LEAD SCORE Case Study

Presented By: Chaitra S and Gaziya Gani Khan

X Education | Data Science Division

Business Problem Solving Approach

- ➔ Frame the Problem with Precision
- ➔ Analyze the Problem
- ➔ Analyze Possible Solutions
- ➔ Implement the Solution
- ➔ Result Explanation: Call to Action

1. Frame the Problem with Precision

Problem Statement

We, X Education, offer online courses to industry professionals. We acquire leads from several websites and search engines. Currently, having 30% lead conversion rate and needs a model to identify potential leads to achieve 80% lead conversion rate through efficient data driven action.

➔ Essentially need to identify:

- The factors behind high potential leads
- Lead Score for every lead to identify whether a lead get converted or not

2. Analyze the Problem

After checking the available structured data, Asked Macro Questions:

- Why do industry professionals need online courses?
- How many courses they need to take to stay relevant in their domain?
- What will be the future of work in different domains and how the rest of the world can change?

Then Asked Micro Questions:

- Why X Education facing low lead conversion rate [38%] issue?
- What kind of courses does X Education offer?

Built Hypothesis:

Given data seems to have even distribution of lead related information hence sales team couldn't distinguish potential leads which end up random calls resulting in low conversion.

Validate the Hypothesis

Sales-team have **high leads** Sales-team receive **different responses** when called

Root-Cause of Issue

Zero idea about whether a lead get **converted or not**

3. Analyze Possible Solutions

As per the hypothesis Sales team is unable to distinguish potential lead thus need

- higher interpretability to solve the problem - sales data have high frequency and high importance for sustainable business operation

therefore, a supervised Logistic Regression ML model would be better for identifying converted lead as the provided data is also sufficient to build a logistic model

Simple EDA can be helpful in identifying converted lead and not-converted lead

Logistic Regression model can be helpful in identifying converted lead

Random forest model or Support vector machine can be used to identify converted lead

Key metrics of Model

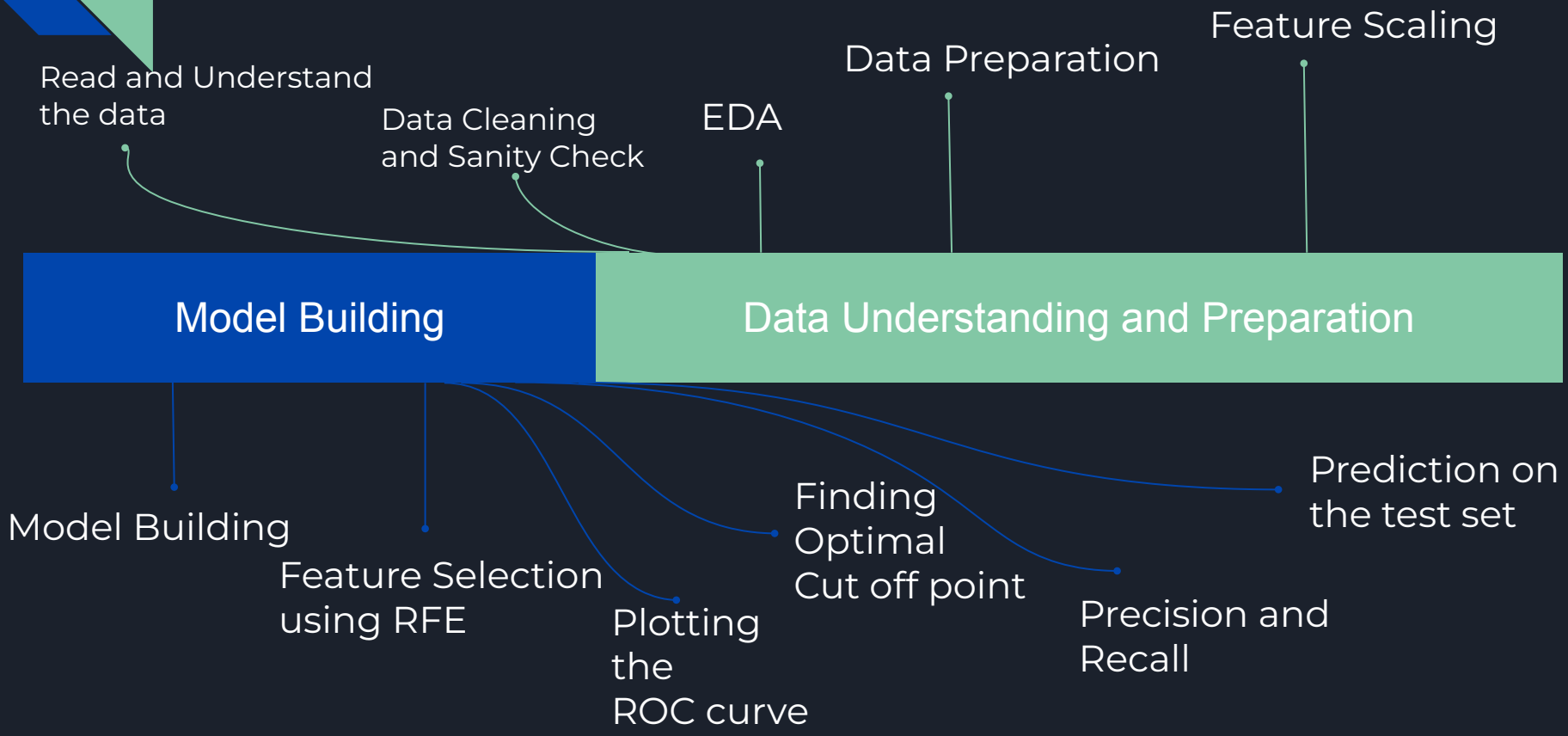
Confusion Matrix **Type I Error, Type II Error Accuracy**

