Lead Scoring Case Study

SUBMITTED BY:
SHUBHAM GAUR
RAHUL DESHPANDE
NITIN SHYAM

Lead Scoring Case Study

Problem Statement:

X Education sells online courses to industry professionals. 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%.

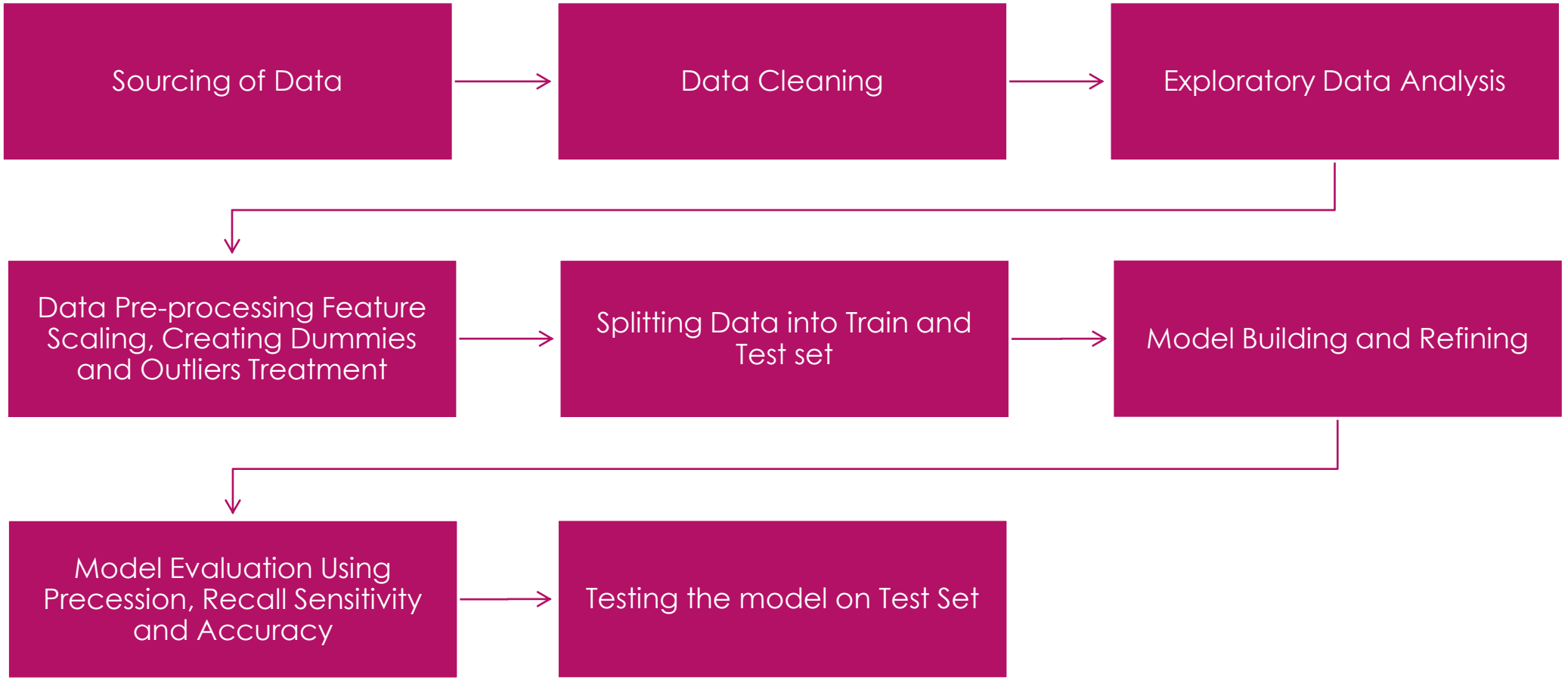
Business Objective

Identifying the most promising leads and tag them "Hot Leads".

Create a Machine Learning model that will assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

Target lead conversion rate should be around 80%

Strategy


Sourcing of Data
Data Cleaning
Exploratory Data Analysis
Data Pre-processing Feature Scaling, Creating Dummies and Outliers Treatment
Splitting Data into Train and Test set
Model Building and Refining
Model Evaluation Using Precession, Recall Sensitivity and Accuracy
Testing the model on Test Set

Problem Solving Methodology

Data Sourcing, Cleaning and Preparation.

