LEAD SCORING CASE STUDY

SUBMITTED BY-
ASHWINI PATIL
NIKITA GUPTA
VIKALP SHARMA

CONTENTS

Introduction

- 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.

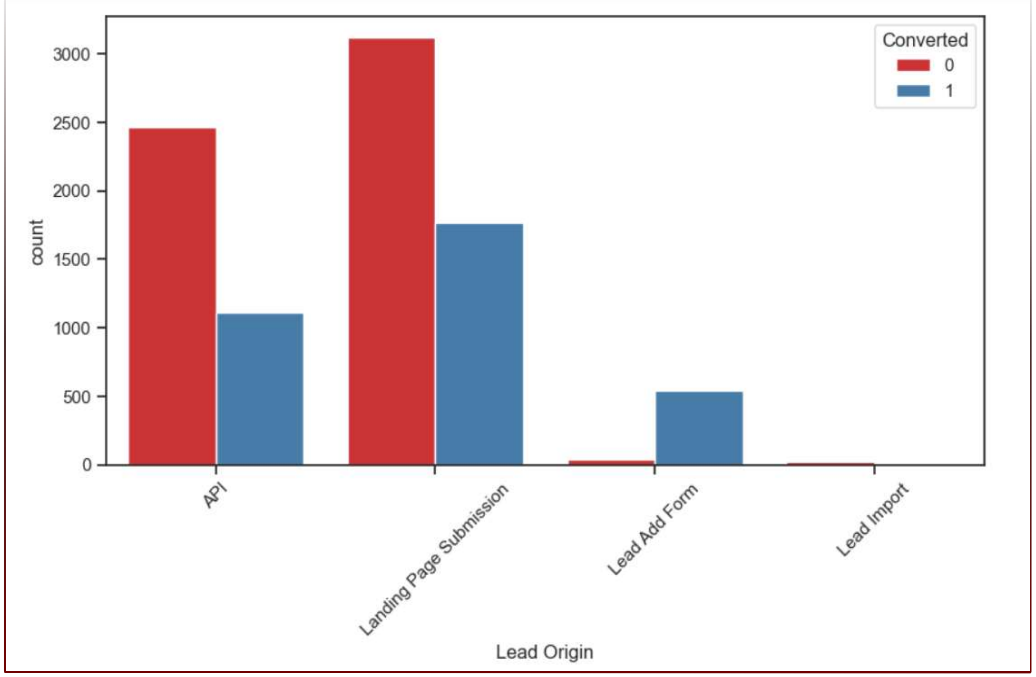
Problem Statement

- 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%.
- 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

