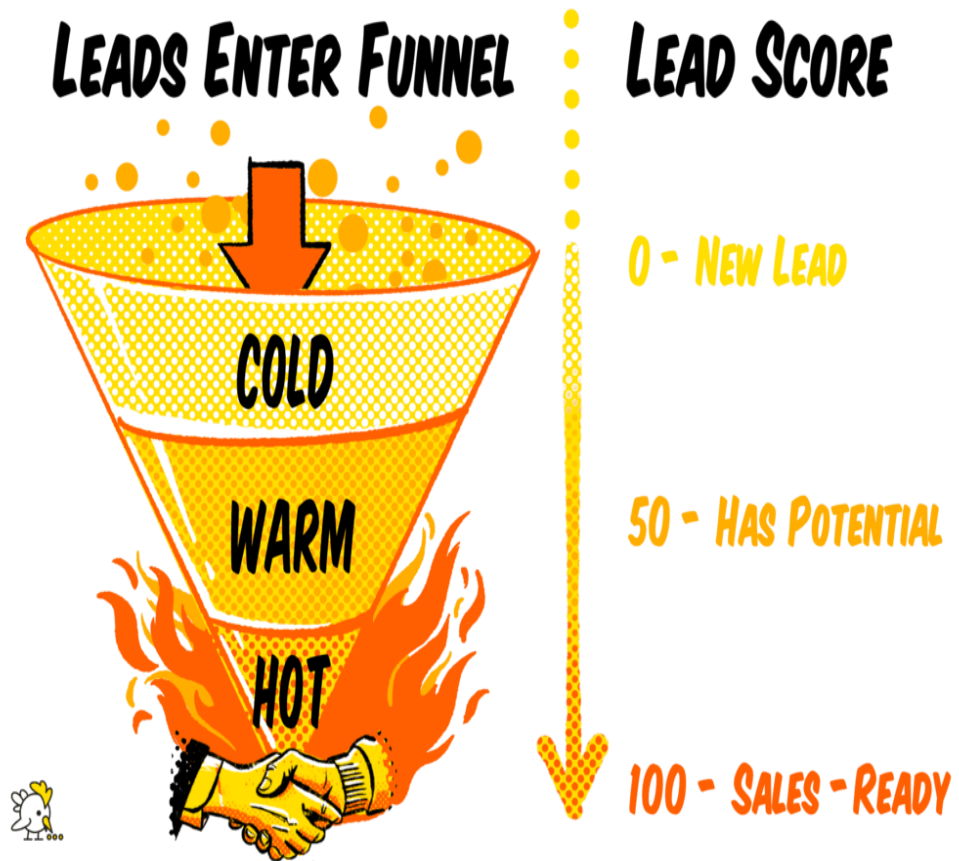


LEAD SCORING CASE STUDY



Prepared By

**Mohit Mittal
Vivek SD**

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.
- X education wants to find the hot leads which can be converted to paying customers with high probability.
- They get leads from various sources and once a lead is identified they are contacted by Sales team. Very few leads out of total leads purchase the course i.e. the conversion rate is very low.
- The company wants to identify the hot leads so that the Sales team can focus solely on them instead of calling everyone. In this way they want to increase their conversion rate.

