Lead Scroing Case Study

Group Members:

1. Chandrakant Jagtap
2. Archana Patil
3. Sirisha G.

Problem Statement:

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads through several websites & search engines where people fill the forms for courses.
- The typical lead conversion rate at X education is around 30%, which 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.

Business Objective:

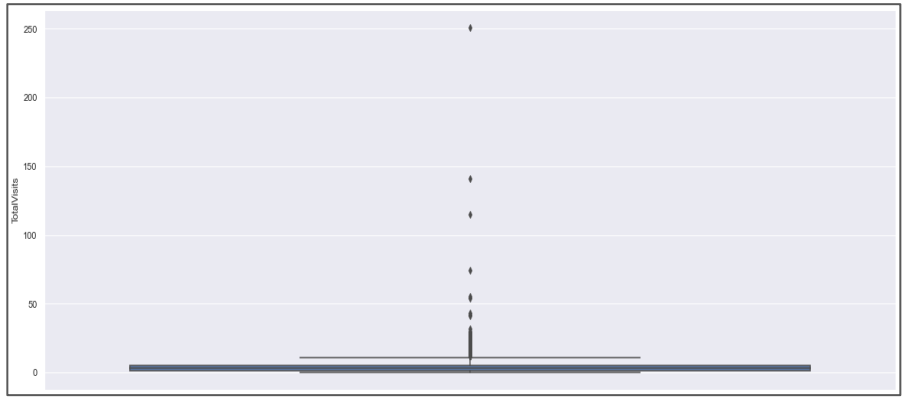
- X education wants to know most promising leads,i.e. the leads that are most likely to convert into paying customers.
- For that company want to build a Model wherein you need to assign a lead score to each of the leads
- The CEO, has given a ballpark of the target lead conversion rate to be around 80%

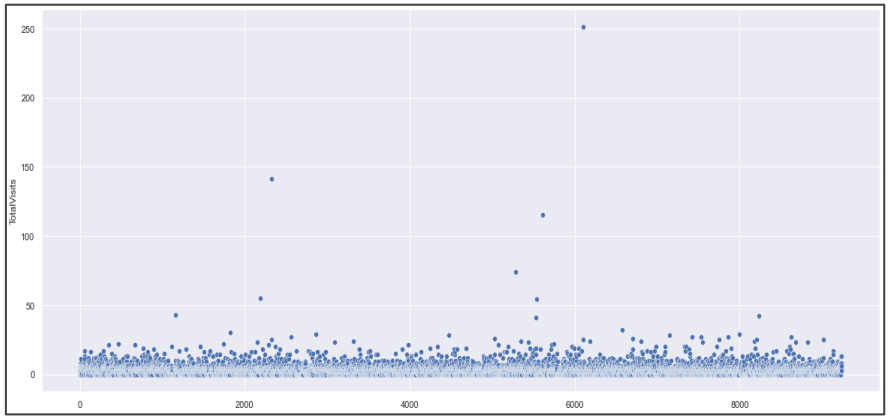
Approach

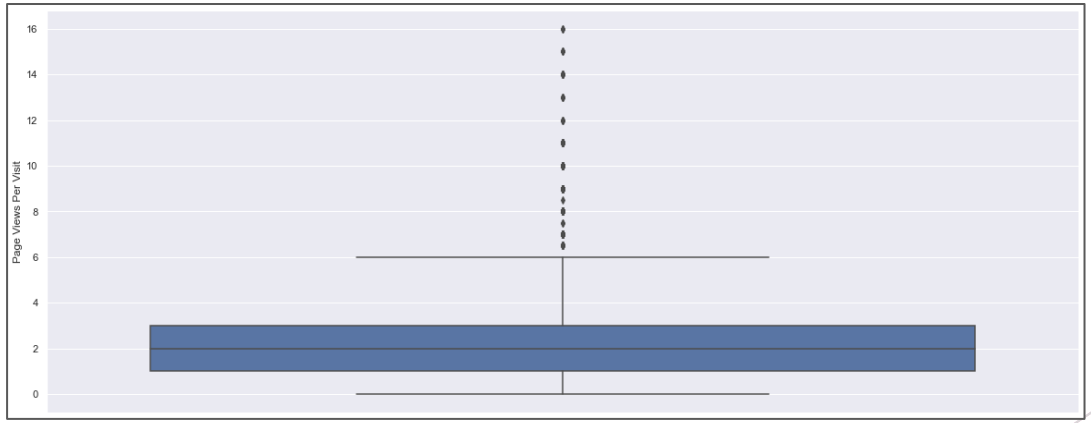
- ❑ Source the data for analysis
- ❑ Reading & Understanding the data
- ❑ Data Cleaning:
  - ➢ Check and handle duplicate data.
  - ➢ Check and handle NA values and missing values
  - ➢ Dropping columns
  - ➢ Imputation of the values, if necessary
  - ➢ Check and handle outliers in data.
- ❑ EDA
  - ➢ Univariate data analysis: value count, distribution of variable etc
  - ➢ Bivariate data analysis: correlation coefficients and pattern between the variables etc.
- ❑ Feature Scaling, Dummy Variables and encoding the data
- ❑ Splitting the data into Test & Train dataset
- ❑ Prepare the data for Modelling
- ❑ Logistic Regression used for the Model Building
- ❑ Model Evaluation – Specificity & Sensitivity or precision recall
- ❑ Making predictions on the test set
- ❑ Conclusions and Recommendations

