

# Lead Scoring Case Study Business Overview

Prepared By –  
Aditya Kumar Dubey

# Contents

- Problem Statement
- Goals & Objectives
- Exploratory Data Analysis
- Model Validation
- Conclusion

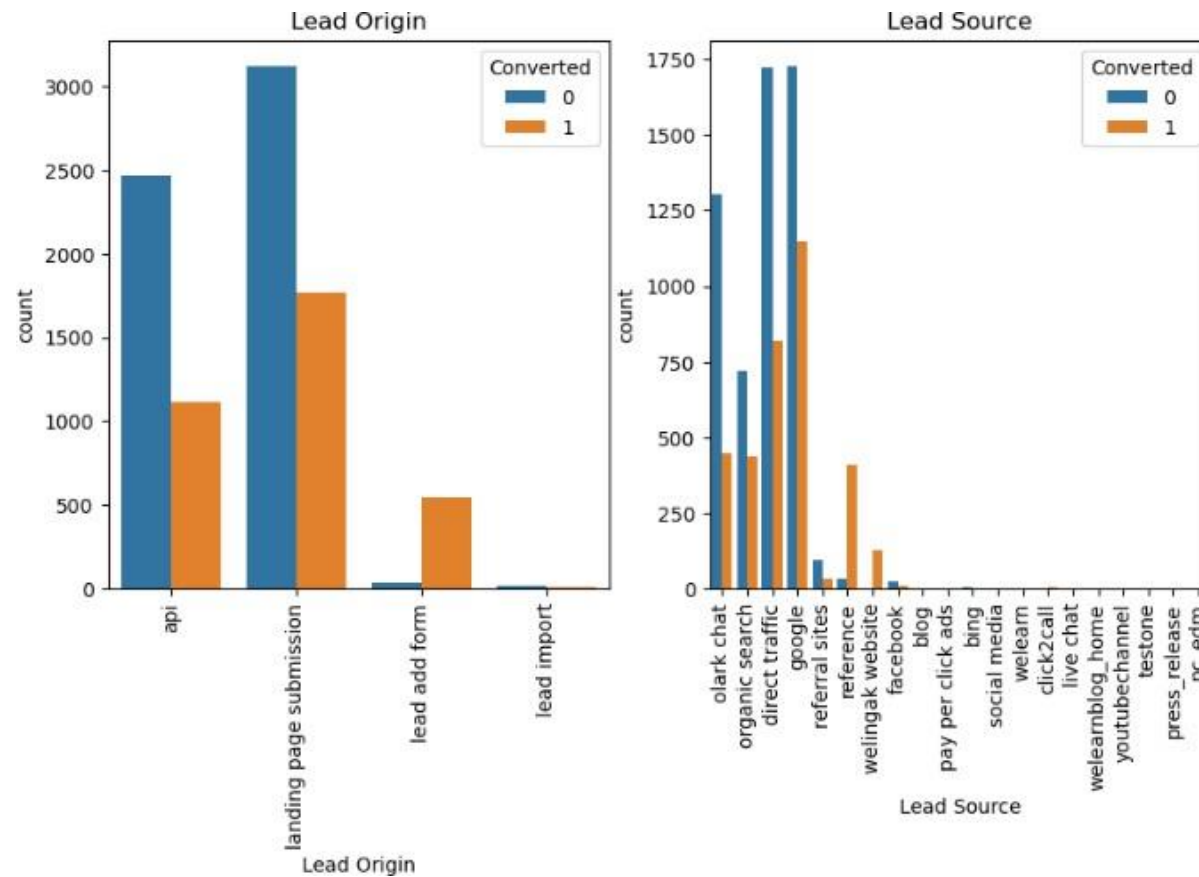
# Problem Statement

- X Education needs help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- The company requires a model where if a customer has a lead score greater than or equal to 80%, he/she will have a higher conversion chance and the customers with a lead score less than 80%, he/she will have a lower conversion chance.

# Goals & Objectives

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by X Education to target potential leads.
- A higher score (equal to or above 80) would mean that the lead is hot, i.e. is most likely to convert whereas a score lower than 80 would mean that the lead is cold and will mostly not get converted.
- There are some more problems presented by X Education, which the model should be able to adjust to if the company's requirement changes in the future.

