Lead Score Case Study for X Education

Prepared By-
Prashanth, Pratibha & Pratima

Lead Score Case Study for X Education

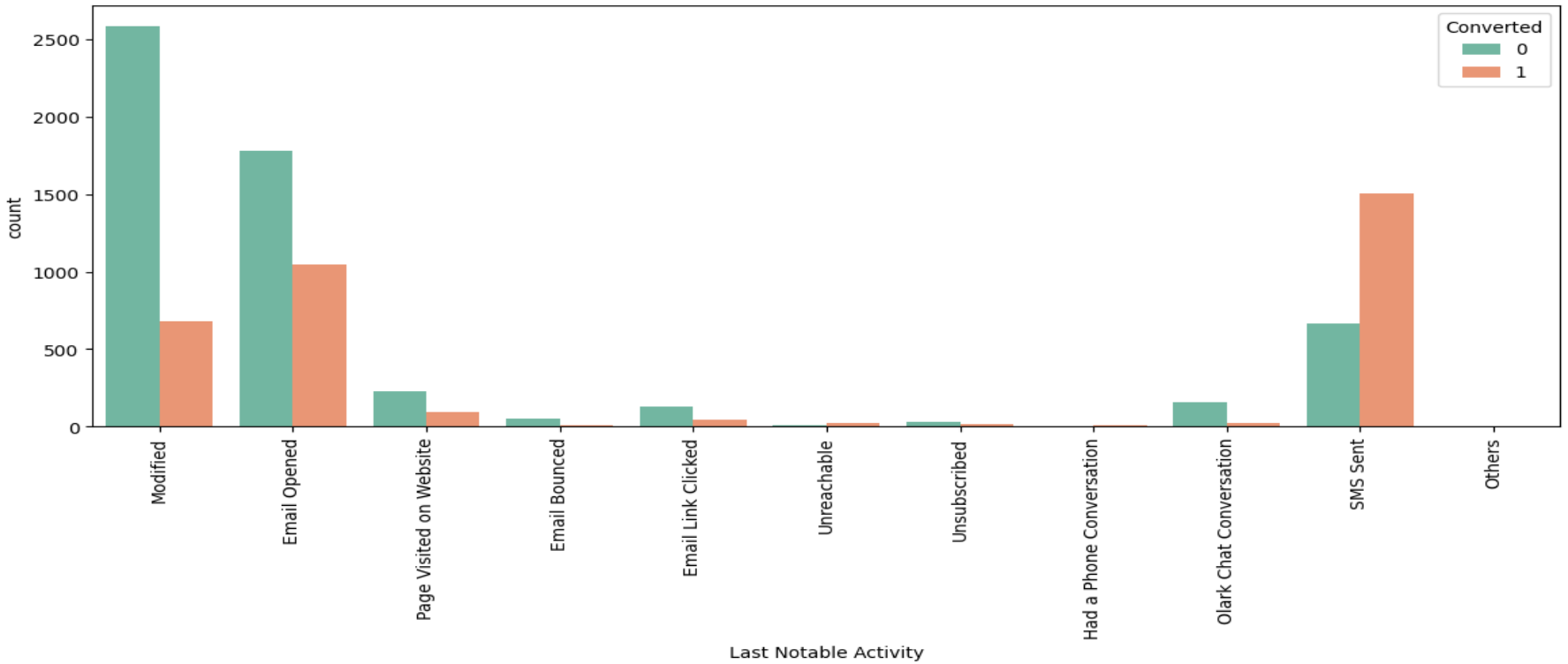
Problem Statement : X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Business Goal: X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy And methods

