# Lead Score Assignment Case Study

Submitted by : Akansha Pruthi,

Hariharan & Lau

# Problem Statement

X Education sells online courses to industry professionals. Many experts who are interested in the courses land on their website and browse for courses. The company advertised using websites and Google search engines. once the visitors land on their website and browse through the various courses and fill up the enquiry form with email and phone number. These enquires converts into leads for X Education. They also get some leads from past recommendations. Once this data is captured by sales team, they start reaching them out through emails and calls. Some leads are converted during the calls, but most are unresponsive.

Through this process, 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**

# Objective

Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

# Index

- Steps Performed in IPYNB
- Understanding the data
- Exploratory Data Analysis ( Univariate, Bi-variate & Multivariate Analysis)
- Model Building, Evaluation & Prediction
- Insights & recommendations
- Conclusion

# Steps Performed in Notebook

- Reading and Understanding the data
- EDA (Exploratory Data Analysis)
- Creating dummy variable
- Splitting the Data into Training and Testing Sets
- Feature Scaling using Min/Max Scaling
- Looking at Correlations
- Feature Selection Using RFE
- Model Building – using Logistic regression
- Model Prediction & Evaluation
- Plotting the ROC Curve ('Receiver Operating Characteristic' Curve)
- Making predictions on the Test set

# Understanding the data

**Data types:** There are 4 float64, 3 int64 and 30 object data types.

**Data shape:** 9,240 rows and 36 columns in the dataset.

**Target variable:** Converted

**Missing values:**

| | |
|---|---|
| Lead Source | 36 |
| Last Activity | 103 |
| Page Views Per Visit | 137 |
| TotalVisits | 137 |
| City | 1420 |
| Specialization | 1438 |
| How did you hear about X Education | 2207 |
| Country | 2461 |
| What is your current occupation | 2690 |
| Lead Profile | 2709 |
| What matters most to you in choosing a course | 2709 |
| Tags | 3353 |
| Asymmetrique Activity Index | 4218 |
| Asymmetrique Profile Index | 4218 |
| Asymmetrique Activity Score | 4218 |
| Asymmetrique Profile Score | 4218 |
| Lead Quality | 4767 |

Figure 1 represents the outliers present in the dataset: TotalVisits, Page Views Per Visit

Table in this slide is the descriptive statistics of numeric variables after treating the outliers and imputing the missing values.

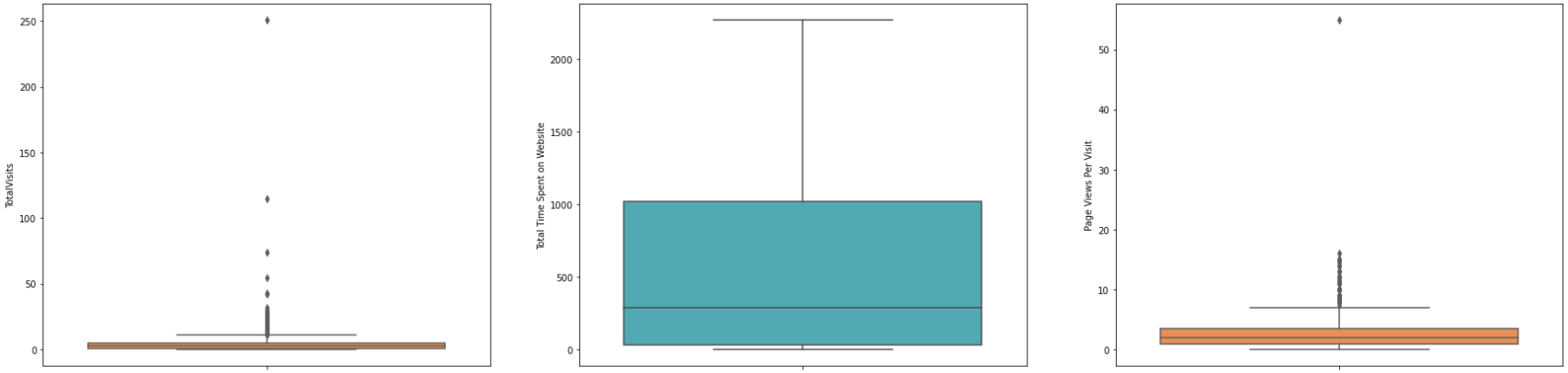



Fig. 1

| | Do Not Email | Converted | TotalVisits | Total Time Spent on Website | Page Views Per Visit | A free copy of Mastering The Interview |
|---|---|---|---|---|---|---|
| **count** | 6226.000000 | 6226.000000 | 6226.000000 | 6226.000000 | 6226.000000 | 6226.000000 |
| **mean** | 0.066817 | 0.480726 | 3.270800 | 530.808063 | 2.347920 | 0.334083 |
| **std** | 0.249724 | 0.499669 | 2.907667 | 565.111571 | 1.828861 | 0.471707 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| **25%** | 0.000000 | 0.000000 | 1.000000 | 30.000000 | 1.000000 | 0.000000 |
| **50%** | 0.000000 | 0.000000 | 3.000000 | 281.000000 | 2.000000 | 0.000000 |
| **75%** | 0.000000 | 1.000000 | 5.000000 | 1020.000000 | 3.250000 | 1.000000 |
| **90%** | 0.000000 | 1.000000 | 7.000000 | 1421.000000 | 5.000000 | 1.000000 |
| **95%** | 1.000000 | 1.000000 | 9.000000 | 1590.750000 | 6.000000 | 1.000000 |
| **99%** | 1.000000 | 1.000000 | 13.000000 | 1848.500000 | 7.000000 | 1.000000 |
| **max** | 1.000000 | 1.000000 | 17.000000 | 2272.000000 | 8.500000 | 1.000000 |

