Lead Scoring Case Study- X Education

Submitted by DSC 65:

Rafida Khannum

Keshavi Gupta

Venkat Aditya

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- our approach to the problem statement

- Data Exploration
- Data Cleaning
- Exploratory Data Analysis
- Data Preparation
- Feature Selection using RFE
- Model Building
- Model Evaluation
- Results on observations
- Conclusion on Hypotheses
- Overall Conclusion

Problem solving methodology

Data Sourcing , Cleaning an Preparation

- Read the Data from Source
- Convert data into clean format suitable for analysis
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.

Feature Scaling and Splitting Train and Test Sets

- Feature Scaling of Numeric data
- Splitting data into train and test set.

Model Building

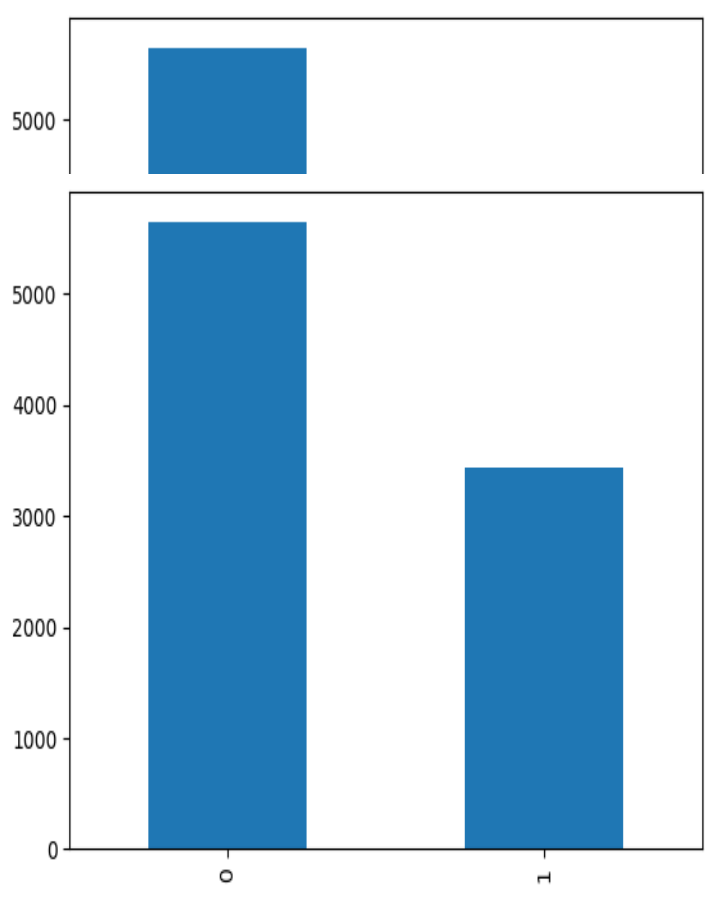
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