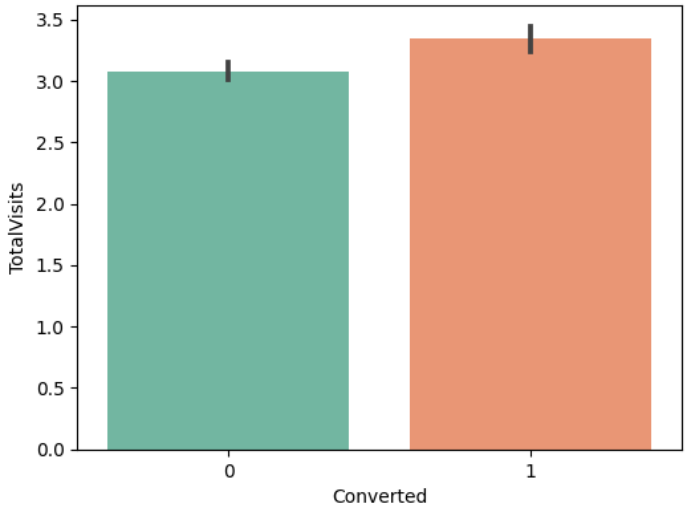
- ✓ **Data Pre-processing and EDA**
  - ✓ The dataset initially contained **9240 rows and 37 columns**.
  - ✓ Identify and treat missing values and outliers
- ✓ **Feature Scaling and Splitting Train and Test Sets**
  - ✓ Feature Scaling of Numeric data
  - ✓ Splitting data into train and test set.
- ✓ **Model Building and Evaluation**
  - ✓ Logistic Regression for building the lead scoring model.
  - ✓ Recursive Feature Elimination (RFE) for select impactful variables.
- ✓ **Predictions and Lead Scoring**
  - ✓ To check accuracy and balance between sensitivity and specificity.
- ✓ **Results and Recommendations**
  - ✓ The model's insights for optimizing resource allocation and focusing on the most promising leads.

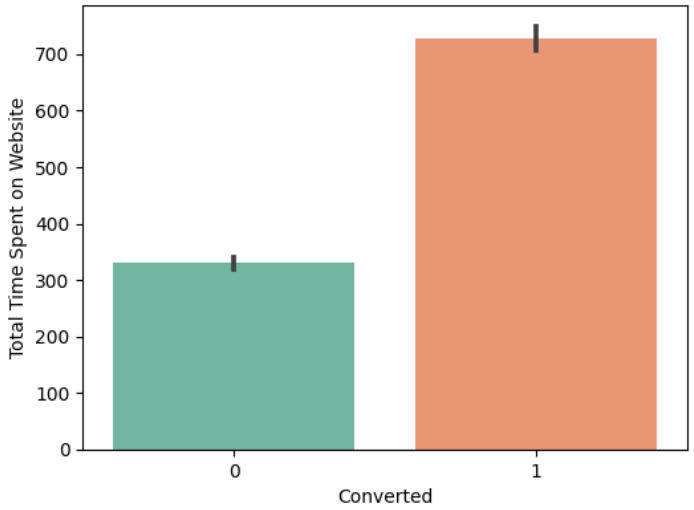
Exploratory data analysis

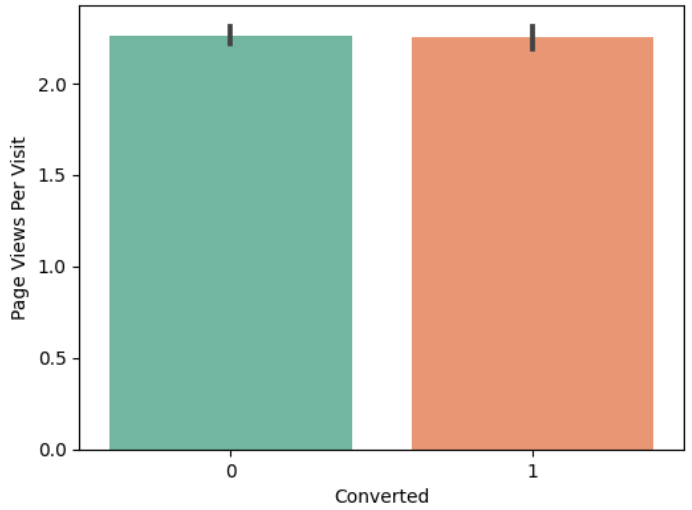


•Maximum lead is generated having 'Modified' and 'Email Opened' as Last Notable Activity but the conversion rate is less
•Last notable activity with label 'SMS sent' has the highest conversion
•Conversion rates for 'Had a phone conversation' is also good, but leads generated are

