X Education - Lead Scoring Case Study

Identification of Hot Leads to focus more on them and thus enhancing the conversion ratio for X Education

Group Members:
Ramya shree T V(7892138625, ramyashree@gmail.com)
Ayushman joshi(7827517323, erayushman@gmail.com)

Background

- X Education , An education company named sells online courses to industry professionals

- Many interested professionals land on their website
- The company markets its courses on several websites 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
- 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%
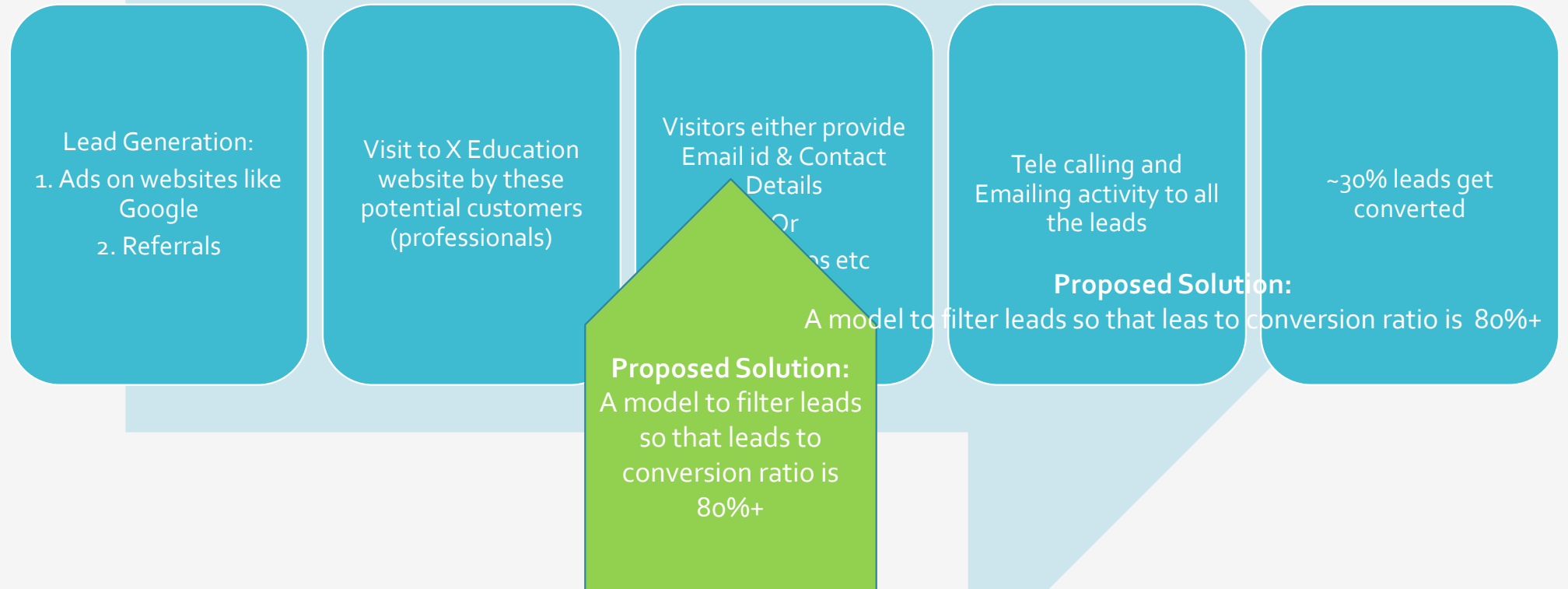
Problem Statement