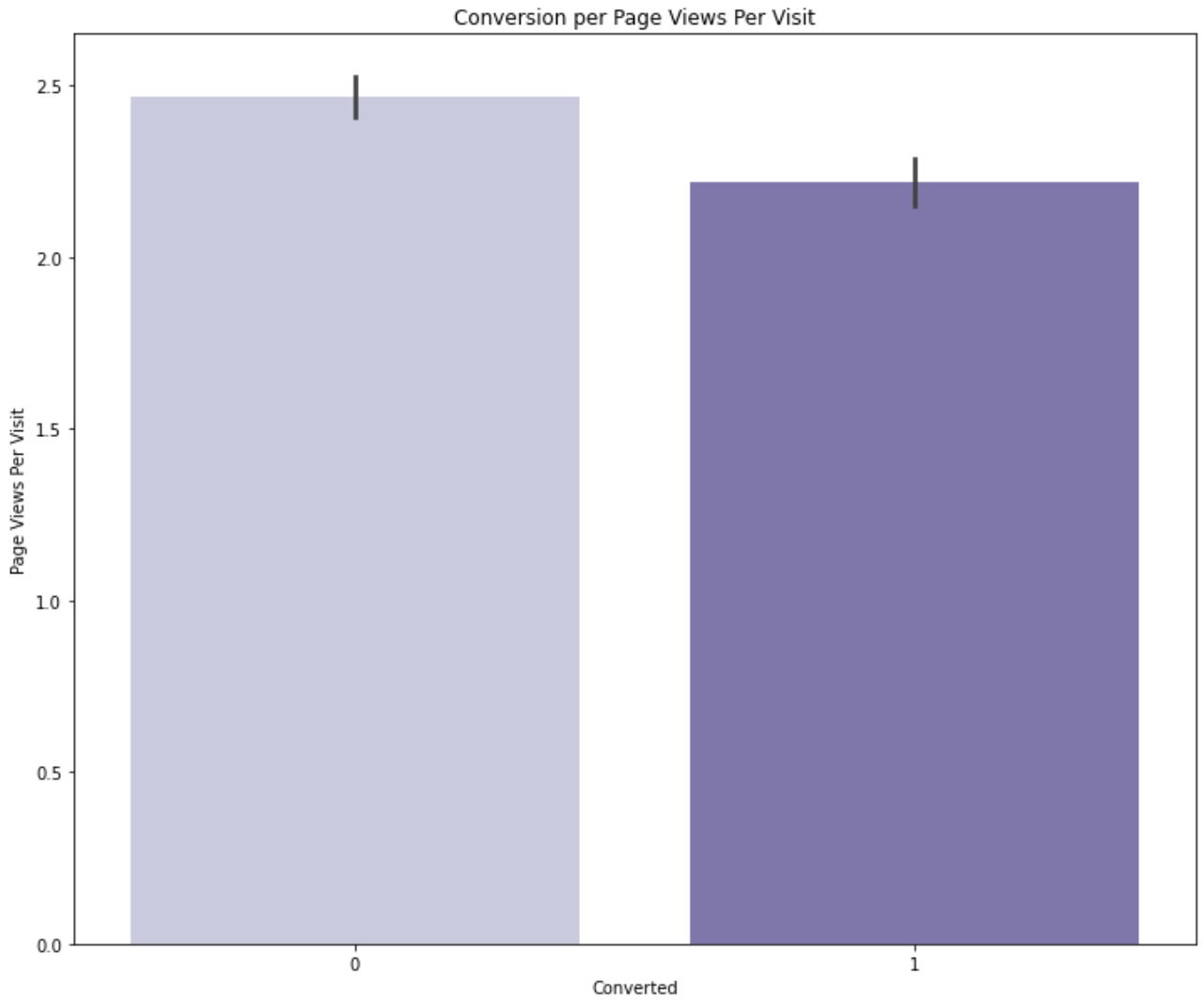
# Exploratory Data Analysis

Univariate Analysis

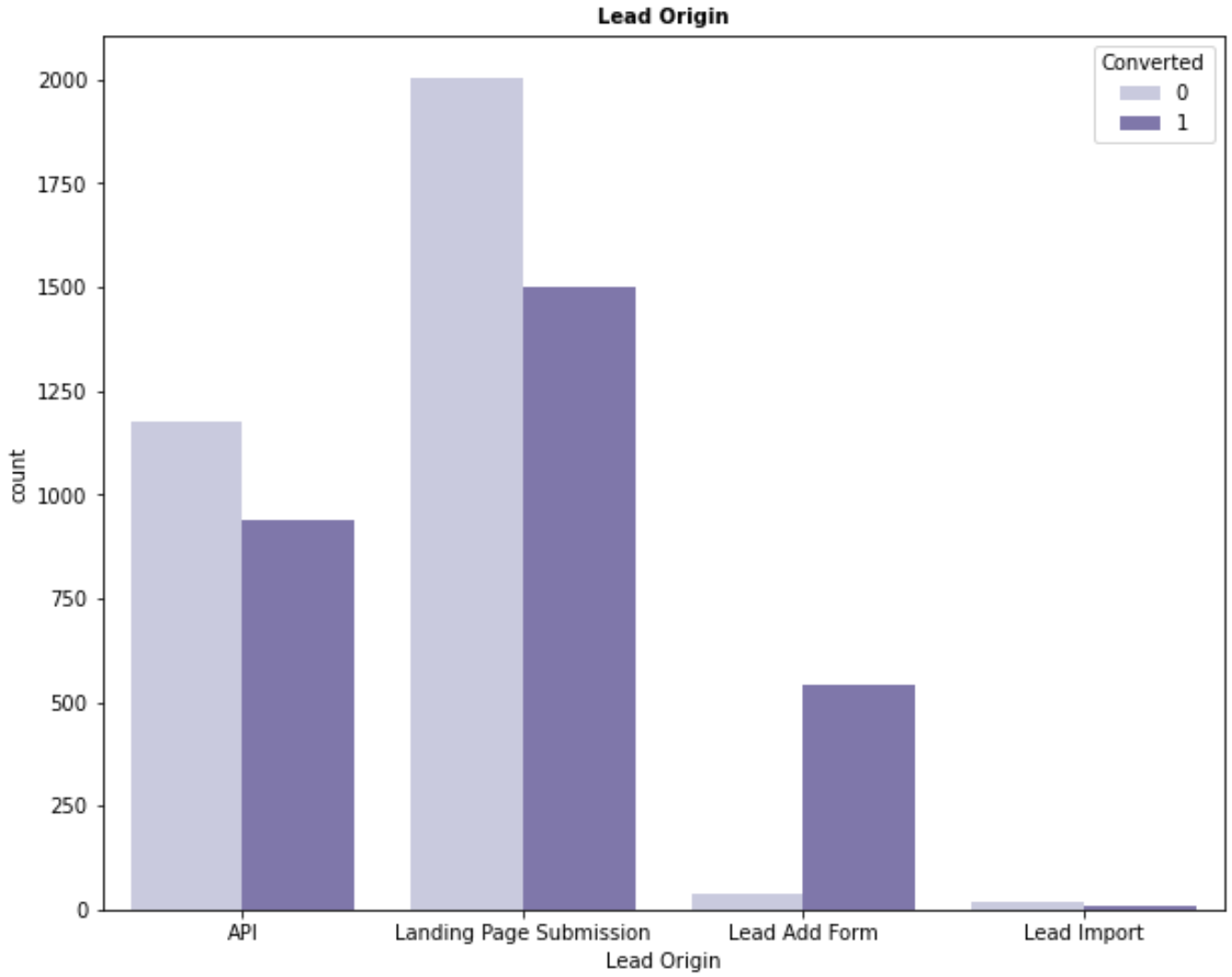
Bi-variate Analysis

Multivariate Analysis

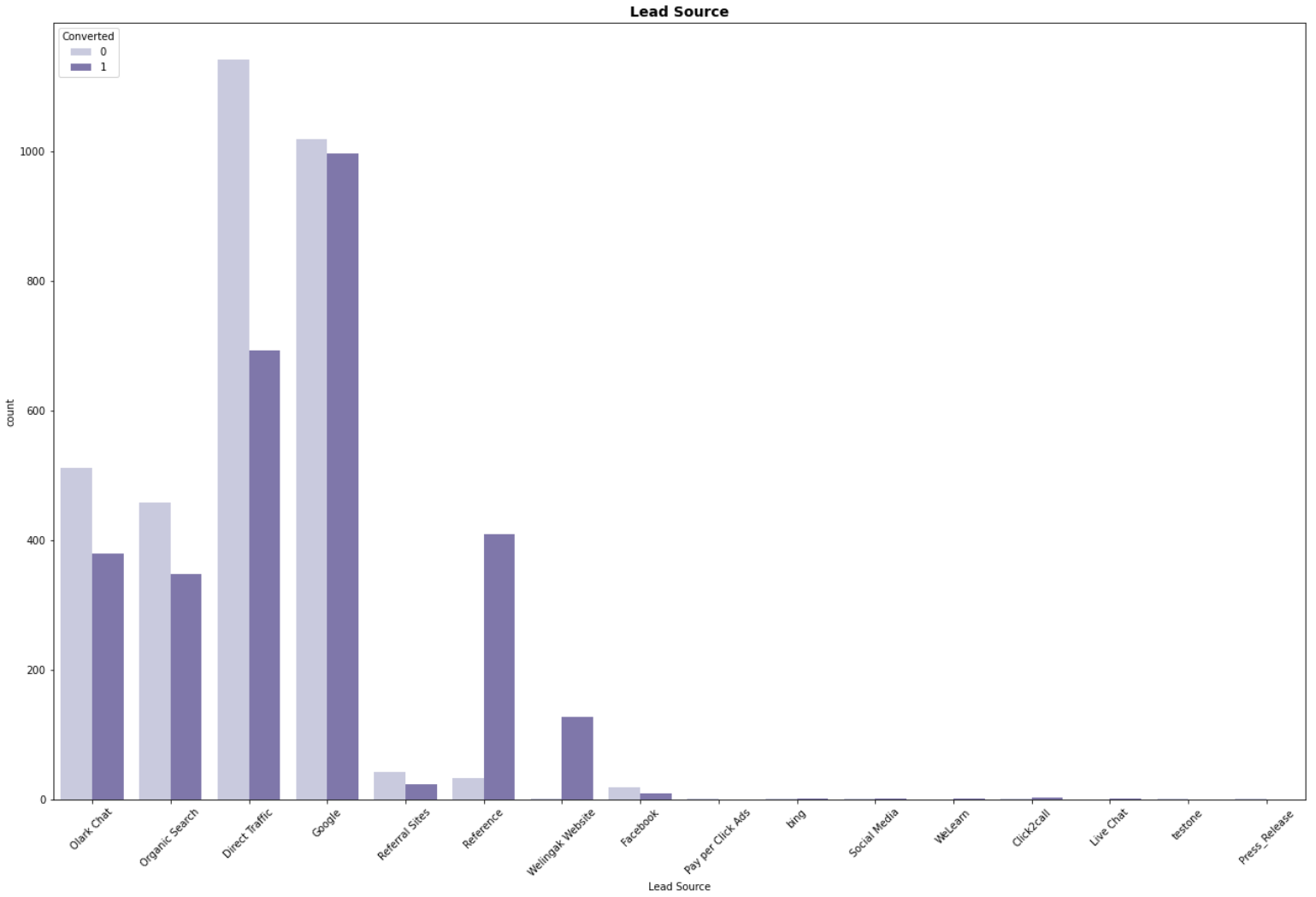
