LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

SUBMITTED BY:

1. DHARANI GOALLA
2. AJAYA AJAYA
3. VARSHA RANI

CONTENTS

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

PROBLEM STATEMENT

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

- The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

BUSINESS OBJECTIVE

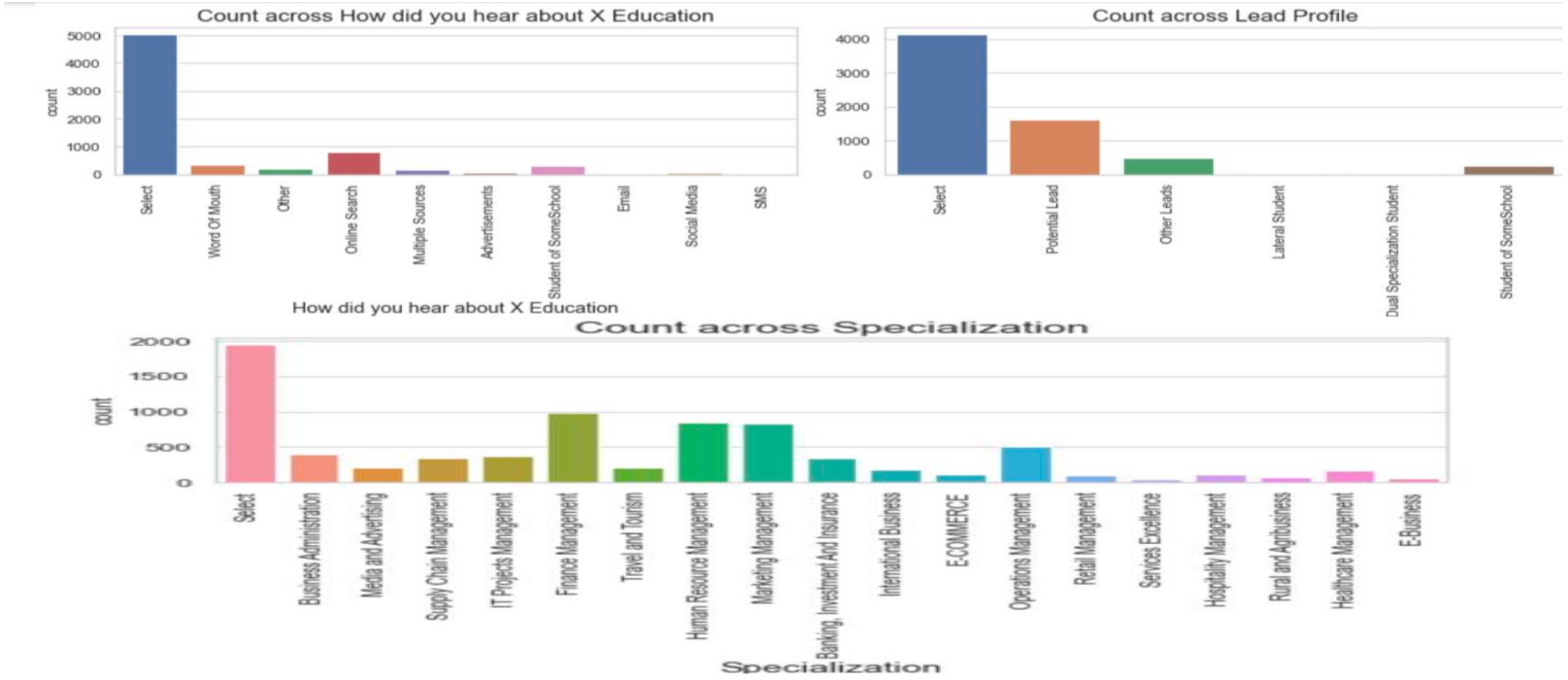
- Lead X has requested the development of a model that assigns a lead score ranging from 0 to 100. This model will enable them to identify "hot" leads and improve their conversion rate.
- The CEO has set a goal of achieving an 80% lead conversion rate.
- In addition to identifying hot leads, Lead X also wants the model to be capable of handling future constraints. This includes determining appropriate actions during peak times, optimizing manpower utilization, and establishing approaches once the conversion target has been met.

