



# Lead scoring case study

---

Identification of Hot Leads to focus more on them and thus enhancing the conversion ratio.

BUSINESS OBJECTIVE

# Libraries used

---

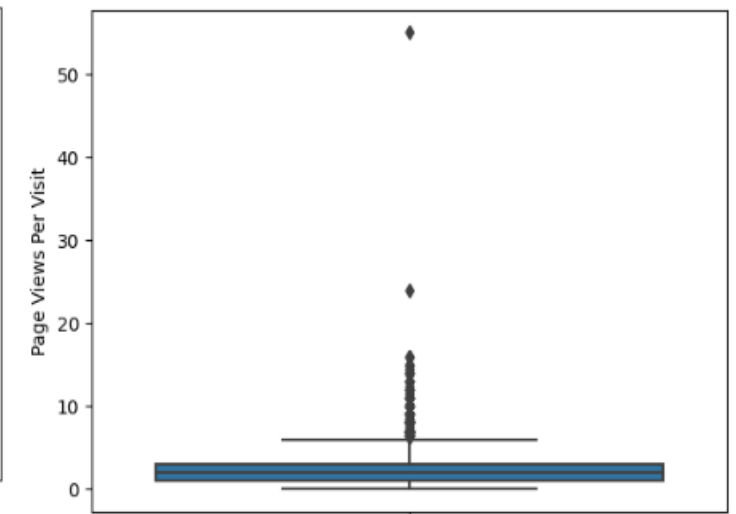
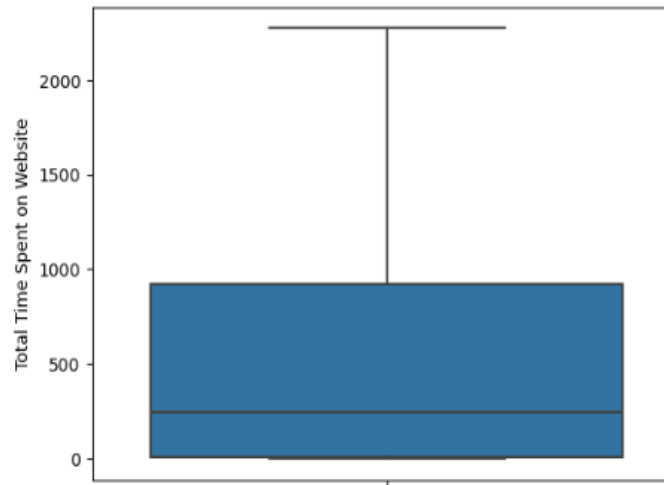
- a) Numpy
- b) Pandas
- c) Matplotlib
- d) Seaborn
- e) Sklearn

# Observations and results

---

## Univariate analysis

1. There is low variation in Page Views Per Visit and Total Visits but higher variation in Total Time Spent on Website.
2. There are a lot of outliers in Page Views Per Visit and Total Visits which needs to be treated before modelling.

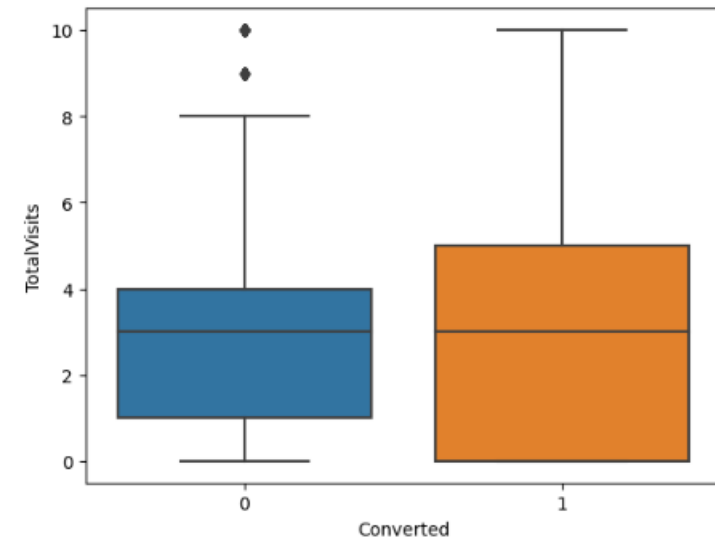


# Observations and results

---

## Bivariate analysis

1. There is positive correlation between Total Time Spent on Website and Conversion.
2. There is some correlation between Conversion and some categorical columns like Lead Origin and Lead Source.
3. Variables Lead Origin, Lead Add Form, Last Activity\_EmailBounced, Lead Source Reference, Total Time Spent on Website have effects on conversion.