Page views per visits are more for leads not converted comparatively to converted.



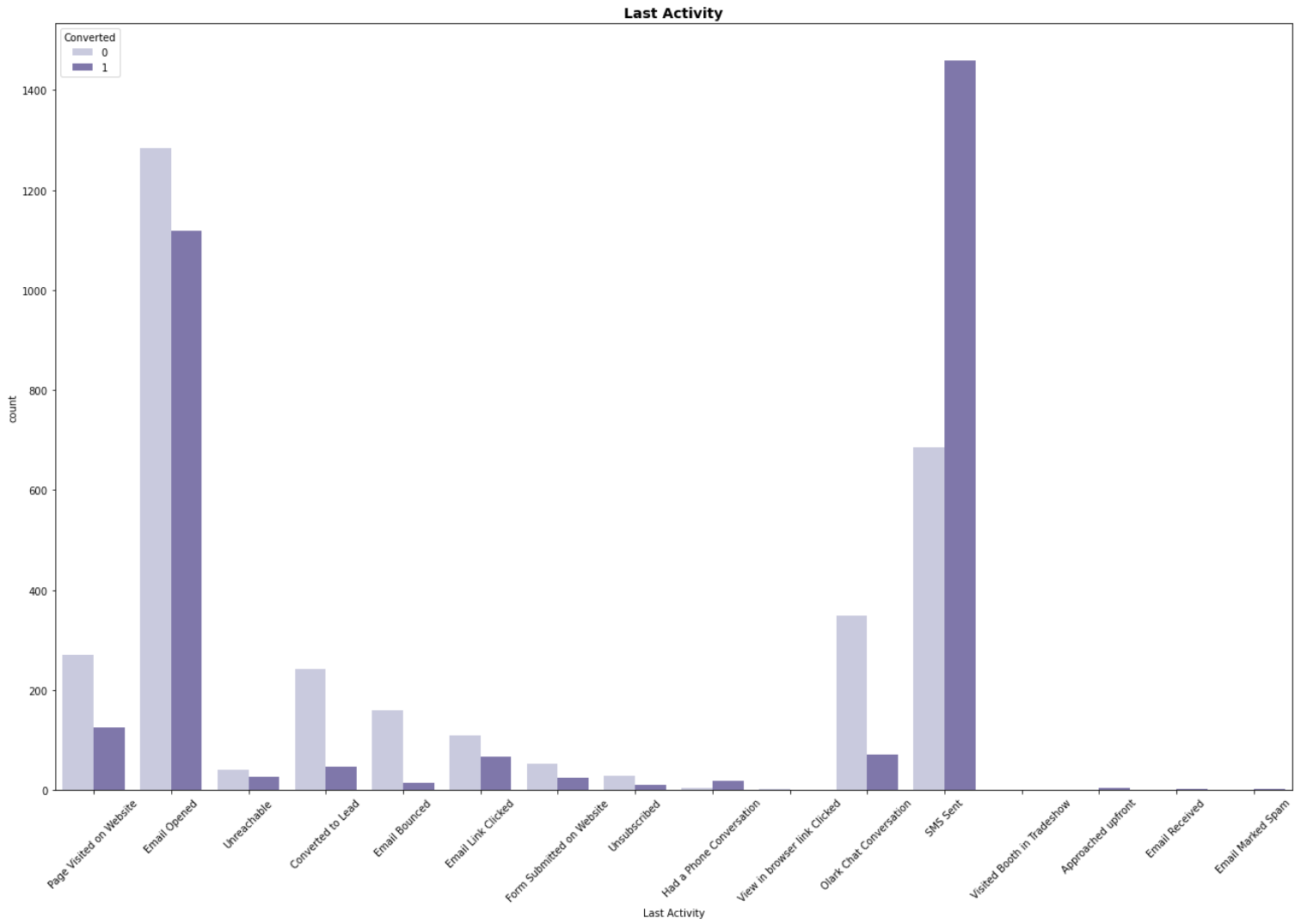
The Lead Origin- Landing Page Submission has the highest conversion rate among others.



