



# LEAD SCORE CASE STUDY

By Nikhil Kapoor

# ASSIGNMENT DETAILS

## Problem Statement:

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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 Goal:

X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%

# SOLUTION METHODOLOGY

## ➤ Data cleaning and data manipulation.

- Check and handle duplicate data.
- Check and handle NA values and missing values.
- Drop columns, if it contains large amount of missing values and not useful for the analysis.
- Imputation of the values, if necessary.
- Check and handle outliers in data.

## ➤ Exploratory Data Analysis

- Univariate data analysis: value count, distribution of variable etc.
- Multivariate data analysis: correlation coefficients and pattern between the variables etc.

## ➤ Feature Scaling & Dummy Variables and encoding of the data.

## ➤ Classification technique: logistic regression used for the model making and prediction.

## ➤ Validation of the model.

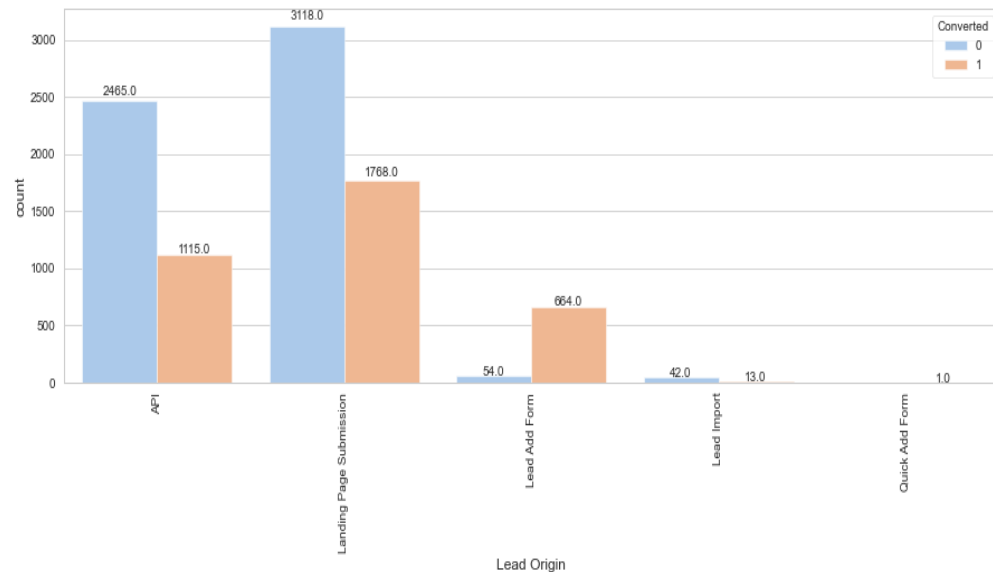
## ➤ Model presentation.

## ➤ Conclusions and recommendations.

# EXPLANATORY DATA ANALYSIS

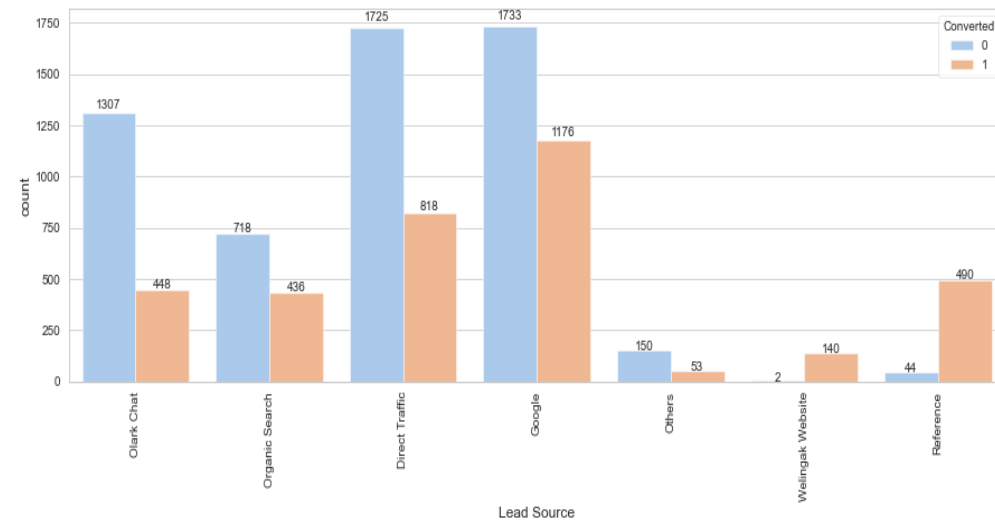
## Lead Origin :

