# Lead Score Casestudy

GROUP MEMBERS :

1. ANIKET VERMA
2. AKSHITA GOEL
3. TANAY DEY

# Problem Statement

X Education offers online courses to professionals in various industries. However, the company faces challenges with its lead conversion rate. For instance, out of 100 acquired leads in a day, only around 30 are successfully converted. To improve efficiency, the company aims to pinpoint the most promising leads, referred to as 'Hot Leads'. By successfully identifying this subset of leads, the lead conversion rate is expected to increase. This shift will enable the sales team to concentrate their efforts on engaging with the potential leads instead of reaching out to every individual.

# Business Objective

X Education is seeking to identify the most prospective leads by developing a model specifically designed to recognize 'hot leads'. The objective is to deploy this model for future use, ensuring its availability and effectiveness in identifying promising leads consistently.

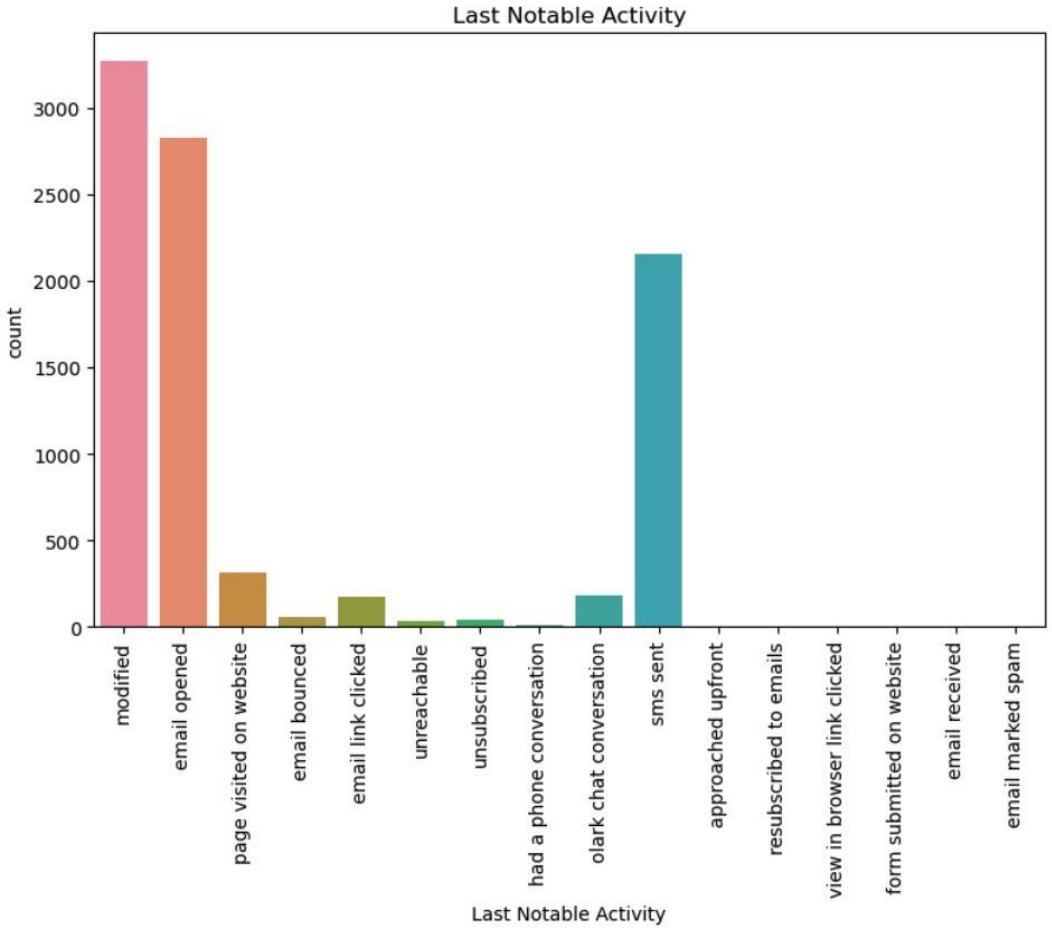
# Solution Methodology

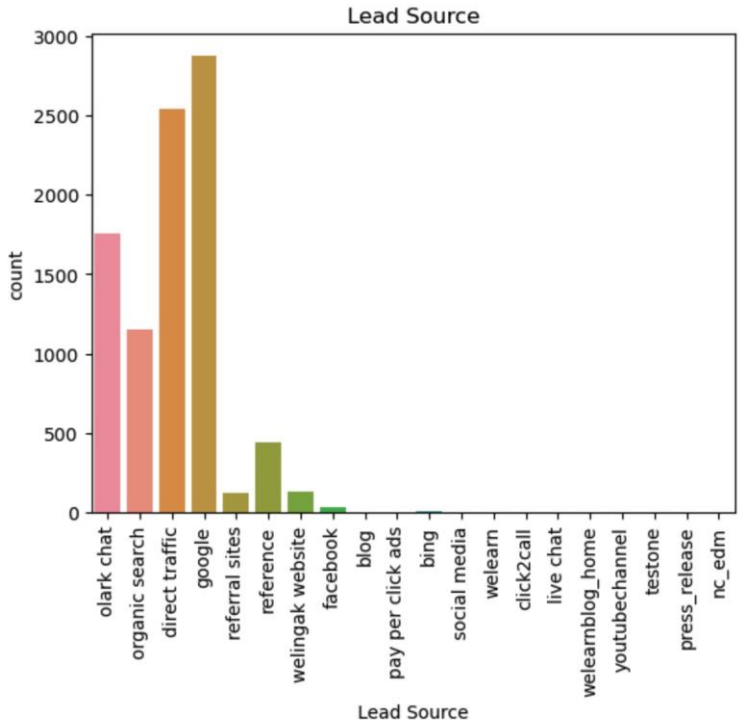
- Data cleaning and data manipulation:
    1. Handling and resolving duplicate data.
    2. Managing NA values and missing values.
    3. Dropping columns with a large number of missing values and no relevance for analysis.
    4. Imputing values as needed.
    5. Addressing outliers in the data.
- Exploratory Data Analysis (EDA):
    - Univariate data analysis:
        - Analyzing value count and variable distribution.
    - Bivariate data analysis:
        - Examining correlation coefficients and patterns between variables.
- Feature Scaling & Dummy Variables and encoding of the data.
- Classification technique:
    - Utilizing logistic regression for model creation and prediction.
- Validation of the model.
- Model presentation.
- Conclusions and recommendations.

