LEAD SCORE CASE STUDY

Logistic Regression Model

By Dipesh 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

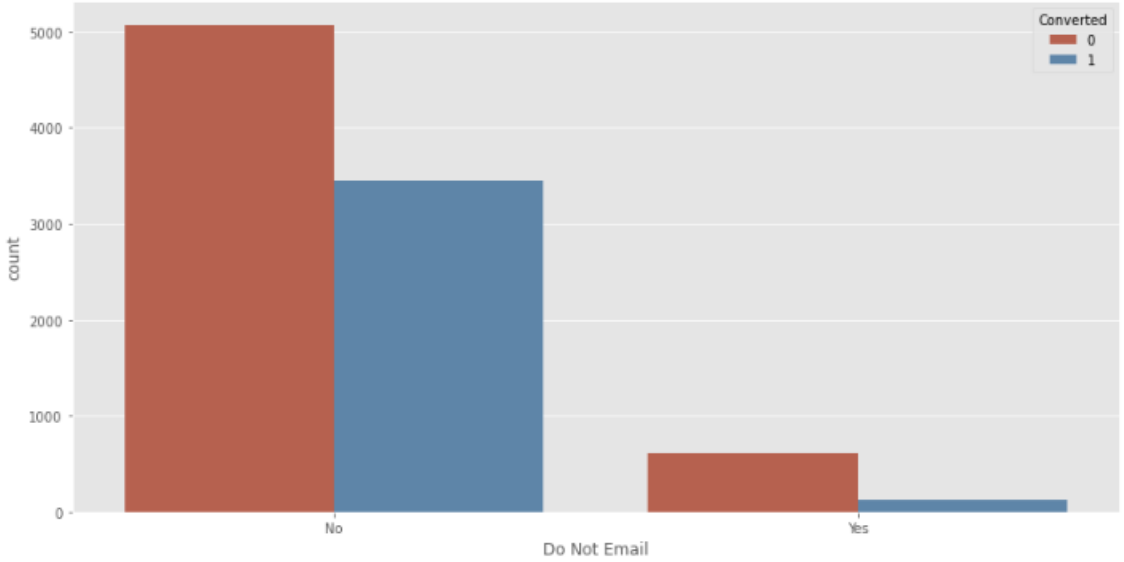
- There are quite a few goals for this case study.
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower 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.

Approach

- 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

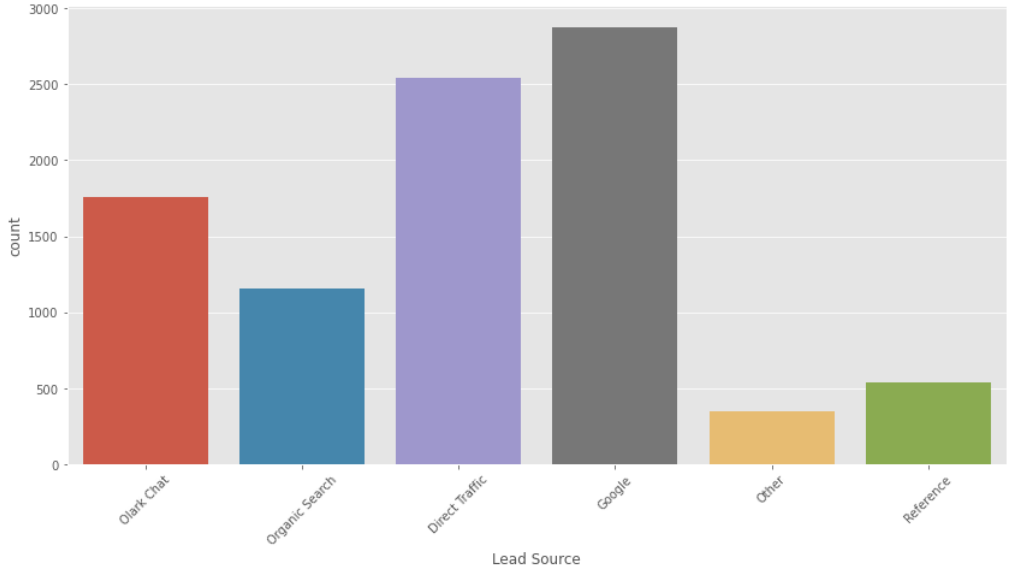
- The analysis reveals that Google searches exhibit a higher conversion rate compared to other lead sources.
- This implies that leads generated through Google searches have a higher likelihood of being converted into paying customers for X Education. Therefore, focusing on optimizing strategies to attract and nurture leads from Google searches can significantly improve the overall conversion rate and success of X Education's marketing efforts.