Outliers:

Total Visits, Page Views per Visit have outliers







EDA

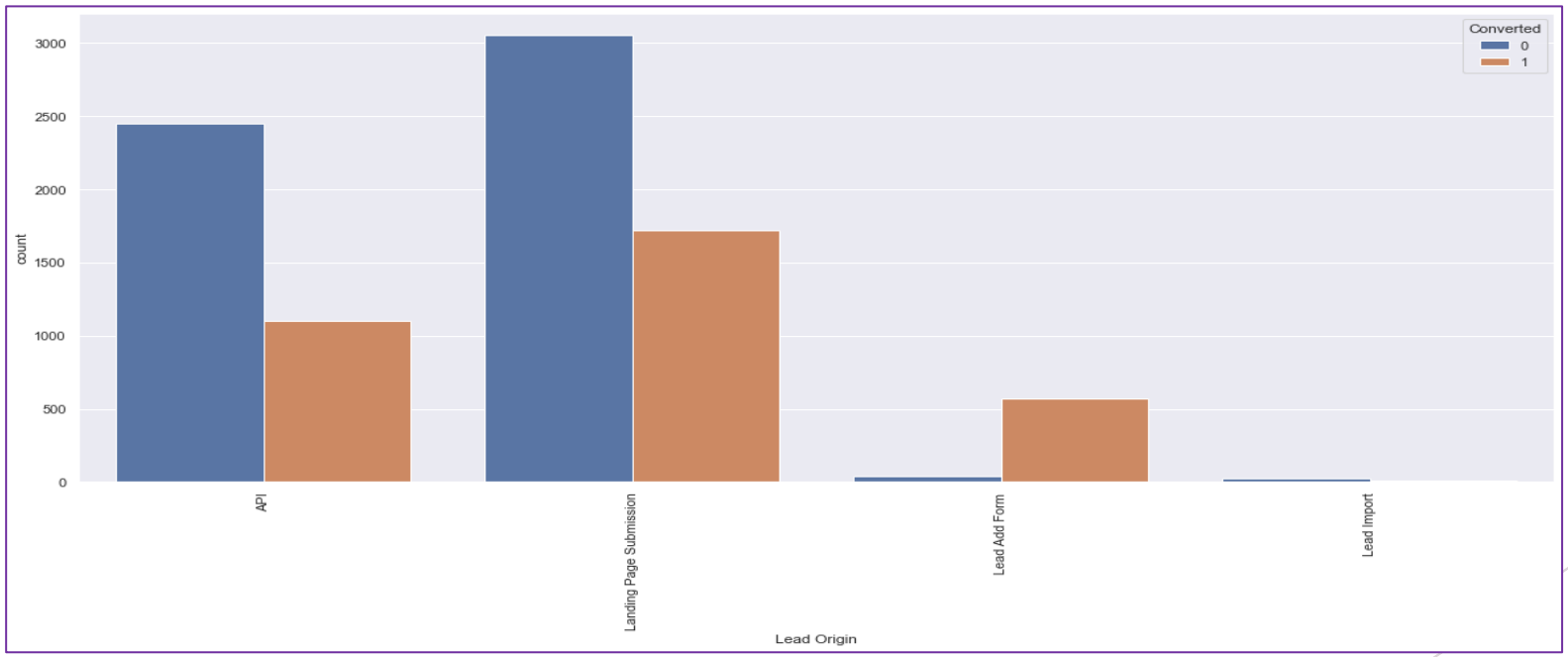
Lead Origin

•API and the Landing Page Submission bring higher number of leads and conversion rate is on the higher side.

