# Lead Scoring Case Study
## Identifying Hot Leads

-- Ananya Mishra
-- Rohit
-- Shashank Jaiswal

# 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 on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Now, although 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. 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.

# Problem Statement

A typical lead conversion process can be represented using the following funnel:

# Objective

1.  To generate an approach & model for identifying potential customers ( determining conversion rate of the data generated persons into 'Hot Leads' ) & selection of the persons whose conversion rate is nearly 80 percent or more .

2.  To build a model where each individual customer is given a 'potential index' based upon various factors to overall find the 'conversion rate probability '.
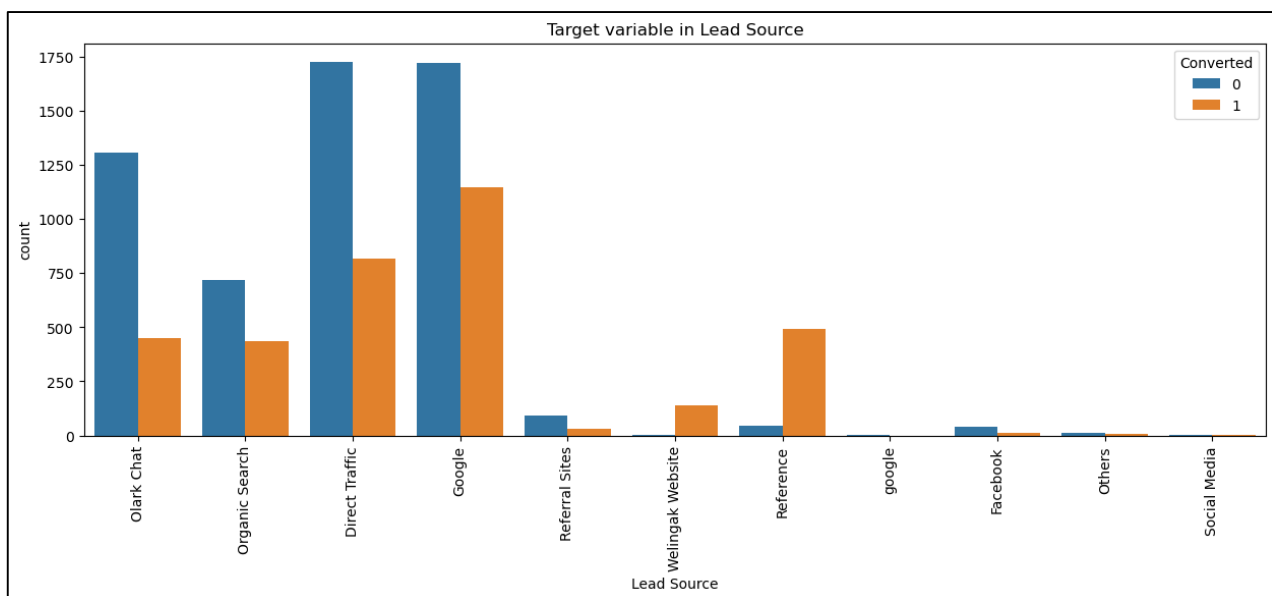
# EDA

Factors Influencing Conversion Rates (Categorical Vs Converted) :

