

Lead Scoring Case Study



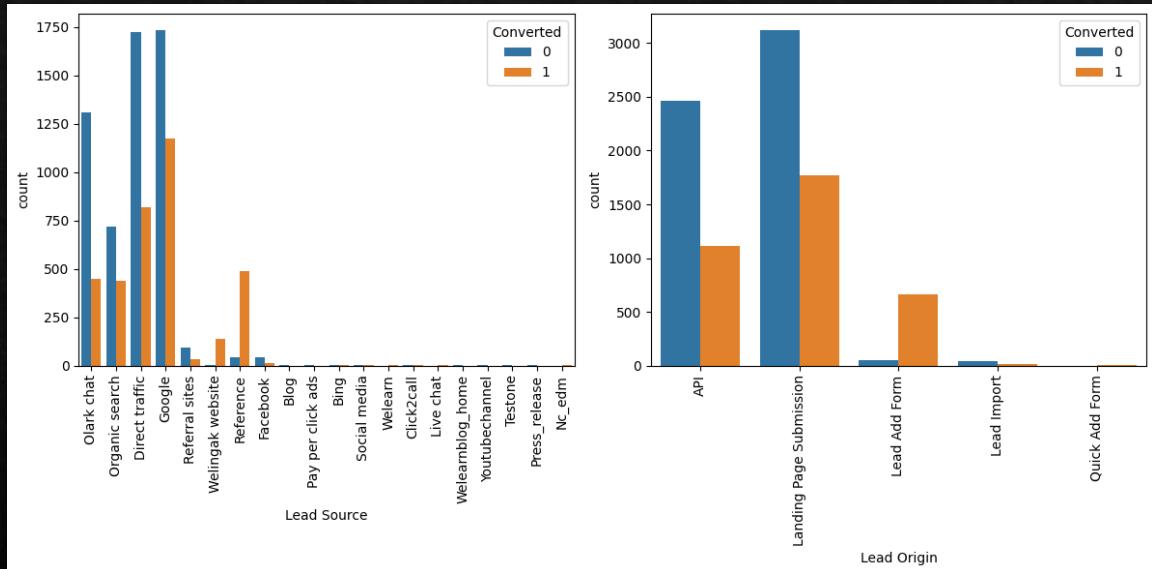
Introduction

- ❖ About X Education:
 - ❖ Industry-aligned online education companies offer courses that are relevant to the students.
 - ❖ Website visits, forms, and referring people can generate leads.
- ❖ Problem Statement:
 - ❖ Low lead conversion rate around 30%.
 - ❖ Even high-potential leads are not attracting the proper focus of the sales team.
 - ❖ An efficient lead scoring model is required to prioritize ‘Hot Leads’.
- ❖ Objective of the Case Study:
 - ❖ A logistic regression model is used to predict the lead conversion probability.
 - ❖ It helps to assign a lead score (0-100) to select promising leads for the sales team to concentrate on.
 - ❖ Improve the conversion rate above 80%.

