CASE STUDY

LEAD CONVERSION

For X Education

Problem Statement

An education company named 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 Goals

X Education requires help in selecting the most promissing goals i.e., leads that are more likely to get converted to paying customers.

The company needs a model wherein a lead score is assigned to each of the leads such that the customers with the higher lead score is more likely top get converted.

The CEO in particular has given a ballpark of the target lead conversion rate to be around 80%

Strategy

- Read the data for analysis
- Clean and Prepare the data
- Exploratory Data Analysis
- Splitting the data into Test and Train Datasets.
- Feature Scaling
- Building a Logistic Regression Model
- Plotting ROC Curve and finding optimal cut-off point
- Assigning Lead Score to all the leads
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and recall
- Applying the best model in test data based on the sensitivity and specificity metrics.

Problem Solving Methodology

Data Sourcing, cleaning and preparation

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

Feature Scaling and Splitting Train and Test Sets

- Feature scaling of numeric data.
- Splitting Data into Train and Test Sets.

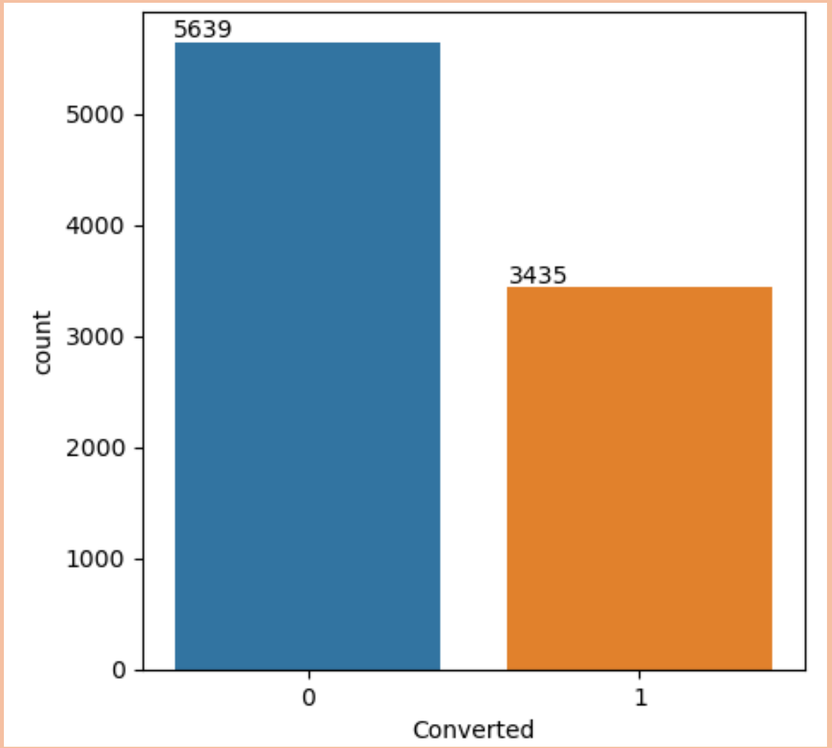
Model Building

- Feature Selection using RFE.
- Determining the optimal model using Logistic Regression.
- Calculate various metrics like Accuracy, Sensitivity, Specificity, Precision and Recall and evaulate the model.

