



Lead Scoring Case Study

- MANU SMRITI

Problem Statement

Develop an efficient model to identify potential leads who are likely to enroll in an online course offered by X Education.

Analysis Approach

Data Cleaning: Load the dataset, understand its structure, and clean the data.

Exploratory Data Analysis (EDA): Check for class imbalance and perform univariate and bivariate analyses.

Data Preparation: Create dummy variables, split the data into training and test sets, and apply feature scaling.

Model Building: Use Recursive Feature Elimination (RFE) for selecting top features, manually refine feature selection, and finalize the model.

Predictions on Test Data: Compare metrics between the training and test sets and assign a lead score.

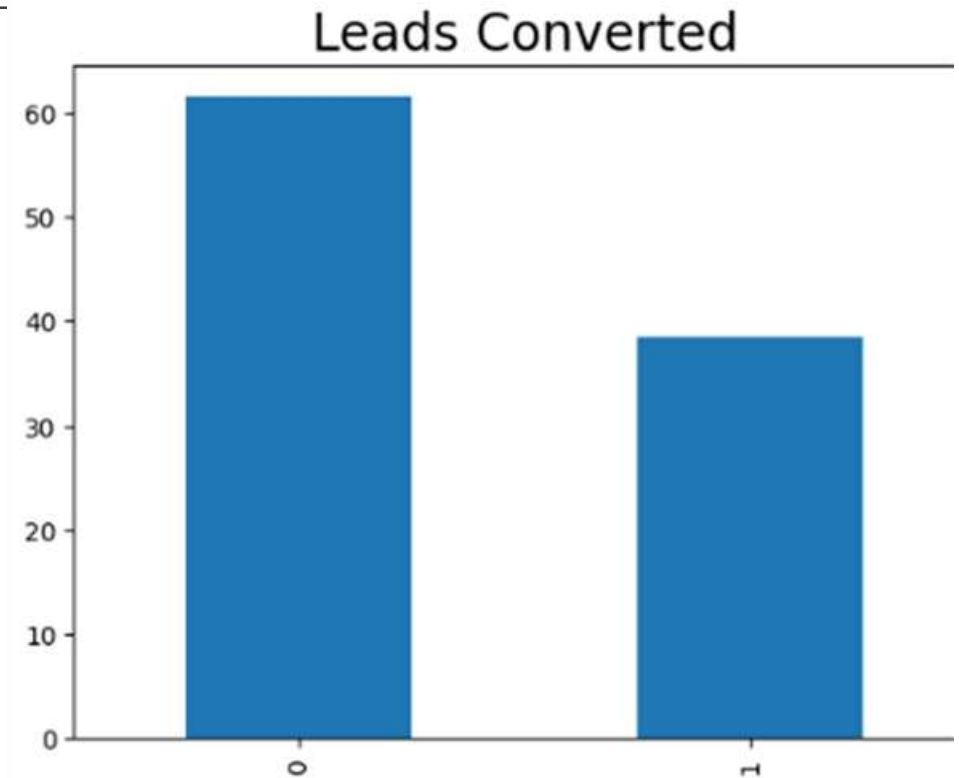
Recommendations: Identify and suggest the top features to focus on for higher conversion rates.

Data Cleaning

1. The "Select" level indicates null values for some categorical variables where customers did not choose any option.
2. Columns with more than 35% null values were removed.
3. Features that do not provide useful insights, such as "country" and "what matters most to you in choosing a course," were dropped.
4. Columns not useful for modeling, like "Prospect ID" and "Lead Number," or those with only one category of response, were eliminated.
5. Null values in categorical columns were replaced with the most frequently occurring values.
6. Null values in numerical columns were replaced with the median values.
7. Outliers in numerical columns were addressed using scaling techniques.
8. Binary categorical variables were mapped to appropriate values.

