

# Lead Scoring Case Study

By:  
Manoj Kumar

# Problem Statement

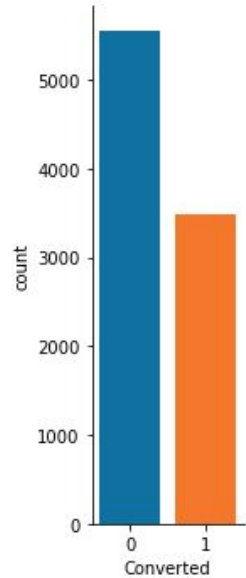
1. For education company named X Education, typical lead conversion is around 30%.
2. Now we need to find the most potential leads, also known as **Hot Leads**.
3. If the set of leads were identified then sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.
4. Target column: Converted

# Strategy

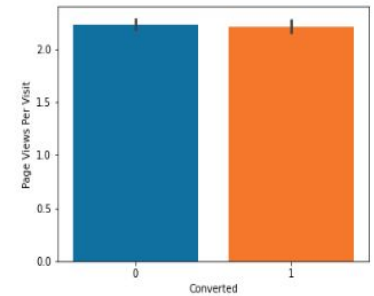
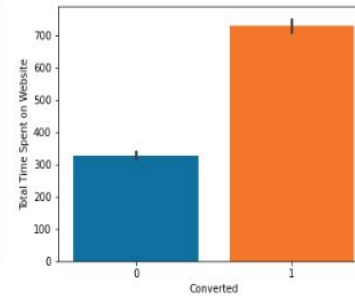
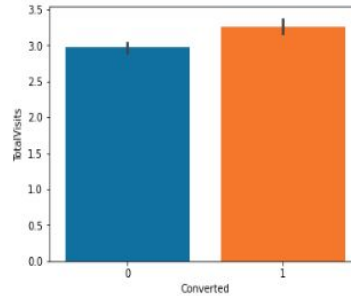
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity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

# Exploratory data analysis

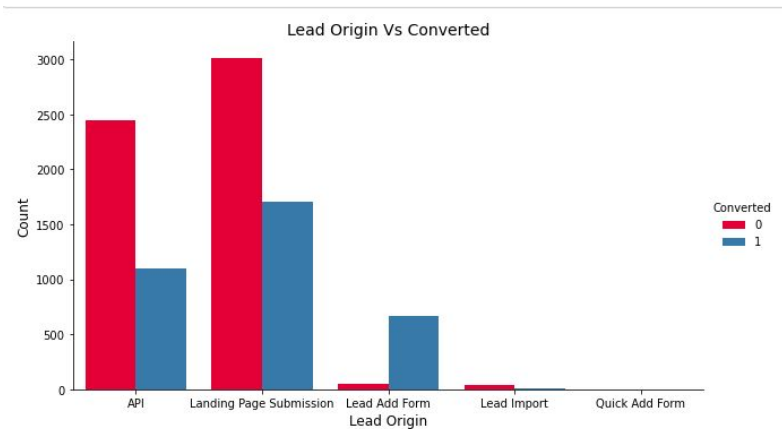
Conversion rate is 39% in total



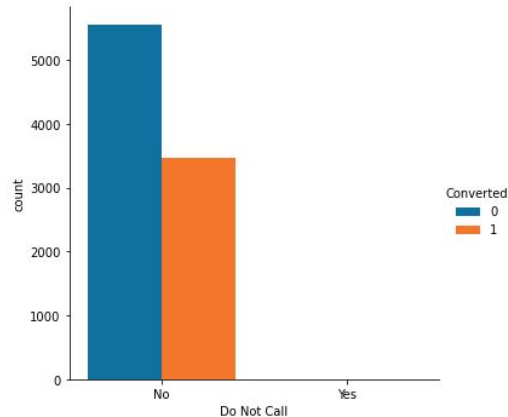
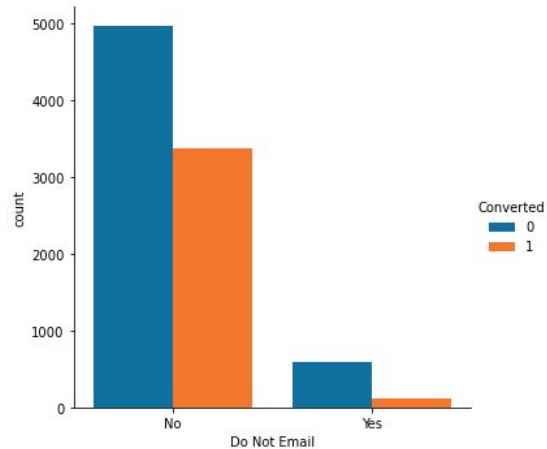
Conversion rates are high for Total Visits, Total time spent on website and Page View per visit.



