Lead Score Case Study

By
- Vishwanath Hiremath
- Raghvendra Kothe

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.

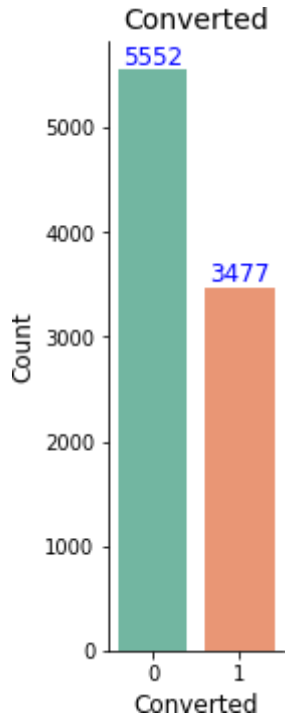
The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

1. Source the data for analysis
2. Clean and prepare the data
3. Exploratory Data Analysis.
4. Feature Scaling
5. Splitting the data into Test and Train dataset.
6. Building a logistic Regression model and calculate Lead Score.
7. Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
8. Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

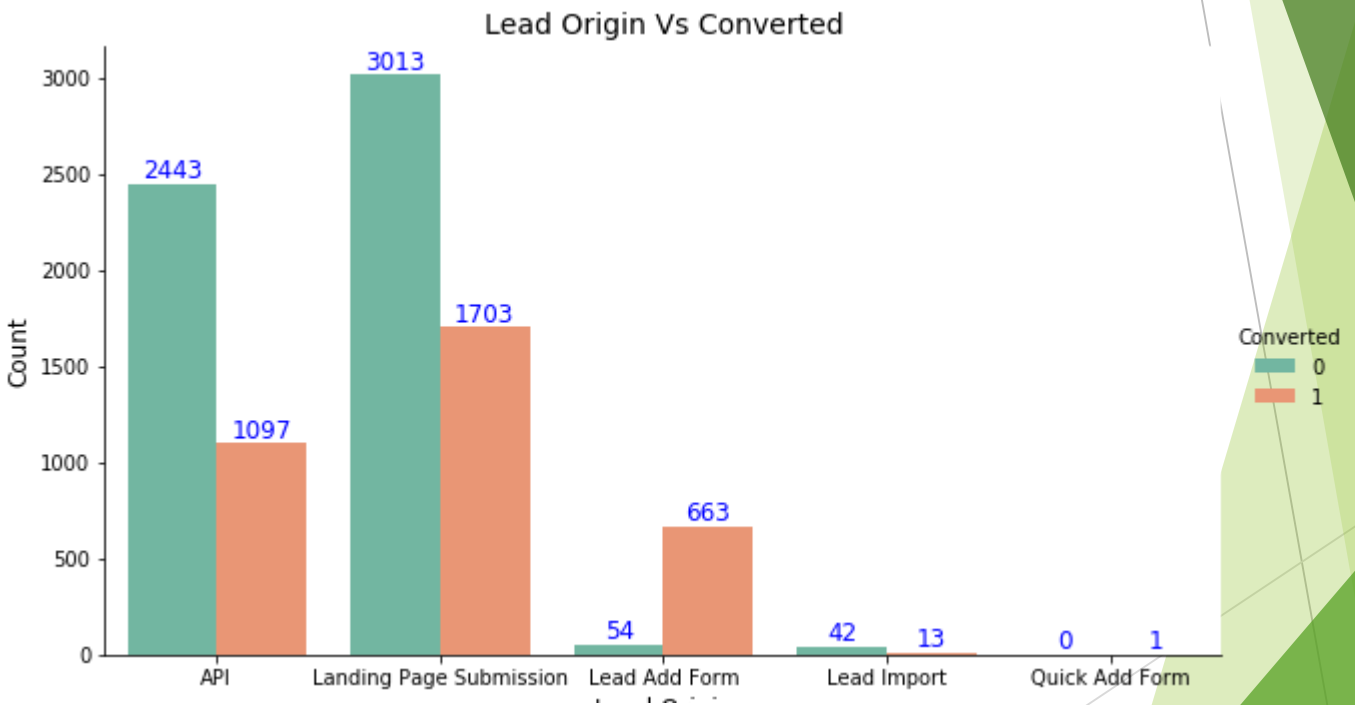
Exploratory Data Analysis

We have around 39% Conversion rate in Total

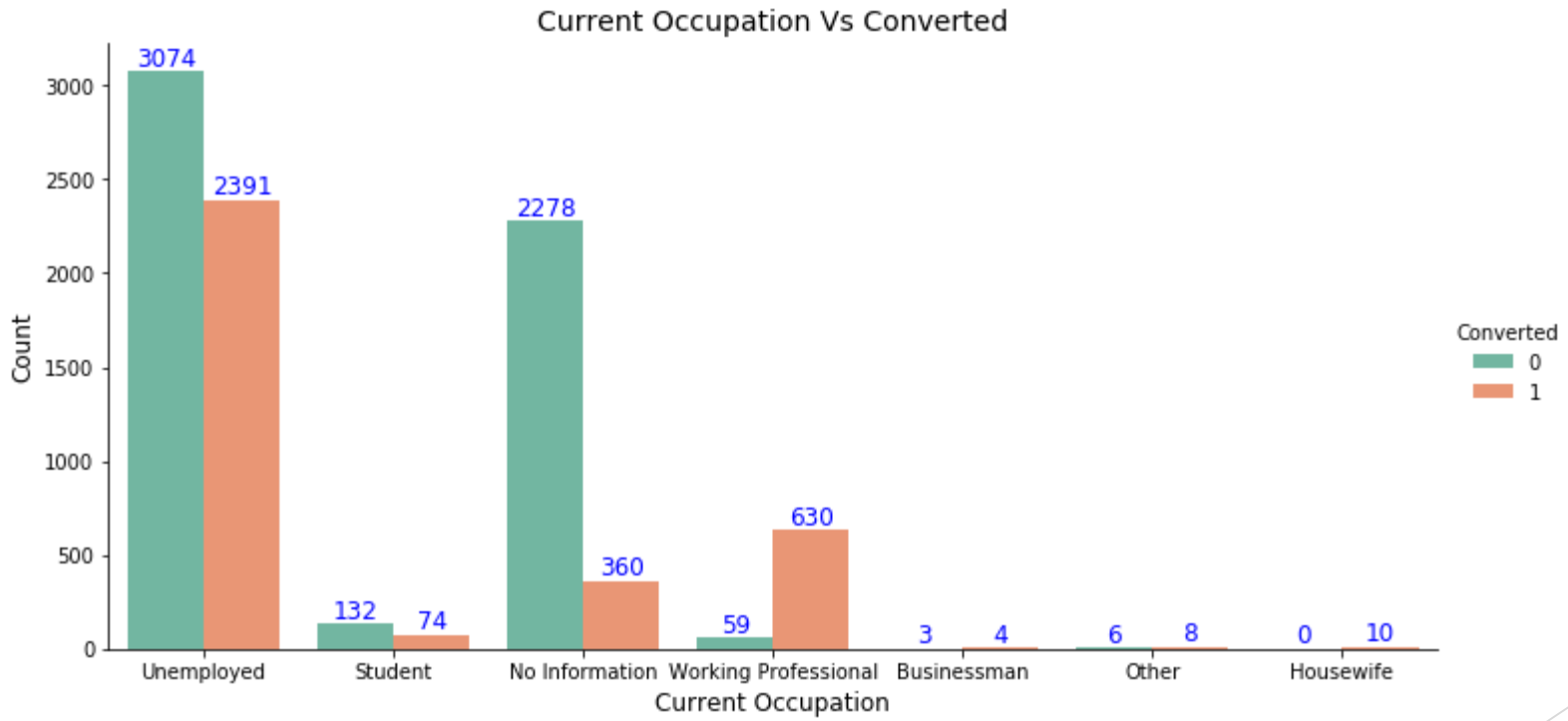


.

