

Lead Scoring Case Study

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

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. A typical lead conversion process can be represented using the following funnel:



Business Goal Continued...

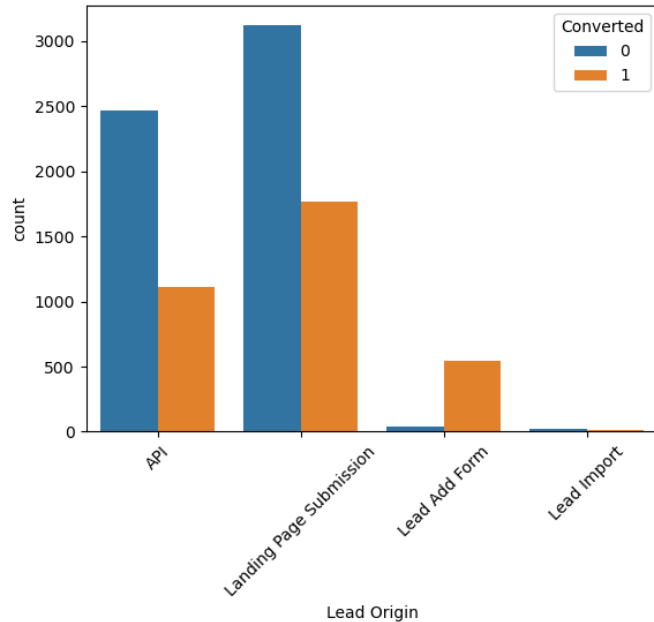
As you can see, there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need 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 CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

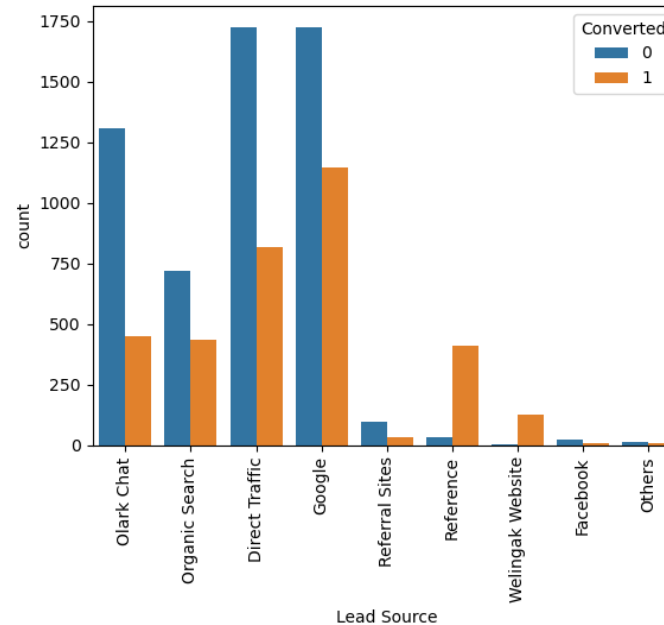
Strategy

- Reading the data
- Cleaning and preparing the data
- EDA
- Train Test Split
- Feature Scaling
- Building a Logistic Regression Model
- Generating a Lead Score
- Evaluating the model
- Choosing the optimal threshold

