Lead Score Case Study

Submitted by :

Lavanya Ravilla

Vasundhara Lakshmi

Aman Mehrotra

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.

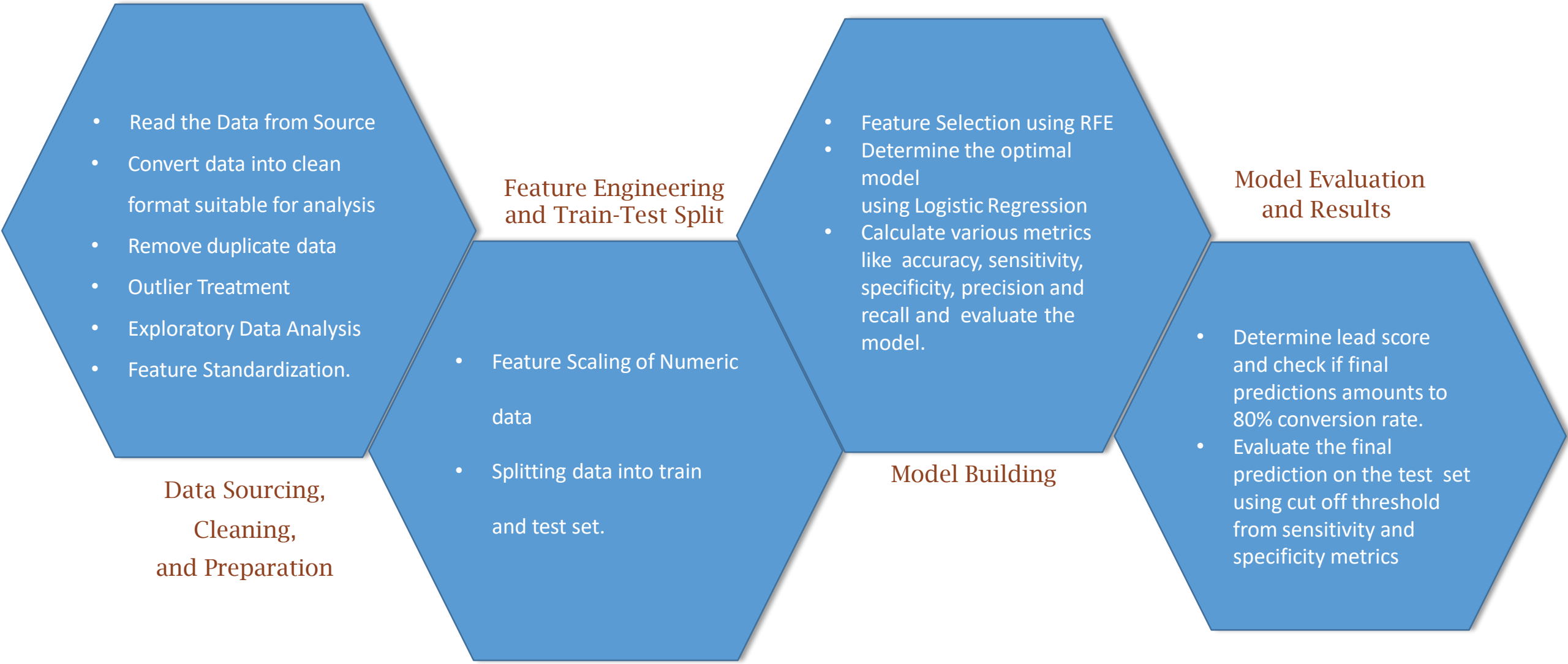
- The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

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

Step by step strategy to handle data preparation and modelling:

- 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.