- Reading Data
- Data Cleaning
- Removing Duplicates
- EDA
- Outliers Treatment

Feature Scaling and Splitting Data into Train and Test set.

- Creating Dummies
- Feature scaling
- Test train Split.

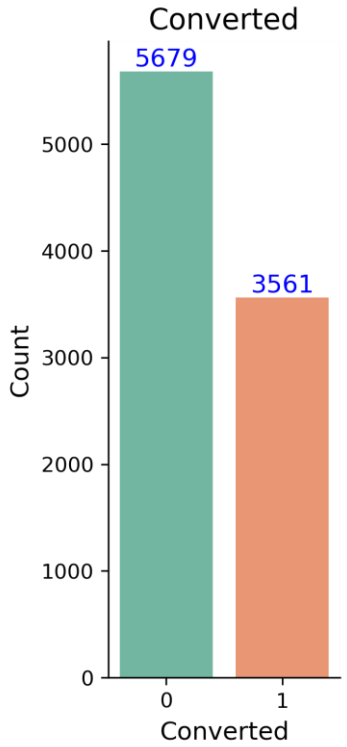
Model Building.

- Creating Logistic Regression Model
- Feature Selection using RFE.
- Determining optimal Model
- Evaluating the model using Precision, Recall, Accuracy and Sensitivity.

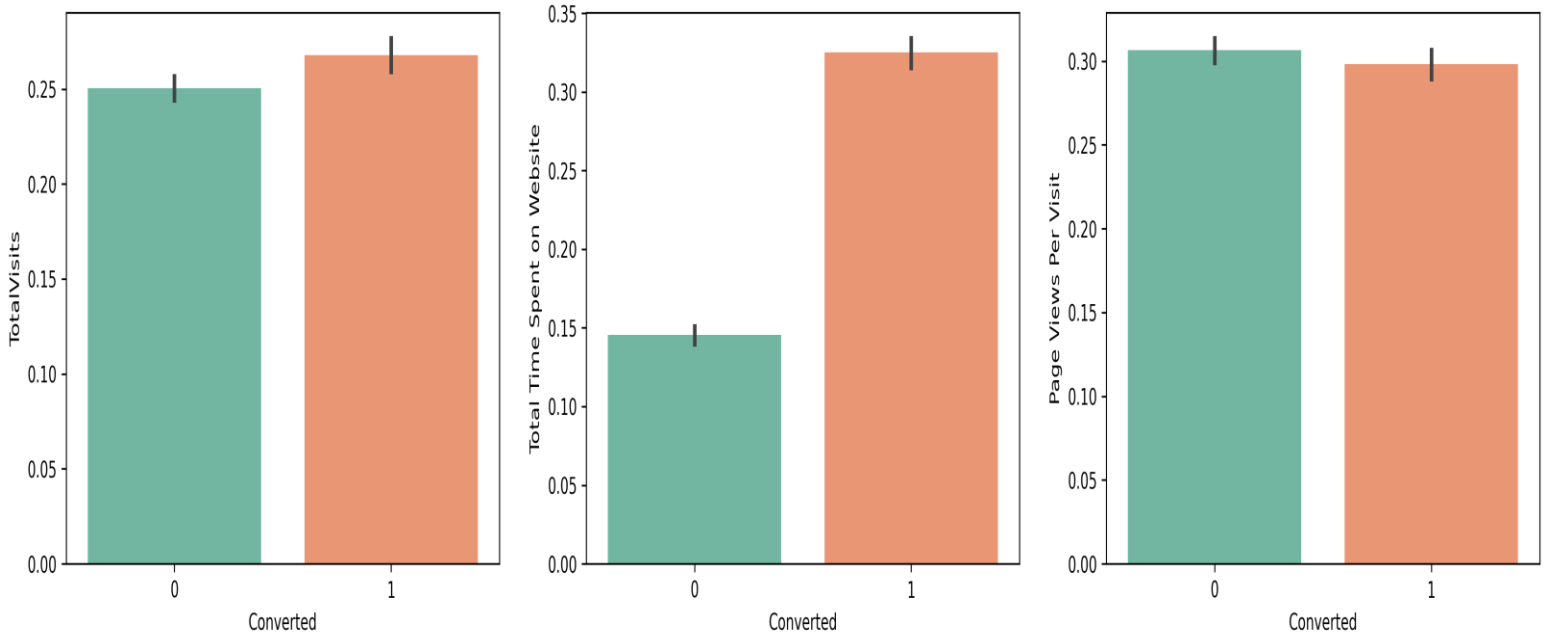
Model Evaluation and Conclusion.

