LEAD SCORE CASE STUDY

Logistic Regression Model

By Jagruti Pimpale

PROBLEM STATEMENT

X Education, an online education company, seeks to boost its lead conversion rate by identifying and prioritizing "Hot Leads" with the highest potential to convert into paying customers.

They engage leads through calls and emails to nurture their interest. By implementing a lead scoring model, the company aims to assign scores to leads based on their likelihood of conversion.

The objective is to focus efforts on leads with higher scores, increasing the overall lead conversion rate. X Education targets an ambitious lead conversion rate of around 80%.

BUSINESS GOAL

There are quite a few goals for this case study.

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.

There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

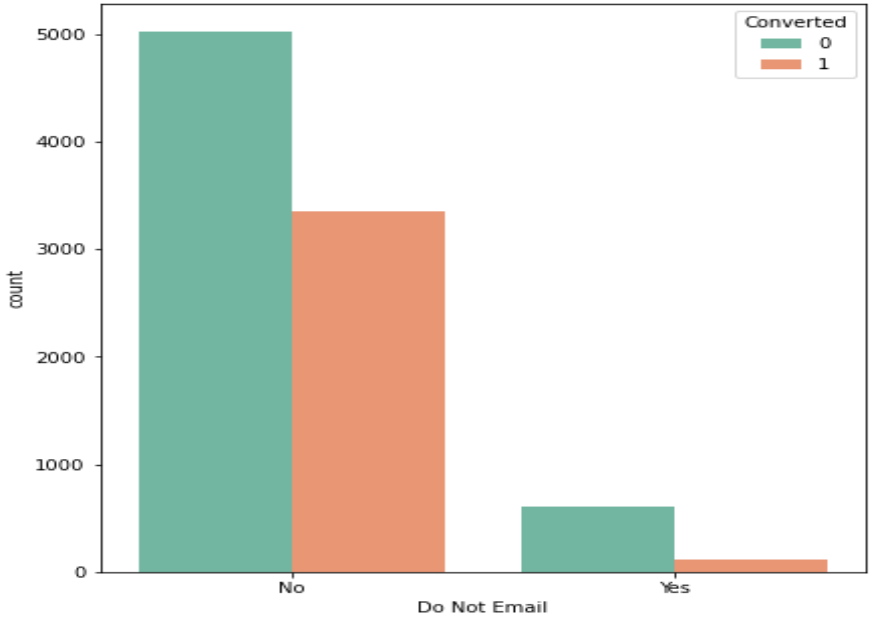
STRATEGY

- Import the acquired data.
- Clean and prepare the data for analysis.
- Perform exploratory data analysis to identify the most influential attributes for lead conversion.
- Scale the features for model building.
- Prepare the data for model training.
- Build a logistic regression model.
- Assign a lead score to each lead based on the model.
- Test the model's performance on the train set.
- Evaluate the model using various measures and metrics.
- Test the model on the test set.
- Measure the accuracy of the model and assess its performance using other relevant metrics.

DO NOT EMAIL VS COUNT

-Google searches have a higher conversion rate compared to other lead sources.

-This indicates that leads generated through Google searches are more likely to be converted into paying customers for X Education.