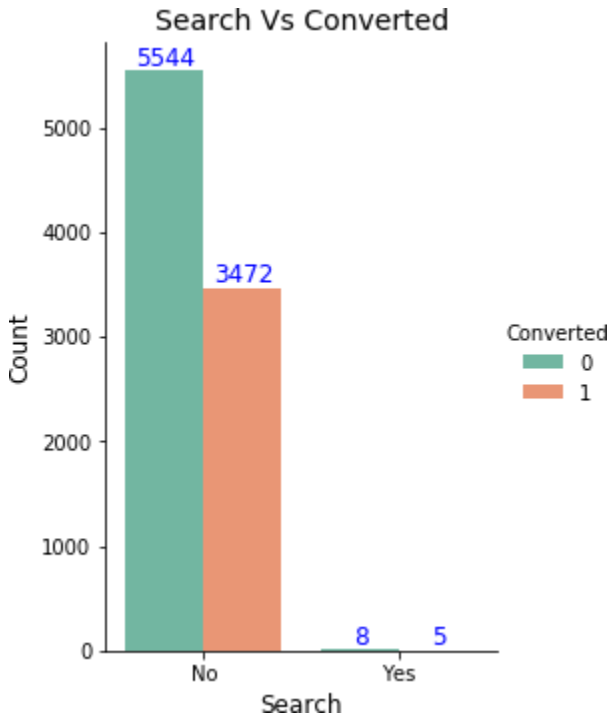
- The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit



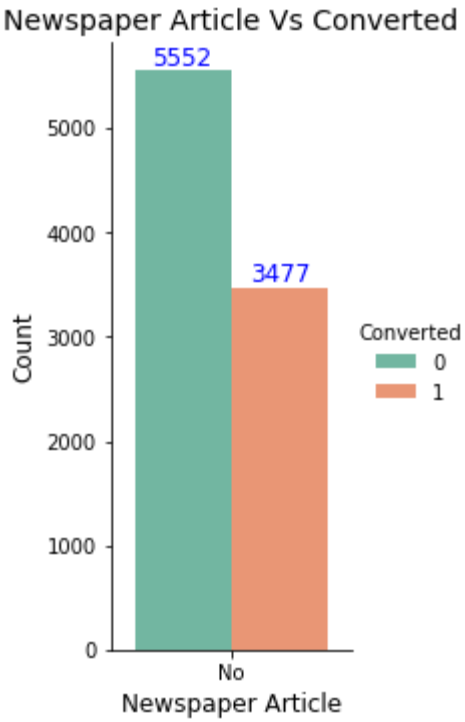
- More conversion happend with people who are unemployed. It can also be noticed from the above data that - Out of 7 business men, 4 got converted - Out 10 housewives, all 10 leads got converted.



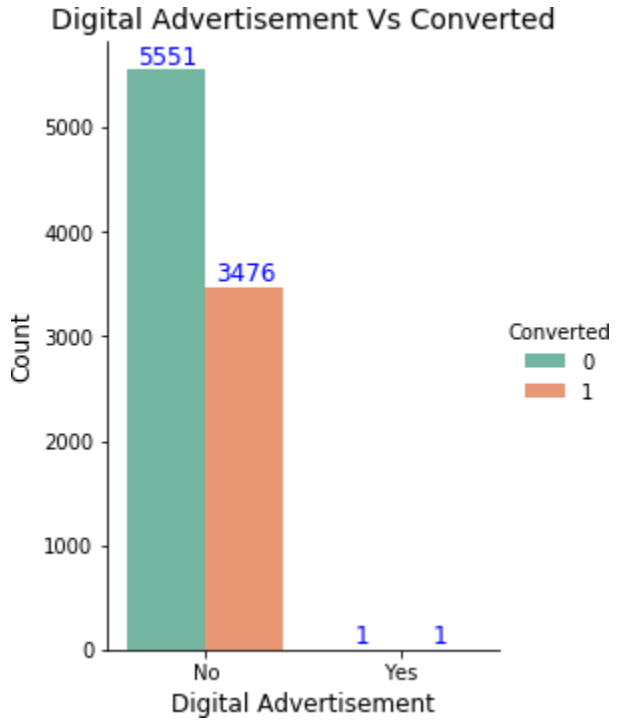
- Conversion rate is high on leads who are not through search