- Determining the "Lead Score" and ensuring that "Hot Leads" have conversion rate of 80%
- Evaluating the model on train set, using Precession and Recall.

Exploratory Data analysis

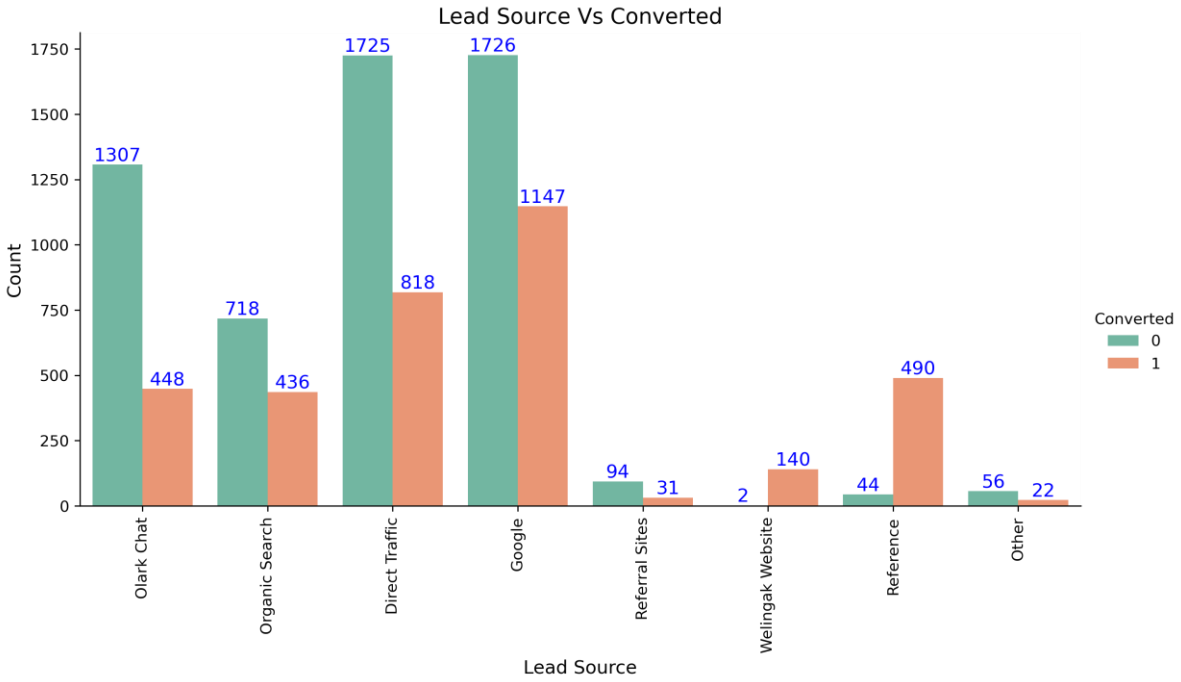


Initial conversion rate is around 39%

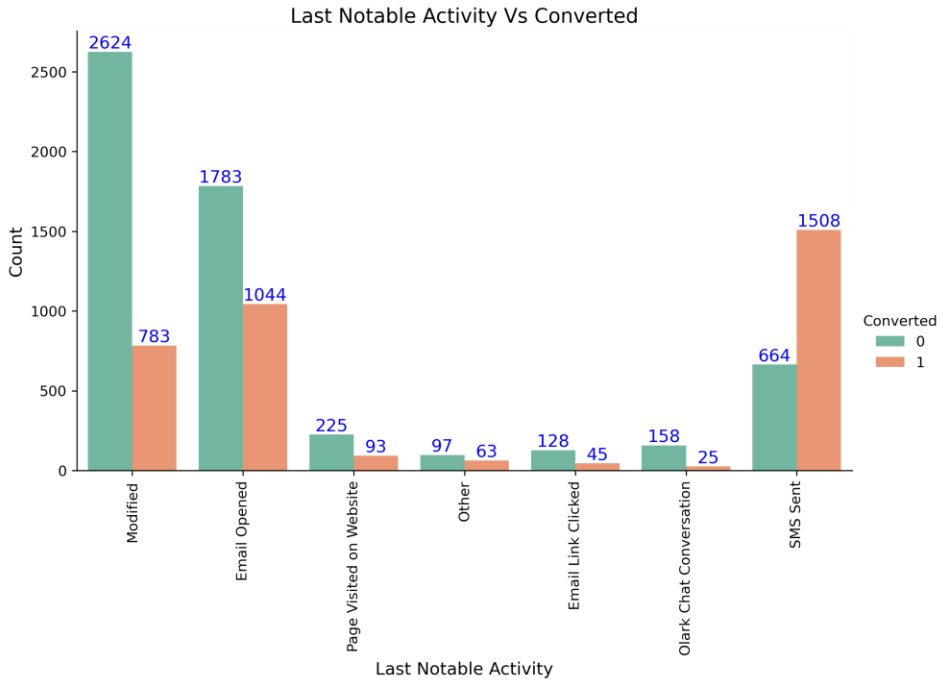


