Leads Scoring Case Study

Presented By: Vedant Patel &Vrushali Rane

Problem Statement

X Education is an online education company with a low lead conversion rate. X Education gets a lot of leads, but only a small percentage of them convert into customers.

Business Objective

- Identify the most promising leads in the database.
- Develop a lead scoring model to predict Hot leads.
- Increase the lead conversion rate to 80%.

Scope

The lead scoring model will be developed using a variety of data sources, lead's demographic information, lead's interactions with the company.

Benefits

- Increased lead conversion rate.
- Improved sales efficiency.
- Increased customer satisfaction.

Approach and Methods

1. Read and understand the data.

2. Drop the variables with a high percentage of missing values, impute missing values, and remove outliers.

3. Perform EDA on the data to get a better understanding of its structure and distribution.

4. Convert the Binary variables to 1 or 0, create dummy variables for categorical variables and drop those original variables.

5. Split the data into a 70% training set and a 30% test set.

6. Use min-max scaling to scale the numerical variables, then use the statsmodels package to create our initial model.

7. Use recursive feature elimination (RFE) to select the most important features in the training set and drop features with p-values greater than 0.05. Calculate the variance inflation factor (VIF) for each of the remaining feature and drop features with VIF greater than 5. Created a dataframe with the converted probability values assumed that a probability value of more than 0.5 means 1 and a probability value of less than 0.5 means 0. Calculate the confusion matrix, the accuracy, sensitivity and specificity of the model.

8. Plot the ROC curve for the selected features.

9. Find the optimal cutoff point for the ROC curve and observe the new accuracy, sensitivity and specificity of the model. Calculate the lead score.

10. Calculate precision and recall metrics on data sets.

11. Make predictions on the test set.

Results of Analysis

❖ Leads who spend more time on the website are more likely to convert.

❖ They are more interested in the company's products or services.

❖ They are taking the time to learn more about them.


Total Time Spent on Website v/s Converted
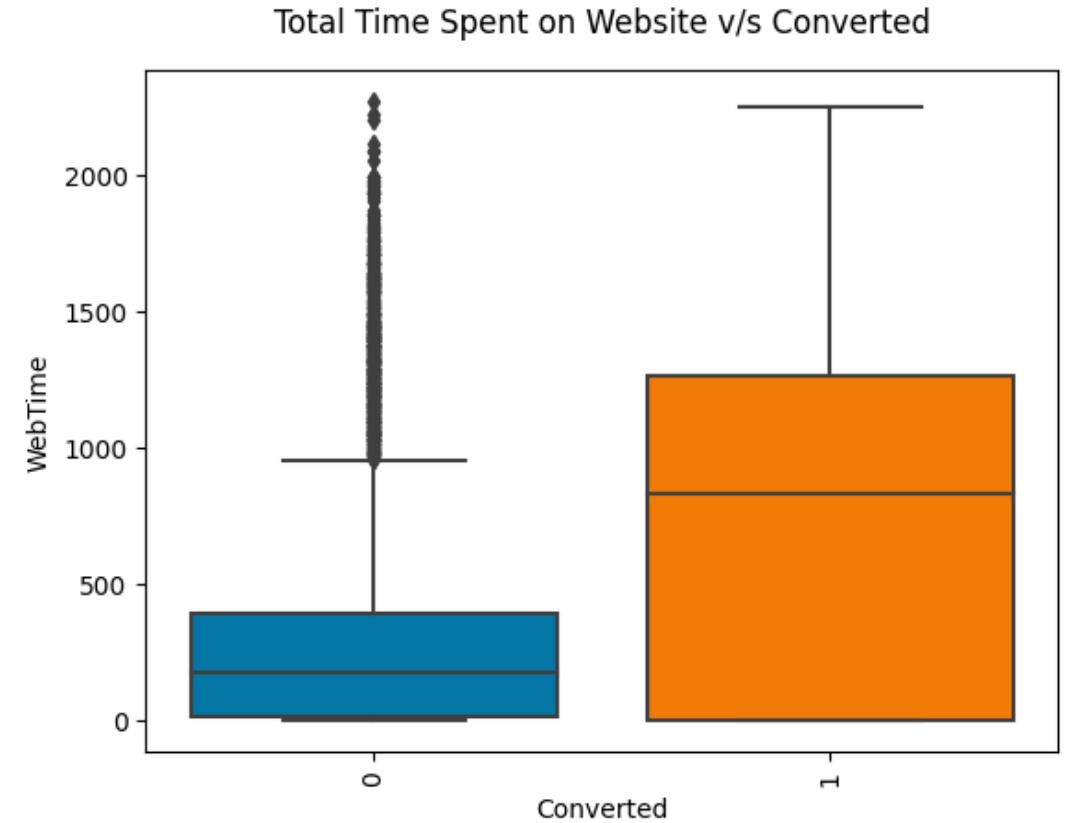
Results of Analysis

