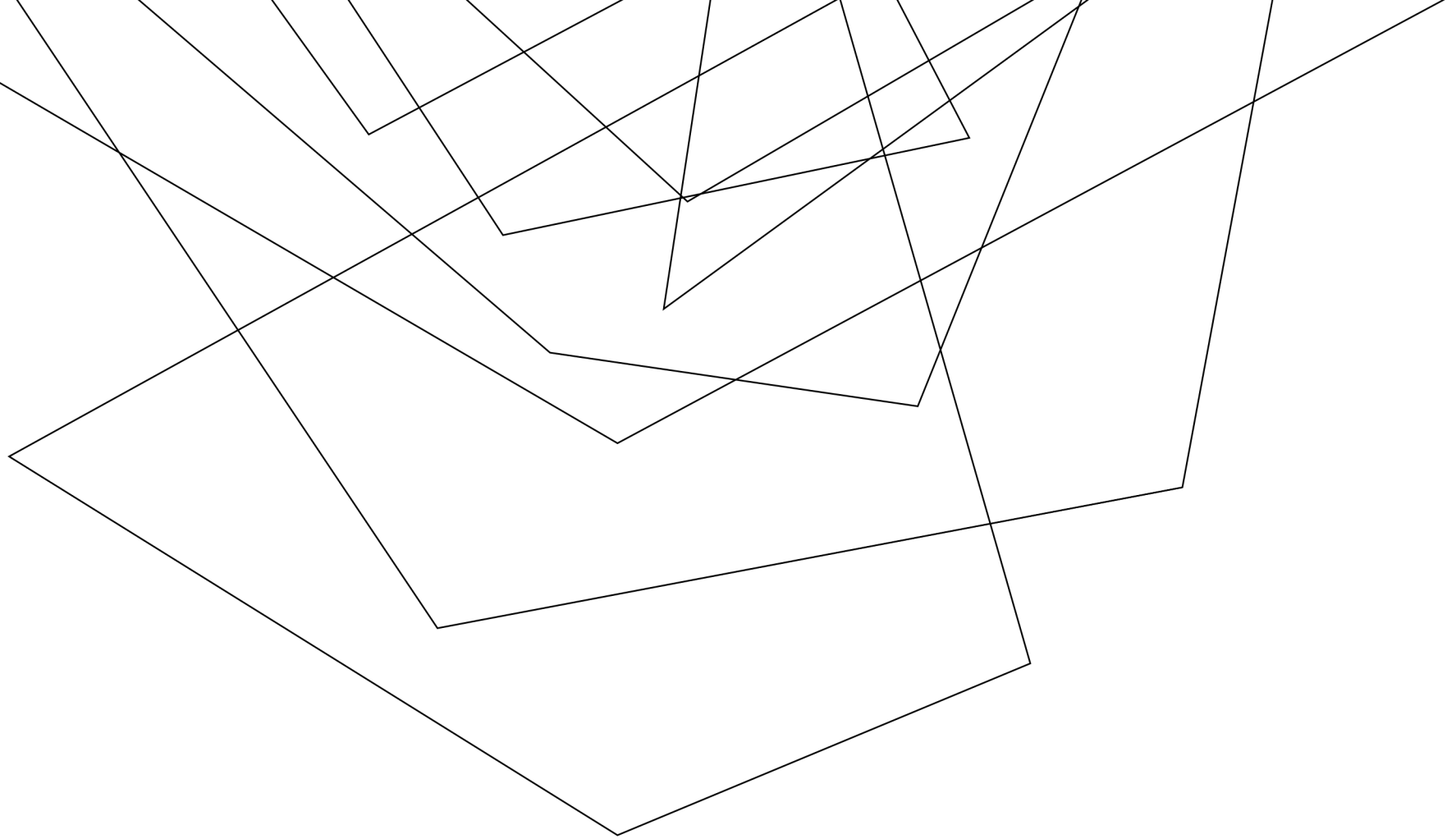


LEAD SCORE CASE STUDY

BUSINESS UNDERSTANDING

Problem Statement :