Recall & Precision

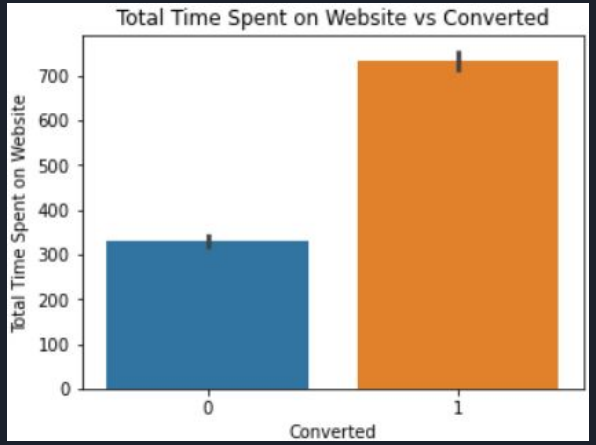
Sensitivity & Specificity

ROC Curve- AUC Score

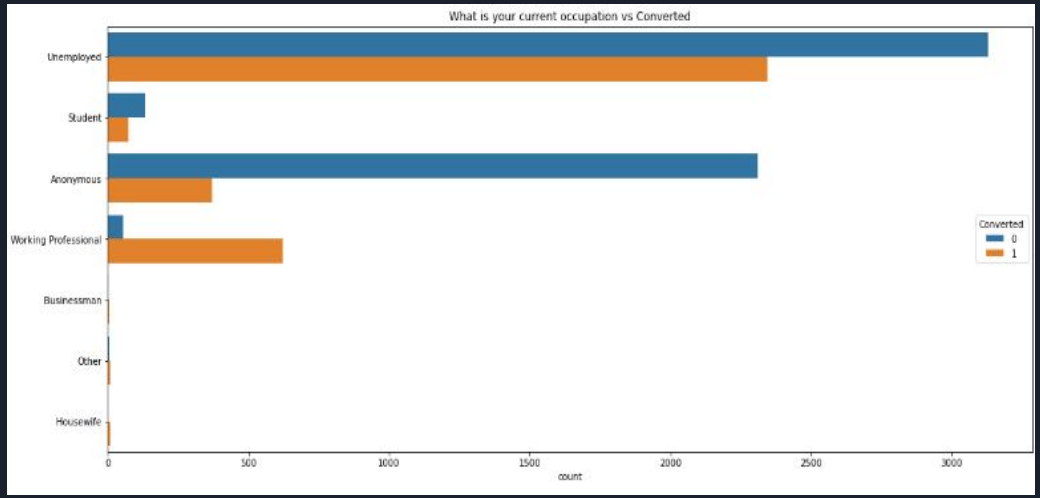
4. Implement the Solution



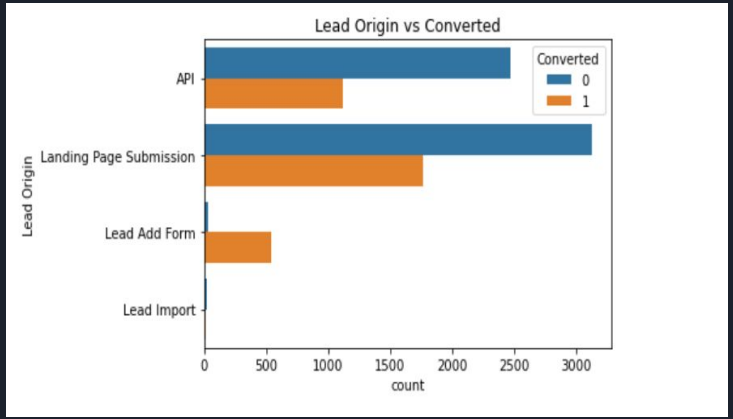
- While EDA identified one of the KPIs as Total Time Spent on Website where converted leads usually spend maximum time on website