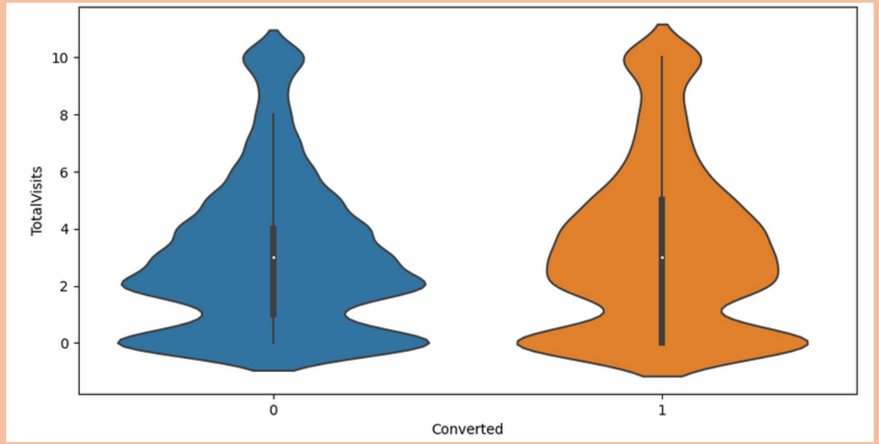
Result

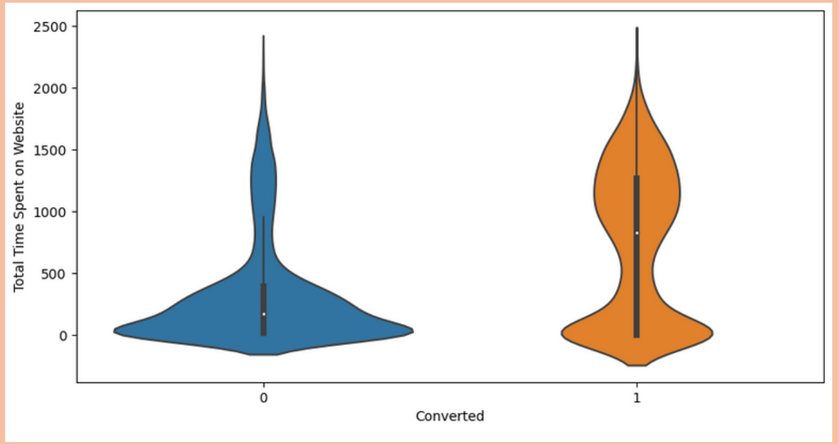
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cutoff threshold from sensitivity and speciticity metrics.

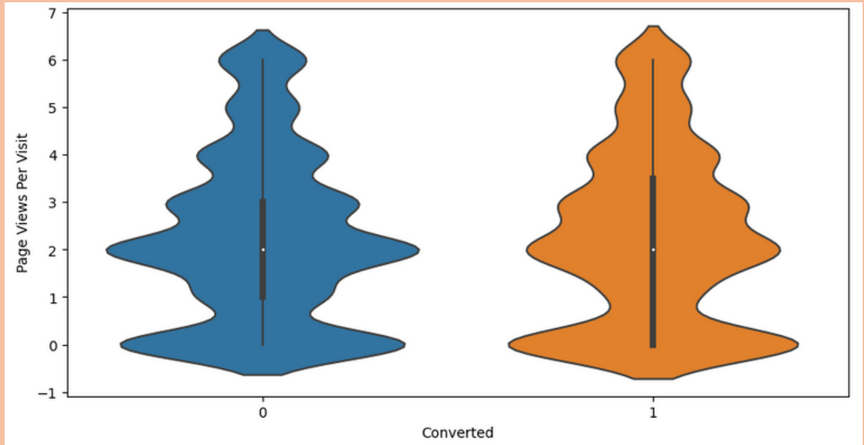
Exploratory Data Analysis



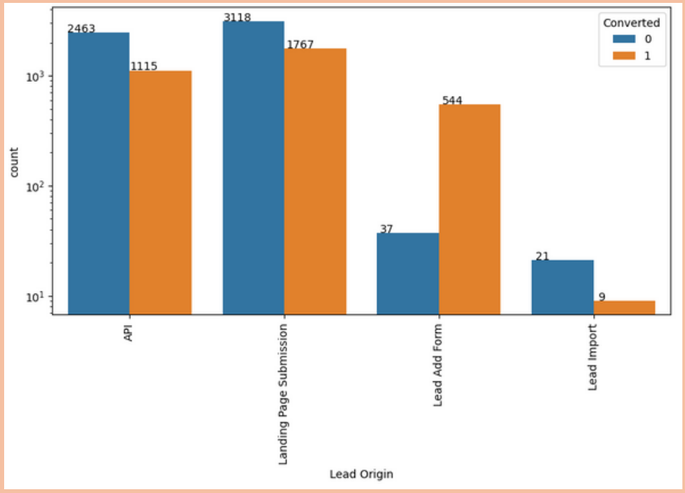
The conversion rate in total is around 38%.