LEAD SOURCE VS COUNT

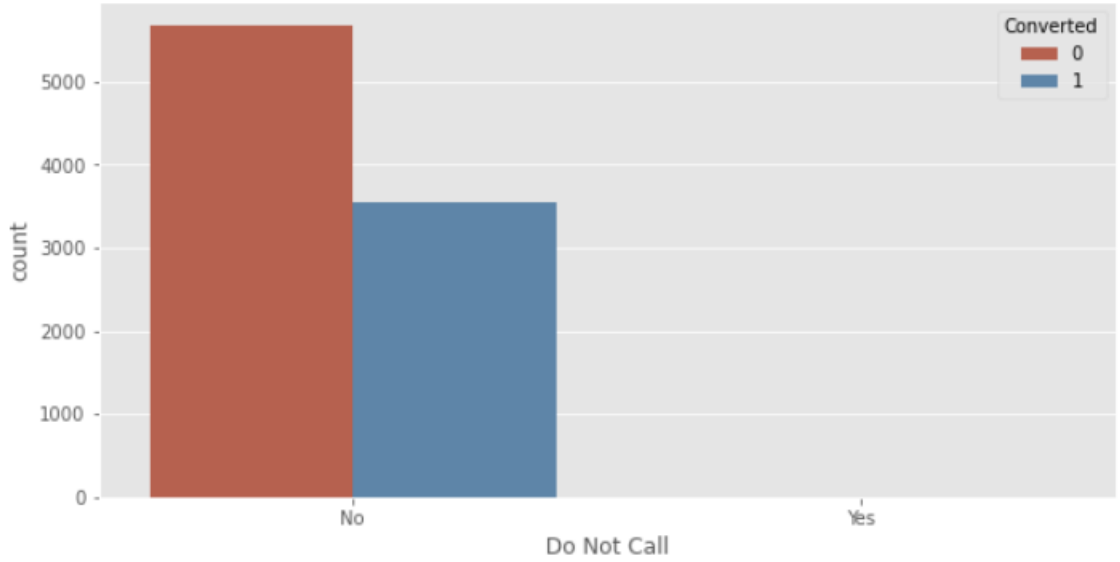
- The analysis indicates that Google searches have a higher conversion rate compared to other lead sources. Additionally, referrals (references) also exhibit a high conversion rate.

  These findings highlight the success of marketing efforts targeted at Google searches and referrals, as they have proven effective in converting leads into paying customers for X Education. By further optimizing strategies for these lead sources, X Education can enhance its lead generation and overall conversion rates.