XEducationCompany's Proble

-      X Education gets a lot of leads but its lead conversion rate is very poor

- To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

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

- We are required 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

- We will help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

  - We are required 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

Lead – Conversion Process
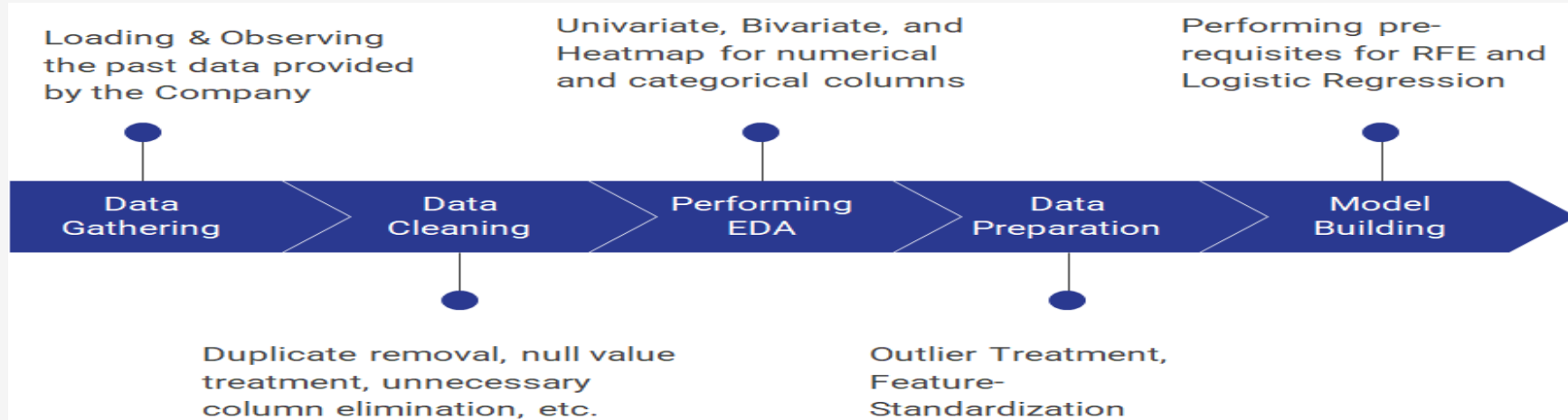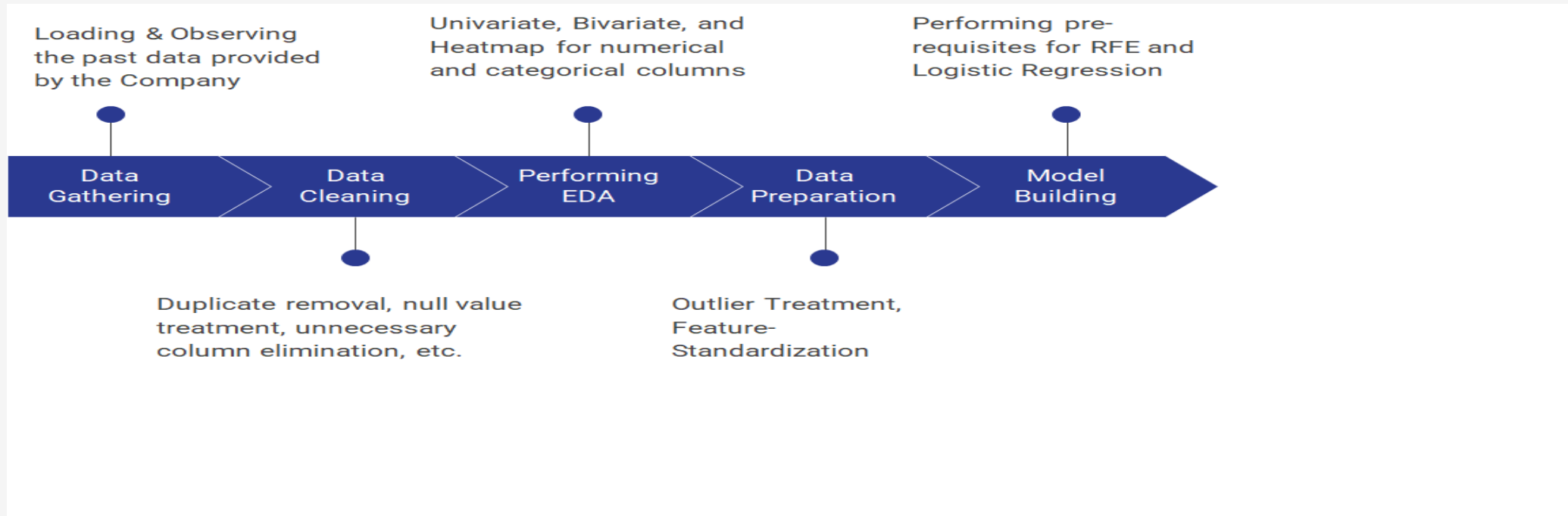
Lead to Conversion process

Lead Generation:
1. Ads on websites like Google