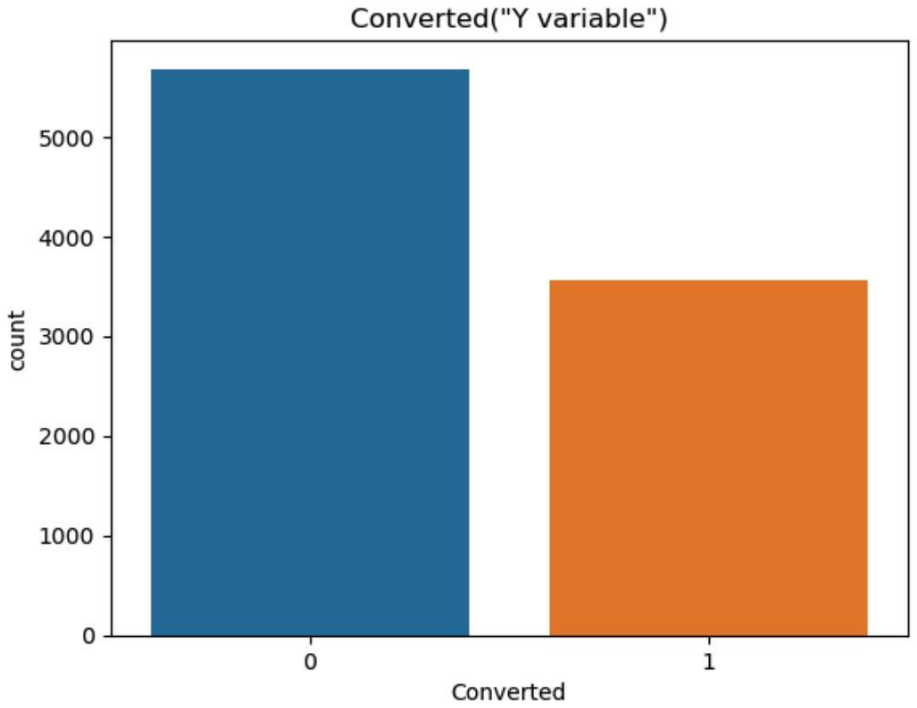
# Data Manipulation

- The dataset consists of 37 rows and 9240 columns.
- Features such as "Magazine," "Receive More Updates About Our Courses," "Update me on Supply Chain Content," "Get updates on DM Content," and "I agree to pay the amount through cheque" have been excluded as they are single-value features.
- The unnecessary columns "Prospect ID" and "Lead Number" have been removed for analysis purposes.
- After examining the value counts for certain object-type variables, we have identified features with insufficient variance, and these have been dropped. Examples of such features include "Do Not Call," "What matters most to you in choosing course," "Search," "Newspaper Article," "X Education Forums," "Newspaper," and "Digital Advertisement."
- Columns with more than 35% missing values, such as 'How did you hear about X Education' and 'Lead Profile,' have also been dropped.