Customers identified via Lead Add Form have a high conversion of 664.

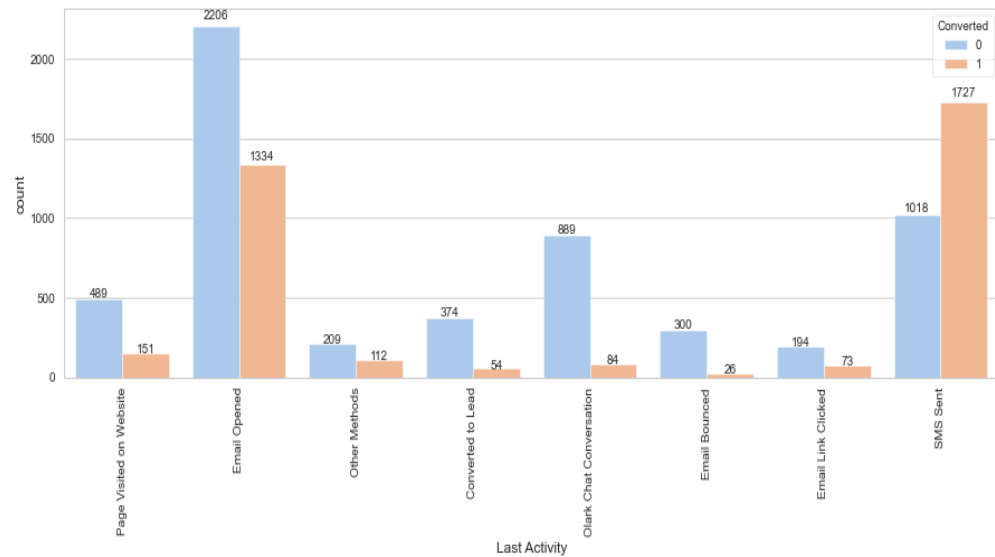


## Lead Source:

1. Leads via Google & Direct Traffic convert less as compared to others.
2. Leads coming via Reference have the highest conversion rate.

