# *Lead Scoring Case Study*

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 company markets its courses through different search engine like Google, through referrals, videos fill up the forms when browsing the courses etc.

- The analysis is done as a part of X Education to help them understand how can we classify different leads and convert them to make sure they are choosing the course.
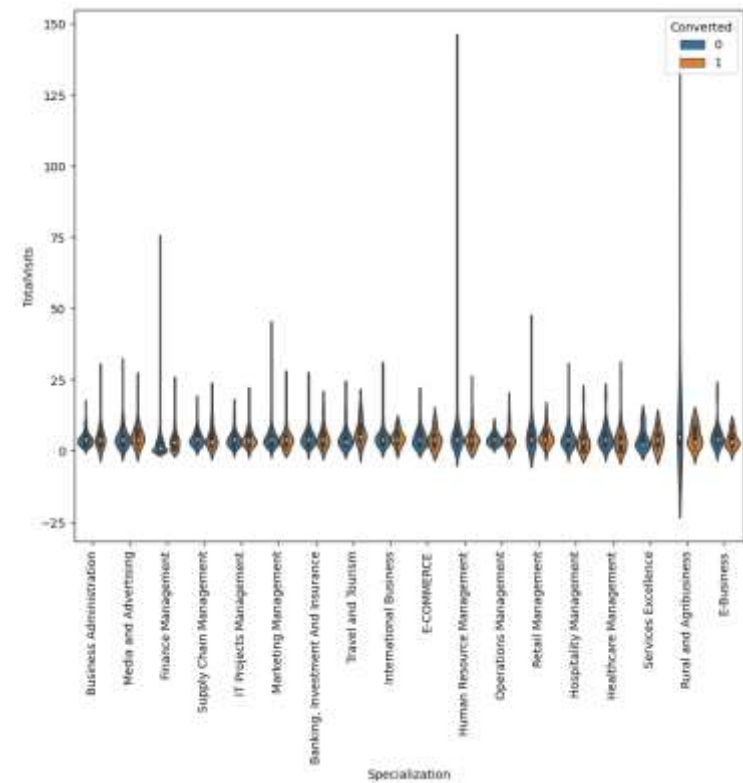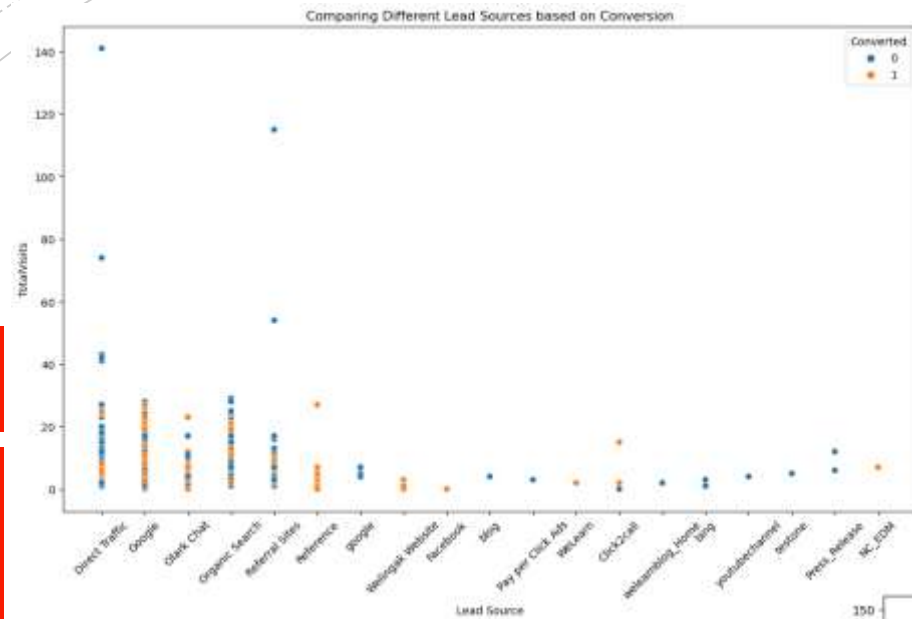
Steps in Analysis