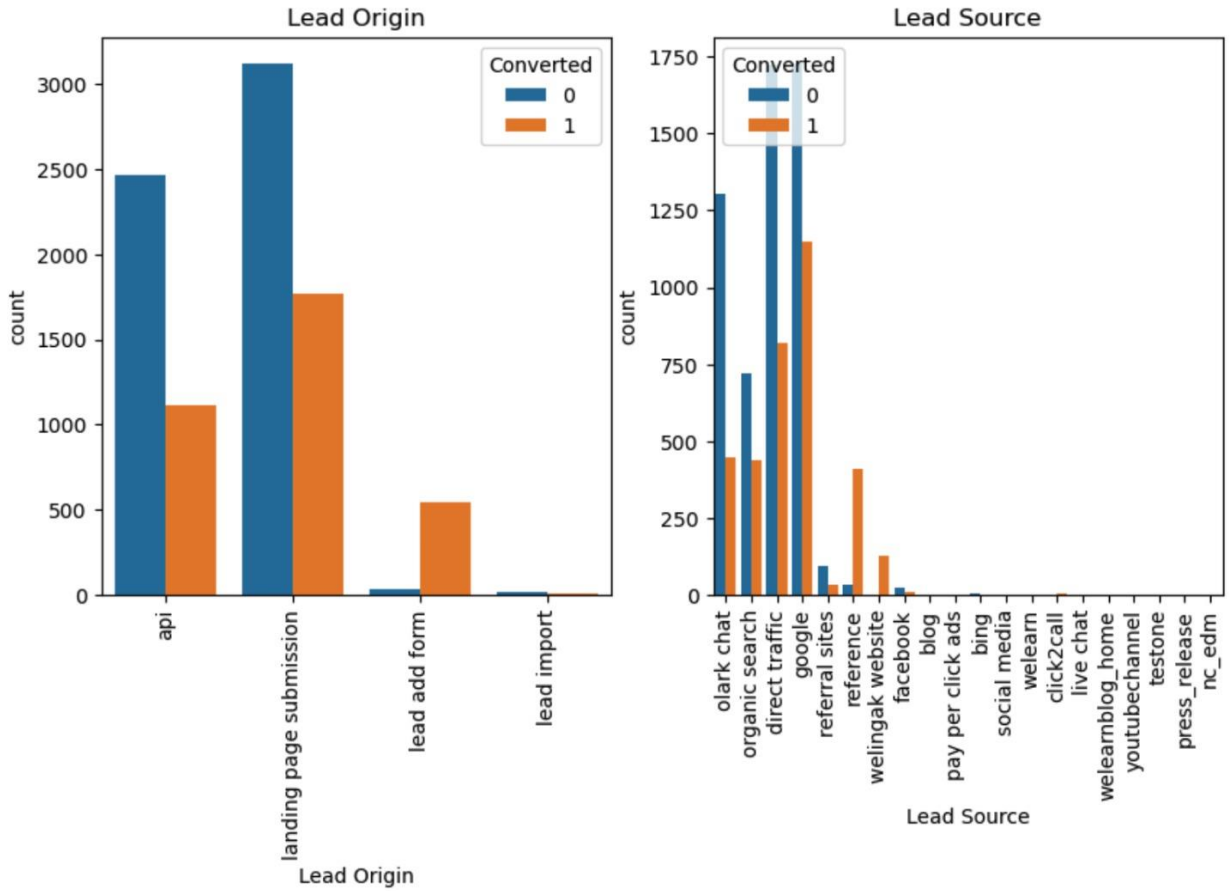
# EDA

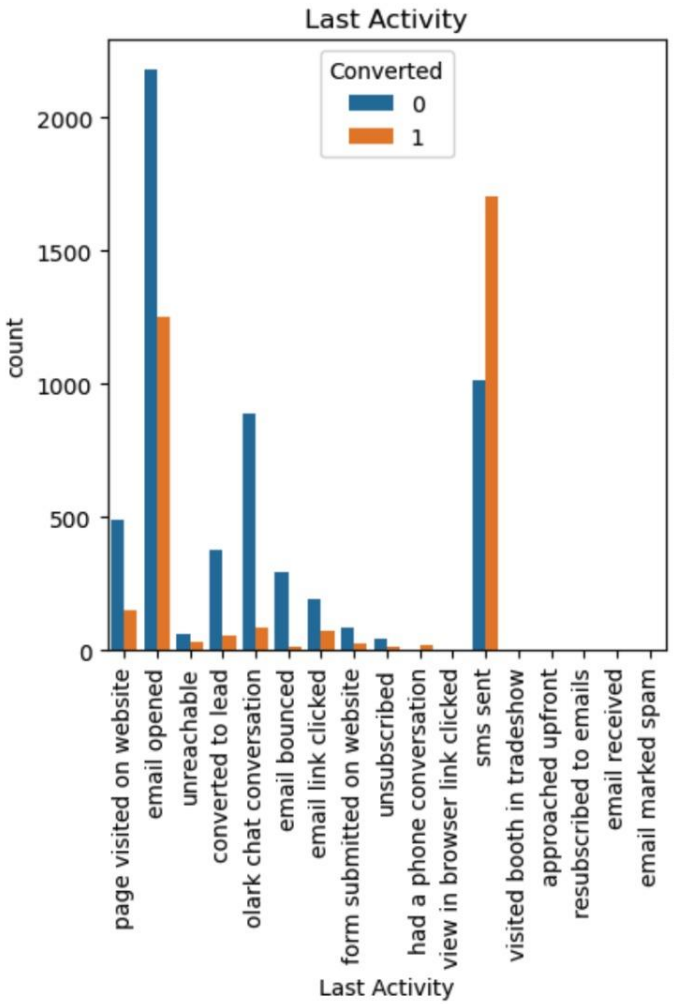

Last Notable Activity
count
3000
2500
2000
1500
1000
500
0
modified
email opened
page visited on website
email bounced
email link clicked
unreachable
unsubscribed
had a phone conversation
olark chat conversation
sms sent
approached upfront
resubscribed to emails
view in browser link clicked
form submitted on website