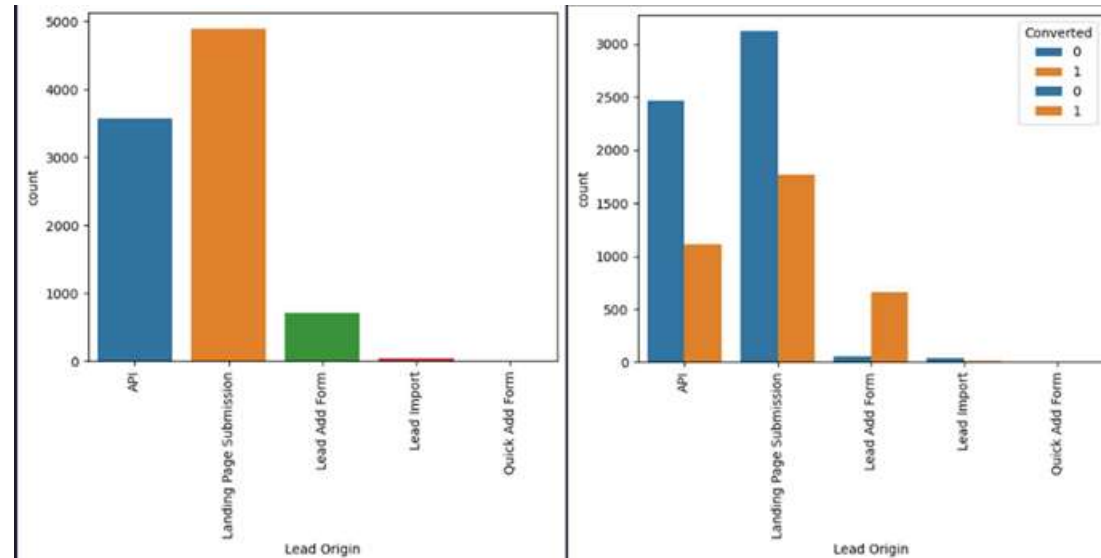
EDA

1. Target variable data is imbalanced
2. The conversion rate is 38.5%, meaning only 38.5% of the people converted.
3. Conversely, 61.5% of the people did not convert into leads.



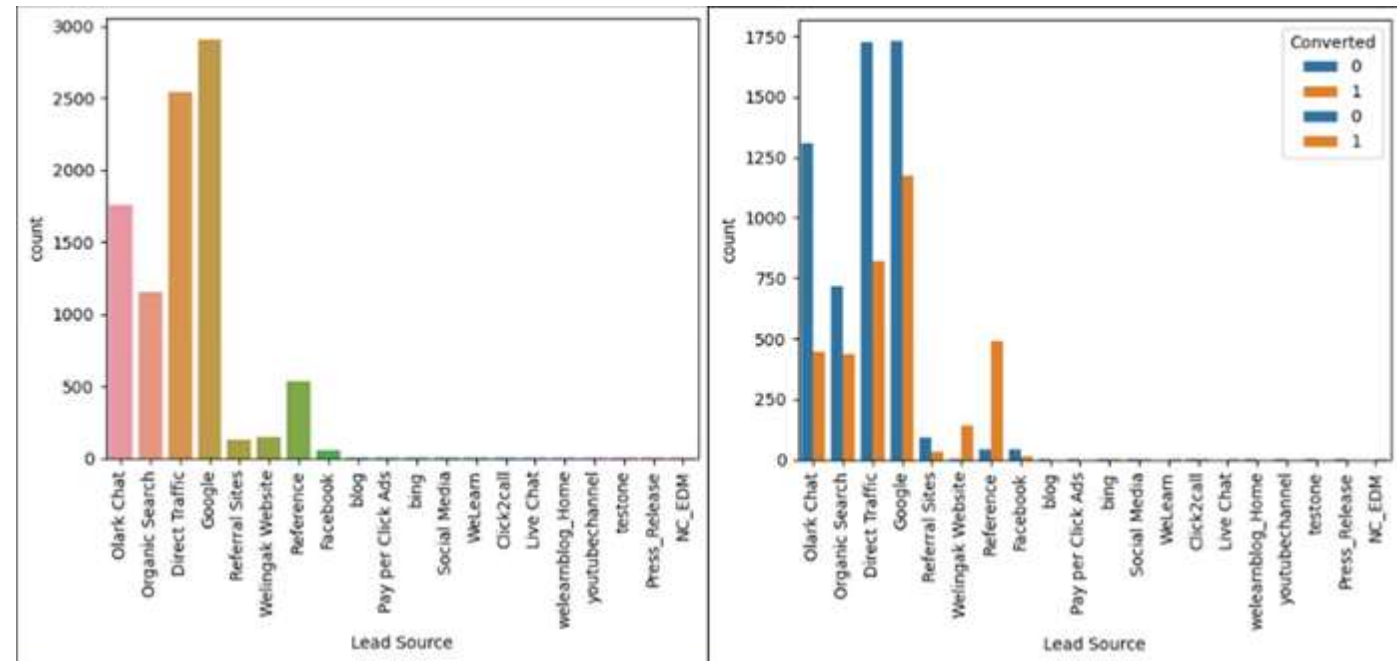
Lead Origin

- The majority of customers were identified through 'Landing Page Submission'.



