LEAD SCORE CASE STUDY

By: Mehak, Akhil Gupta

Problem Statement

An education company named X Education sells online courses to industry professionals.

Although X Education gets a lot of leads, its lead conversion rate is only 30%.

The company wants to increase lead conversion rate to 80%

Goal

- Build a logistic regression model to assign a **lead score** between 0 and 100 to each of the leads which can be used by the company to target potential leads.

- A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

Approach

Exploratory Data Analysis	Data Preparation	Build Model on Train Set	Making Predictions on Test Set & Model evaluation
Handling 'Select', Missing values, Outliers, Merging small frequencies, Visualizations like Count plots, boxplots, pair plots and correlation heatmap	Binary encoding, Dummy variable creation, Test Train split, Standardization, Determining X and y	Use RFE to get top 15 features, using stasmodels build model and eliminate predictors with p-value > 0.05 or VIF > 5. Building ROC Curve. Find the optimal cuttoff using metrics like Accuracy, Sensitivity and Specificity	Standardization of numerical columns, predict lead conversion pribability and assign Lead score to each lead, calculate metrics metrics like Accuracy, Sensitivity and Specificity

Factors impacting Lead Conversion

	Cofficient	Sign
Tags_Closed by Horizzon	7.674683	+
Tags_Lost to EINS	6.058750	+
Tags_Will revert after reading the email	4.811041	+
Lead Source_Welingak Website	4.337418	+
Tags_Ringing	3.451202	-
Tags_switched off	3.302062	-
Tags_Already a student	2.927412	-
Last Activity_SMS Sent	2.100536	+
Tags_Interested in other courses	1.930141	-
Last Notable Activity_Modified	1.703421	-
Last Notable Activity_Olark Chat Conversation	1.592738	-
Lead Origin_Lead Add Form	1.561728	+
Lead Source_Olark Chat	1.397307	+
Total Time Spent on Website	1.091531	+
Do Not Email	0.955641	-

How these factors impact Lead conversion rate?

- Determine Potential leads i.e. 'HOT' leads that can get converted.

- Measures to take to increase the lead conversion rate.

Tags
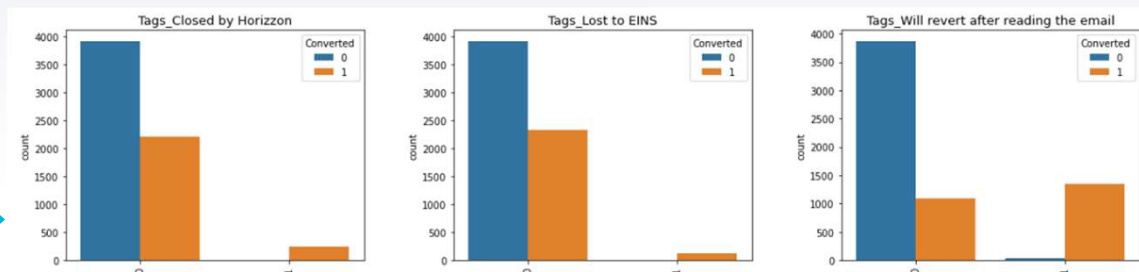
- Closed by Horizon
- Lost to EINS
- Will revert after reading the mail

Have a very high conversion rate. Company needs to focus on these tags to increase the lead conversion rate.
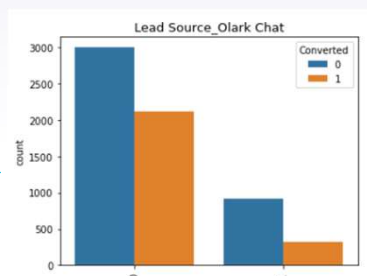
Negative impact on Conversion

- Already a student
- Interested in other courses
- Switched off/ Ringing

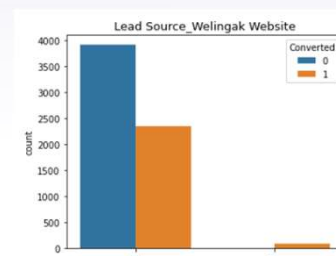Have a low conversion rate. No need to focus on these leads.
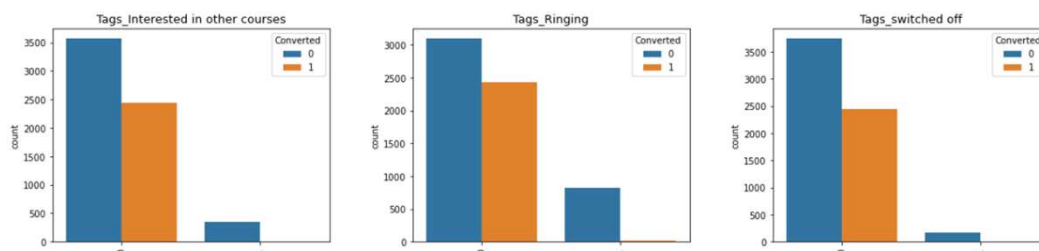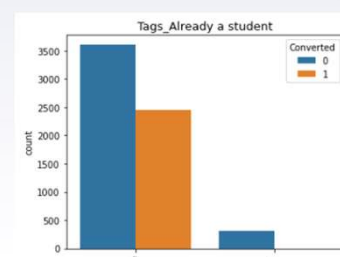
Lead Source

