LEAD_SCORING_CASESTUDY

- Shubham Sharma

COMPANY'S PROBLEM

- X Education is a company that provides industry professionals with online courses. The company advertises its courses on numerous popular websites, including Google. X Education wants to find the most promising leads that can be turned into paying customers.
- Despite the fact that the company generates a large number of leads, only a small percentage of them are converted into paying customers, and the company would like to have a higher lead conversion rate. Leads are generated through a variety of channels, such as email, website advertisements, Google searches, and so on. By approaching those leads who expressed interest in taking the course, the company had a 30% conversion rate throughout the entire process of converting leads into customers. In terms of assisting conversions, the implementation process of lead generation attributes is inefficient.

What's our goal

- The company needs to create a model for identifying the most promising leads. Each lead should be assigned a lead score that indicates its potential. The higher the lead score, the more likely the lead will be converted; the lower the lead score, the less likely the conversion will be. The model should have an 80% or higher lead conversion rate.

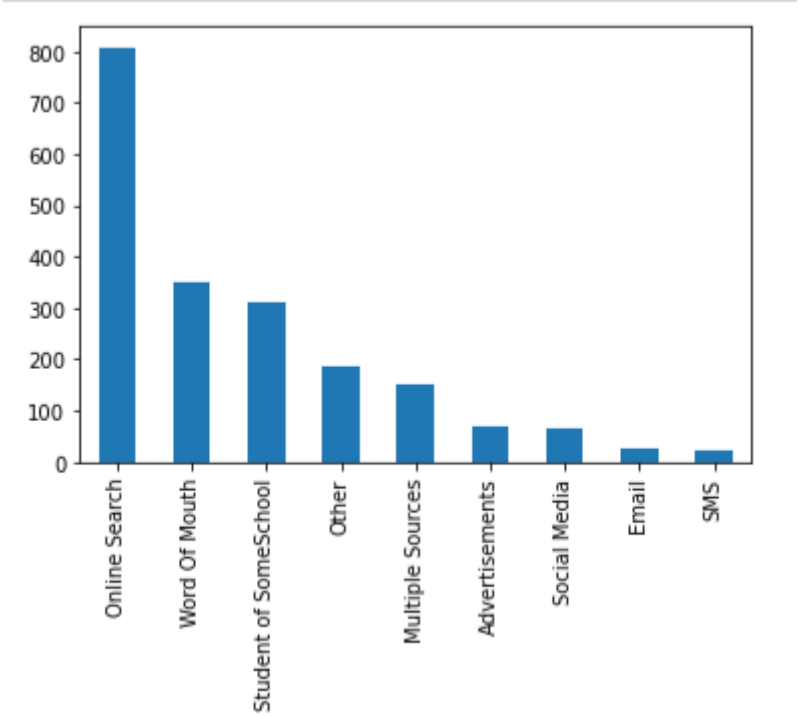
IMPORTANT RESULTS

- We have dropped the column who have null values more than 30%.
- We have dropped Country and City column because Country only contains maximum data from India.
- City contains the maximum data from Mumbai and adjoining areas.
- We are doing EDA to check more columns and their relationship.

EXPLORATORY DATA ANALYSIS

Univariate Data Analysis

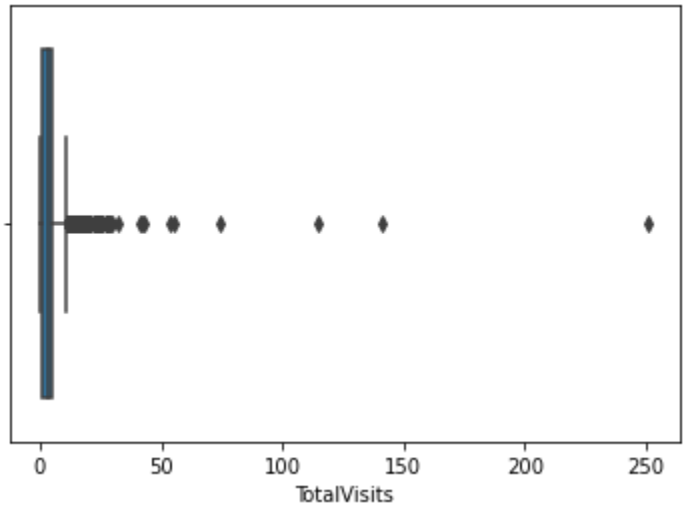
Online Search shows high conversions as compared to others.