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

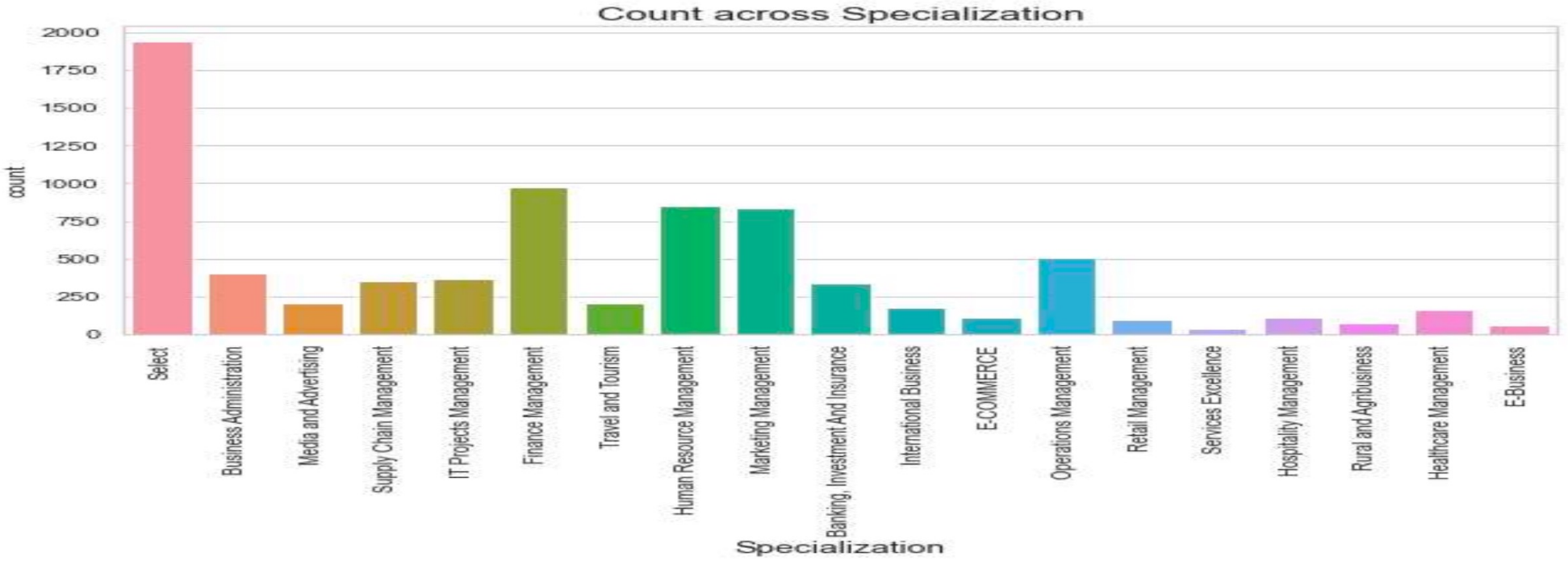
EDA – DATA CLEANING

- There are a few columns in which there is a level called 'Select' which is taking care