Average "Total Time Spent on Website" by converted leads is more than twice that of non converted leads.

Exploratory Data analysis

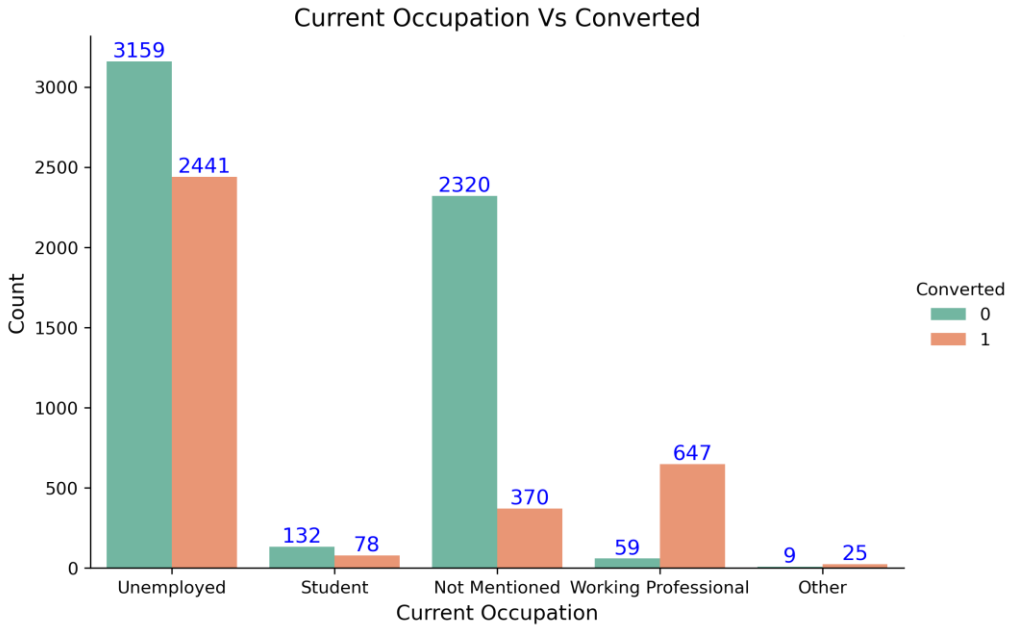


Rate of conversion from "Refrence" is around 92% and "Welingark Website" is 98.5% , but total no of cases less than 1%

