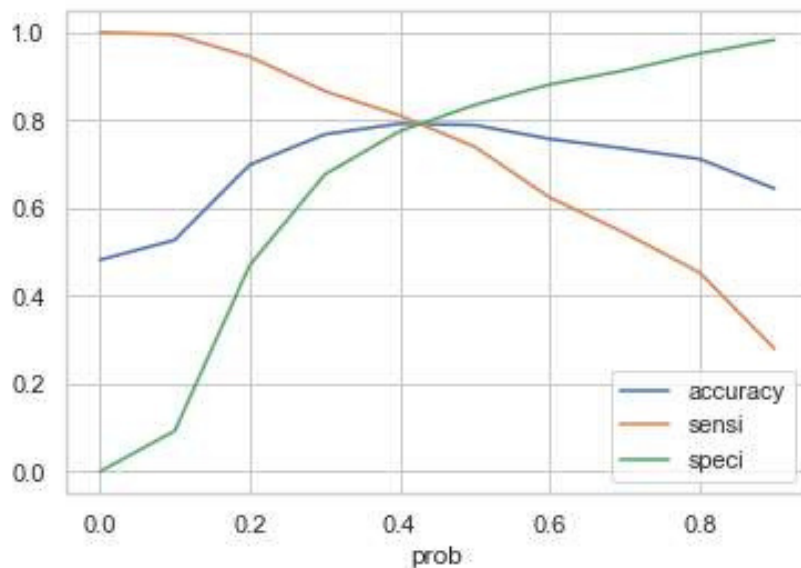Target variable in What is your current occupation

Correlation

- There is no correlation between the variables

Model Evaluation

ROC Curve

- 0.42 is the tradeoff between Precision and Recall -
- Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 42 % to be 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_OlarkChat
- Specialization_Others
- Lead Origin_LeadAdd Form
- Lead Source_WelingakWebsite
- Total Time Spent on Website
- Lead Origin_LandingPage Submission
- What is your current occupation_Working Professionals
- Do Not Email

Conclusion

- We observe that the conversion rate for API and Landing Page submissions is around 30-35%, which is close to the average. However, the conversion rate is significantly lower for leads from the Lead Add Form and Lead Import. This suggests that we should prioritize focusing on leads generated through API and Landing Page submissions.
- Leads who spend more time on the website are more likely to convert.
- We observe that the majority of leads are generated through Google and direct traffic. However, the highest conversion rates come from referrals and the Welingak website.
- The most common last activity among leads is opening an email, while the highest conversion rate is seen with leads who received an SMS. The majority of leads are unemployed, but the highest conversion rates occur among working professionals.