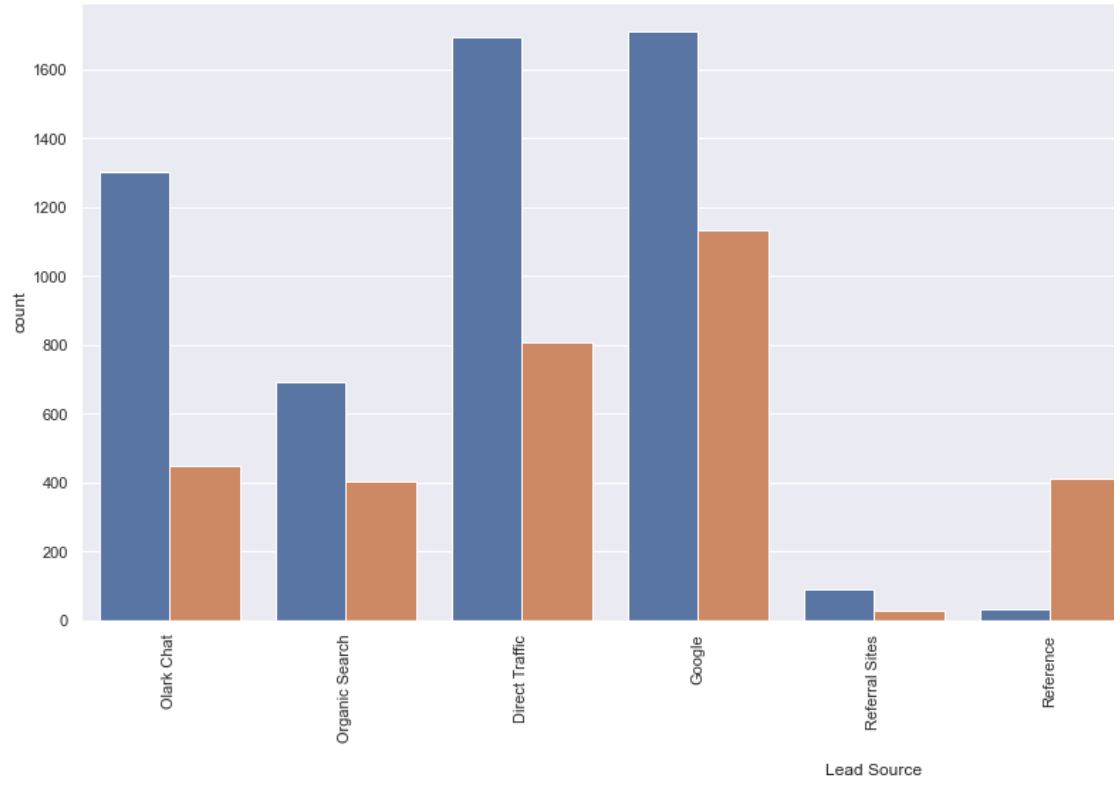
•Lead Add Form has a very high conversion rate but the count is not that big to make a decision.

•Lead Import Relatively vary low compared to other categories .

•In order to improve overall lead conversion rate, we have to improve lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



