Lead Score Case Study

Logistic Regression

Group member:

Mr.Anurag Srivastava

Ms.Ishwari Mote

Ms.Raina Jain

Problem Statement

X Education is an organization which provides online courses for industry professional. The company marks its courses on several popular websites like google.

X Education wants to select most promising leads that can be converted to paying customers.

Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion. Leads come through numerous modes like email, advertisements on websites, google searches etc.

The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

Goal

- ➔ The company requires a model to be built for selecting most promising leads.
- ➔ Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted, the lower it is the lesser the chances of conversion
- ➔ The model to be built in lead conversion rate is around 80% or more.

STRATEGY

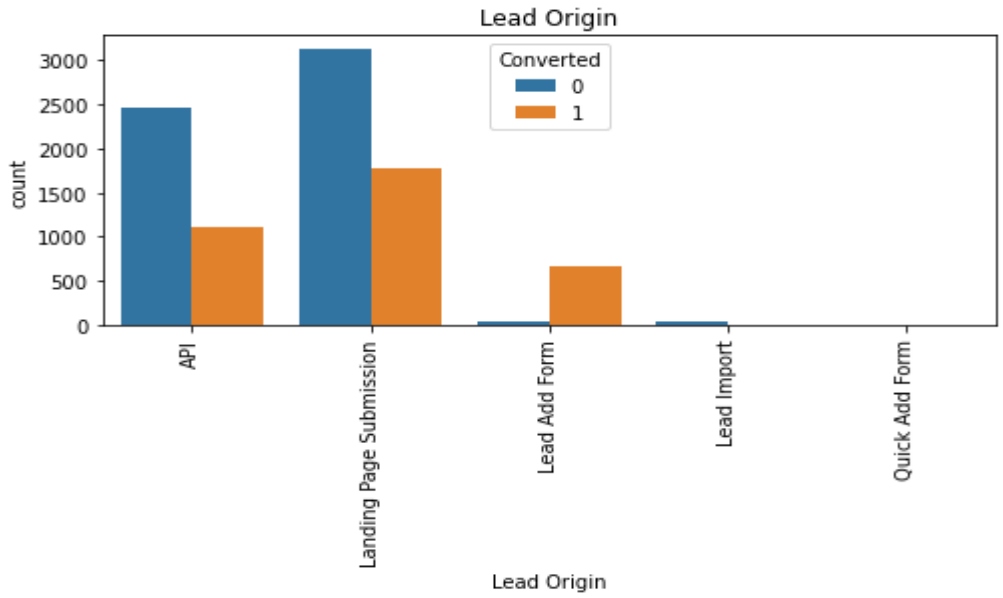
- Import data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evaluation

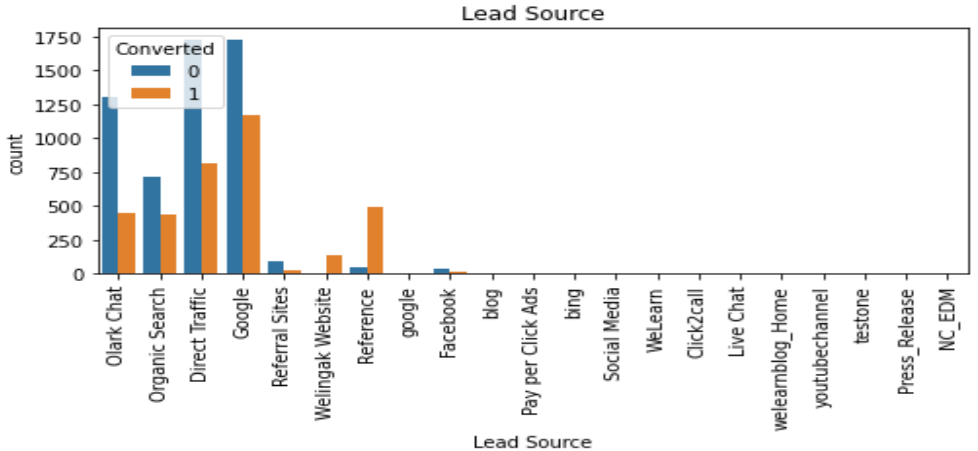
Top Factors that Impacted Conversion of Leads

- ❏ Lead Origin
- ❏ Lead Source
- ❏ Do Not Email
- ❏ Converted
- ❏ TotalVisits
- ❏ Total Time Spent on Website
- ❏ Page Views Per Visit
- ❏ Last Activity
- ❏ Country
- ❏ Specialization
- ❏ How did you hear about X Education
- ❏ Tags
- ❏ Lead Quality
- ❏ Lead Profile
- ❏ City
- ❏ Asymmetrique Activity Index
- ❏ Asymmetrique Profile Score
- ❏ Last Notable Activity

The majority of leads came from landing page content, about 3000 of them converted, and then through APIs, about 2500 converted.

