

# **Lead Score 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%.

## **Strategy**

- ↪ Source the data for analysis
- ↪ Clean and prepare the data
- ↪ Exploratory Data Analysis
- . ↪ Feature Scaling
- ↪ Splitting the data into Test and Train dataset.
- ↪ Building a logistic Regression model and calculate Lead Score.
- ↪ Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall
- ↪ Applying the best model in Test data based on the Sensitivity and Specificity Metrics

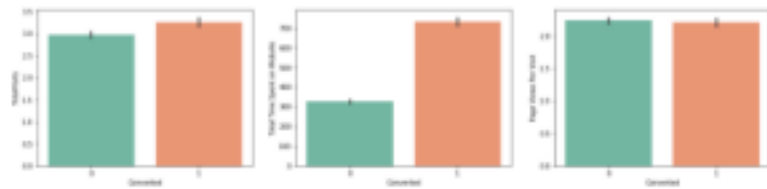
## **Problem solving methodology**

## Exploratory Data Analysis

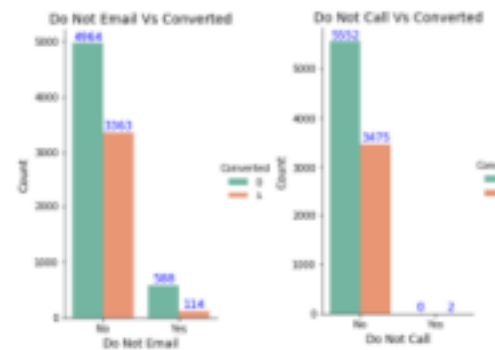
We have around 39% Conversion rate in Total



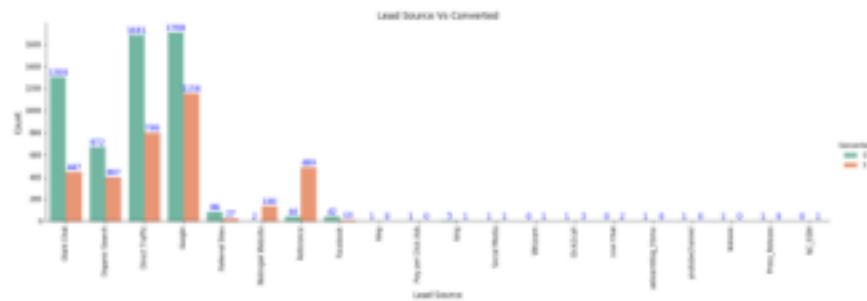
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit



Major conversion has happened from Emails sent and Calls made



Major conversion in the lead source is from Google



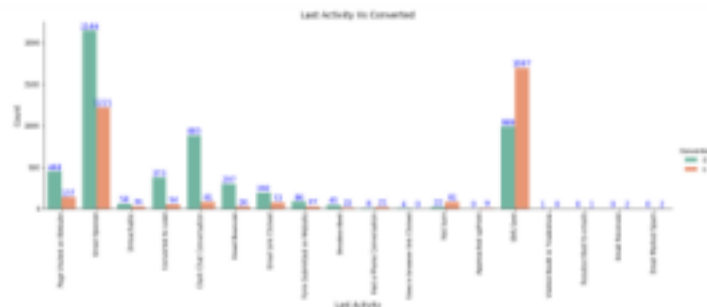
Not much impact on conversion rates through Search, digital advertisements and through recommendations



More conversion happened with people who are unemployed



Last Activity value of SMS Sent' had more conversion.