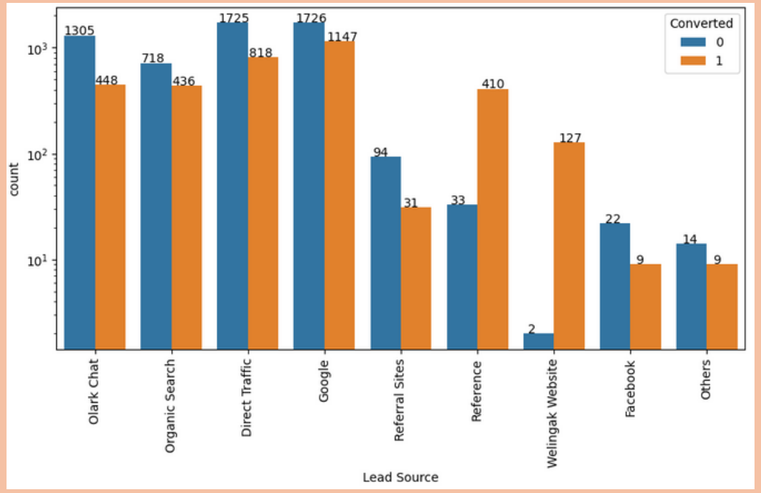




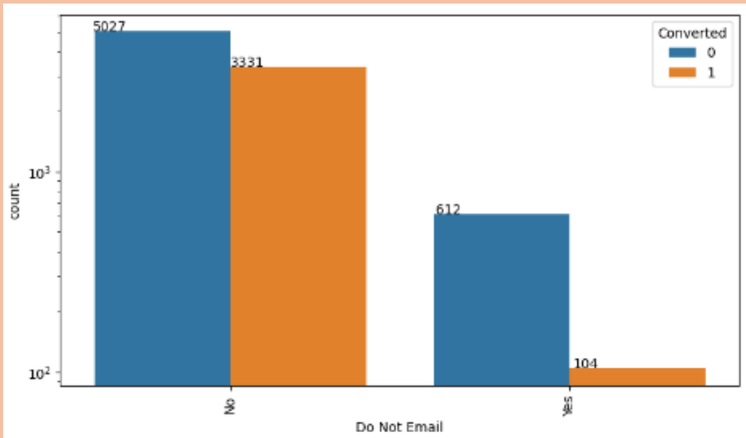
The conversion Rate was highest for the 'Total Time Spent on the Website' and there was no notable difference between the median of converted and not converted for 'Total Visits' and 'Page Views Per Visit'.

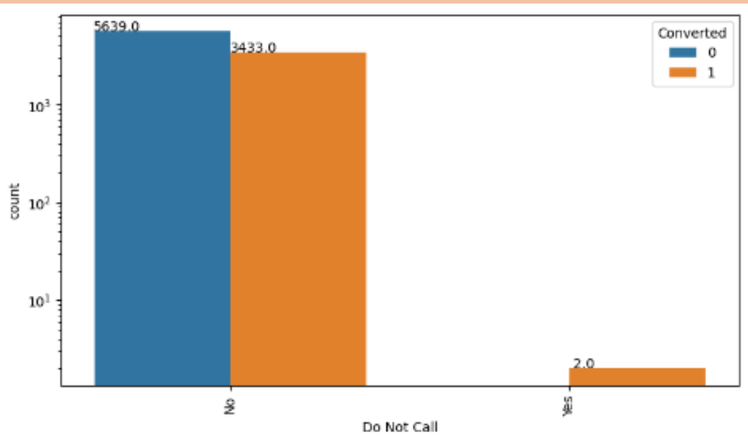


In Lead origin maximum conversion happended from 'Landing Page Submission'

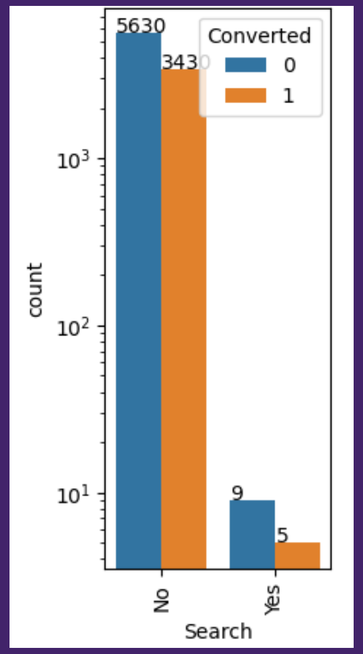


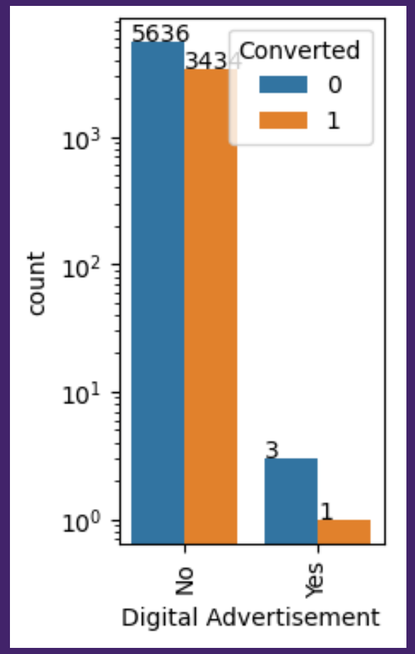
Major conversion in the 'Lead Source' is from 'Google'.