- Importing the libraries

- Reading the dataset

- Data Cleaning

- EDA

- Data Visualisation

- Model Building

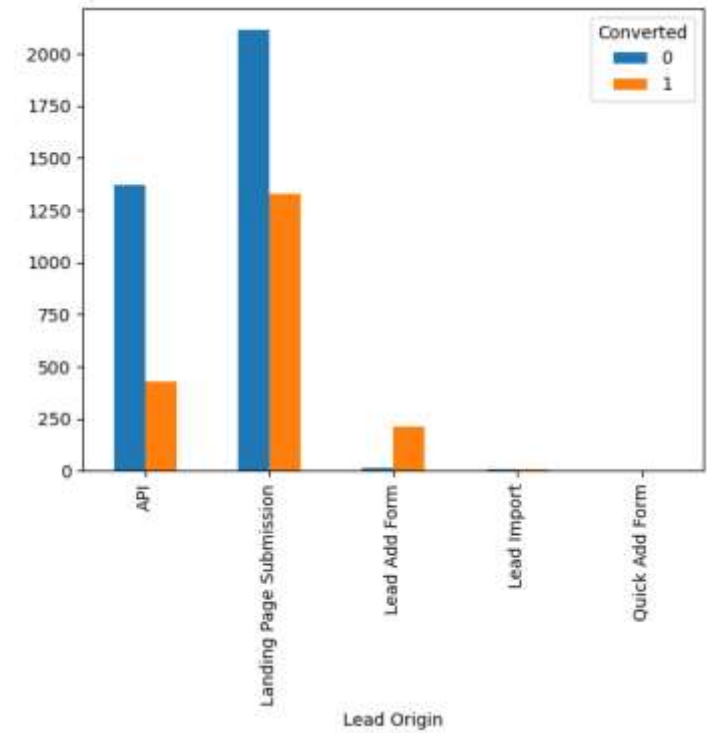