1. Origin of Lead
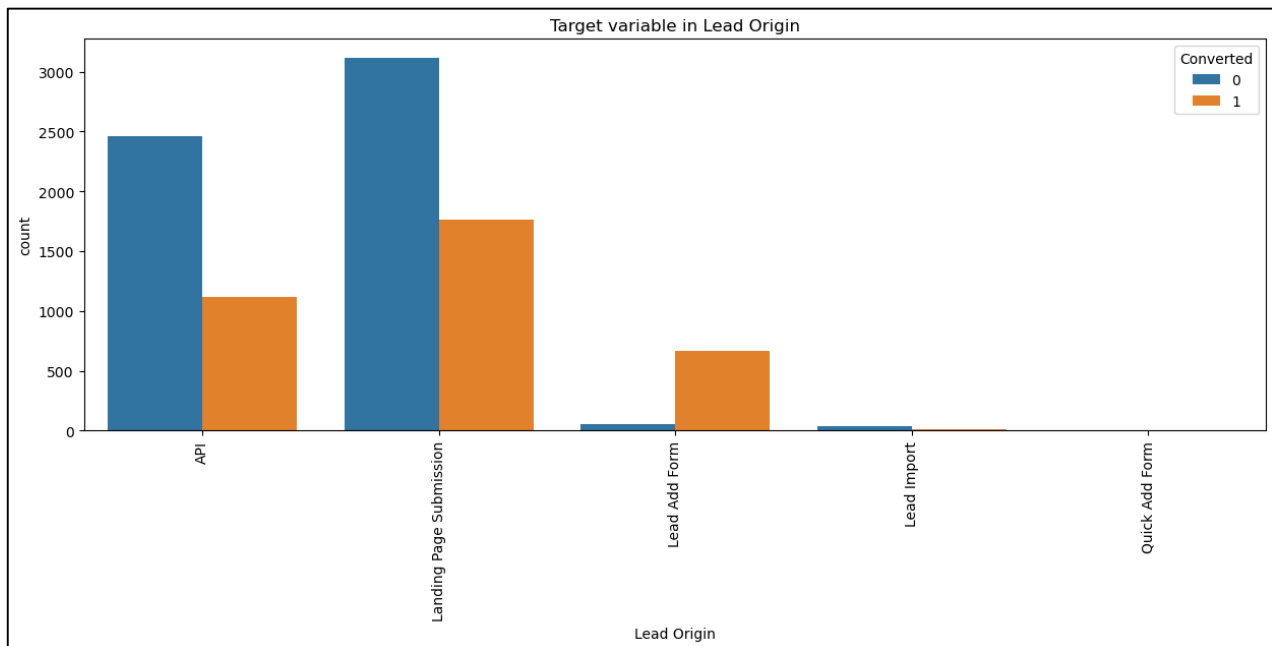


- It is evident from the dataset , Landing Page submission constitute most amount of leads .
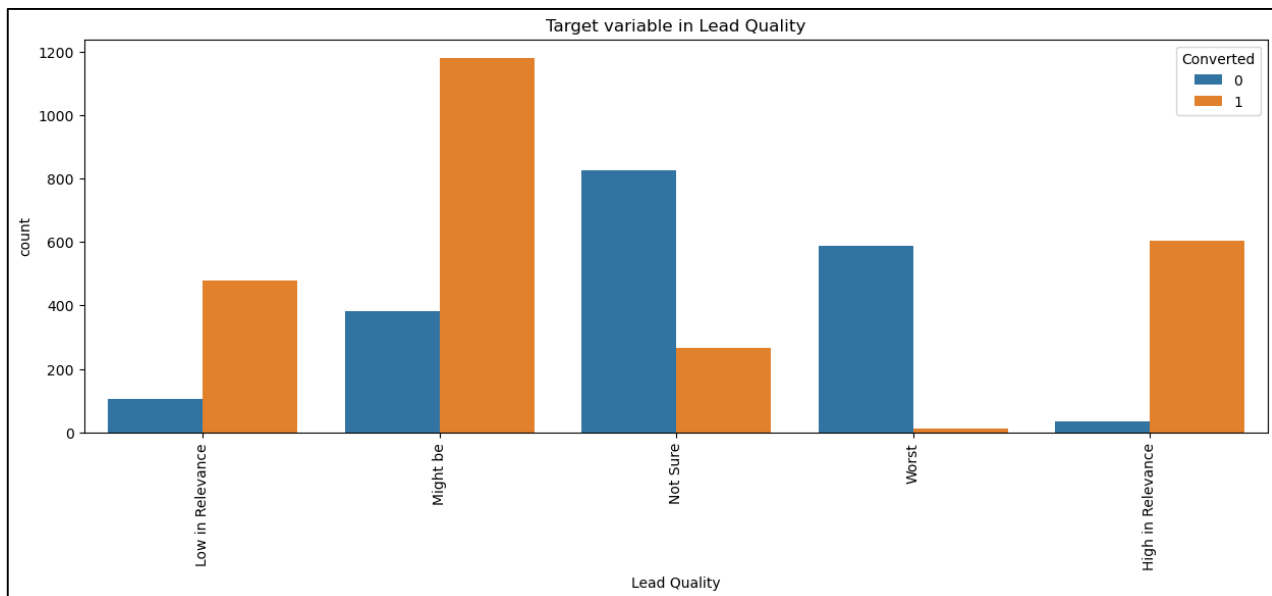- Customers through 'Lead Add Form' has the highest conversion factor.