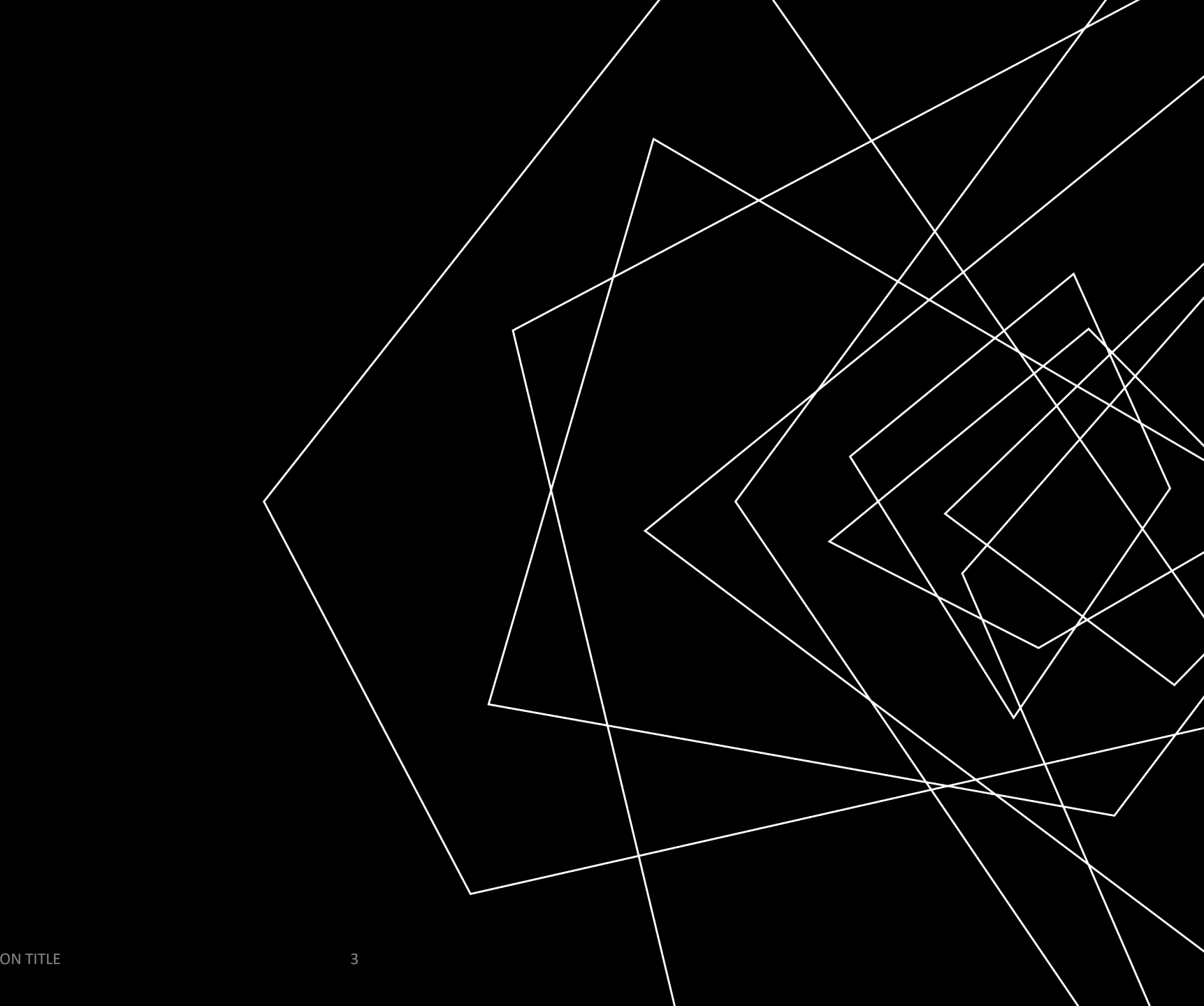
- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form 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%.

