

# Lead Scoring Case Study

By - Arindam Bhattacharya & Satyendra Subedi

# Problem Statement

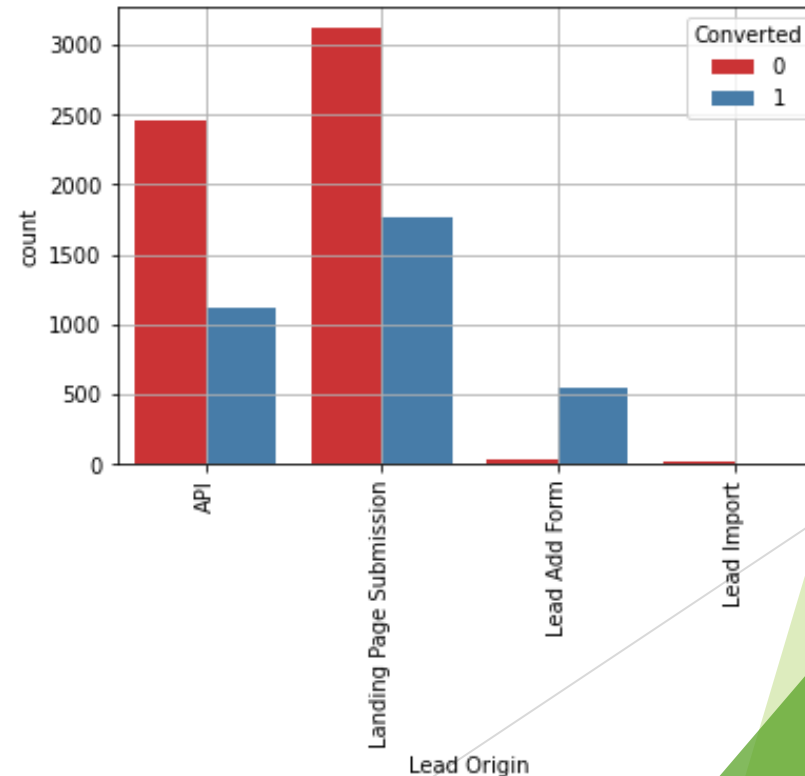
- ▶ An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- ▶ The objective of the case study is to build a model wherein we need to assign a lead score 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%.

# Approach to Analysis

- ▶ Data was imported and cleaned
- ▶ Missing values were dealt with by removing columns with a high number of missing values eg. 'How did you hear about X Education', 'Lead Profile' etc.
- ▶ Some columns 'Magazine', 'Receive More Updates About Our Courses' had only one output 'No'. So they were also removed.
- ▶ The columns 'Search', 'Newspaper Article' had a very high degree of data imbalance, so they were removed from the analysis
- ▶ **Assumption:** Hot Leads are the leads who have a lead score of 80 or above