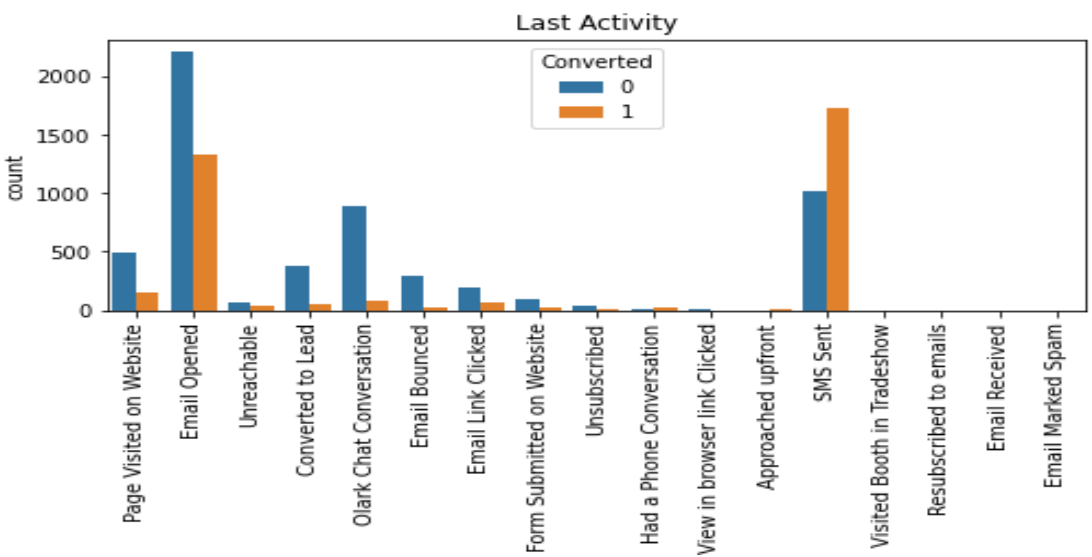
Most leads were sourced from Google and Direct Traffic , with 1750 of leads converted, with organic search, and Olark chat converting around 750, and 1250 respectively.

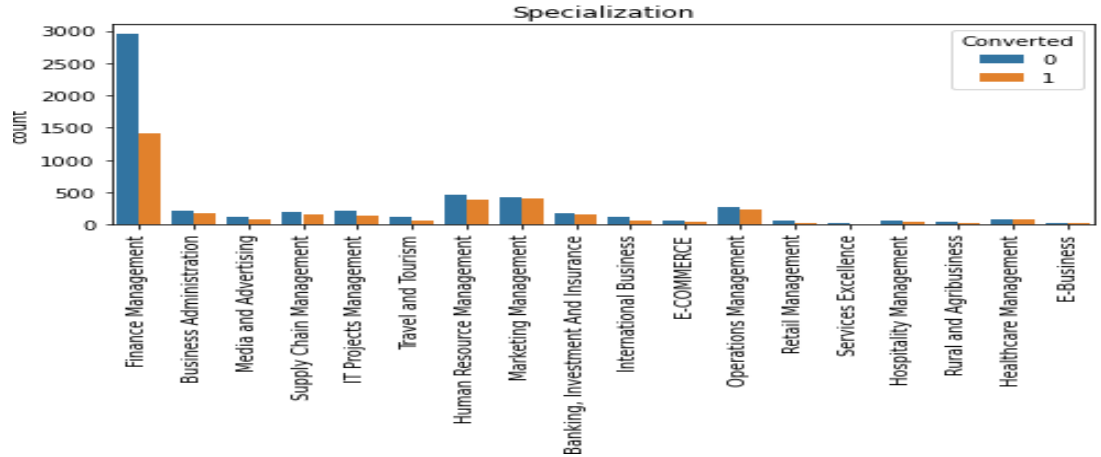




Most of the lead have their Email opened as their last activity. Conversion rate for leads with last activity as SMS Sent is almost 1000 count.

Leads with Finance management count is more than 2500, and Marketing Management is about 500.