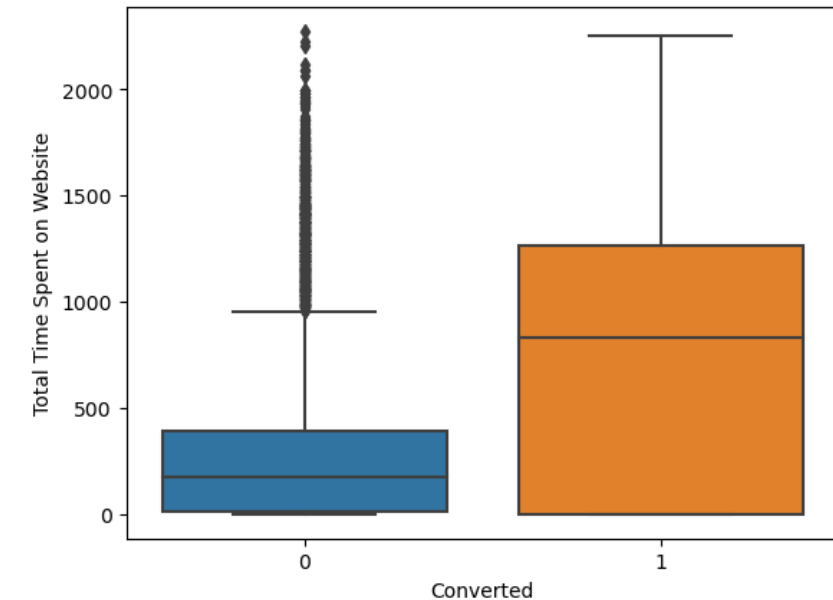
Exploratory Data Analysis



Business needs to focus on converting leads originating from API and Landing Page Submission and increase the count of leads from Lead Add Form

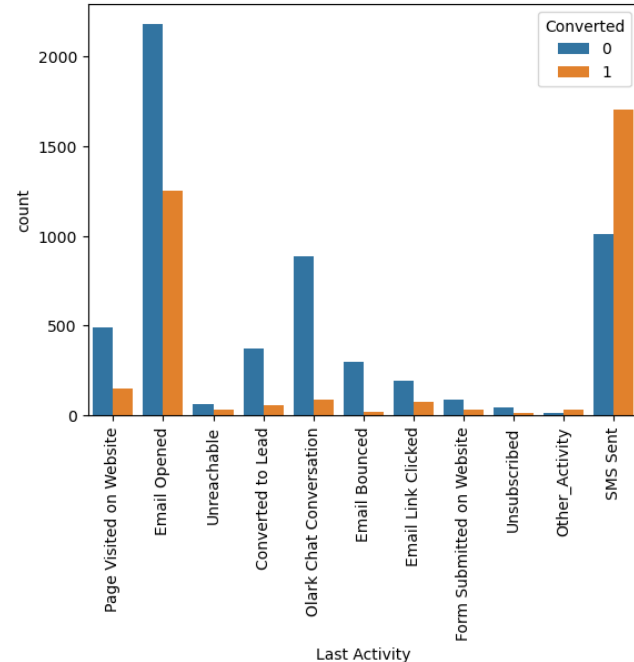


Business should focus on improvement in lead conversion for Olark Chat, Organic Search, Direct Traffic and Google. Whereas Reference and Welingak Website have a good conversion rate so the business should try to generate more leads from them.

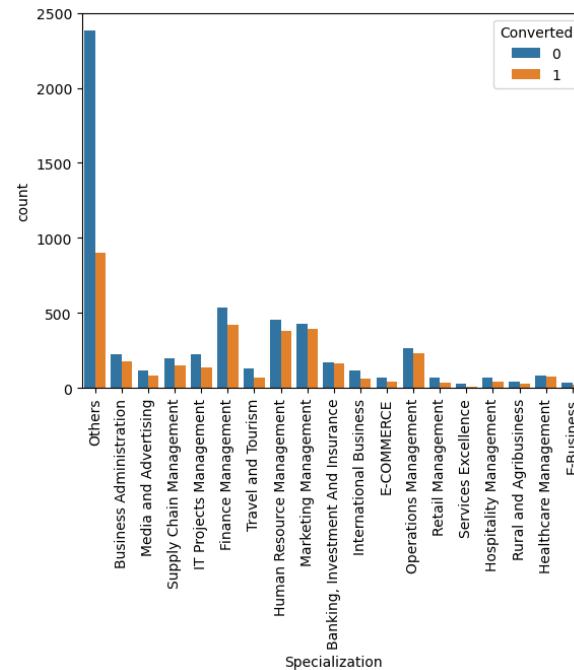


Business should plan their website such that people spend more time on it as we can see from the above plot time spend is highly related to conversion.