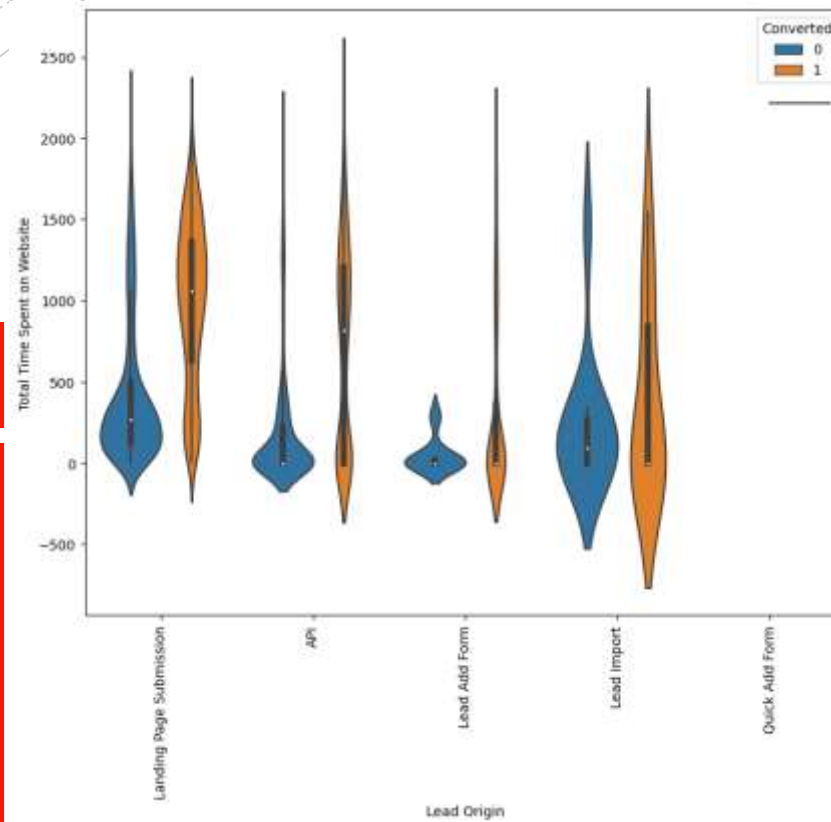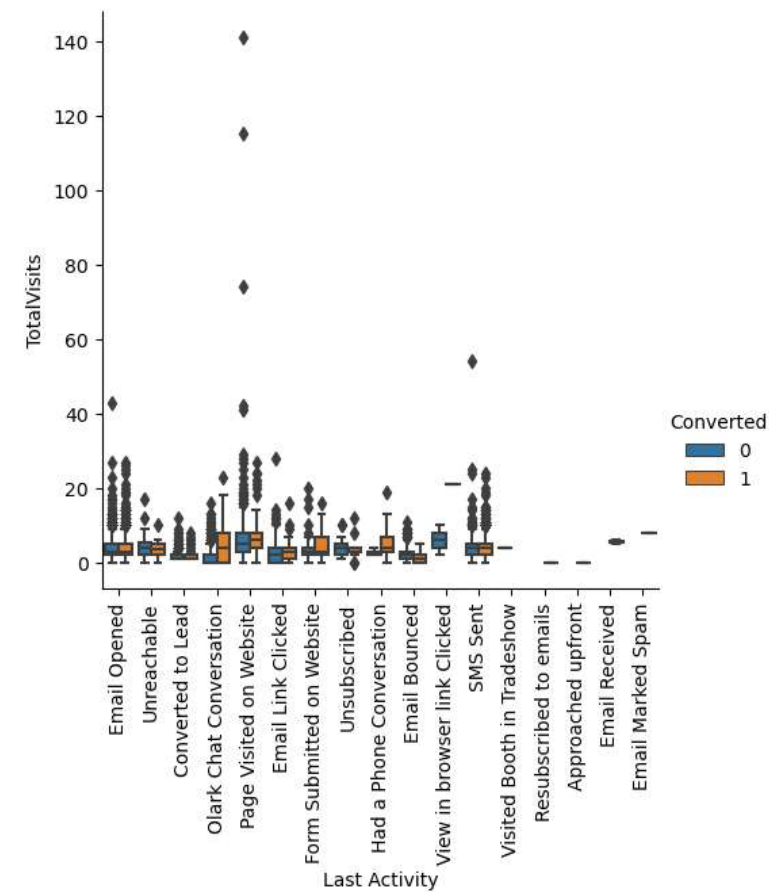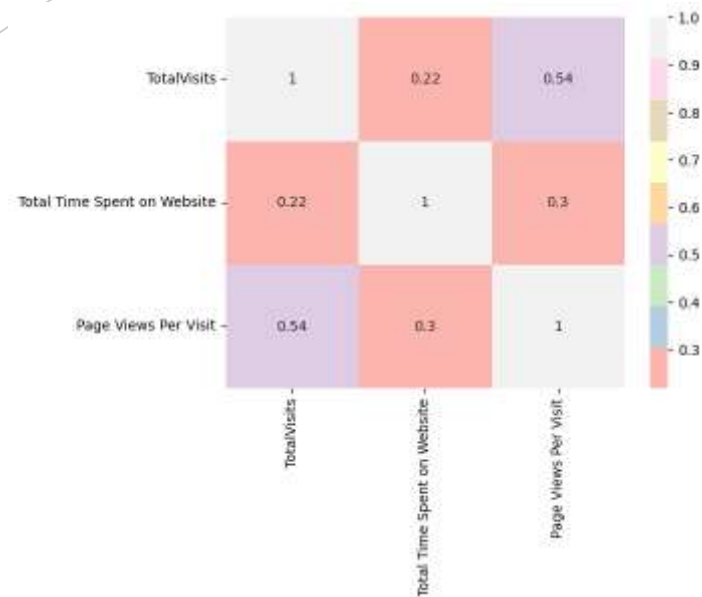- Predictions and evaluation of the model set