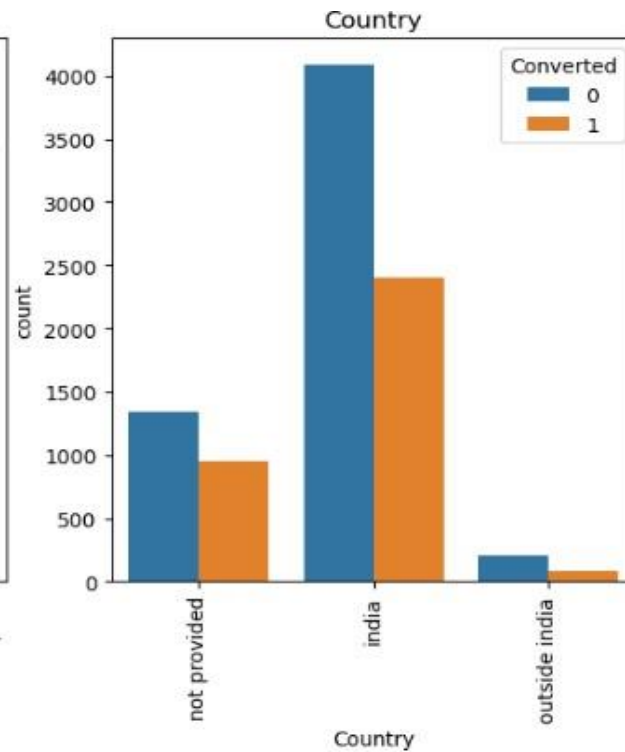
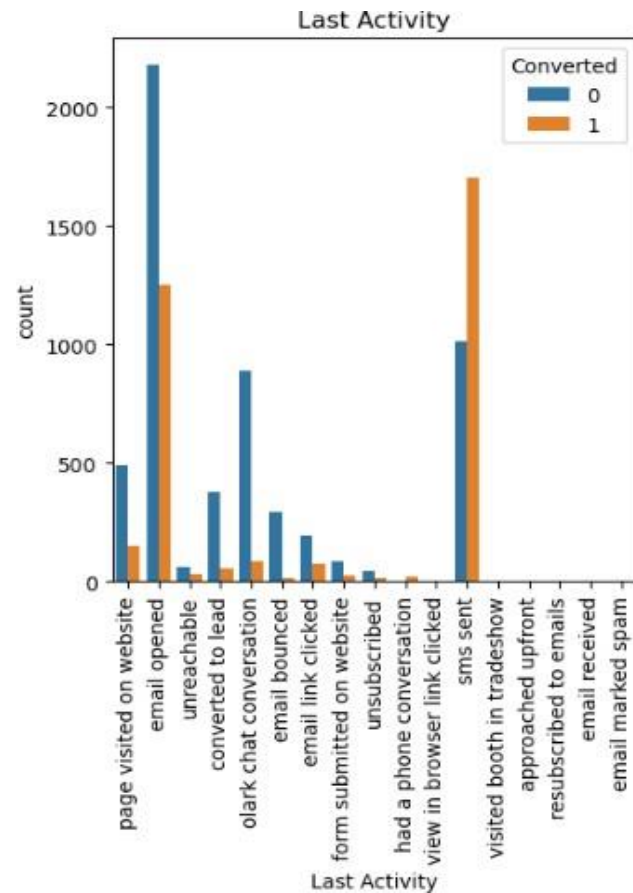
# Exploratory Data Analysis



From the graphs it is quite evident that –

- Most leads originated are from Landing Page Submission and API.
- Highest Lead Source is obtained from Google Search, followed by Direct Website Traffic & Olark Chat.