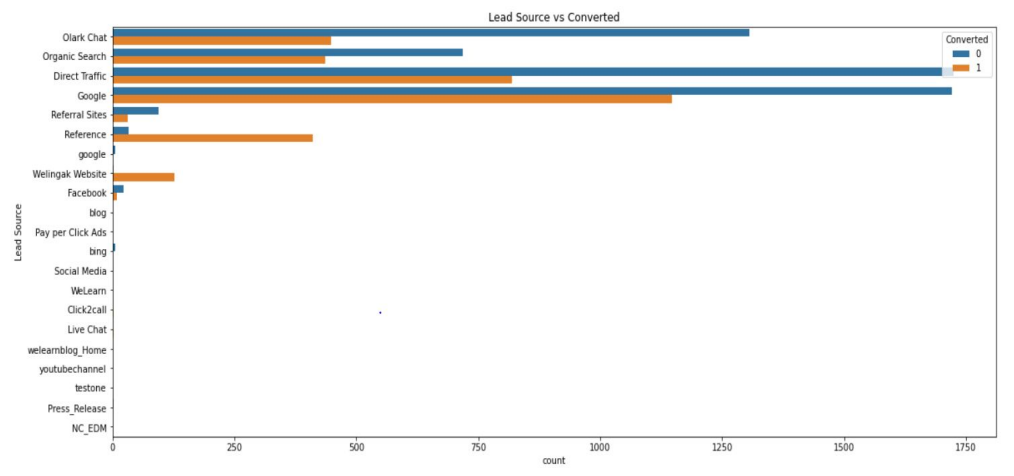
- While EDA identified another KPI as Working Professional people as this occupation has maximum ratio of converted leads to not-converted



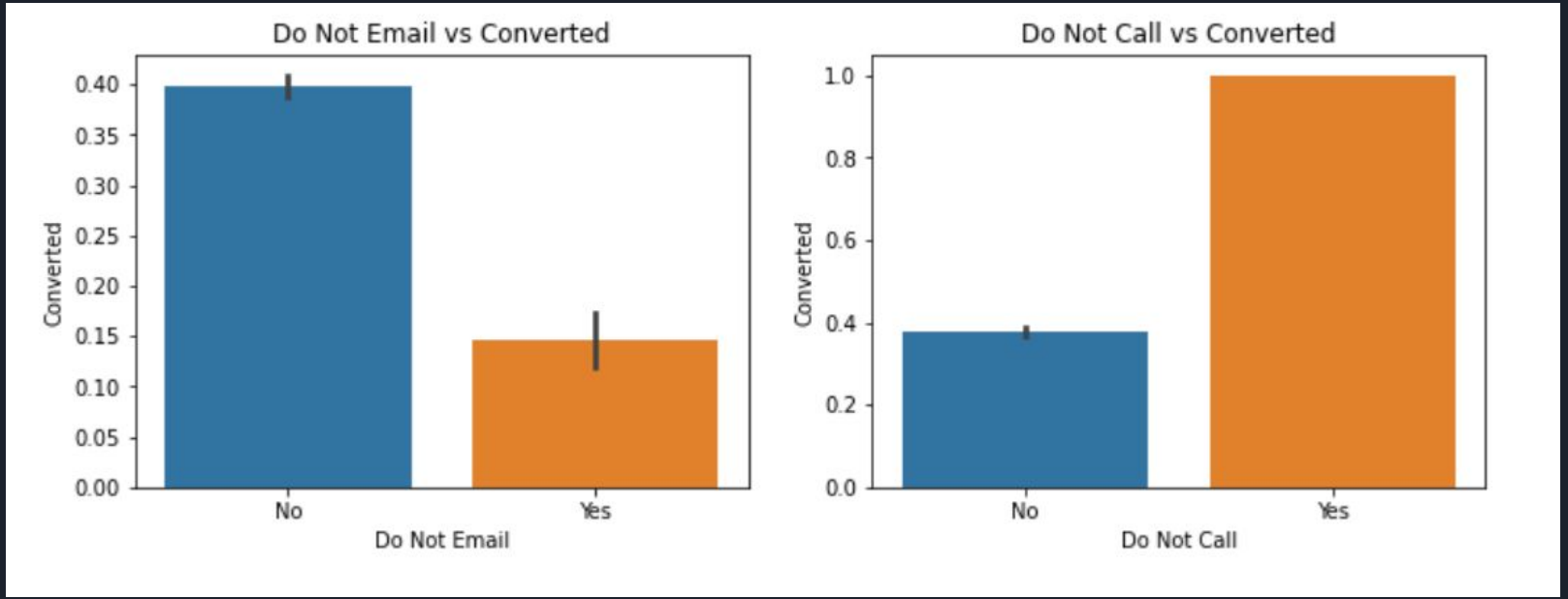
- While EDA identified another KPIs as Landing Page Submission and API where there is maximum lead conversion rate