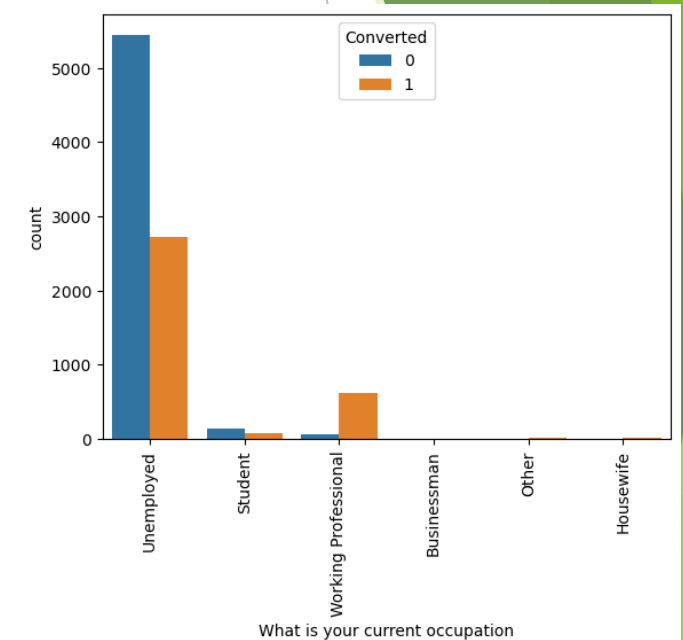
Exploratory Data Analysis Continued...



Email opened and SMS sent are the categories to watch out for

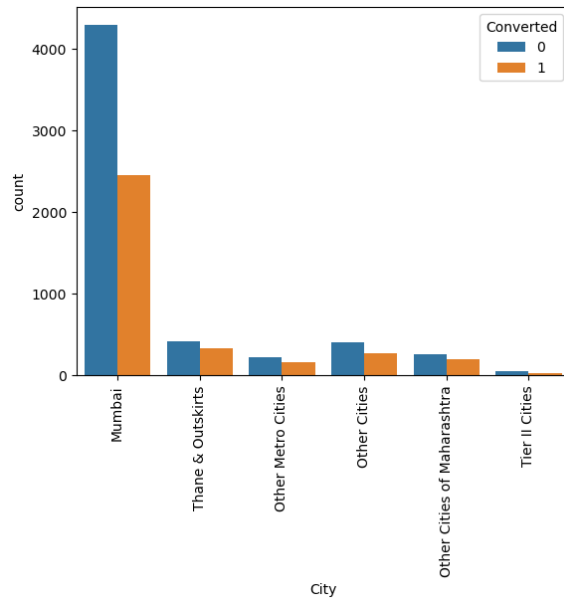


Business should focus on Specializations having high conversion rate.

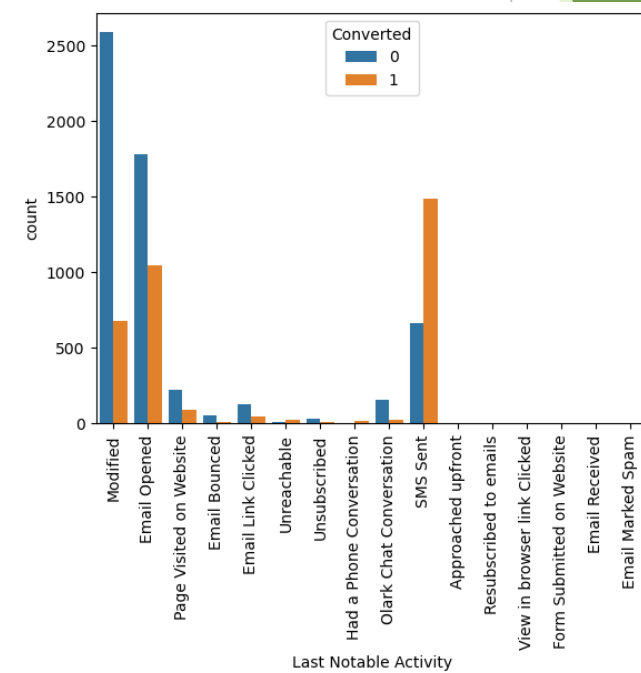


Business can focus on generating more leads from working professionals and try to convert more of unemployed people.

Exploratory Data Analysis Continued...



Business should focus on leads from Mumbai.



Business can see the fruitful categories and can target them.