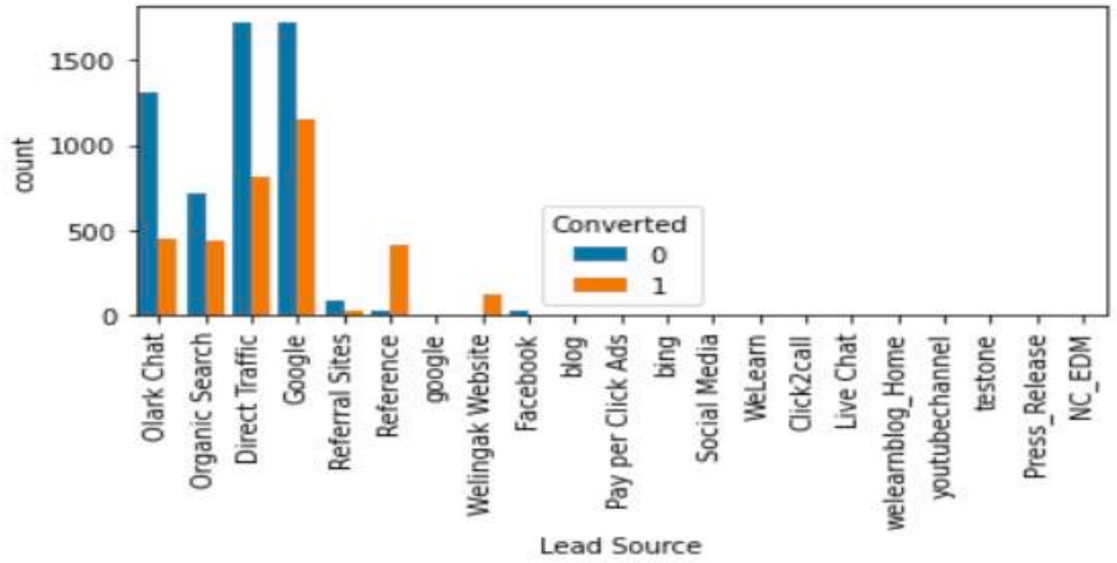
SPECIALIZATION

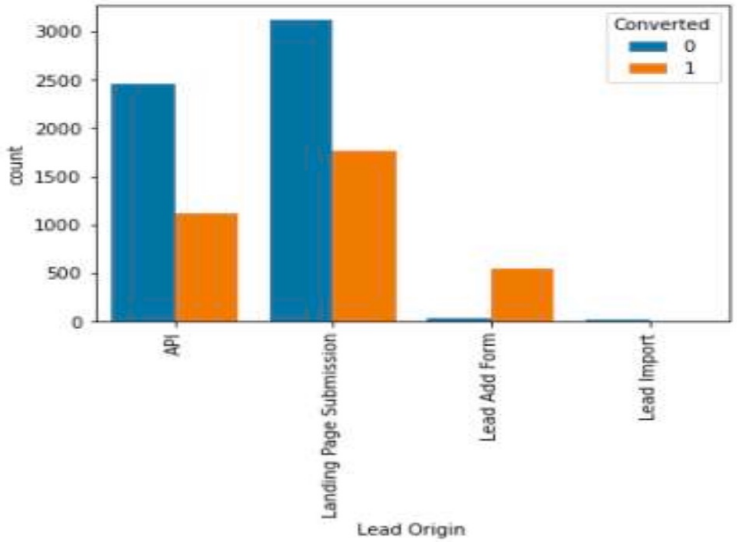
- Leads from HR, Finance & Marketing management specializations are high probability to convert



LEAD SOURCE & LEAD ORIGIN

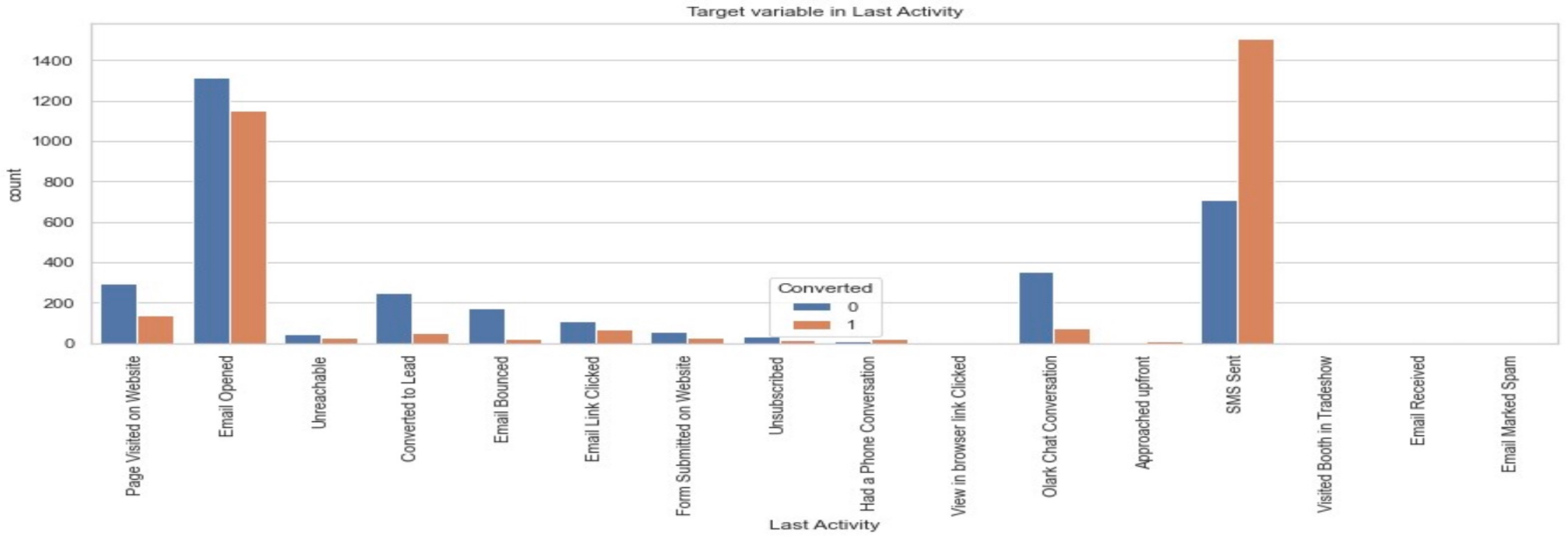
- In lead source the leads through google & direct traffic high probability to convert
- Whereas in Lead origin most number of leads are landing on submission