2. Referrals

Visit to X Education website by these potential customers (professionals)

Visitors either provide Email id & Contact Details
Or
...s etc

Proposed Solution:
A model to filter leads so that leads to conversion ratio is 80%+

Tele calling and Emailing activity to all the leads

Proposed Solution:
A model to filter leads so that leas to conversion ratio is 80%+

~30% leads get converted

Solution

For our Problem Solution, the crucial part is to accurately identify hot leads.

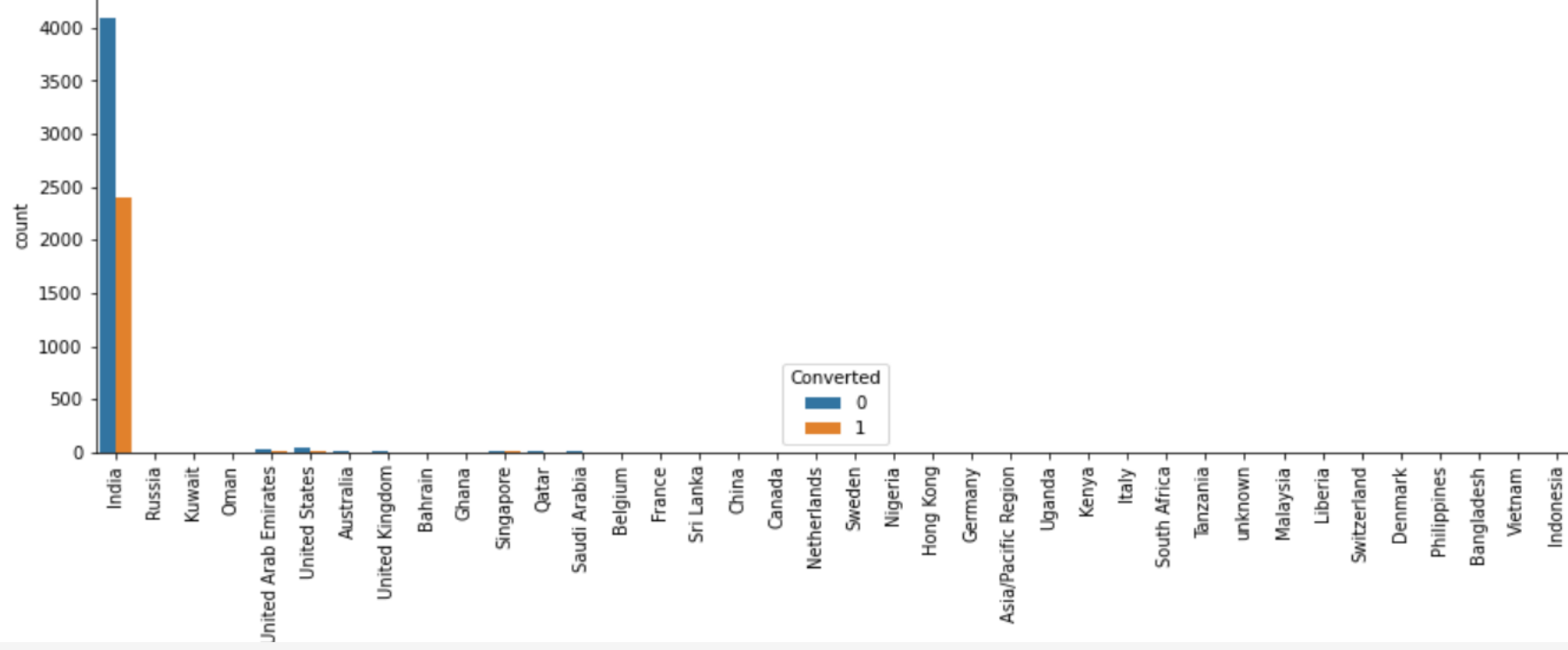The more accurate we obtain the hot lead, the more chance we get of higher conversion ratio.

Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

Implementation

Loading & Observing the past data provided by the Company

Univariate, Bivariate, and Heatmap for numerical and categorical columns

Performing pre-requisites for RFE and Logistic Regression

Data Gathering

Data Cleaning

Performing EDA

Data Preparation

Model Building

Duplicate removal, null value treatment, unnecessary column elimination, etc.

Outlier Treatment, Feature-Standardization

Plots (Visualization)

As we can see the Number of Values for India are quite high (nearly 97% of the Data), this column can be dropped
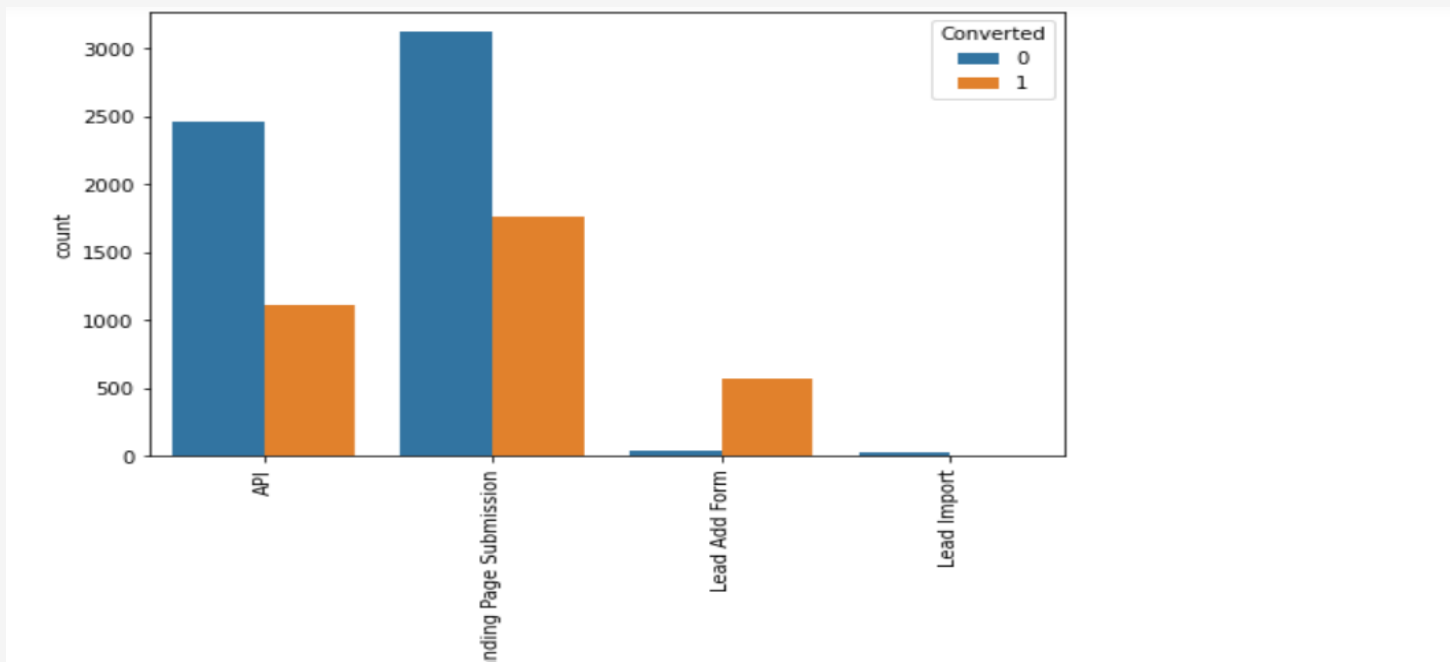
Inference

API and Landing Page Submission bring higher number of leads as well as conversion.

Lead Add Form has a very high conversion rate but count of leads are not very high.

Lead Import and Quick Add Form get very few leads.