email received
email marked spam
Last Notable Activity

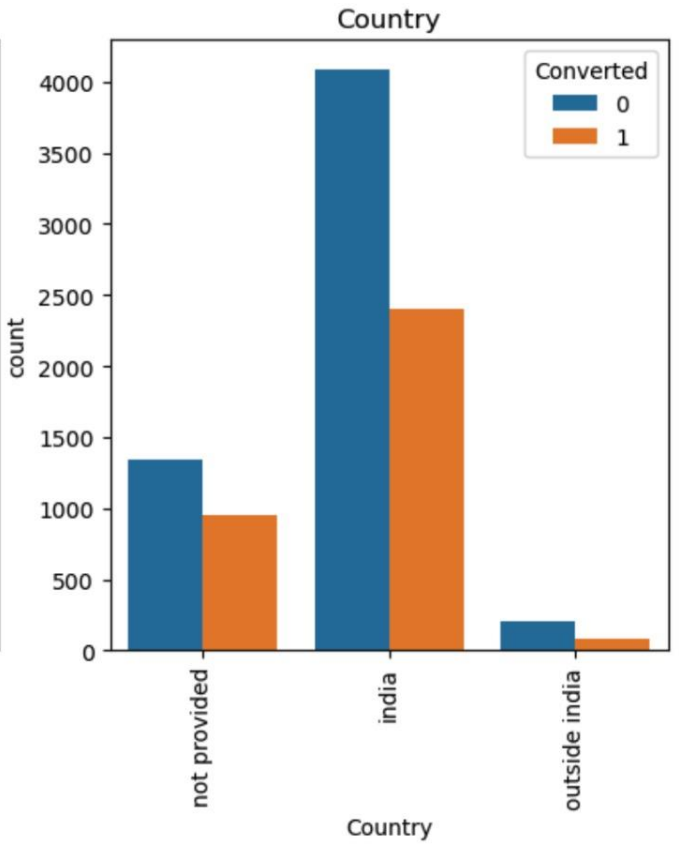

Lead Source
count
3000
2500
2000
1500
1000
500
0
olark chat
organic search
direct traffic
google
referral sites
reference
welingak website
facebook
blog
pay per click ads
bing
social media
welearn
click2call
live chat
welearnblog_home
youtubechannel
testone
press_release
nc_edm
Lead Source