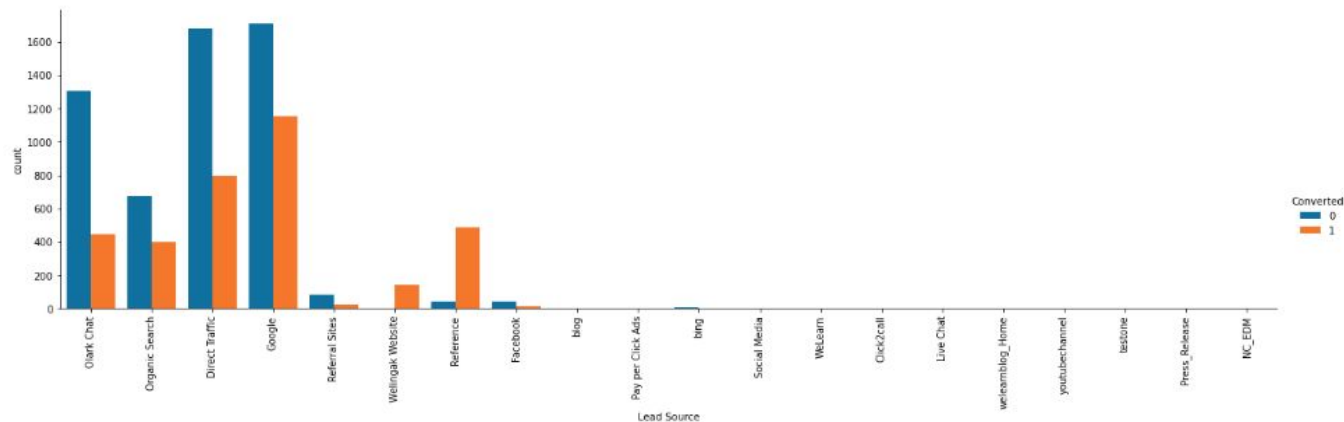
In Lead Origin, maximum conversion happened from Landing Page Submission



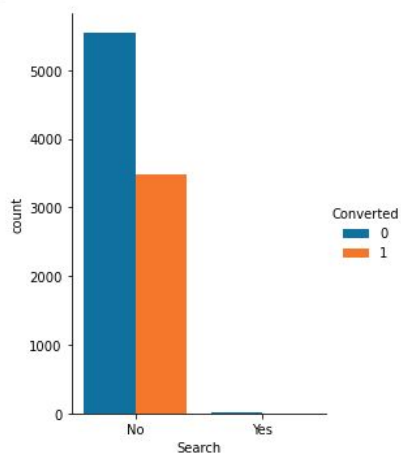
Major conversion has happened from Emails sent and Calls made



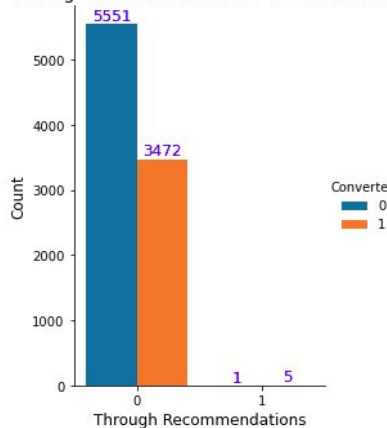
Major conversion in the lead source is from Google



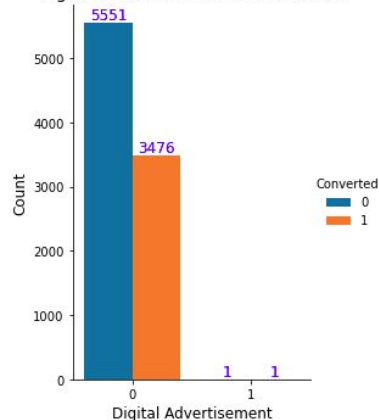
Not much impact on conversion rates through Search, digital advertisements and through recommendations