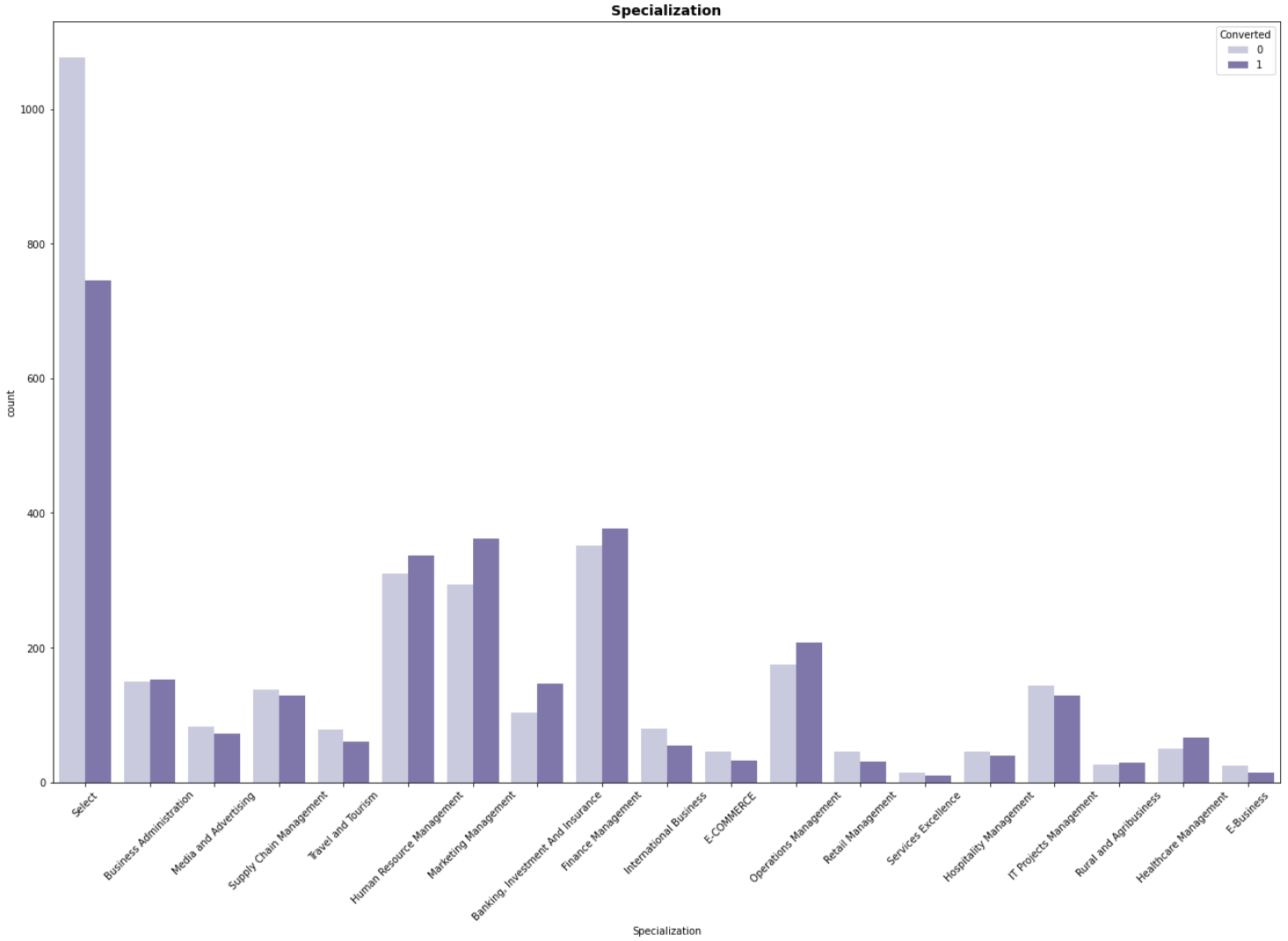
Google has the highest conversion rate.



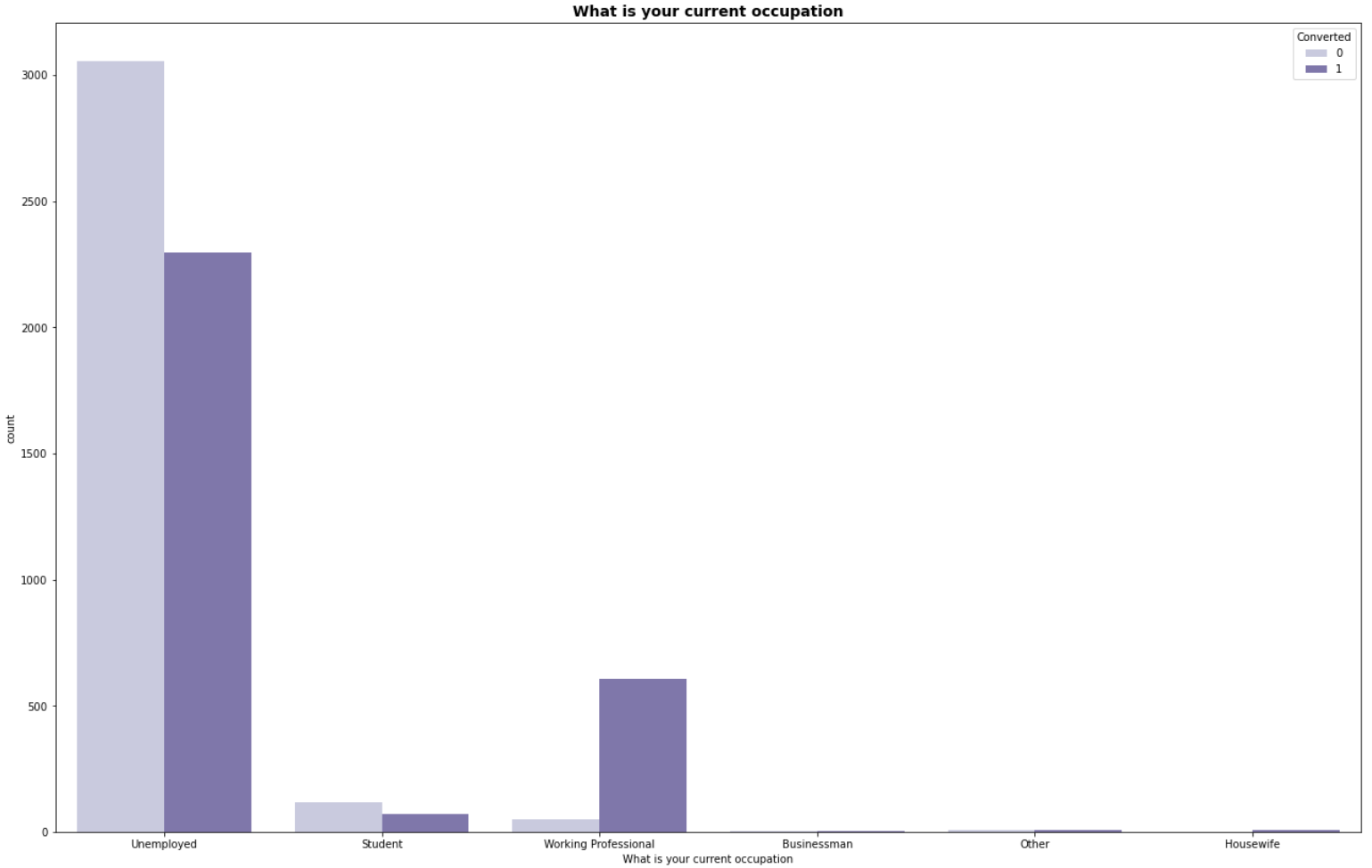
Leads whose Last Activity was SMS sent had the best conversion rate.