2. Lead Source



- Target Lead Source should be 'reference' as its conversion rate is very high .
- Google & Direct traffic are also a very reliable source for Lead generation & conversion .
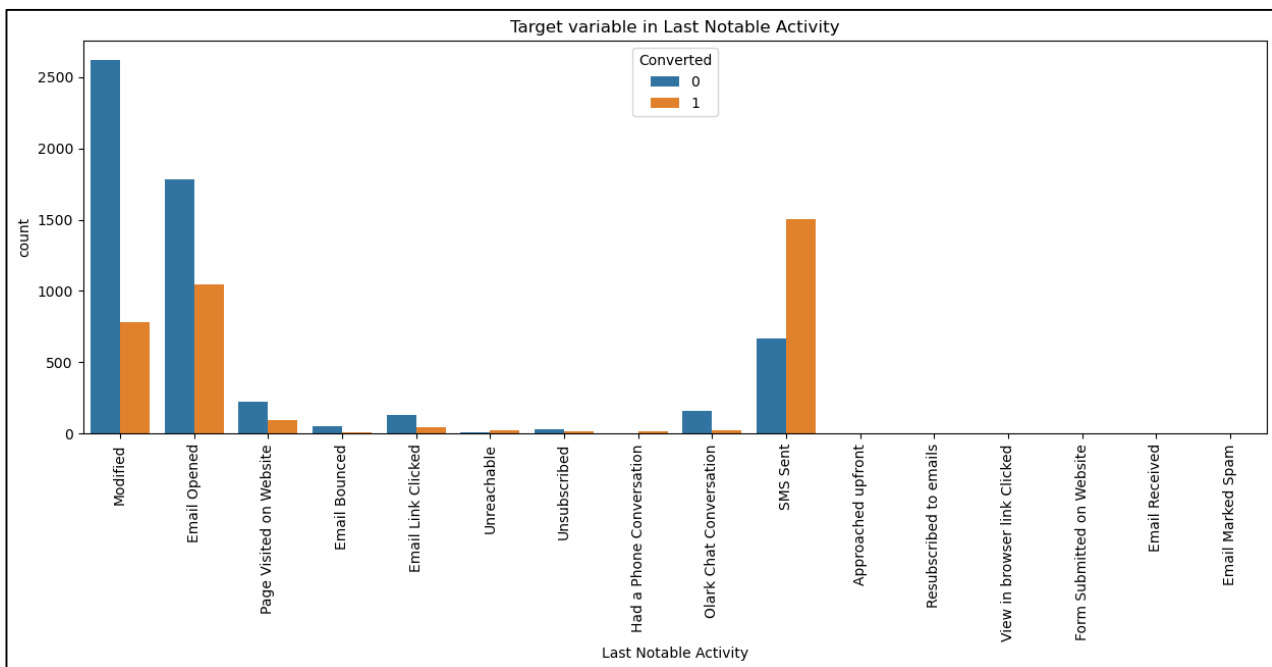
# EDA