DO NOT CALL VS COUNT

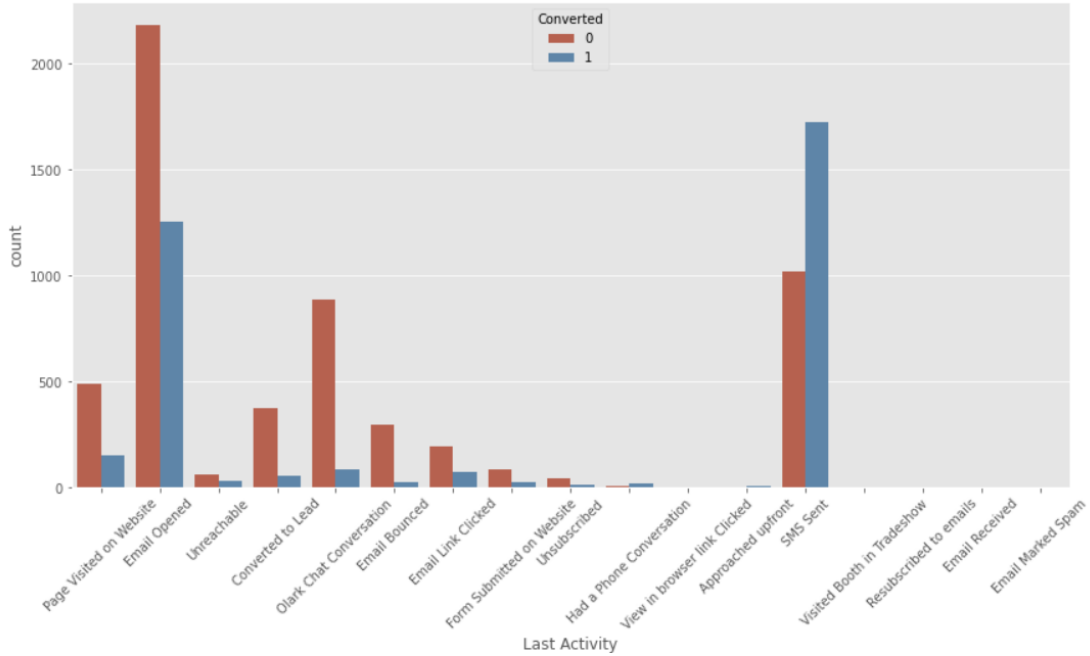
- According to the graph, the majority of leads prefer not to be informed through phone communication.
- This suggests that phone calls may not be the preferred method of communication for most leads. X Education should consider alternative channels such as email or SMS messages to effectively engage with leads and increase the chances of conversion.



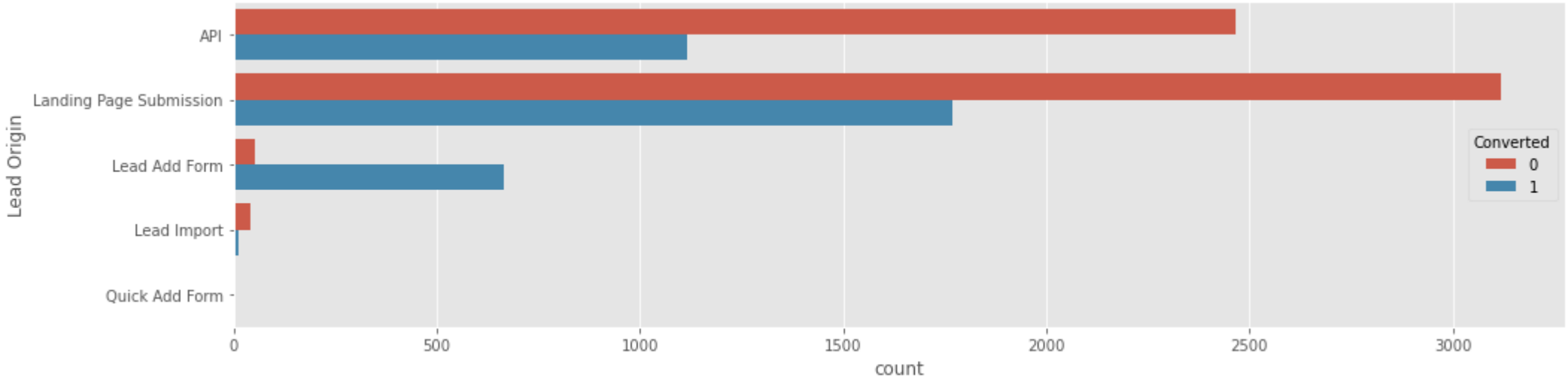
LAST ACTIVITY VS COUNT

- 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.



LEAD ORIGIN VS COUNT



- The graph clearly shows that the landing page submission method is preferred by the majority of leads, while the API, lead add form, and lead import methods have lower usage.

This valuable insight can be used by X Education to optimize their lead generation process.

