Lead Score Case Study

Submitted by :

Deepak Gupta

Dhemath Dhanraj

Satish

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.

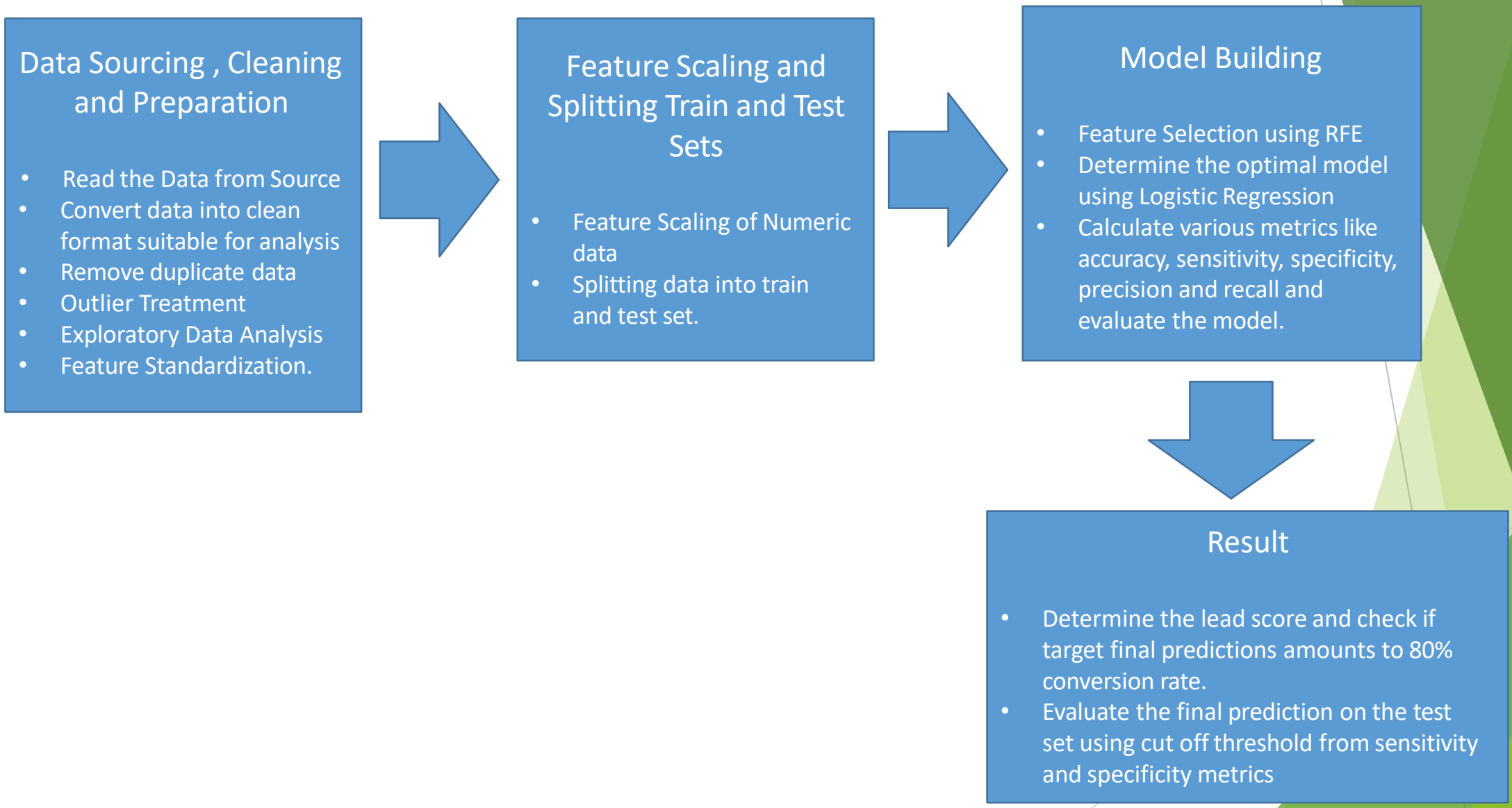
The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