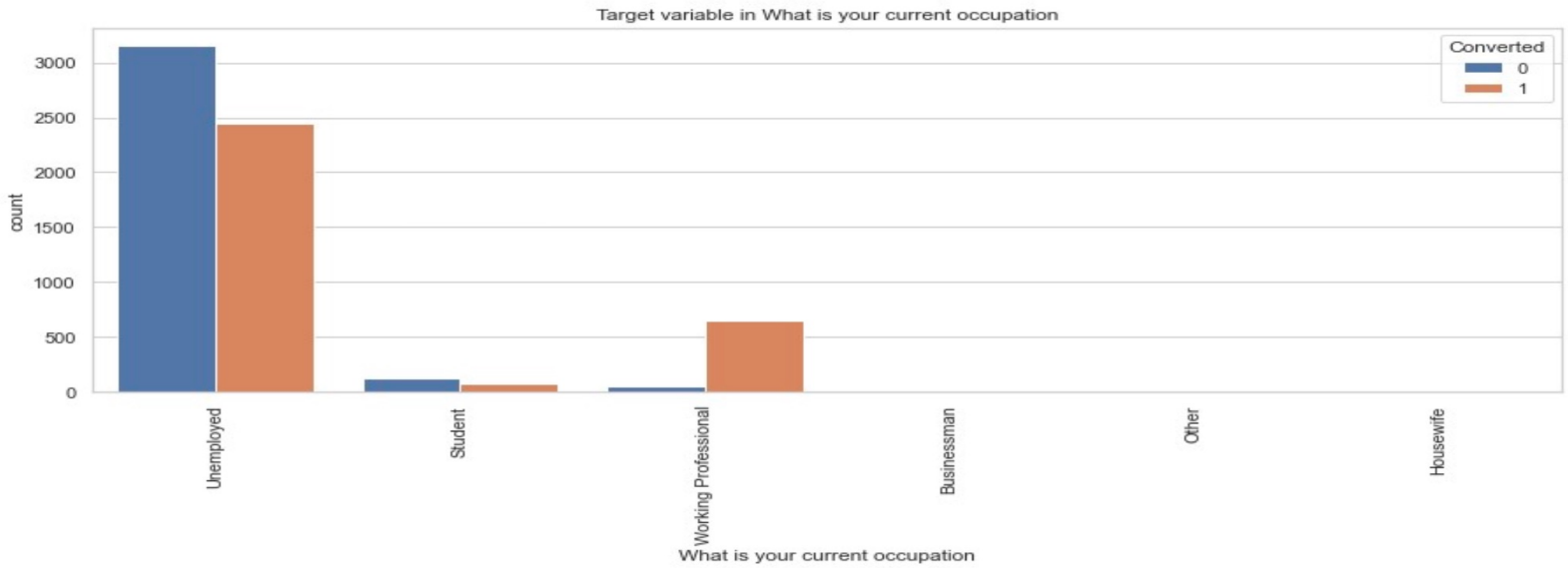
LAST LEAD ACTIVITY

- 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.