EDA

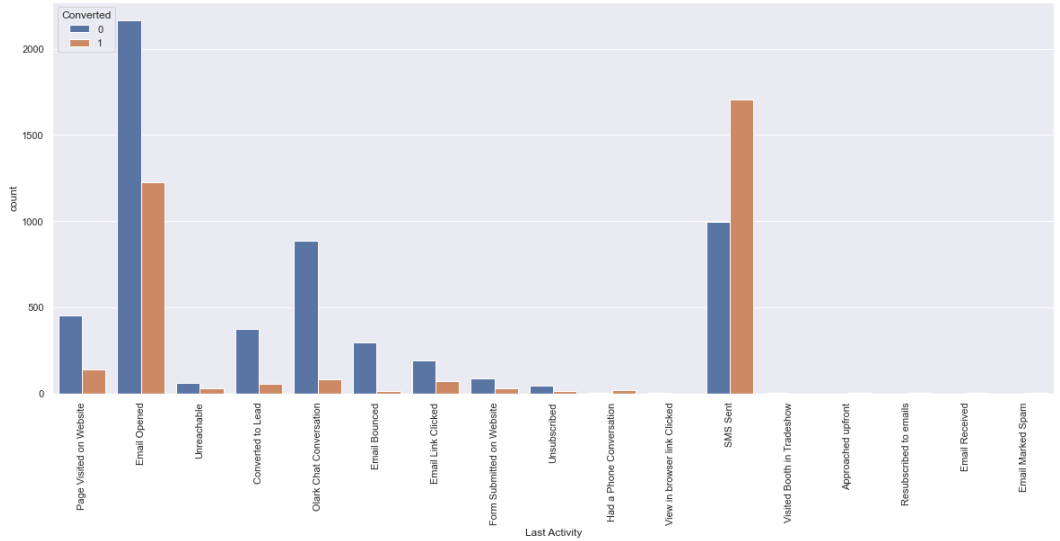


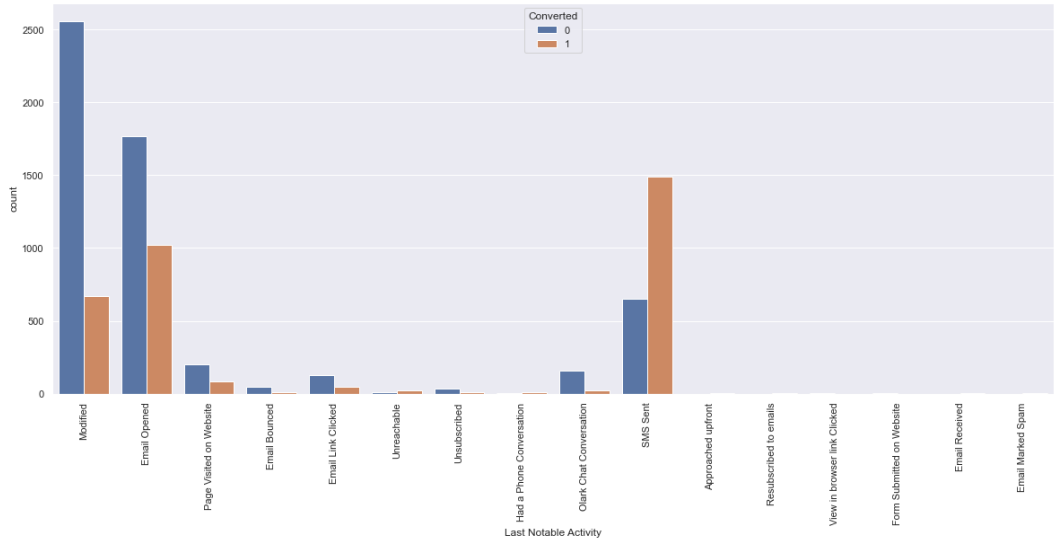
- **Lead Source**
- Most of leads are generated by Google and Direct traffic.
- Conversion Rate of reference leads and leads through welingak website is high.
- To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

EDA

Last Activity & Last Notable Activity

- SMS Sent have best conversion as compared to other options
- Email opened have most number of Leads