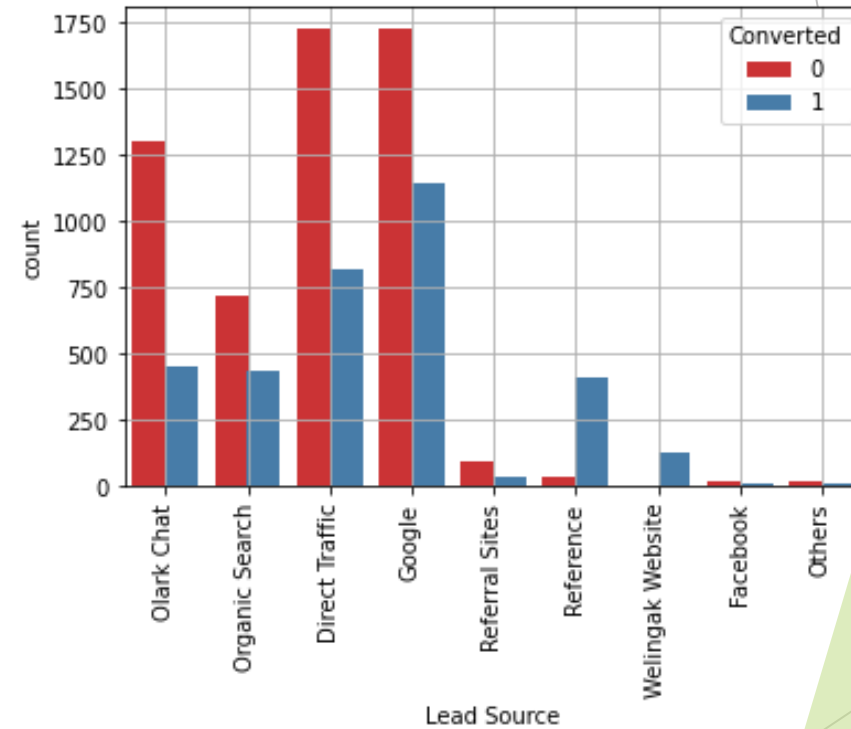
# Exploratory Data Analysis

1. **Converted:** This is the target variable, Indicates whether a leads has been successfully converted (1) or not (0). The leads conversion rate is approximately 38%
2. **Lead origin:**
  1. API and Landing Page Submission have a 30-35% conversion rate but the count of leads originated from them are considerable.
  2. Leads Add Form has more than 90% conversion rate but the count of leads is not very high.
  3. Leads Import has a lower count.