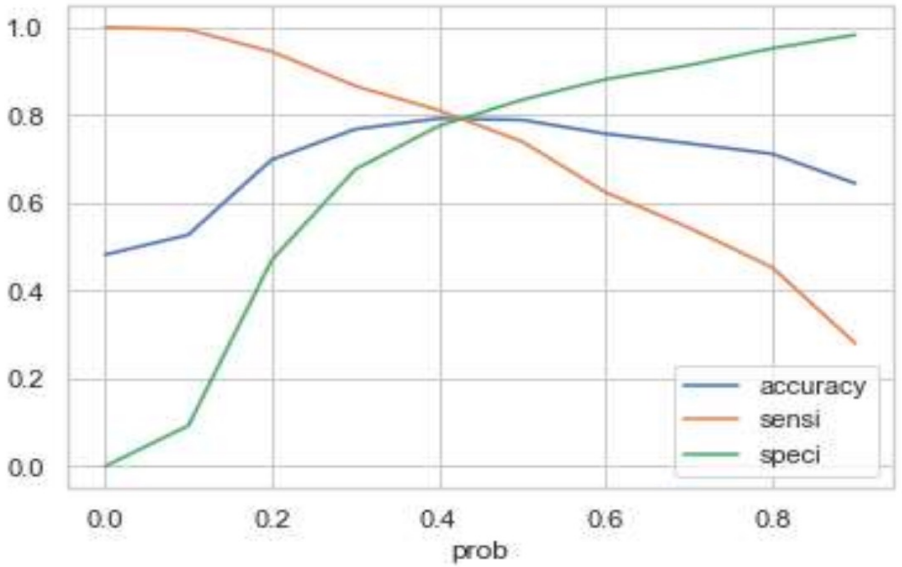


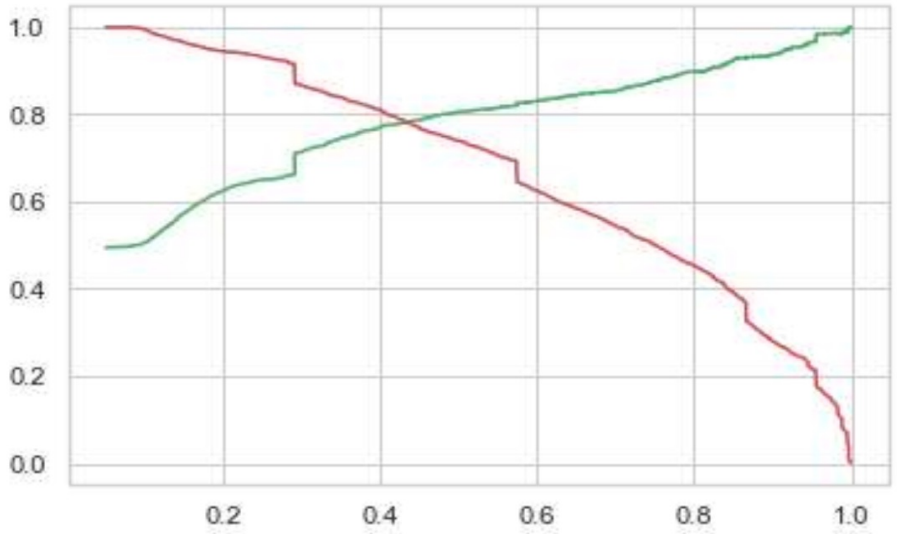
CORRELATION

- There is no co-relation 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%

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 Lead Add forms and Lead imports. Therefore, it is evident that we should prioritize our focus on leads originating from API and Landing page submissions.
- Furthermore, we have noticed that a majority of the leads are generated through Google/direct traffic. The highest conversion ratio is observed for leads generated through references and the Welingak website.
- Moreover, leads who spend more time on the website have a higher likelihood of conversion.
- In terms of the last activity, the most common one is email opened, while the highest conversion rate is observed for leads who received SMS messages. Additionally, a large portion of the leads are unemployed, but the maximum conversion rate is seen among working professionals.