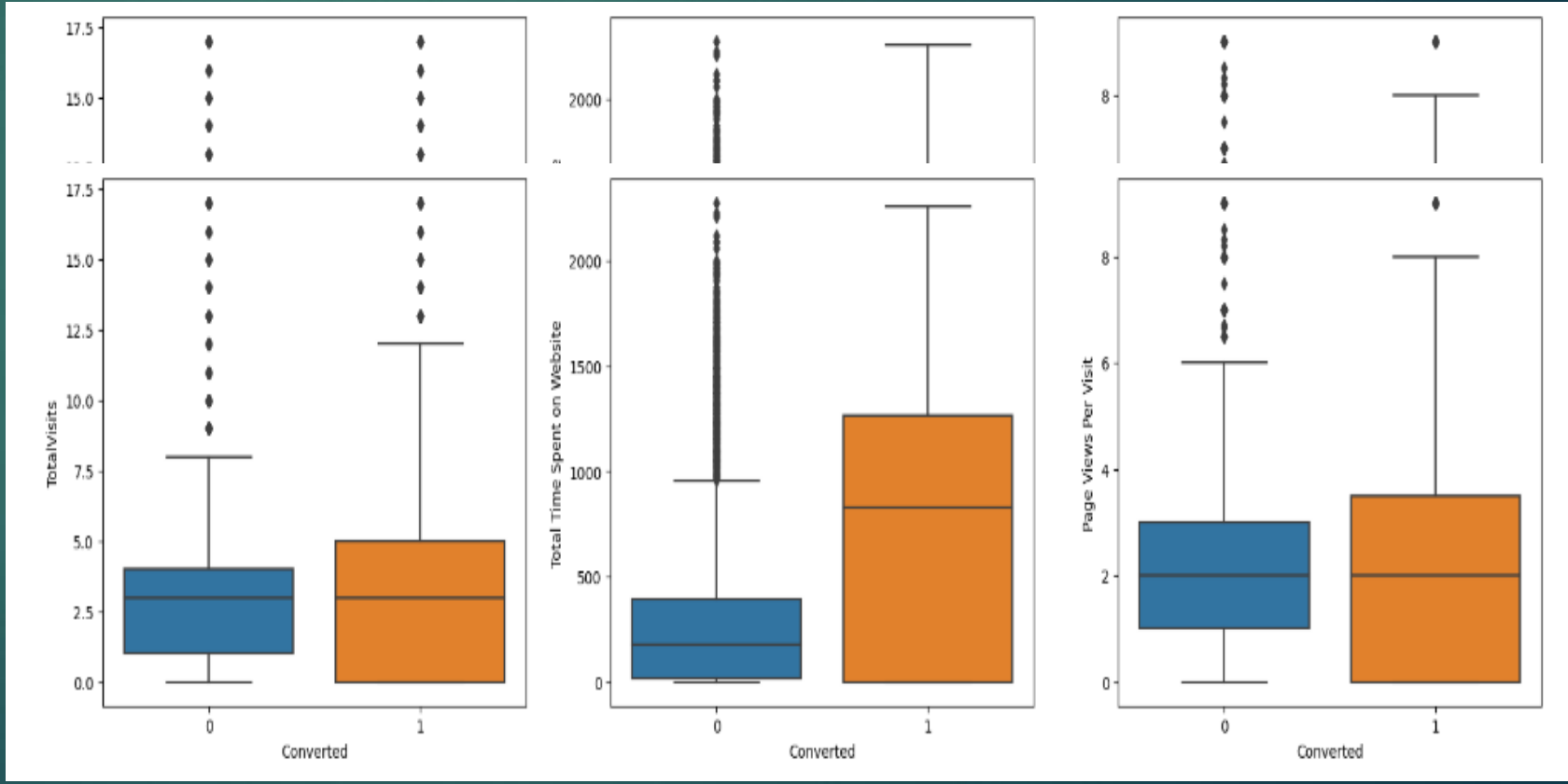
Result

- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

Exploratory Data Analysis