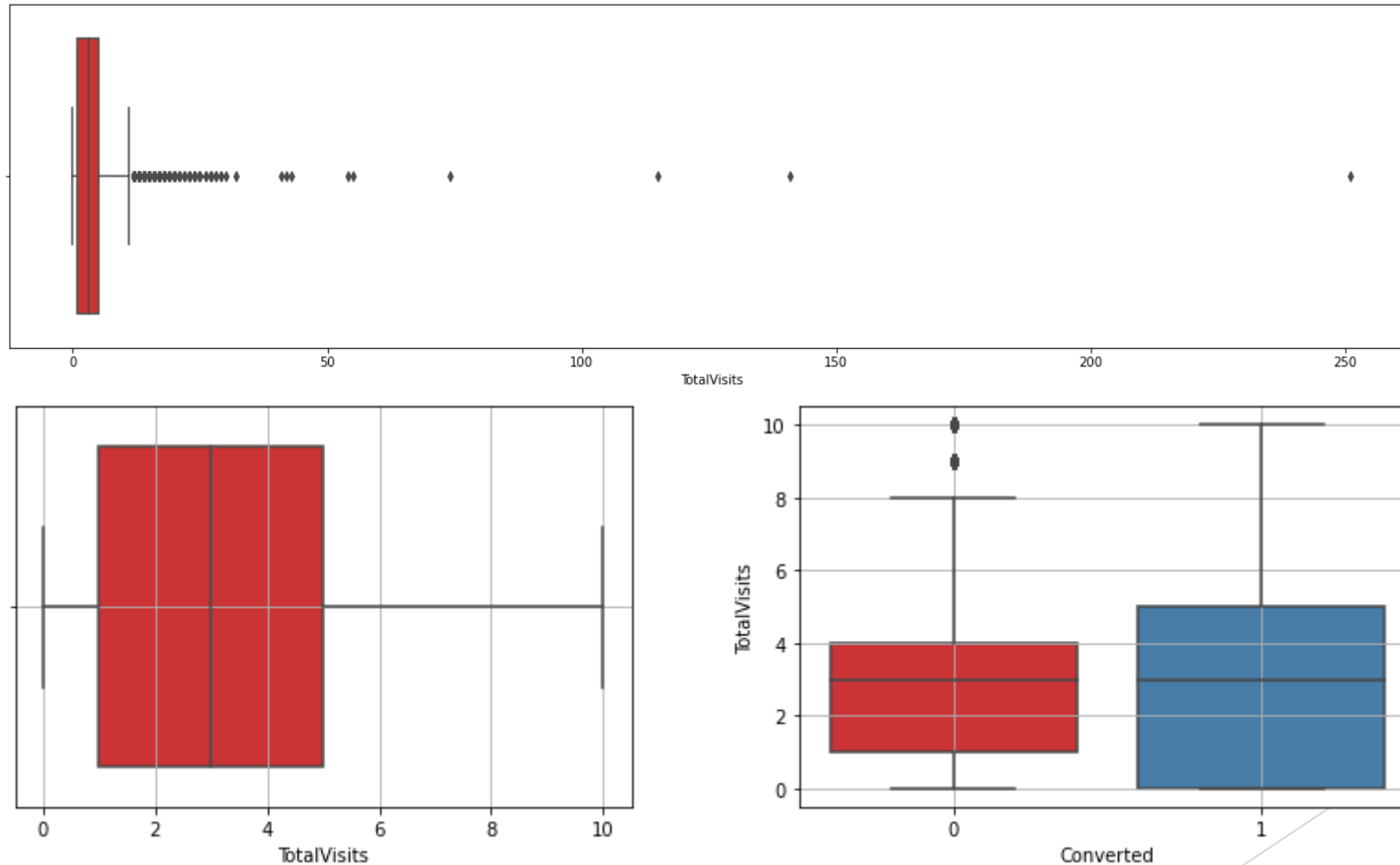


## ► Lead Source

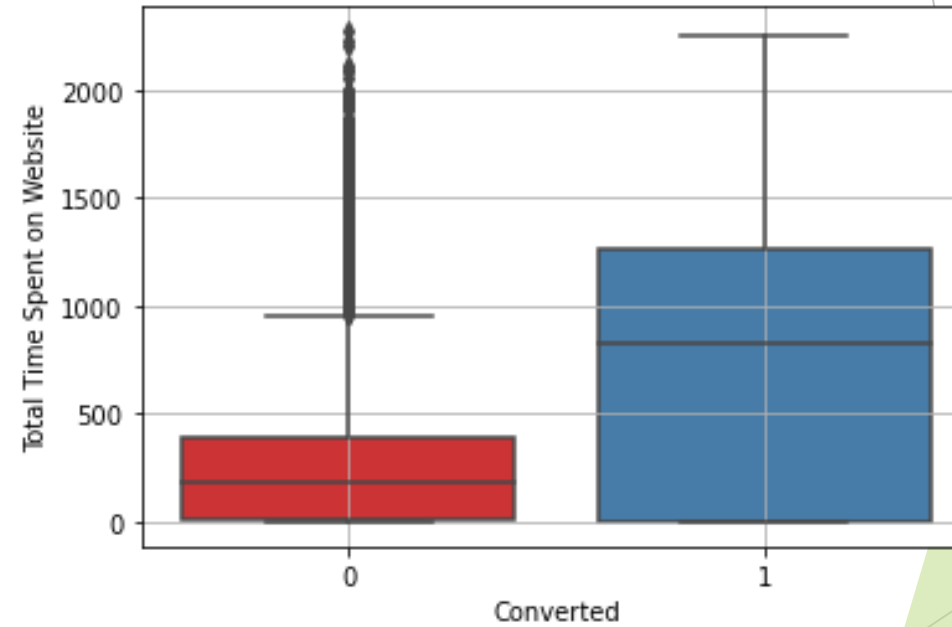
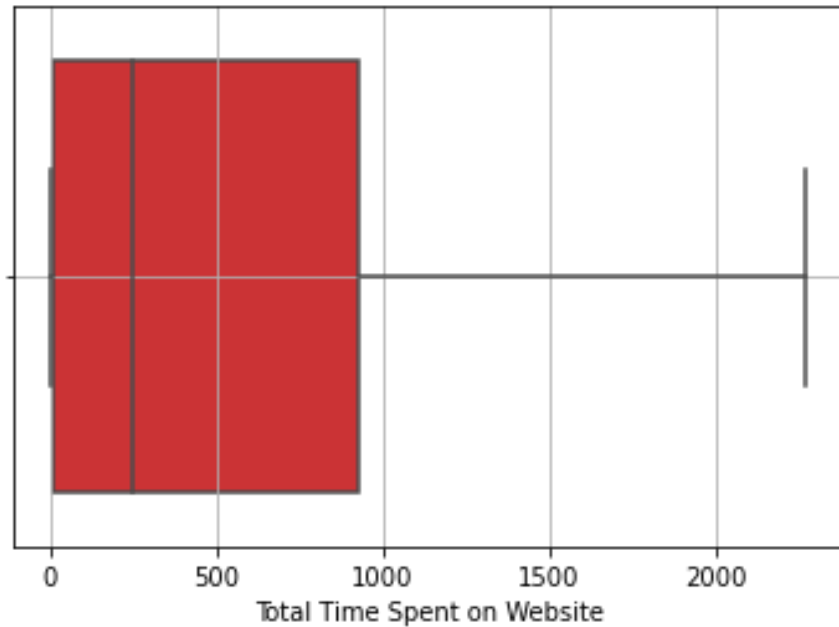
1. Google and Direct traffic generated the maximum number of leads.
  2. Conversion Rate of reference leads and leads through the Welingak website is high.
- **To improve overall lead conversion rate, the focus should be on improving lead conversion from olark chat, organic search, direct traffic, google leads and generate more leads from reference and the Welingak website.**



- **Total Visits:** There are a few data points that are outside the normal range, so the data points were capped to the 95<sup>th</sup> percentile. However, not much can be concluded from this as the median for both converted and non-converted are the same