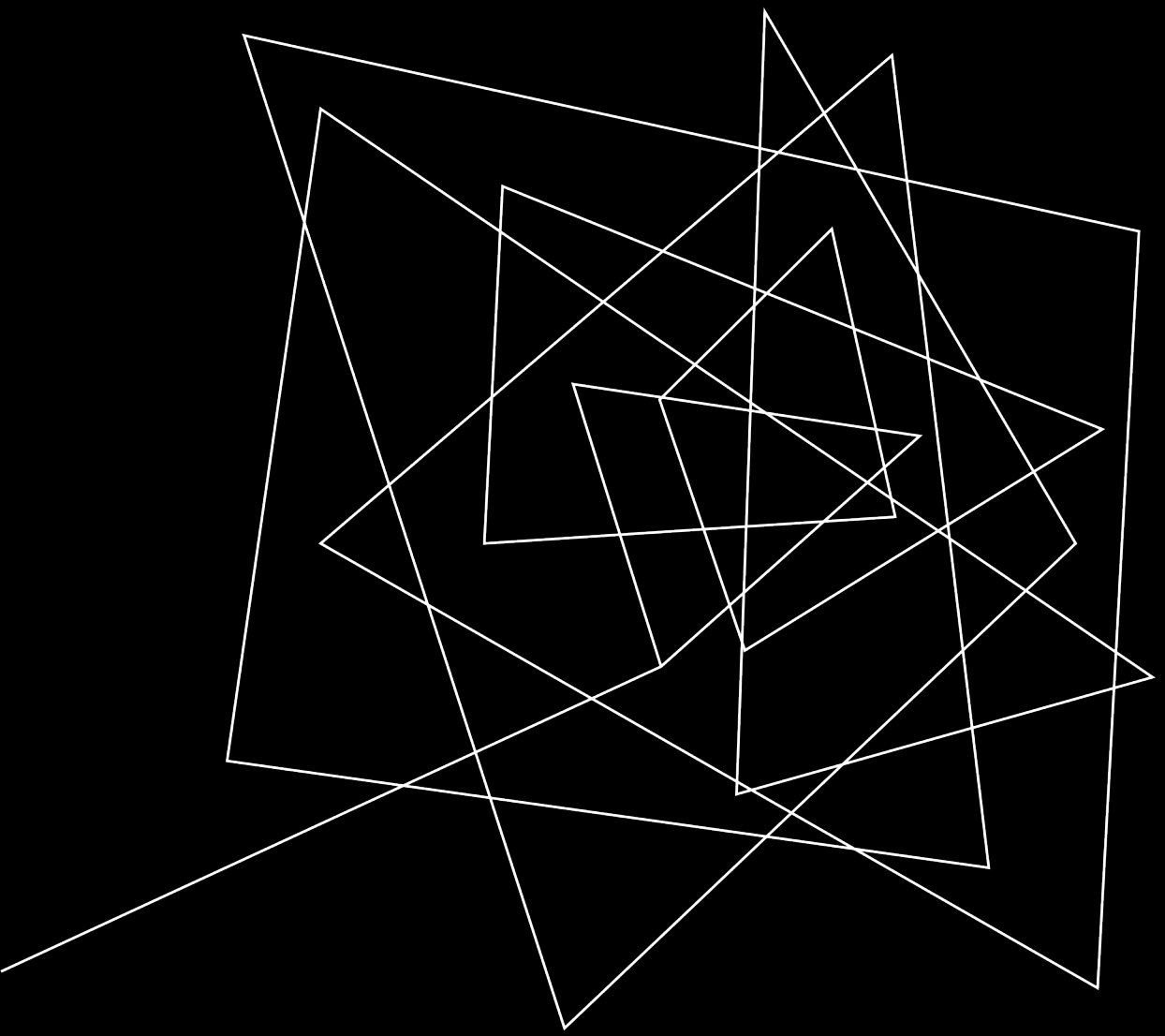
Business Goal:

- 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%

PROBLEM SOLVING STEPS

1. Data Loading and Cleaning
2. EDA
3. Data Preparation
4. Modeling
5. Prediction





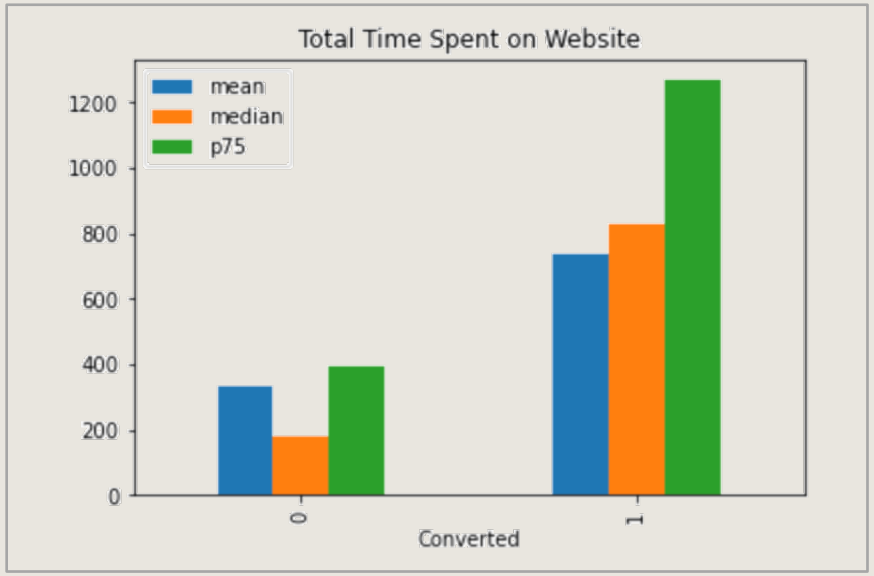
EXPLORATORY DATA ANALYSIS

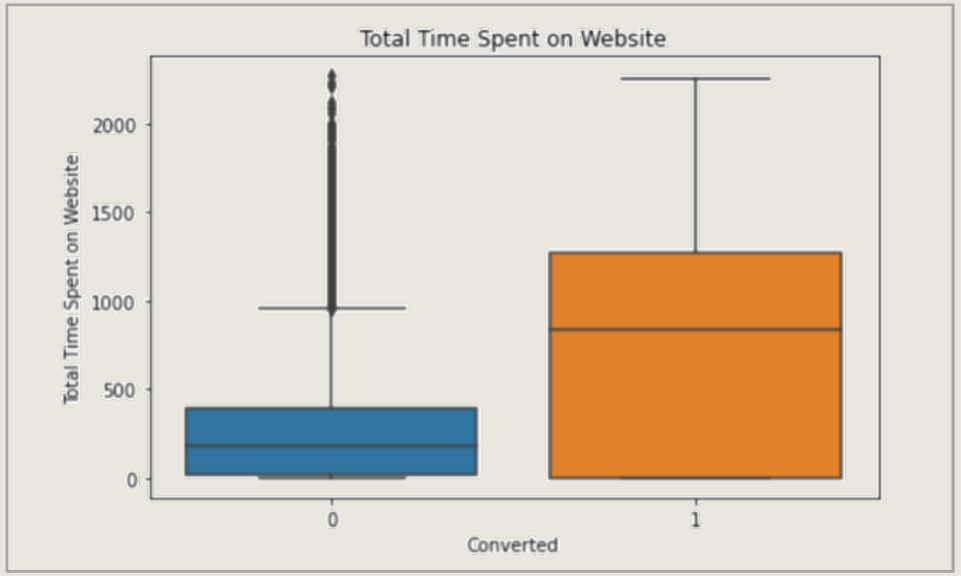
(EDA)

ANALYSIS OF CONTINUOUS DATA

Analysis

Looking at the visualization, we can clearly see that 'Total Time Spent on Website' is directly impacting conversions. More time the users spend on the Website, the higher is the conversion rate





Suggestion

It would be great if X Education could improve their website to increase total time spent

ANALYSIS OF CATEGORICAL DATA

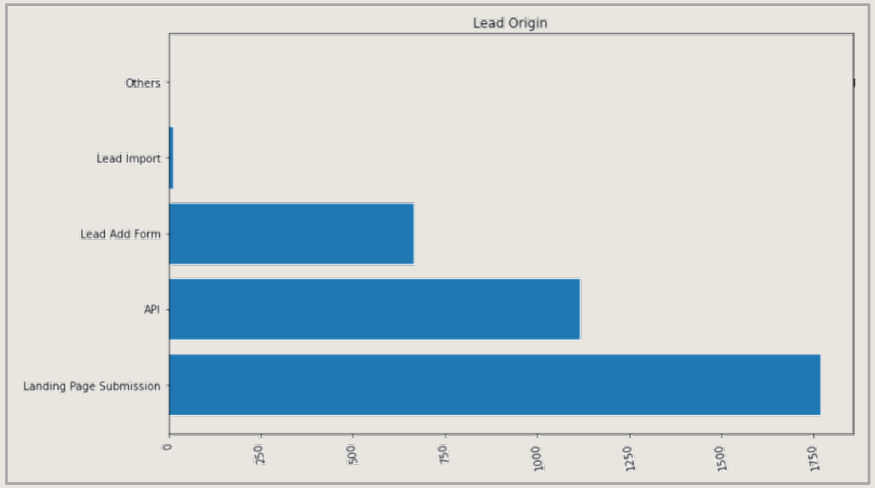
Analysis

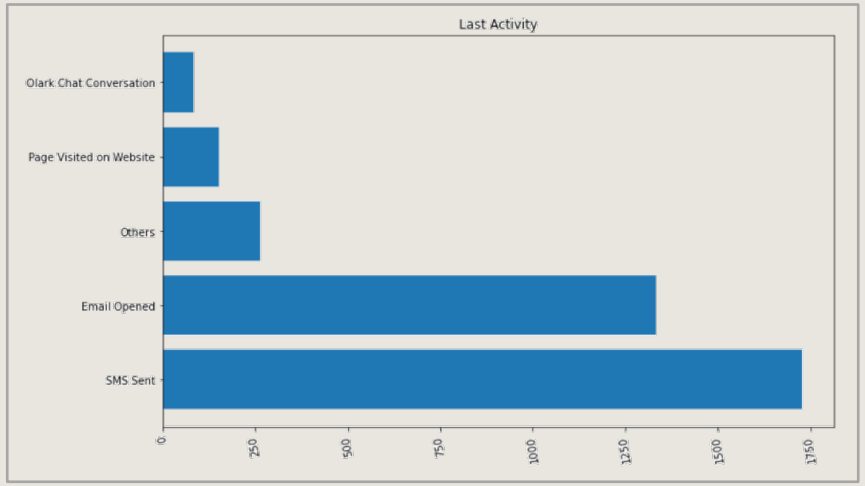
Looking at the visualizations, we can infer that Conversion rate is high when :