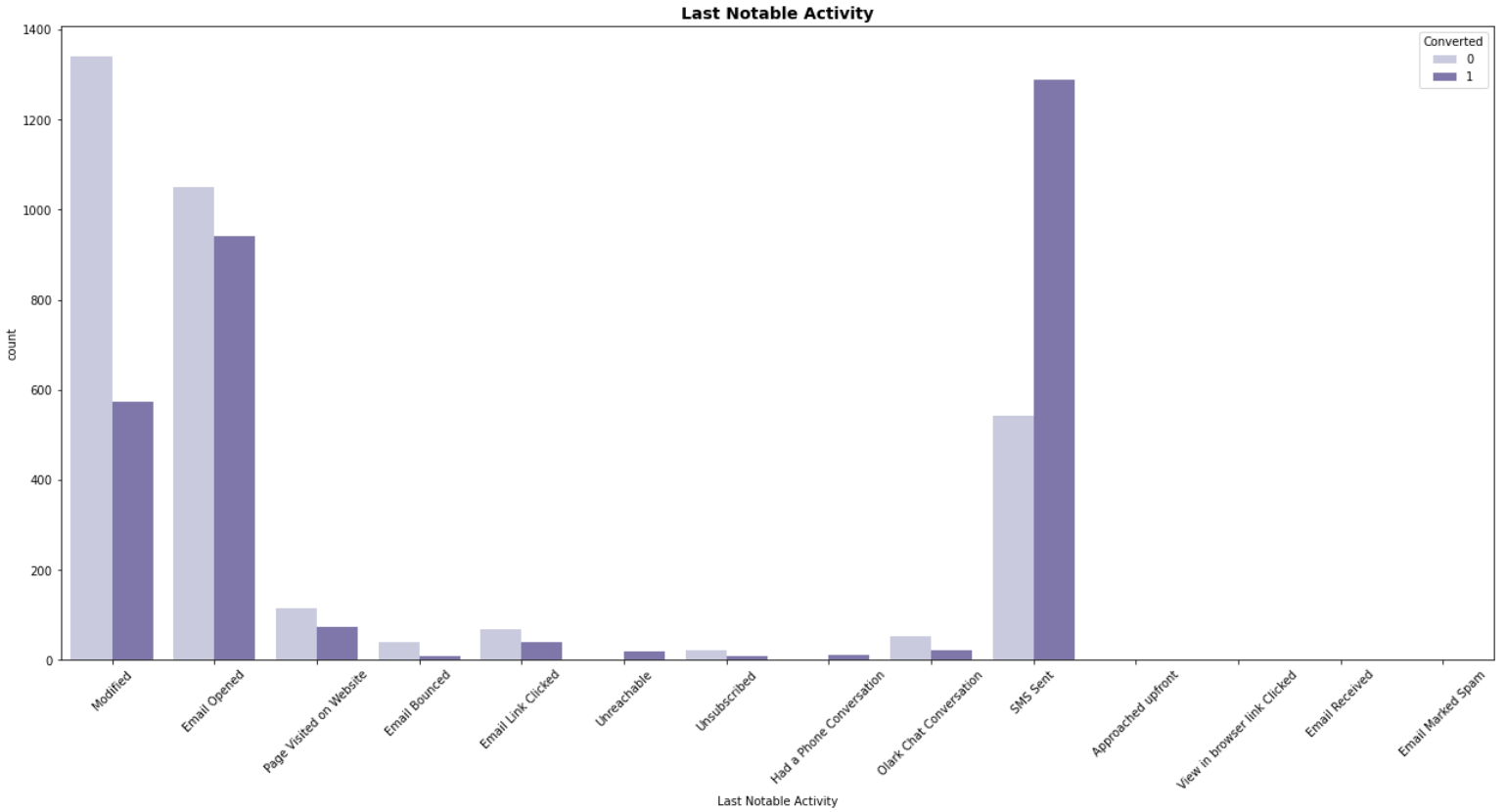
Lead from Specialization who are unknown/Select columns has the highest rate of conversion.