- Olark Chat

To increase the conversion rate for the Company must increase the number of representatives on Olark chat, so that the potential customers are attended and not missed out.
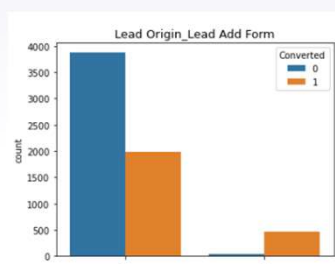
- Welingak Website

As compared to Olark chart, there are less customers on Welingak website. So the website must be promoted to attract more leads.
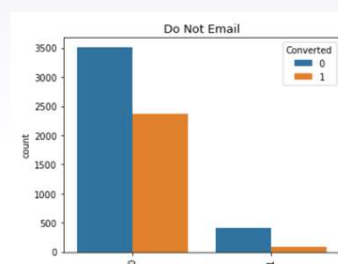
Lead Origin

Lead Add Form

More focus should be given to leads who originated from 'Add form'.
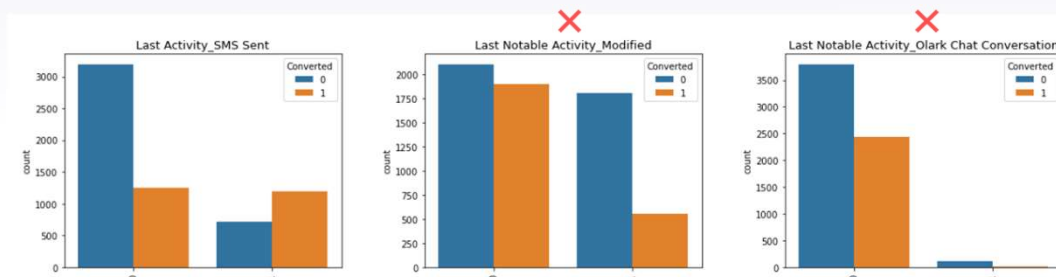
Do not email

Negative impact on Conversion

Company needs to focus more on leads who have opted for mails to increase the lead conversion.
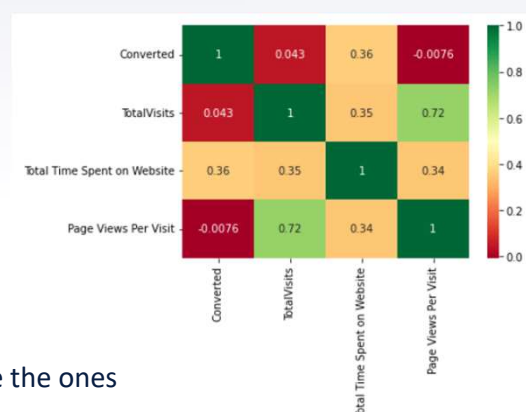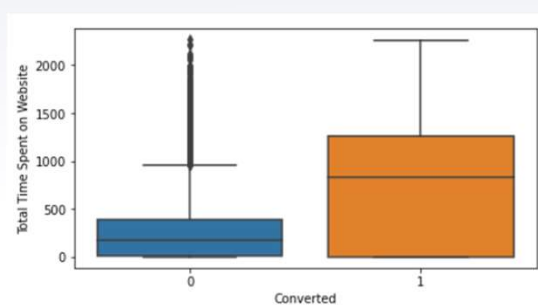
Last Activity/ Notable Activity

Company needs to track the last activity of the leads.

If the last activity is **SMS sent**, there is a high chance that lead will get converted.

If the last notable activity is Modified or Olark chat conversion, the conversion rate is lower.

Total time Spent on Website

- The leads spending more time on the website are the ones to be focused on.
- Total time spent on website is highly correlated to the lead conversion
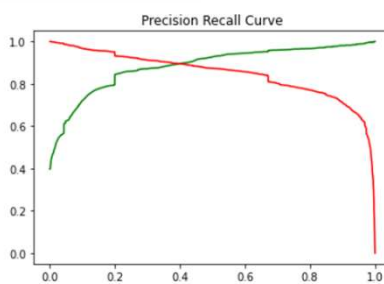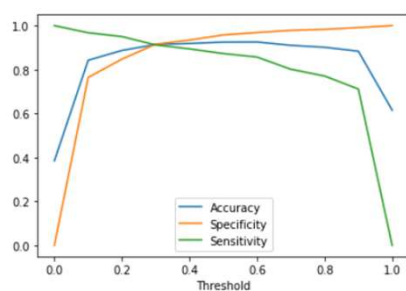
Metrics for the model

| Correctly predict Converted as Converted | + | Reduce 'False Negatives' | → | Maximize Sensitivity to get HOT leads |

Train set

Accuracy: 0.9129
Sensitivity: 0.9215
Specificity: 0.9076
Precision: 0.8617

Test set

Accuracy: 0.916
Sensitivity: 0.9189
Specificity: 0.9142
Precision: 0.8699

Lead Score

Train set (lead_score_train_df)

Test set (test_lead_df)

- Lead score (0- 100) column is present for all the leads.

- The leads with lead score greater than 30, can be targeted as they have high probability of getting converted.

Conclusion

Focus on leads with
- Current status of lead (tag) : Closed by Horizzon, Lost to EINS, Will revert after reading the mail
- When the lead source was Welingak website or Olark chat
- When the lead origin is Lead add form.
- When the last activity was SMS Sent
- Total time spent on the website.

Do not focus on leads with
- Current status of lead (tag) : Switched off, Ringing, Already a student, Interested in other courses
- Last notable activity: Olark chat conversation, Activity Modified

Thank you !