- PRESENTEDBY:
 - RAKESH MAITY
 - ANURADHA
 - AJINKYA SURVASE

LEAD SCORING CASE STUDY





PROBLEM

STATEMENT:

named X Education sells online courses to industry professional:



√ While X Education gets a lot of leads, its lead conversion rate is a mere 30%.



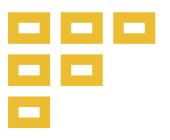
✓ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone



√ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

BUSINESS OBJECTIVE:





✓ Build a logistic regression model to identify hot leads with a ballpark of the target lead conversion rate of ~80%

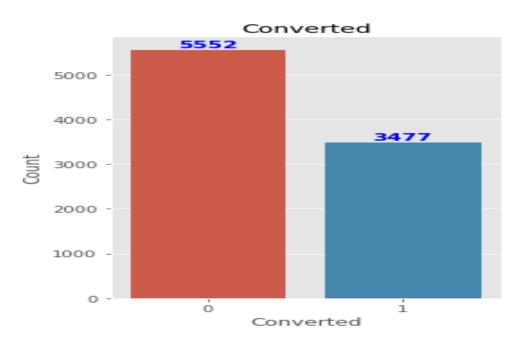
√ Model should be flexible and should be able to incorporate company's future requirements.



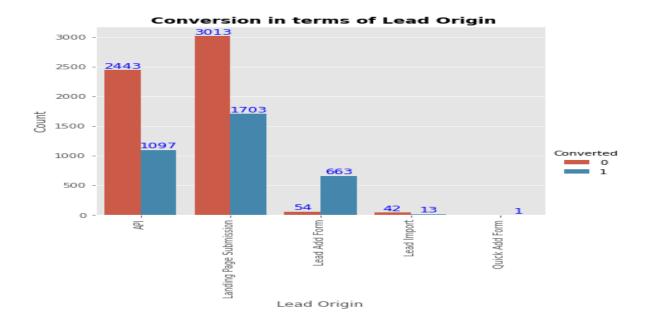
• OVERALL APPROACH:

- ✓ READING AND UNDERSTANDING THE DATA
- ✓ DATA CLEANING AND MANIPULATION
- ✓ EXPLORATORY DATA ANALYSIS
- ✓ TEST TRAIN SPLIT
- ✓ DUMMY VARIABLE CREATION AND FEATURE SCALING
- ✓ BUILDING A LOGISTIC REGRESSION MODEL
- ✓ MODEL EVALUATION: SENSITIVITY, SPECIFICITY AND PRECISION
- ✓ PREDICTIONS ON THE TEST SET
- ✓ INSIGHTS AND RECOMMENDATIONS

EXPLORATORY DATA ANALYSIS



Lead conversion is around 39% for Company

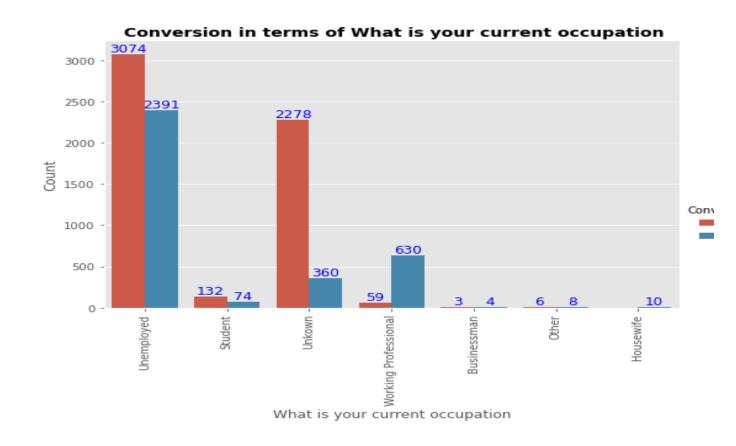


LEAD ORIGIN:

- Leads from 'API' and 'Landing Page Submission' are high; conversion rates are poor at 31% and 36% respectively.
- Though leads from 'Lead Add Form' is low, the conversion rate is very high at ~92%.