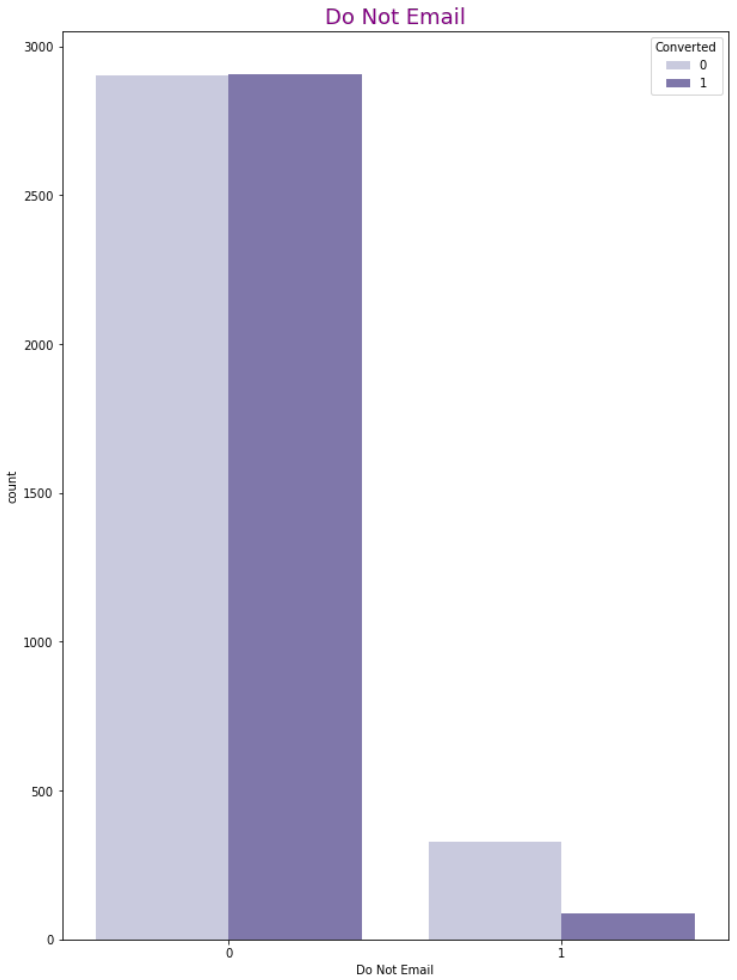
Person who are unemployed has the highest conversion rate comparatively to working professional.

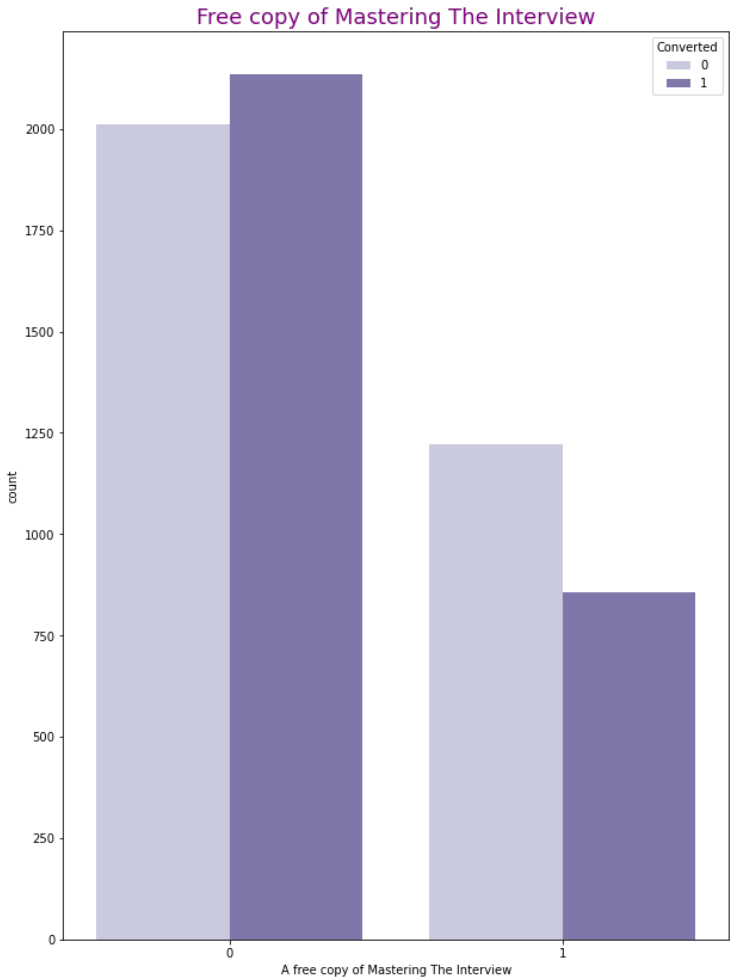


Students whose Last Notable Activity was found to be SMS Sent had the best conversion rate.