By focusing on improving the landing page submission method and finding ways to increase engagement with the other methods,

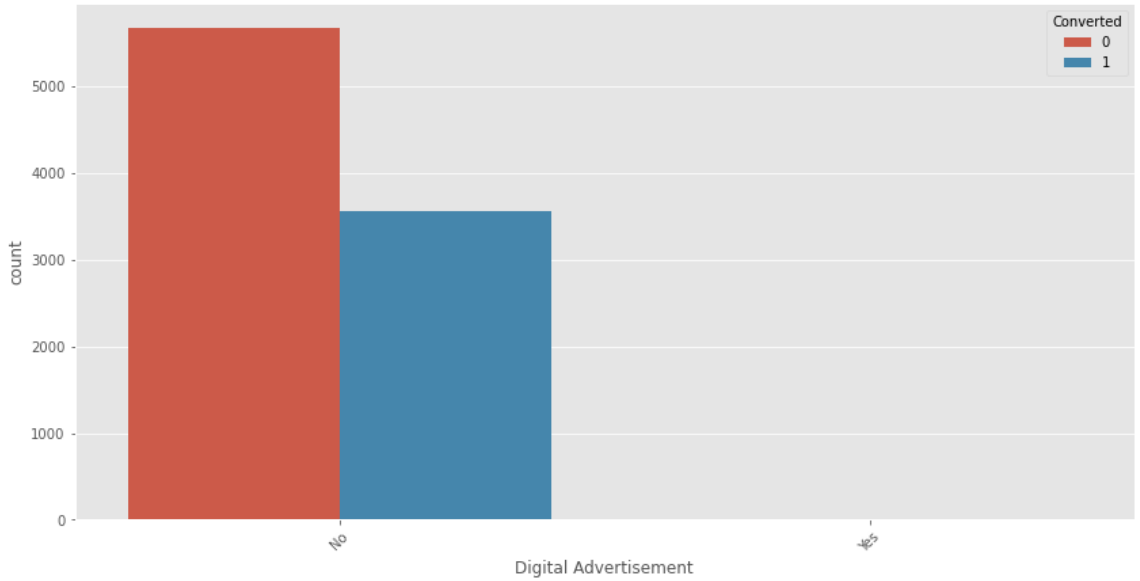
X Education can enhance lead conversion and overall performance..

DIGITAL ADVERTISEMENTS VS COUNT

The graph clearly indicates that digital advertisements do not yield 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.

Based on these findings, it is recommended that X Education reevaluate their digital advertising strategies and explore alternative lead generation methods that have shown higher conversion rates.