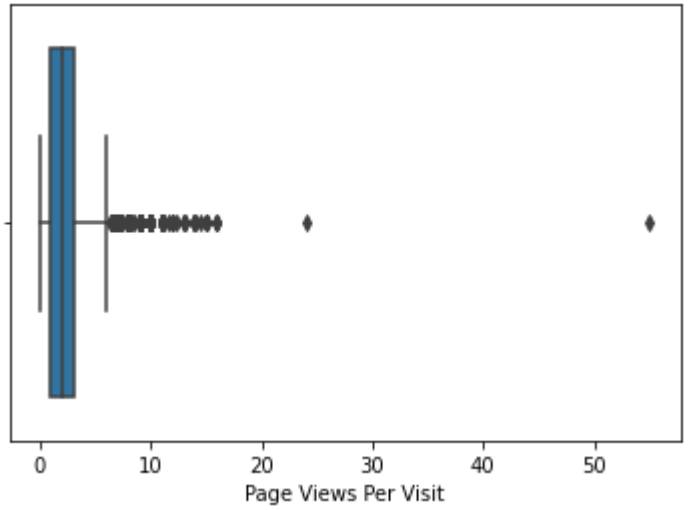


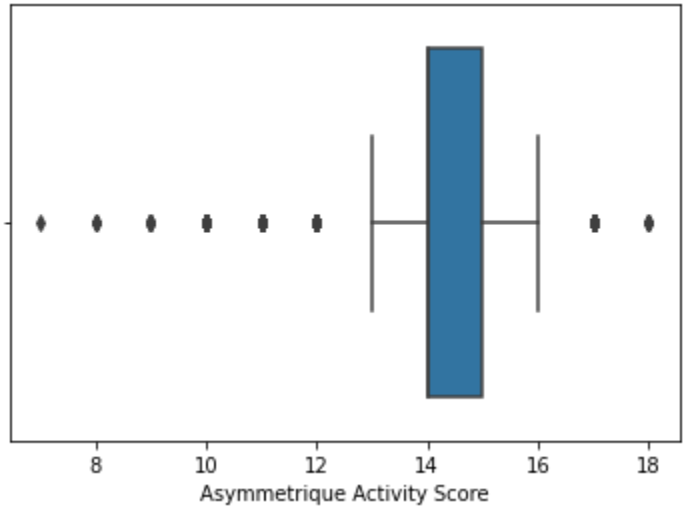
PIE CHART - LEAD QUALITY

There are a large number of people who are interested in the course but are unsure about it. There is a lot of room for improvement in order to convert them.

OUTLIERS ANALYSIS





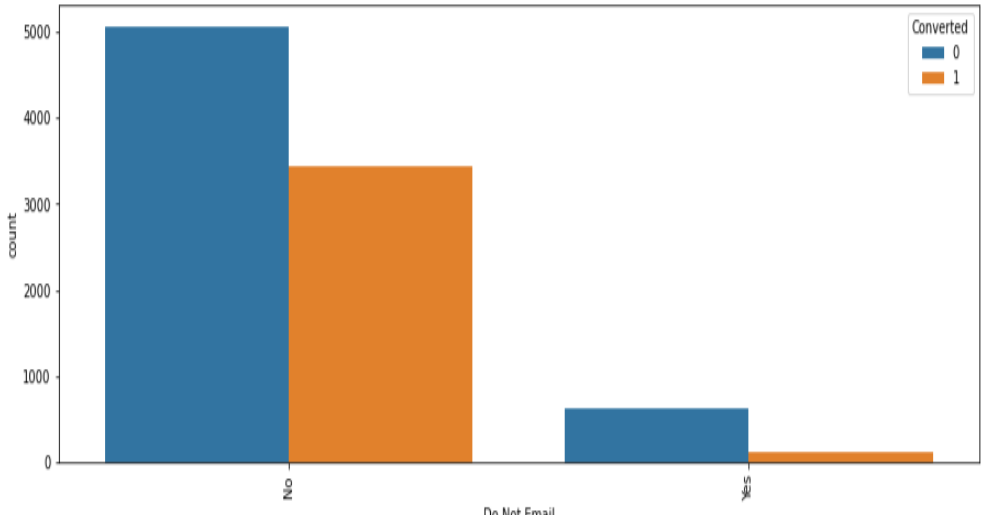


Outliers present in Totalvisits, Page views per visit and Asymmetric Activity Score.

