

# Lead Scoring Case Study using Logistic Regression

Submitted by:

- Hemangini Bhatt
- Harshitha Praturi
- Harikrishnan R



# 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. 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%.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

# Goals to achieve:

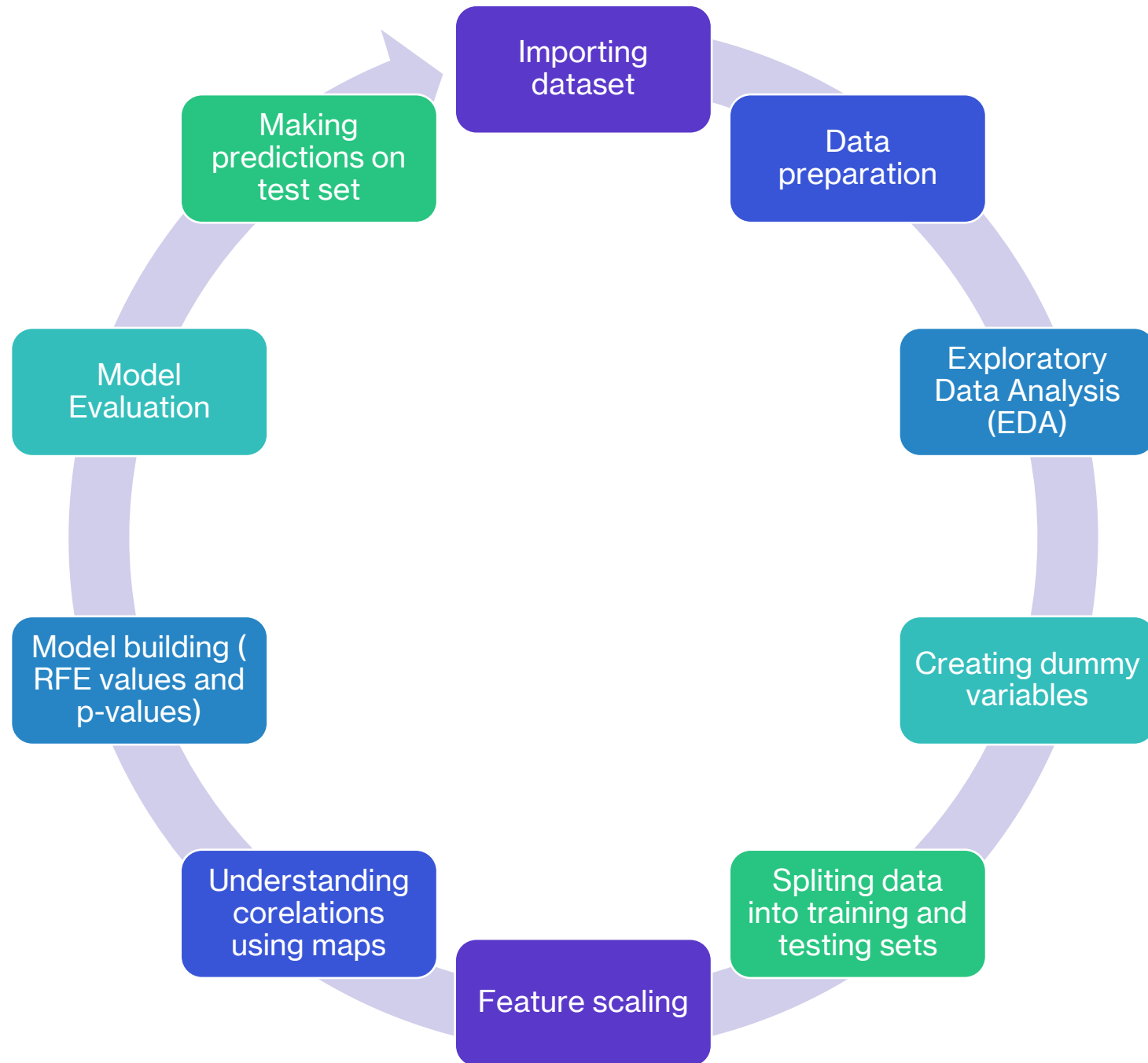


Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

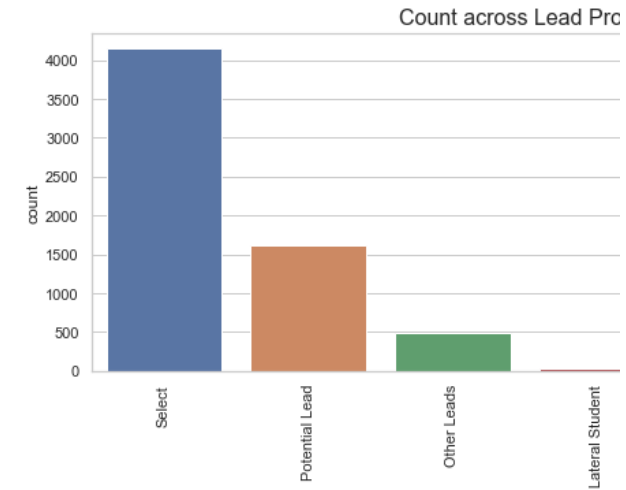
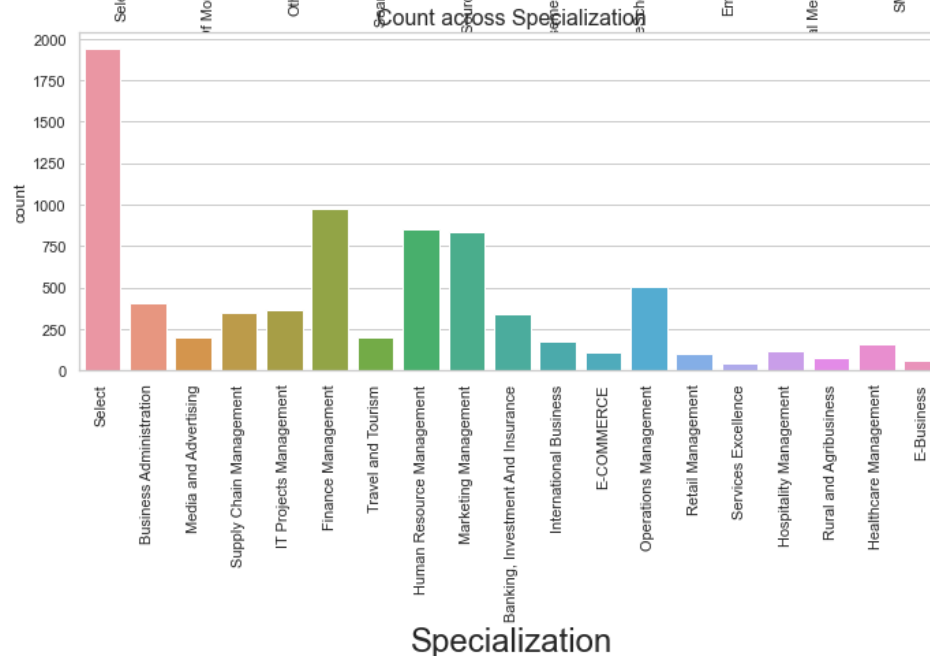
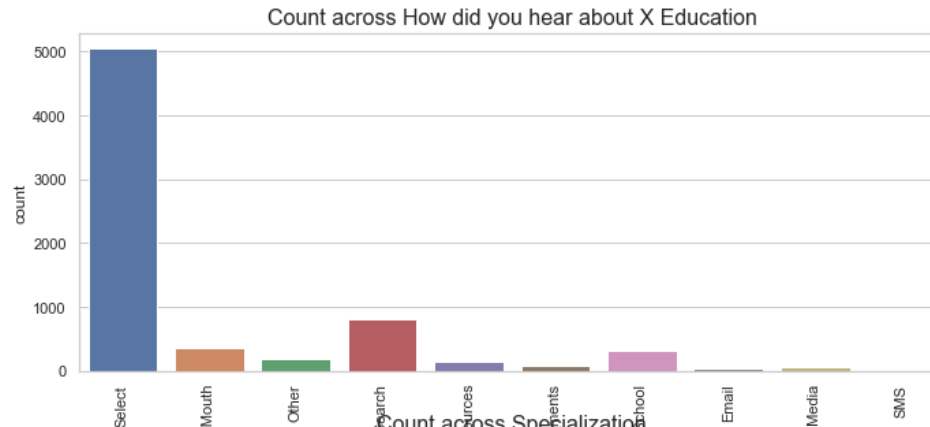