## 3. Lead Quality



Target variable in Lead Quality

- The category 'Not sure' is having a very high conversion rate , hence the company should be focused on this category .
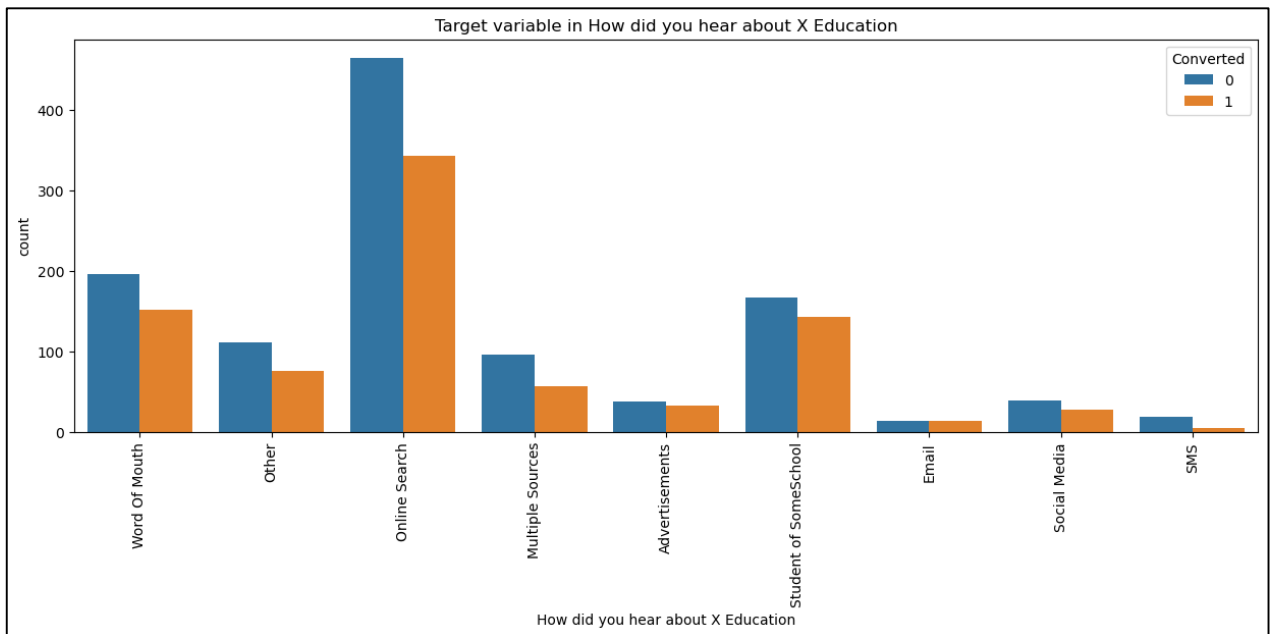- Contrary to the word 'Worst' this category also exhibits a high conversion rate .