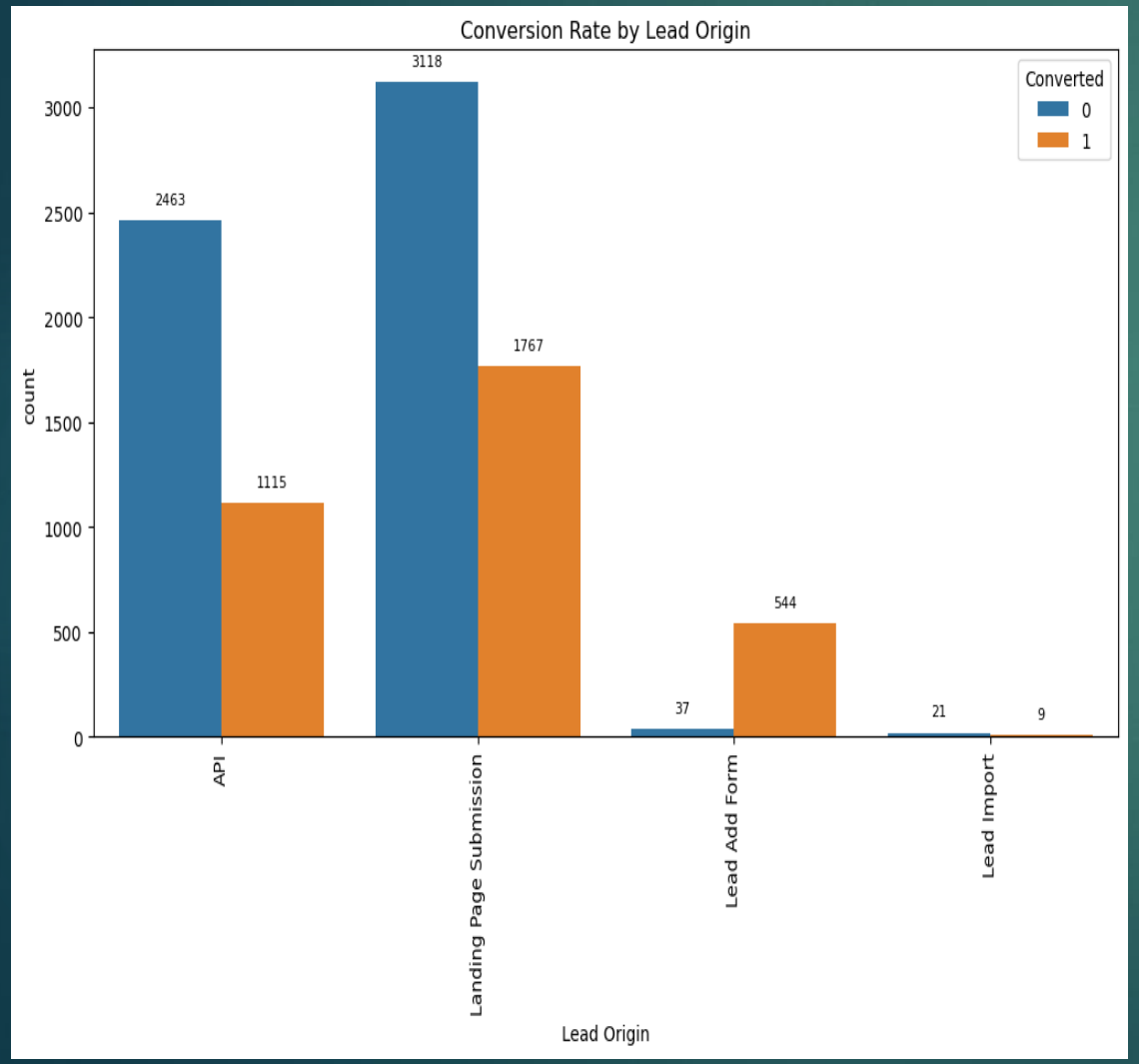
- We have around 37.85% Conversion rate in Total
- The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