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

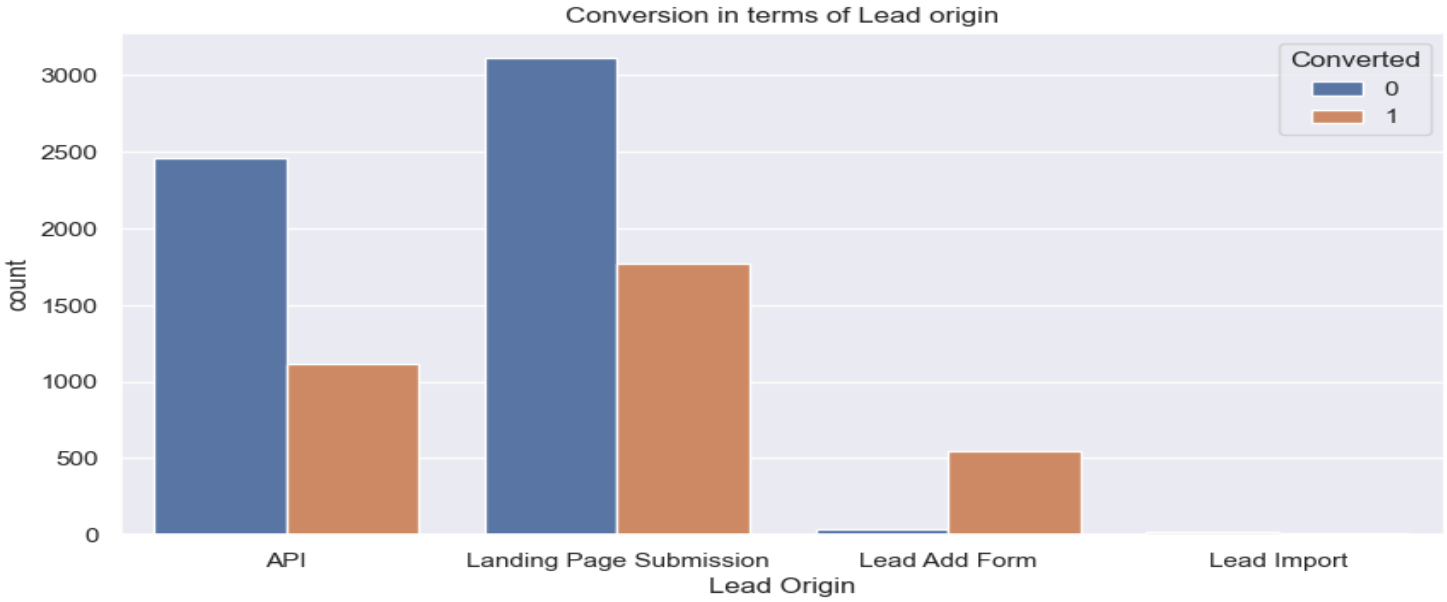
Strategy

- 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



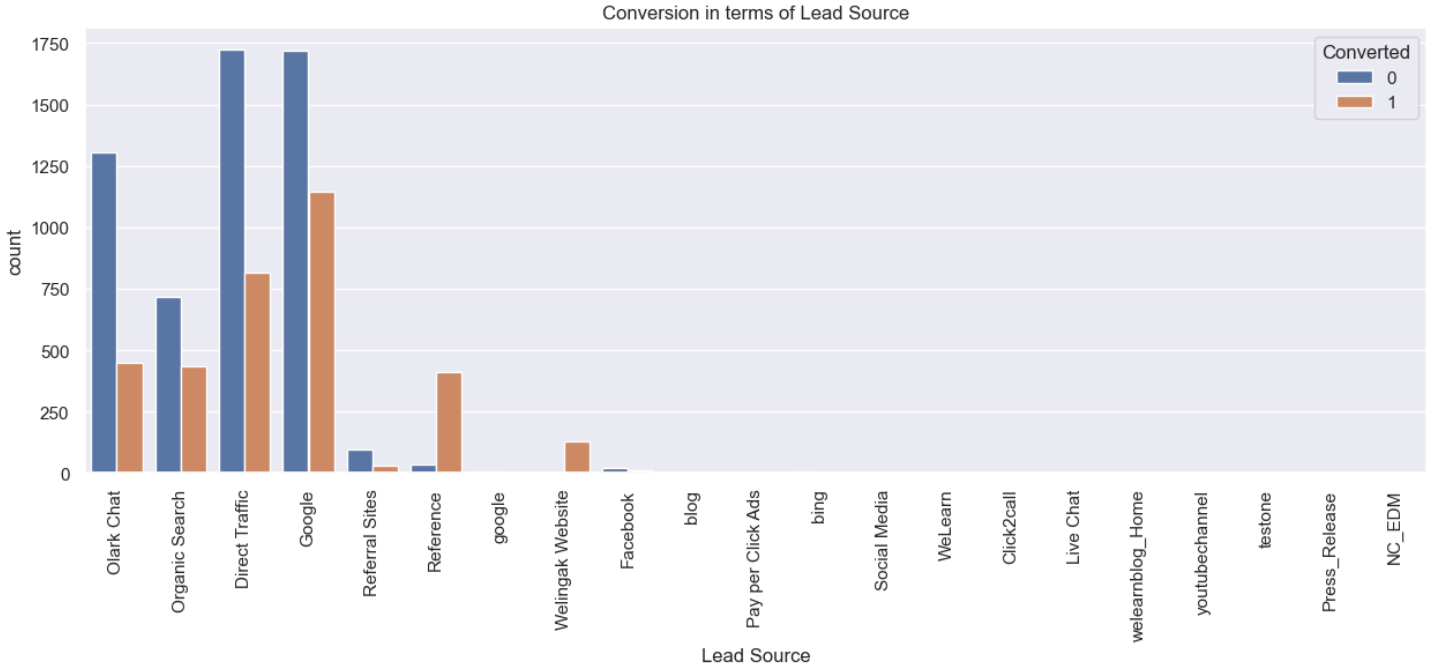
Exploratory Data Analysis



- From the above plot and Lead origin conversion summary, we can infer that:
- Lead Add Form has the highest conversion rate at 94%
- API and Landing Page Submission have 31% and 36% conversion rate but they generate maximum leads counts.
- Lead Import has the least amount of conversions and leads count.
- To improve overall lead conversion rate, focus should be on improving lead conversion rate of API and Landing Page Submission. Also,generate more leads from Lead Add form since they have a very good conversion rate

.