Dataset and Solution Approach

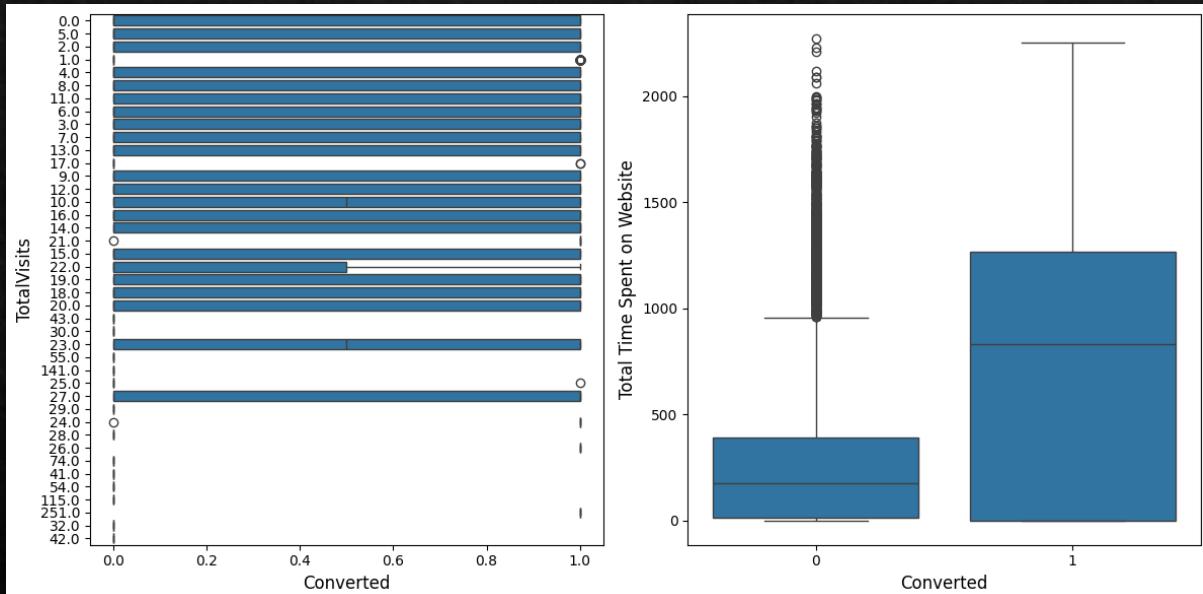
- ❖ Dataset Overview:
 - ❖ The dataset has approximately 9,000 records, about lead source, website activities, engagement.
 - ❖ Contains numerical & categorical features influencing the conversion of leads.
- ❖ Presentation Structure:
 - ❖ Data Analysis & Cleaning
 - ❖ Feature Selection & Model Building
 - ❖ Key Findings & Business Insights
 - ❖ Strategic Recommendations

Dataset Exploration



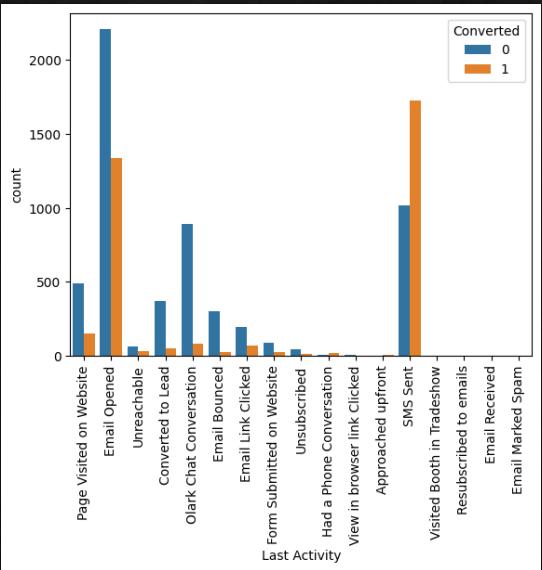
- ❖ Lead Source & Lead Origin Analysis is the focus of this work.
- ❖ The lead generation is highest from Google and Direct Traffic, but their conversion rate is not the highest.
- ❖ Reference and leads from the Welingak Website have the highest conversion rates.
- ❖ They have a fair number of leads from API and Landing Page Submission, but their conversion rate is still 30%.
- ❖ Its high conversion rate makes it a valuable source, and it generates few leads, which makes it a good source.
- ❖ The volume and conversion for lead import are low and can be deprioritized.
- ❖ Focus on boosting conversion rates on API and Landing Page Submission and increasing leads from Lead Add Form to enhance efficiency.

Dataset Exploration

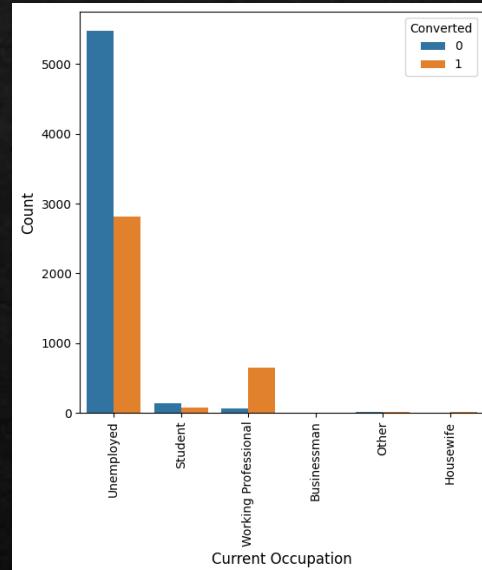


- ❖ The more time a user spends on the website, the higher the chances they will be converted.
- ❖ So that the time of the Users on websites can be increased, some features are added which makes the website more attractive for User.