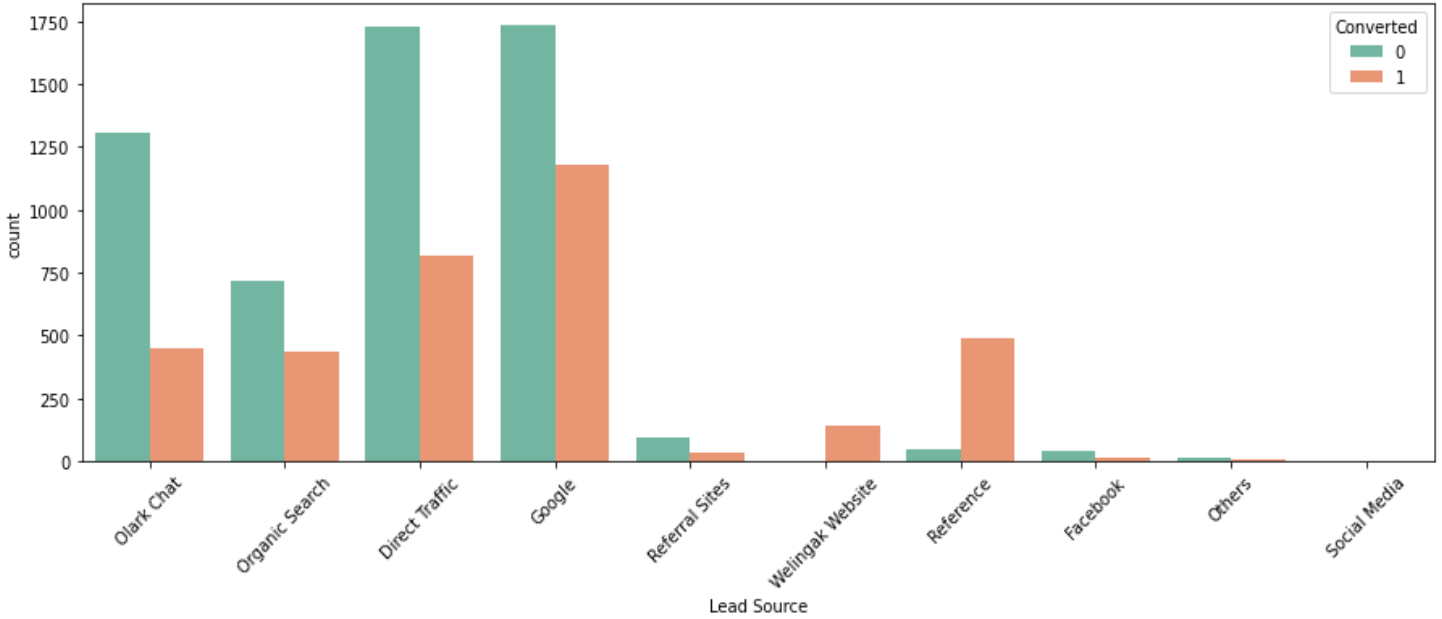


LEAD SOURCE VS COUNT

-Google searches have a higher conversion rate compared to other lead sources.

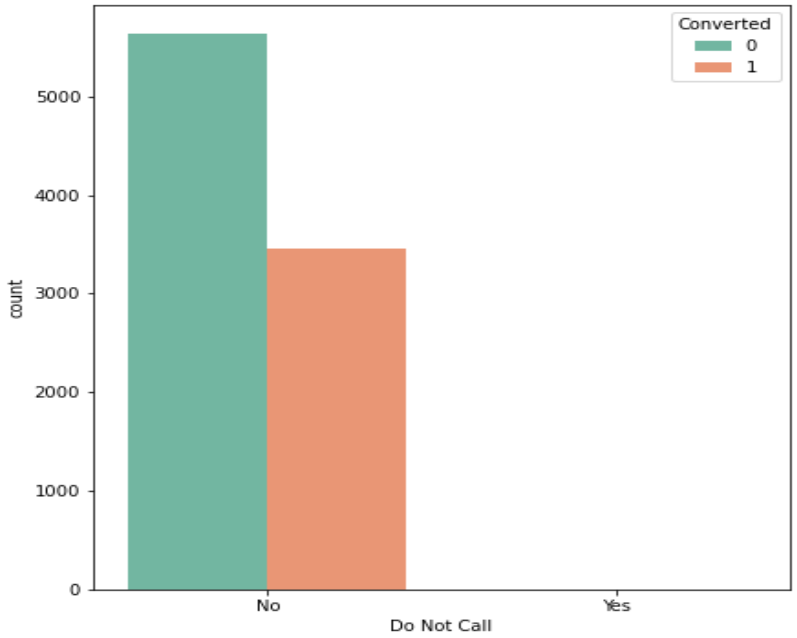
-References (referrals) also show a high conversion rate.

-Marketing efforts targeting Google searches and referrals have been successful in terms of lead conversion.