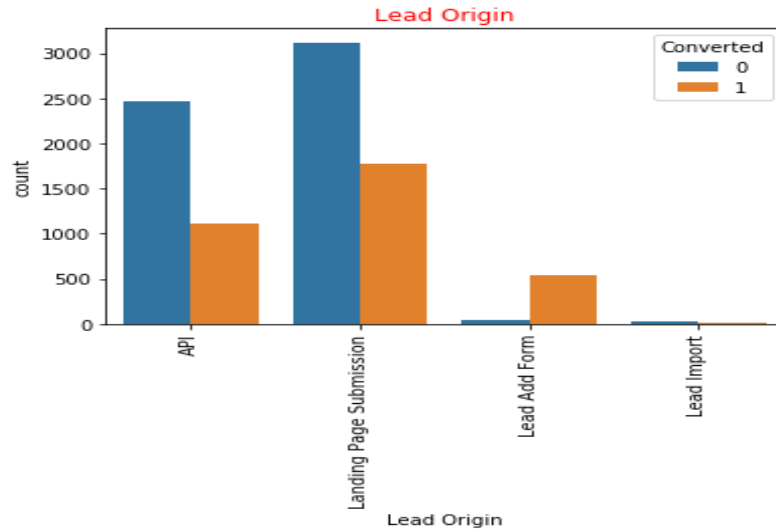
Business Goal/Expectation

- Company wants us to build a model for identifying the potential hot leads
- The model should assign a lead score to all the leads such that the customer with higher lead score has higher conversion chance and customer with lower lead score has lower conversion chance.
- Expected lead conversion rate from model is around 80%

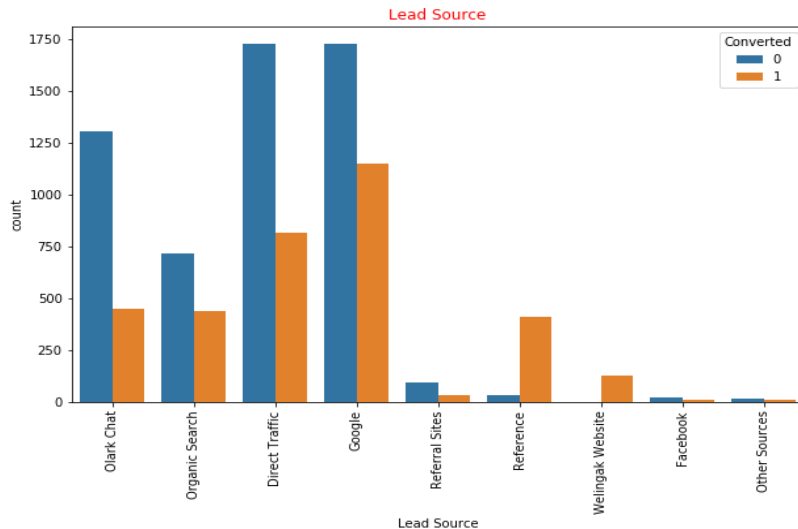
Approach

- Importing and understanding the data
- Cleaning the data and preparing it for Analysis
- Exploratory Data Analysis
- Splitting data into training set and test set
- Scaling
- Building logistic regression model with the best features
- Evaluating the model on training set
- Finding the optimal cut-off to get the best accuracy, sensitivity and specificity.
- Evaluating the model on test set
- Assigning lead score to leads based on model created