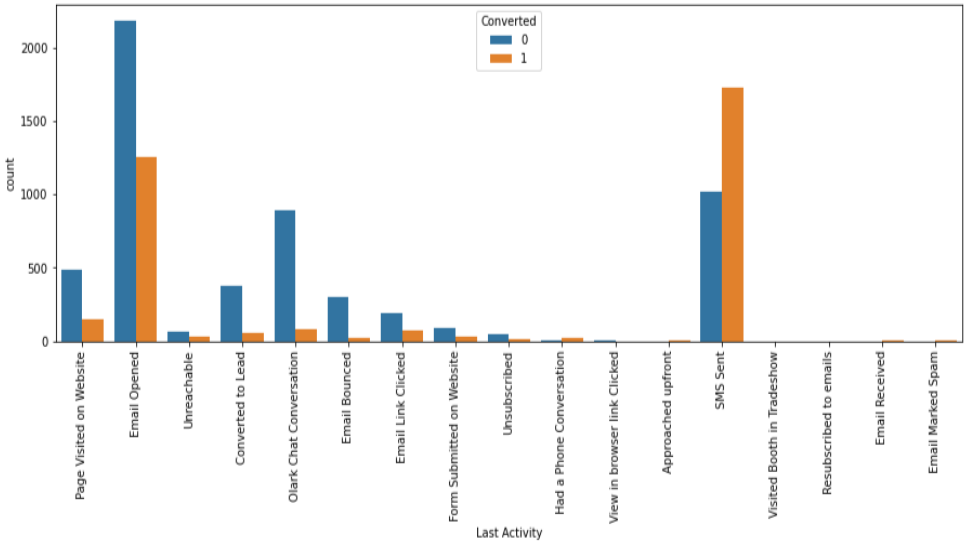
BIVARIATE ANALYSIS

Do not Email vs Converted

People who want to get updates on Email has high conversion.