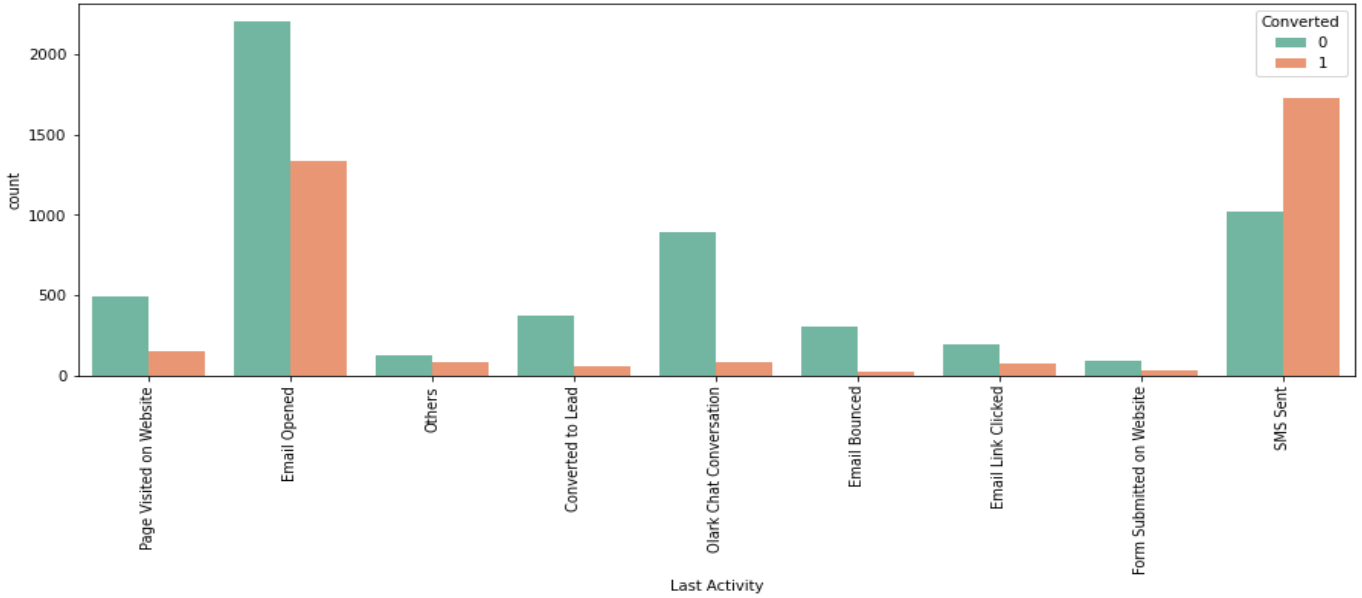
DO NOT CALL VS COUNT

This graph indicated most of leads prefer not to informed through phone.



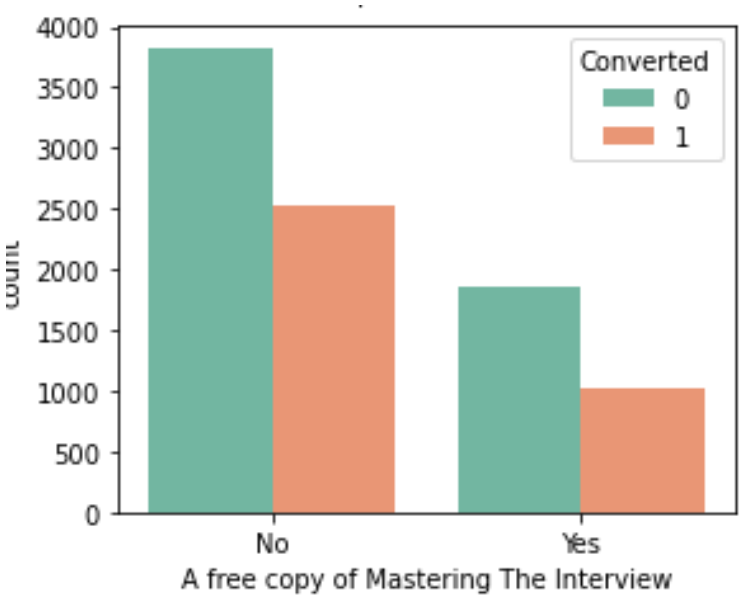
LAST ACTIVITY VS COUNT

According to the graph analysis, SMS has shown to be a promising method for generating higher confirmed leads. Additionally, email has also demonstrated high conversion rates.



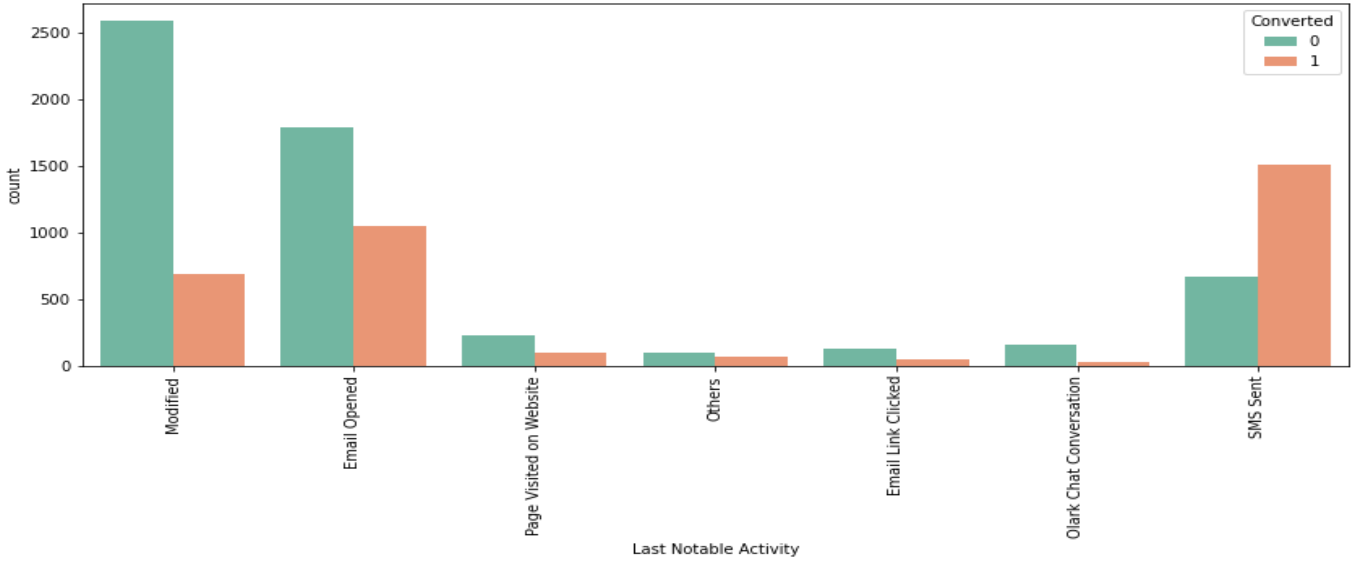
A FREE COPY OF MASTERING THE INTERVIEW VS COUNT

The data shows that leads generally prefer fewer copies of interviews compared to graphs. This indicates that providing less information through interviews is more appealing to leads.