## 4. Notable Activity



Target variable in Last Notable Activity

- People who read emails , have a high conversion and application rate .
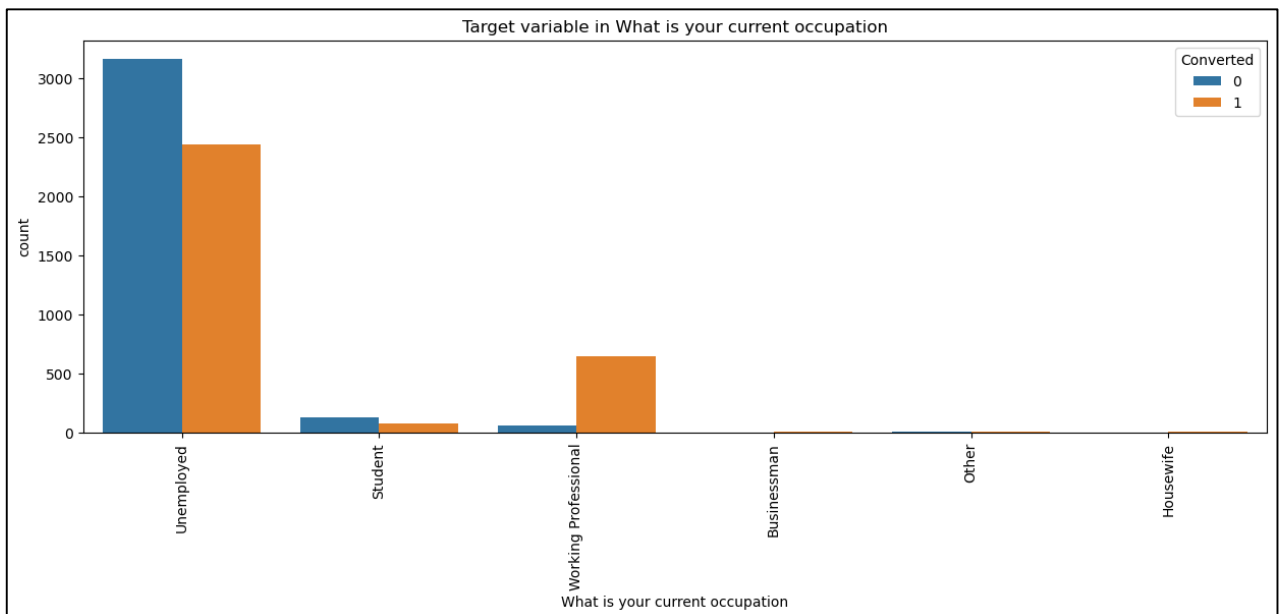- SMS is also very effective target variable category .

# EDA

## 5. Mode of Advertisement



Target variable in How did you hear about X Education

- Online Search , Word of mouth & Student of same school are the best advertisement source for lead generation which is also having high conversion rate.
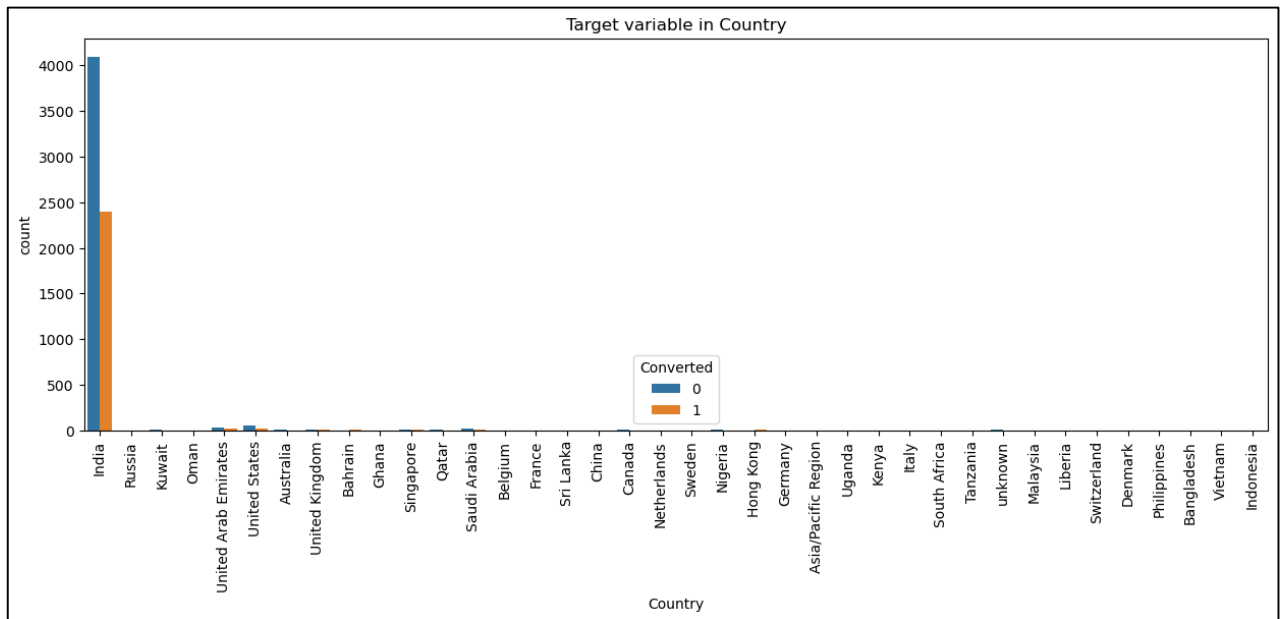
## 6. Customer Occupation