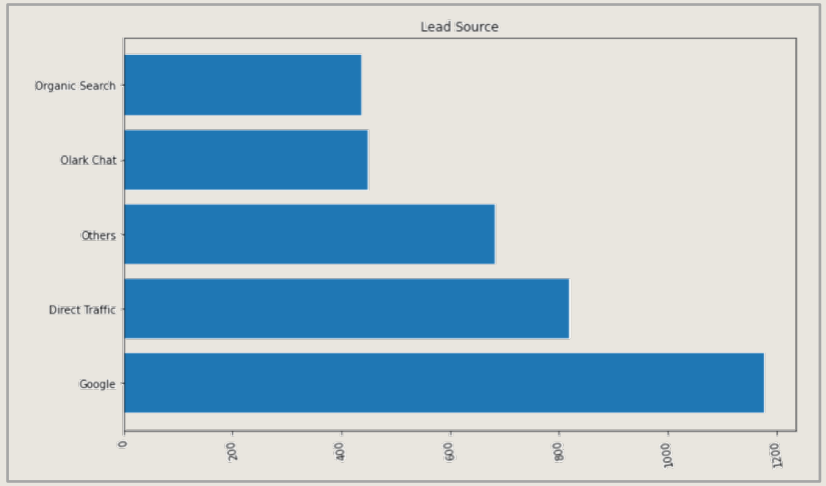
- The last activity is 'SMS sent'