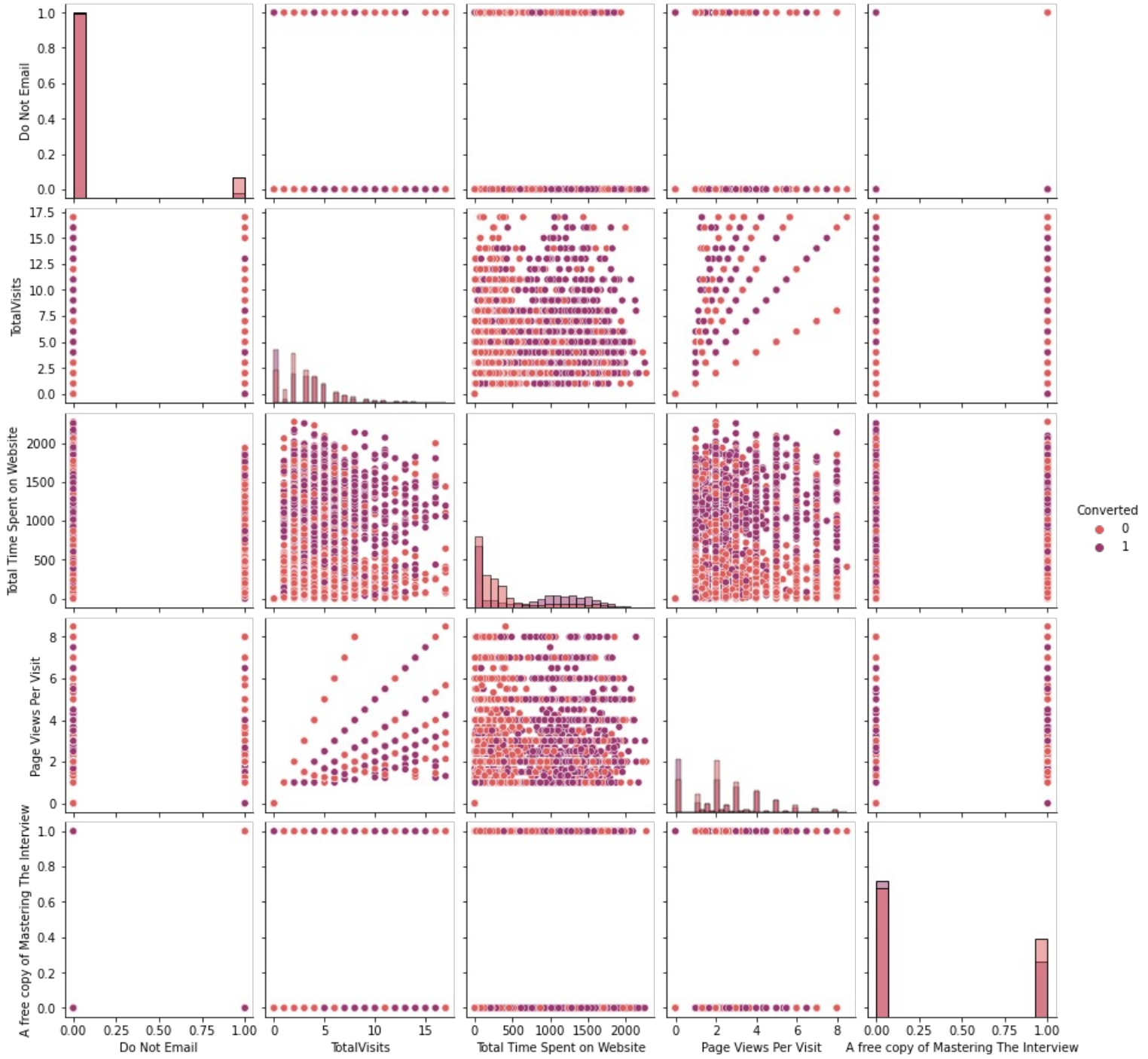


# Bi-variate Analysis



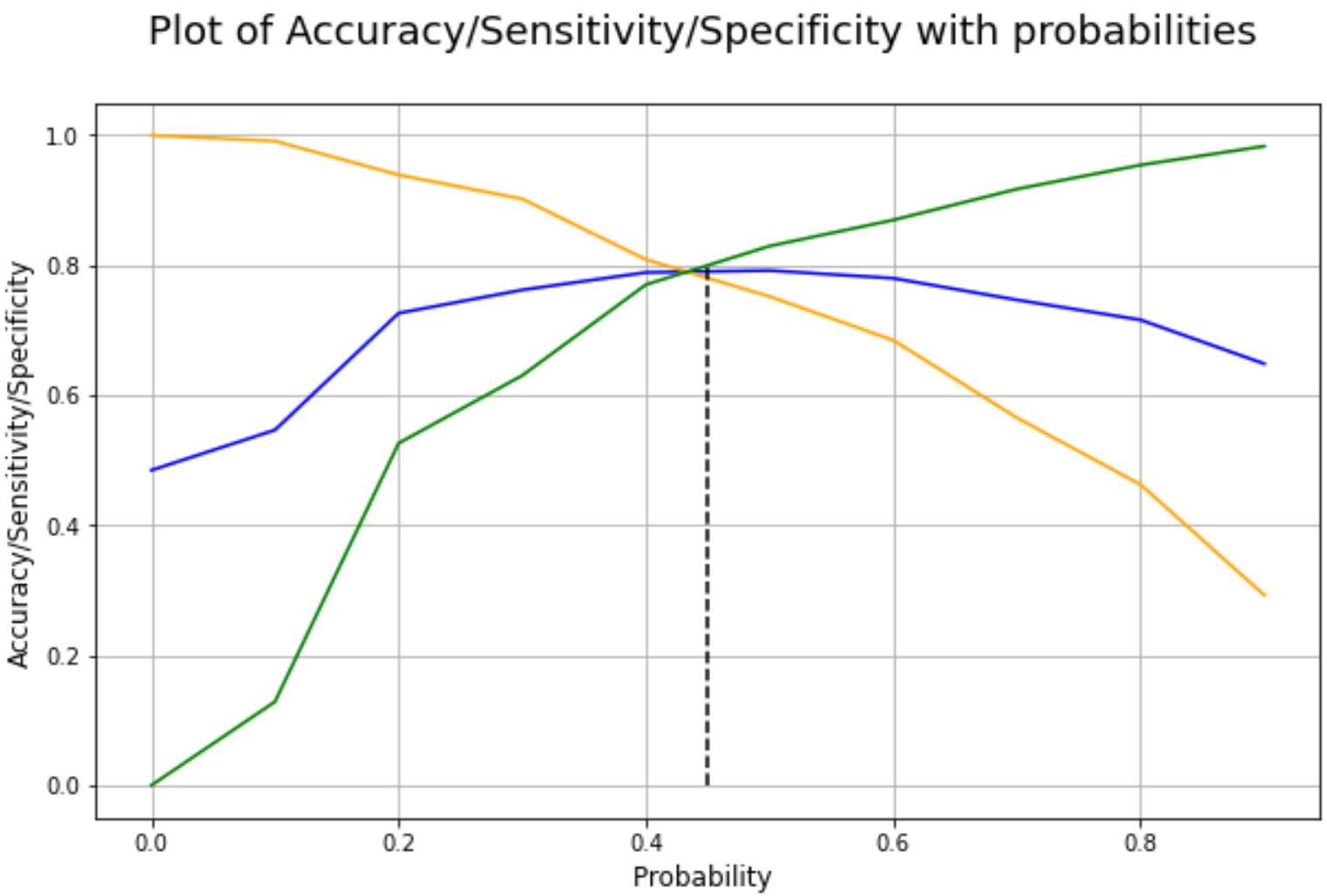



# Pair Plot ( Multivariate Analysis)



# Correlation Heat Map

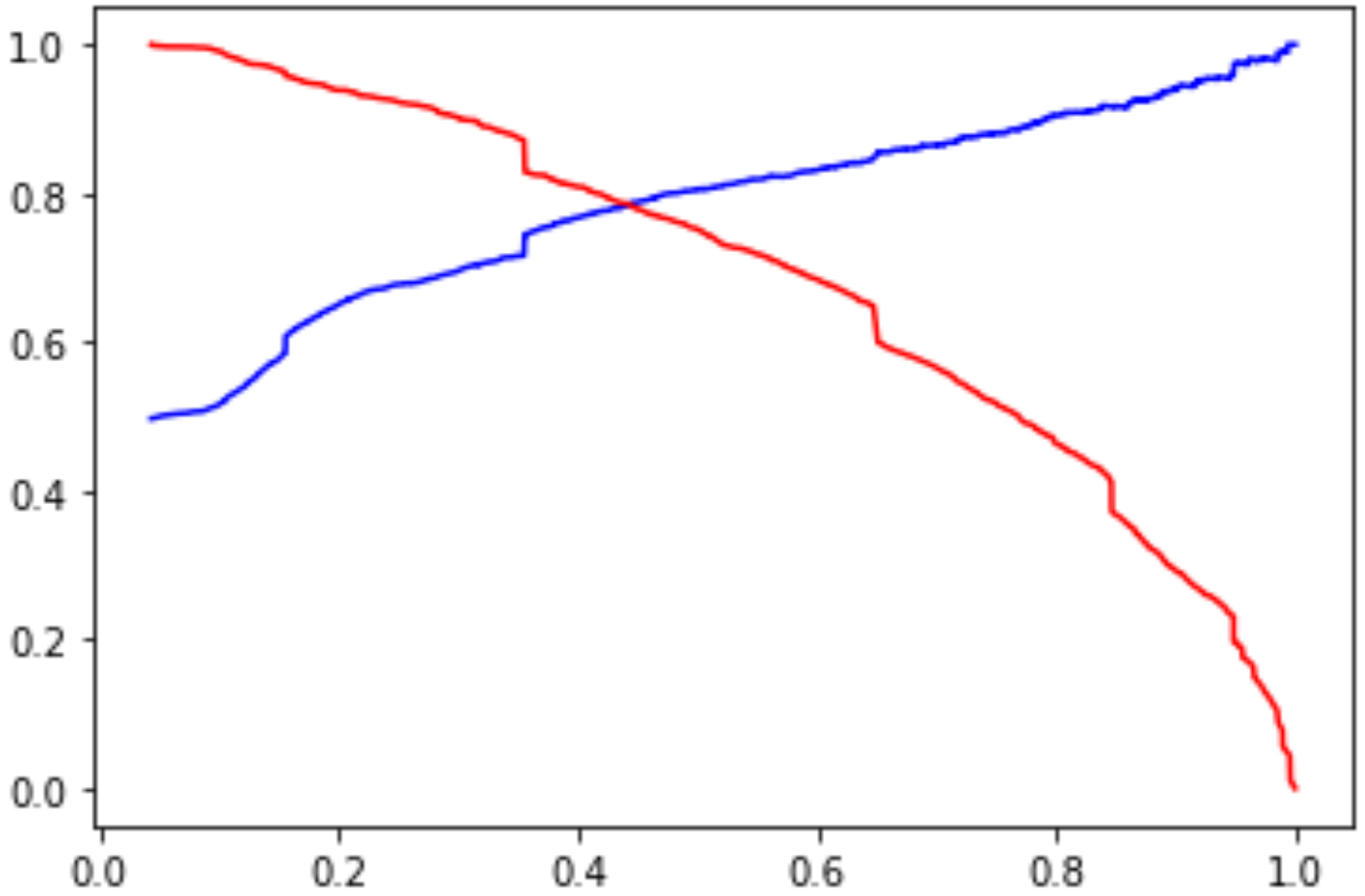
# MODEL EVALUATION TRAIN AND TEST SET



## TRAIN SET:

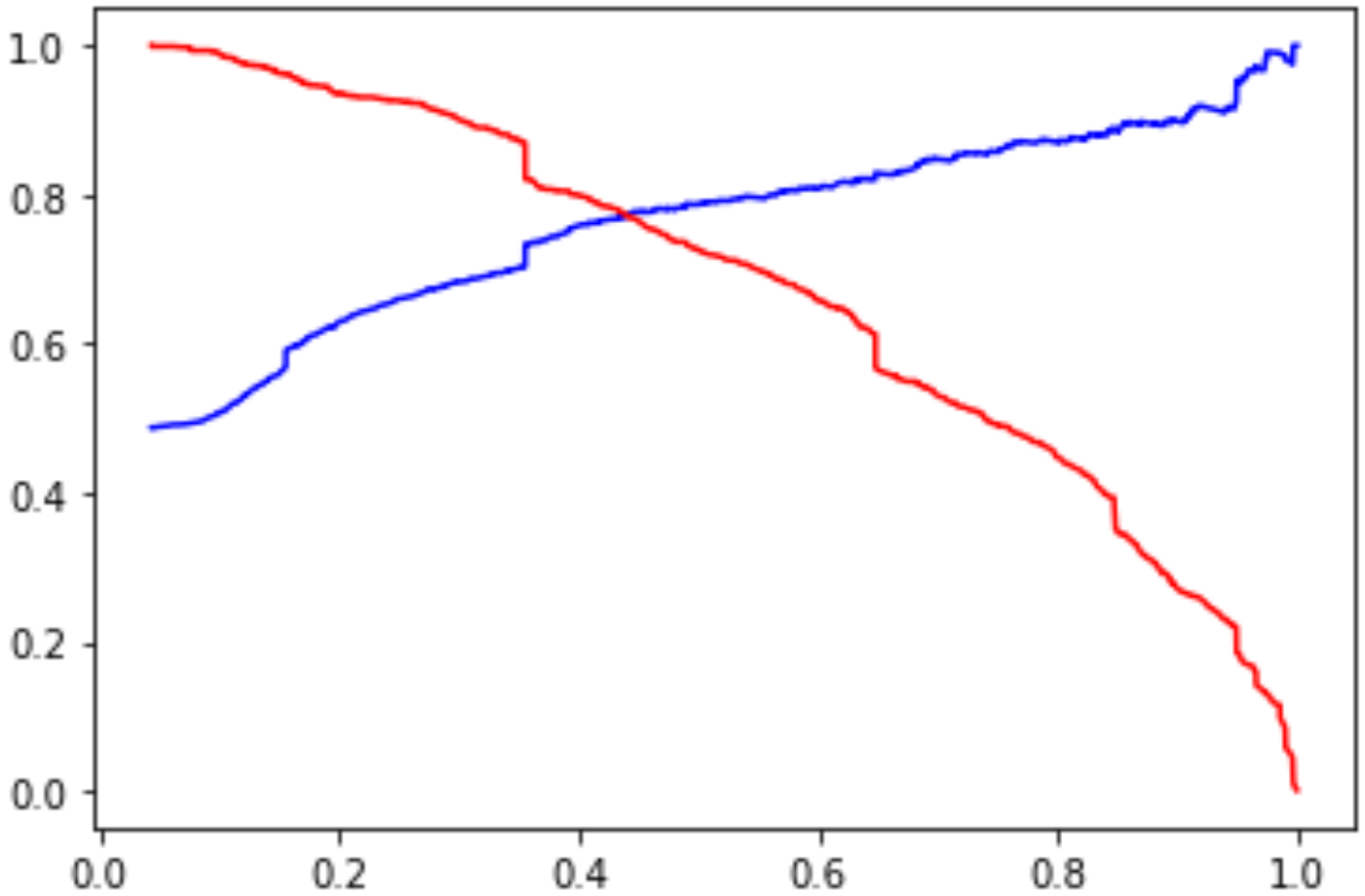
Accuracy = 0.79072
Sensitivity = 0.79071
Specificity = 0.79073
Precision = 78%
Recall =79%

The cut-off value is approximately 0.43



## TEST SET:

Accuracy = 0.785
Sensitivity = 0.799
Specificity = 0.773
Precision = 77%
Recall = 78%



# Conclusion

- The elements that are crucial for prospective Leads include Current employment status is "unemployed”.
- "Total visits," "LeadOrigin Lead Add Form," "Latest Activity as SMS delivered," and "Total time spent on the website" are some examples.

# Recommendations

- Regular data collection, model running, and lead updates are all beneficial. It's generally accepted that the optimal time to call prospective leads is shortly after they express interest in your courses.
- Email is just as effective as cold calling, so it's a good idea to mail the leads in addition to making phone calls to remind them.
- You can save a lot of time by limiting the number of call efforts to two to four and boosting the frequency with which you use other media, such as Google ads or emails, to stay in touch with leads.
- By concentrating on hot leads, which have a low contact rate but high conversion rates, we have a better chance of bringing in more value to the company.