



# Lead Scoring Case Study

Kapeesh Kaul | Santu Mukherjee | M SivaRamaKrishna

# Problem Statement

- Build a model to 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
- The model is to be built with a lead conversion rate of around 80% or more.

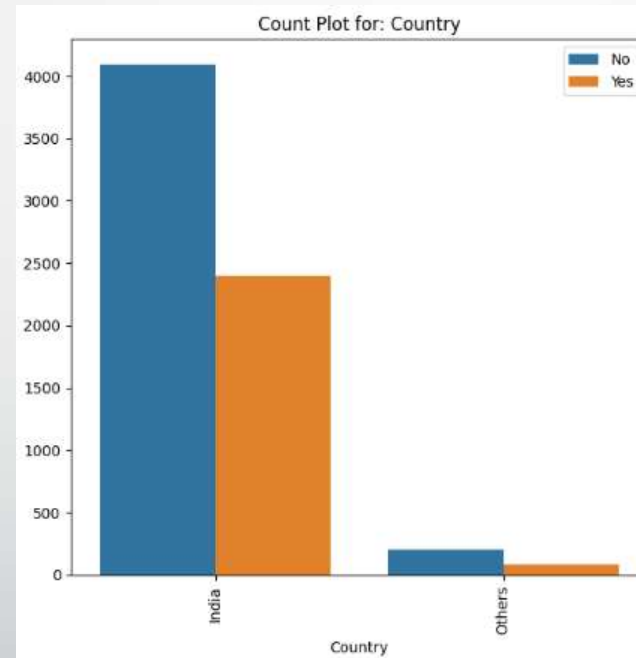
# Strategy

- Data Cleaning and EDA
- Splitting and Scaling the data
- Feature Selection
- Model Building and Evaluation Metrics
- Testing and Performance Evaluation
- Assigning the Score Variable