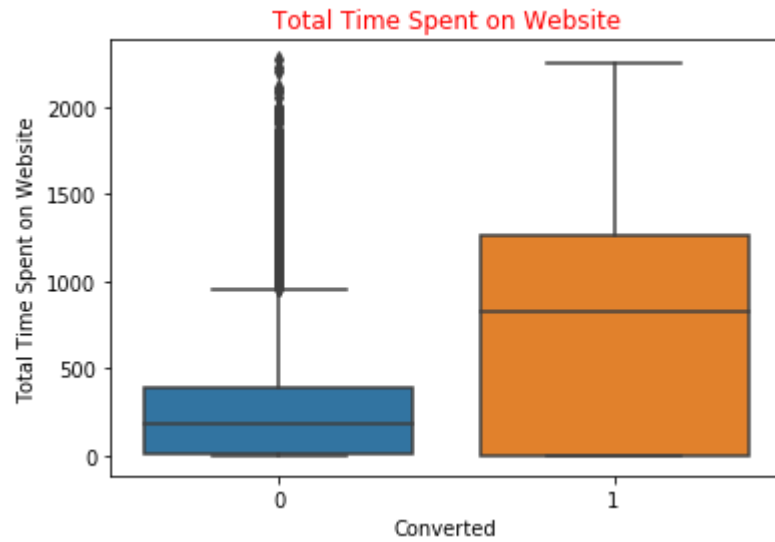
Exploratory Data Analysis Insights



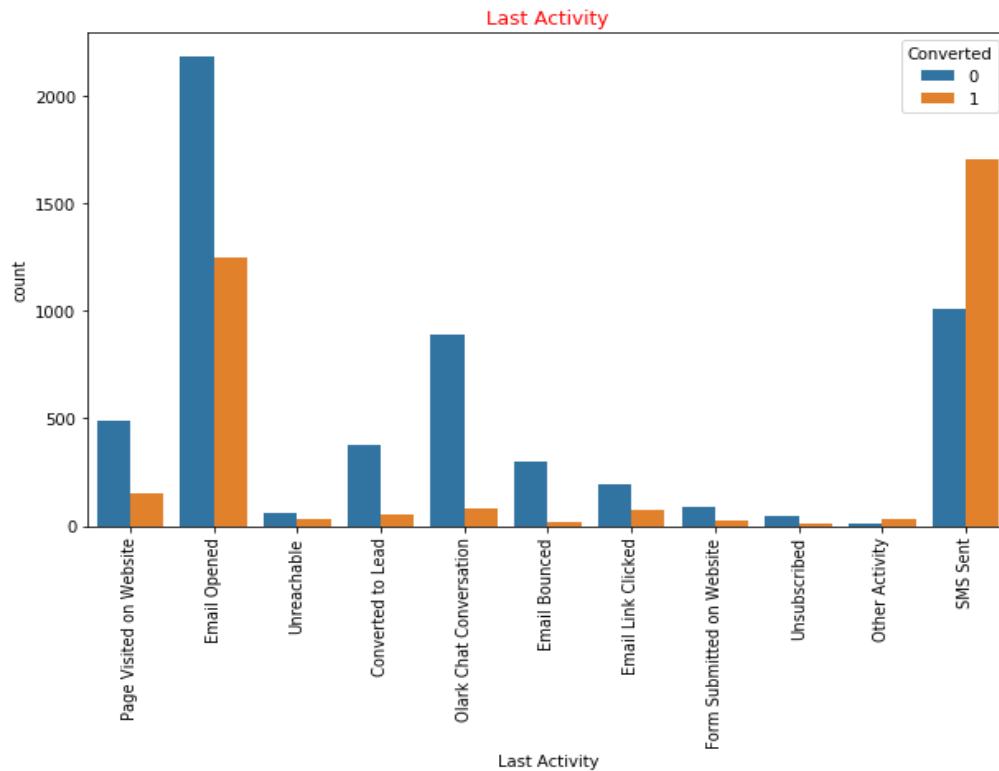
- Best Conversion rate among different categories of Lead Origin is that for Lead Add Form. More focus can be given to increase leads from this.
- Also API and Landing Page Submission generates lot of leads but their conversion rate is not very impressive.



- We get a lot of leads from Google and Direct traffic on website.
- Conversion rate is best for References and from Welingak Website. More advertising can be done to land people on this website and increase the number of leads.