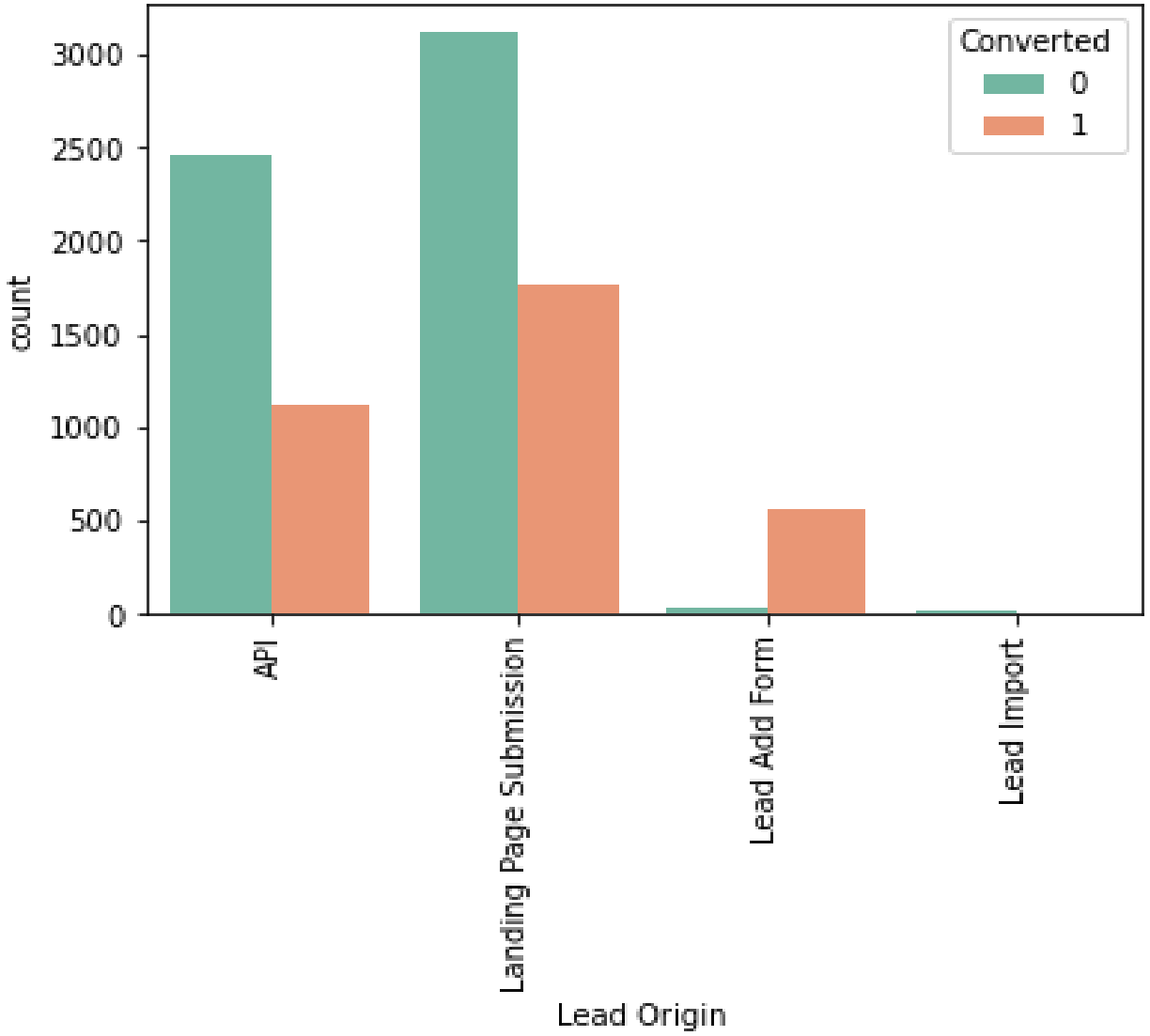
LAST NOTABLE ACIVITY VS COUNT

- SMS: Graph indicates significant lead conversion through SMS, highlighting its effectiveness as a communication channel.
- Emails: Graph shows considerable lead conversion through email communications, emphasizing the importance of email marketing strategies.