Target variable in What is your current occupation

- Unemployed people tend to take the course more and have a very strong potential as a conversion factor .
- Working Professional people are the second category to be considered.

# EDA

## 7. Country


Target variable in Country

- India is having the clear dominance in percentage of Customers .
- UAE , USA & Saudi Arabia have some scope for customers as well.

## 8. Specialization


Target variable in Specialization
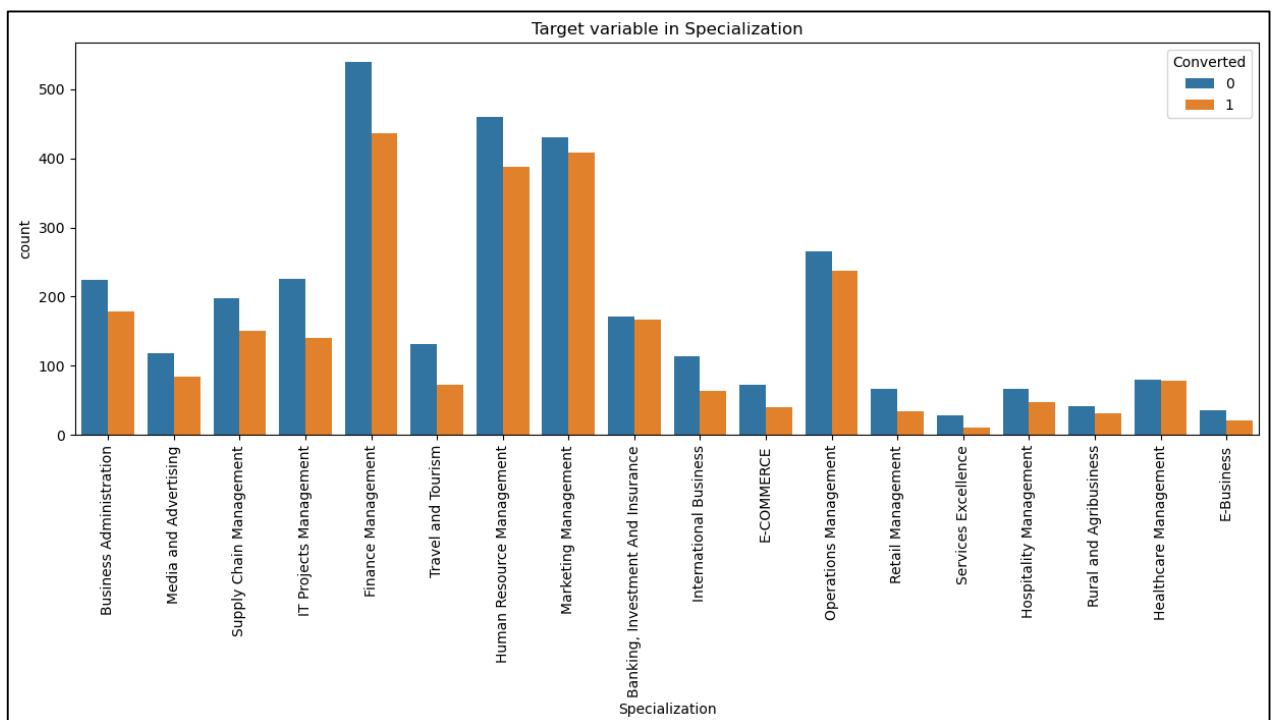
- The target segment should be people specializing in 'Management ' field , it is also evident from the graph above that it is having decent conversion rate as well.