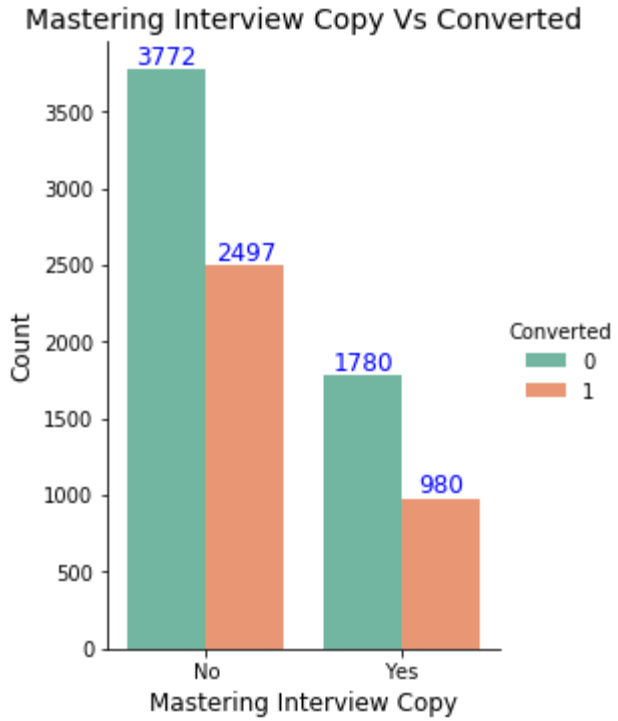
- Since "Newspaper Article" column now has only one value for all rows - "No" , it is safe to drop this column