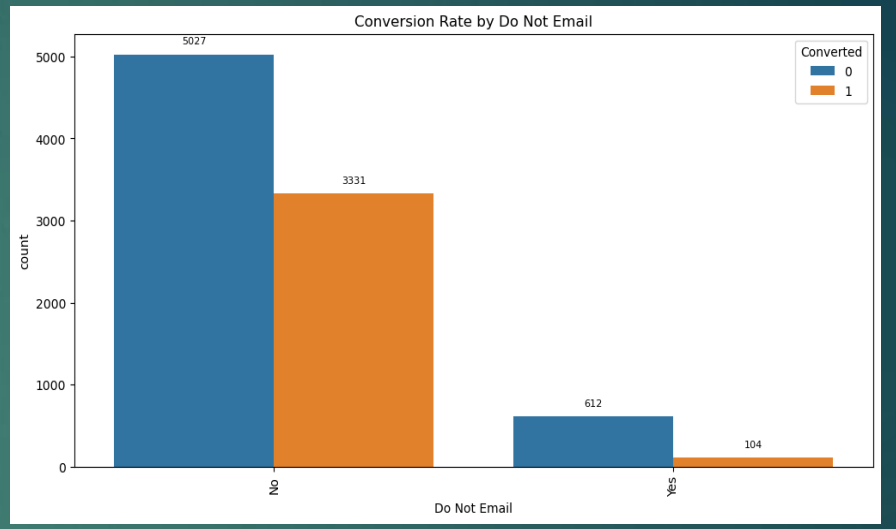


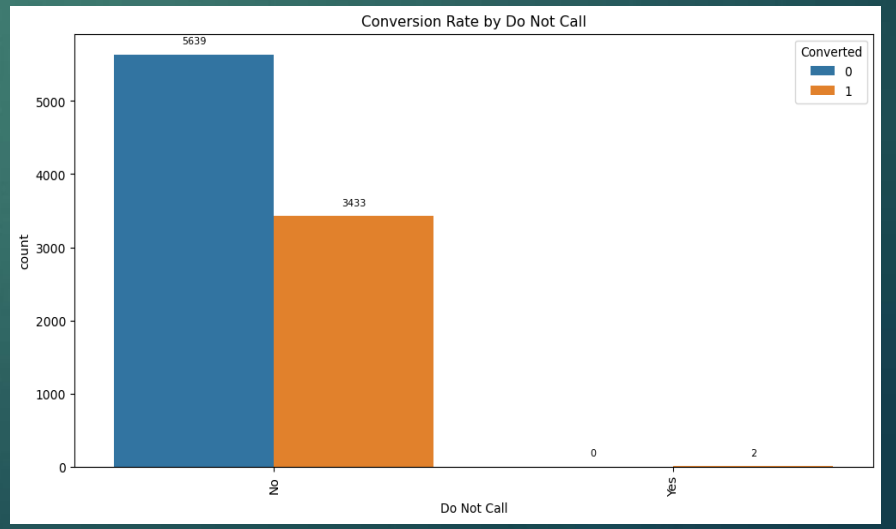


Exploratory Data Analysis- contd.

- In Lead Origin, maximum conversion happened from Landing Page Submission
- Major conversion has happened from Emails sent and Calls made