# EDA

Multi Co-Linearity among variables :



The dependent variables do not show any strong multi co-relationship between themselves .

# FACTORS RESPONSIBLE IN DRIVING LEADS
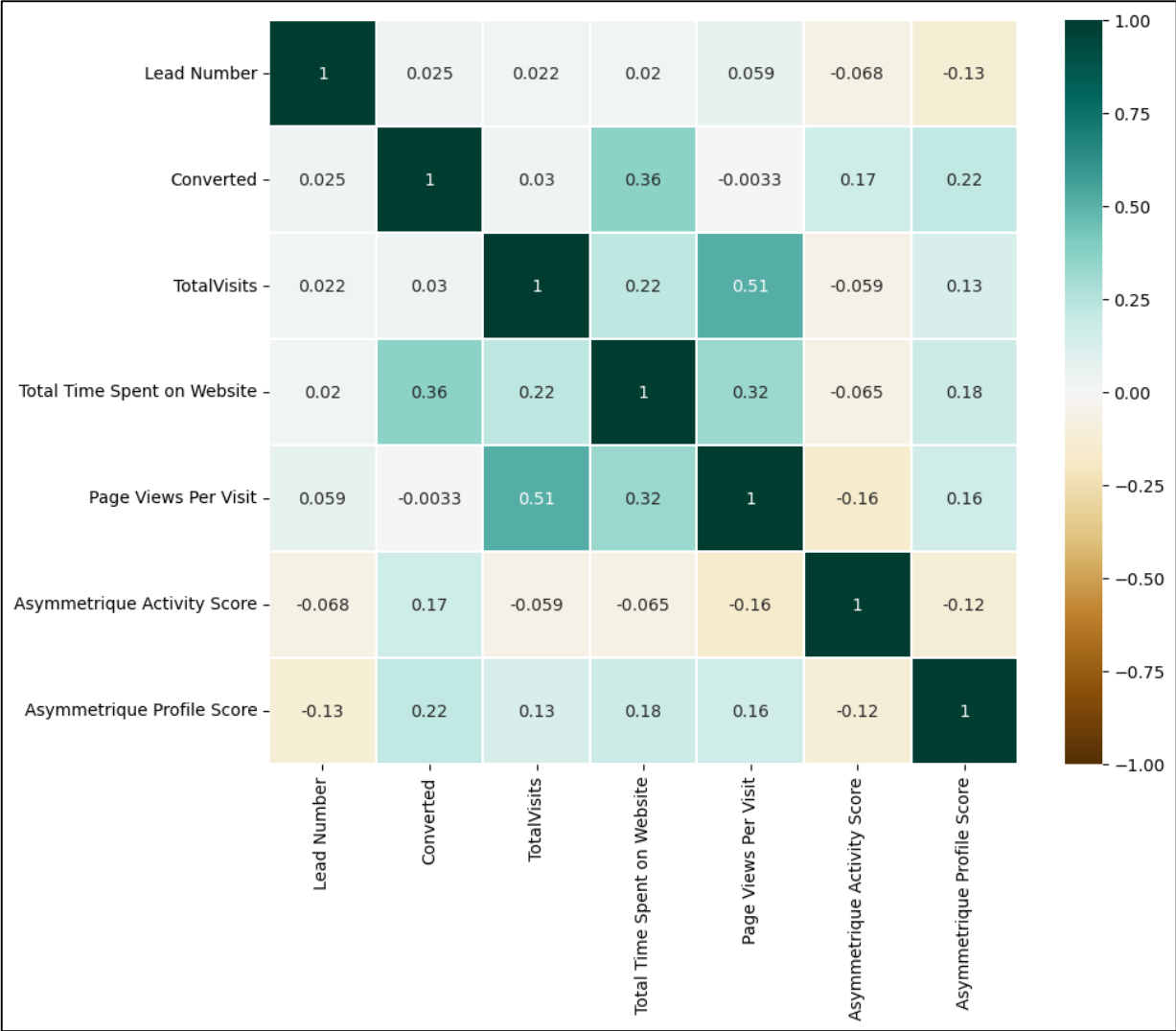
| Dep. Variable: | Converted | No. Observations: | 4461 |
|---|---|---|---|
| Model: | GLM | Df Residuals: | 4449 |
| Model Family: | Binomial | Df Model: | 11 |
| Link Function: | Logit | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -2079.1 |
| Date: | Tue, 03 Jan 2023 | Deviance: | 4158.1 |
| Time: | 20:39:51 | Pearson chi2: | 4.80e+03 |
| No. Iterations: | 7 | Pseudo R-squ. (CS): | 0.3642 |
| Covariance Type: | nonrobust | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.2040 | 0.196 | 1.043 | 0.297 | -0.179 | 0.587 |
| TotalVisits | 11.1489 | 2.665 | 4.184 | 0.000 | 5.926 | 16.371 |
| Total Time Spent on Website | 4.4223 | 0.185 | 23.899 | 0.000 | 4.060 | 4.785 |
| Lead Origin_Lead Add Form | 4.2051 | 0.258 | 16.275 | 0.000 | 3.699 | 4.712 |
| Lead Source_Olark Chat | 1.4526 | 0.122 | 11.934 | 0.000 | 1.214 | 1.691 |
| Lead Source_Welingak Website | 2.1526 | 1.037 | 2.076 | 0.038 | 0.121 | 4.185 |
| Do Not Email_Yes | -1.5037 | 0.193 | -7.774 | 0.000 | -1.883 | -1.125 |
| Last Activity_Had a Phone Conversation | 2.7552 | 0.802 | 3.438 | 0.001 | 1.184 | 4.326 |
| Last Activity_SMS Sent | 1.1856 | 0.082 | 14.421 | 0.000 | 1.024 | 1.347 |
| What is your current occupation_Student | -2.3578 | 0.281 | -8.392 | 0.000 | -2.908 | -1.807 |
| What is your current occupation_Unemployed | -2.5445 | 0.186 | -13.699 | 0.000 | -2.908 | -2.180 |
| Last Notable Activity_Unreachable | 2.7846 | 0.807 | 3.449 | 0.001 | 1.202 | 4.367 |

