

Lead Scoring Case Study

Submitted By: Prachee Patel

Priyanka Bansal

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

- 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. They have process of form filling on their website after which the company
- that individual as a lead.
- ☐ 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, this means 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

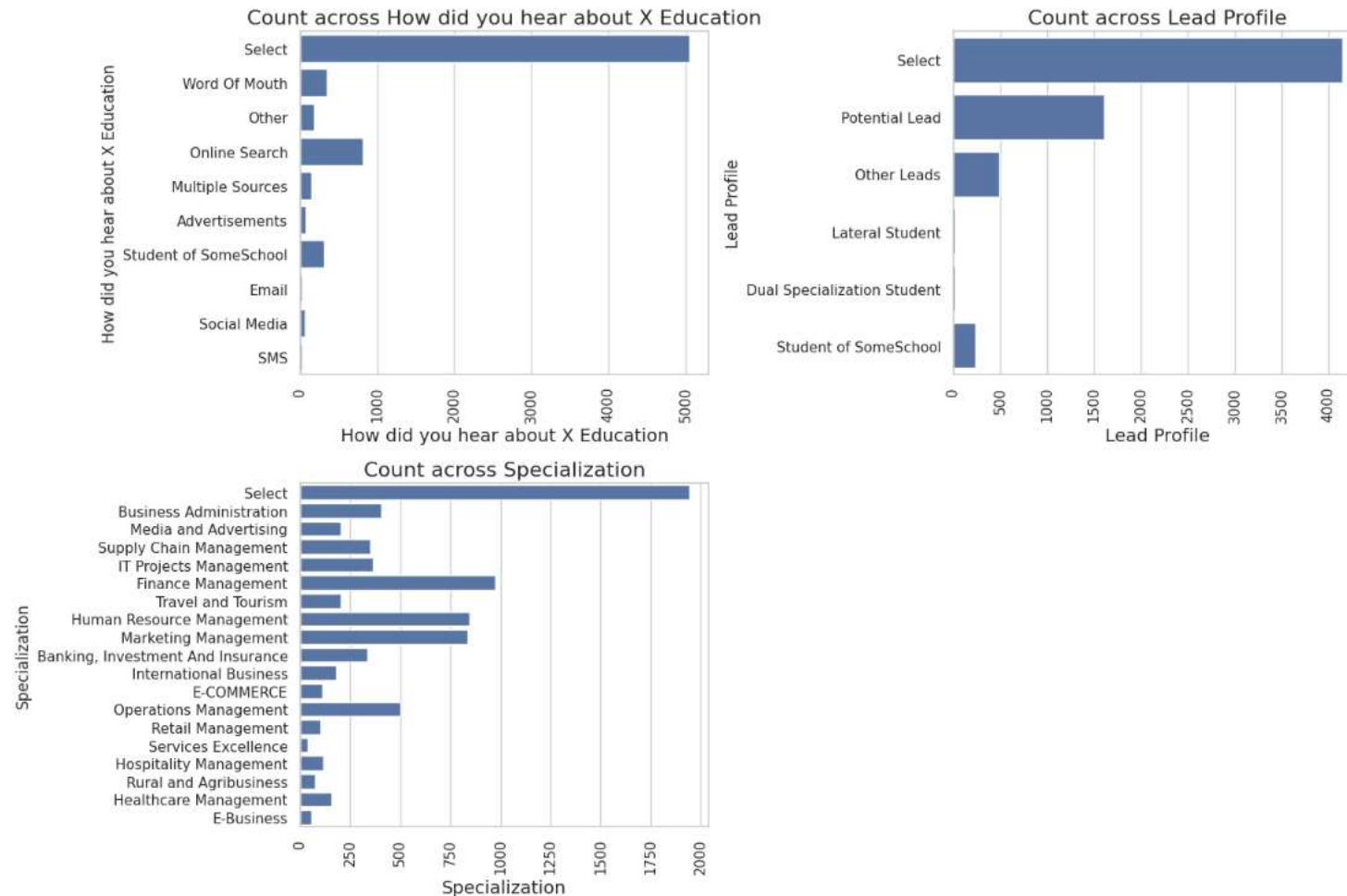
- Business Objective

- ☐ Lead X wants us to build a model to give every lead a lead score between 0 -100 .
- So that they can identify the Hot leads and increase their conversion rate as well.
- ☐ The CEO want to achieve a lead conversion rate of 80%.
- ☐ They want the model to be able to handle future constraints as well like Peak time
- actions required, how to utilize full man power and after achieving target what
- should be the approaches.