Plots

Plots

Comparing Different Lead Sources based on Conversion
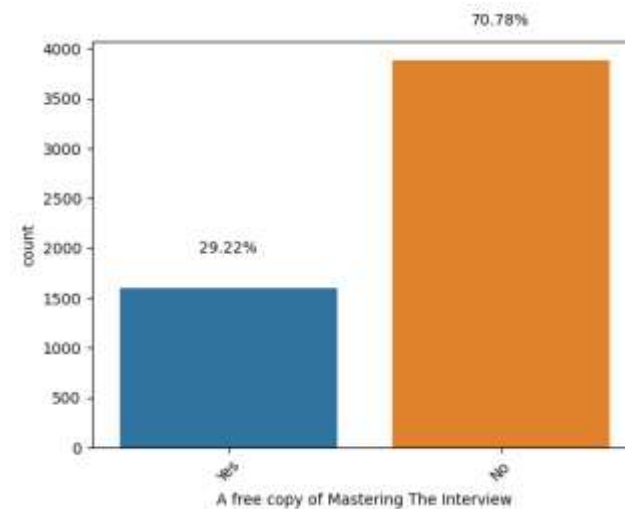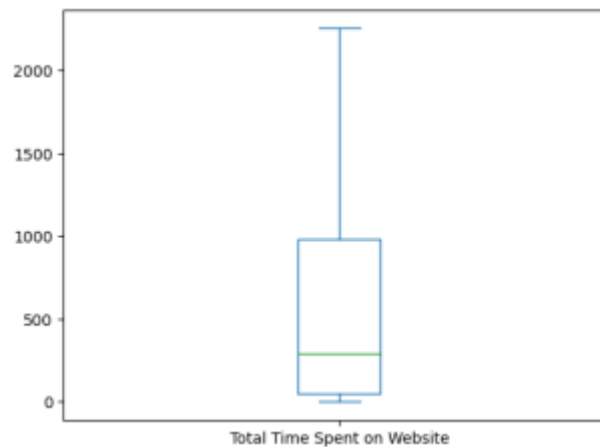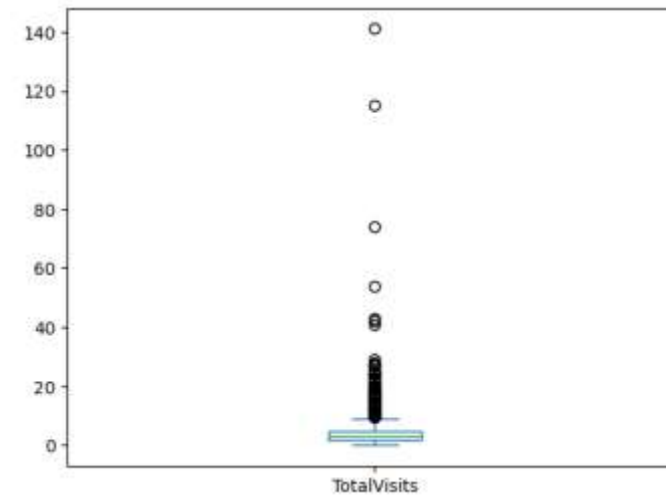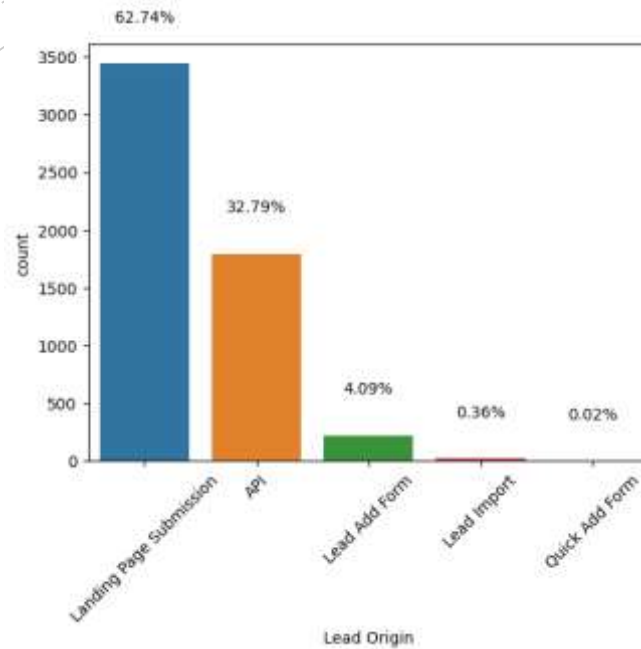
Plots

Plots

Evaluating the model

- We have used confusion matrix for train and test data.

- Different Evaluation metrics are used: Accuracy, Precision, etc.

- Accuracy: 79% on train data and 78.5% on test data

- Precision: 66% on train data and 67% on test data

- Recall or Sensitivity: 83% on train and test data

- Specificity: 76% on train data and 75% on test data

- The area under the ROC curve is 87%.

- The initial cut-off is 0.5 and final cut-off is 0.3.

Summary

The significant variables on the final model are:

- Do Not Email
- Lead Source - Olark Chat
- Last Activity - Email Bounced
- Specialization - Finance Management
- Last Activity – Olark Chat Conversation
- Total Time Spent On Website
- Last Activity – SMS Sent
- Last Activity – Unsubscribed
- Last Activity – Converted to Lead

Summary

- We can use more variables for analysis and come up with more models that could help the business.

- We can also use different analysis techniques like where there are more categorical values trying them to limit them and simplify them for better analysis.