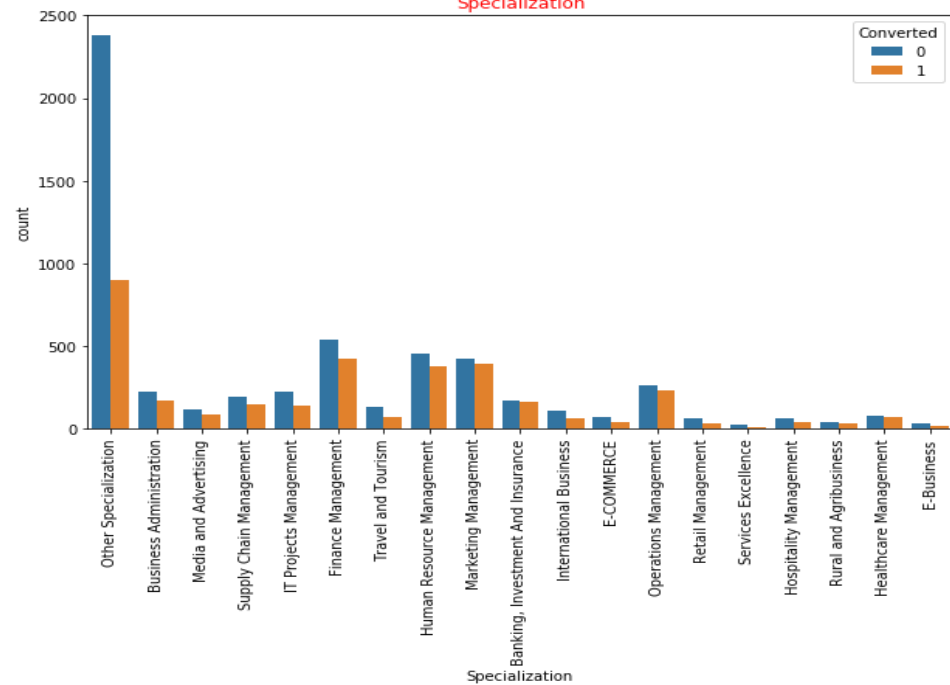


Leads who spend more time on website are likely to have better chance of getting converted than who spend less time.



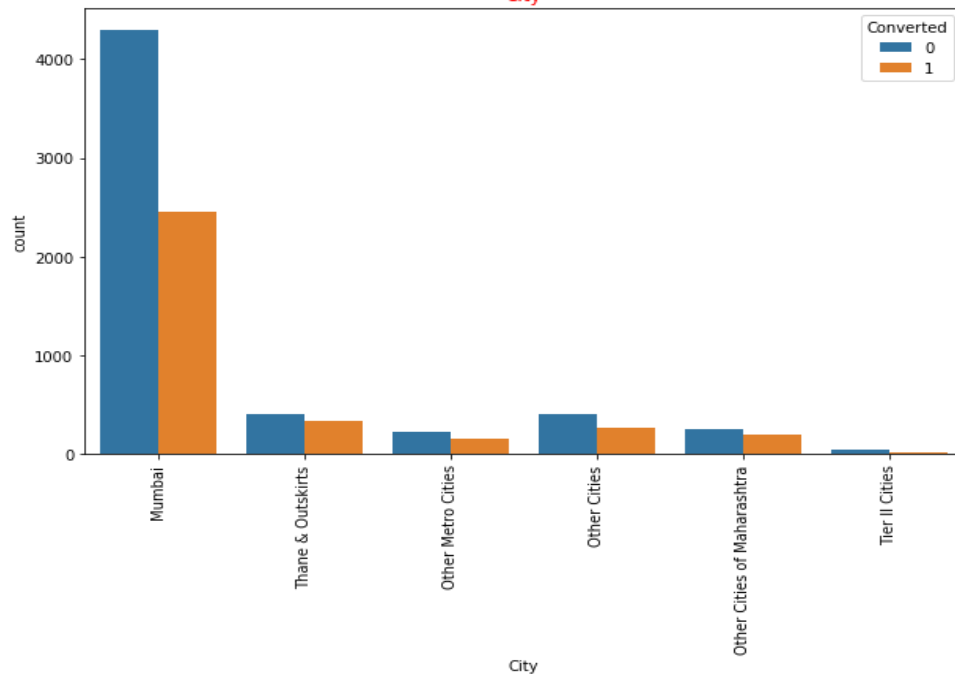
- Leads whose last activity is recorded as SMS Sent have a good conversion rate.
- Lot of leads have their last activity as Email Opened.