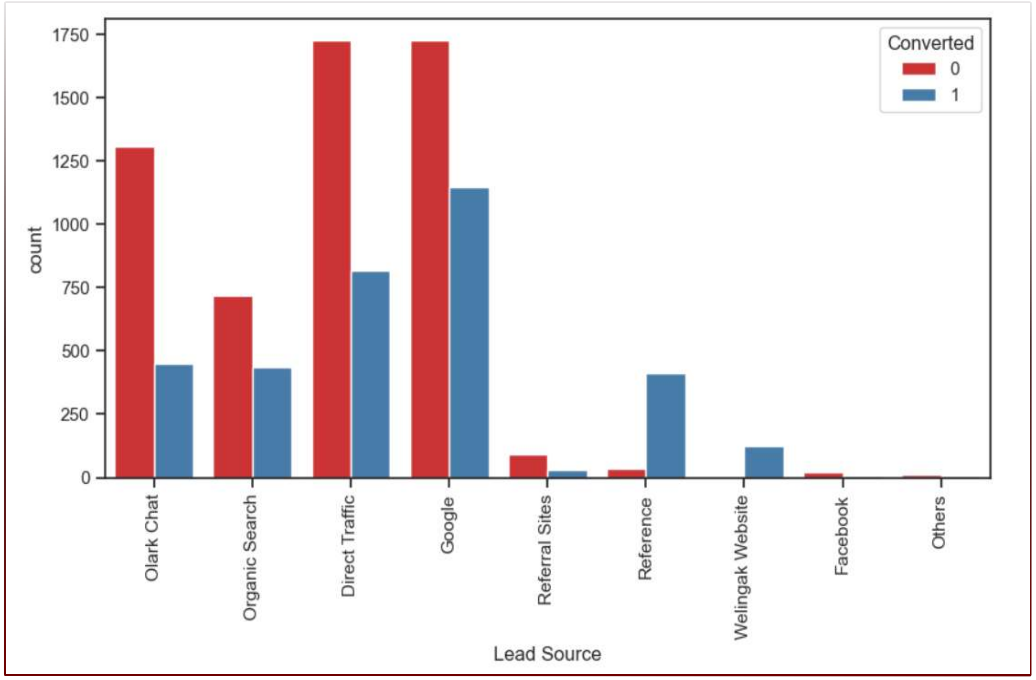
Business Problem

- 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.
- Building 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%.