LEAD ORIGIN VS COUNT

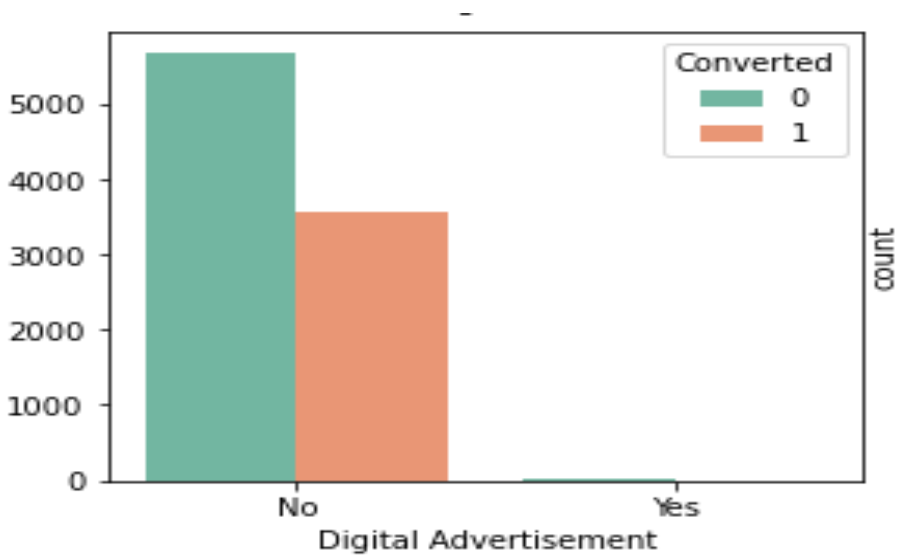
-The graph demonstrates that the landing page submission method is the most popular among leads, while the API, lead add form, and lead import methods have relatively lower usage.
-This information can help X Education optimize their lead generation process by focusing on the landing page submission method and potentially improving the other methods to increase lead engagement and conversion.



DIGITAL ADVERTISEMENTS VS COUNT

-According to the graph, digital advertisements do not appear to generate promising leads.

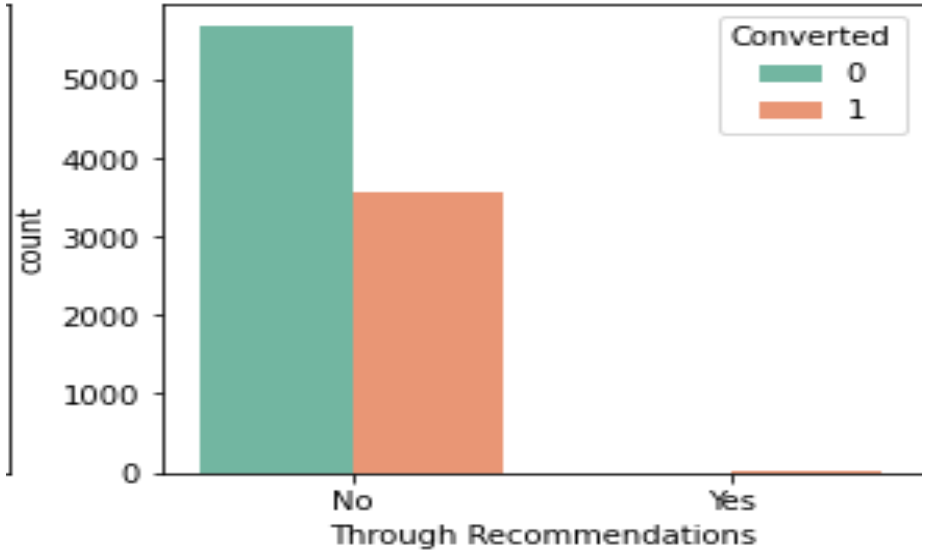
-The data suggests that the conversion rate or effectiveness of digital advertisements in attracting potential customers is relatively low compared to other lead generation methods.



THROUGH RECOMMENDATION VS COUNT

-Based on the graph, leads generated through recommendations do not seem to have a significant impact in terms of conversion.

-The data indicates that the effectiveness of recommendations in attracting potential customers and converting them into paying customers is relatively low compared to other lead generation methods.