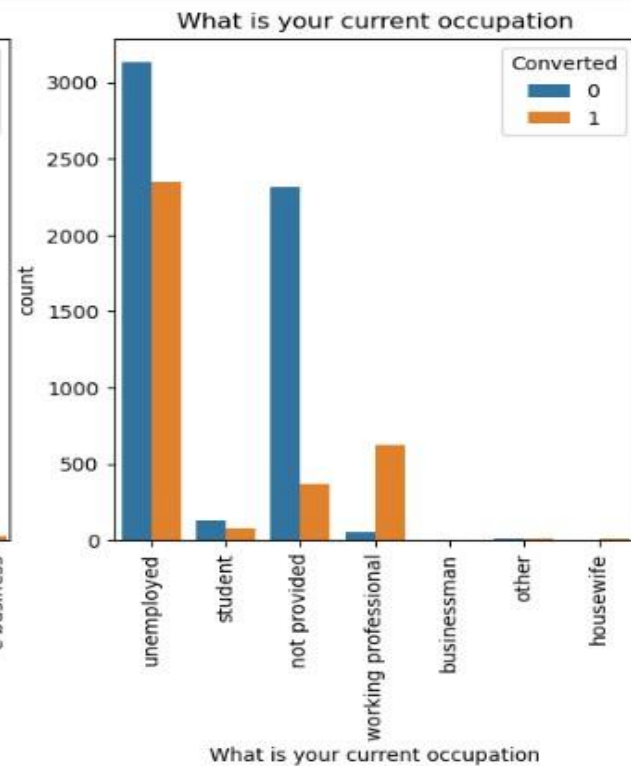
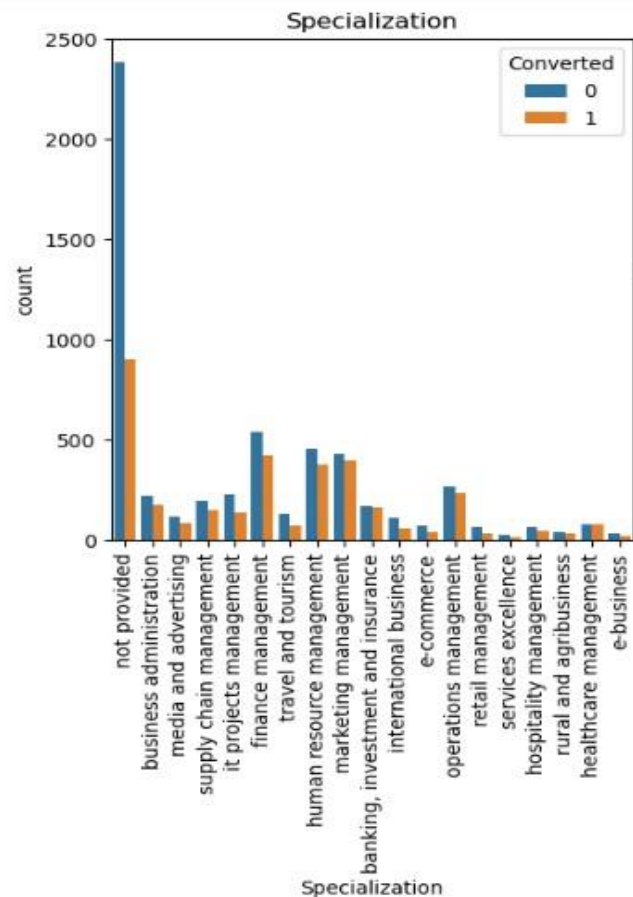
# Exploratory Data Analysis



From the graphs it is quite evident that –

- Most conversions are established by either text messages or emails.
- Most conversions are made for Learners from India.

# Exploratory Data Analysis

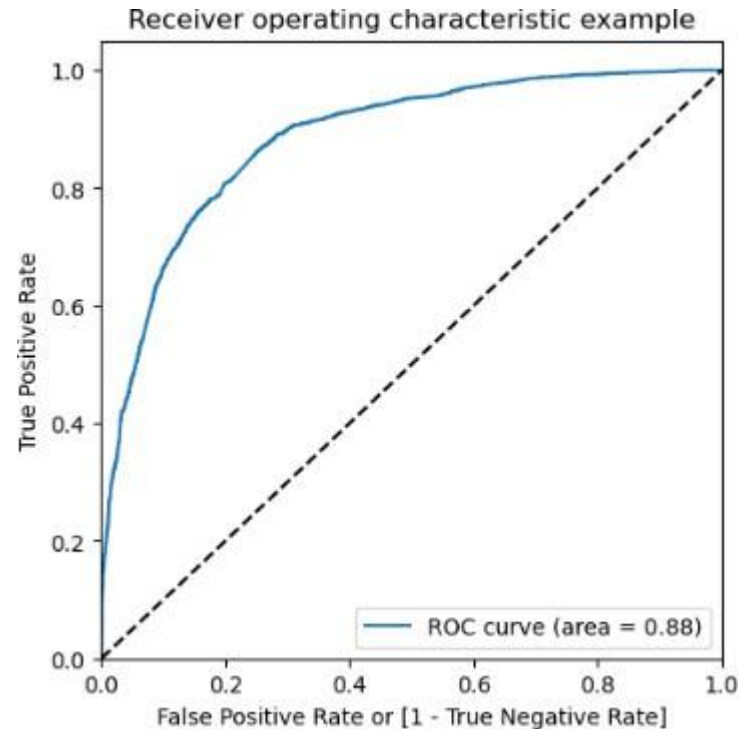


From the graphs it is quite evident that –

- Most Specializations chosen by learners are Finance Management, HR Management & Marketing Management.
- Most conversions are made for Learners who are currently Unemployed.

# Model Validation

- ROC Curve obtained with area = 0.88 (Ideal = 1).
- Values obtained for Accuracy, Sensitivity & Specificity should be within the range of 5%.



	Train Dataset	Test Dataset
Accuracy	80.32%	80.79%
Sensitivity	80.37%	81.31%
Specificity	80.28%	80.5%
Precision	75.4%	73.24%
Recall	75.9%	76.61%



# Conclusion

- Variables that matter the most in finding potential conversions are –
  1. Total Visits
  2. Total time spent on website
  3. Lead Origin – Lead Add Form
  4. Lead Sources – Olark Chat, Welingak Website
  5. Do Not Email – Yes
  6. What is your current occupation? – Other, Student, Unemployed & Working Professional
- Keeping this in mind, X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mindset and buy their courses.