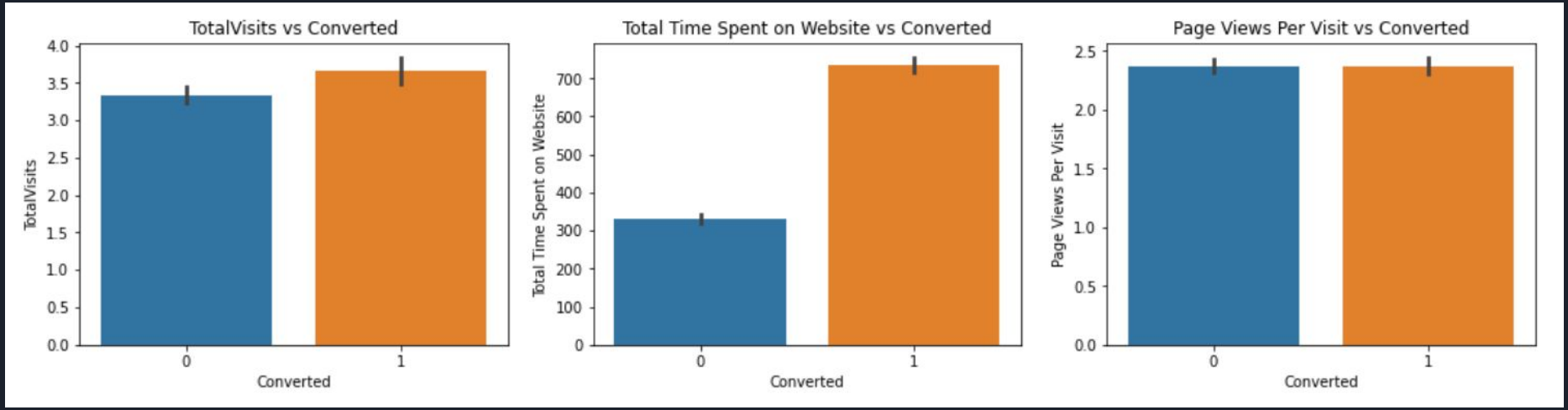
- We noted by the analysis of Lead Source with the Converted that the maximum conversion is from the type Google followed by Direct Traffic



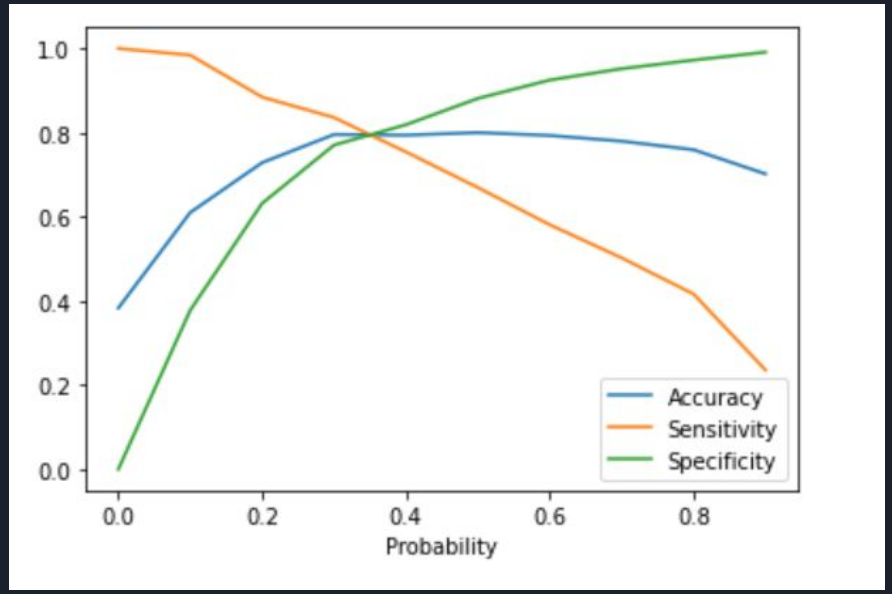
From the below plots we can note that the conversion rate of Do Not call is higher when compared to Do Not Email