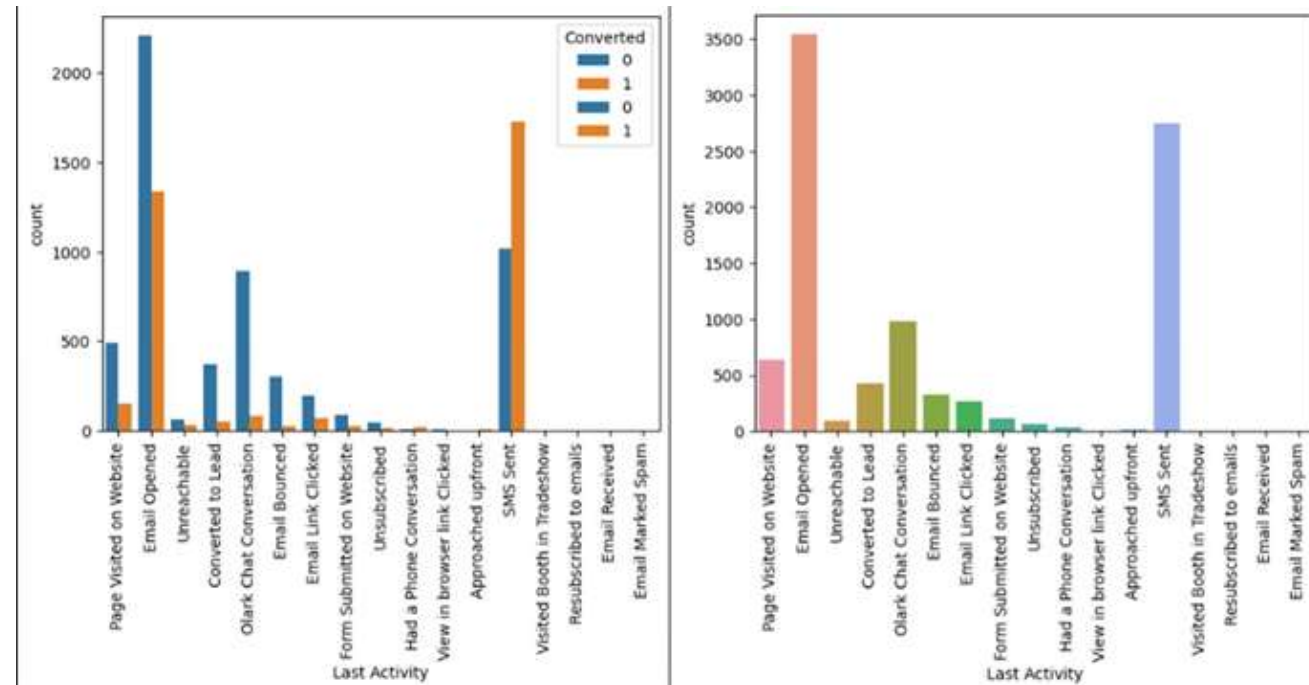
Lead Source

Most number of lead sources are from Google and Direct Traffic



Last Activity

Most customers' last activities were 'SMS sent' and 'Email opened'.



