# Lead Scoring Assignment

# PROBLEM STATEMENT

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- In order to increase the lead conversion rate, the company first should 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.

Business Objective:

-- X education wants to know most promising leads.

-- For that they want to build a Model which identifies the hot leads.

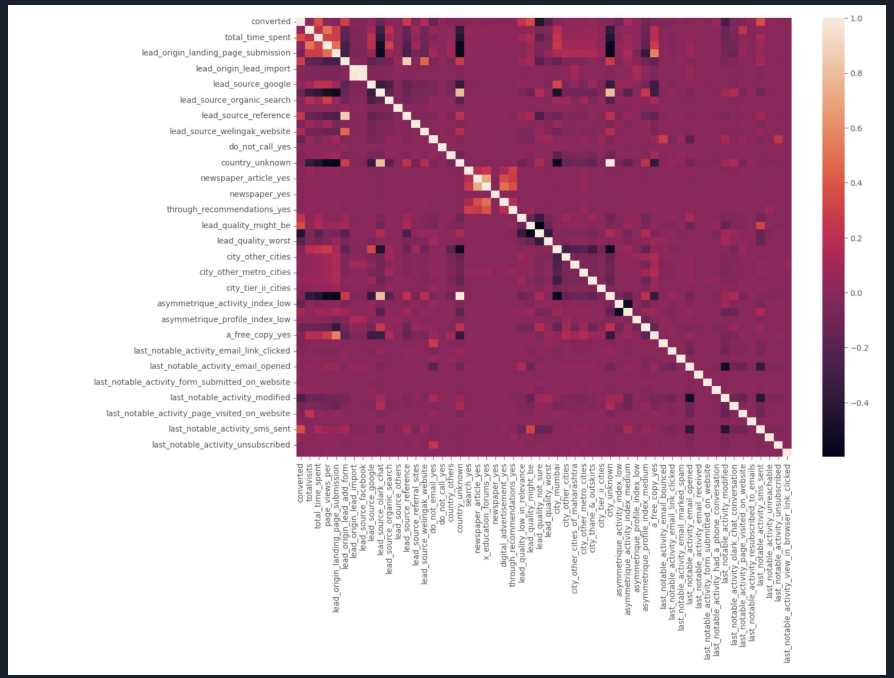
-- Deployment of the model for the future use

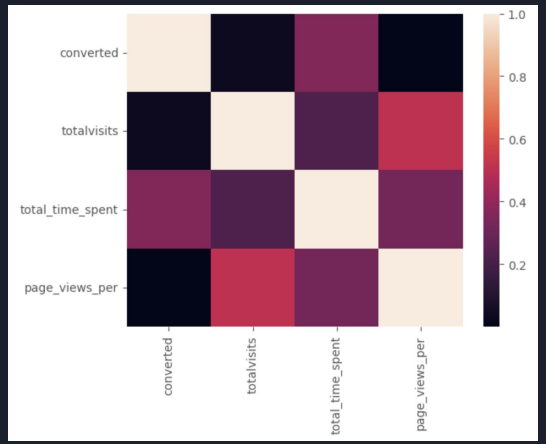
# Solution Methodology

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and pvalues)
- Model Evaluation
- Making predictions on test set