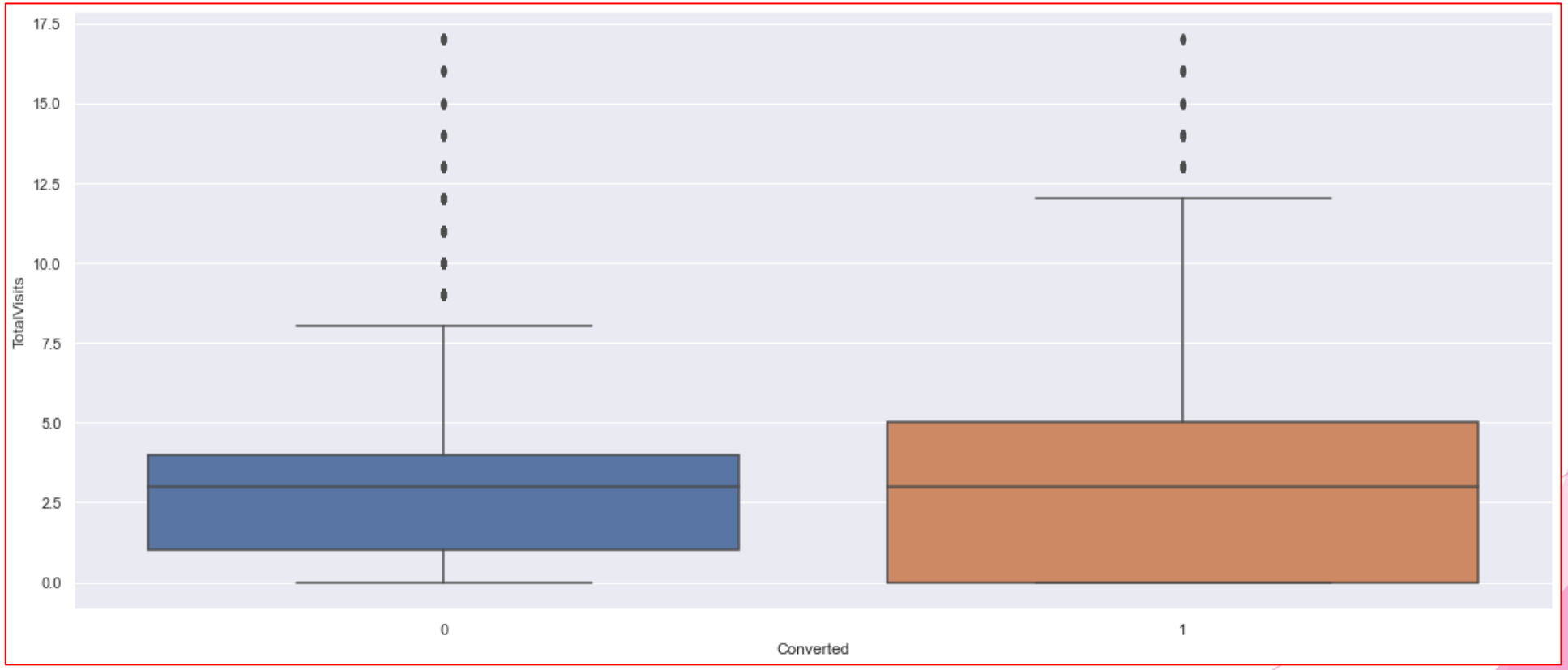




EDA

Continuous variables

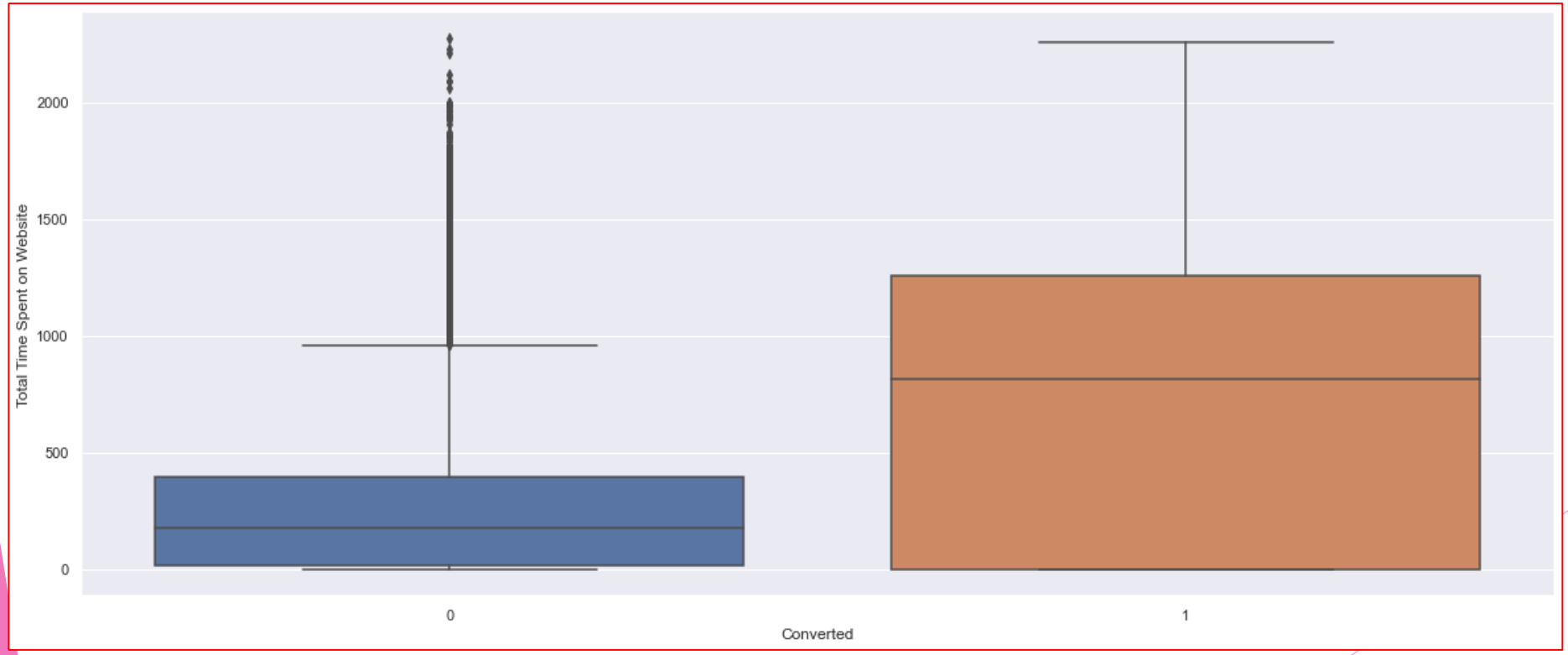
Total Visits



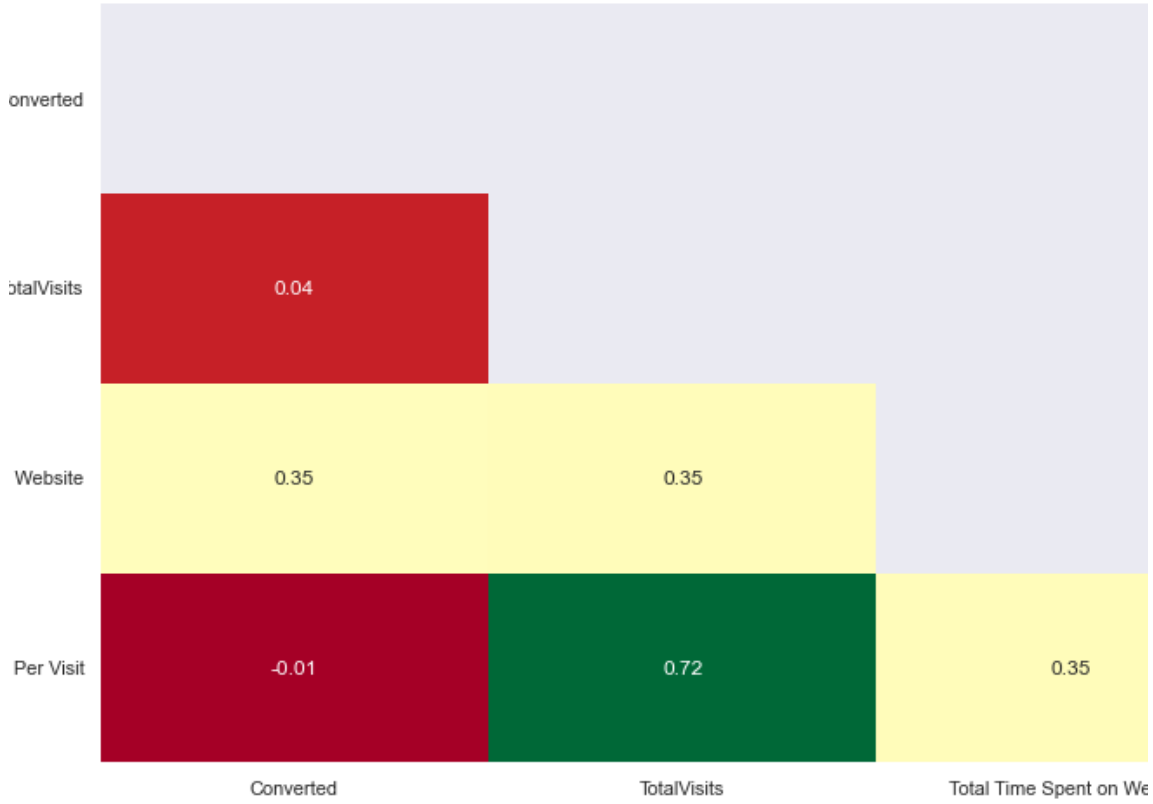
EDA

Continuous variables