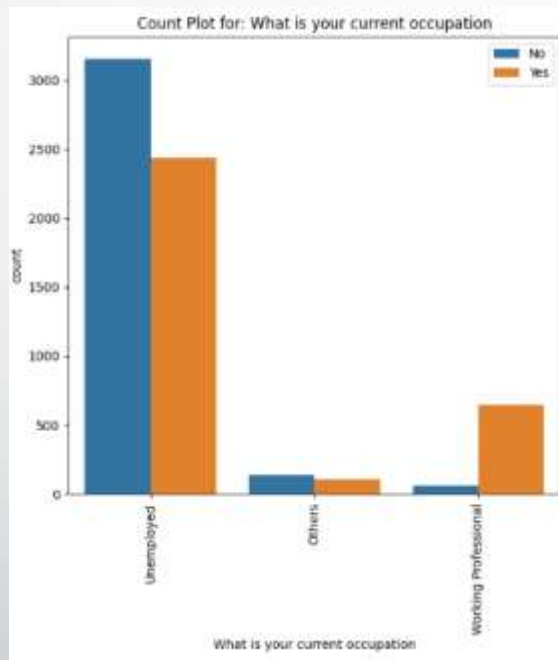


Insights From EDA

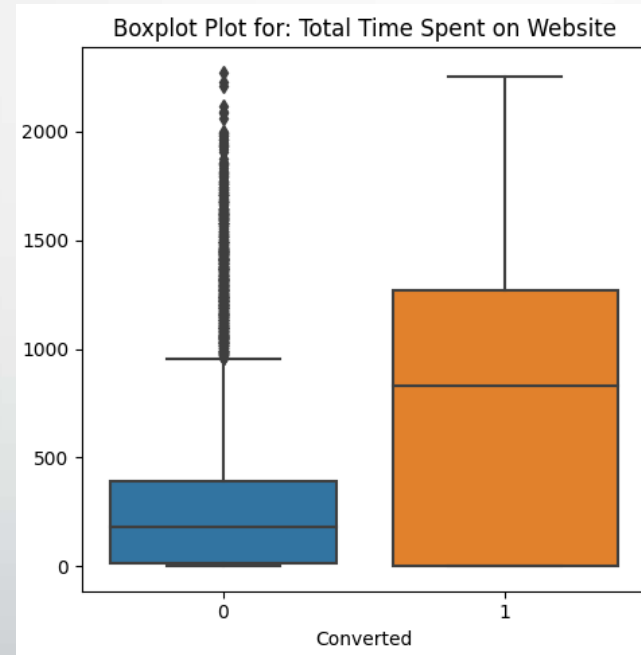
Most of the leads are from India



Working professionals have a high ratio of converted leads



A large share of leads who do not convert spend less time of the website





# Building the Model



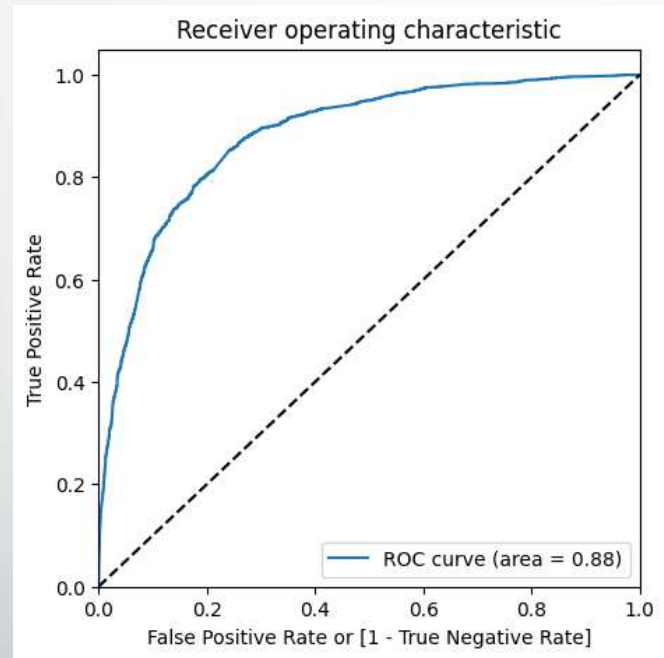
# Model Parameters

	coef	std err	z	P> z	[0.025	0.975]	Features	VIF
const	-2.3903	0.124	-19.238	0.000	-2.634	-2.147	2 Page Views Per Visit	3.43
TotalVisits	5.9395	2.207	2.691	0.007	1.613	10.266	3 Lead Origin_Landing Page Submission	3.03
Total Time Spent on Website	4.3585	0.163	26.785	0.000	4.040	4.677	11 What is your current occupation_Unemployed	2.67
Page Views Per Visit	-4.7812	1.273	-3.756	0.000	-7.276	-2.287	1 Total Time Spent on Website	2.05
Lead Origin_Landing Page Submission	-0.4482	0.090	-4.979	0.000	-0.625	-0.272	0 TotalVisits	1.96
Lead Origin_Others	-2.0026	0.504	-3.974	0.000	-2.990	-1.015	5 Lead Source_Olark Chat	1.58
Lead Source_Olark Chat	0.7193	0.127	5.662	0.000	0.470	0.968	14 Last Notable Activity_SMS Sent	1.56
Lead Source_Others	1.5037	0.176	8.533	0.000	1.158	1.849	13 Last Notable Activity_Others	1.53
Lead Source_Reference	3.0139	0.216	13.935	0.000	2.590	3.438	10 Last Activity_Page Visited on Website	1.47
Last Activity_Olark Chat Conversation	-1.7399	0.174	-10.006	0.000	-2.081	-1.399	8 Last Activity_Olark Chat Conversation	1.45
Last Activity_Others	-1.2331	0.128	-9.618	0.000	-1.484	-0.982	12 What is your current occupation_Working Profes...	1.39
Last Activity_Page Visited on Website	-1.2177	0.174	-6.987	0.000	-1.559	-0.876	9 Last Activity_Others	1.38
What is your current occupation_Unemployed	1.0014	0.081	12.311	0.000	0.842	1.161	7 Lead Source_Reference	1.26
What is your current occupation_Working Professional	3.5504	0.197	17.987	0.000	3.164	3.937	6 Lead Source_Others	1.21
Last Notable Activity_Others	0.8239	0.156	5.273	0.000	0.518	1.130	4 Lead Origin_Others	1.15
Last Notable Activity_SMS Sent	1.2402	0.081	15.329	0.000	1.082	1.399		

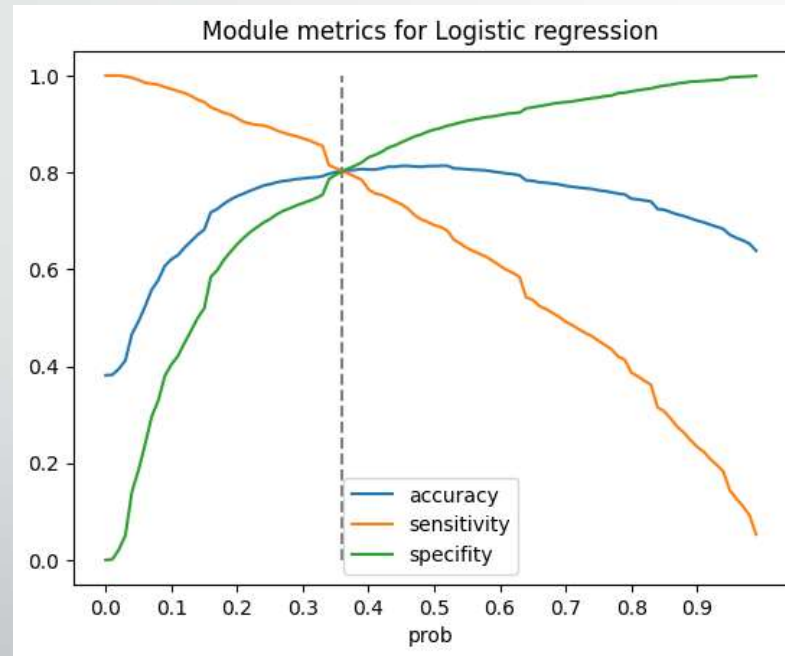
# Key Insights from the Model

- The higher the number of total visits made by the user, the higher the chance that the lead gets converted to a paying customer.
- The more the number of pages a lead visits in a single visit, the lesser the chance of them converting.
- The higher the time the lead spends on the website, the higher the chance that they convert to a paying customer