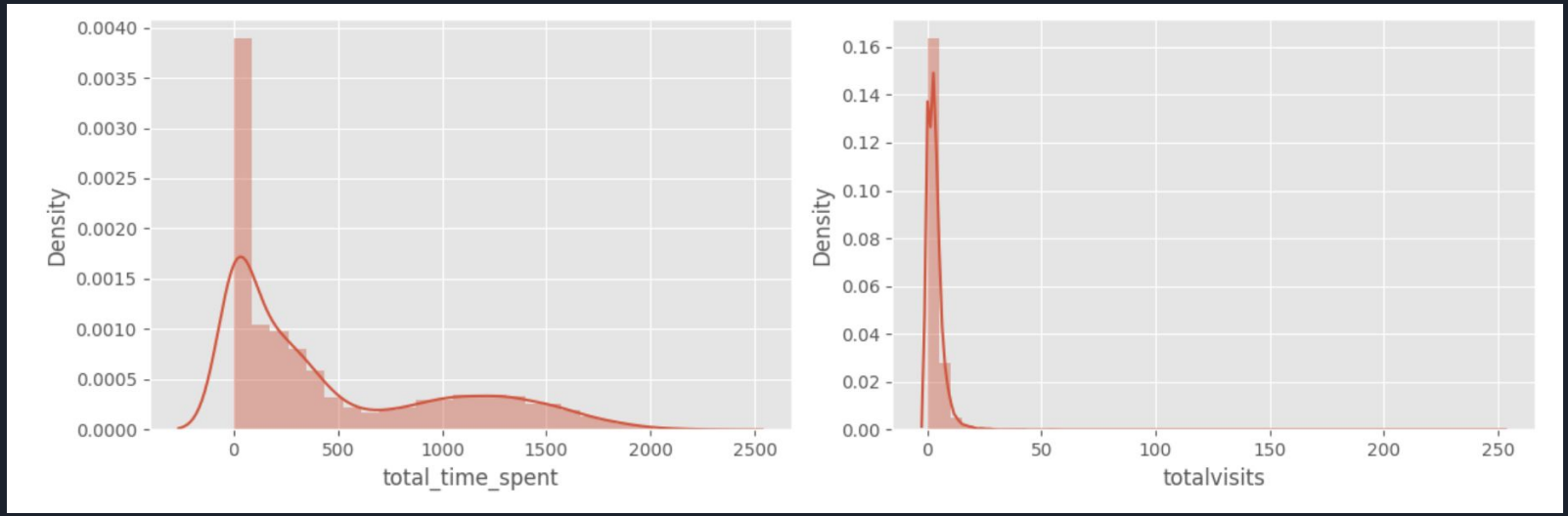
# EDA

Correlation HeatMap -

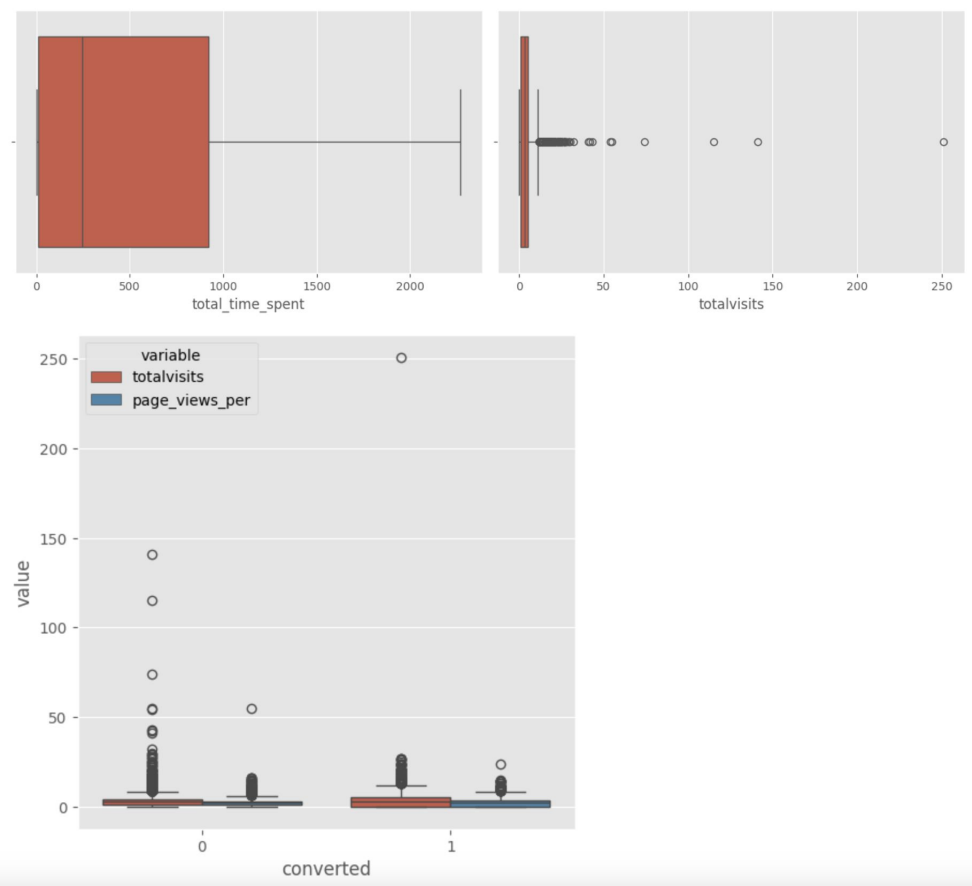


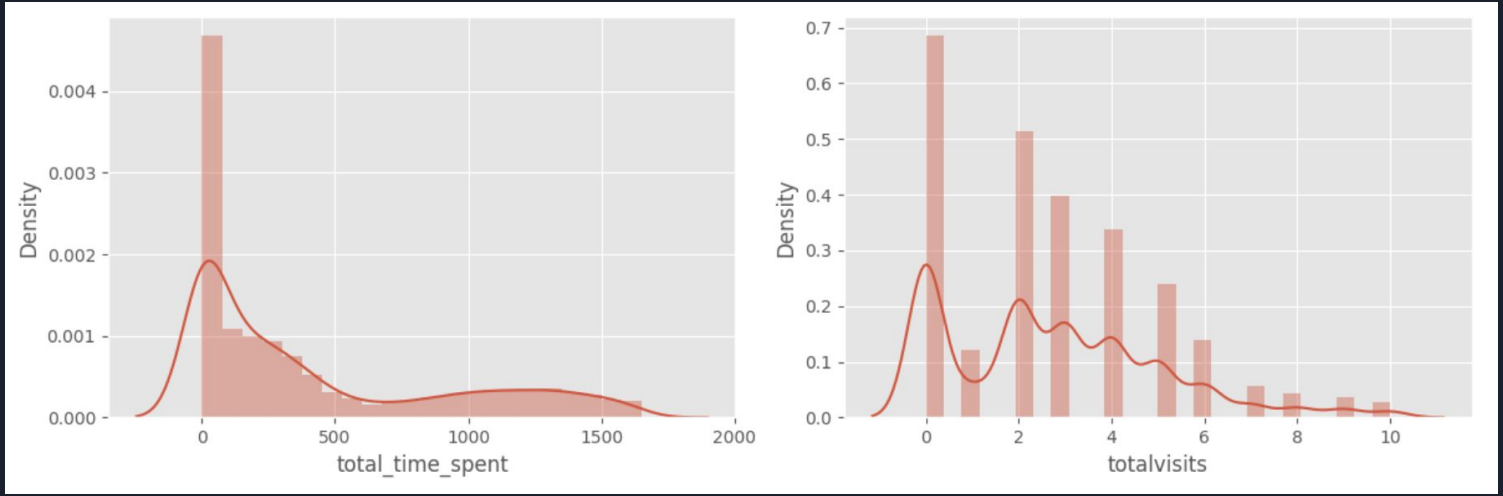
# Univariate and Outlier -



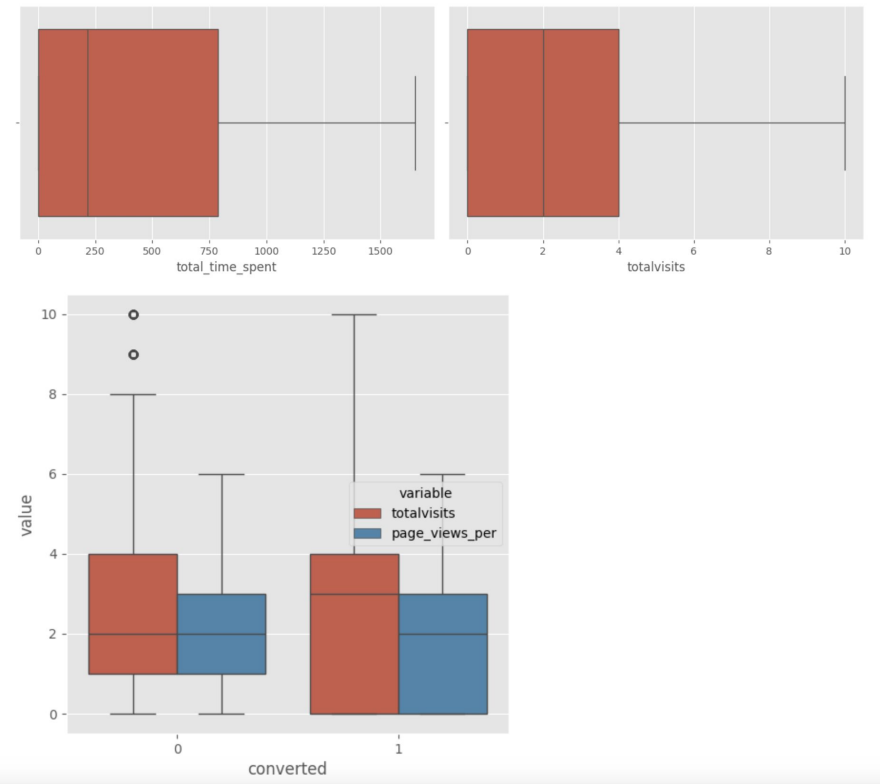