- The lead origin is from from Landing Page Submission

- The lead source is Google



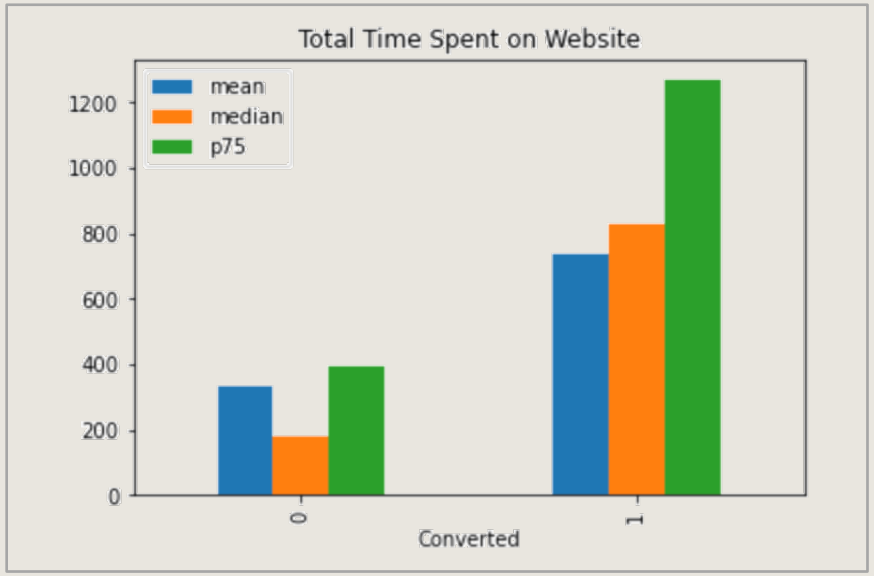


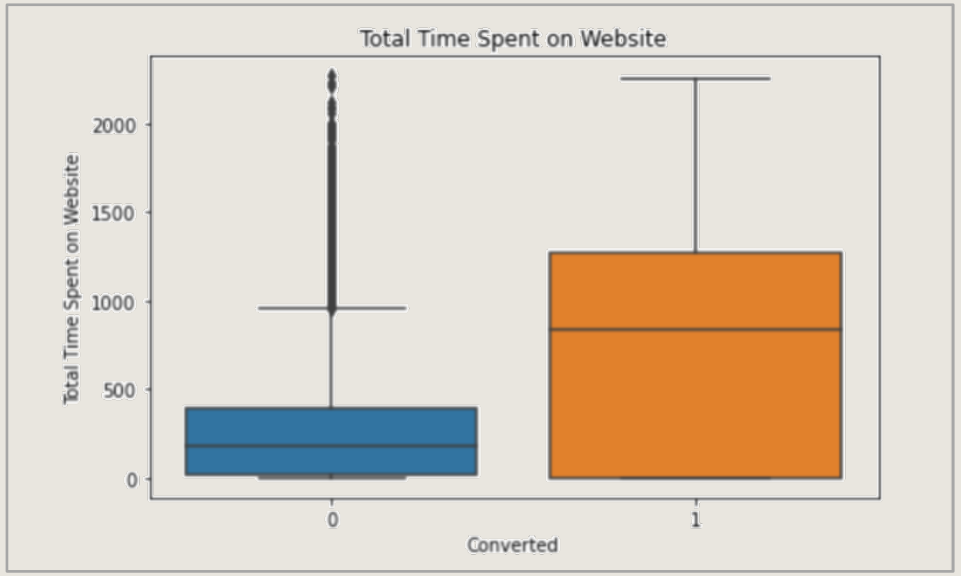


ANALYSIS OF CONTINUOUS DATA

Analysis

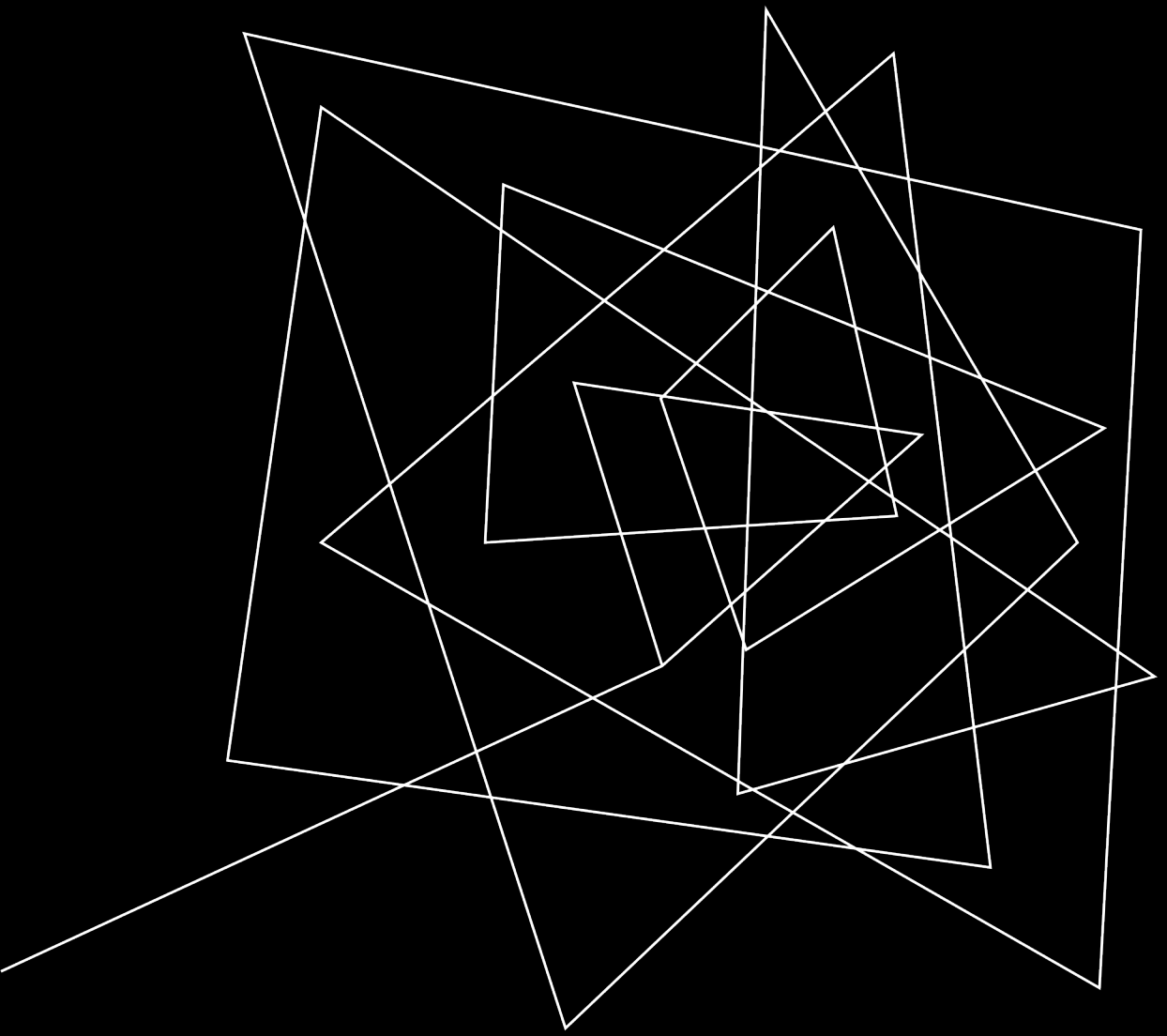
Looking at the visualization, we can clearly see that 'Total Time Spent on Website' is directly impacting conversions. More time the users spend on the Website, the higher is the conversion rate





Suggestion

It would be great if X Education could improve their website to increase total time spent