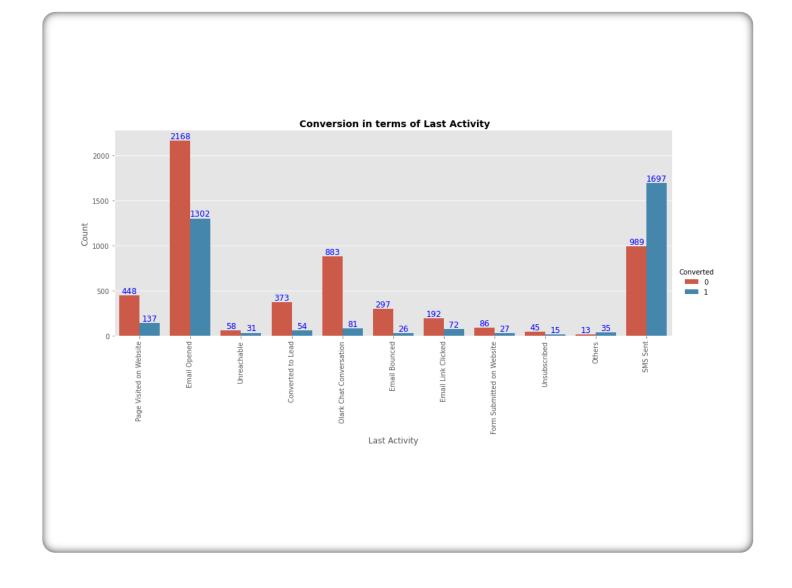
EXPLORATORY DATA ANALYSIS

- Working Professionals have the highest conversion rate at ~91% followed by students at ~86%.
- Unemployed leads have low conversion rates but they generate maximum leads counts.



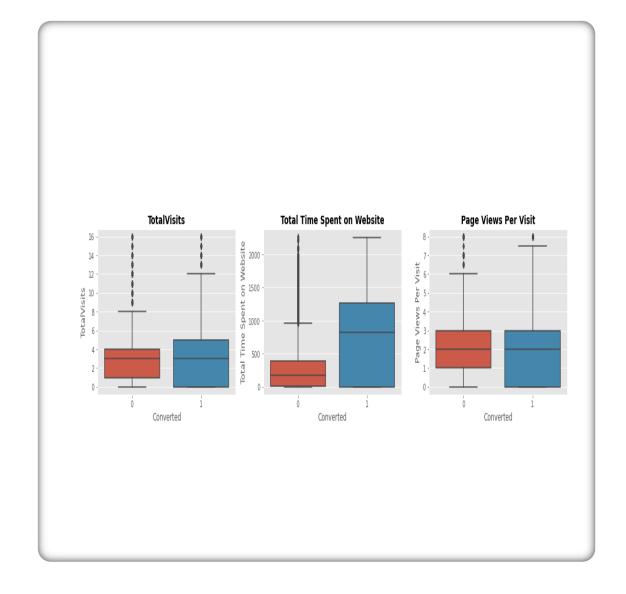
EXPLORATORY DATA ANALYSIS

• LAST ACTIVITY: - Maximum leads are generated from people with last activity -Email opened and SMS sent. -Conversion rate is highest for SMS Sent (~63%), where as it is only ~38% for Email Opened. - Olark chat conversation and Page Visited on Website generates significant number of leads but their conversion rates are extremely low at 8% and 38% respectively