❖ Leads who come from the lead add form are more likely to convert.

❖ They have already expressed interest in the company's products or services by filling out the form.

Lead Origin Vs Converted
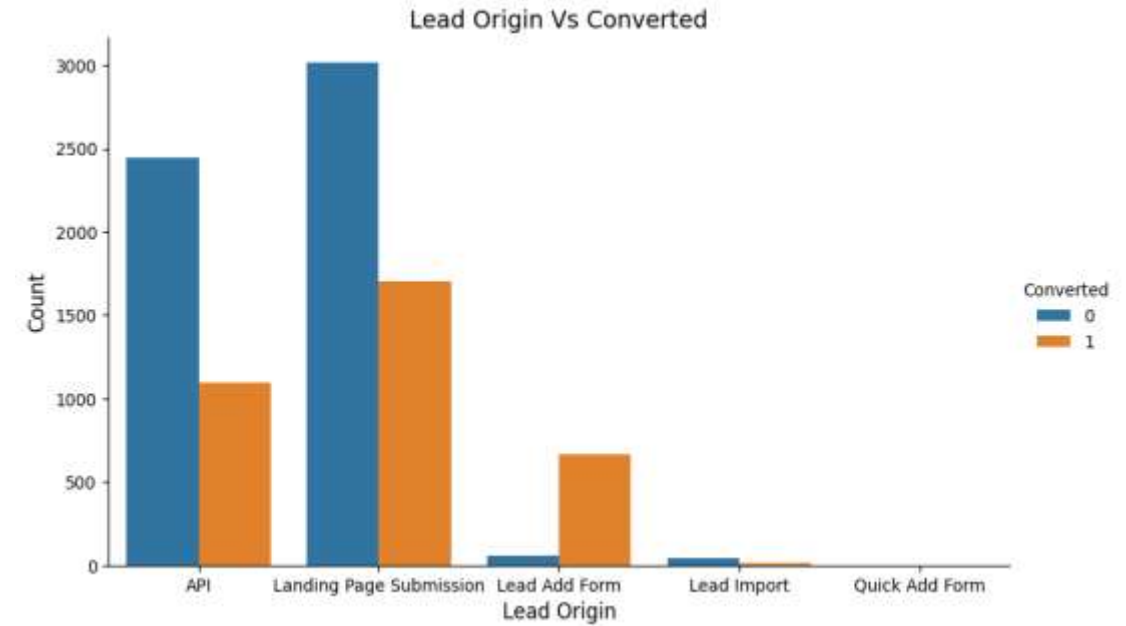
Results of Analysis

❖ Leads who have had a phone conversation with a sales representative are more likely to convert .

❖ They have had the opportunity to learn more about the company and its products or services from a sales team who can answer their questions and address their concerns.