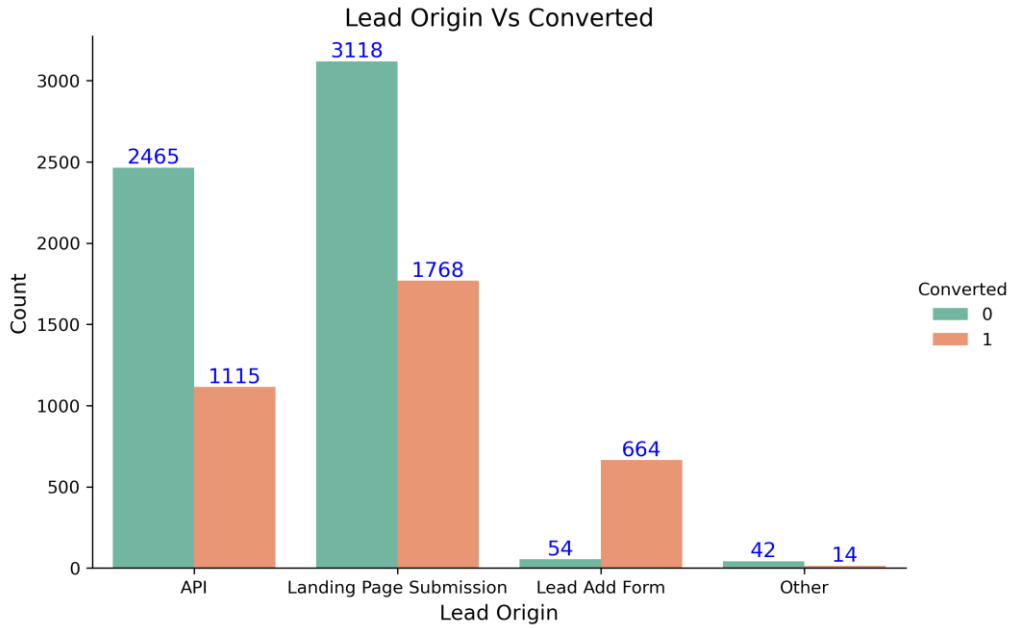


conversion rate is around 70% for "SMS Sent"

Exploratory Data analysis



Conversion rate for Working Professional" is around 90%.

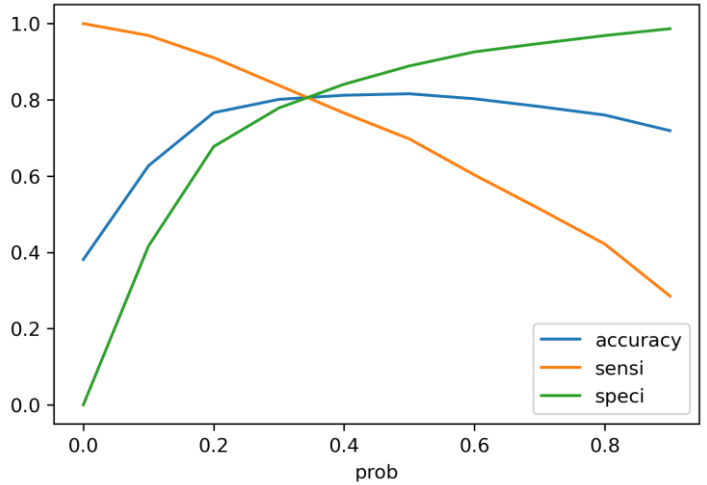


Maximum conversion happened from Landing Page Submission
Rate of conversion for "Lead Add Form" is 92.5%.

Attributes Impacting Conversion Rate

S.no	Attribute	Impact Factor
1.	Total Time Spent on Website	4.50
2.	Lead Origin_Lead Add Form	3.80
3.	What is your current occupation_Working Professional	3.56
4.	Lead Source_Welingak Website	2.10
5.	Last Notable Activity_SMS Sent	1.91
6.	What is your current occupation_Other1	1.80
7.	Lead Source_Olark Chat	1.51
8.	What is your current occupation_Unemployed	1.00
9.	What is your current occupation_Student	0.99
10.	TotalVisits	0.98
11.	Last Notable Activity_Other	0.91
12.	Last Notable Activity_Email Opened	0.70
13.	City_Not Mentioned	-0.37
14.	A free copy of Mastering The Interview	-0.46
15.	Specialization_Hospitality Management	-1.00

Model Evaluation on Train Dataset

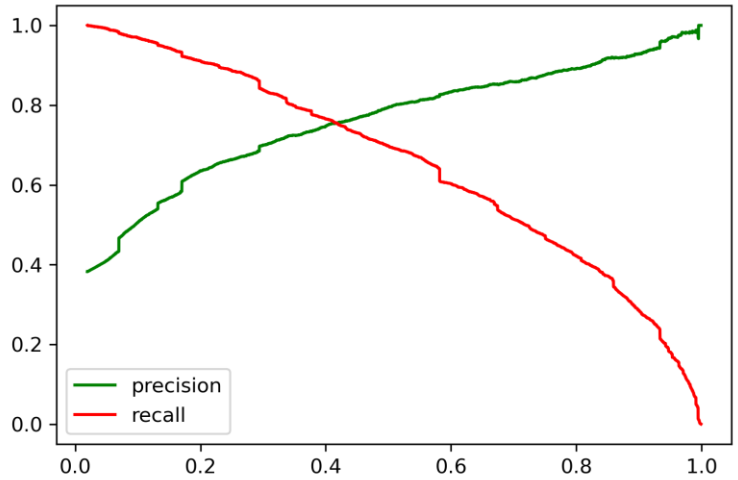


The graph depicts optimal cutoff of .34 based on Accuracy, Sensitivity and Specificity