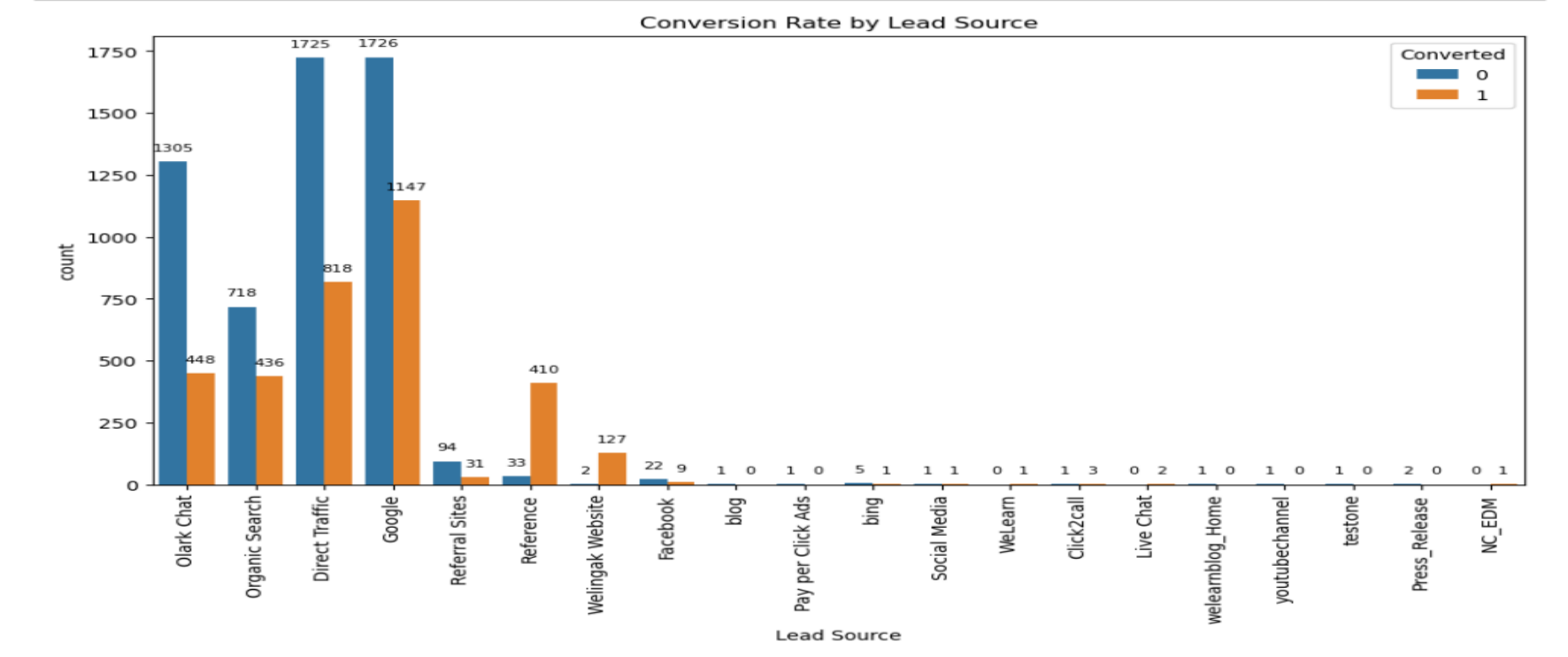




Exploratory Data Analysis- contd.

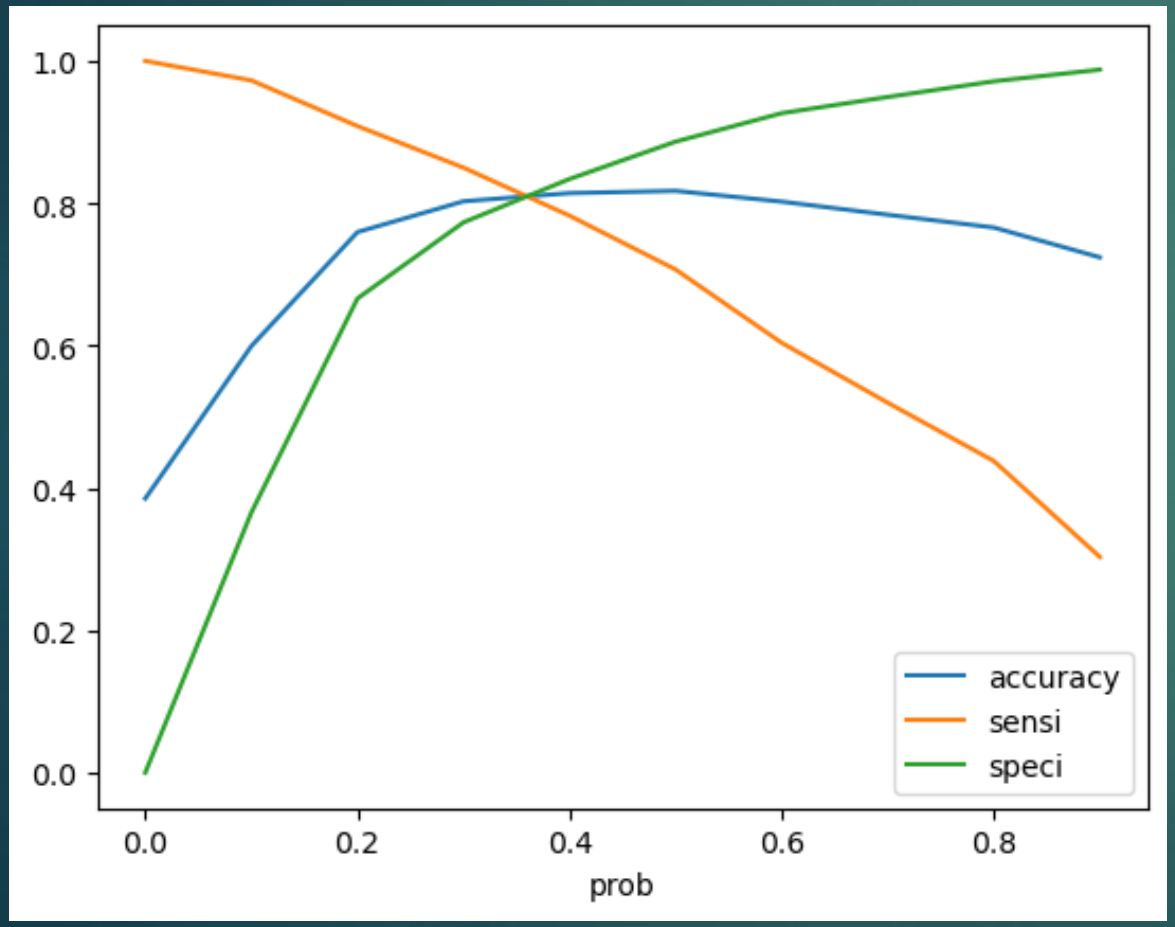
- As highlighted from the visual below, Google and Direct traffic generate the highest number of leads. The conversion rate is particularly high for leads from Referrals and the Welingak website.