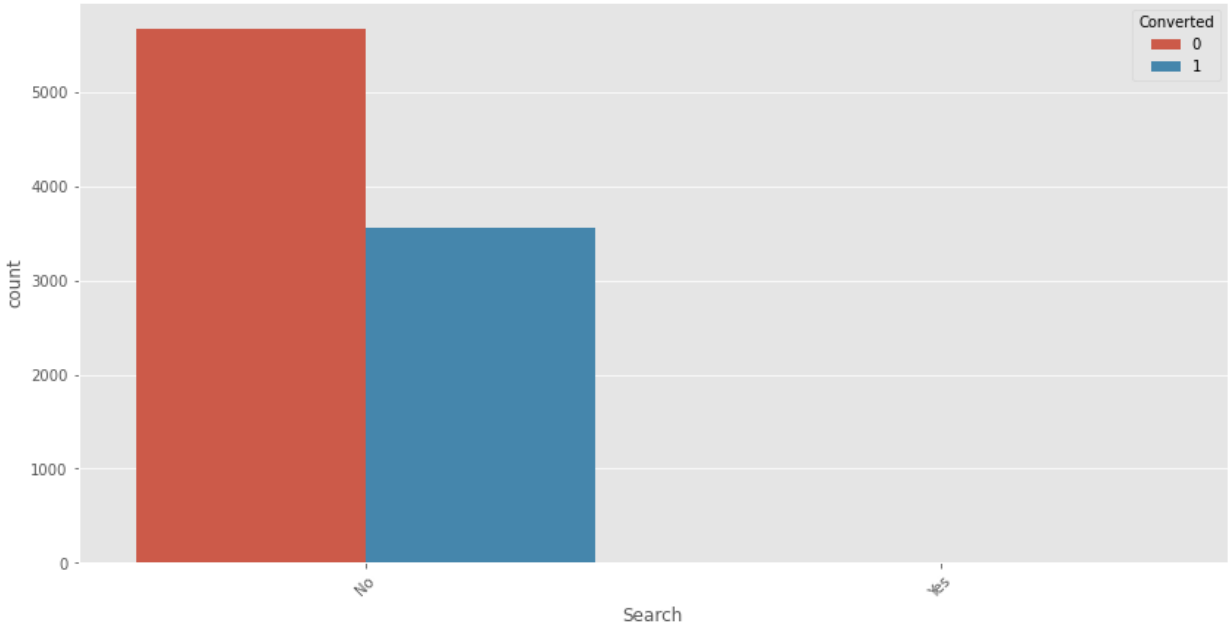


SEARCH VS COUNT

-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.



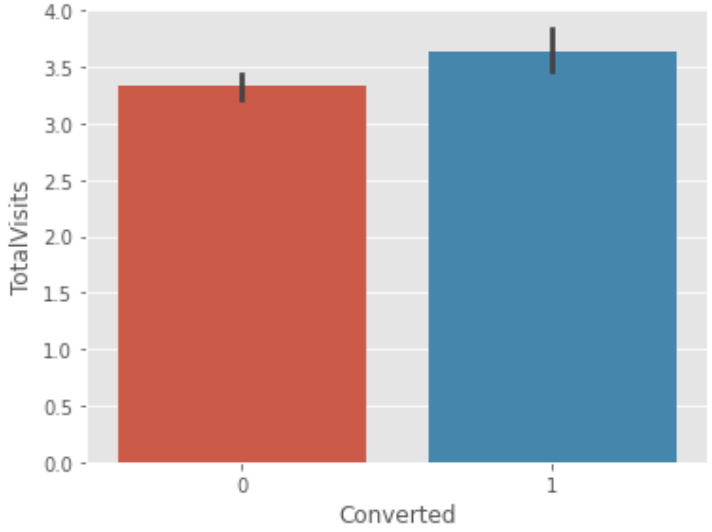
TOTAL TIME SPENT ON WEBSITES VS CONVERTED

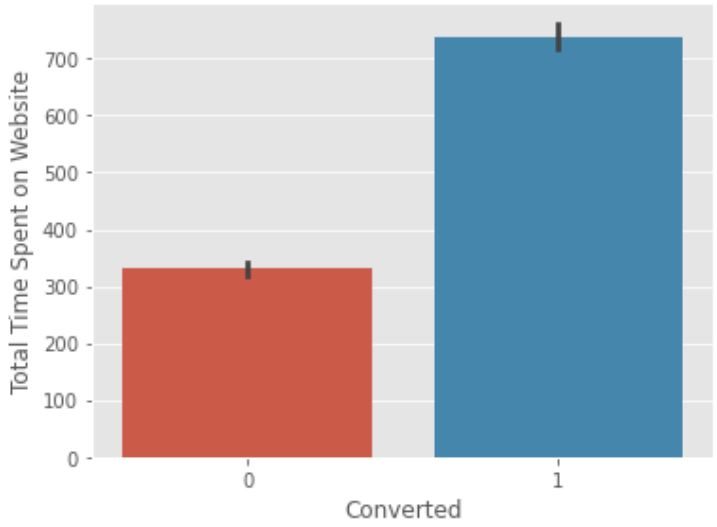
-According to the graph, a larger number of people are spending more time on the website compared to the average time.

