



# Lead Scoring Case Study

**Submitted by**

Tanvi Acharya

Kunwar Singh Sohal

Praveen M



# Lead Score Case Study for X Education

## 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 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%.



# Strategy Implemented

- 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



**Data Sourcing , Cleaning and  
Preparation**

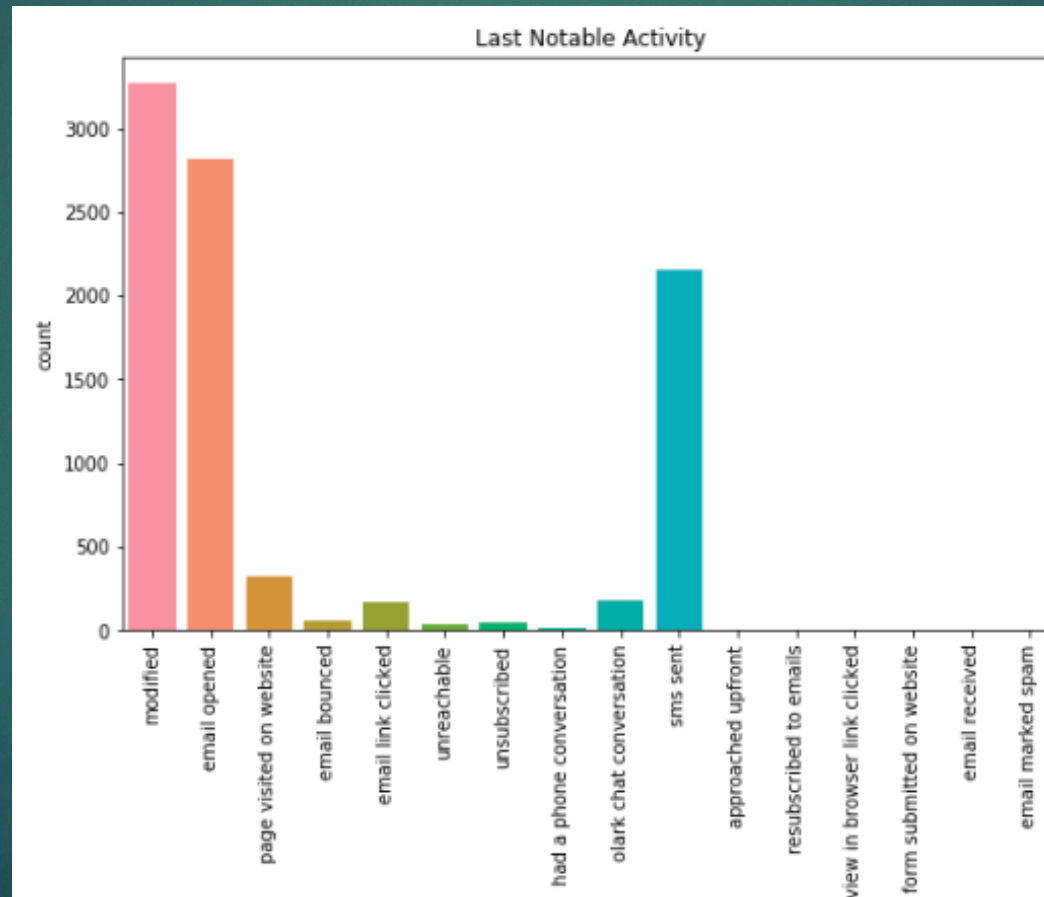


# Data Manipulation

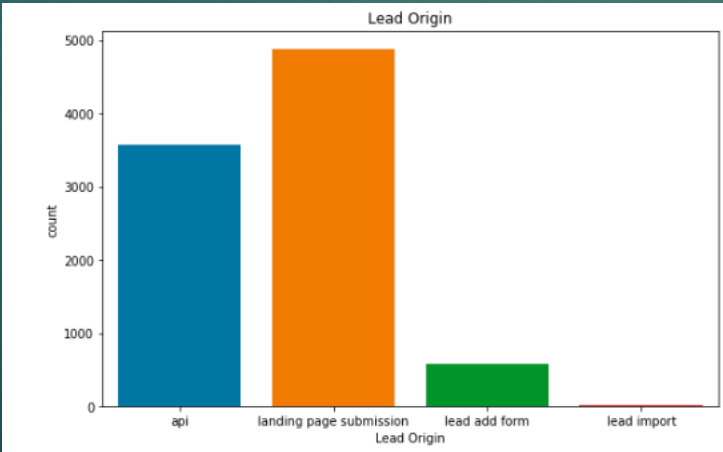
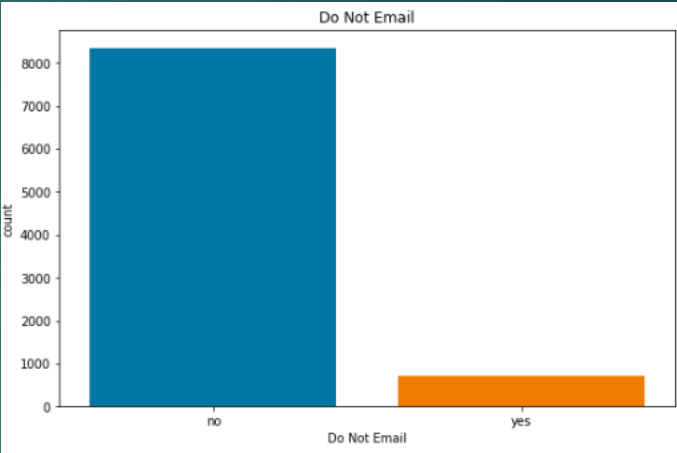
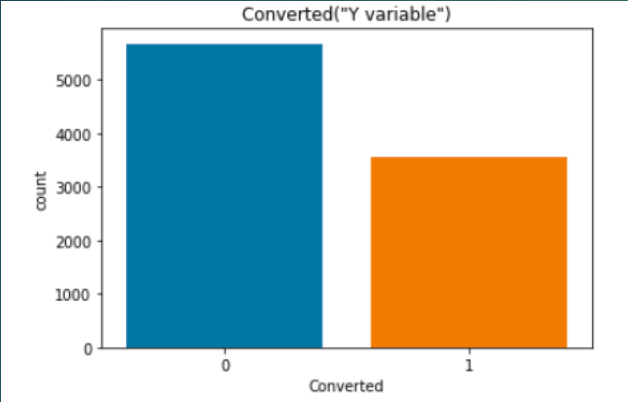
- Total Number of Rows =37, Total Number of Columns =9240.
- Single value features like “Magazine”, “Receive More Updates About Our Courses”, “Update me on Supply”
- Chain Content”, “Get updates on DM Content”, “I agree to pay the amount through cheque” etc. have been dropped.
- Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis.
- After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: “Do Not Call”, “What matters most to you in choosing course”, “Search”, “Newspaper Article”, “X Education Forums”, “Newspaper”, “Digital Advertisement” etc.
- Dropping the columns having more than 35% as missing value such as ‘How did you hear about X Education’ and ‘Lead Profile’.