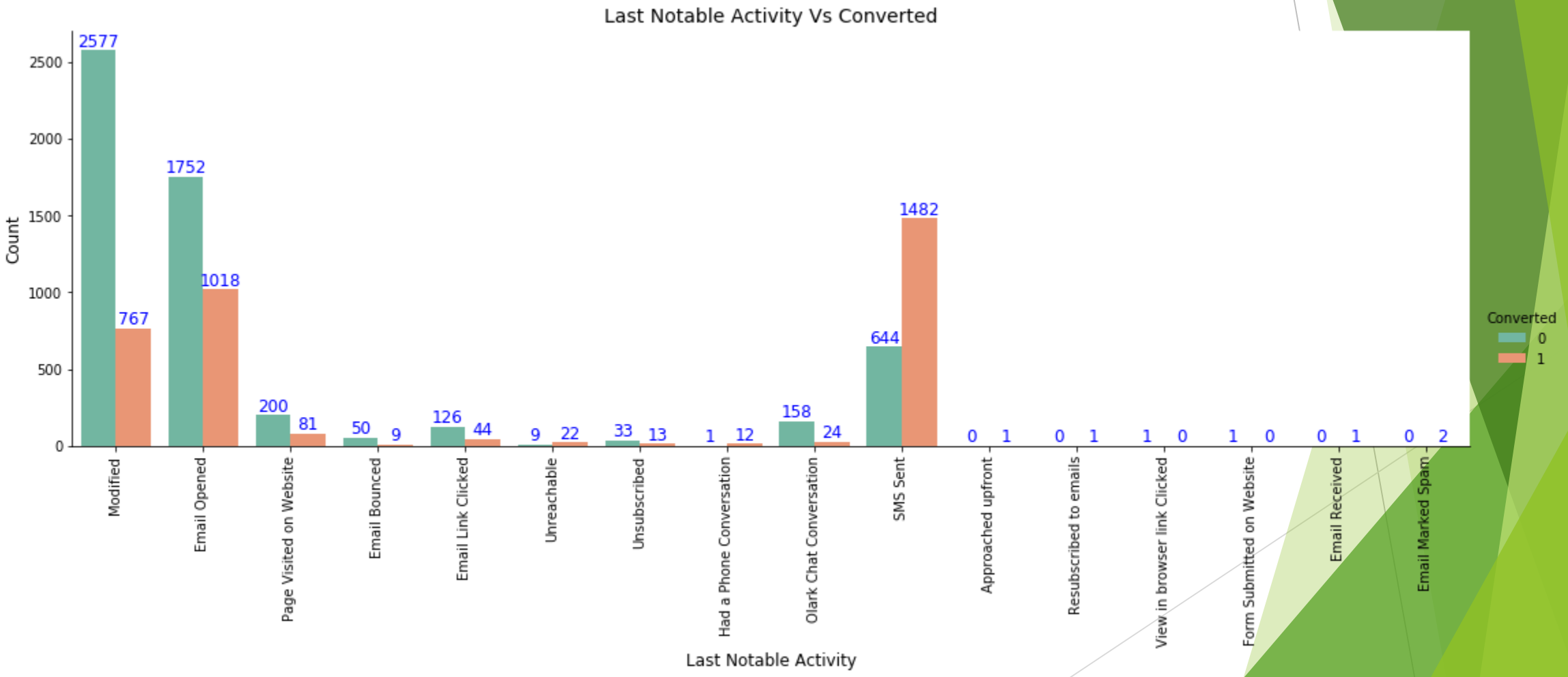
- It can be noticed above that there were 2 leads that came from digital advertisement of which one lead got converted



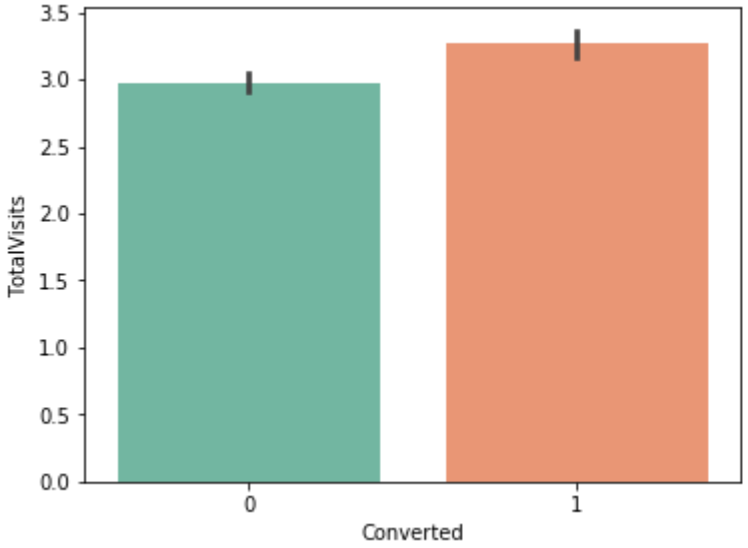
- Conversion rate is high on leads who do not want a free copy of Mastering Interviews