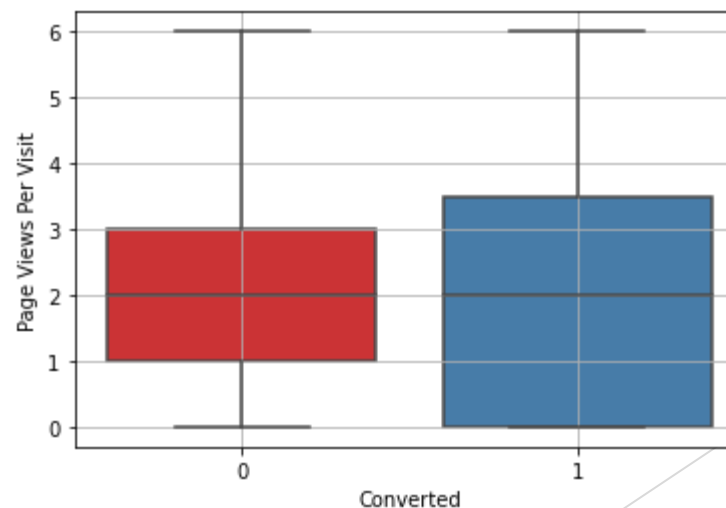
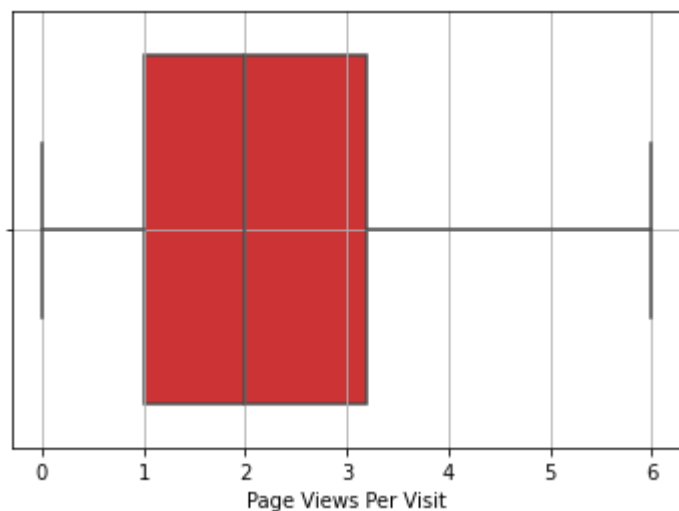
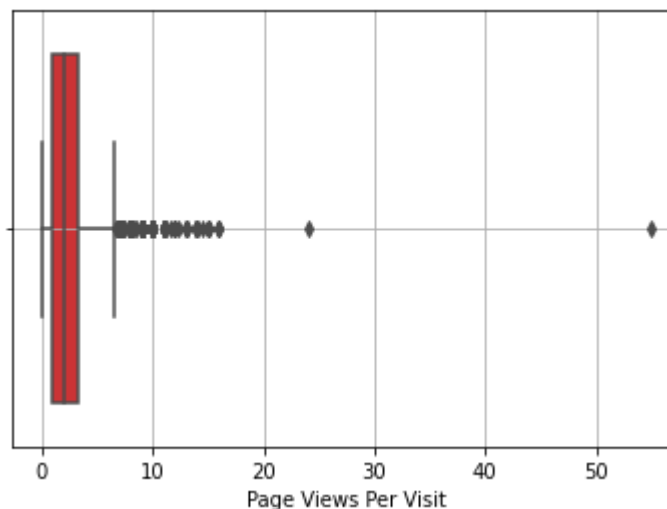


- **Total Time Spent on Website:** Leads spending more time on the website are more likely to be converted.
- So, the website should be made more engaging to make leads spend more time.