MODEL EVALUATION

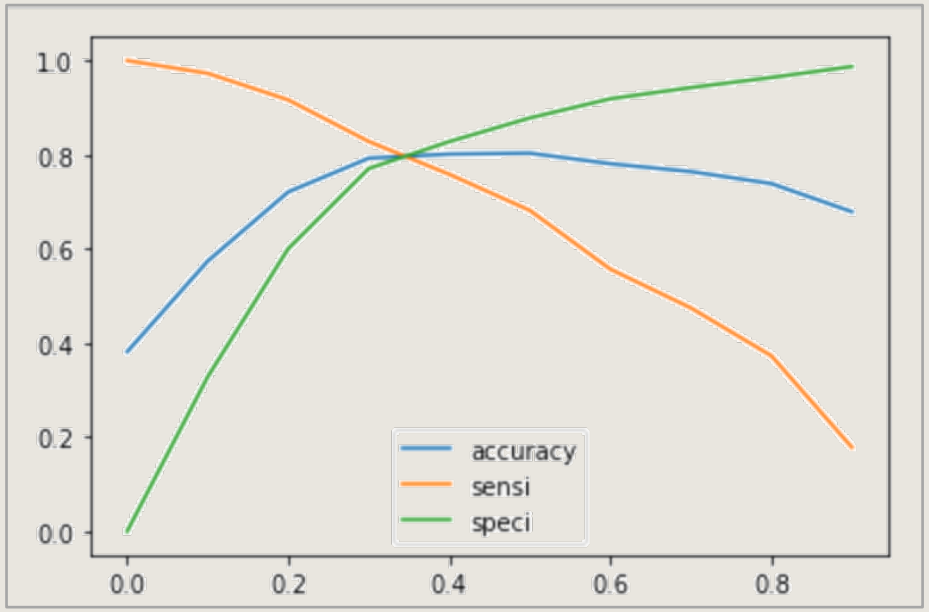
Train Dataset

VARIABLES IMPACTING THE CONVERSION RATE

1. Page Views Per Visit
2. TotalVisits
3. Total Time Spent on Website
4. Do Not Email
5. Last Activity_Olark Chat Conversation
6. Last Activity_Others
7. Last Activity_Page Visited on Website
8. Lead Source_Google
9. Lead Source_Olark Chat
10. Lead Origin_Lead Add Form
11. Last Notable Activity_Modified
12. Last Notable Activity_Others
13. Last Notable Activity_SMS Sent

SENSITIVITY AND SPECIFICITY ON TRAIN DATA SET

The graph depicts an optimal cut off of **0.34** based on Accuracy, Sensitivity and Specificity