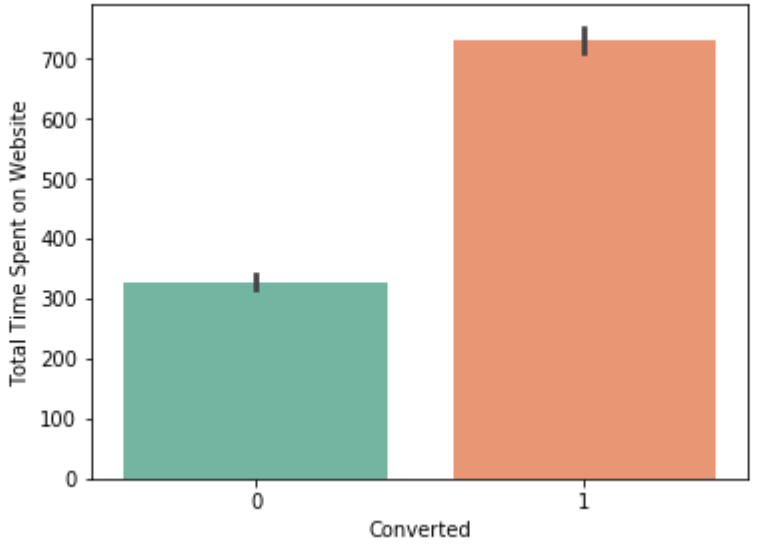


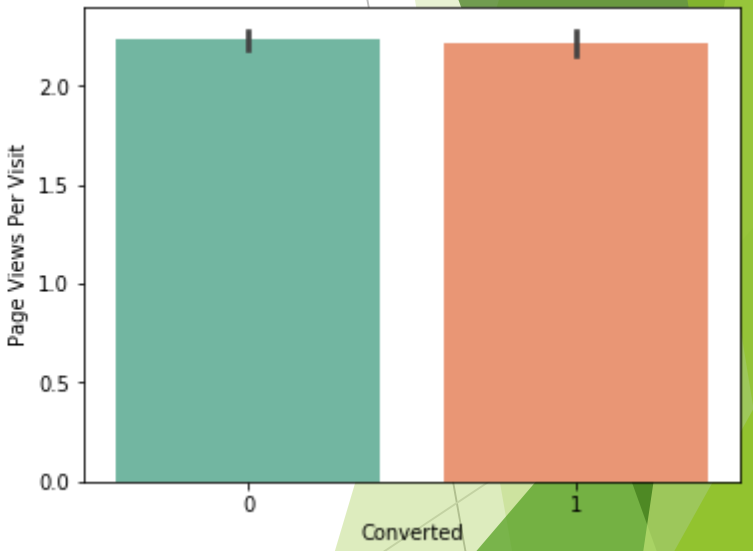
- It can be noticed that the conversion rate is high for "SMS Sent"



- The conversion rated were high for Total Visits, Total Time Spent on Website and Page Views Per Visit