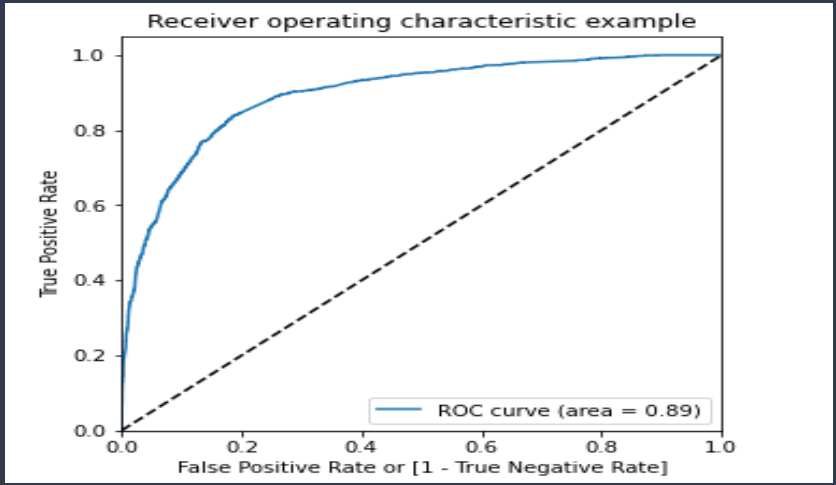


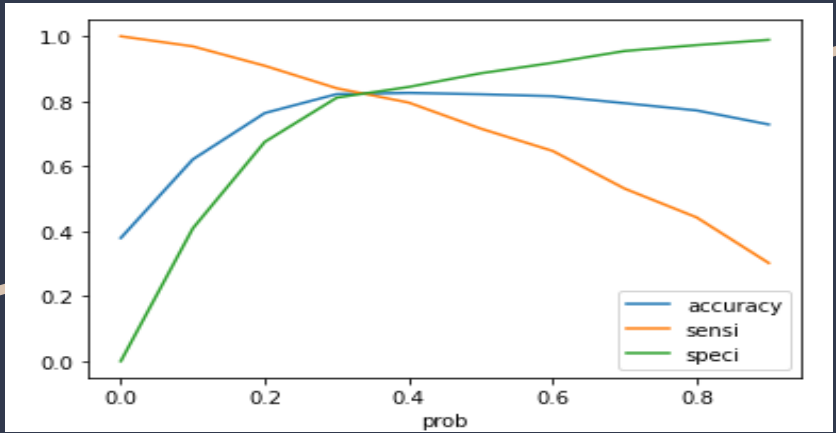


Model Building

- Splitting into train and test set
- Scale variables in train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate variables based on high p-values
- Check VIF value for all the existing columns
- Predict using train set
- Evaluate accuracy and other metric
- Predict using test set
- Precision and recall analysis on test predictions

Model Evaluation (Train)

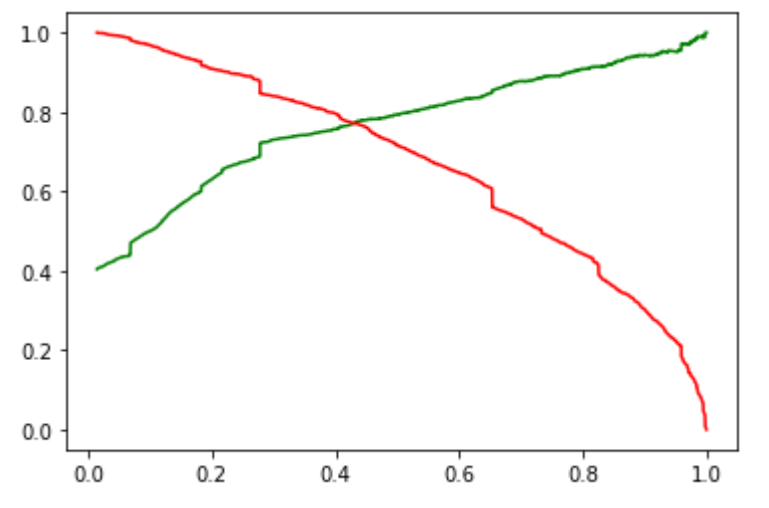




- Finding Optimal Cut off Point
- Optimal cut off probability is that
- probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.35

- 82% Accuracy
- 82 %Specificity
- 82 %Sensitivity

Trade of Curve between Precision and recall



- Precision Value is 79%
- Recall Value is 71%

Model Evaluation (Test)

- 81% Accuracy
- 82% Specificity
- 80% Sensitivity

Conclusion

- The company should make calls to the leads coming from the lead Source "Welingak Websites" and "Chat" 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 make calls to the leads coming from the lead Origin "Add form" 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.


Thank You !