Exploratory data analysis

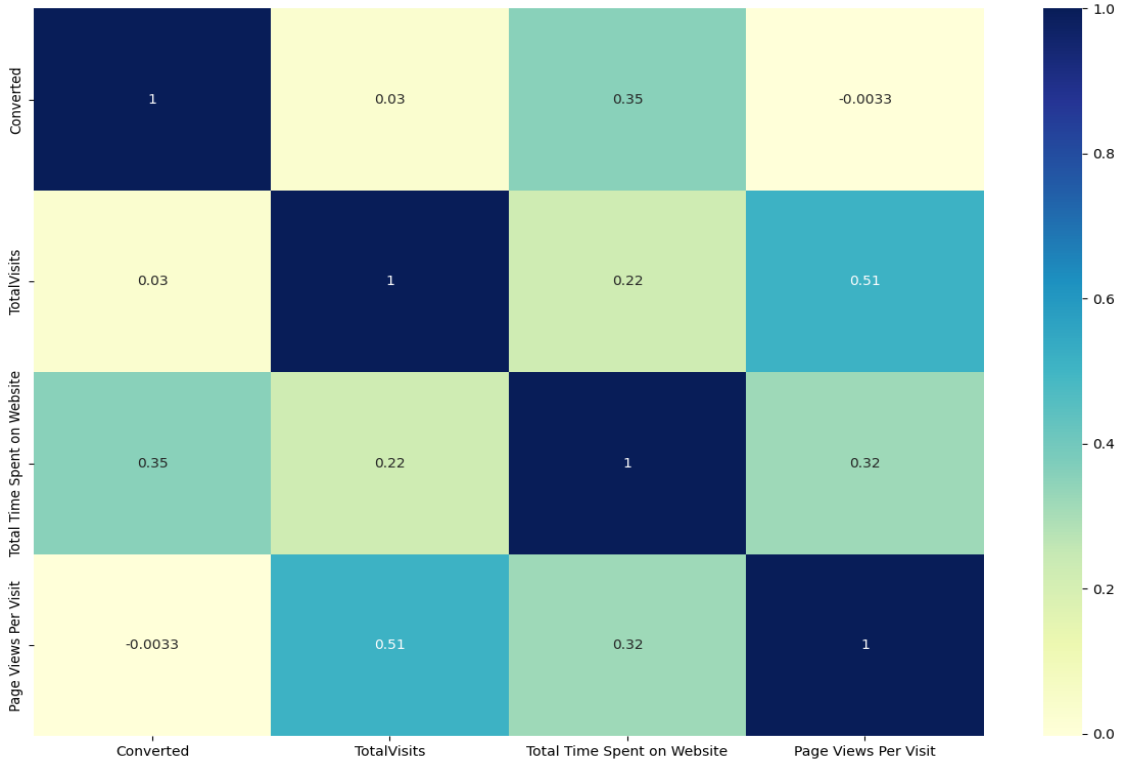






•Conversion rate is high for 'Total Visits' and 'Page Views Per Visit'

Correlation between variables



•**Total Time Spent on Website**: With a correlation coefficient of 0.35 with 'Converted', this variable has a moderate positive correlation, indicating that leads spending more time on the website are more likely to convert.

•**Page Views Per Visit**: The correlation coefficient of 0.51 with 'TotalVisits' suggests that these two variables move together; however, its relationship with 'Converted' is negligible, as shown by the coefficient of -0.0033.

•**TotalVisits**: Despite a higher number of visits being slightly correlated with an increased time on the website (0.22), its direct correlation with 'Converted' is weak (0.03), suggesting that the number of visits alone is not a strong predictor of conversion.

Model Building and Evaluation

Logistic Regression was used for building the lead scoring model.

Recursive Feature Elimination (RFE) was employed to select impactful variables.

The model underwent several iterations for optimization, focusing on improving metrics like accuracy, sensitivity (recall), and specificity.