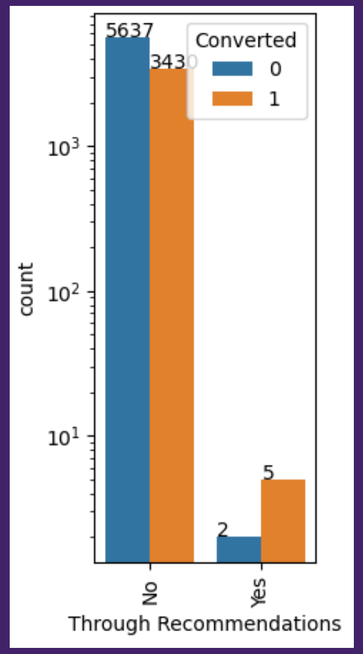




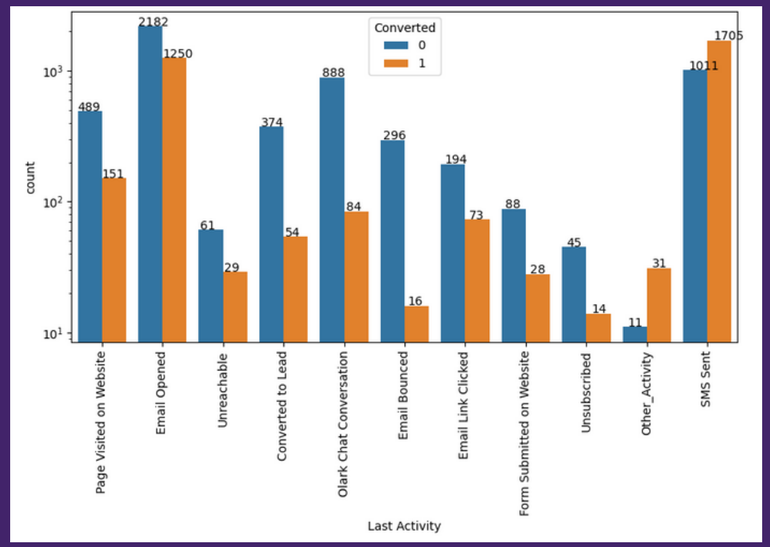
Major conversion has happened from 'Email Sent' and 'Calls Made'







Not much impact on conversion rates through 'Search', 'Digital Advertisments' and 'Through Recommendations'



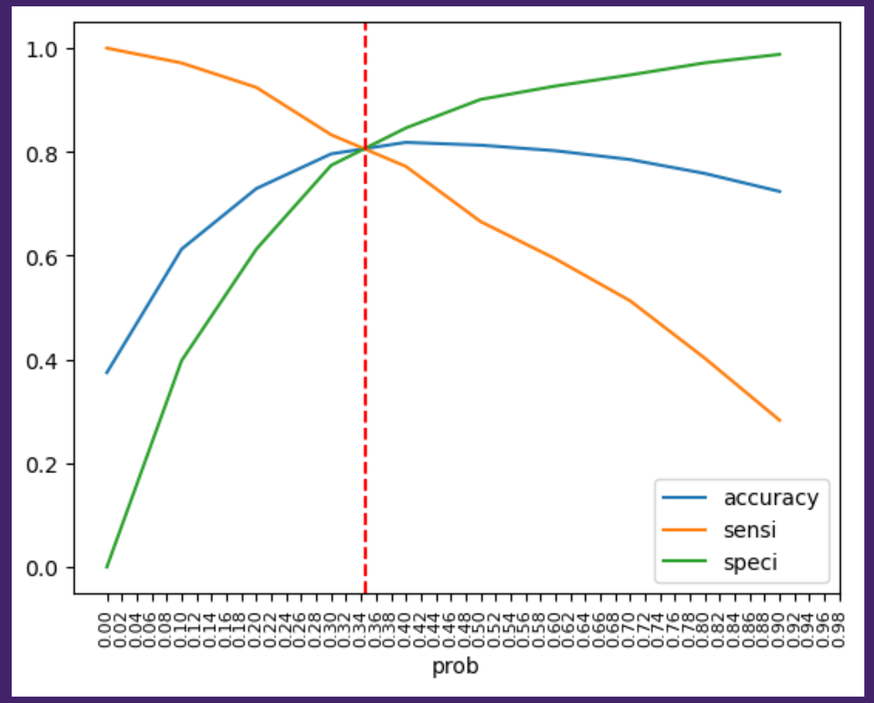
'Last Activity' value of 'SMS_Sent' had more conversion .