Boxplots before outlier removal -



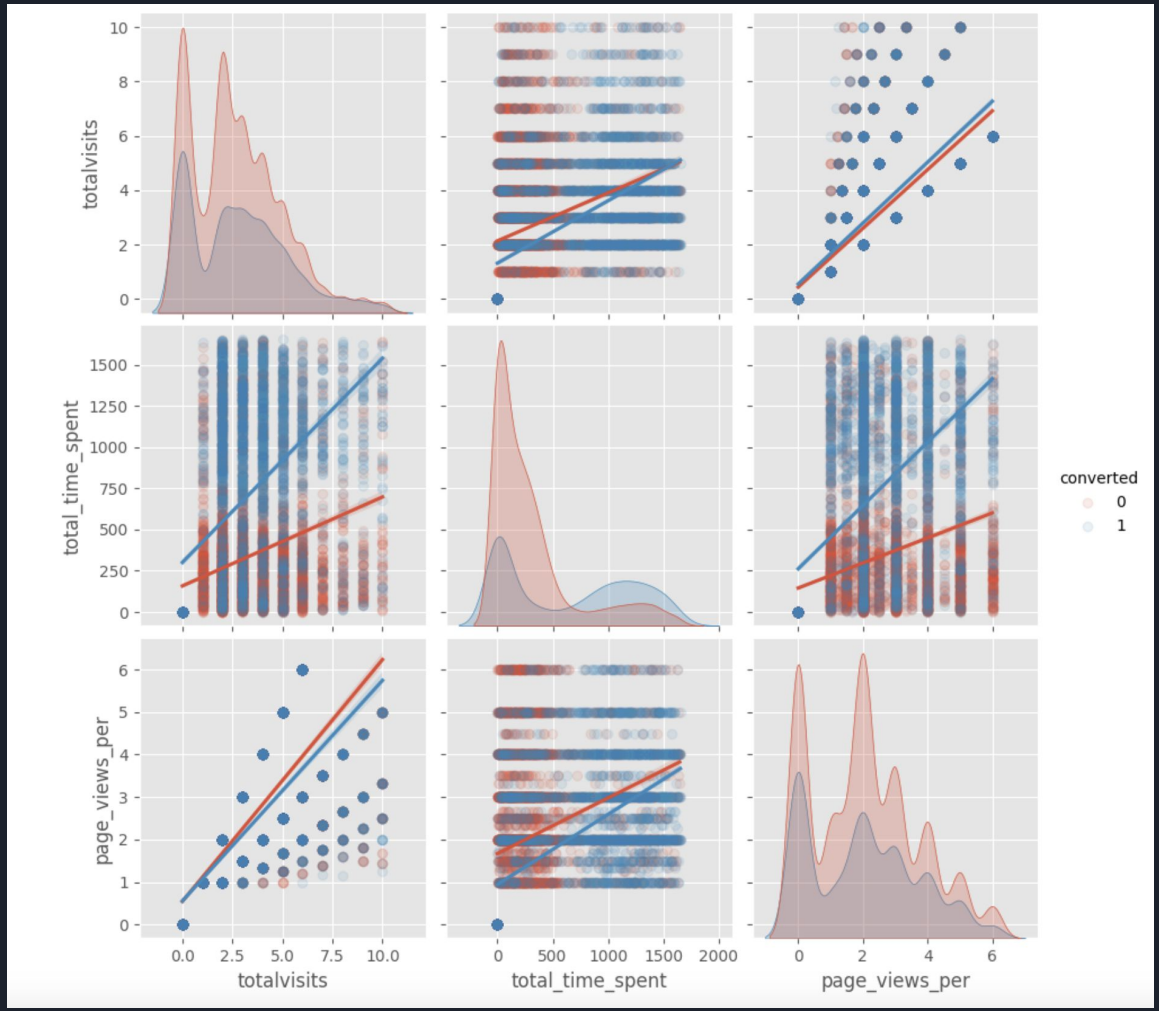
After Outlier removal -



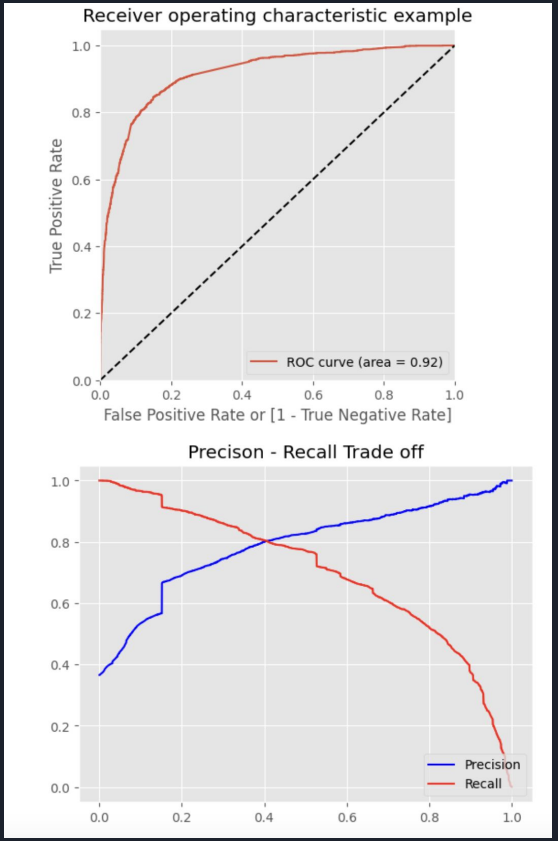
Boxplots after outlier removal -

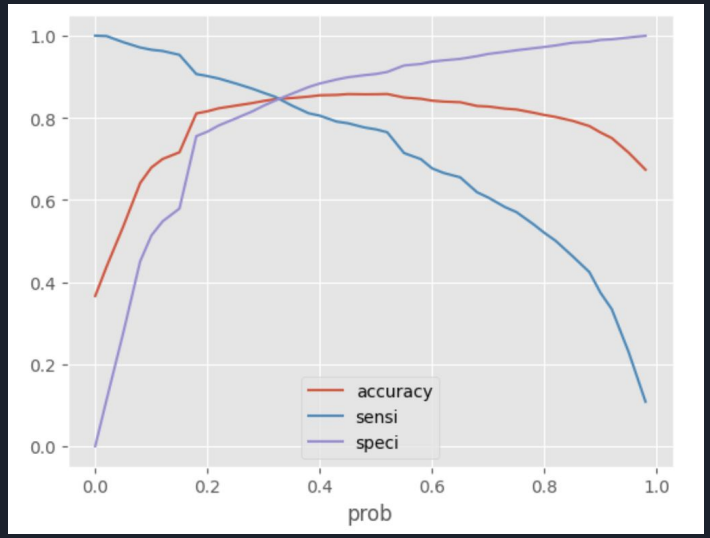


# Bivariate Analysis-



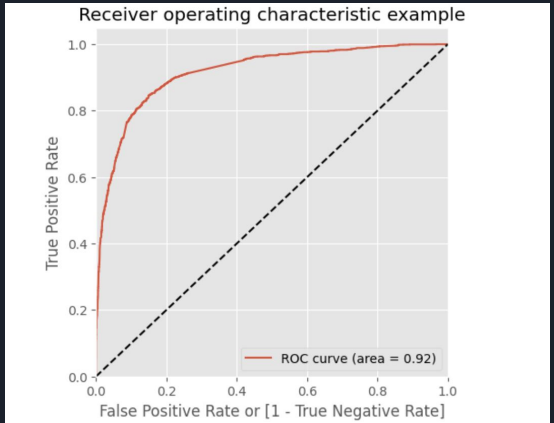