There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

# Approach appointed:

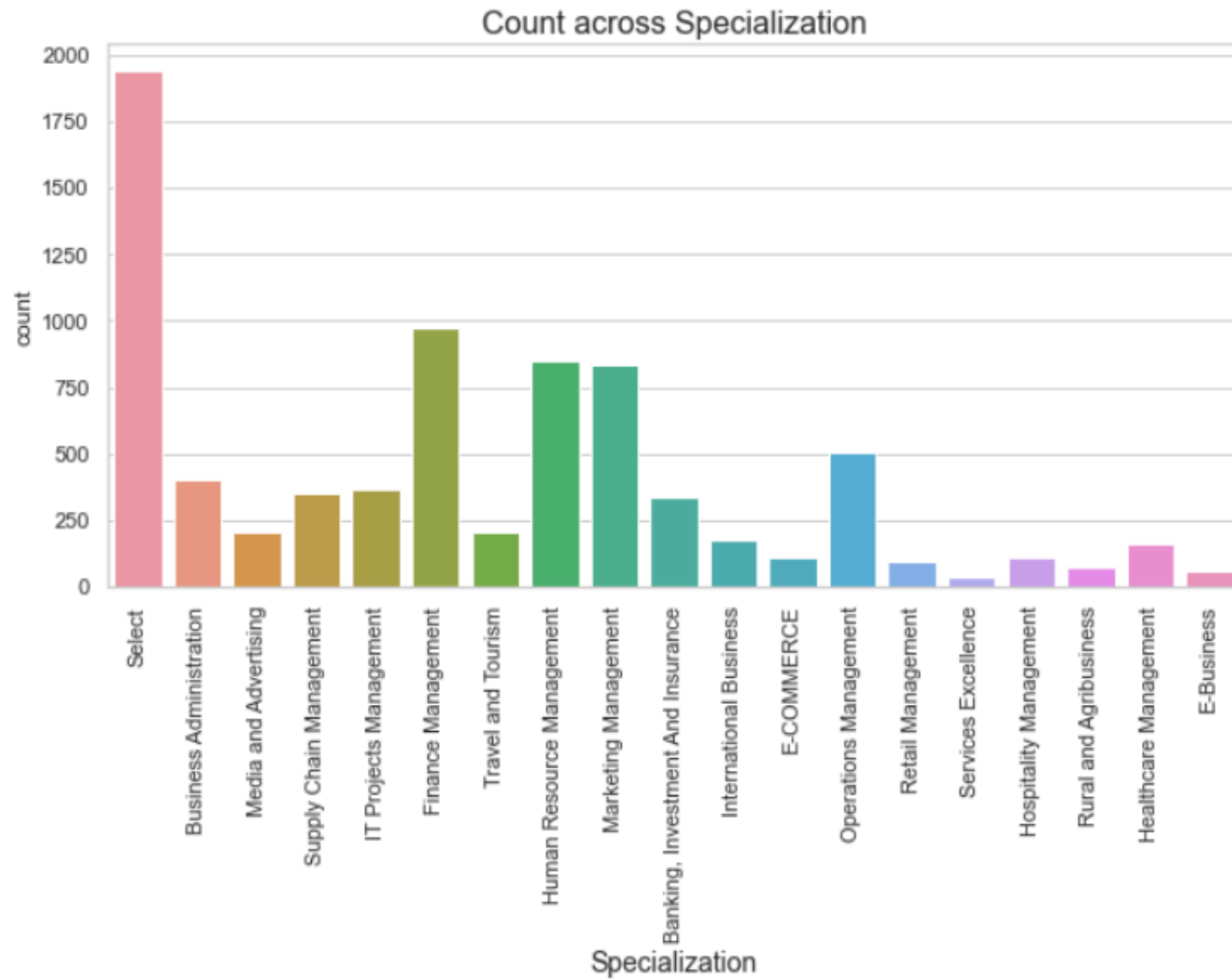


# Data Cleaning



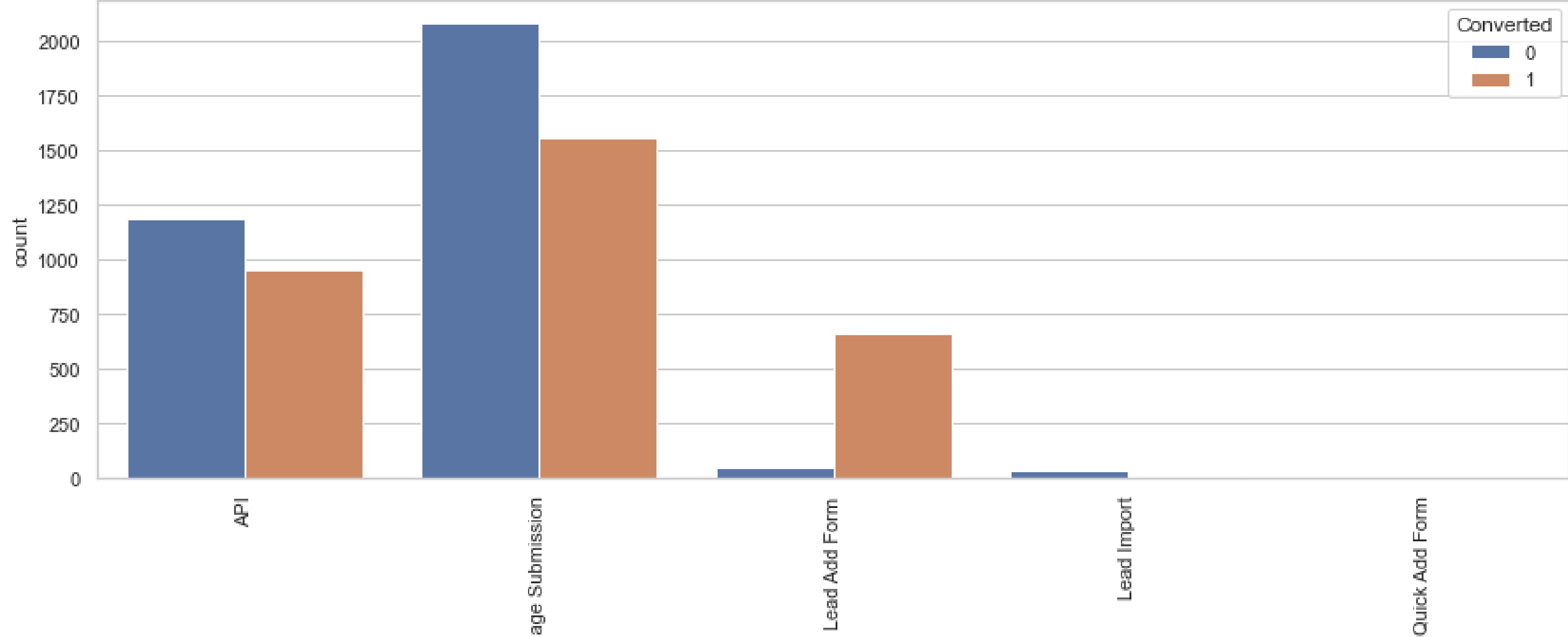
Lead Profile

There are few columns that are yet to be selected or haven't been selected and hence have the word 'Select' in the columns, which is removed as it doesn't provide any value.