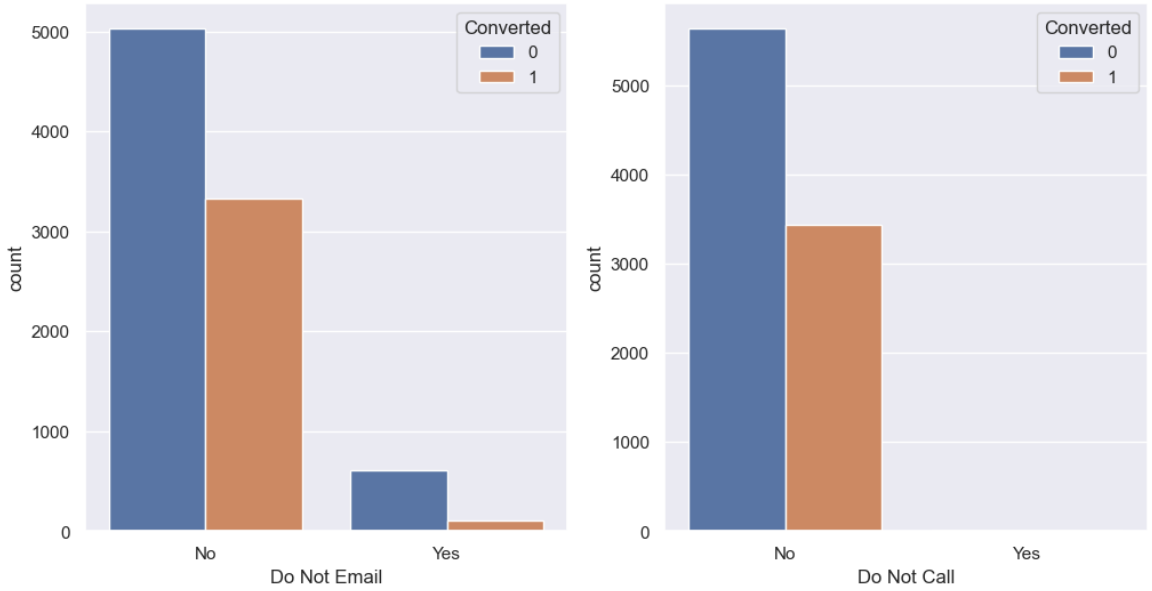
Exploratory Data Analysis



- Few Lead sources have very low count.Hence we can merge them into common category 'Others'
- Also , there are 2 categories with same name Google and google.Hence replace google with Google to have a single category

.

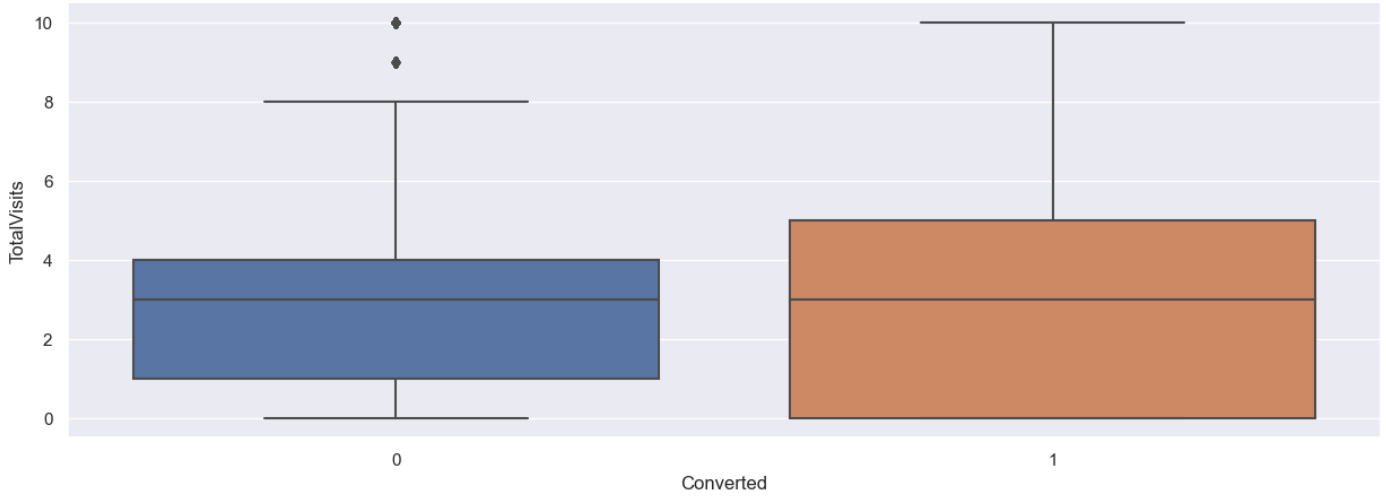
Exploratory Data Analysis



- Around 99% of customers do not like to be called or receive emails about the course.

.