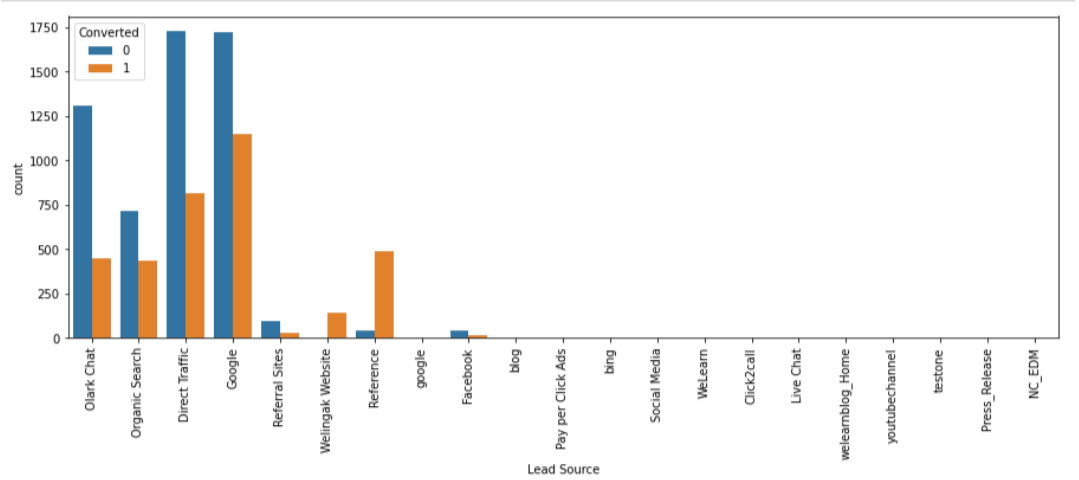


Last Activity Vs Converted



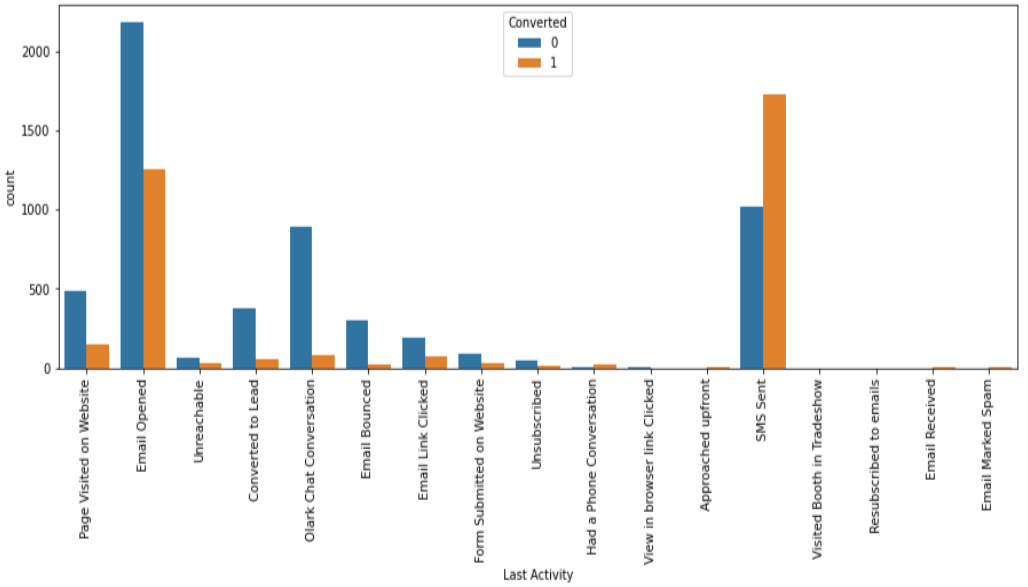
Lead Source Vs Converted

It clearly shows that Direct Searches and Google has high conversions while references has higher conversion rate.



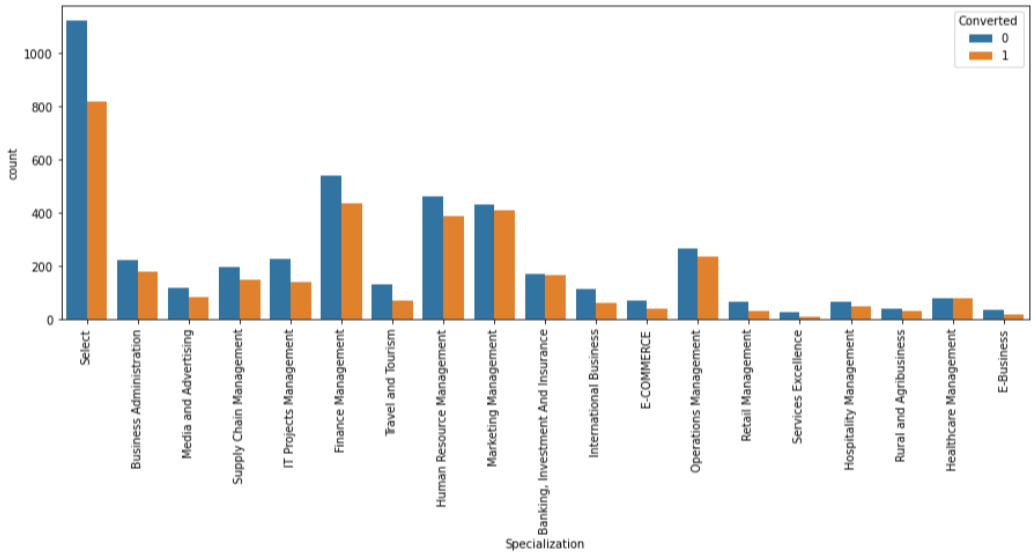
Total Visits Vs Converted

Email and Chat shows high conversion rate while SMS has high conversions.