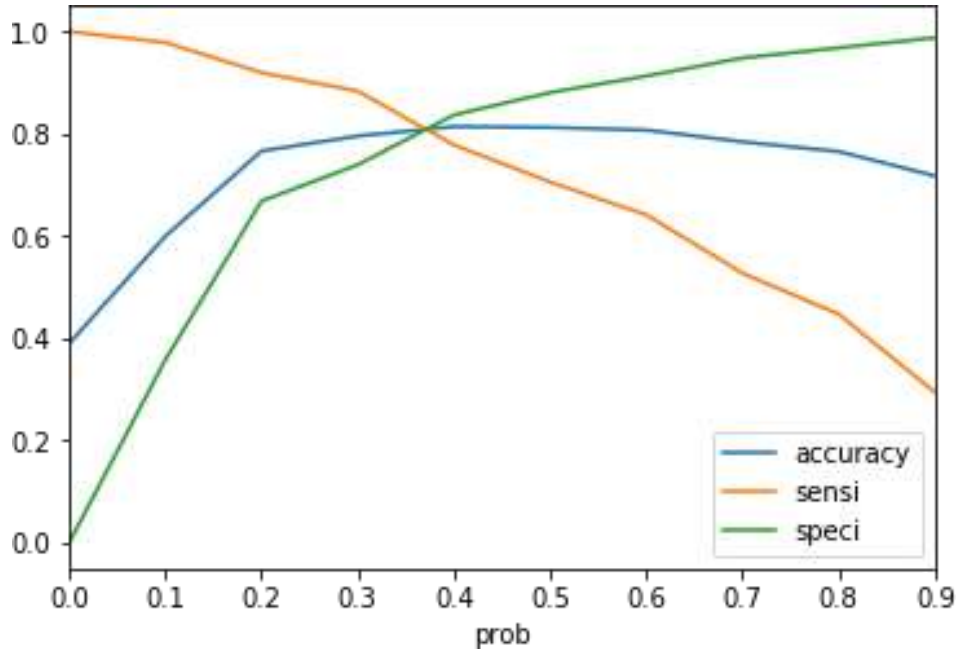


Variables Impacting the Conversion Rate

- **Do Not Email**
- **TotalVisits**
- **Total Time Spent on Website**
- **Page Views Per Visit**
- **LeadOrigin_Lead Add Form**
- **LeadSource_Olark Chat**
- **LeadSource_Welingak Website**
- **LastActivity_Email Opened**
- **LastActivity_Olark Chat Conversation**
- **LastActivity_SMS Sent**
- **CurrentOccupation_No Information**
- **CurrentOccupation_Working Professional**
- **LastNotableActivity_Had a Phone Conversation**
- **LastNotableActivity_Modified**
- **LastNotableActivity_Unreachable**

Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity



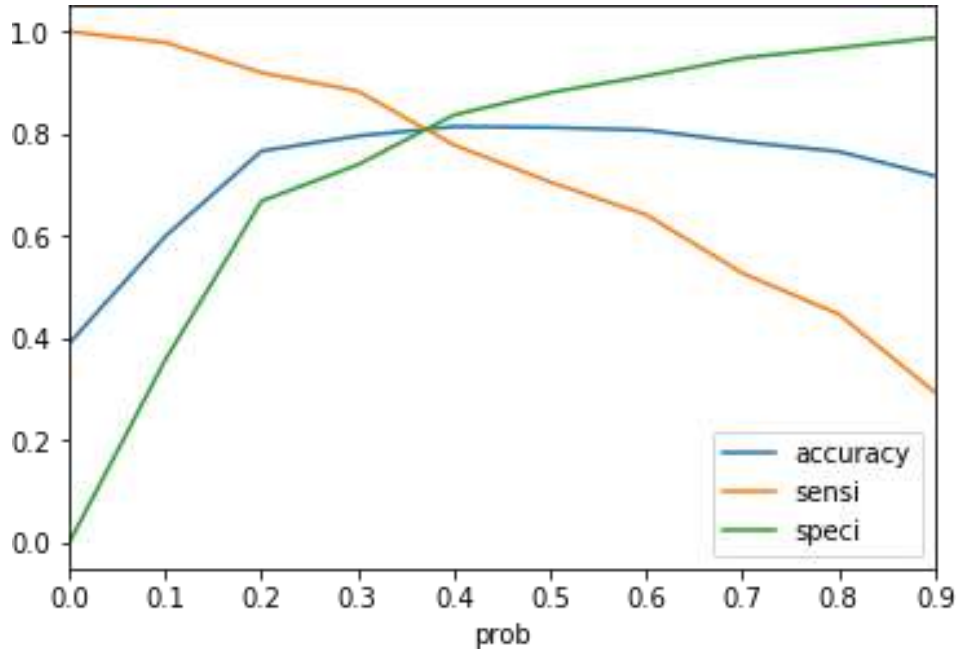
Confusion Matrix

3166	692
491	1971

- Accuracy – 81.28%
- Sensitivity - 80.05 %
- Specificity – 82.06 %
- False Positive Rate – 17.9 %
- Positive Predictive Value – 74.01 %
- Positive Predictive Value – 86.57%

Model Evaluation - Sensitivity and Specificity on Test Data Set

The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity



Confusion Matrix

1393	301
203	812

- Accuracy – 81.39%
- Sensitivity - 80 %
- Specificity – 82.23 %

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the Optimal-cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 81%, 80% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is 80%
- Hence overall this model seems to be good.