- To boost the overall lead conversion rate, efforts should be directed toward improving the conversion rates for Olark chat, Organic Search, Direct traffic, and Google leads, while also increasing the volume of leads from referrals and the Welingak website.



Model Evaluation - Sensitivity and Specificity on Train Data Set

- The graph depicts an optimal cut off of 0.34 based on Accuracy, Sensitivity and Specificity



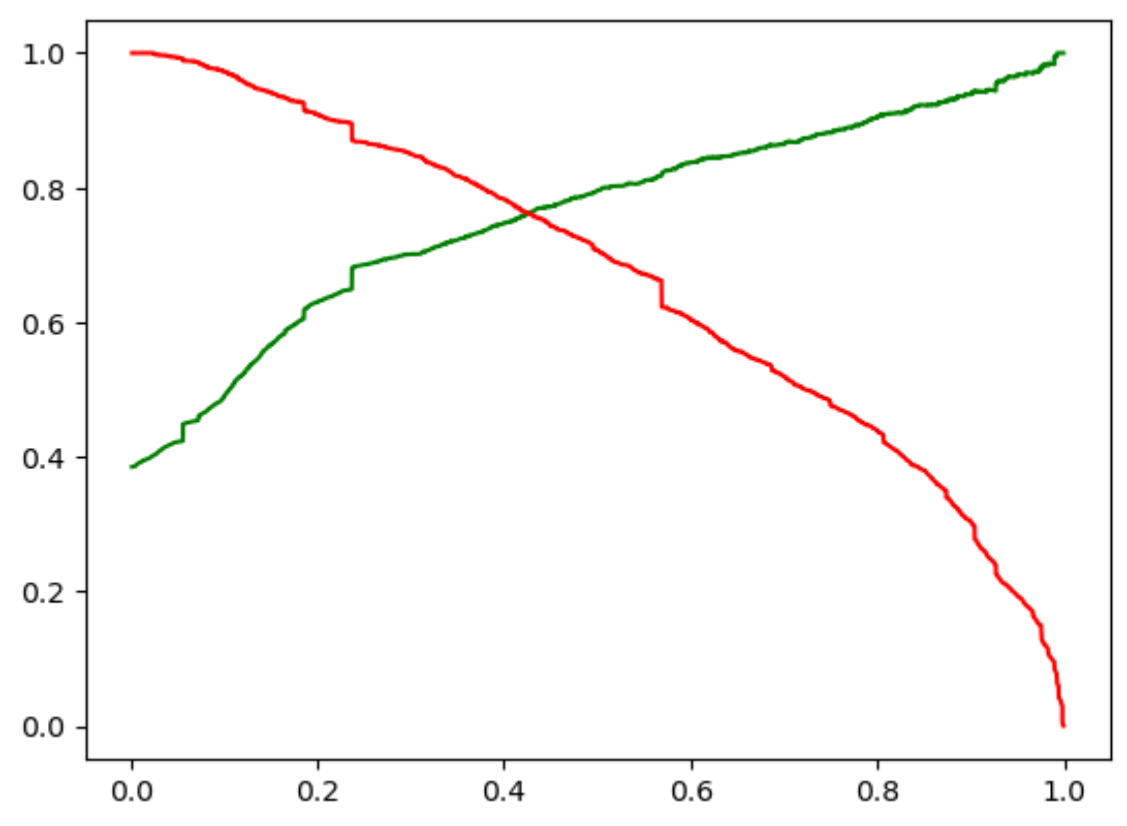
Confusion Matrix

3462	443
717	1729

- Accuracy – 81.73%
- Sensitivity – 70.68%
- Specificity – 88.65%
- False Positive Rate – 11.34 %
- Positive Predictive Value – 79.60 %
- Negative Predictive Value – 82.84%

Model Evaluation- Precision and Recall on Train Dataset

- The graph depicts an optimal cut off of 0.42 based on Precision and Recall



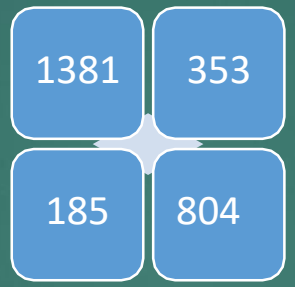
Confusion Matrix

3462	443
717	1729

- Precision - 80 %
- Recall - 71 %

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix



- Accuracy : 80.2%
- Sensitivity : 81.2%
- Specificity : 79.6%

Results on Observations

Comparing the Train and Test performance metrics-

Observation on Train data-

- We observe the below values after running the model on train data:
  - Accuracy: 80.8%
  - Specificity: 82.5%
  - Sensitivity: 79.7%

Observation on Test data-

- After running the model on the Test Data , we obtain
  - Accuracy : 80.2%
  - Sensitivity : 81.2%
  - Specificity : 79.6%

Inference from Model Results:

- **Alignment with Business Goal:**
  The model achieves an accuracy of approximately 80% on both train and test data, closely aligning with the business objective of reaching an 80% lead conversion rate. This suggests that the model is effective in identifying potential leads that are likely to convert.

- **Model Performance Consistency:**
  The performance metrics—specificity, sensitivity, and accuracy—are consistent between the train and test datasets, which indicates that the model generalizes well and is not overfitting. This consistency gives confidence that the model's predictions are reliable and can be trusted to guide business decisions.

- **Sensitivity and Specificity Balance:**
  The model's sensitivity (ability to correctly identify leads that will convert) and specificity (ability to correctly identify leads that will not convert) are well-balanced. This balance is crucial for optimizing sales efforts, ensuring that the sales team focuses on genuine 'Hot Leads' while minimizing time spent on less promising ones.

- Impact on Sales Strategy:
  With this model, the sales team can prioritize leads with a higher likelihood of conversion, directly addressing the inefficiency problem stated in the business problem. By improving focus on high-probability leads, the model supports the strategic goal of enhancing lead conversion rates and thereby increasing overall revenue.

Recommendations based on our analysis

- Focus on High-Conversion Lead Sources:
  - Welingak Website and Reference sources have strong positive coefficients, indicating leads from these sources are more likely to convert. Prioritize follow-ups with these leads.
- Target Working Professionals:
  - Leads identified as Working Professionals show a high likelihood of conversion. Direct more sales efforts towards this demographic.
- Leads with High Website Engagement:
  - Leads that have spent more time on the website are more engaged and show a higher likelihood of converting. These leads should be prioritized for follow-up calls.
- Engage with Olark Chat Leads:
  - Leads originating from Olark Chat also demonstrate a higher likelihood of conversion, making them a valuable target for follow-up.
- Avoid Follow-Ups with Already Converted Leads:
  - Leads whose last activity was 'Converted to Lead' are already on the path to conversion and may not need immediate follow-up.