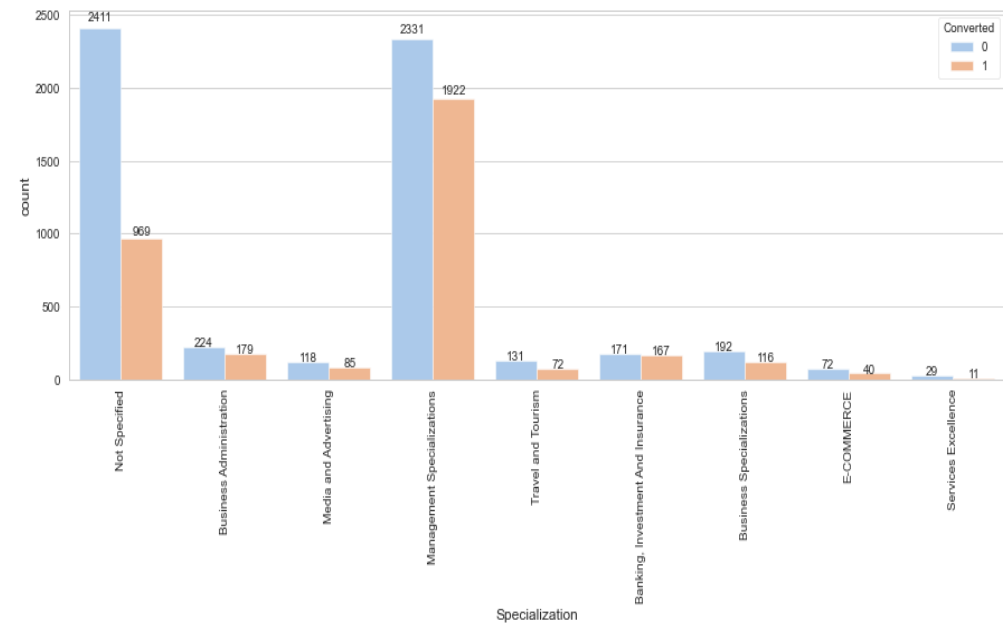


**Last Activity:** Customers with last activity as SMS sent have a conversion of 1727.



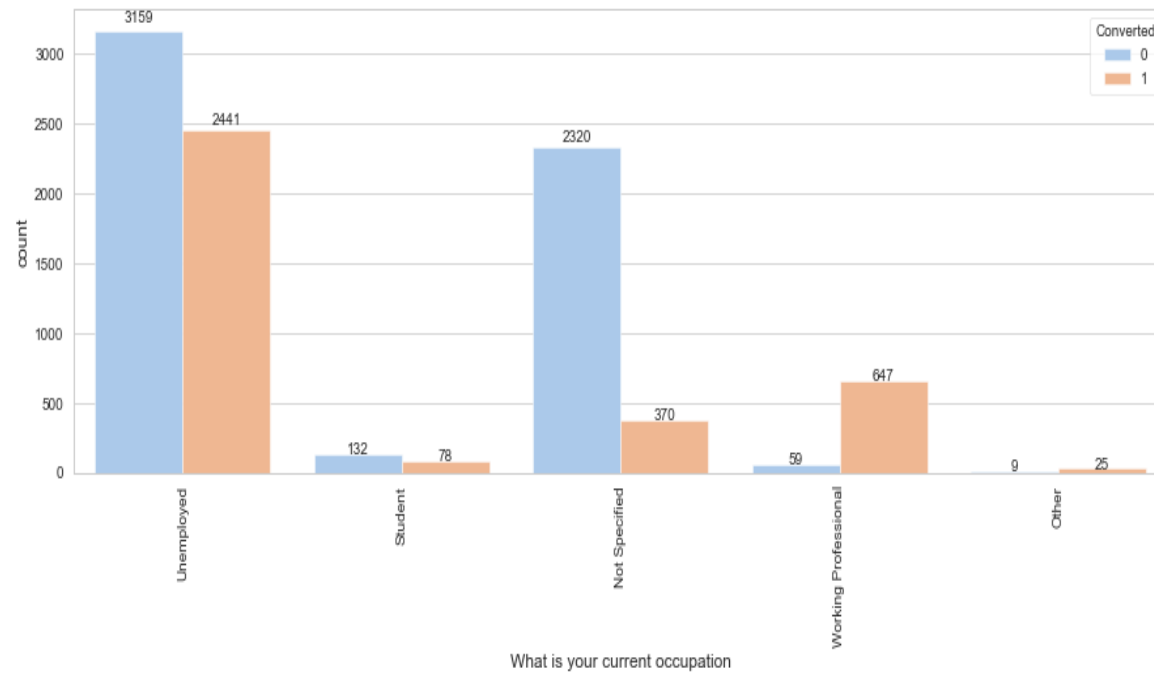
### Specilization:

1. Management Specialization have most conversion of 1922.
2. Services Excellence has least conversion 29.



### What is your current occupation:

1. Out of 5600 unemployed only 2441 converted.
2. While out of 706 Working Professionals, 647 converted.



# MULTIVARIATE ANALYSIS

Suggests that the 'TotalVisits' is fairly correlated to 'Page Views Per Visit'.