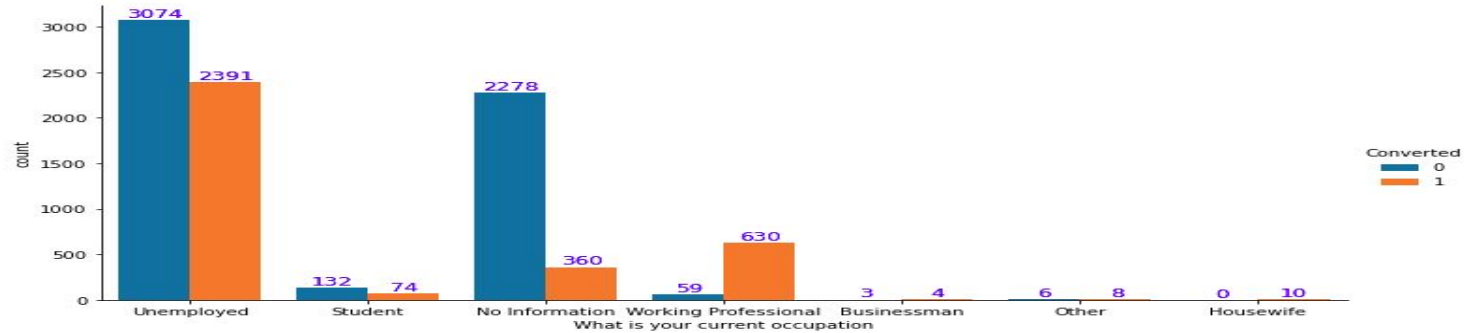
Through Recommendations Vs Converted



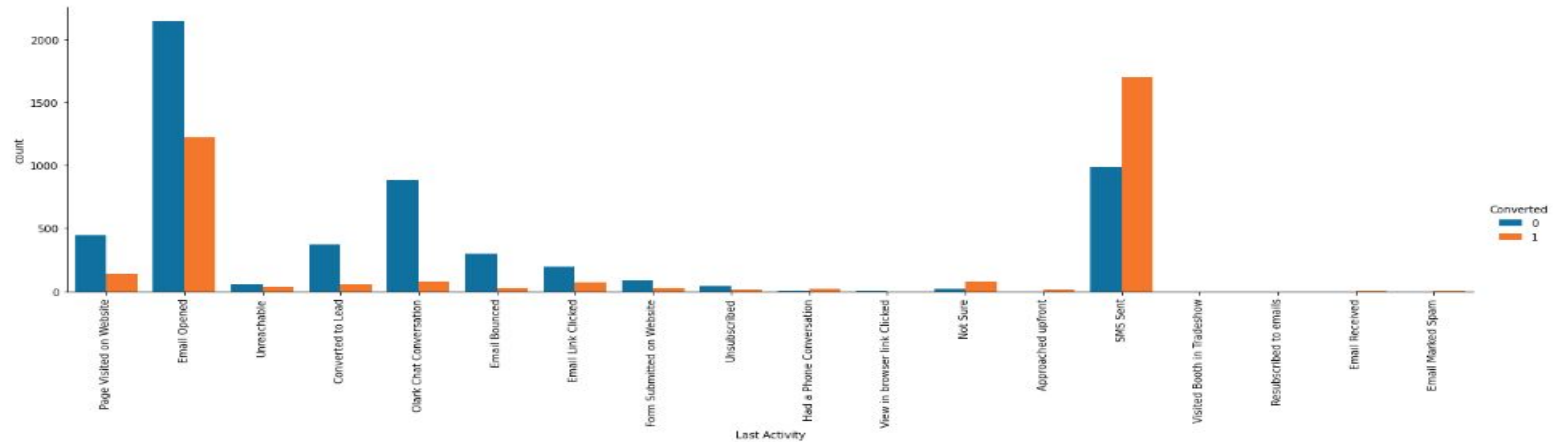
Digital Advertisement Vs Converted



More conversion happened with people who are unemployed



Last Activity value of 'SMS Sent' had more conversion



# Variables Impacting the Conversion Rate

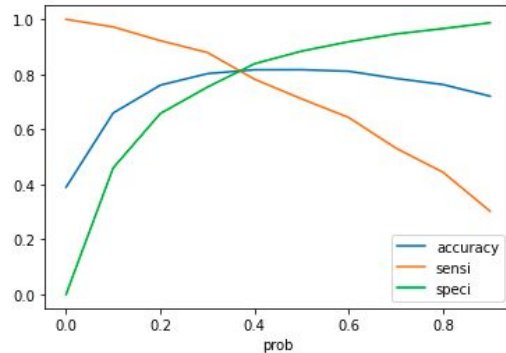
- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent
- Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - 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

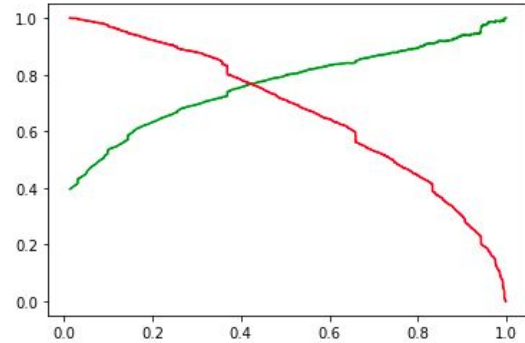


```
[3166, 692]  
[ 491, 1971]
```

- Accuracy- 82%
- Sensitivity- 71%
- Specificity - 88%
- Positive predictive value- 80%

# Model Evaluation- Precision and Recall on Train Dataset

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



# 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%, 79% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- The top 3 variables that contribute for lead getting converted in the model are
  - Total time spent on website
  - Lead Add Form from Lead Origin
  - Had a Phone Conversation from Last Notable Activity