Below features are most important ones which are responsible for leads conversion :

1.Total Visits

2. Total Time Spent on Website

3. Lead Origin_Lead Add Form

4. Lead Source_Olark Chat

5. Do Not Email_Yes

6. Last Activity_Had a Phone Conversation

7. Last Activity_SMS Sent

8. What is your current occupation_Student

9. What is your current occupation_Unemployed

9. Last Notable Activity_Unreachable

# MODEL MATRICS

| Test Data | | |
|---|---|---|
| Confusion Metrics | | |
| | Not Converted Leads | Converted Leads |
| Not Converted Leads | 801 | 195 |
| Converted Leads | 213 | 703 |
| Accuracy | 78.60% | |
| Precision | 78.28% | |
| Recall | 76.74% | |

| Train Data | | |
|---|---|---|
| Confusion Metrics | | |
| | Not Converted Leads | Converted Leads |
| Not Converted Leads | 1852 | 460 |
| Converted Leads | 479 | 1670 |
| Accuracy | 78.95% | |
| Precision | 78.40% | |
| Recall | 77.71% | |

The Model seems to predict the Conversion Rate very well. We should be able to help the education company select the most promising Leads or the Hot Leads..

# CONCLUSION

- From the model , it is conclusive that the following Categories are very important for considering the factor of conversion :

1. Total Visits : The more number of visits by a customer depicts more interest towards the course .

2. Total time spent on website : As inferred from the EDA analysis , the time spent on website is directly proportional to the conversion rate .

3. Lead Origin Lead Add Form : The origin of Lead plays a major role in determining his inclination towards the course .

- For the growth of the company , they need to focus more on the customers with a higher rate of conversion factor (Greater than 75 %)