Dataset Exploration



- ❖ There is a high count of 'Email Opened' and the maximum conversion rate for 'SMS sent'.
- ❖ Increase the conversion of those whose last activity was 'Email Opened', call to these leads to increase their count as well as increase the number of leads which have last activity as 'SMS sent'.



- ❖ High conversion rate and more number of Unemployed leads as compared to any other category.
- ❖ Since it is easy to reach out to them via different social sites, this should help increase the number of working professional leads and thereby increase the overall conversion rate.

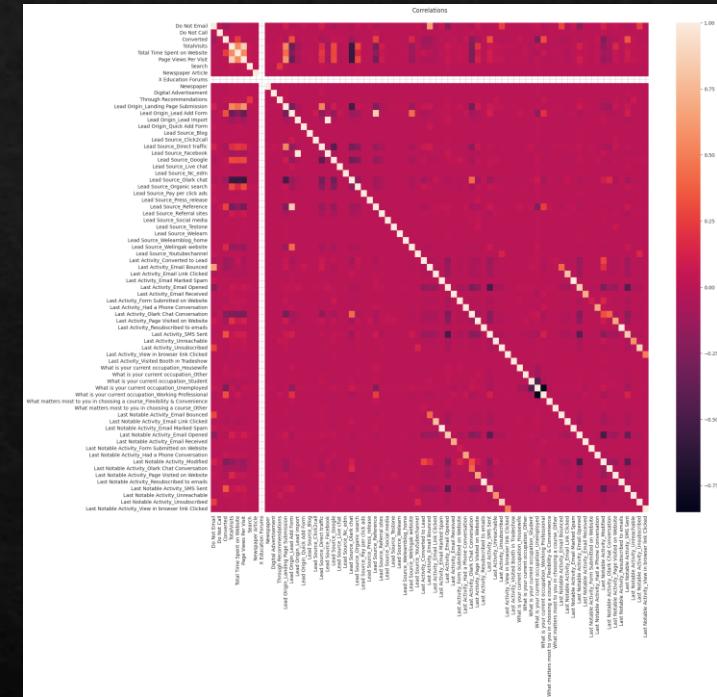
Data Preparation

```
# Splitting the dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.70,
                                                    random_state=100)

0.4.2 Feature Standardization

scaler = StandardScaler()

# Scaling the Total Time Spent on Website for the easy analysis.
X_train[['Total Time Spent on Website']] = scaler.fit_transform(X_train[['Total
Time Spent on Website']])
```



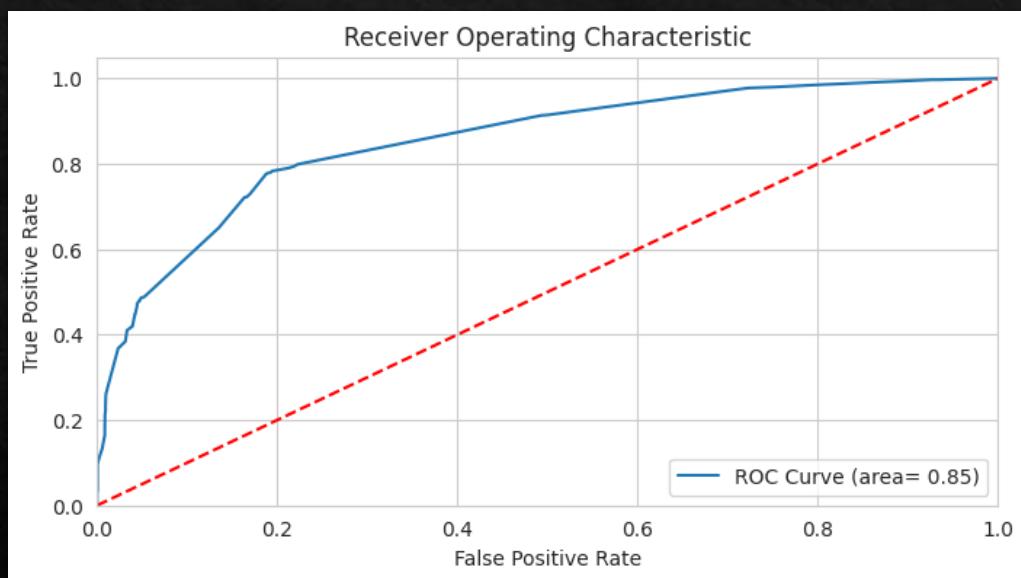
- ❖ This was split into train and test 70-30.
- ❖ A standard scaler was applied to features.
- ❖ An effect of variables was observed in multicollinearity, and they were removed.