Last Notable Activity Vs Converted
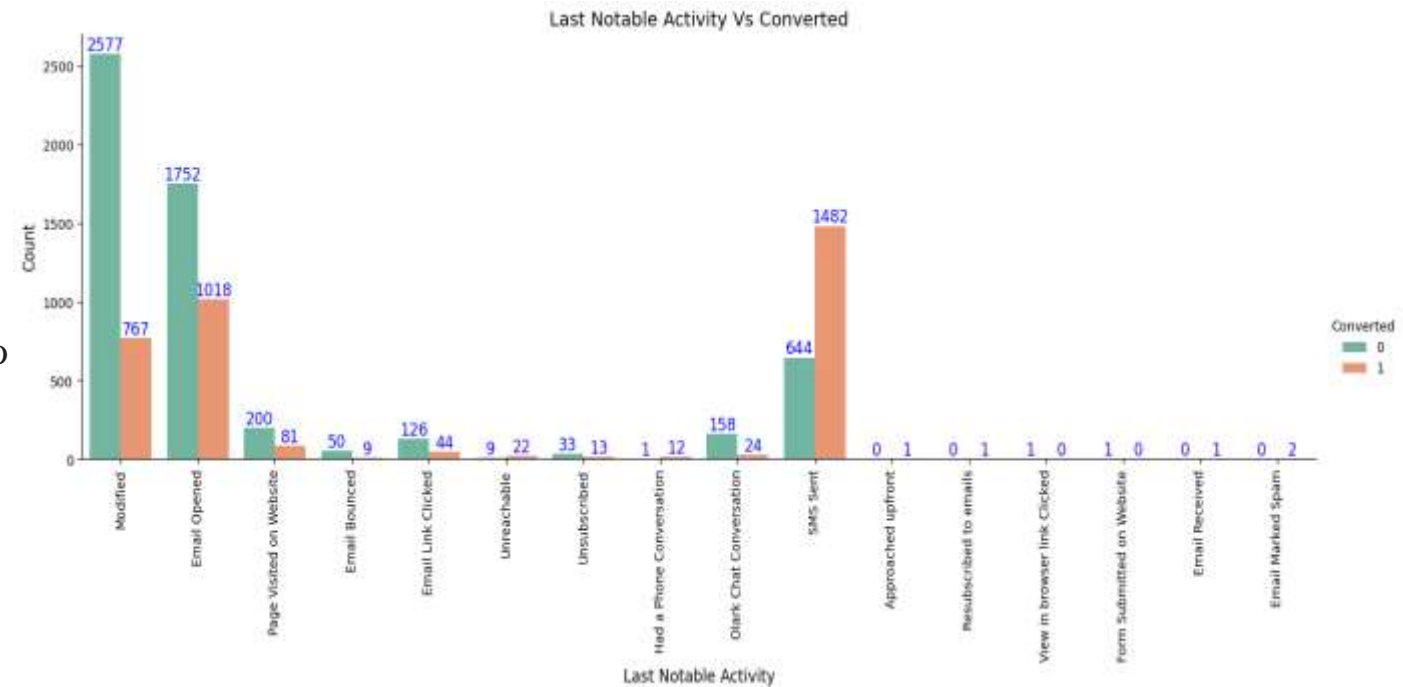
Results of Analysis

❖ Leads who are working professionals are more likely to convert into customers than other people.

❖ They have disposable income, which means they have money to spend on your products or services.

❖ They are more likely to be interested in your products or services because they are more likely to be in need of them as it will help them in their career growth.


What is your current occupation Vs Converted
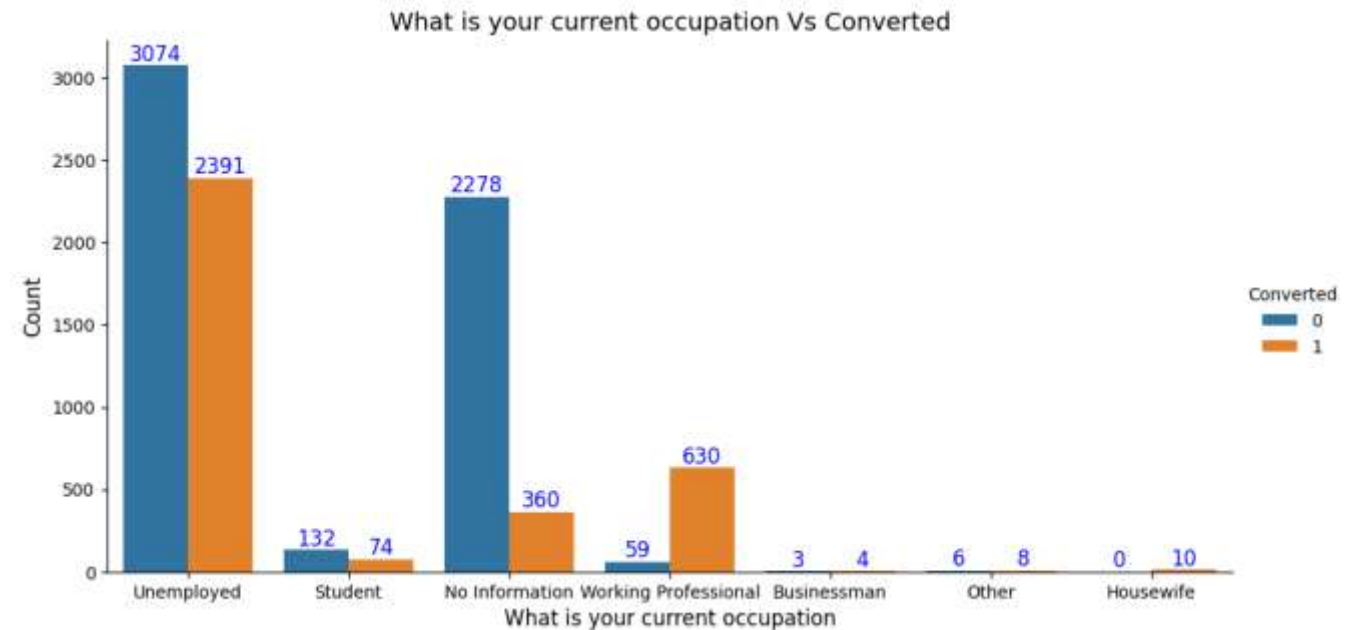
Results of Analysis