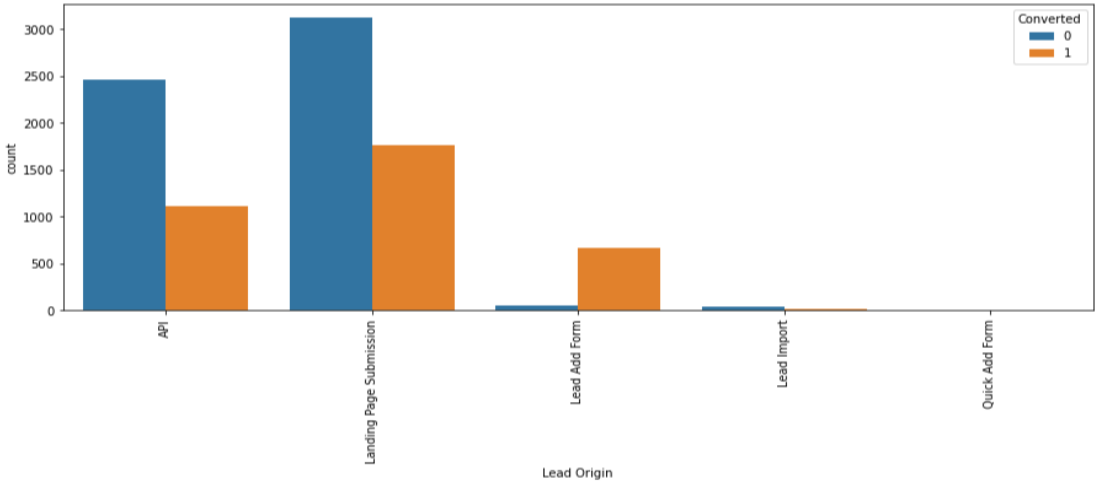
Specializations Vs Converted

Most people don’t know about specialization while people from Finance, HRD & Marketing can be promising leads.