# Building and Measuring Model Performance

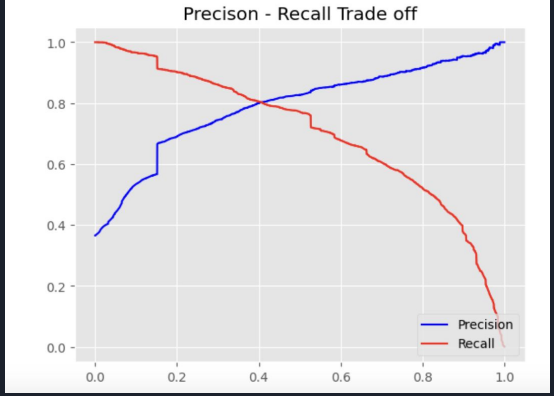



Getting Optimum cut-off value - 0.32

New predicted values based on cut-off -



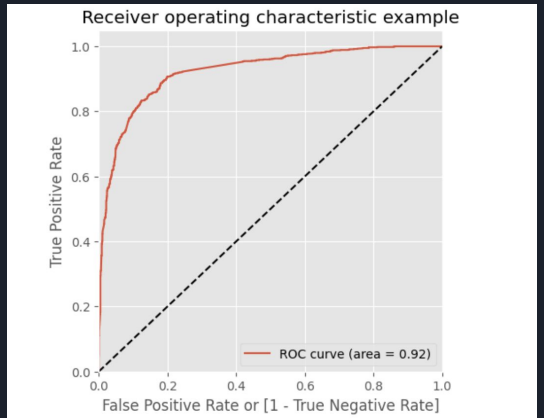

Observations-

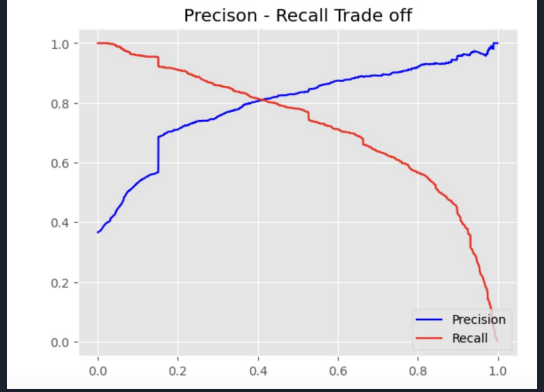
Sensitivity (Recall): 85%

Specificity: 84%

Precision: 75%

# Measuring Performance on Test Set




Observations-

Sensitivity (Recall): 85%

Specificity: 84%

Precision: 76%

# Conclusion

- We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore we can intervene that we need to focus more on the leads originated from API and Landing page submission.
- We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.

# THANK YOU