Total Time spent on Websites



Correlation



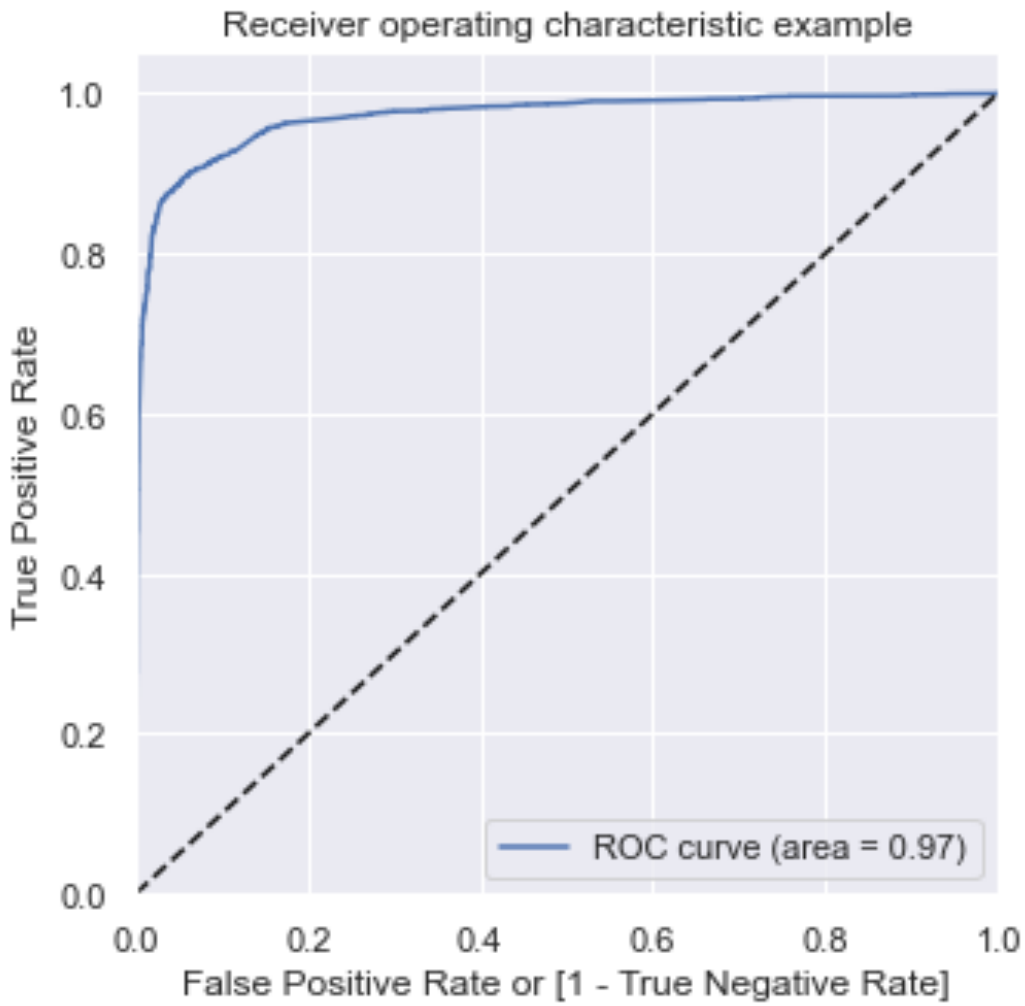
- Total Visits are having strong co relation to the Total Time Spent on website
- Conversion is negatively correlated to page view pe visits

Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Accuracy of around 92%
- sensitivity of around 87%
- specificity of around 96%