In order to improve overall lead conversion rate, we have to improve lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form.
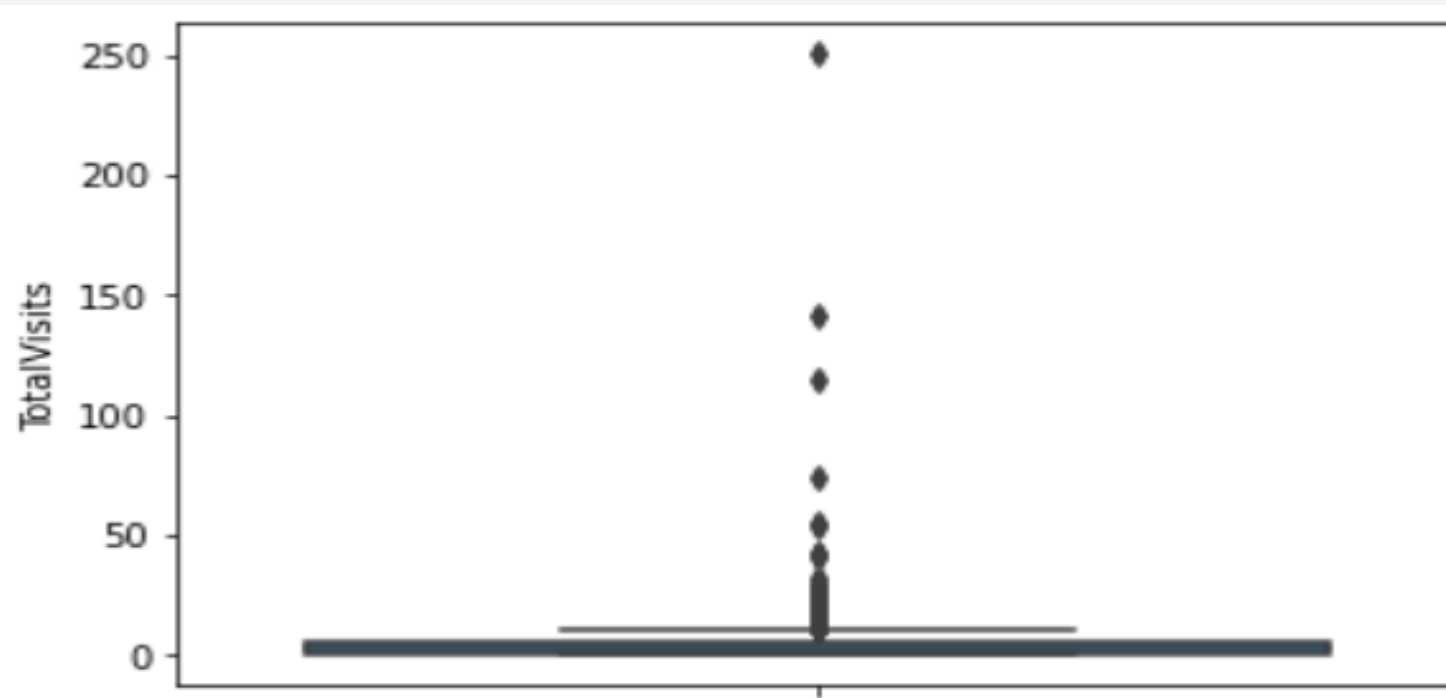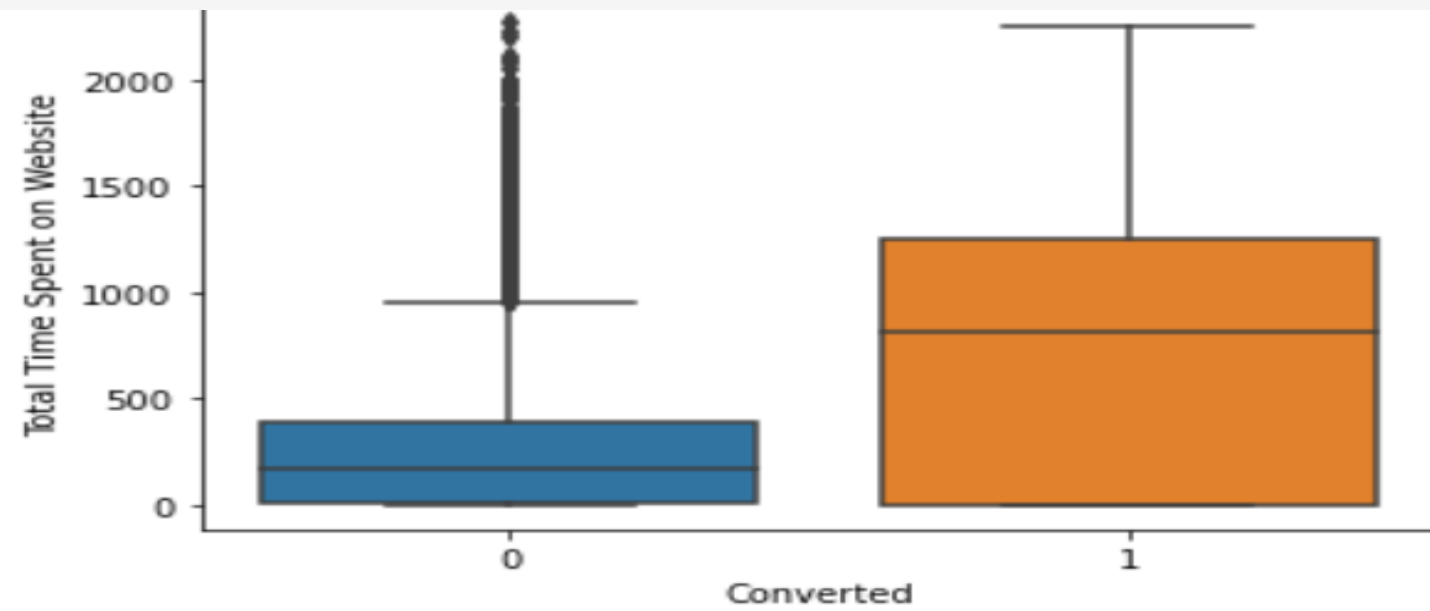
We can see presence of outliers here