Recommendations based on our analysis- contd.

- Reconsider Leads with Specific Last Activities:
  - Leads whose last activity was 'Olark Chat Conversation' or 'Email Bounced' show lower conversion likelihood. Consider deprioritizing follow-ups with these leads.
- Re-evaluate Leads from Certain Origins and Specializations:
  - Leads originating from 'Landing Page Submission' and those with a specialization of 'Others' are less likely to convert. Resources might be better allocated elsewhere.
- Respect Do Not Email Preferences:
  - Leads that have opted 'Do not Email' as 'Yes' are significantly less likely to convert. Focus on alternative communication methods for these leads or deprioritize them.
- Assess Leads with Specific Notable Activities:
  - Leads whose last notable activities were 'Page Visited on Website', 'Email Link Clicked', or 'Modified' also show lower conversion potential. Consider deprioritizing these leads in follow-up efforts.

Conclusion on Hypotheses

Hypothesis 1: Leads that spend more time on the website are more likely to convert.

- **Validated:** The analysis supports this hypothesis, as leads with higher engagement (spending more time on the website) show a higher likelihood of conversion. This is reflected in the positive coefficient for the "Total Time Spent on Website" variable, indicating a positive relationship between website engagement and conversion likelihood.

Hypothesis 2: Certain lead sources (e.g., referrals, specific ads) have higher conversion rates.

- **Validated:** The model's output shows that certain lead sources, such as "Welingak Website" and "Reference," have strong positive coefficients, suggesting higher conversion rates. This confirms that not all lead sources are equal in conversion potential, and targeting those with higher conversion rates can improve overall efficiency.

Hypothesis 3: Demographic factors like occupation and country might influence conversion likelihood.

- **Partially Validated:** The analysis confirms that occupation, specifically leads identified as "Working Professionals," is a significant factor influencing conversion likelihood, as indicated by its positive coefficient. However, the hypothesis regarding country influence wasn't explicitly tested or observed in the provided data, so further analysis would be needed to explore that aspect.

Overall Conclusion:

- The data analysis strongly supports the first two hypotheses, providing clear evidence that website engagement and certain lead sources are crucial factors in predicting lead conversion.
- The third hypothesis is partially validated, with occupation being an influential factor, though further exploration is needed for other demographic variables like country.
- These insights can be directly applied to refine X Education's lead conversion strategy, making it more targeted and effective.

Overall Conclusion

Based on the analysis and model performance, several key insights can be drawn:

- **Model Effectiveness:** The model demonstrates strong predictive accuracy, aligning closely with the business objective of improving lead conversion rates to 80%. The consistency in performance metrics between the training and test datasets indicates that the model is reliable and generalizes well, making it a solid tool for guiding sales strategies.

- **Targeted Sales Efforts:** The model's output has identified 386 leads with a higher likelihood of conversion. By focusing on these leads, X Education can optimize sales efforts and improve efficiency, addressing the current inefficiency problem where the sales team spends time on leads with lower conversion potential.

- **Validated Hypotheses:** The analysis supports the hypotheses that,
  - Leads spending more time on the website are more likely to convert. Certain lead sources, like referrals and specific online channels, have higher conversion rates.
  - Demographic factors, such as occupation, play a significant role in conversion likelihood.
  - **Strategic Recommendations:** The recommendations based on the model's output provide actionable steps for X Education to enhance their lead conversion strategy. By prioritizing leads from high-conversion sources, targeting working professionals, and focusing on engaged leads, the company can better allocate resources and increase its conversion rate.

- **Potential for Business Impact:** By implementing these targeted strategies, X Education can significantly increase its lead conversion rate, moving closer to the desired 80%. This improvement will not only increase revenue but also optimize the sales team's efforts, ensuring that time and resources are spent on the most promising leads.

Overall, the model provides a clear pathway to achieving the business goal of a higher lead conversion rate, supporting strategic decision-making and enabling the company to make more informed calls on which leads to pursue.