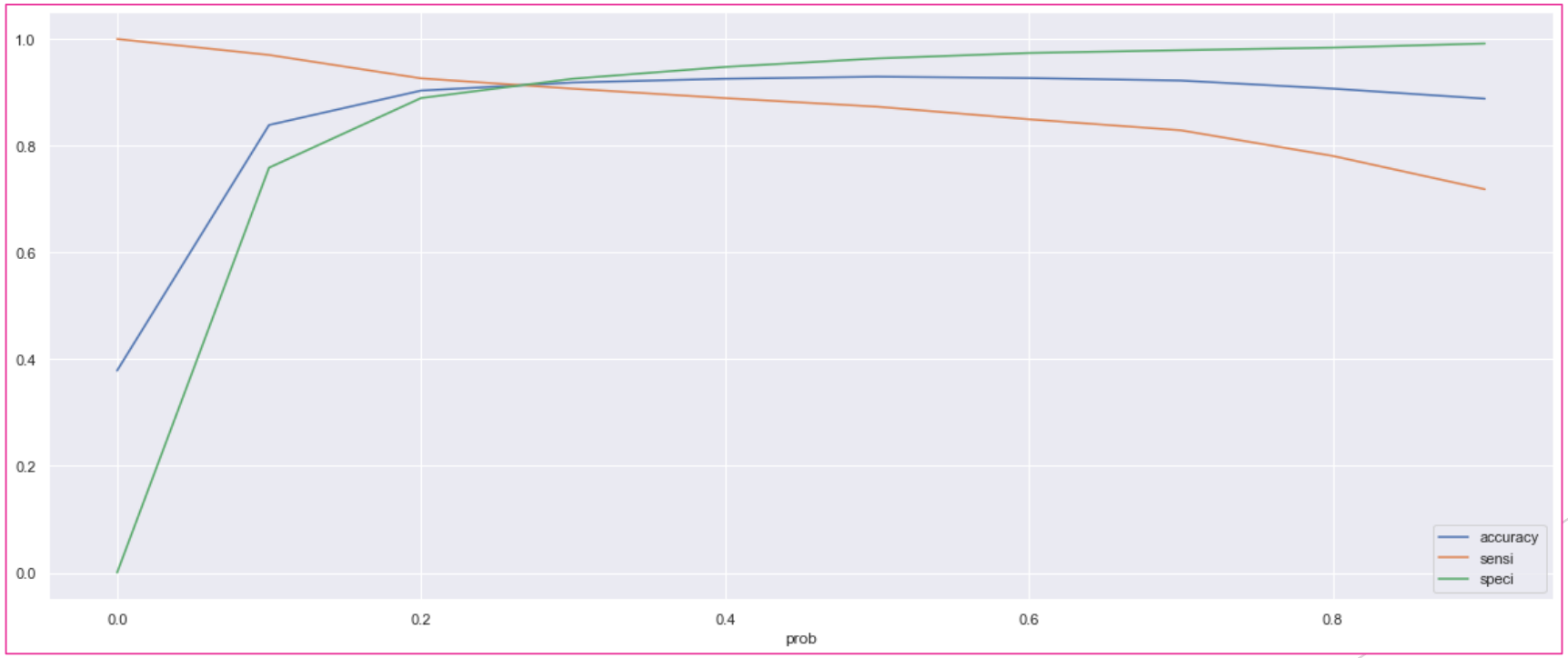
ROC Curve

- The ROC Curve should be a value close to 1
- The area under ROC curve is 0.97 which is a very good predictive model



Plotting accuracy sensitivity and specificity for various probabilities

- Optimal cut off probability is that probability where we get balanced sensitivity and specificity. optimal cut off is at 0.30



Comparing values Train Vs Test

Train
- Accuracy : 91.84%
- Sensitivity : 90.68%
- Specificity : 92.55%

Test Data
- Accuracy : 91.10%
- Sensitivity : 90.68%
- Specificity : 92.55%

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model

Conclusion

It was found that the variables that mattered the most in the potential buyers are (In descending order)

- The total time spend on the Website.
- Total number of visits.
- When the lead source was:
  - a.Olark Chat b.Welingak website
- When the last activity was:
  - a.Converted to Lead b. Email Bounced
  - c.Olark Chat Conversation
- When the lead origin is Lead add format.
- Lead Profile is Student of Some school & Unknown
- Tags assigned to customers indicating the current status of the lead:
  - a. Busy b. Closed by Horizzon c. Lost to EINS
  - d. Ringing e. Will revert after reading the email
  - f. Switched off
- Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Thank you !!