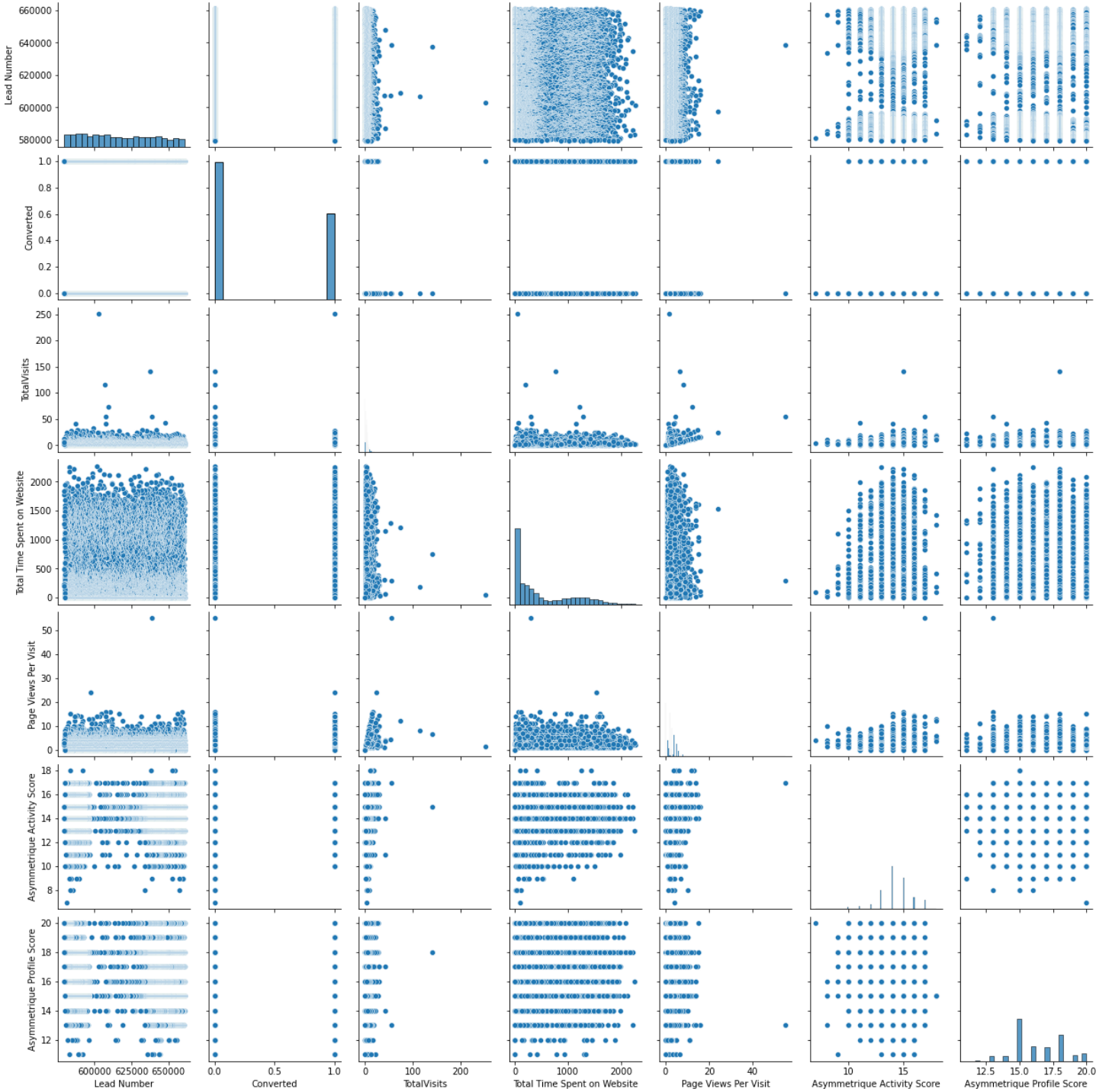


Lead Origin Vs Converted

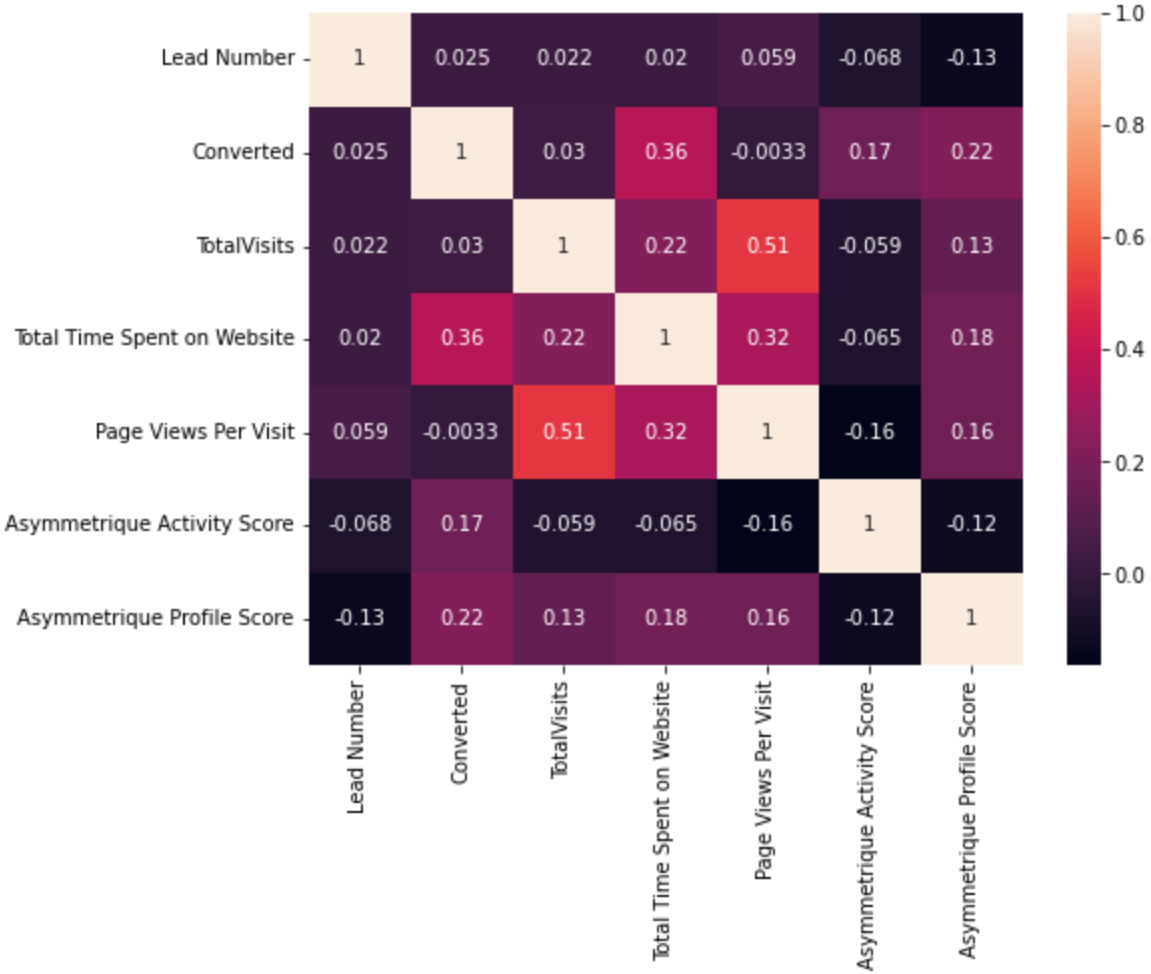
It clearly shows that landing page has high conversions as compare to others.