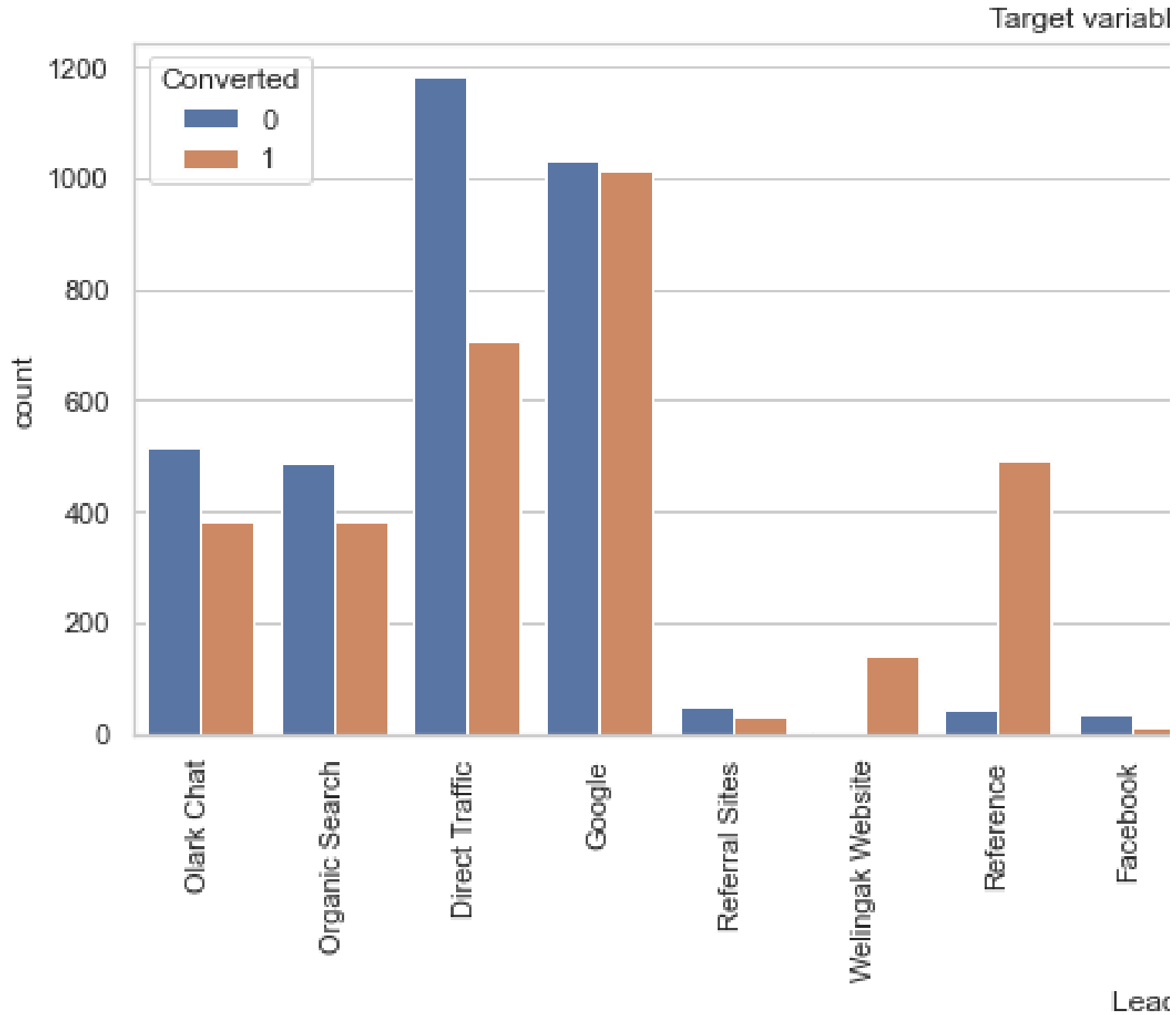


# Conversion rates

Analysis results here show that, most people opt for the course for better career prospects, the majority leads are from Finance Management, Human Resources and Marketing Management followed by Operations Management.