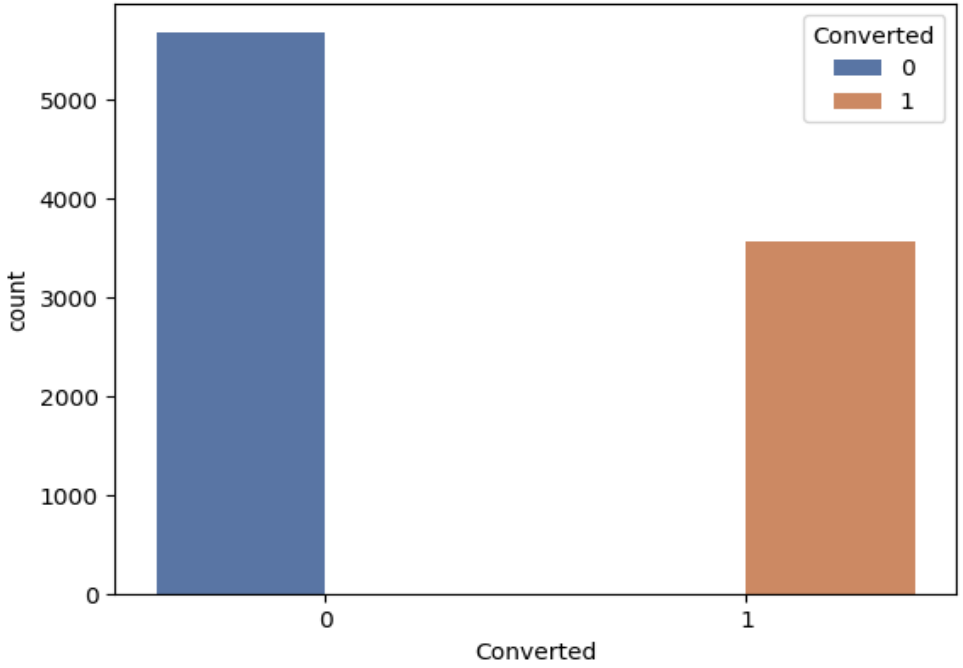
Methodology deployed:

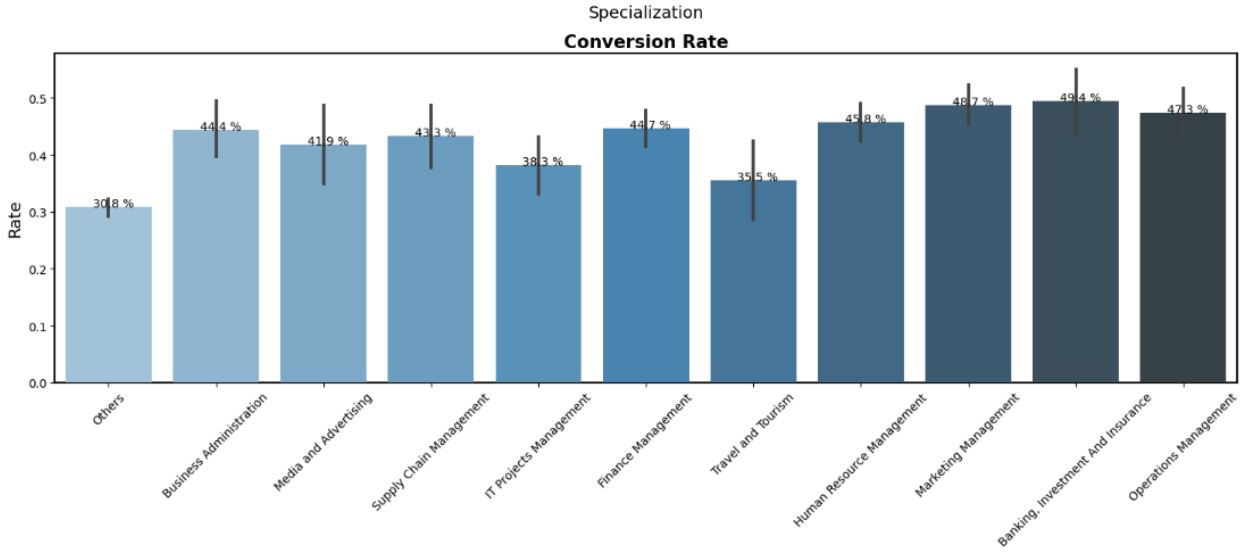


Insights from EDA:

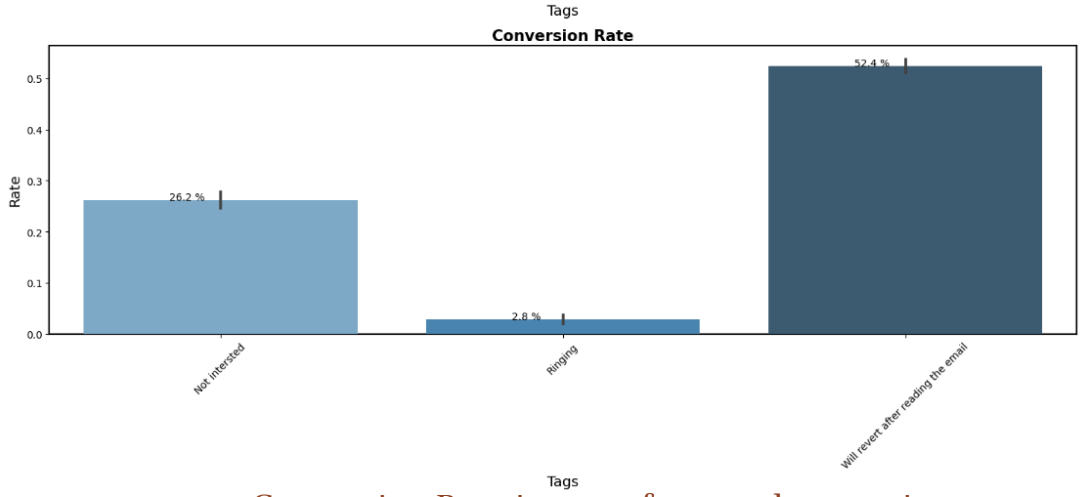
Conversion Rate is close to 39% as per the current data

```
Values by absolute number
 0    5679
1    3561
Name: Converted, dtype: int64
Values by %
 0    61.461039
1    38.538961
Name: Converted, dtype: float64
```



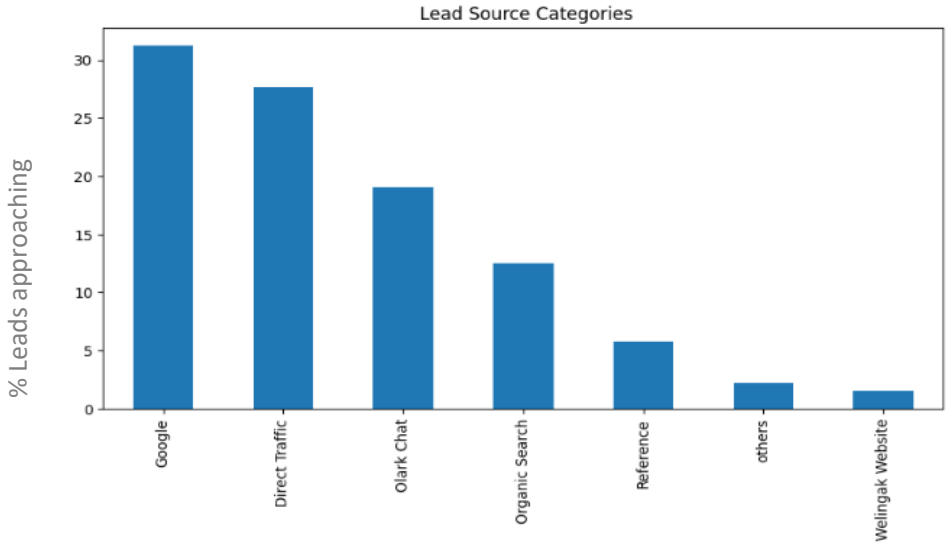


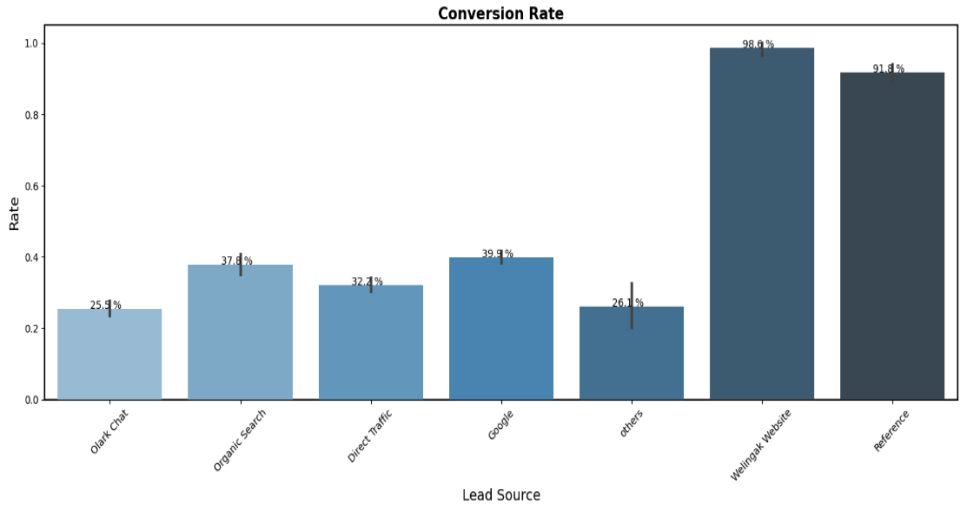
Conversion Rate is more for people opting for specialization in marketing, finance and banking sectors