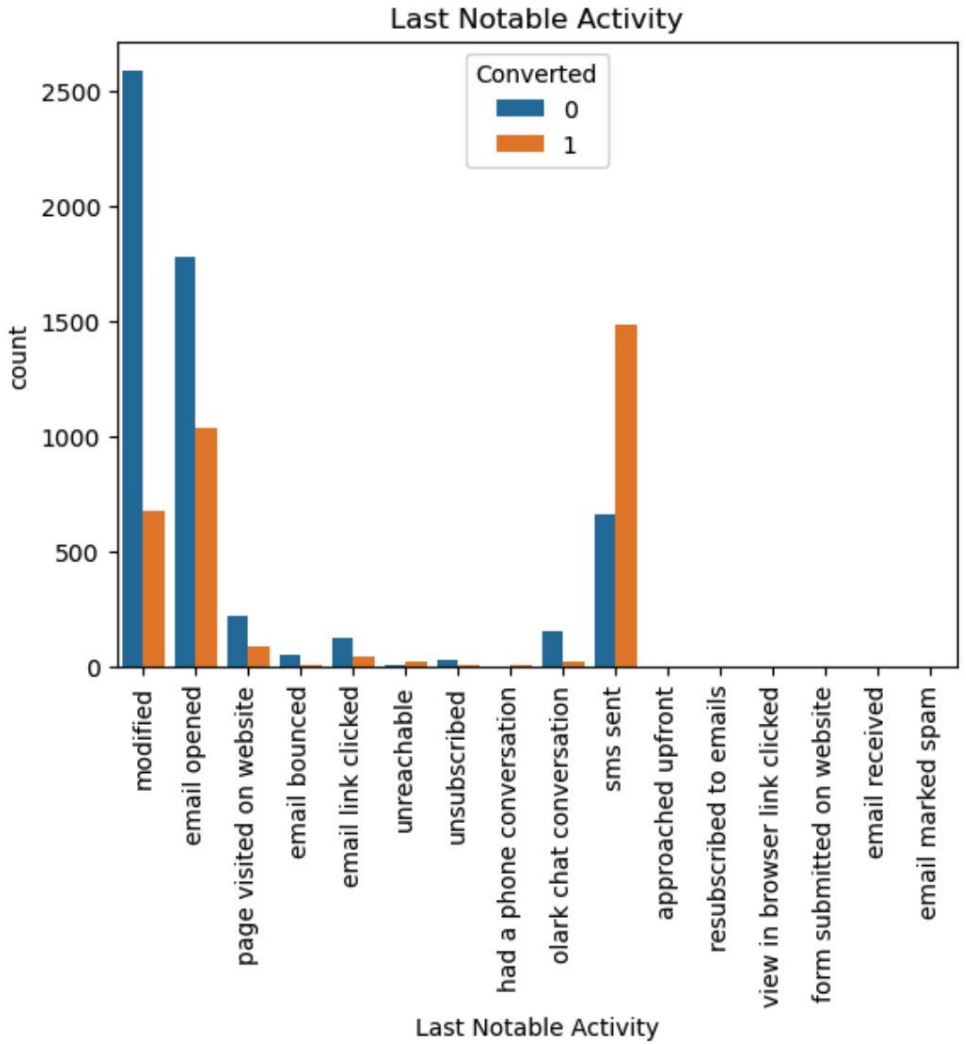

Converted("Y variable")
count
5000
4000
3000
2000
1000
0
0
1
Converted

# Categorical Variable Relation




Last Activity
Converted
0
1
count
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow
approached upfront
resubscribed to emails
email received
email marked spam
Last Activity


Country
Converted
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
not provided
india
outside india
Country