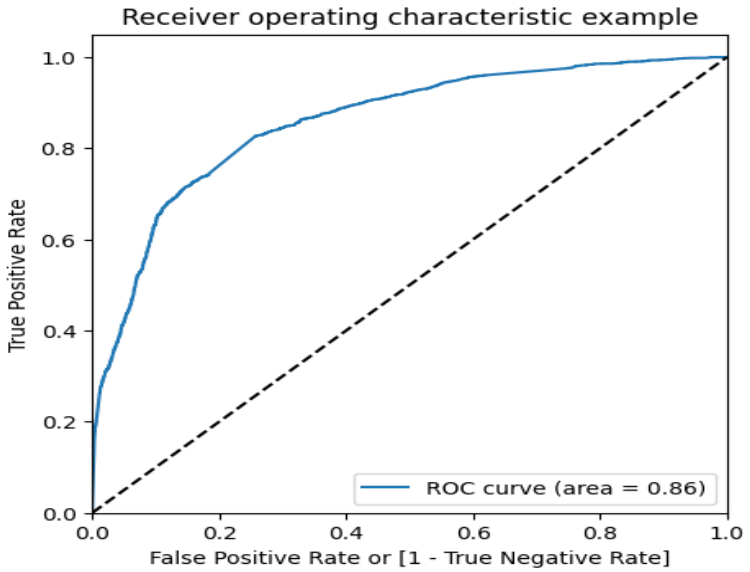
Key performance metrics for the final model on the training and test set were:

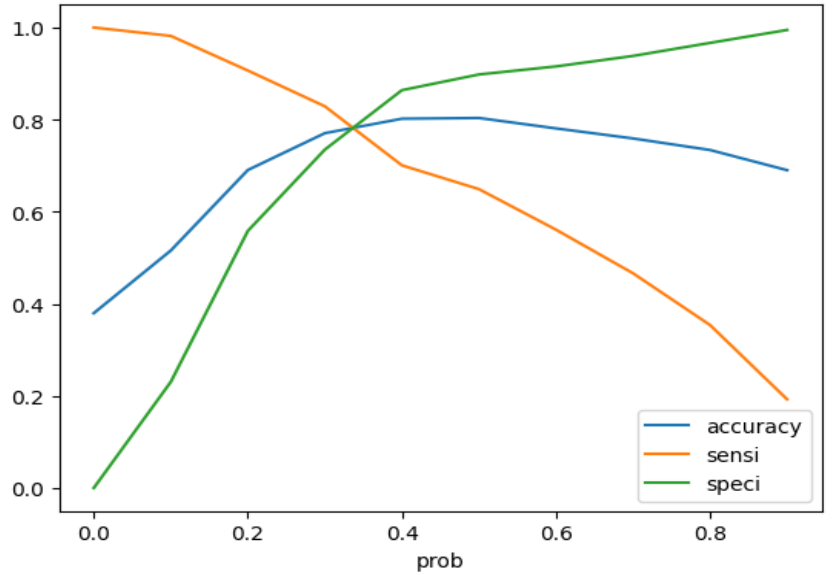
| Train Set | Test Set

|----------|-------------------|

- Accuracy | 77.05% | 77.52% |
- Sensitivity | 82.89% | 83.01% |
- Specificity | 73.49% | 74.13% |

Predictions and Lead Scoring





- The model's predictions on the test set were consistent with the training set, maintaining a high level of accuracy and balance between sensitivity and specificity.
- The **ROC curve value was 0.86**, indicating a strong predictive capability.
- The **optimal cutoff point for lead conversion probability was identified as 0.3**.
- Leads with a score higher than 85 were considered 'hot leads', with a higher likelihood of conversion.

Results and Recommendations

The final predictive model achieved around **83% conversion accuracy**, closely aligning with the CEO's target of an 80% lead conversion rate.

A total of **336 'hot leads'** were identified, which should be prioritized by the sales team for contact.

The model's insights can be used to optimize resource allocation and focus on the most promising leads.

Suggestions for future strategy

Prioritize High-Engagement Leads: Focus on leads that have spent a significant amount of time on the website, as they demonstrate higher interest and engagement.

Target Specific Lead Origins and Sources: Give priority to leads coming from add forms and high-converting sources like Google, Direct Traffic, and Organic Search.

Focus on Working Professionals: Since working professionals have a higher likelihood of conversion, tailor marketing and follow-up strategies specifically for them.

Use Lead Scoring Effectively: Implement a robust lead scoring system to identify and prioritize leads with the highest potential for conversion.

Optimize Communication Strategies: Customize communication based on the lead's behavior and interaction with the website, using personalized emails and targeted content