Pairplot


Lead Number
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Asymmetrique Activity Score
Asymmetrique Profile Score

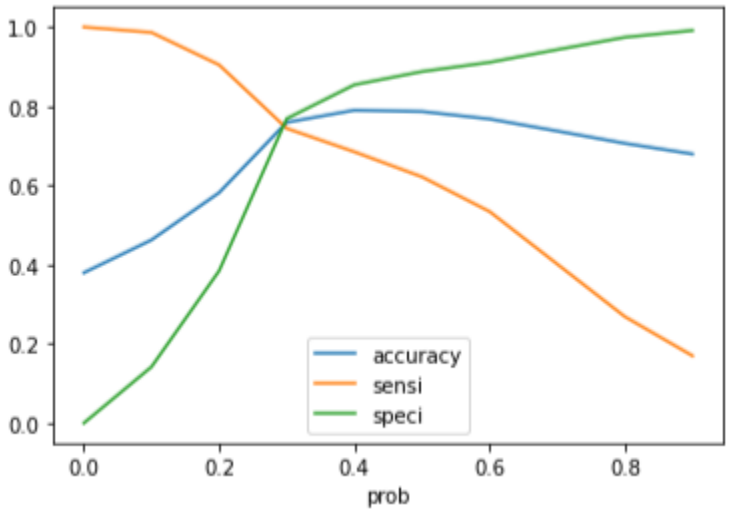
Heatmap


Lead Number
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Asymmetrique Activity Score
Asymmetrique Profile Score
1 0.025 0.022 0.02 0.059 -0.068 -0.13
0.025 1 0.03 0.36 -0.0033 0.17 0.22
0.022 0.03 1 0.22 0.51 -0.059 0.13
0.02 0.36 0.22 1 0.32 -0.065 0.18
0.059 -0.0033 0.51 0.32 1 -0.16 0.16
-0.068 0.17 -0.059 -0.065 -0.16 1 -0.12
-0.13 0.22 0.13 0.18 0.16 -0.12 1
1.0
0.8
0.6
0.4
0.2
0.0

BUILDING THE MODEL

- Scale the variables and split it into train and test.
- Built the first Model using RFE.
- RFE will eliminate the less relevant variables and build the next model.
- Eliminate those variables which have high P-values.
- By checking the VIF's of all the existing variables and predict using train set.
- Evaluate accuracy and other metric Predict using test set and check precision and recall analysis on test predictions.

EVALUATE THE MODEL(TRAIN)



Accuracy Sensitivity & Specificity

- Accuracy – 79
- Sensitivity – 80
- Specificity – 78

- Precision – 84
- Recall - 45

Model Evaluation on (Test)

- Accuracy – 78
- Sensitivity – 69
- Specificity – 83

CONCLUSION

EDA

- Potential Leads are much higher as compared to other elements.
- Direct traffic and Google searches had high conversion while references has high conversion rate.
- Most leads prefer to get updates via Email.
- SMS shows higher conversion rate whole Email and chat shows high Conversions.
- Higher the total visits higher the chances of being a hot lead or potential lead.
- People who might want the course and not sure about it high in numbers. There is great scope to work on them in order to convert them.

Logistic regression Model

- The model shows high close to 78% accuracy
- The threshold has been selected from Accuracy, Sensitivity, specificity measures and precision, recall curves.
- The model shows 69% sensitivity and 83% specificity
- The model finds correct promising leads and leads that have less chances of getting converted
- Overall this model proves to be enhance the conersions and conversion rate.