3240	762
Confusion Matrix	
482	1984

Accuracy - 80.7%
Sensitivity - 80.4%
Specificity - 80.9%
Precision - 72.2%

Precession less than required 80%



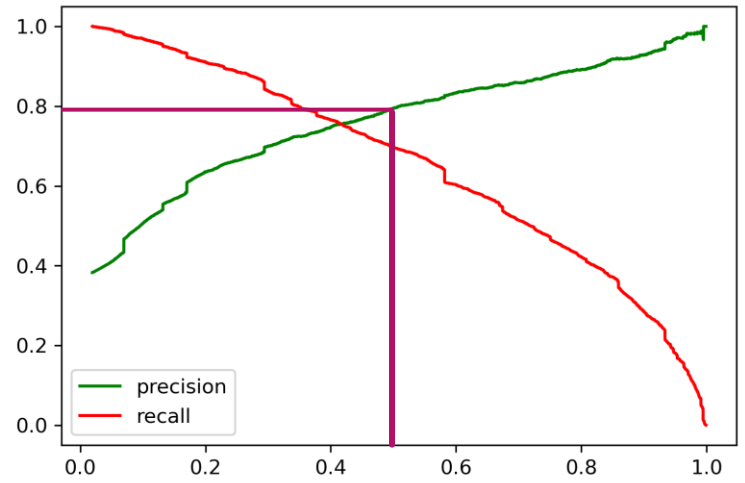
The graph depicts optimal cutoff of .42 based on Precision and Recall

3404	598
Confusion Matrix	
613	1853

Accuracy - 80.7%
Sensitivity - 75.4%
Specificity - 85.0%
Precision - 75.4%

Precession less than required 80%

Model Evaluation on Train Dataset



The graph depicts precision at cutoff of .49 which is around 79% close to required highlighted in red

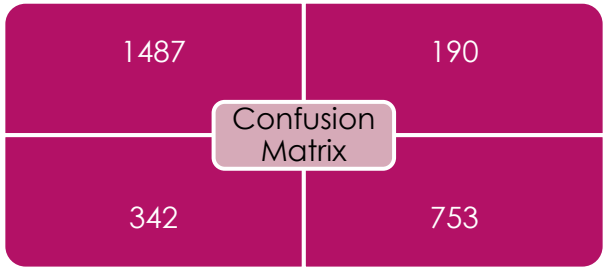
3240	762
Confusion Matrix	
482	1984

Accuracy - 81.5%
Sensitivity - 70.4%
Specificity - 88.4%
Precision - 78.9%

Model Evaluation on Test Dataset

Evaluation Metrics for test Set for cutoff of .49

Accuracy -	81.5%
Sensitivity -	68.7%
Specificity -	88.7%
Precision -	79.8%



No overfitting: Accuracy of test set is similar to train set

Precision of 79.8% at cut-off of .49 close to required 80%

Conclusion and Findings

- Top three variables as predicted by the model are
  - Total Time Spent on Website
  - Lead Add Form from Lead Origin
  - Working Professional from What is your current occupation
- From business perspective, X education need to focus on running campaigns that direct user to landing page for submission and their target should be working professionals who are looking for job change or carrier change as an edge.
- The conversion rate for leads coming from reference is around 92%, therefore X education should consider incentivizing referral.
- Conversion rate for leads coming through Welingark website is around 98.5% compared to 40% for leads coming through Google.
- Conversion rate for leads originating from add form is 92.5%.
- While we have checked both Sensitivity-Specificity, and precession-recall metrics, for optimal cut-off, the conversion rate in both cases was less than 80% for hot leads, therefore optimal cutoff of .49 was conspired to get the required conversion rate of 80%.
- Accuracy, sensitivity, specificity and precession of test and train set are comparable therefor the model is not overfitting on the train data.
- If the lead score as predicted by the model is grater than 49 the lead should be considered as “hot lead” with high probability of conversion.
- Overall performance of the model seems good.