❖ The ROC curve area is a useful measure for comparing the performance of different models on the same dataset.

❖ An ROC curve area of 90% is considered to be excellent.

❖ The model is able to distinguish between the positive and negative classes very well

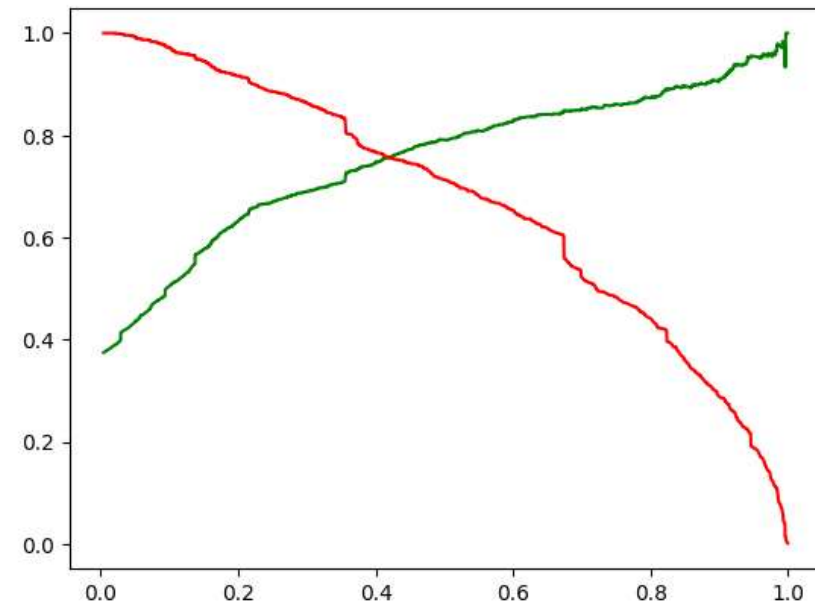Receiver operating characteristic example
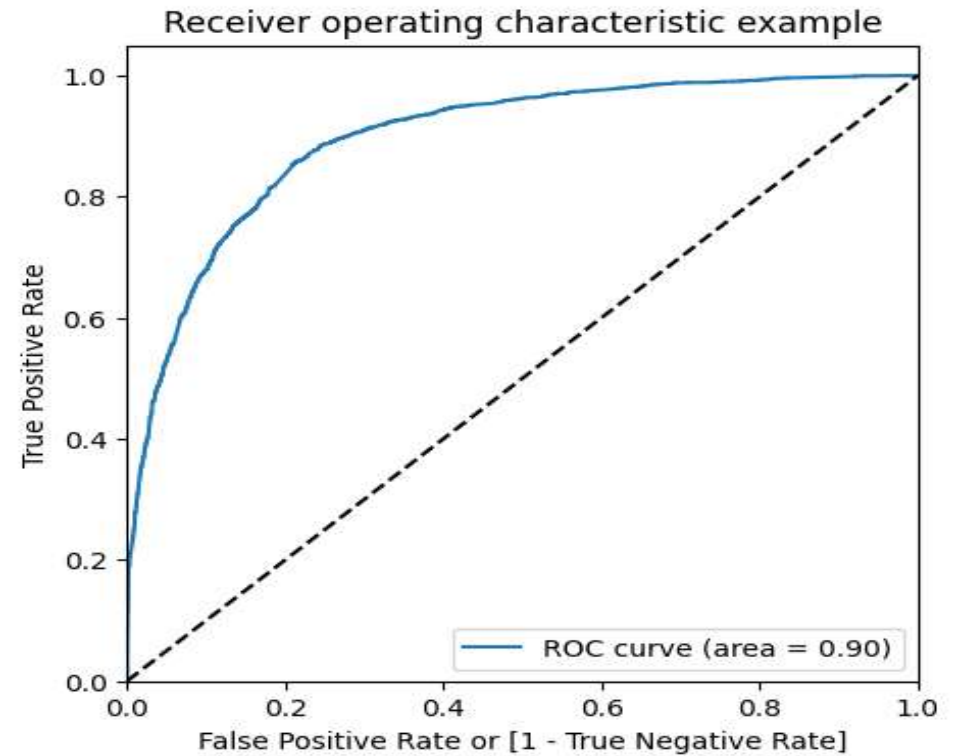
Results of Analysis