Accuracy – 80 %

Sensitivity - 80 %

Specificity – 80 %

False Positive Rate - 20 %

Positive Predictive Value - 70 %

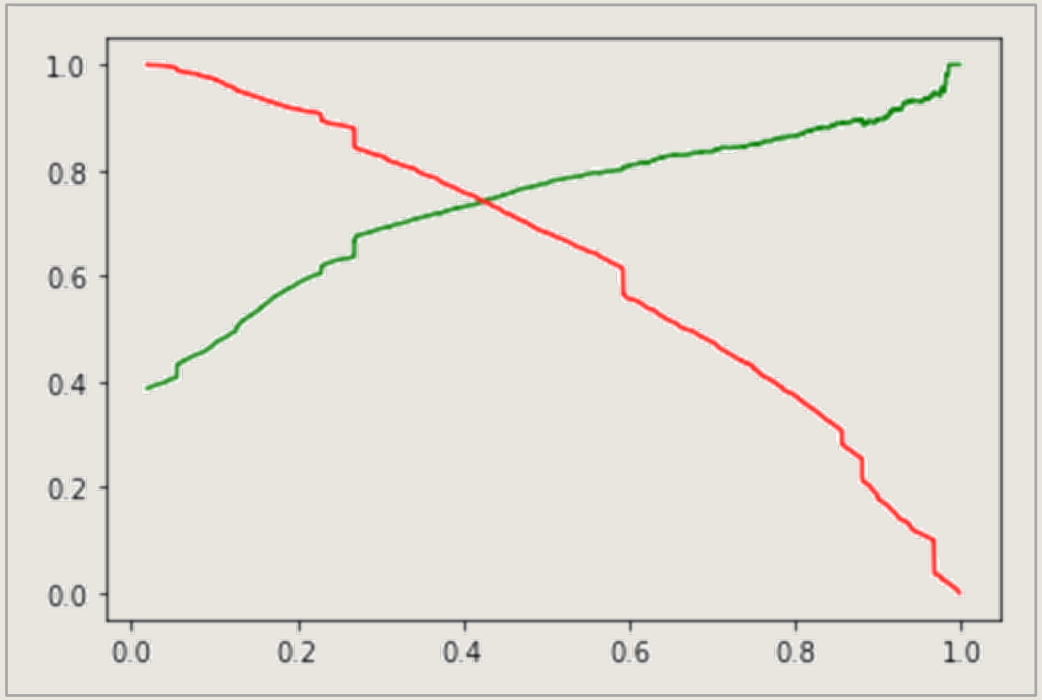
Negative Predictive Value – 86 %

PRECISION AND RECALL ON TRAIN DATASET

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

Precision – 78 %

Recall - 68 %



MODEL EVALUATION TEST DATASET

Accuracy – 80 %

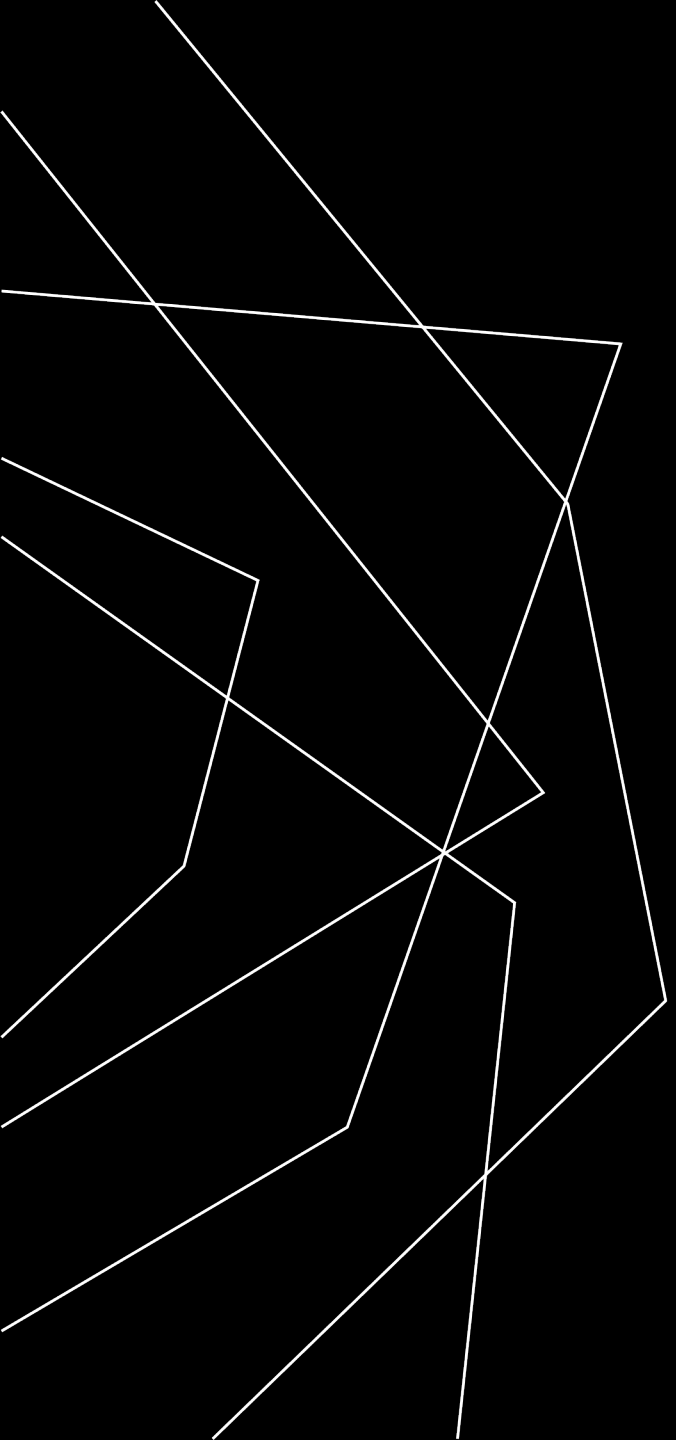
Sensitivity - 80 %

Specificity – 80 %

Conversion Rate - 80%

CONCLUSION

- The top 3 variables that contribute for lead getting converted in the model are
  - Total Time Spent on Website
  - Lead Origin_Lead Add Form
  - Last Notable Activity_SMS Sent
- It would be great if X Education could improve their website to increase total time spent because that leads to conversions
  - Looking at the visualizations, we can infer that Conversion rate is high when :
  - The last activity is 'SMS sent’
  - The lead origin is from from Landing Page Submission
  - The lead source is Google
  - In addition to this, consultation selling model i.e. engaging phone conversations is the go-to-market strategy for X education
- **MODEL -** 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 80%, 80% and 80% 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 around 80%
  - Hence overall this model seems to be good



THANK YOU

ANKIT TAMBE