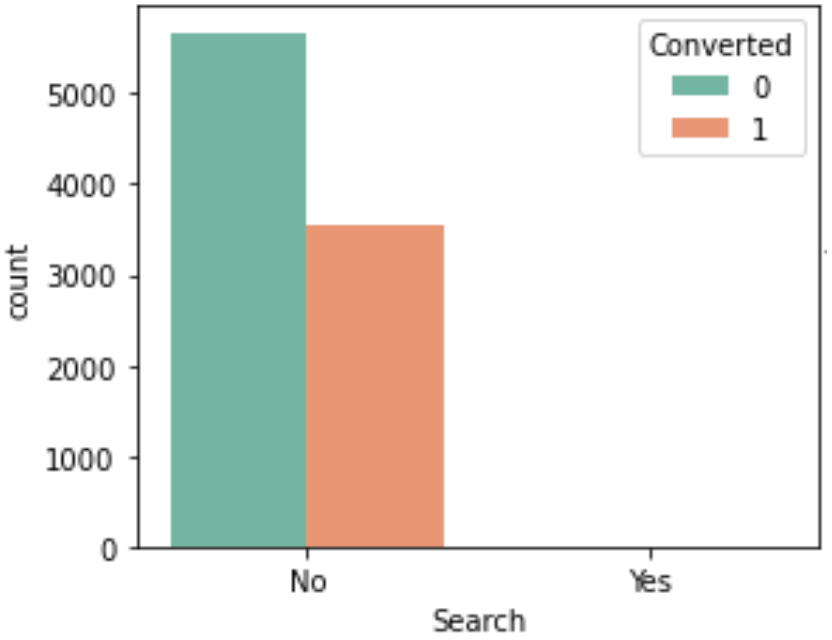


SEARCH VS COUNT

-Based on the graph comparing searches and their counts, it appears that searches are not a reliable source of leads.

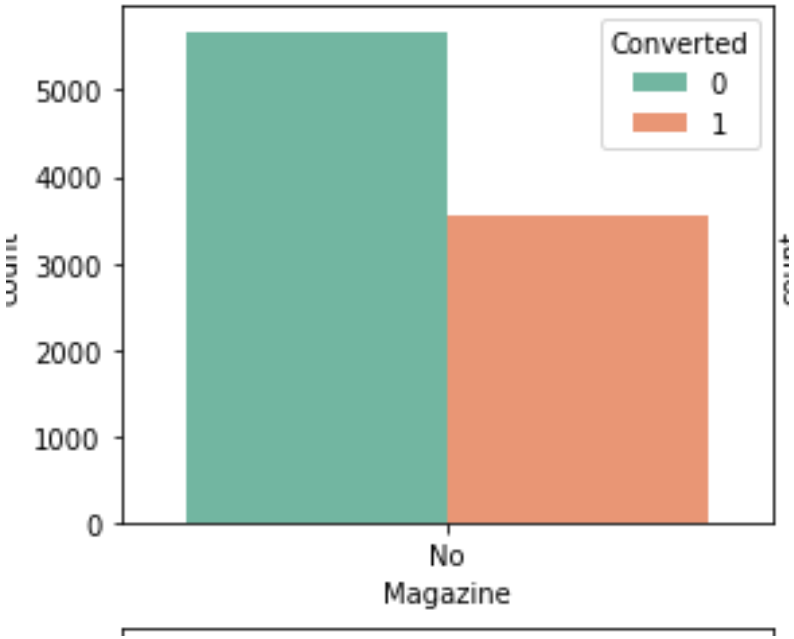
-The data shows that the number of searches is high, but the conversion rate is low, with zero conversions in some cases.

-This suggests that leads generated through searches may not result in successful conversions for X Education.



MAGAZINE VS COUNT

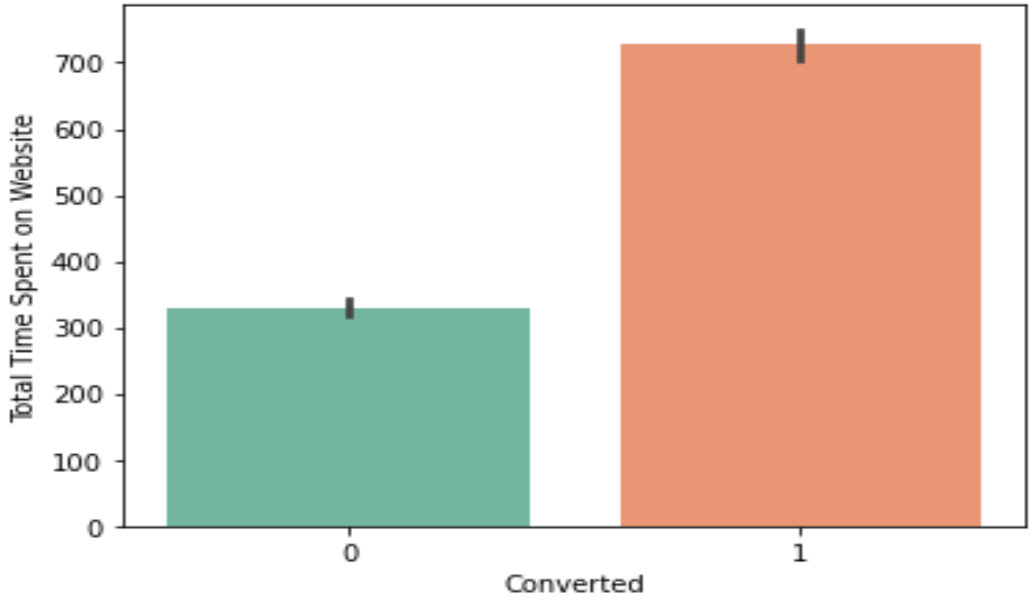
-Magazines do not demonstrate a higher conversion rate. The data indicates that there are no conversions