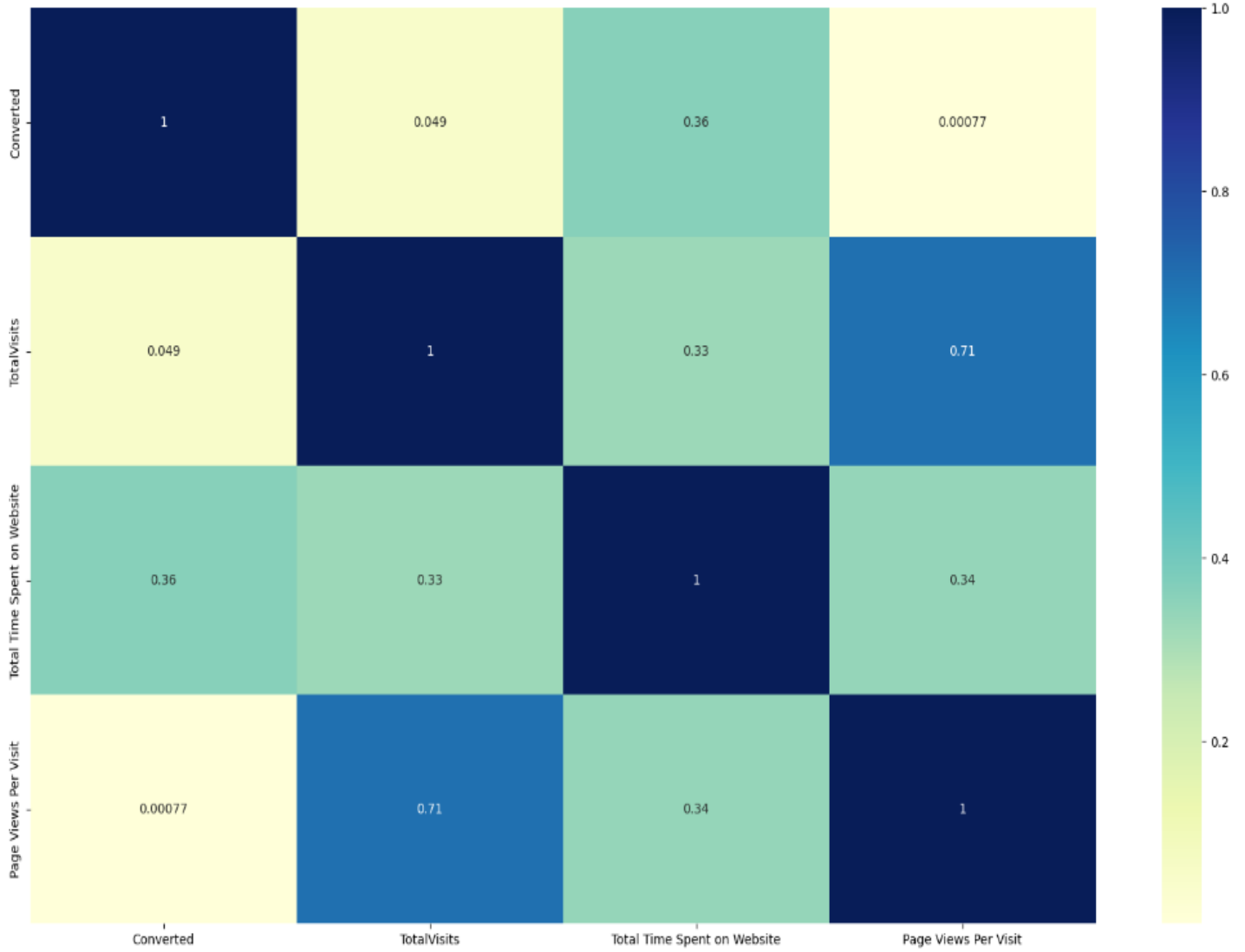


Conversion Rate is more for people reverting with email

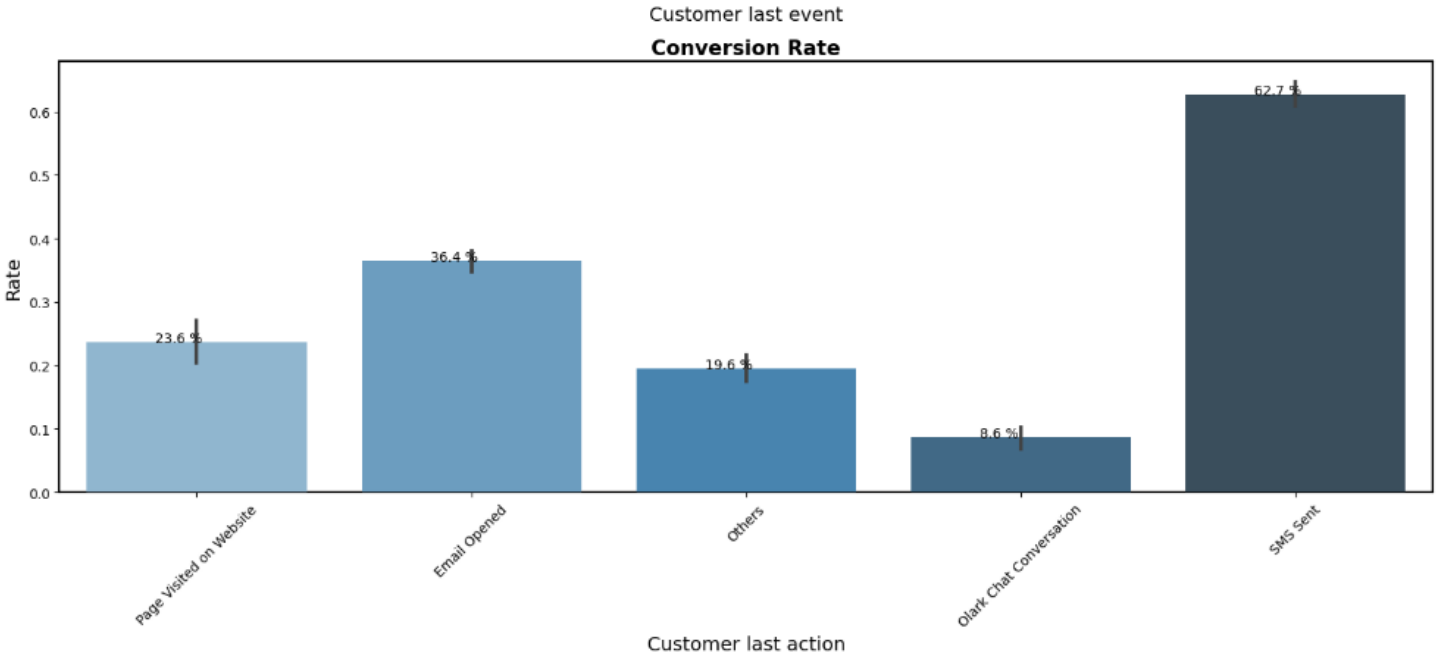
Major lead source is Google but most conversions happened through Welingak Website and References