Trade-off between Precision and Recall (Train Set)
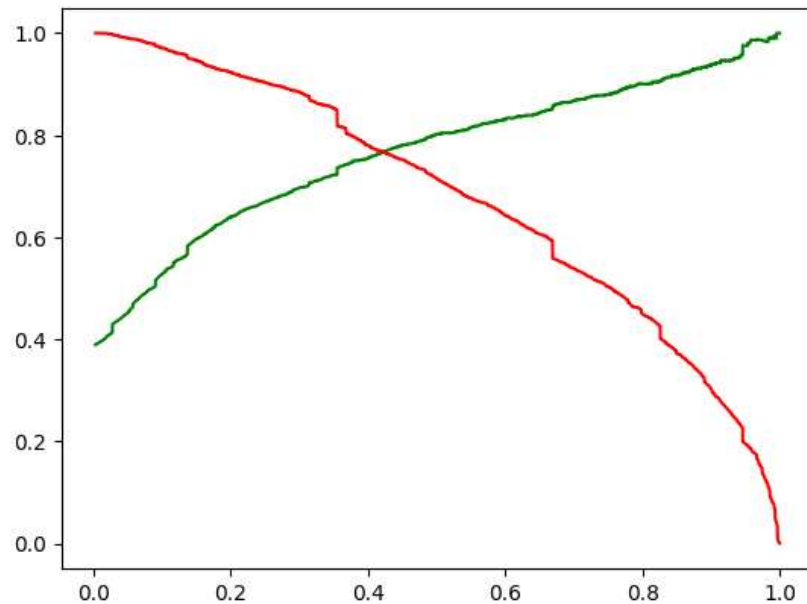
Trade-off between Precision and Recall (Test Set)

The trade-off between precision and recall on the train and test set is similar. The model is not overfitting the training data and it is able to generalize well to new data which is important for a model to be useful in the real world.

Conclusion

In conclusion, the data analysis reveals several key patterns that influence lead conversion for X Education, the online education company.

Optimizing X Education's lead conversion involves focusing on specific strategies:

1. Website Engagement: Improve website content and user experience to retain visitor interest and raise conversion rates.

2. Lead Form: Manage and respond promptly to leads from the lead form, as these exhibit higher conversion potential.

3. Phone Conversations: Engage in personalized phone conversations to address queries and build rapport, enhancing conversion odds.

4. "Do Not Call" Leads: Respect communication preferences; consider alternate methods like email or SMS for better outcomes.

5. Initial Interest: Leads showing upfront interest might benefit from targeted follow-ups instead of direct calls.

6. Demographic Tailoring: Understand diverse demographics to tailor communication, especially for segments like housewives.

7. Repeat Leads: Prioritize previously converted leads; use automated methods like SMS or email to foster further conversions.

Implementing these strategies could significantly boost X Education's conversion rates.