Reviewing the Model

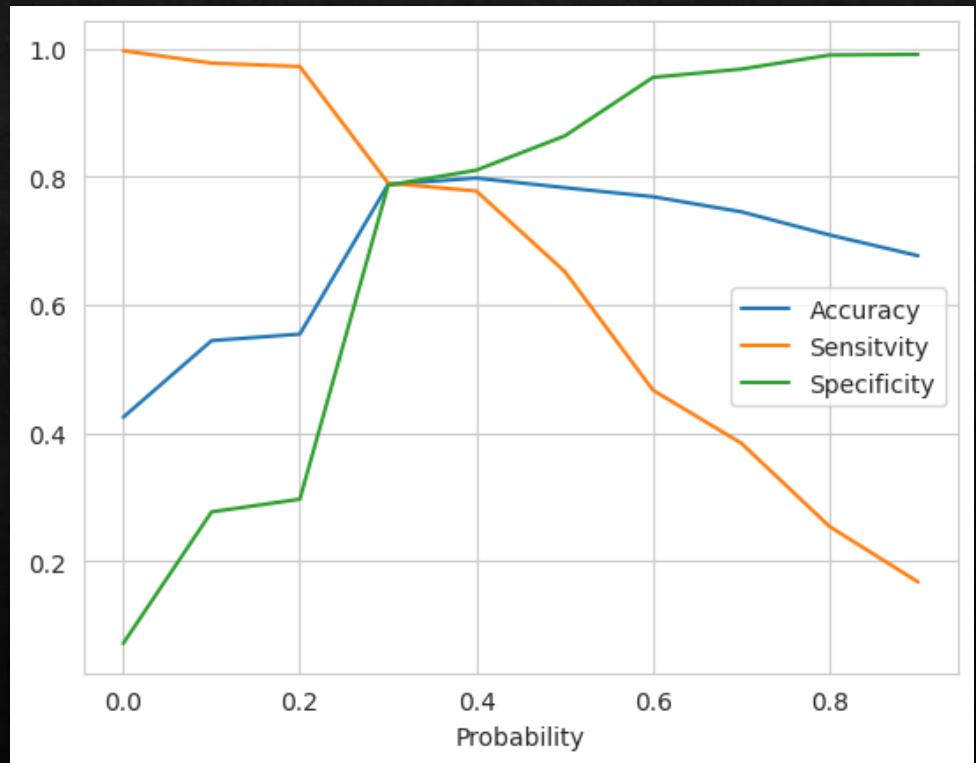
Dep. Variable:	Converted	No. Observations:	6363
Model:	GLM	Df Residuals:	6299
Model Family:	Gaussian	Df Model:	63
Link Function:	Identity	Scale:	0.14609
Method:	IRLS	Log-Likelihood:	-2876.8
Date:	Mon, 17 Feb 2025	Deviance:	920.22
Time:	10:50:36	Pearson chi2:	920.
No. Iterations:	3	Pseudo R-squ. (CS):	0.4649
Covariance Type:	nonrobust		

- ❖ 4 models were built to compare performances.
- ❖ The p-values of many attributes are insignificant, and the analysis attempts the RFE to eliminate features.
- ❖ Some of the features have high p-values. Removing them is necessary to obtain the desired value.

Evaluating the Model (Training Data)



- ◆ The model is good at predicting the curve area, and it's closer to the left border than the right border. It is getting the accuracy up to 85%.



- ◆ The perfect probability cutoff was 0.3.

Evaluating the Model (Training Data)

```
Accuracy score in predicting train dataset : 0.7883074021687884  
Precision score in predicting train dataset: 0.6960464272760246  
Recall score in predicting train dataset: 0.7903624382207578
```

- ❖ Train Dataset Performance:
 - ❖ Accuracy: 78.8%. This accuracy shows good performance for the model and could be useful in real life scenarios.
 - ❖ Sensitivity: 79.0% - While the model identifies most of the converting leads, some false positives are also being made.
 - ❖ The model is able to well identify conversions but a slightly lower precision suggests that the model misclassifies some non converting leads as possible conversions.