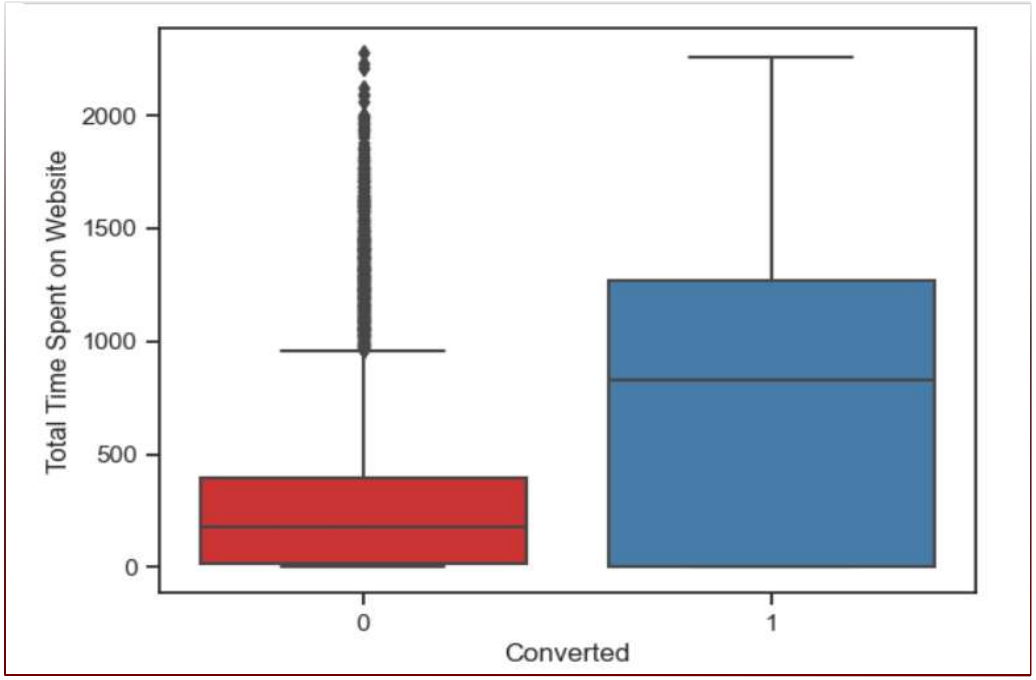
Exploratory Data Analysis



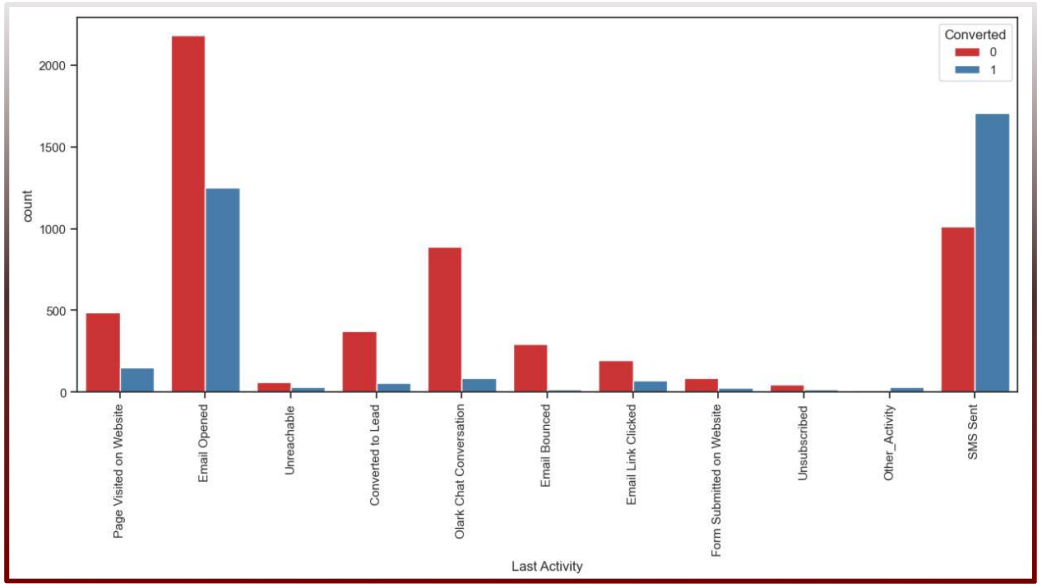
- API and Landing Page Submission have 30-35% conversion rate.
- Lead Add Form has more than 90% conversion rate.
- Lead Import are very less in count.



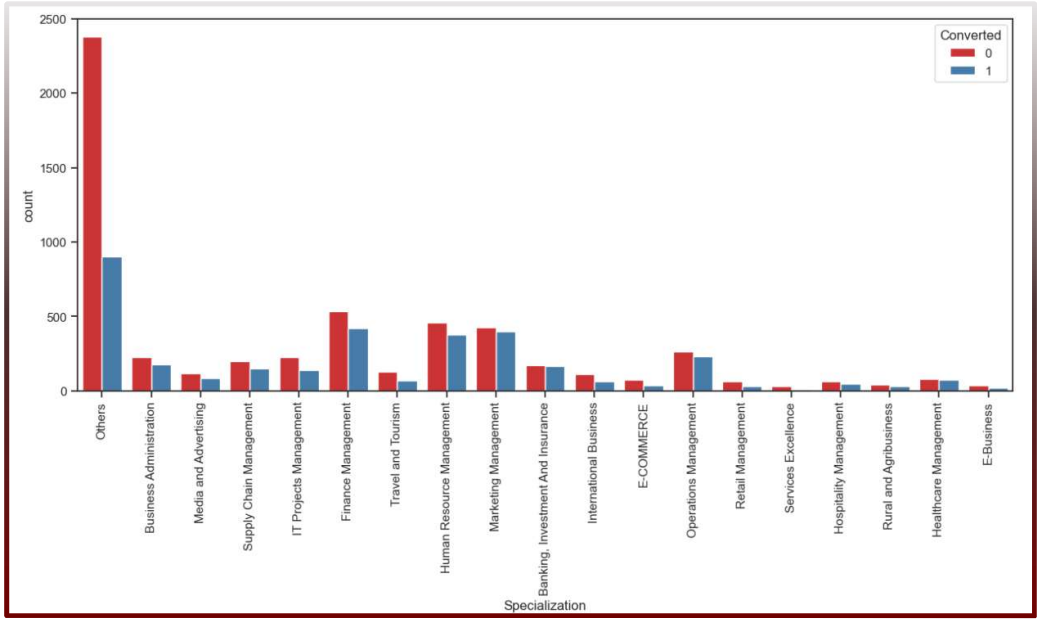
- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.