Specialization



We can observe that leads with Management related specialization show good conversion rates. More focus can be given to generate more leads from Management related profiles.

City



- It can be observed that maximum number of leads are generated from Mumbai City and they also show good conversion rate.
- Hence more leads can be generated from Mumbai city by contacting more people in the area.

Model Building Approach

- Splitting data into train and test set
- Initial Model built with all the selected feature
- Feature selection using RFE(Top 20)
- Building Subsequent models based on these features
- Eliminating features based on p-value and VIF and rebuilding the model with remaining features
- Evaluating accuracy and other metrics on training set
- Finding the optimal cut-off using intersection of accuracy, sensitivity and specificity on graph.
- Evaluating the model on test data
- Assigning lead score to the leads.

Final Model Features and Coefficients

	coef	std err	z	P> z	[0.025	0.975]
const	-0.3620	0.143	-2.523	0.012	-0.643	-0.081
Total Time Spent on Website	1.0408	0.042	24.902	0.000	0.959	1.123
Lead Origin_Landing Page Submission	-1.3454	0.134	-10.063	0.000	-1.607	-1.083
Lead Origin_Lead Add Form	3.6438	0.276	13.202	0.000	3.103	4.185
Lead Source_Welingak Website	1.7255	0.788	2.188	0.029	0.180	3.271
Last Activity_Email Bounced	-1.7596	0.394	-4.464	0.000	-2.532	-0.987
Last Activity_SMS Sent	1.4465	0.085	16.975	0.000	1.279	1.613
Specialization_Other Specialization	-1.2098	0.137	-8.859	0.000	-1.477	-0.942
Tags_Busy	0.7746	0.236	3.282	0.001	0.312	1.237
Tags_Ringing	-3.3931	0.263	-12.913	0.000	-3.908	-2.878
Tags_Will revert after reading the email	1.2797	0.093	13.774	0.000	1.098	1.462
Last Notable Activity_Modified	-1.1910	0.084	-14.177	0.000	-1.356	-1.026
Last Notable Activity_Olark Chat Conversation	-1.5595	0.330	-4.733	0.000	-2.205	-0.914
What is your current occupation_Working Professional	2.4740	0.217	11.416	0.000	2.049	2.899

Final model contains the following features and it also shows the coefficients of each feature:

“Lead Origin_Lead Add Form” tend to be the most effective feature with most weightage according to coefficient.

Other important features are
“What is your current occupation_Working professional”
“Lead Source_Welingak Website”
“Last Activity_SMS sent”
“Total time spent on Website”

Evaluation metrics Result(train set)

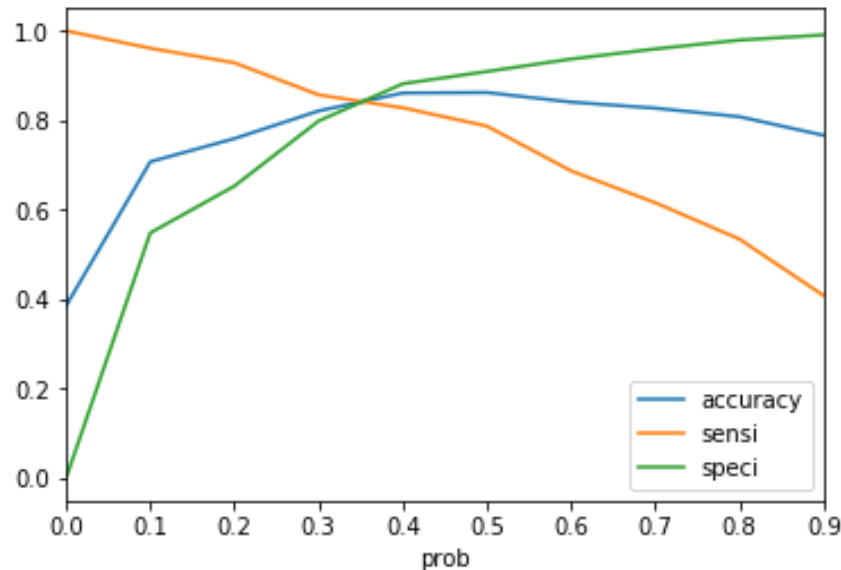
Confusion Matrix

```
[[3322  583]
 [ 385 2061]]
```