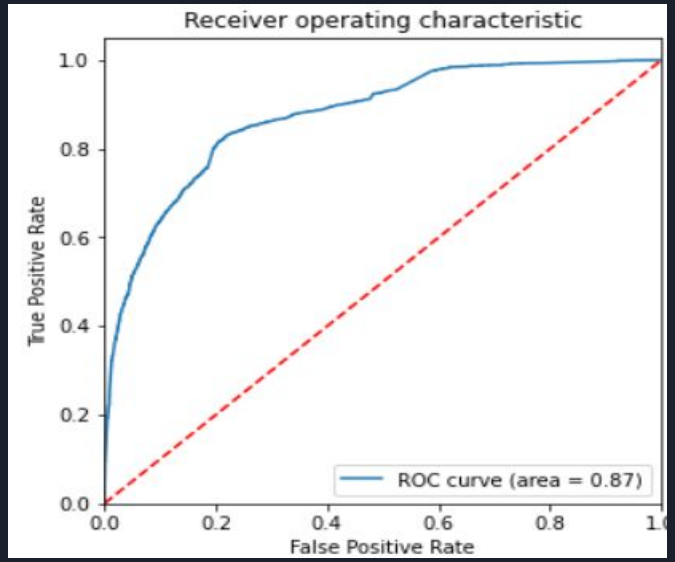
From the below plots we can note that the conversion rate were high for TotalVisits, Total Time Spent on Website and Page Views Per Visit

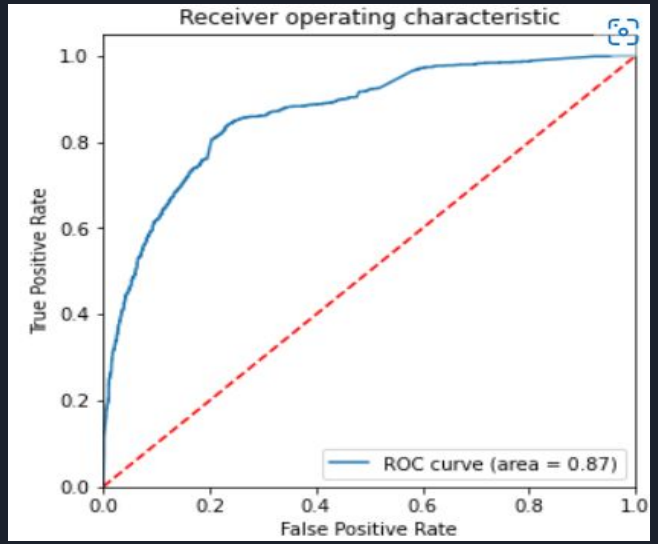


Accuracy, Sensitivity and Specificity on Train Data Set



From the above plot, we can note that the graph predicts approximately an optimal cut off of 0.3 based on accuracy, specificity and sensitivity





- ROC curve (train_set), classifiers give curve closer to the top-left corner indicate a better performance. Hence model has high sensitivity which help achieve the 80% target

- ROC curve in test_set, classifiers also give curve closer to the top-left corner indicate a better performance. Thus model is ready to implement

5. Result Explanation: Call to Action

Notable data driven inferences drawn from the data analysis and logistic regression to help sustain X Education revenue

- Identification of working professionals and Businessman class as highest hot leads help achieve the target conversion rate
- Removal of push emails help attract more leads with less cost as around 90% non-converted leads opted for not to email
- While investments should be made to improve the quality of calls as around 95% leads converted opted for call service
- Sales team should only make calls based on Lead score provided by model

This would help achieve around 80% lead conversion rate and give X Education a competitive edge over others

Thank you !