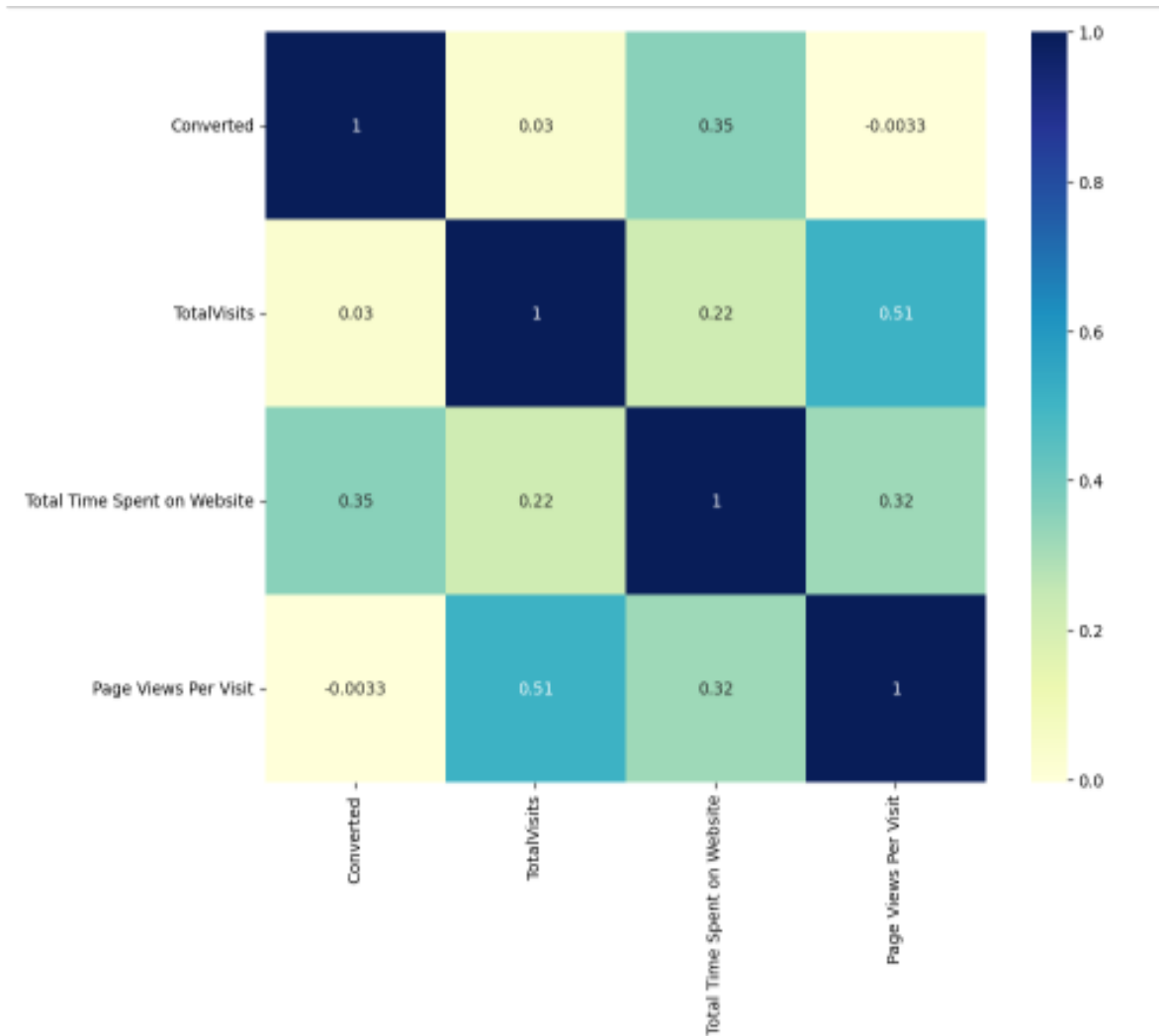


# Observations and results

---

## Correlation Analysis (Heatmap)

1. There is positive correlation between TotalTime Spent on Website and Conversion.
2. There is almost no correlation in Page ViewsPer Visit and TotalVisits with Conversion



# Final model summary

---

<b>Accuracy</b>  Overall accuracy on Test set: 0.786	<b>Sensitivity</b>  sensitivity of our logistic regression model:0.733
<b>Specificity</b>  specificity of our logistic regression model:0.823	<b>Top 3 variables to be focused for conversion</b> <ul style="list-style-type: none"><li>• Total Time Spent on Website</li><li>• Last Notable Activity SMS Sent</li><li>• Total Visits</li></ul>

# Conclusions

---

1. Our Logistic Regression Model is decent and accurate enough.
2. X education company needs to focus on the following aspects to improve conversion rate:
  - a) Increasing user engagement on website.
  - b) Increment in sending SMS notifications.
  - c) Increasing Total Visits by advertising.
  - d) Improving chat service.