- **Page Views Per Visit:**

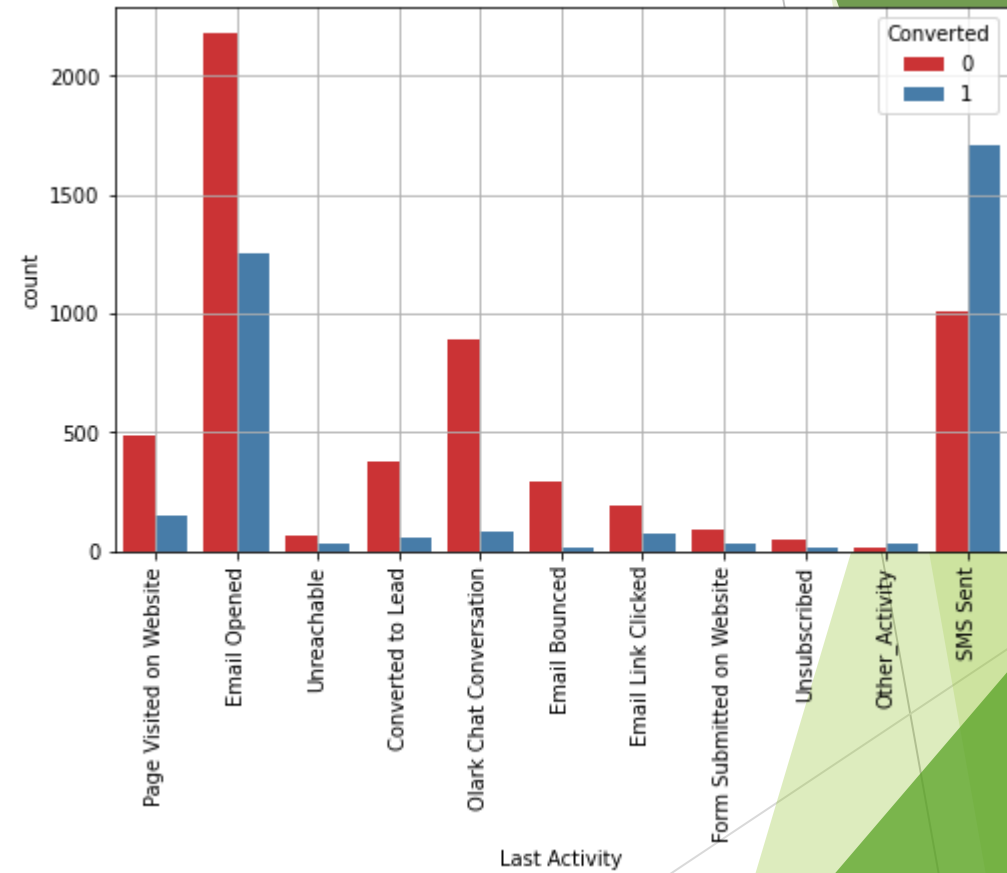
- There were outliers in this variable, so the data has been capped to the 95th percentile
- The median for both converted and non-converted are similar, so not much can be inferred.





## ► Last Activity

1. Most of the leads have 'Email Opened' as their last activity.
2. Conversion rate for leads with last activity as SMS Sent is almost 60%.



# Model Building and Evaluation

- ▶ Data was split for training and testing in a 70:30 ratio
- ▶ The final model (Model No.5) was selected which had 11 variables
- ▶ Multicollinearity in the model was checked using Variance Inflation Factor
- ▶ The optimal cut-off point for the sensitivity-specificity check was calculated and 0.34 was chosen.

# Results

The values obtained after Training & Testing:

Train Dataset:

- Accuracy: 81.16 %
- Sensitivity: 81.92 %
- Specificity: 80.69 %

Test Dataset:

- Accuracy: 80.60 %
- Sensitivity: 80.58 %
- Specificity: 80.62 %

Thus, we have achieved our goal of getting a ballpark of the target lead conversion rate of 80%. The model seems to predict the Conversion Rate very well and is not overfitting.

# Results

The most important features of the final model are:

	Coefficients
Lead Origin_Lead Add Form	3.817505
What is your current occupation_Working Professional	2.607298
Last Activity_Other_Activity	2.173572
Last Activity_SMS Sent	1.293022
Total Time Spent on Website	1.096510
Lead Source_Olark Chat	1.076184
const	-0.108721
Last Notable Activity_Modified	-0.902378
Last Activity_Olark Chat Conversation	-0.952169
Specialization_Others	-1.118892
Lead Origin_Landing Page Submission	-1.125353
Do Not Email	-1.482795

# Recommendations

The company should make calls to the below leads as they are more likely to get converted:

- ▶ Coming from the lead sources "Welingak Websites" and "Reference".
- ▶ Those who are "working professionals"
- ▶ Those who spend "more time on the websites"
- ▶ Coming from the lead sources "Olark Chat"
- ▶ Whose last activity is "SMS Sent"

The company should not make calls to the below leads as these are not likely to get converted

- ▶ Whose last activity is "Olark Chat Conversation".
- ▶ Those having lead origin as "Landing Page Submission".
- ▶ Those having Specialization as "Others".
- ▶ Those who chose "Do not Email" as "yes".

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect. The shapes are layered, with some appearing more prominent than others, and they extend from the right and bottom edges towards the center.

Thank You!