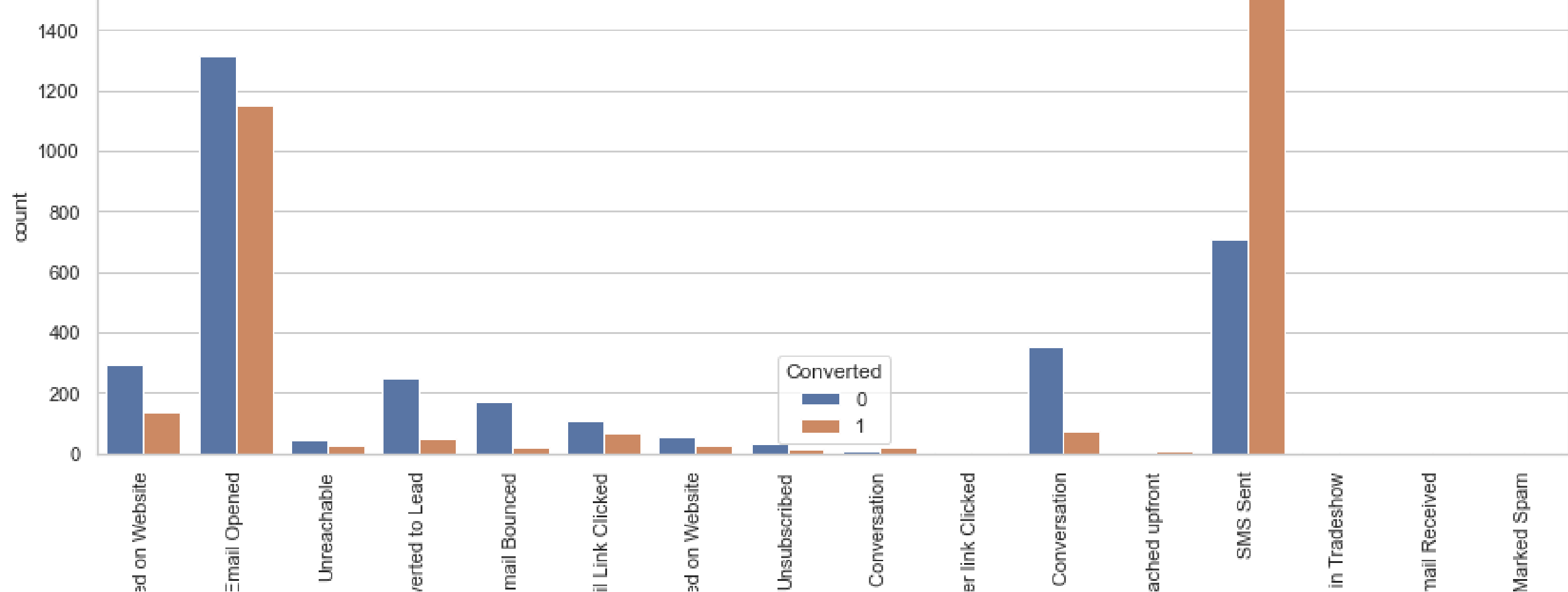


**Target Variables in lead origin**



**Targeted  
Variables  
in lead source**



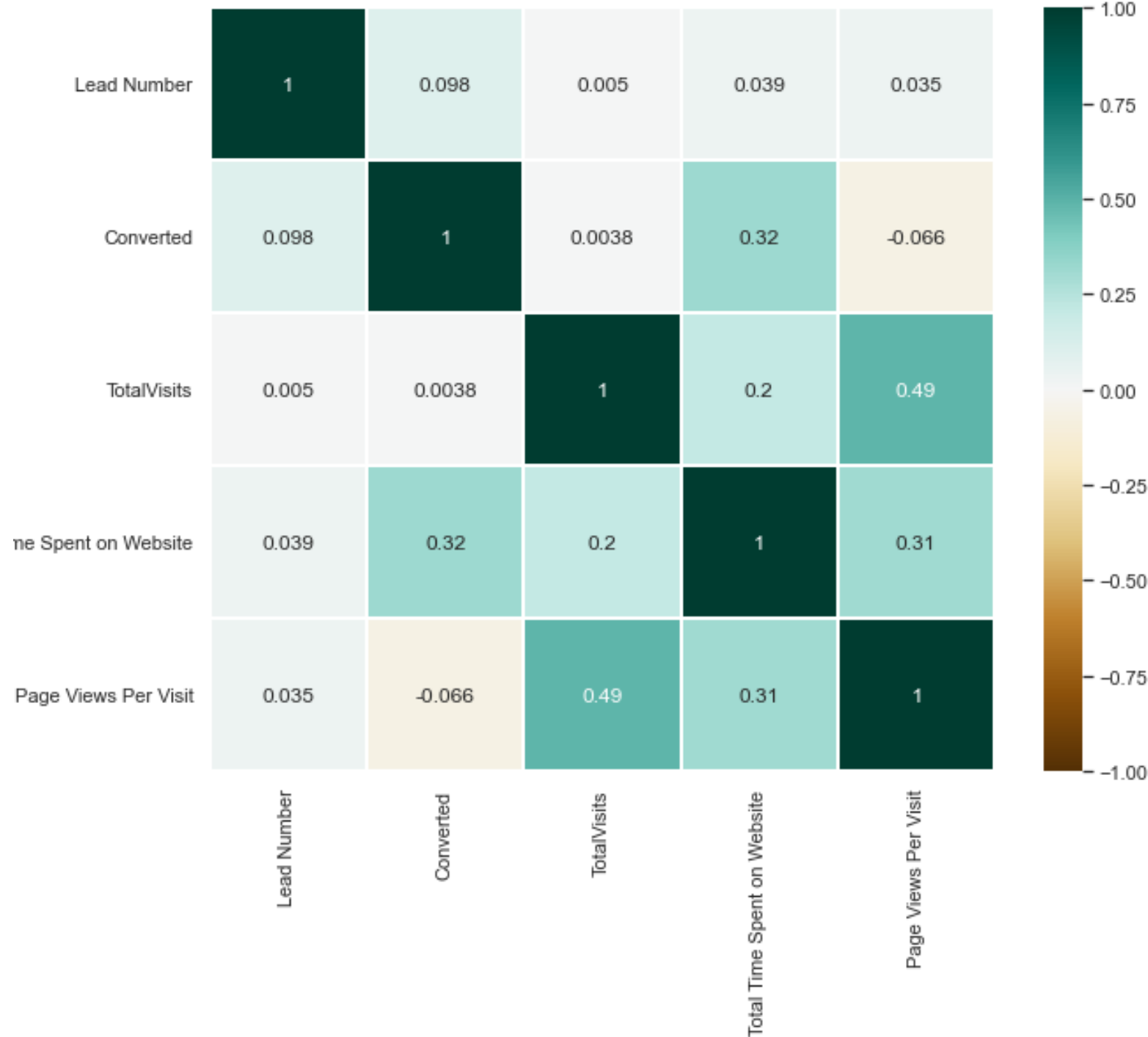


## Conversion rates based on last activity

- It has been brought forth that conversion rates are higher when the customer has been approached through mail and sms. Both SMS and e-mail have proven to be effective for leads.