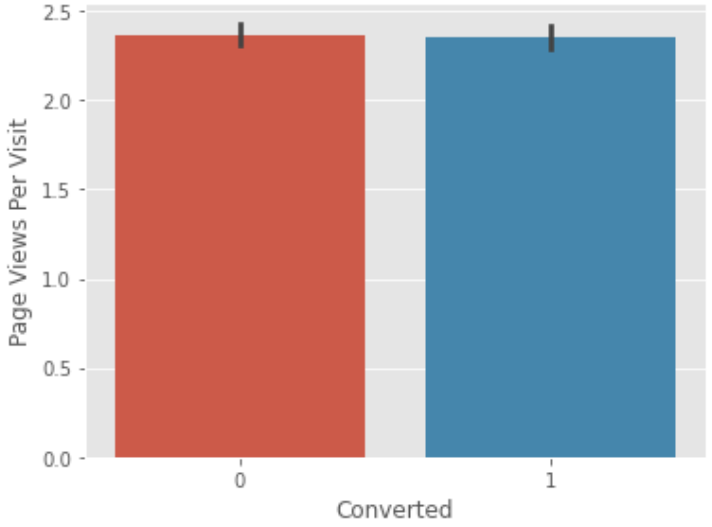
-This indicates that these leads, who are spending more time on the website, have the potential to be promising leads for X Education.

TOTAL VISITS VS CONVERTED

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







Model Building Steps

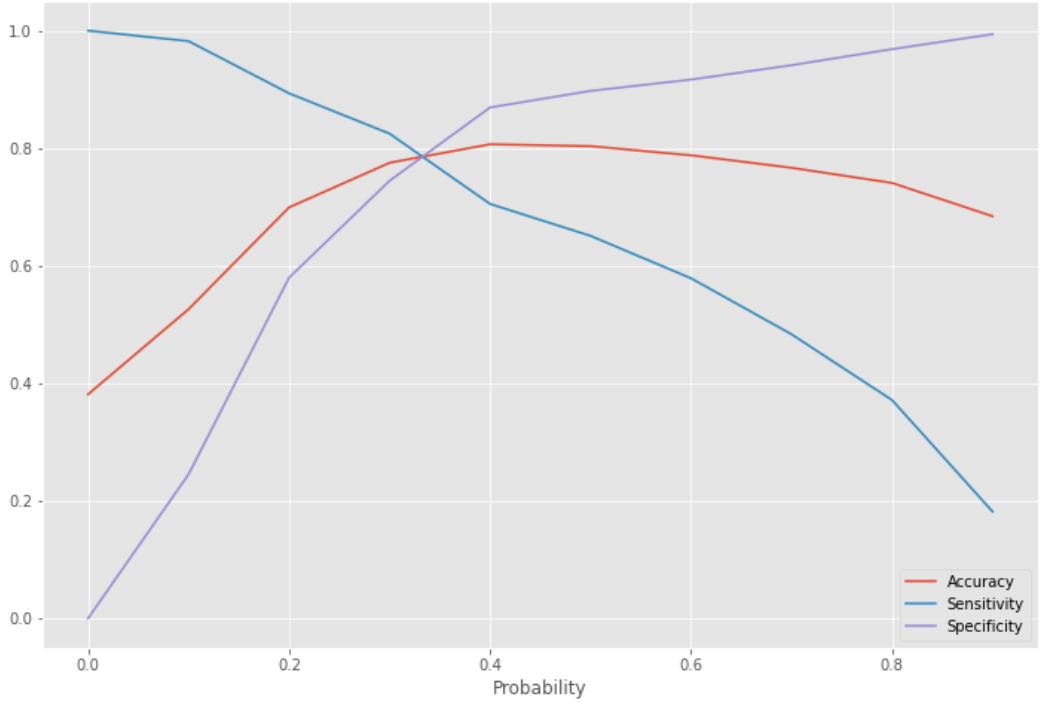
1. The dataset was divided into a train set and a test set to facilitate model training and evaluation.
2. Scaling techniques were applied to the variables in the train set to ensure consistency and prevent any bias.
3. The initial model was constructed using the train set, serving as a baseline for further refinement.
4. Recursive Feature Elimination (RFE) was employed to identify and eliminate less relevant variables, thereby improving the model's performance.
5. A subsequent model was developed after eliminating variables based on high p-values, focusing on the most significant predictors.
6. The Variance Inflation Factor (VIF) was examined for all remaining columns to detect and address multicollinearity issues.
7. Predictions were generated using the train set to evaluate the model's accuracy and other performance metrics.
8. The model's performance was assessed using various measures and metrics to determine its effectiveness on the train set.
9. Predictions were also made using the test set to assess the model's generalization capabilities and performance on unseen data.
10. Precision and recall analysis were conducted on the test set predictions to gain insights into the model's ability to correctly identify positive instances and avoid false positives.