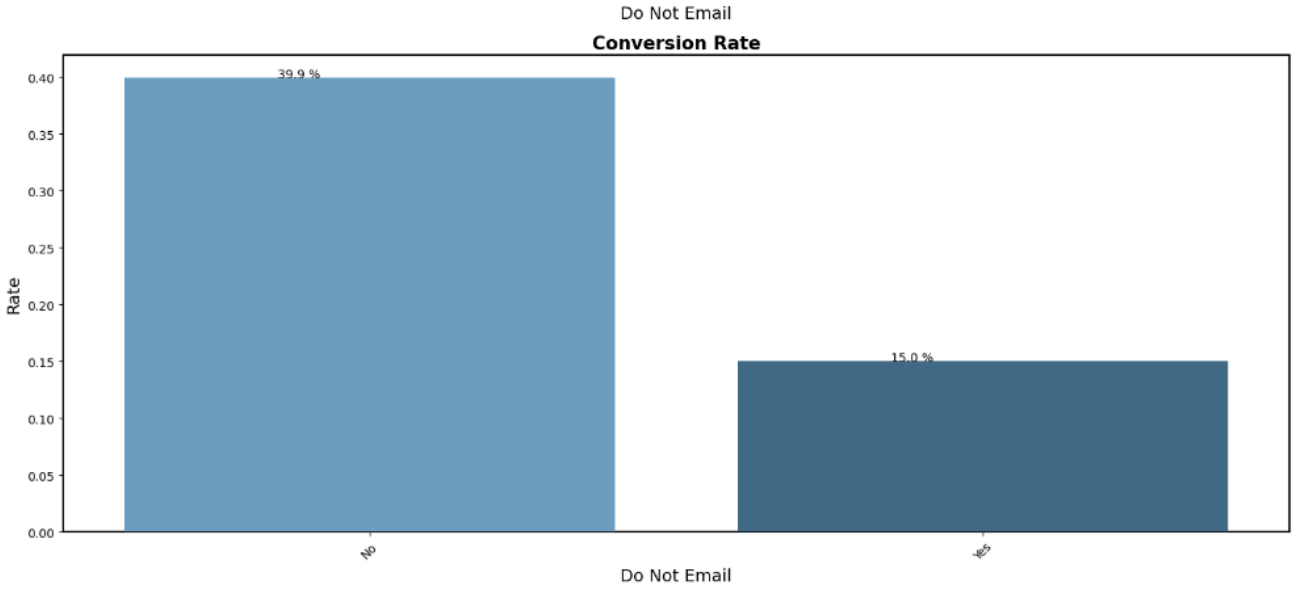


Leads are more likely to get converted if they are spending more time on website



Conversion rate is more for users with last action as SMS sent

Conversion rate is more for users who do no revert back with email

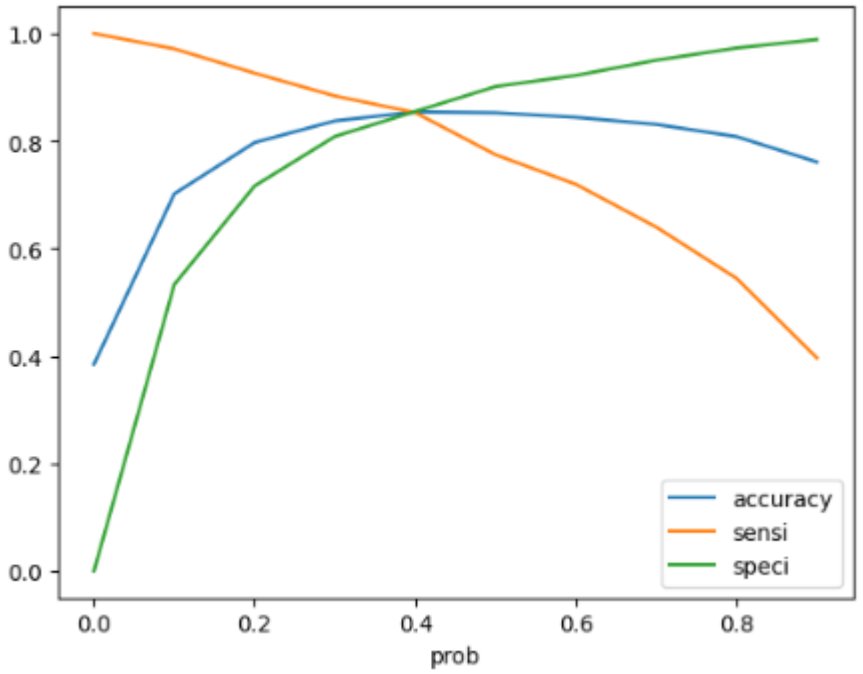


Variables impacting the conversion rate :

- TotalVisits,
- Do Not Email
- Total Time Spent on Website
- Lead Source_Olark Chat,
- Lead Source_Reference,
- Lead Source_Welingak Website,
- Last Activity_Olark Chat Conversation,
- Last Activity_Others,
- Specialization_Travel and Tourism,
- What is your current occupation_Working Professional,
- what matters most to you in choosing a course_Other,
- Tags_Ringing,
- Last Notable Activity_SMS Sent

Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.4 based on Accuracy, Sensitivity and Specificity



Confusion Matrix

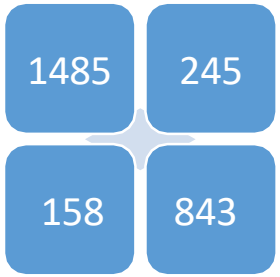
3352	569
361	2090

Evaluation Metrices

- Accuracy – 85.4%
- Sensitivity – 85.27 %
- Specificity – 85.48 %
- False Positive Rate – 14.51 %
- Predicted Conversions – 85.27%

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix



Evaluation Metrices

- Accuracy – 85.20%
- Sensitivity – 84.21 %
- Specificity – 85.83 %
- False Positive Rate – 14.16 %
- Predicted Conversions – 84.21%

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 85%, 84% and 85% 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 85% (in train set) and 84% in test set
- The top 3 variables that contribute for lead getting converted in the model are
  - Total time spent on website
  - Lead Source
  - What is your current occupation

Hence overall this model seems to be working up to the expectations.