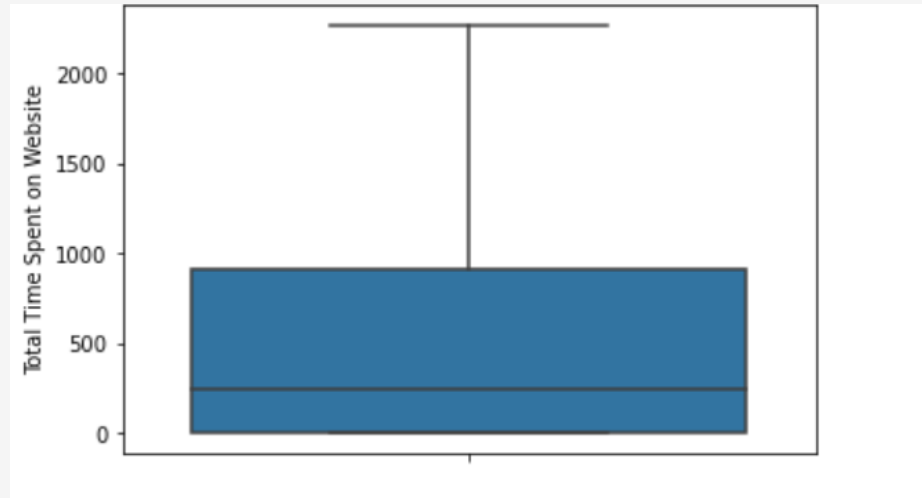
Since there are no major Outliers for the above variable we don't do any Outlier Treatment for this above Column
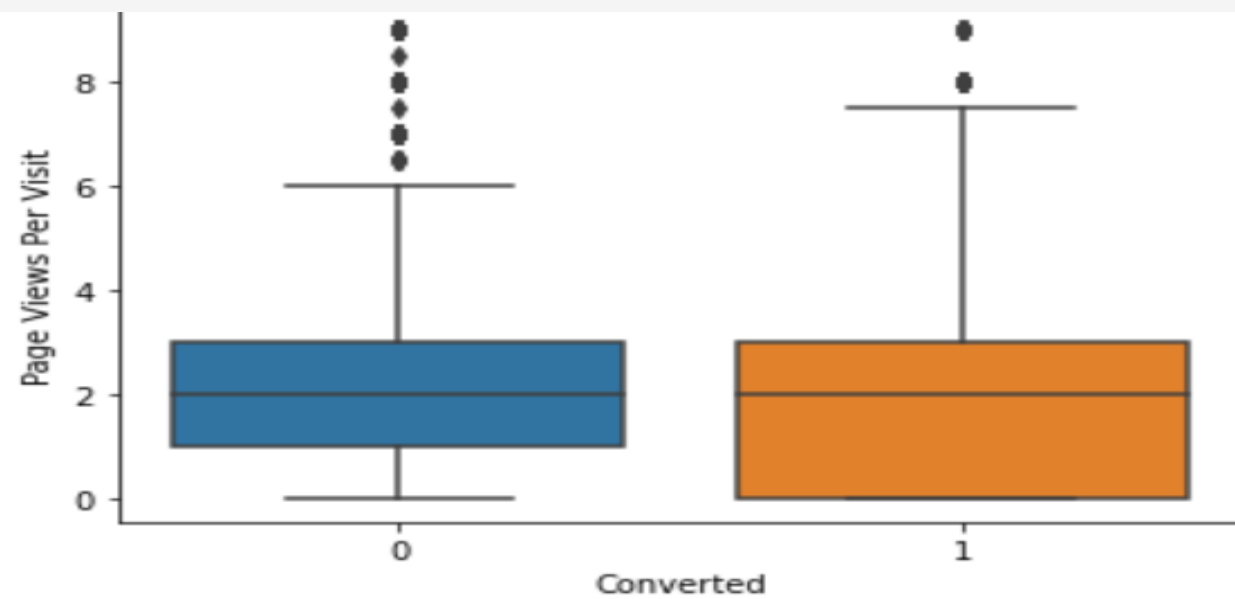
:
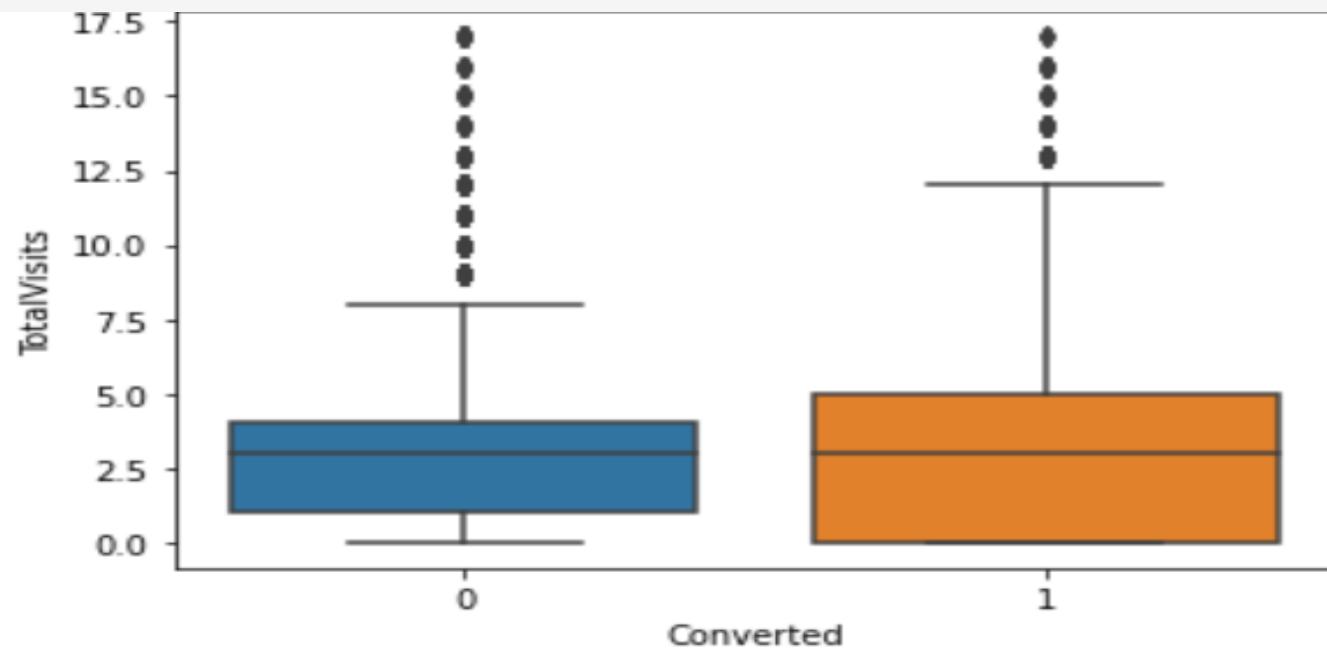
Inference