TOTAL TIME SPENT ON WEBSITES VS CONVERTED

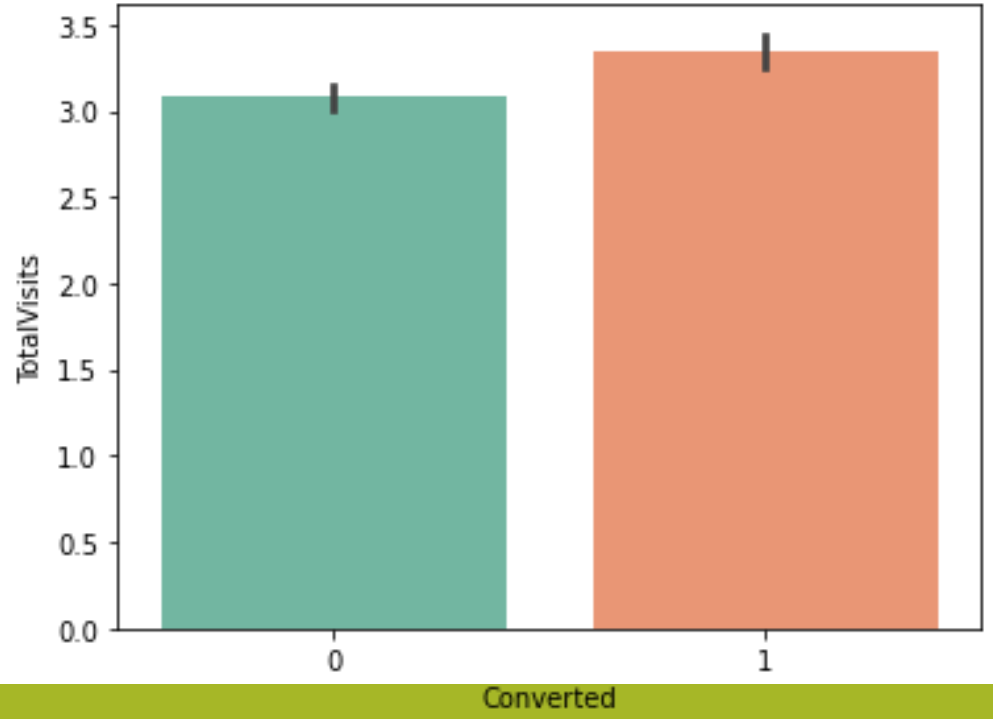
-According to the graph, a larger number of people are spending more time on the website compared to the average time.

-This indicates that these leads, who are spending more time on the website, have the potential to be promising leads for X Education.



TOTAL VISITS VS CONVERTED

-The graph suggests that leads with higher total visits have a slightly higher chance of being promising leads.



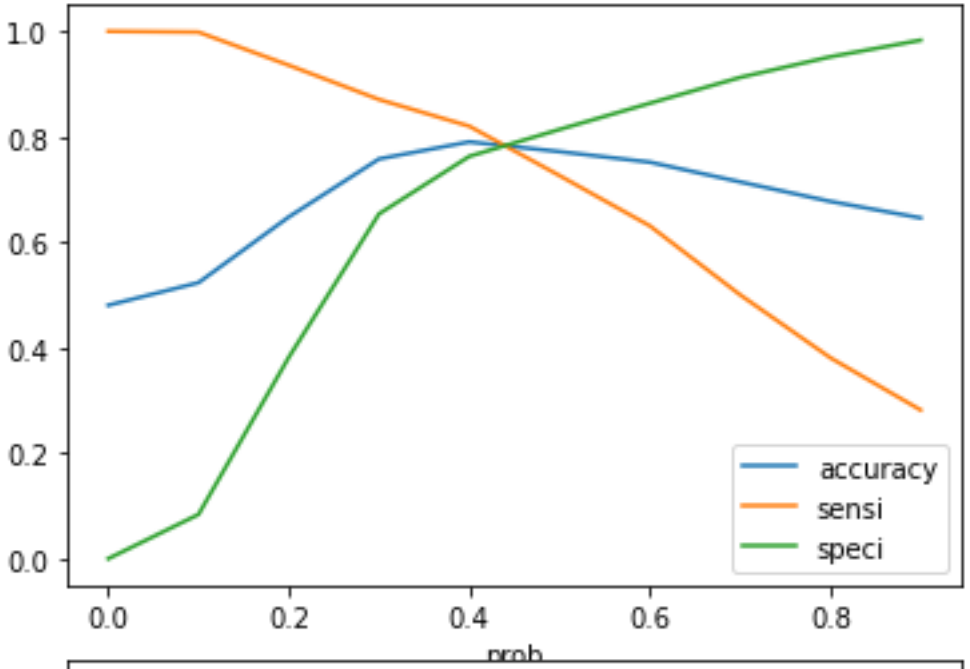
To build the model, the following steps were followed:

1. The dataset was split into a train set and a test set.
2. The variables in the train set were scaled to ensure consistency.
3. The first model was built using the train set.
4. Recursive Feature Elimination (RFE) was used to eliminate less relevant variables and refine the model.
5. The next model was built after eliminating variables based on high p-values.
6. The Variance Inflation Factor (VIF) was checked for all the remaining columns to assess multicollinearity.
7. Predictions were made using the train set.
8. The accuracy and other metrics were evaluated to assess the model's performance on the train set.
9. Predictions were also made using the test set.
10. Precision and recall analysis were conducted on the test set predictions to further evaluate the model's performance.