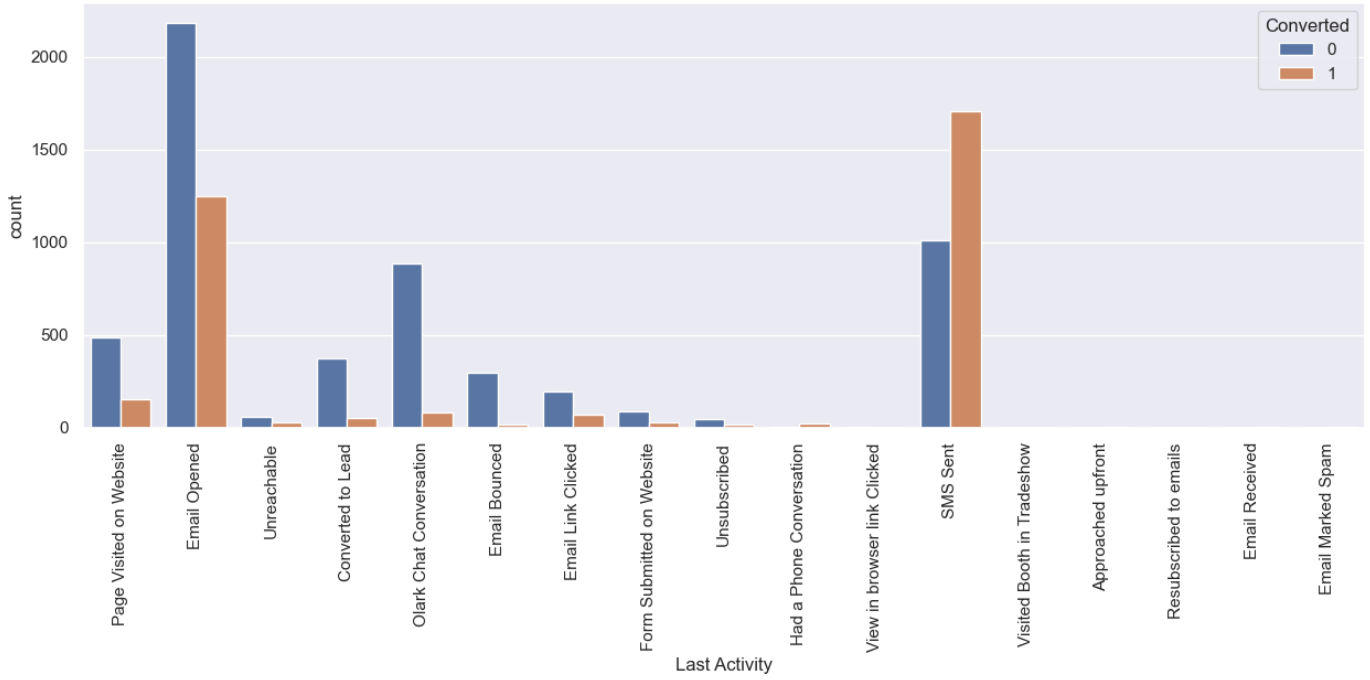
Exploratory Data Analysis



- Median for converted and non-converted leads are same.
- People who visits the platform have equal chances(50-50) of applying and not applying for the course.

.

Exploratory Data Analysis



- Maximum leads are generated from people with last activity - Email opened and SMS sent.
- Conversion rate is around 63% and 36% .
- Least leads are generated from people with last activity - Approached upfront,Email Marked Spam,Resubscribed to emails ,emails received,View in browser link Clicked,Visited Booth in Tradeshow
- olark chat conversation and Page Visited on Website generates significant number of leads but their conversion rate is around 9% and 24%.

Conclusion

- The sales team of the X-Education should focus on the leads having lead origin - lead add form , occupation - Working Professional , Lead source - Wellingak website.
- Hot Leads are identified as 'Customers having lead score above 35. Sales Team of the company should first focus on the 'Hot Leads'
- The 'Cold Leads'(Customer having lead score <= 35) should be focused after the Sales Team is done with the 'Hot Leads'.
- There are many important variables like city, specialization , occupation which can potentially explain Conversion better.It is important for the management to make few of these information mandatory to fill , so that we can use in our model and build important decisions for the business.
- We have high recall score than precision score. Hence this model has an ability to adjust with the company's requirements in coming future.
- High Sensitivity will ensure that almost all leads who are likely to Convert are correctly predicted where as high Specificity will ensure that leads that are on the brink of the probability of getting Converted or not are not selected.
- It's better to focus least on customers who do not want to be called about the course.
- If the Last Notable Activity is Modified, he/she may not be the potential lead.