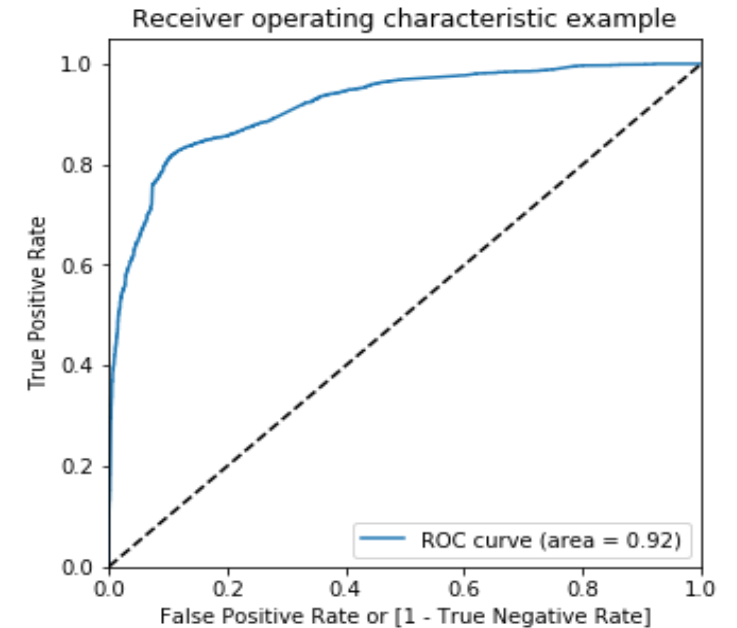
Accuracy - 84.75%

Sensitivity - 84.26%

Specificity - 85.07%



We get a optimal cut-off of 0.35 from the intersection point of Accuracy, sensitivity and specificity.



ROC curve for training data shows curve at top left corner and ROC curve area of 0.92, indicating good model.

Evaluation metrics Result(Test Set)

Confusion Matrix

```
[[1457  277]
 [ 175  814]]
```

Accuracy - 83.40%

Sensitivity - 82.30%

Specificity - 84.02%

Lead Score of Top 10 rows in test set

Prospect ID	Converted	Converted_prob	final_predicted	Lead_Score
3271	0	0.285524	0	29
1490	1	0.982201	1	98
7936	0	0.254316	0	25
4216	1	0.762393	1	76
3830	0	0.261244	0	26
1800	1	0.780140	1	78
6507	0	0.012962	0	1
4821	0	0.209163	0	21
4223	1	0.979632	1	98
4714	0	0.276561	0	28

Conclusions and Recommendations

- **Landing page Submission and API** generates large number of leads. We can try improving the **lead add form** to generate more leads as the conversion rate is good for add form.
- We can encourage people to provide more **references** as they have good conversion rate and also we can improve the **Welingak website** marketing so that more people land on the website and generate more leads. We can offer good rewards for referrals to increase references.
- Leads who **spend more time on website** are likely to get converted. We can make the website more appealing and interactive to engage more leads. Also we can have demo lesson videos related to course on website so that people get fair understanding and are more interested to take up the course.
- We can focus more on **SMS based communications** to communicate with leads. Emails can be considered as the other option for communication.
- We can focus more on the **management related profiles** to get more leads as they show better conversion rates.
- We tend to get most number of **leads from Mumbai city**, hence we can focus more on Mumbai based profiles to get more leads and in turn better conversion chances.
- In our model we are getting a accuracy of around 83% which indicates we will be able to identify the hot leads and reduce the efforts of the sale team and in turn increase the Conversion rate.