Variables Impacting Conversion Rate

- Page Views Per Visit
- Lead Origin_Landing Page Submission
- Lead Source_Olark Chat
- Last Activity_Email Opened
- Specialization_Others
- TotalVisits
- Last Activity_SMS Sent
- Last Activity_Olark Chat Conversation
- Lead Source_Reference
- Total Time Spent on Website
- Do Not Email
- What is your current occupation_Working Profes...
- Last Activity_Email Link Clicked
- Lead Source_Welingak Website
- Last Activity_Unsubscribed
- Last Activity_Unreachable
- What is your current occupation_Student
- Last Activity_Other_Activity
- Specialization_Hospitality Management

Model Evaluation - Sensitivity and Specificity on Train Dataset

The graph depicts an optimal cut off of 0.345 based on Accuracy, Sensitivity and Specificity.



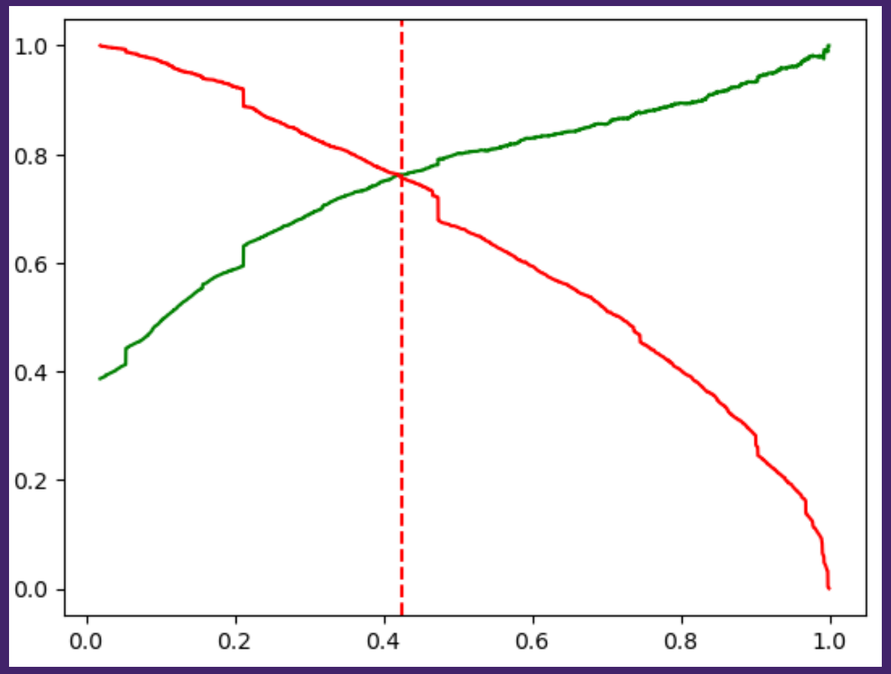
Confusion Matrix

3690	850
520	2199

- Accuracy: 81%
- Sensitivity: 80%
- Specificity: 81%
- False Positive Rate: 18%
- Postive Predictive Value: 72%
- Negative Predictive Value: 87%

Model Evaluation - Precision and Recall on Train Dataset

The Graph depicts an optimal cutoff of approximately 0.42 based on precision and recall.



- **Precision : 80%**
- **Recall : 66%**

Model Evaluation - Sensitivity and Specificity on Test Dataset

Confusion Matrix

887	212
142	574

- **Accuracy: 80%**
- **Sensitivity: 80%**
- **Specificity: 73%**

Recommendation

- The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.
- The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.
- The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.
- The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.
- The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.
- The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.

Conclusion

- **While we have checked both sensitivity, specificity as well as precision and recall metrics , we have considered the optimal cutoff based on sensitivity and specificity for calculating the final prediction.**
- **Accuracy , Sensitivity and Specificity values of test set are around 80%, 80% and 73 % respectively, which are approximately closer to the respective values calculated using trained set.**
- **The top three variables that contribute for lead getting converted in the model are**

1. Welingak Website from Lead Source
2. Working Professionals from What is your current occupation
3. SMS Sent from Last Activity

- **Overall we have build quite a decent model.**

Thank You