Data Preparation

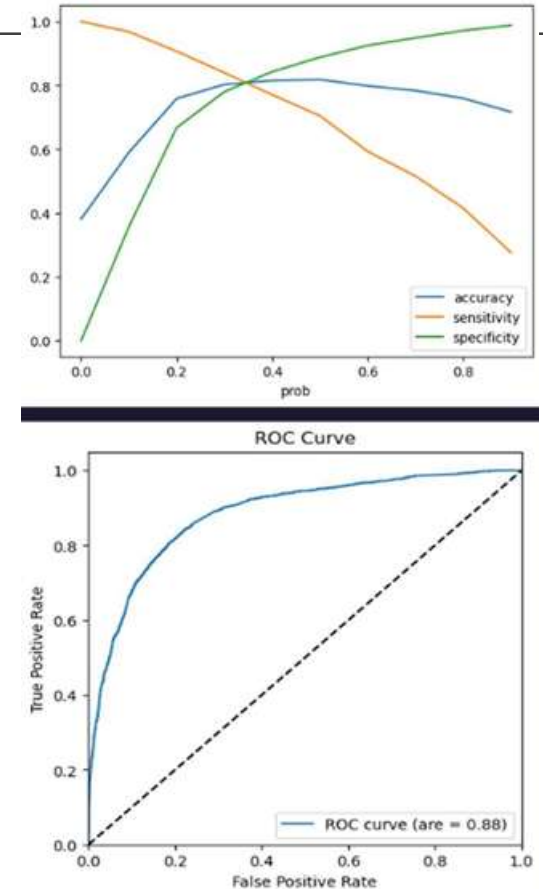
1. Binary categorical columns were already mapped to 1 and 0 in previous steps.
2. Created dummy variables for categorical features with more than two levels.
3. Split the data into training and test sets.
4. Applied min-max scaling to the features for normalization.

Model Building

1. Used Recursive Feature Elimination (RFE) to select the top 20 features.
2. Manually tuned the model to finalize 16 features.
3. Dropped features with p-values higher than 0.05 one by one, rebuilding the model each time.
4. Ensured the Variance Inflation Factor (VIF) for all features is within an acceptable range (<5).
5. The final model is referred to as Model 6.

Model Evaluation

1. The CEO of X Education has set a target lead conversion rate of around 80%.
2. Sensitivity is the key metric to identify 80% of potential leads and meet this target.
3. From the accuracy, sensitivity, and specificity curve, a probability cutoff of 0.35 is chosen to distinguish between leads who will convert and those who will not.
4. With this cutoff value, a sensitivity of 81% is achieved.
5. The Area Under the Curve (AUC) for this model is approximately 0.88, indicating strong performance.
6. These evaluation metrics are based on the training data. When the model is applied to the test dataset, it achieves similar AUC and sensitivity values, demonstrating its reliability.



Recommendations

1. Increasing lead conversion is paramount for the growth and success of X Education. To achieve this goal, we've built a regression model to pinpoint the most influential factors affecting lead conversion.
2. Prioritize features with the highest positive coefficients in our marketing and sales strategies to enhance lead conversion rates.
3. Identify features with negative coefficients as areas needing improvement to optimize lead conversion efforts further.

	coef
const	-0.7988
Do Not Email	-1.3394
TotalVisits	9.0912
Total Time Spent on Website	4.5516
Page Views Per Visit	-3.8015
Lead Origin_Lead Add Form	3.6853
Lead Source_Olark Chat	1.0670
Lead Source_Welingak Website	1.9524
Last Activity_Converted to Lead	-1.0660
Last Activity_Email Bounced	-1.1244
Last Activity_Olark Chat Conversation	-1.2396
What is your current occupation_Working Professional	2.8023
Last Notable Activity_Email Link Clicked	-1.9036
Last Notable Activity_Email Opened	-1.3407
Last Notable Activity_Modified	-1.6887
Last Notable Activity_Olark Chat Conversation	-1.4766
Last Notable Activity_Page Visited on Website	-1.8566

To Increase Lead Conversion Rates

- Focus on features with positive coefficients to optimize marketing strategies.
- Develop targeted campaigns to attract high-quality leads from top-performing sources.
- Specifically, total visits and total time spent on the website show very high positive coefficients. Increasing advertising on platforms like Welingak could significantly increase lead acquisition.
- Implement incentives for successful referrals that convert into leads to encourage more referrals.
- Target working professionals, who not only exhibit high conversion rates but also have the financial capacity to afford higher fees.
- Working professionals are actively seeking opportunities to enhance their skills through online courses, presenting a promising opportunity for X Education.

To identify areas of improvement

- Evaluate features with negative coefficients thoroughly.
- Assess these features to determine areas where improvements can be made.

Thank You!!