Testing the Model

```
Accuracy score in predicting test dataset : 0.7961129446277961  
Precision score in predicting test dataset: 0.6976024748646559  
Recall score in predicting test dataset: 0.845360824742268
```

- ❖ Slightly higher than the train set, the accuracy of this model is an accuracy of 79.6%. This means it also generalizes well.
 - ❖ 69.8% precision - The model does a good job of correctly identifying converting leads.
 - ❖ ROI: 3.78x - A higher ROI shows that there is less risk while getting more sales from the audience.
 - ❖ A higher recall on the test set does suggest the model flags good potential leads well but is at the math of precision: even though leads might still convert into sales, some that don't convert might still be contacted.
- ❖ Key Takeaways:
 - ❖ The model works as well on both datasets with minimum overfitting.
 - ❖ This leads to high recall and is, therefore, useful for aggressive lead conversion strategies.
 - ❖ Fine-tuning the probability threshold reduces the false positives for reducing precision.
 - ❖ This model is reliable in making the sales team focus on the prospects that have the most potential.

Predicting Lead Score

Converted	Converted_Probability	ID	Predicted	Lead Number	Lead Score
6906	1	0.953337	6906	1	596625
1873	0	0.075041	1873	0	641392
771	0	0.066225	771	0	652477
4495	0	0.106299	4495	0	616581
9061	1	0.750819	9061	1	580888

- ❖ If your conversion likelihood is high (an example would be a score of 95, 75), then these leads should be called first.
- ❖ If the lead score is a low one (8, 7, 11), the probability of conversion is low, so the sales team makes it a deprioritized lead.
- ❖ This scoring system allows for focusing on the heads that are more likely to convert, improving efficiency and the conversion rate.

Conclusion

- ❖ Effective Lead Scoring Model:
 - ❖ A logistic regression model to assign lead scores was developed so the sales team could better prioritize their high-potential leads.
 - ❖ This model had an accuracy of 79.6% and a great recall (or 84.5%), so most conversion leads were identified.
- ❖ Key Findings:
 - ❖ Lead Source and Lead Origin Matters: Google, Direct Traffic, Reference, and Welingk Website (lead source) has the highest impact on the conversion.
 - ❖ A Lot of Time Spent on Website is Important: Leads who spend more time on the website have a significantly higher possibility to convert.
 - ❖ The lead score allows the sales team to concentrate on the "hot leads" (high scores) and not on low-scoring leads (cold leads).
- ❖ Business Impact:
 - ❖ Lead conversion strategies that are more effective and efficient, so one is wasting sales effort.
 - ❖ With higher conversion rates, X Education is bringing itself closer to its 80% conversion target.
 - ❖ Allocating time and effort to higher producing leads, improving resource utilization to have the right amount of solved leads in the pipeline at one point in time.

Recommendations

- ❖ Put High Scoring leads first for faster conversions
 - ❖ The leads with a lead score above 75 should receive most of the focus sales because they have a higher conversion possibility.
 - ❖ Get sales reps assigned to chase down these leads as fast as possible.
- ❖ Improve Engagement for Mid-Scoring Leads (40-75)
 - ❖ These range leads hold loads of potential but may take a bit more work.
 - ❖ To get closer to the conversion, you can do targeted email campaigns, personalized follow-ups, and limited-time offers.
- ❖ Cutting down Efforts on Low Scoring Leads (40% or less)
 - ❖ If a lead scores low, the person will unlikely convert, so try not to follow up with him excessively.
 - ❖ To keep engagement while not spending high costs, think about mail marketing emails as opposed to sales calls.

Recommendations

- ❖ Get more out of Lead Quality from Key Sources
 - ❖ Increase marketing investment on Google, Reference, and Welingak Website as they have the highest conversions.
 - ❖ Utilizing these leads for possible improvements to engaging API and Landing Page Submission leads, as both have a high volume but low conversion rates.
- ❖ Leverage Website Behavior Insights
 - ❖ Those leads who spend more time on the website are more likely to convert.
 - ❖ Evaluate the successful and unsuccessful methods of interactive content, testimonials, and demo videos to encourage further website engagement over a longer period.
- ❖ Optimize Sales Call Strategy
 - ❖ When there is a high demand, try to send aggressive follow-ups for the predicted conversions (lead score over 50).
 - ❖ When demand is low, scale back on accidental phone calls and put effort into building long-term sales leads.

Thank You.