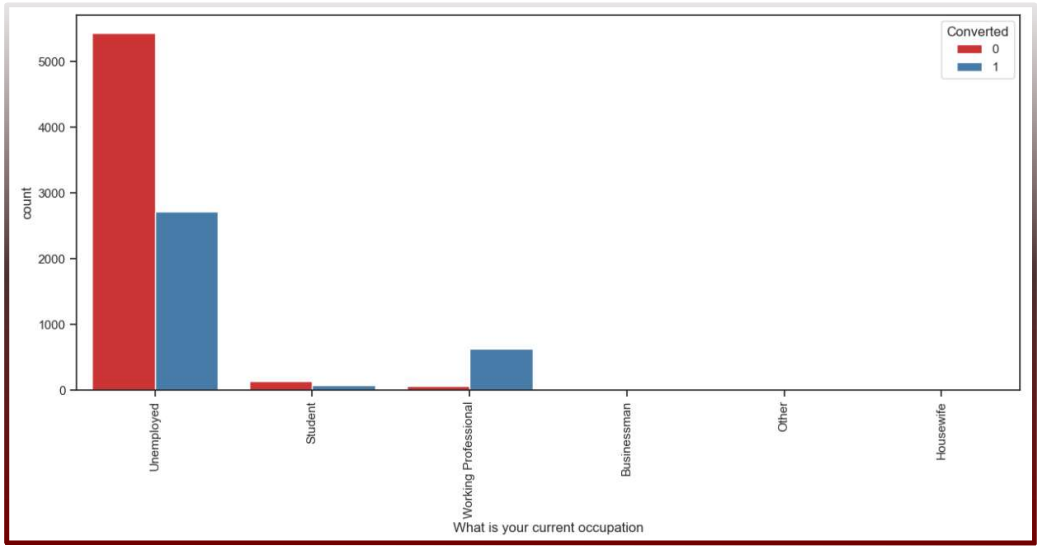
- Leads spending more time on the website are more likely to be converted.



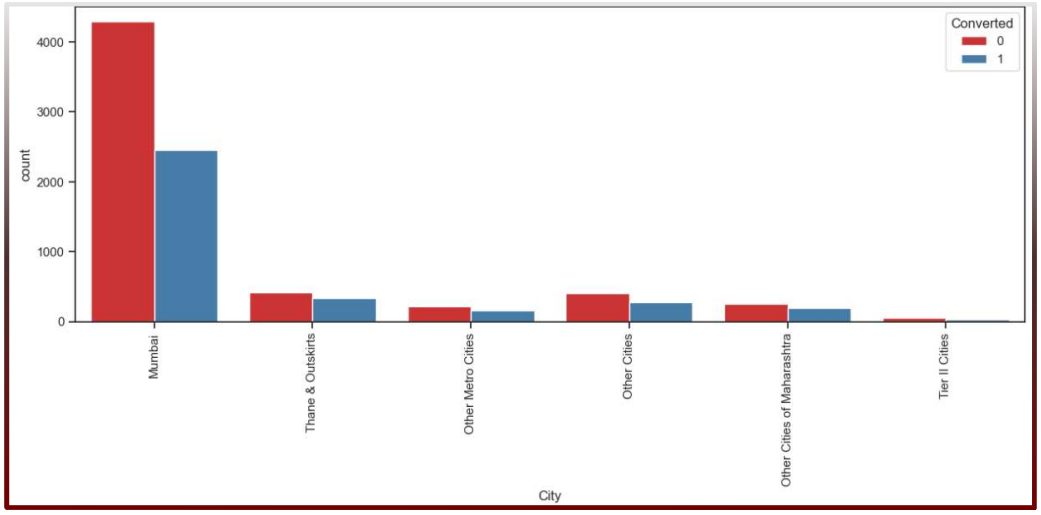
- Most of the leads have their Email opened as their last activity.
- Conversion rate for leads with last activity as SMS Sent is almost 50%.