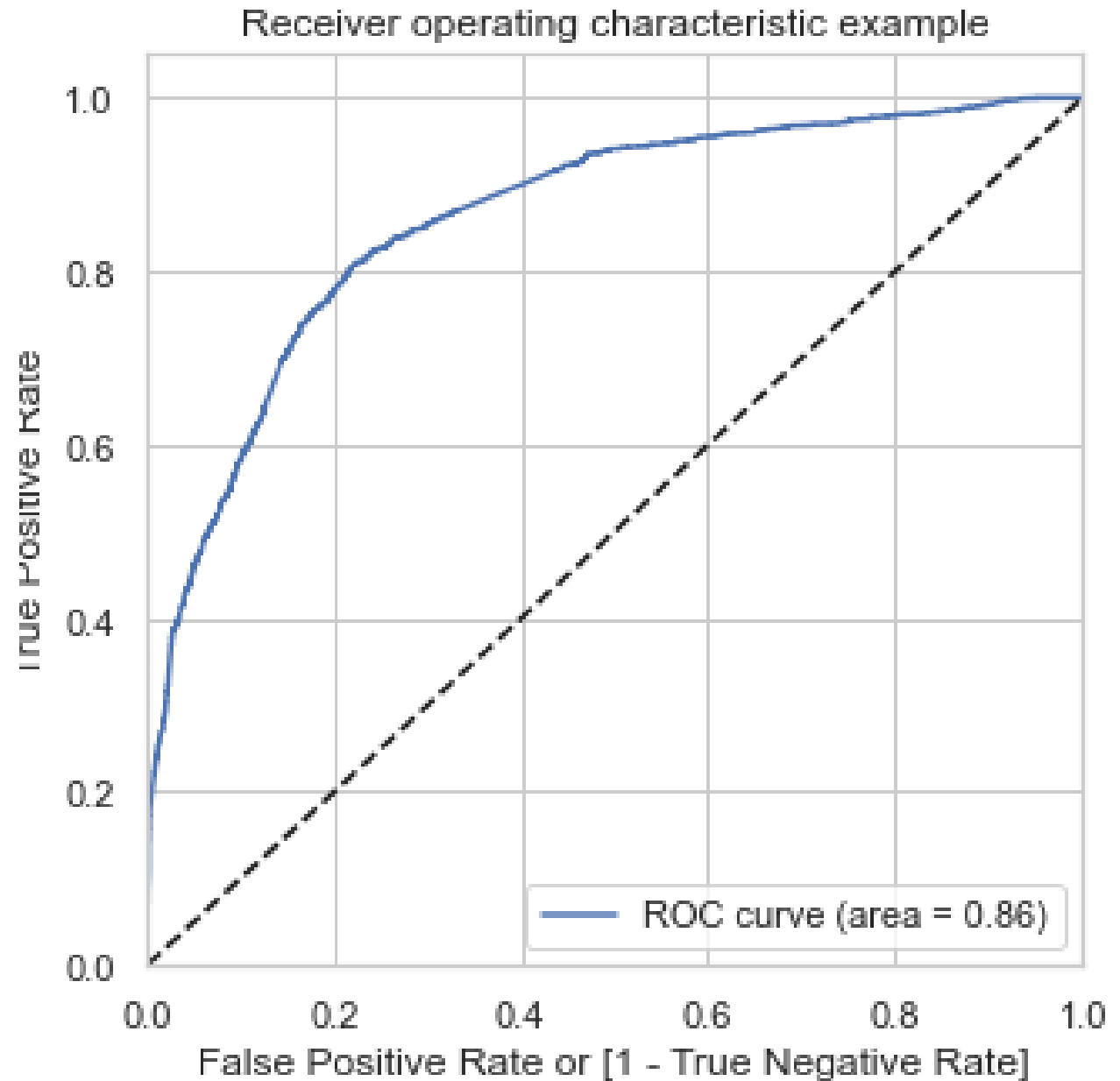
# Corelation

- Total time spent on website has the highest correlation with converted variable.
- Total time spent hence is an influential factor