Median for converted and not converted leads are the close. Nothng conclusive can be said on the basis of Total Visits
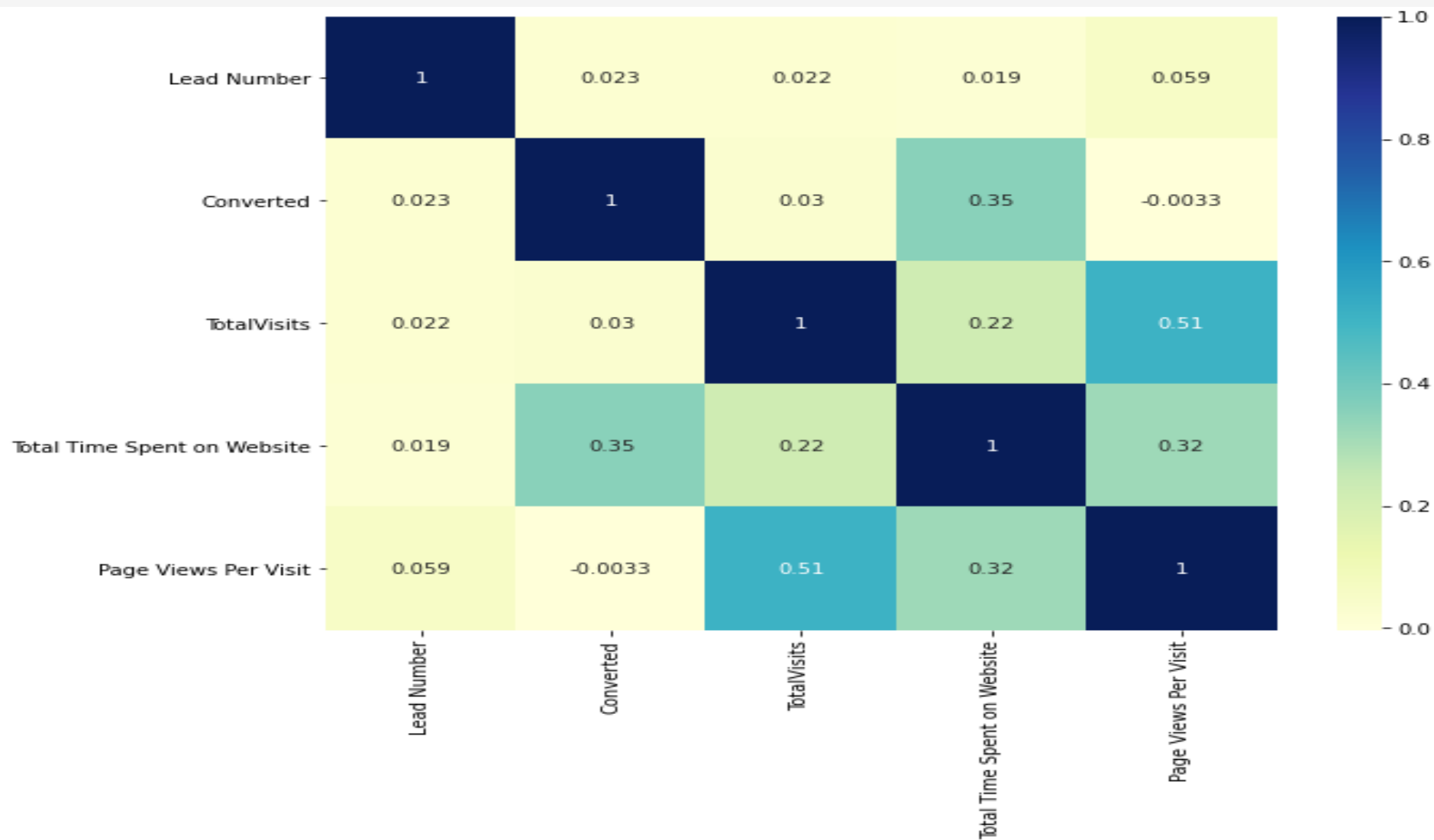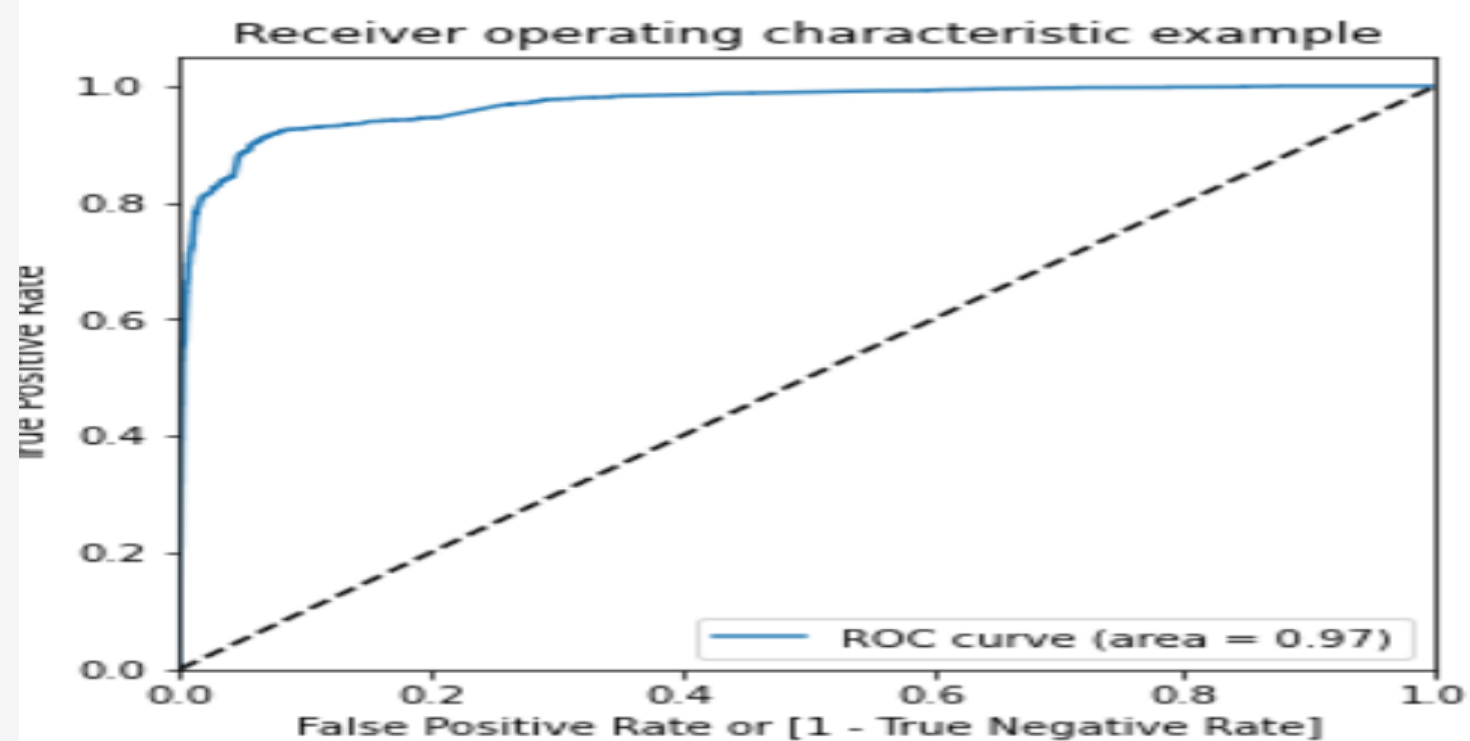
Inference