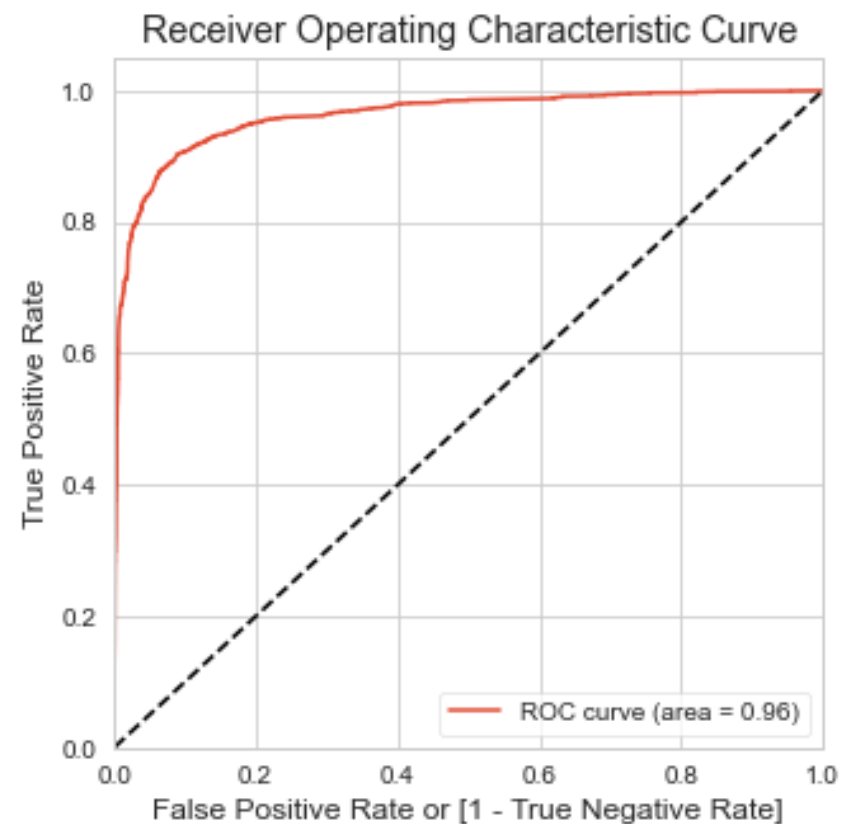


# VARIABLE IMPACTING MOST

- 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 Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity – Unreachable