MODEL EVALUATION (TRAIN)

ACCURACY SENSITIVITY AND SPECIFICITY
- Accuracy : 77.05%
- Sensitivity :82.89%
- Specificity : 73.49%

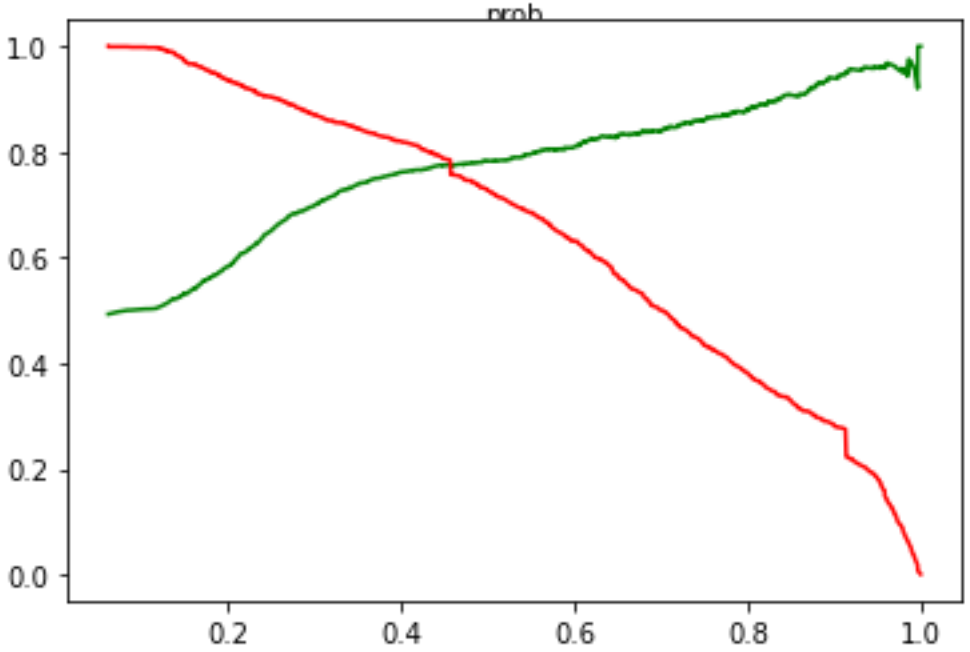
PRECISION AND RECALL
- 69.9% Precision
- 87% Recall



MODEL EVALUATION (TEST)

ACCURACY SENSITIVITY AND SPECIFICITY
- Accuracy : 77.52%
- Sensitivity :83.01%
- Specificity : 74.13%

PRECISION AND RECALL
- 72% Precision
- 89% Recall



In conclusion, the analysis of the data revealed the following insights:

1. Leads who spend more time than the average on the website have shown higher chances of becoming promising leads. Therefore, targeting and approaching these individuals can be beneficial in improving conversions.

2. SMS messages have demonstrated a significant impact on lead conversion. Incorporating SMS communication as part of the marketing strategy can be effective in increasing conversion rates.

3. Landing page submissions have proven to be a valuable source for generating leads. Paying attention to landing page design and optimization can help attract more leads.

4. Certain professions, such as marketing management and human resources management, have exhibited higher conversion rates. Focusing on individuals with these professional backgrounds can lead to more promising leads.

5. Referrals and incentives for referring leads have shown potential for higher conversion rates. Implementing referral programs and offering incentives can encourage existing customers to refer more leads, thereby increasing conversion opportunities.

6. Sending alert messages or providing relevant information has been associated with higher lead conversion rates. Utilizing this approach in communication with leads can be effective in increasing conversions.

Regarding the logistic regression model:

- The model achieved a high accuracy rate, close to 81%, indicating its ability to correctly predict conversions.
- The threshold for classification was determined using accuracy, sensitivity, specificity, precision, and recall curves.
- The model demonstrated 76% sensitivity and 83% specificity, further validating its accuracy in identifying promising leads.
- Overall, the model effectively identifies potential leads with a high likelihood of conversion, making it a reliable tool for X Education.

In summary, by leveraging the insights from the EDA and implementing the logistic regression model, X Education can optimize its lead generation and conversion strategies, resulting in higher conversion rates and business growth.