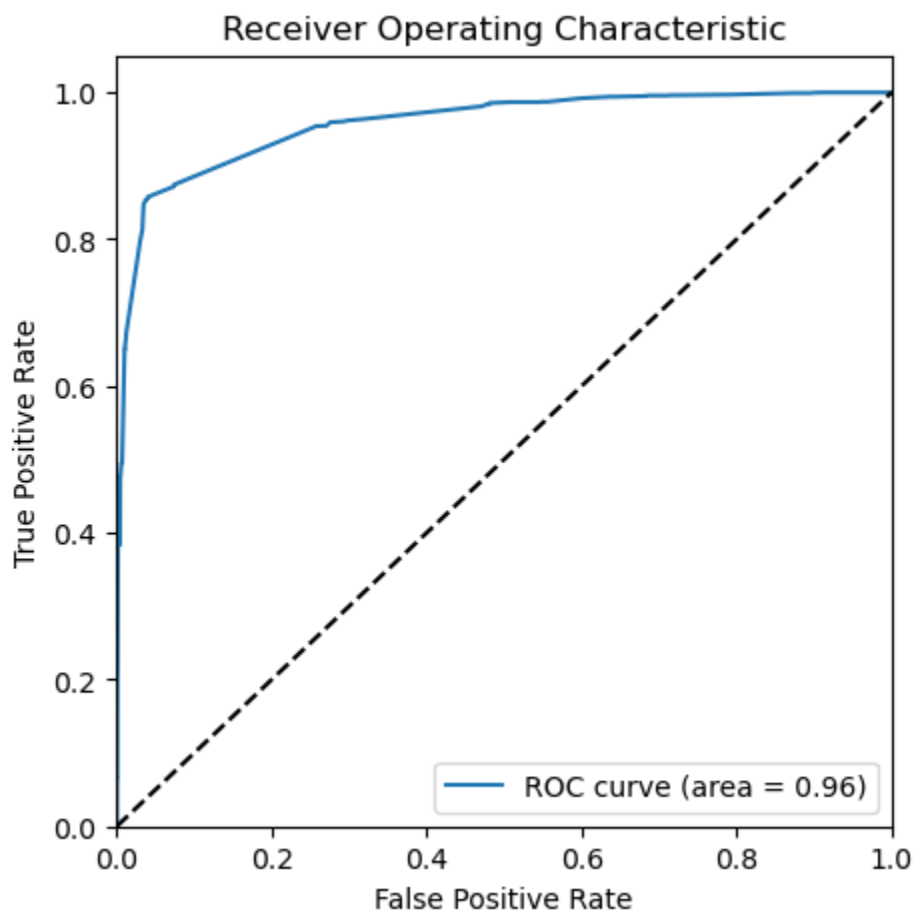


## Variables Impacting the Conversion Rate

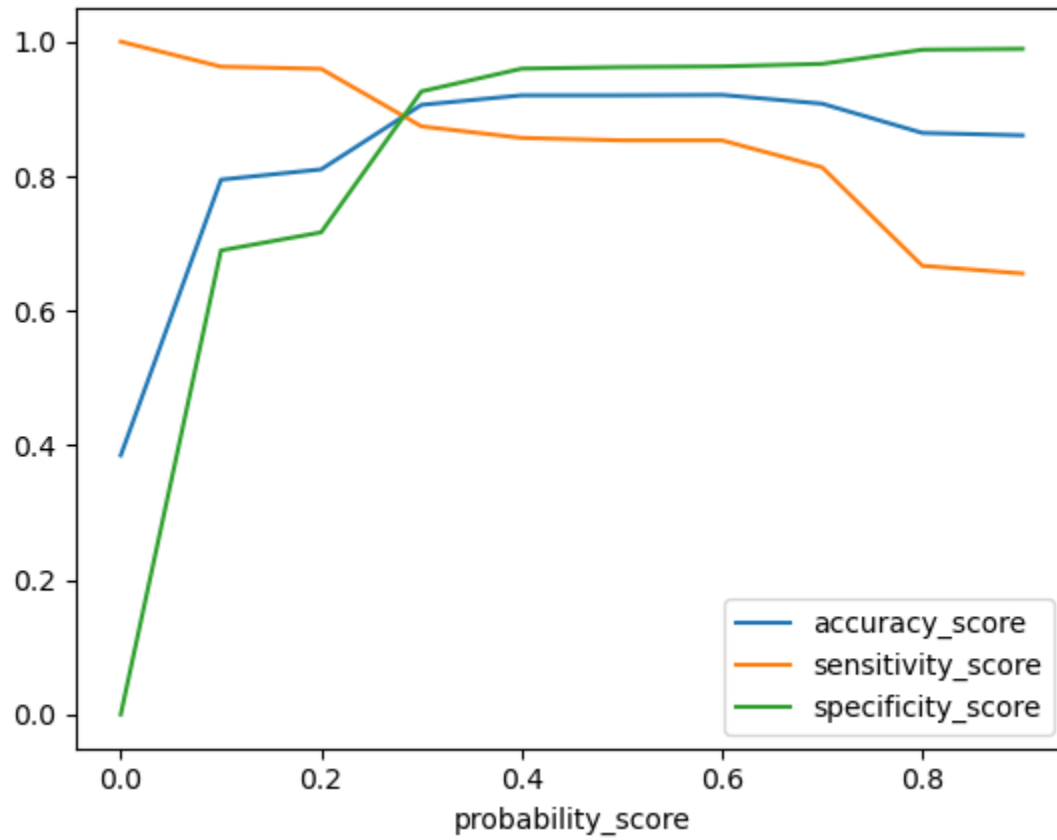
- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form

- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professiona
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

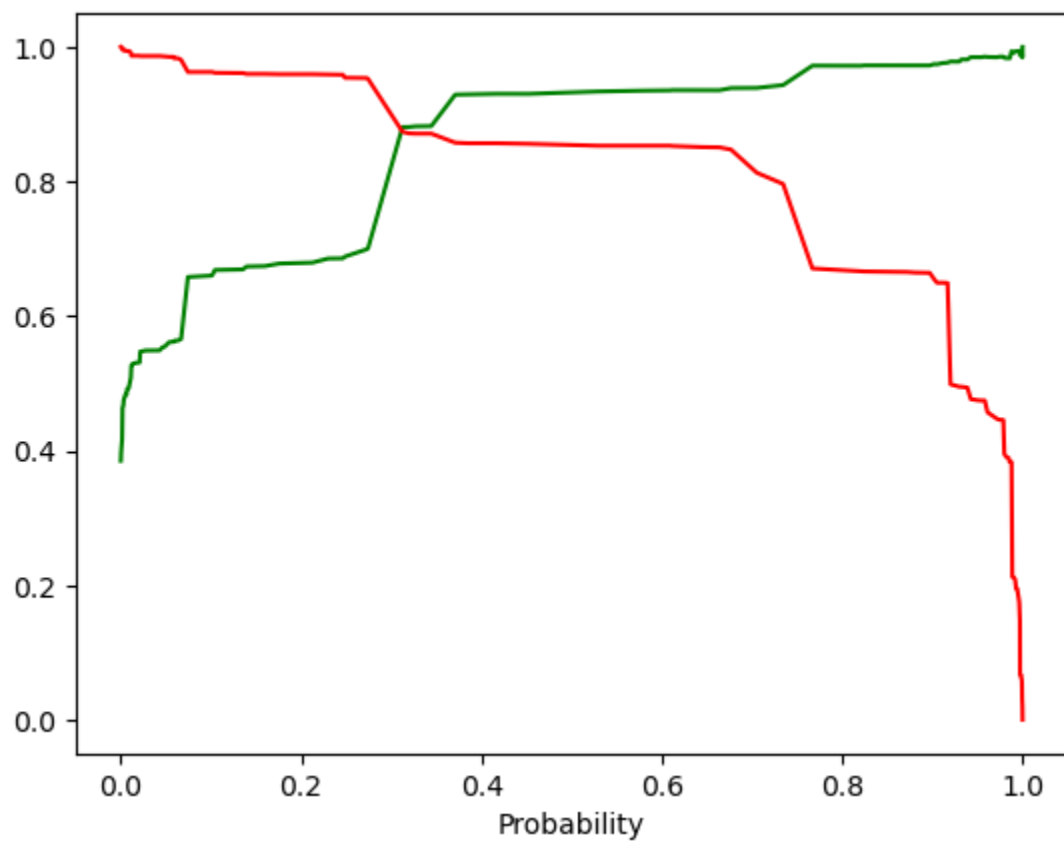
### ROC Curve



## Optimal Cutoff Point



Precision-Recall Trade off



## Conclusion Final Model:

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.

Training Data Results					
Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score	
0	3009	0	0.274201	1	27
1	1012	0	0.319386	1	32
2	9226	0	0.002149	0	0
3	4750	1	0.734474	1	73
4	7987	1	0.994050	1	99

Test Data Results					
Prospect ID	Converted	Convert_Probability	Convert_predicted	Lead_Score	
0	3271	0	0.274201	1	27
1	1490	1	0.978665	1	98
2	7936	0	0.274201	1	27
3	4216	1	0.996277	1	100
4	3830	0	0.274201	1	27

### -----Model Evaluation Metrics-----

Confusion Matrix:

[[1260 474]

[ 67 922]]




Accuracy: 0.8013220712449505

Sensitivity: 0.9322548028311426

Specificity: 0.726643598615917

Precision: 0.660458452722063

- The top 3 variables that contribute for lead getting converted in the model are

-  Total time spent on website
-  Lead Add Form from Lead Origin
-  Had a Phone Conversation from Last Notable Activity

- Hence overall this model seems to be good.

**Submitted by:**

**Venkatagirish Pamidigantam**

**Khushboo Lokhande**

**Aqsa Khan**