- 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

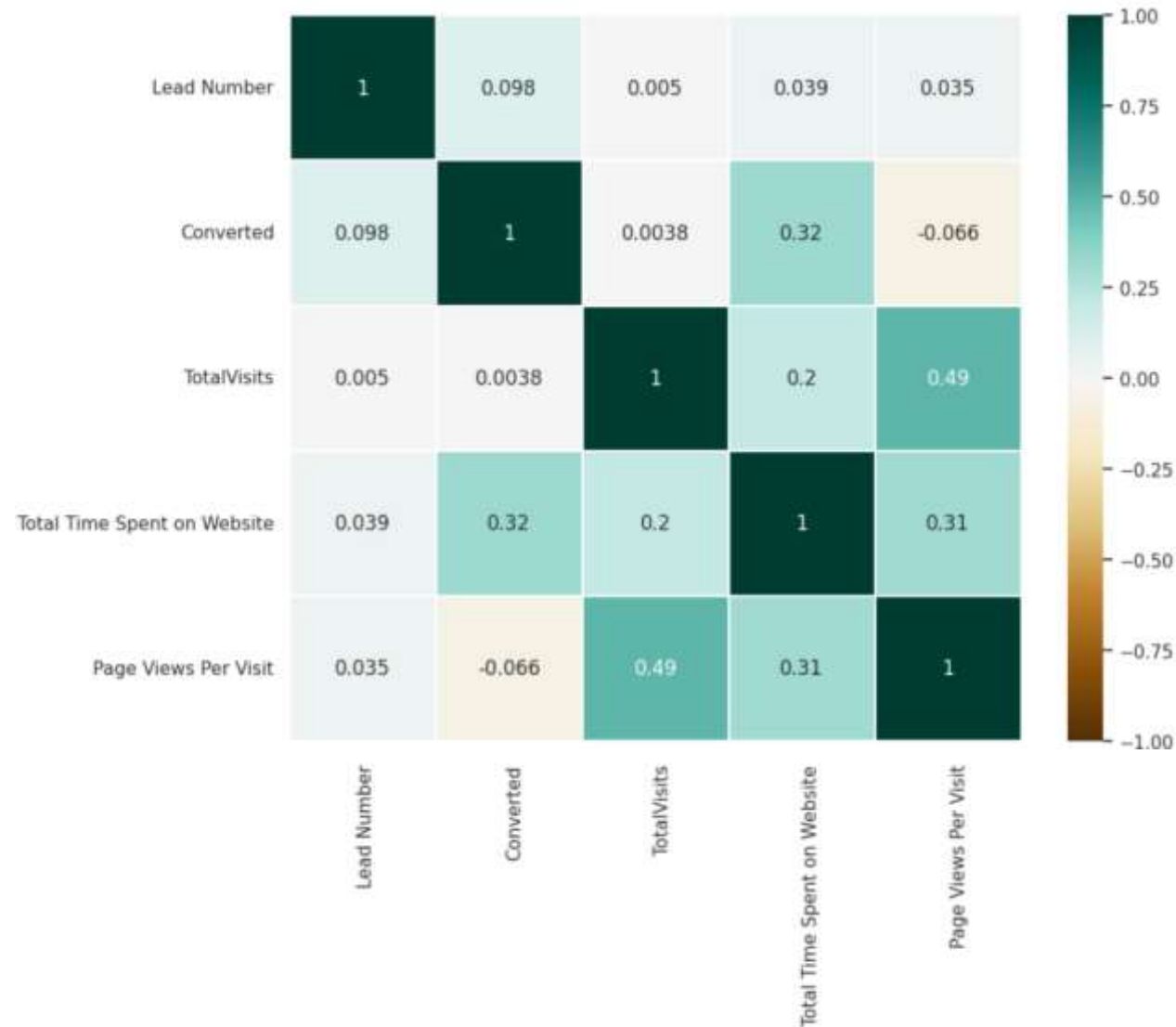
EDA-Data Cleaning

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



Correlation

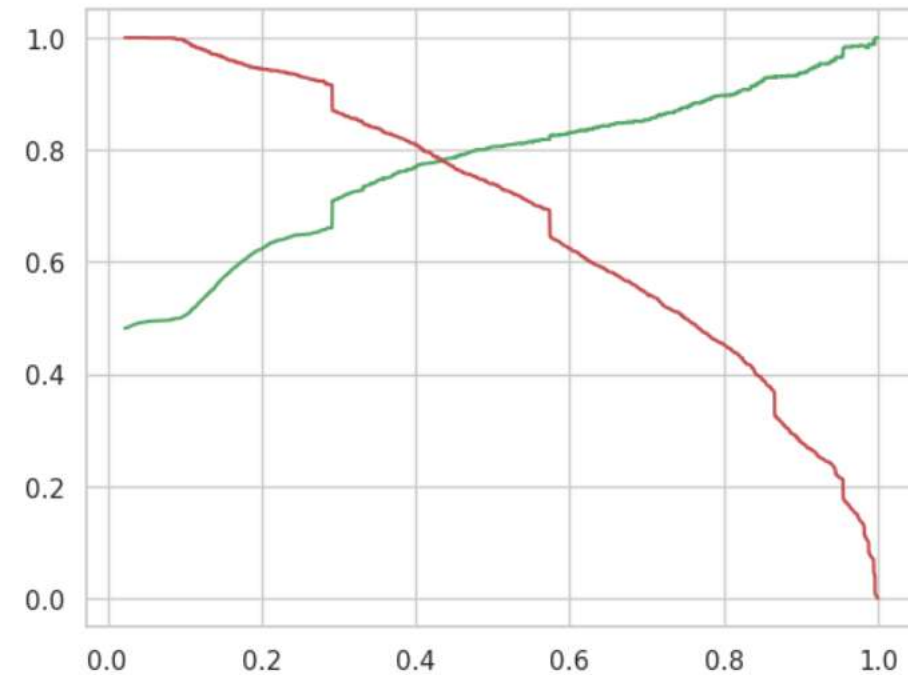
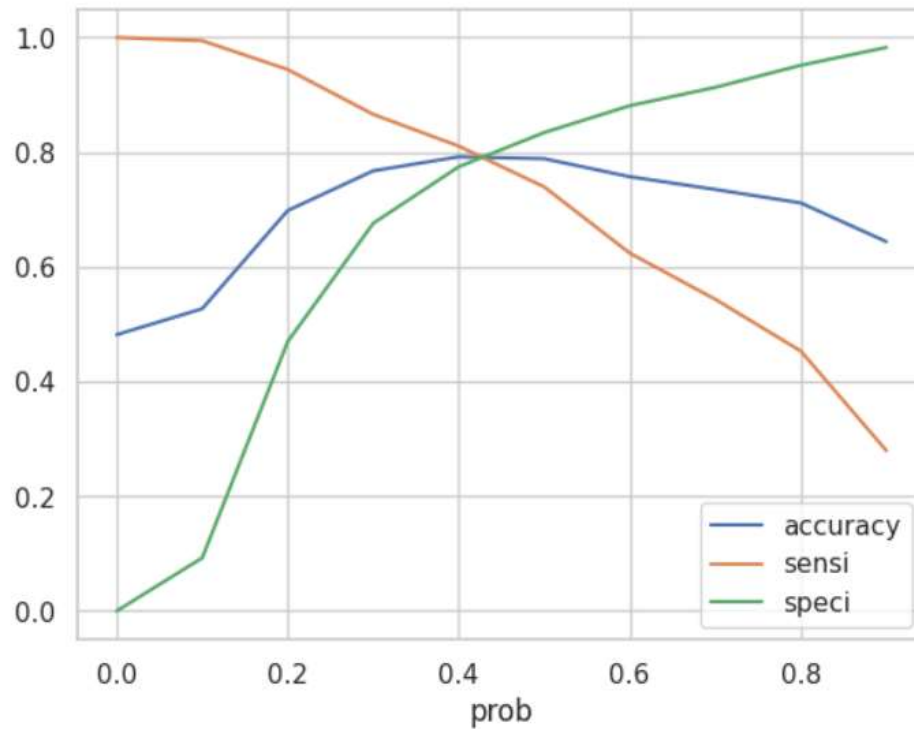
- There is no correlation between the variables



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



Observation

- Train Data

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

- Test Data

Accuracy : 80%

Sensitivity : 77%

Specificity : 80%

Observations

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

- We see that the **conversion rate is 30-35%** (close to average) for **API and Landing page submission**. But **very low for Lead Add form and Lead import**. Therefore we can intervene that we need to **focus more on the leads originated from API and Landing page submission**.
- We see **max number of leads** are generated by **google / direct traffic**. **Max conversion ratio** is by **reference and welingak website**.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. **highest rate = SMS Sent**.
- Max are unemployed. Max conversion with working professional.