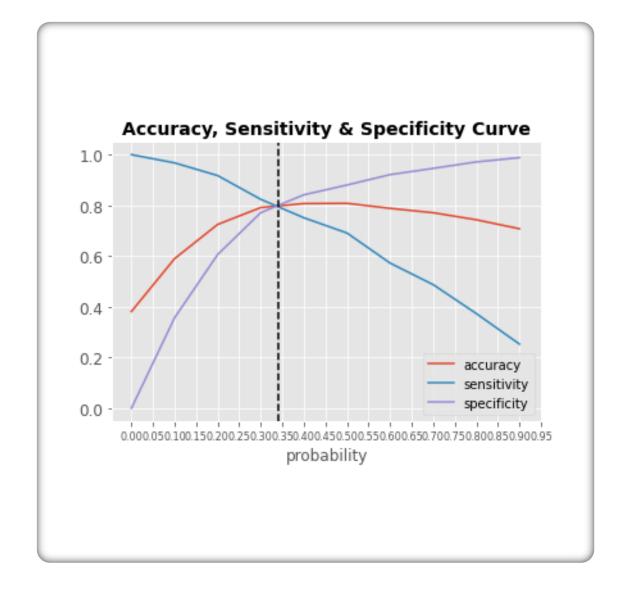
EDA

- - Total Visits: Median for converted and non-converted leads are same. People who visits the platform have equal chances(50-50) of applying and not applying for the course.
- - Total Time Spent on Website: People spending more time on website have more chances of opting for a course People who spend less time on the website didn't opt for any courses.
- Page Views Per Visit: Median for converted and non-converted leads are same.

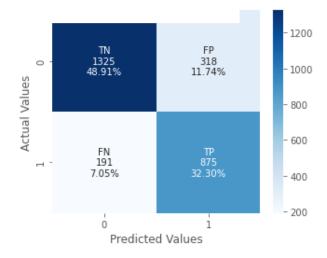


FINDING THE OPTIMAL CUT-OFF POINT:

• The above graph depicts the optimal probability cutoff of 0.34 based on accuracy, specificity and sensitivity analysis.



TRAIN DATA



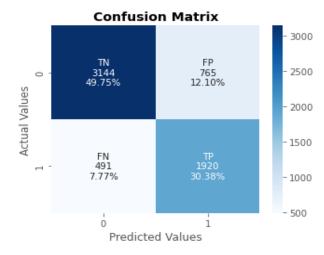
Model Accuracy : 80.1 % Model Sensitivity : 79.6 % Model Specificity : 80.4 % Model Precision : 71.5 %

MODEL EVALUATION:

✓ There is only ~2% difference on train and test data's performance metrics, implying that the final model didn't overfit training data and is performing well.

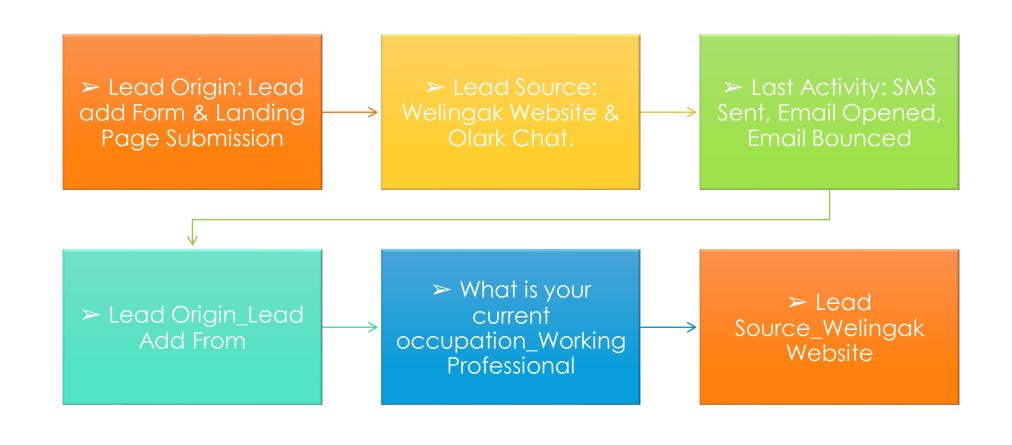
✓ High Sensitivity will ensure that almost all leads who are likely to Convert are correctly predicted.

TEST DATA



Model Accuracy : 81.2 %
Model Sensitivity : 82.1 %
Model Specificity : 80.6 %
Model Precision : 73.3 %

INSIGHTS AND RECOMMENDATIONS: VARIABLES THAT SHOULD BE FOCUSED



Thank you