- Focus should be more on the Specialization with high conversion rate.



- Working Professionals going for the course have high chances of joining it.
- Unemployed leads are the most in numbers but has around 30-35% conversion rate.

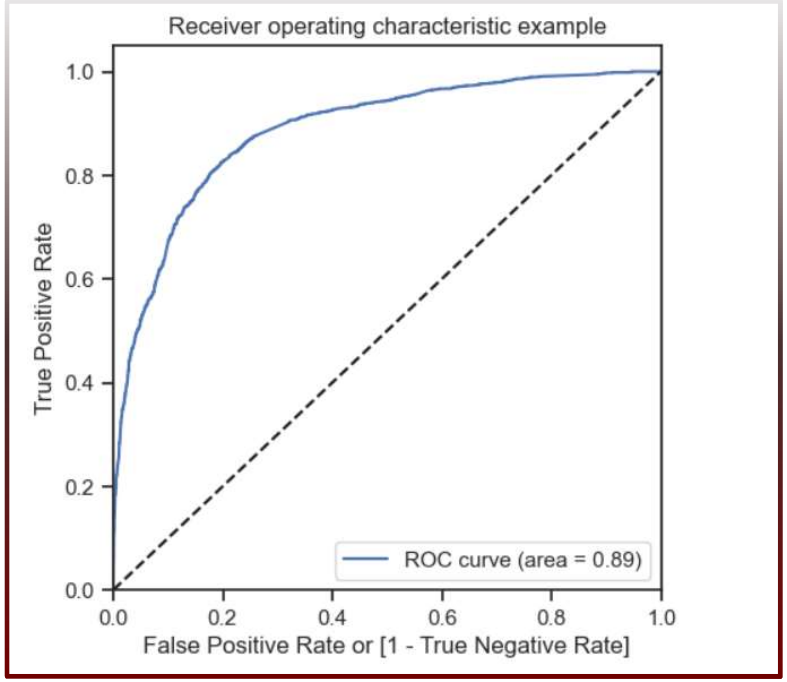


- Most leads are from Mumbai with around 50% conversion rate.

Model Evaluation

	Features	VIF
9	Specialization_Others	2.16
3	Lead Source_Olark Chat	2.03
11	Last Notable Activity_Modified	1.78
2	Lead Origin_Landing Page Submission	1.69
6	Last Activity_Olark Chat Conversation	1.59
8	Last Activity_SMS Sent	1.56
1	Total Time Spent on Website	1.29
4	Lead Source_Reference	1.24
10	What is your current occupation_Working Profes...	1.18
0	Do Not Email	1.13
5	Lead Source_Welingak Website	1.09
7	Last Activity_Other_Activity	1.01

- The P values of all variables is 0 and VIF values are low for all the variables, model-9 is our final model. We have 12 variables in our final model.



- Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.

	Prospect ID	Converted	Converted_prob	final_predicted	Lead_Score
1	1490	1	0.969057	1	97
8	4223	1	0.916621	1	92
16	1946	1	0.924467	1	92
21	2461	1	0.992551	1	99
23	5822	1	0.997991	1	100
...	...	...	...	...	...
2694	1566	1	0.947723	1	95
2699	6461	1	0.961562	1	96
2703	5741	1	0.908283	1	91
2715	6299	1	0.871977	1	87
2720	6501	1	0.854745	1	85

368 rows × 5 columns

- So there are 368 leads which can be contacted and have a high chance of getting converted. These are termed as Hot Leads

Observations

- 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 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.

Conclusion

- The logistic regression model is used to predict the probability of conversion of a customer.
- The final model shows 80.4% accuracy, recall of 70.6% and precision 79.4%
- The model worked well on train and test dataset
- The model looks good and is able to correctly identify the correct leads which has high chances of conversion using lead score prediction.
- In business terms, this model has the capability to adjust with the company's requirements in coming future.

THANK YOU