By following these steps, a logistic regression model was built and evaluated to assign lead scores and identify potential leads with higher conversion chances.

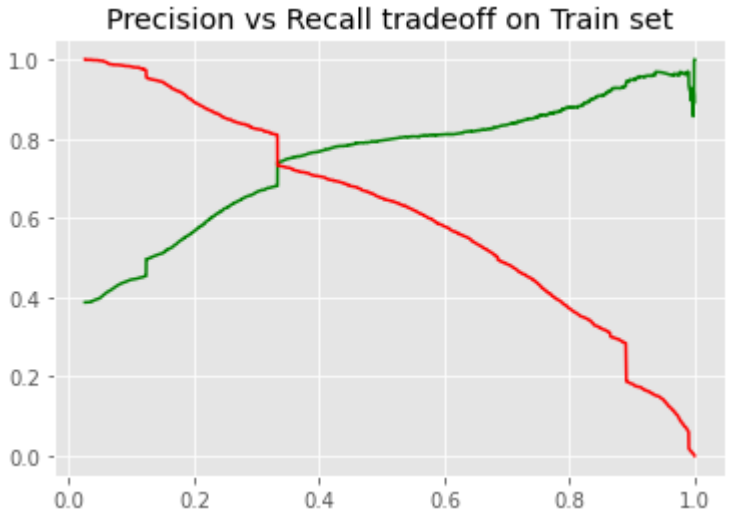
MODEL EVALUATION (TRAIN)

ACCURACY SENSITIVITY AND SPECIFICITY
- 80.9% Accuracy
- 77.6% Sensitivity
- 82.9% Specificity

PRECISION AND RECALL
- 73.4% Precision
- 77.6% Recall



MODEL EVALUATION (TEST)



ACCURACY SENSITIVITY AND SPECIFICITY
- 80.1% Accuracy
- 75.5% Sensitivity
- 83.1% Specificity

PRECISION AND RECALL
- 74.4% Precision
- 75.5% Recall

Conclusion

1. Leads who spend more time on the website than the average have shown a higher likelihood of becoming promising leads. Targeting and engaging these individuals can significantly improve conversion rates.

2. SMS messages have proven to be a powerful tool for lead conversion. Incorporating SMS communication into the marketing strategy can effectively increase conversion rates.

3. Landing page submissions have emerged as a valuable source for generating leads. By focusing on optimizing landing page design and user experience, X Education can attract more leads and boost conversion rates.

4. Certain professions, such as marketing management and human resources management, exhibit higher conversion rates. Targeting individuals from these professional backgrounds can lead to more promising leads.

5. Referrals and incentivizing lead referrals have shown potential in driving higher conversion rates. Implementing referral programs and providing incentives can encourage existing customers to refer more leads, ultimately increasing conversion opportunities.

6. Sending alert messages or providing relevant information has been associated with higher lead conversion rates. Leveraging this approach in communication with leads can be a powerful strategy to improve conversions.

Regarding the logistic regression model:

- The model achieved an impressive accuracy rate of approximately 81%, indicating its capability to make accurate predictions on conversions.
- The threshold for classification was determined using accuracy, sensitivity, specificity, precision, and recall curves, ensuring a well-rounded evaluation.
- The model demonstrated a sensitivity of 76% and specificity of 83%, further validating its accuracy in identifying promising leads.
- Overall, the logistic regression model effectively identifies potential leads with a high likelihood of conversion, making it a valuable tool for X Education's lead generation efforts.

In summary, by leveraging the insights gained from the exploratory data analysis and implementing the logistic regression model, X Education can optimize its lead generation and conversion strategies. This, in turn, will lead to higher conversion rates, business growth, and success in the education industry.