Variables having an impact on conversion

Column

Prospect ID

Lead Origin

Lead Source

Do Not Email

Do Not Call

Converted

TotalVisits

Total Time Spent on Website

Page Views Per Visit

Last Activity

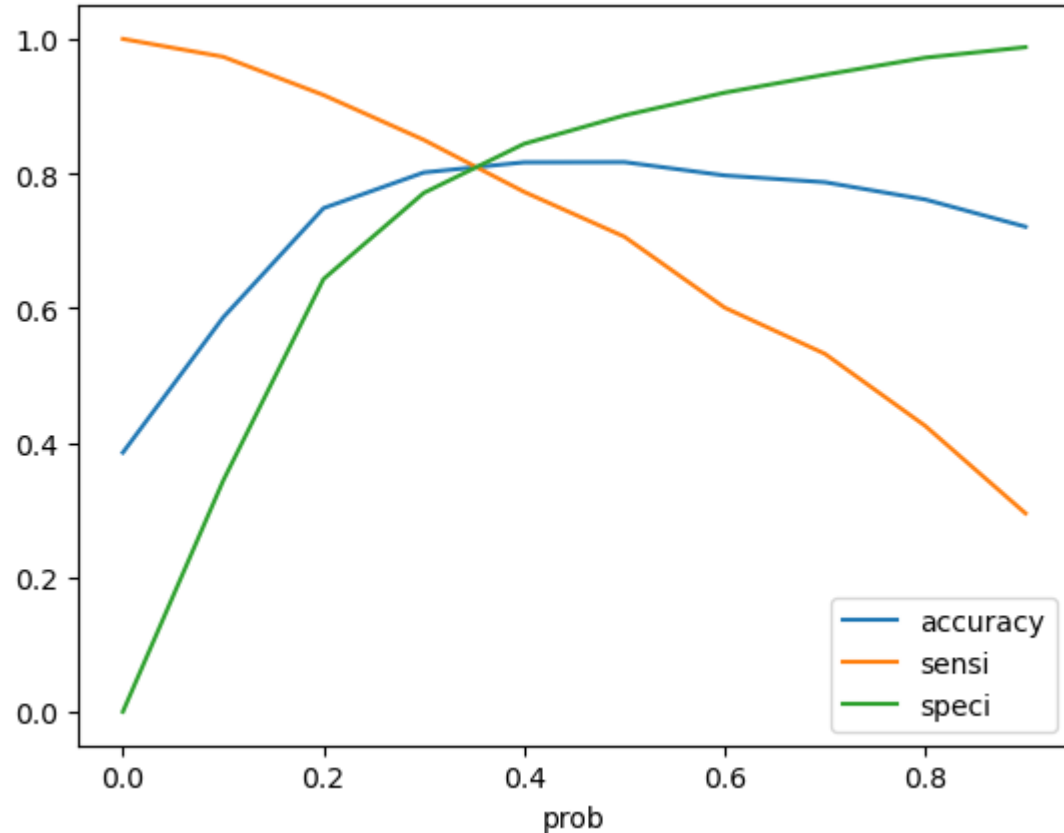
Specialization

What is your current occupation

City

Last Notable Activity

Prediction on Train Data Set



Accuracy : 0.8108959219020627

Sensitivity : 0.8172526573998364

Specificity : 0.8069142125480153

False Positive Rate : 0.19308578745198463

Positive Predictive Value : 0.7261169633127498

Negative predictive value : 0.8757643135075042

Prediction on Test Data Set

1396	338
193	796

Confusion Matrix

Accuracy : 0.8049944913698127

Sensitivity : 0.8048533872598584

Specificity : 0.8050749711649365

False Positive Rate : 0.19492502883506344

Positive Predictive Value : 0.7019400352733686

Negative predictive value : 0.8785399622404028

Results

Comparing the values obtained for Train & Test:

Training Data: Accuracy : 81.0 % Sensitivity : 81.7 % Specificity : 80.6 %

Test Data: Accuracy : 80.4 % Sensitivity : 80.4 % Specificity : 80.5 %

Thus we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . The Model seems to predict the Conversion Rate according to the requirement. Business can be confident on using this model.

Recommendations

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.
- The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.
- The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.
- The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.
- The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.
- The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.
- The company should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.
- The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.
- The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.
- The company should not make calls to the leads whose Last Notable Activity is 'Modified' as they are not likely to get converted.

Lead Source_Welingak Website	5.811465
Lead Source_Reference	3.316598
What is your current occupation_Working Professional	2.608292
Last Activity_Other_Activity	2.175096
Last Activity_SMS Sent	1.294180
Total Time Spent on Website	1.095412
Lead Source_Olark Chat	1.081908
const	-0.037565
Last Notable Activity_Modified	-0.900449
Last Activity_Olark Chat Conversation	-0.961276
Lead Origin_Landing Page Submission	-1.193957
Specialization_Others	-1.202474
Do Not Email	-1.521825

Final features along with coefficients.

Thank you