# ROC Curve



# Finding the optimal Cutoff by plotting the sensitivity-specificity graph



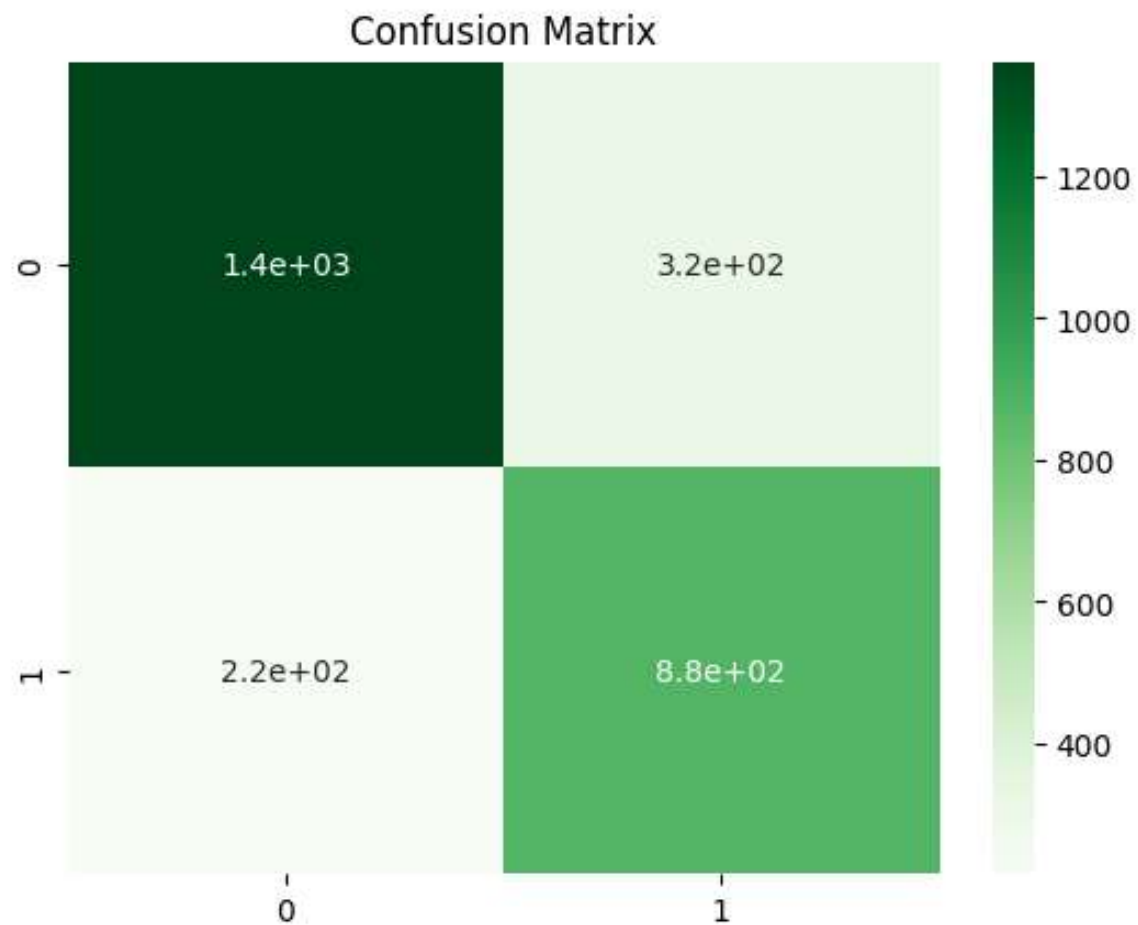
The optimal cutoff is 0.36

## Training Performance metrics

Metric	Value
accuracy	0.802566
sensitivity	0.802514
specificity	0.802599
F1 Score	0.756064



Evaluating the model using the test set



Confusion  
Matrix

# Performance Metrics on the Test Set

Metric	Value
Accuracy	0.8084415584415584
Sensitivity	0.8027397260273973
Specificity	0.8121645796064401
Precision	0.7361809045226131
Recall	0.8027397260273973
F1 Score	0.7680209698558322



# Assigning the Score Variable

```
In [53]: data_test['score'] = score  
         data_test.score.head(10)
```

```
Out[53]: 4269    57.034199  
         2376    94.610093  
         7766    63.364053  
         9199     8.244914  
         4359    83.548922  
         9186    54.921772  
         1631    44.765284  
         8963    28.207767  
         8007     2.655939  
         5324    38.230438  
         Name: score, dtype: float64
```