Leads spending more time on the website are more likely to be converted. Website should be made more engaging to make leads spend more time.
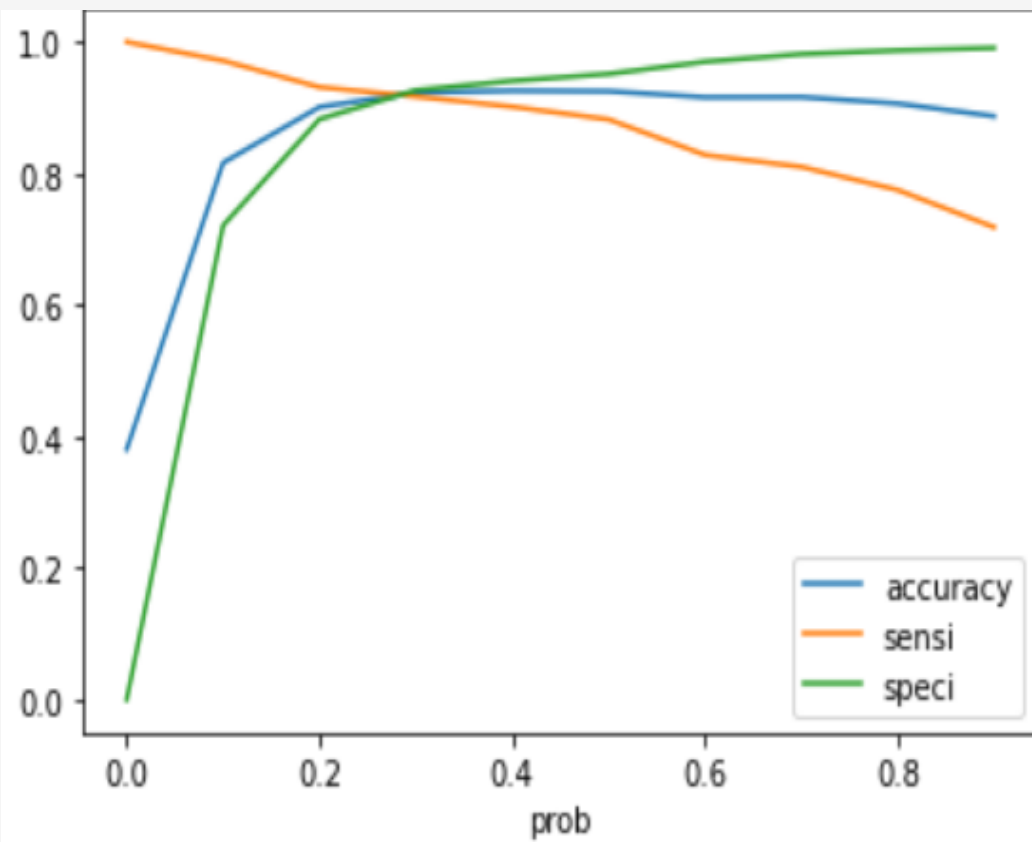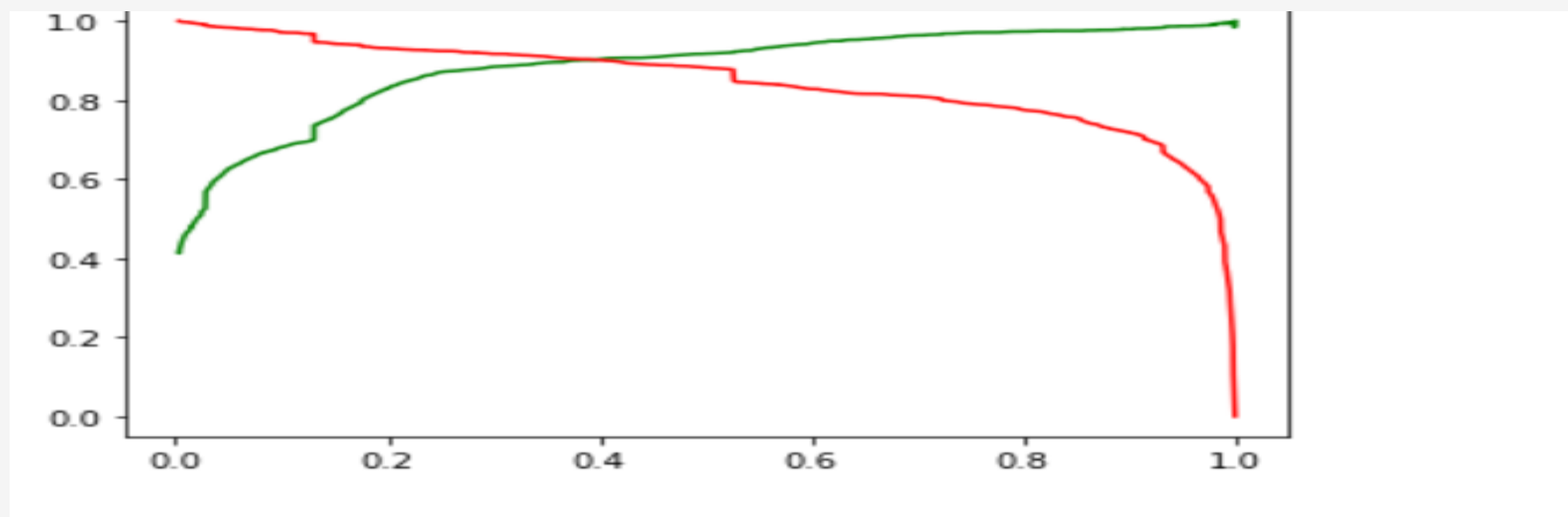
Inference

Median for converted and unconverted leads is the same. Nothing can be said specifically for lead conversion from Page Views Per Visit

Linear Regression Final Model Parameters

Area under ROC = 0.97

Intermediate cut-off = 0.38

Final cut-off = 0.42

Inference / Conclusion

Conclusion 1 (LR Model)

Our Logistic Regression Model is decent and accurate enough, when compared to the model derived using PCA, with 92.79 % Accuracy on Test Set, 91.70% Sensitivity and 92.66 % Specificity. We can vary these parameters by varying the cut-off value and thus predict Hot leads based on scenarios like availability of extra resources and vice-versa.

Conclusion 2 (Recommendation

X Education Company needs to focus on following key aspects to improve the overall conversion rate:

- Increase user engagement on their website since this helps in higher conversion
- Increase on sending SMS notifications since this helps in higher conversion
- Get Total visits increased by advertising etc. since this helps in higher conversion
- Improve the Olark Chat service since this is affecting the conversion negatively