# Exploratory Data Analysis

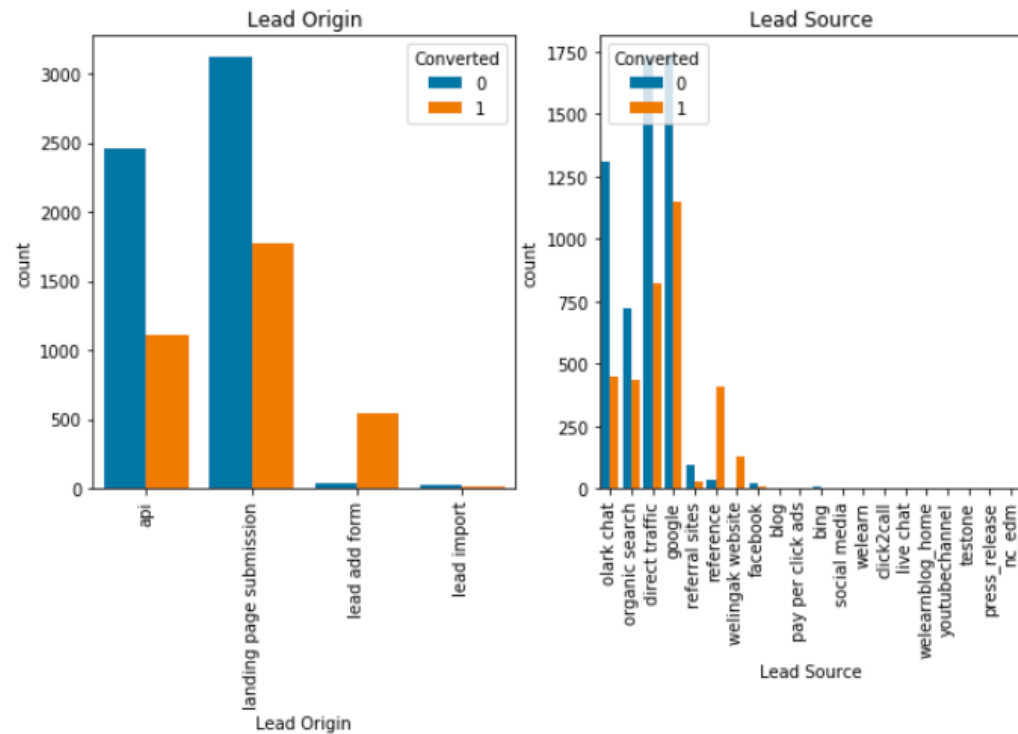


In Lead Origin, maximum conversion happened from Landing Page Submission

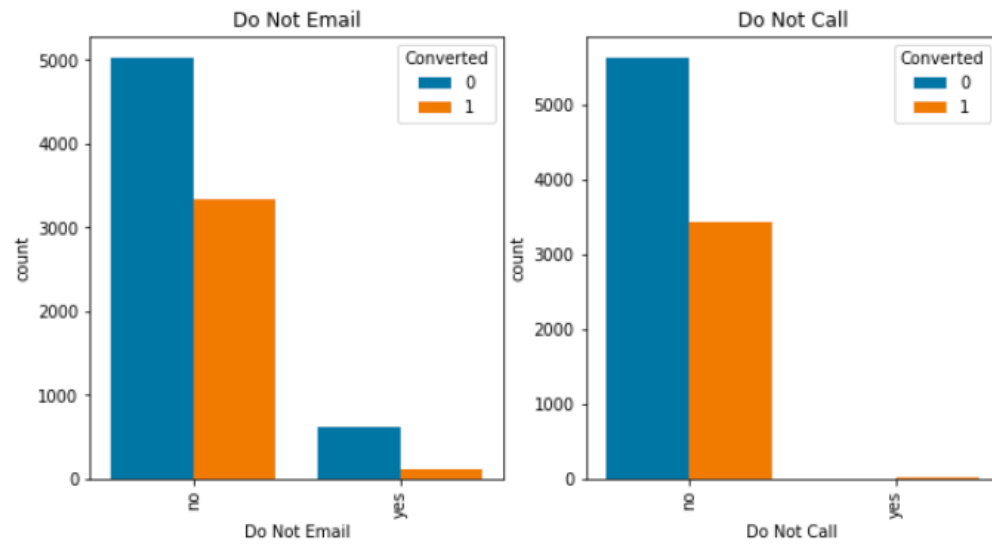




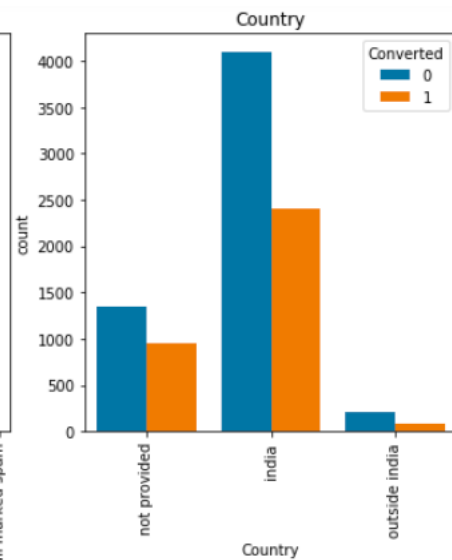
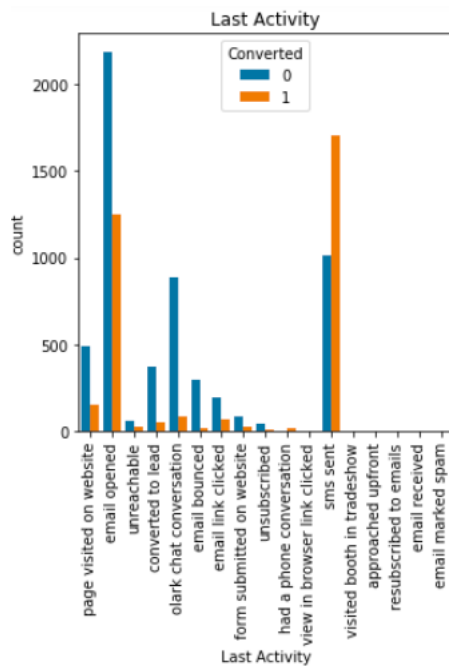
# Categorical Variable Relation







# Categorical Variable Relation



# Categorical Variable Relation



# Data Conversion

- Numerical Variables are Normalised
- Dummy Variables are created for object type variables
- Total Rows for Analysis: 8792
- Total Columns for Analysis: 43

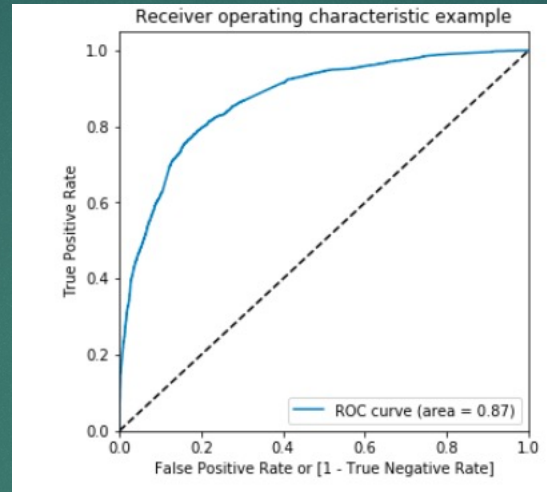
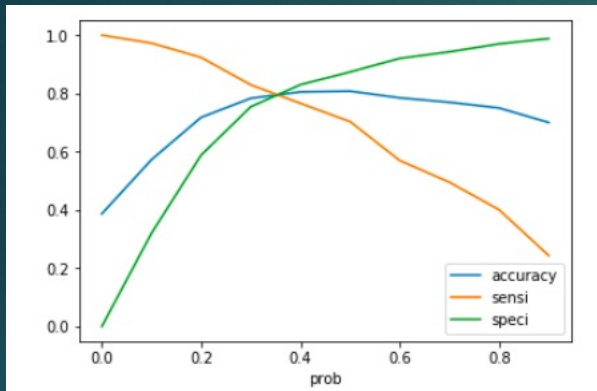


# Model Building

- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p-value is greater than 0.05 and vifvalue is greater than 5
- Overall accuracy 81%



# ROC Curve



- Finding Optimal Cut off Point
- Optimal cut off probability is that
- probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.35.



# Conclusion

- It was found that the variables that mattered the most in the potential buyers are (In descending order) :
  - ✓ The total time spend on the Website.
  - ✓ Total number of visits.
  - ✓ When the lead source was:
    - i. Google
    - ii. Direct traffic
    - iii. Organic search
    - iv. Website
  - ✓ When the last activity was:
    - i. SMS
    - ii. Olark Chat Conversation
  - ✓ When the lead origin is Lead add format.
  - ✓ When their current occupation is as a working professional. Keeping these in mind the X Education can flourish as they have a very high chance to get