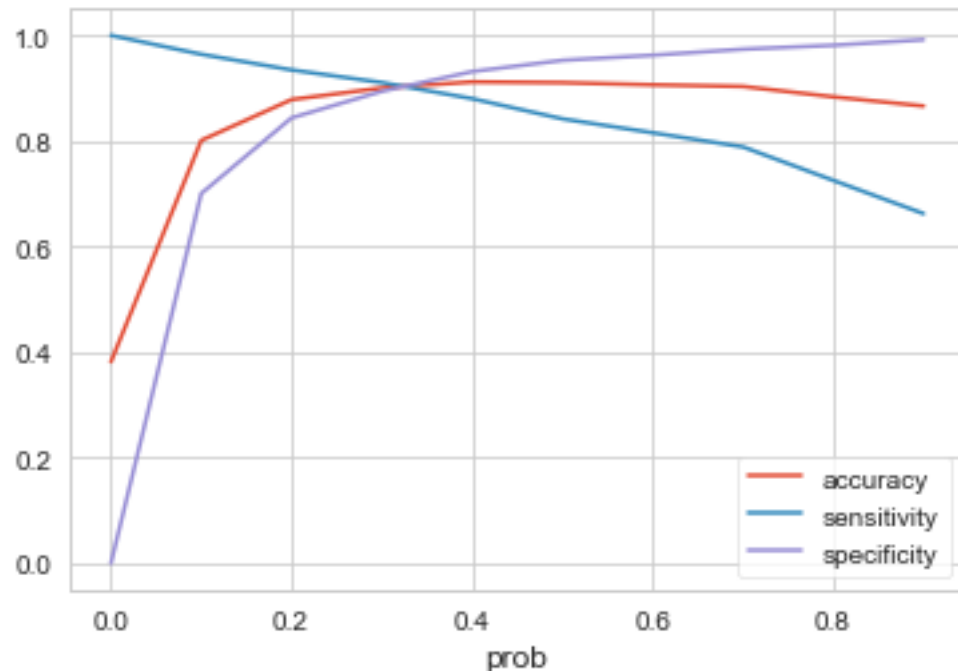


# ROC CURVE



# MODEL EVALUATION – TRAIN DATASET

There is an optimal cut-off at 0.37



Confusion Matrix

# Actual/Predicted	Not_Converted	Converted
# Not_Converted	3518	417
# Converted	217	2211

Accuracy – 90%

Sensitivity – 91%

Specifity – 89%

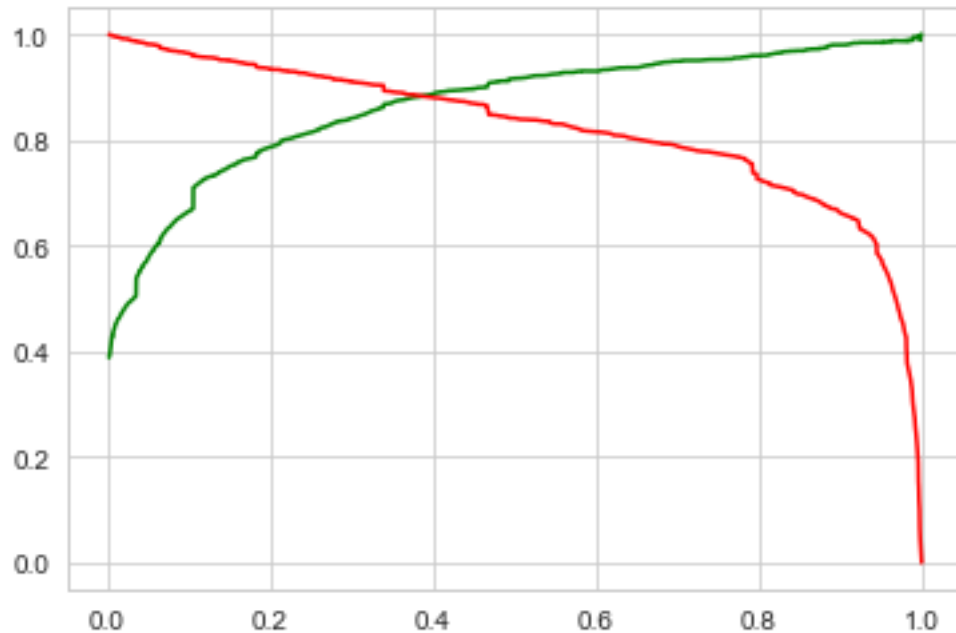
False Positive Rate – 11%

Positive Predictive Value – 84%

Negative Predictive Value – 94%

# MODEL EVALUATION — TRAIN DATASET (PRECISION & RECALL)

Optimal cutoff at 0.37 based on  
Precision & Recall.



➤ Precision — 84

➤ Recall — 91

# MODEL EVALUATION — TEST DATASET

## Confusion Matrix

# Actual/Predicted	Not_Converted	Converted
# Not_Converted	1512	148
# Converted	89	978

Accuracy — 91%

Sensitivity — 92%

Specificity — 91%

# CONCLUSION

- People who spend more time on the website or with more no. of visits convert more.
- When the last activity was **SMS** or **Olark chat conversation** the conversion was more.
- Working professionals have a high chance of conversion.
- While Reference are the best form of leads there are significant conversions from Google, Direct Traffic, Organic search & Welingak website.

**X Education are not converting enough of the Unemployed leads. They can probably come up with a strategy to tap on this segment.**