# Model Evaluation ROC curve

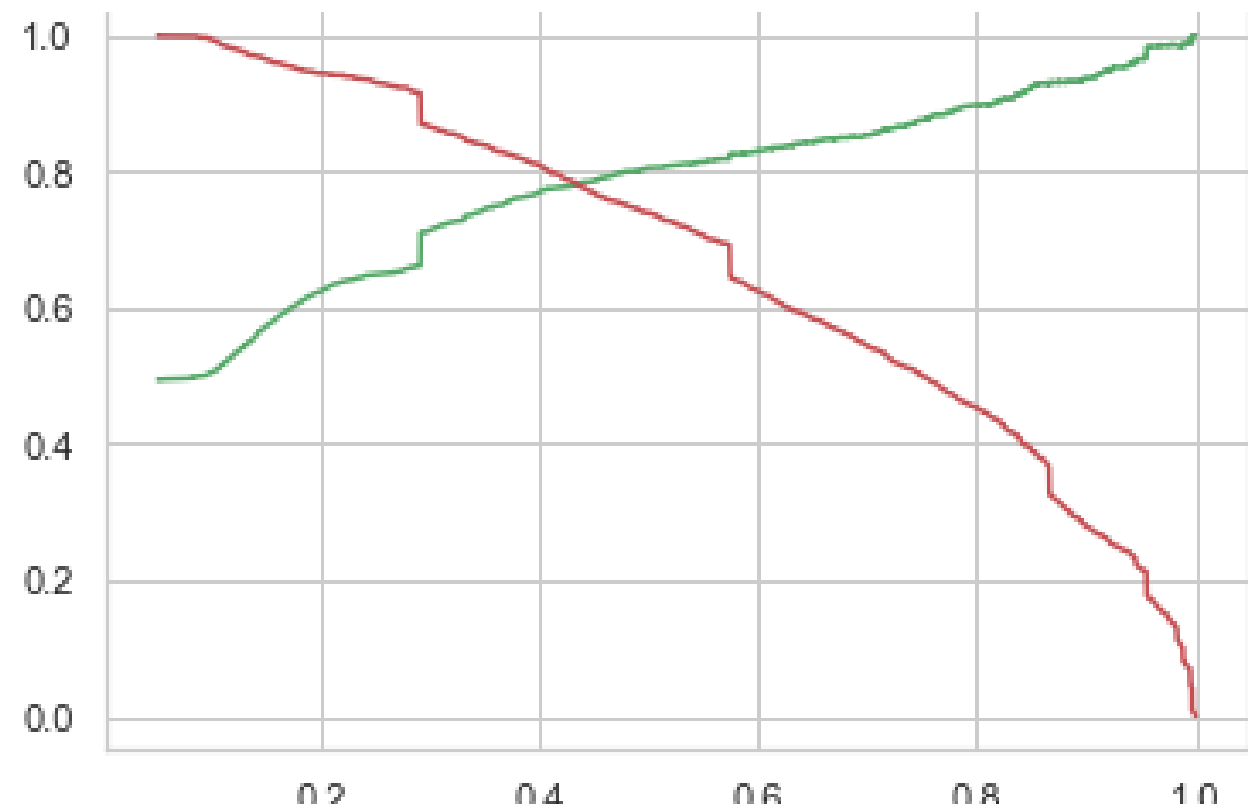
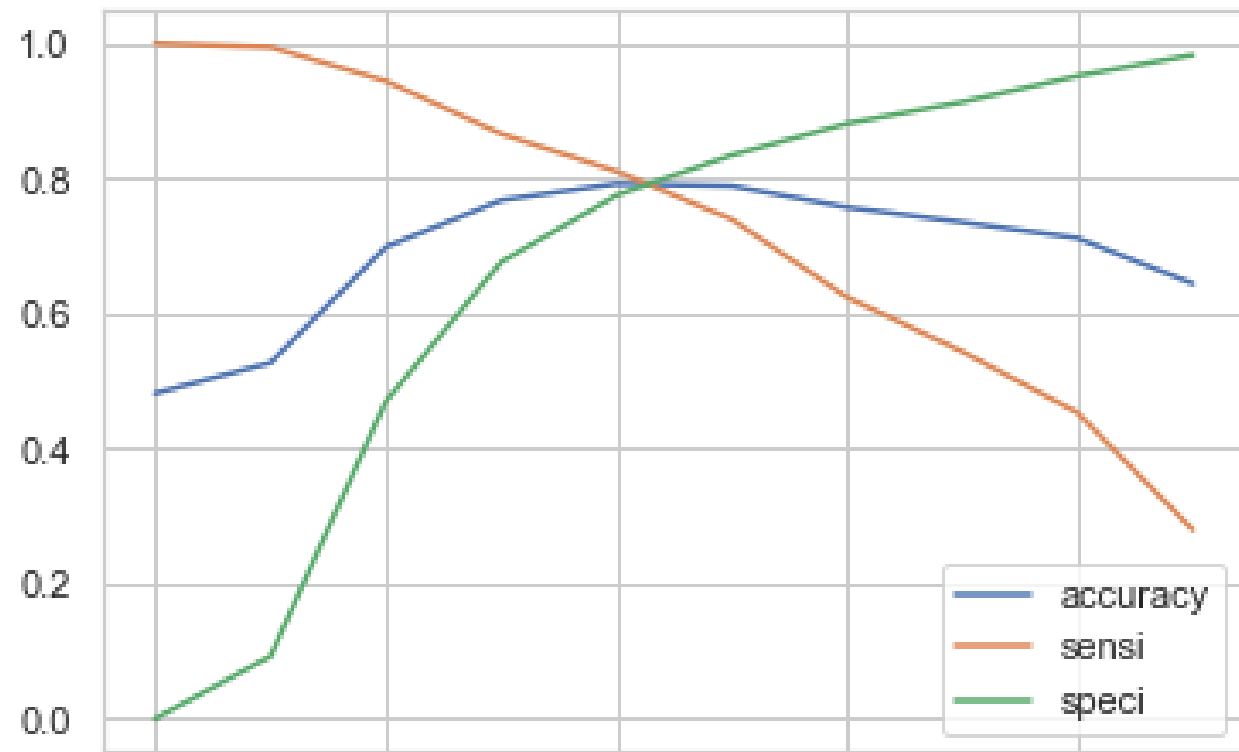
Area under the curve is 0.86  
which is good outcome.



# Model Evaluation

## Precision and Recall

- 0.42 is the trade-off between precision and recall.
- Hence it is safe to say that any lead with conversion probability over 42% is a hot lead.



# Model Observations

- Training Data accuracy : 78.4%
- Training Data Specificity: 78.9%
- Training Data Sensitivity: 77.9% ~ 78%
- Testing Data Accuracy : 78.6%
- Testing Data Specificity : 78.9%
- Testing Data Sensitivity : 77.9%

# Conclusion

Leads who have spend most amount of time on the website are most likely to convert.

Most of the people who have opted for the course is because they wanted better career prospects.

It is observed that most number of leads are generated by google/ direct traffic.

Most common last activity opened is e-mail, hence mails have good conversion rates.

Another method with good conversion rate is SMS.

Most of the people opting for the course are unemployed or are working in Human resources or finance management.