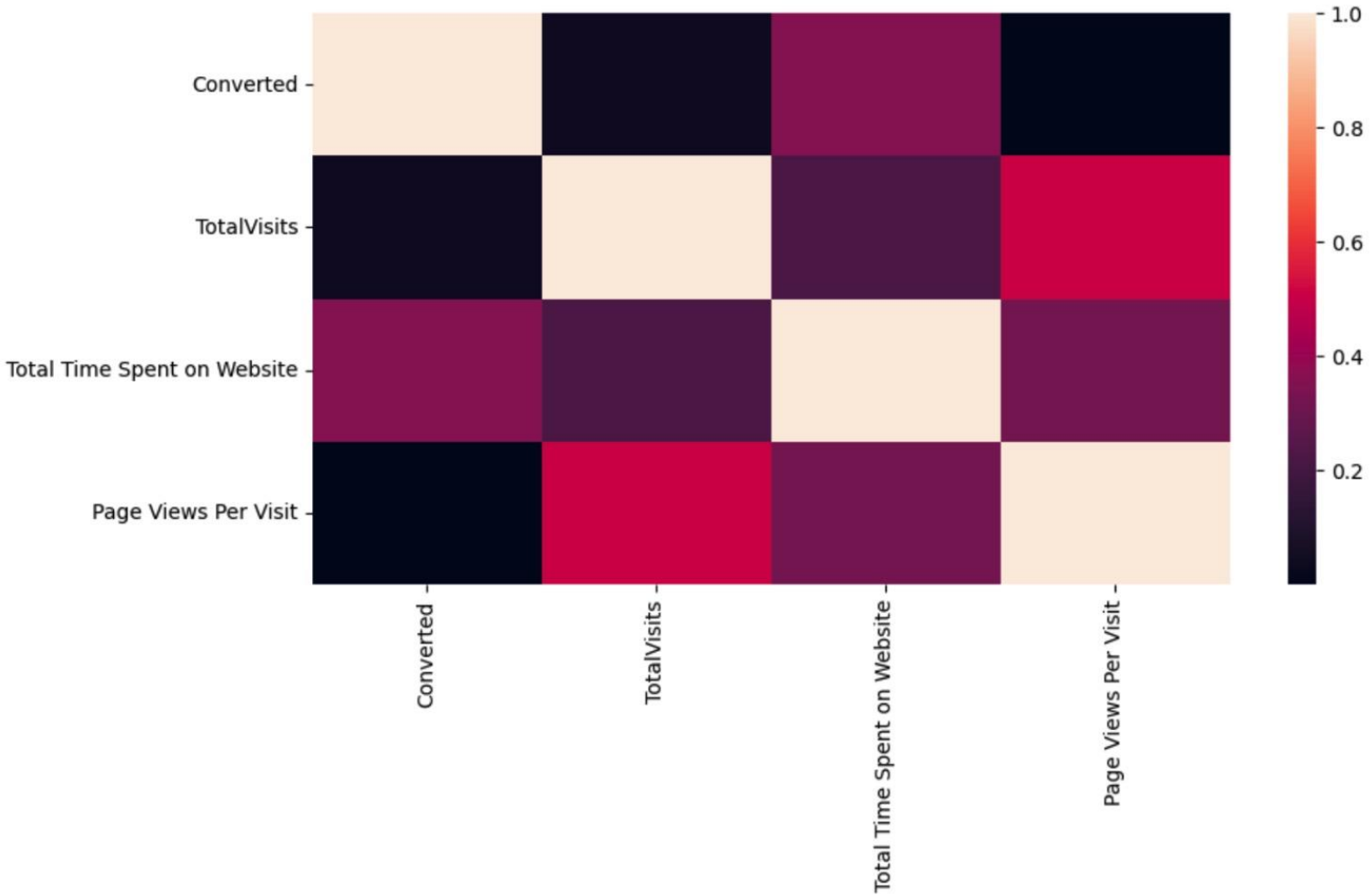

Last Notable Activity
Converted
0
1
count
2500
2000
1500
1000
500
0
modified
email opened
page visited on website
email bounced
email link clicked
unreachable
unsubscribed
had a phone conversation
olark chat conversation
sms sent
approached upfront
resubscribed to emails
view in browser link clicked
form submitted on website
email received
email marked spam
Last Notable Activity

# Data Conversion

- The numerical variables have been normalized.
- Dummy variables have been generated for object-type variables.
- The dataset for analysis consists of 8792 rows.
- There are 43 columns included in the analysis.

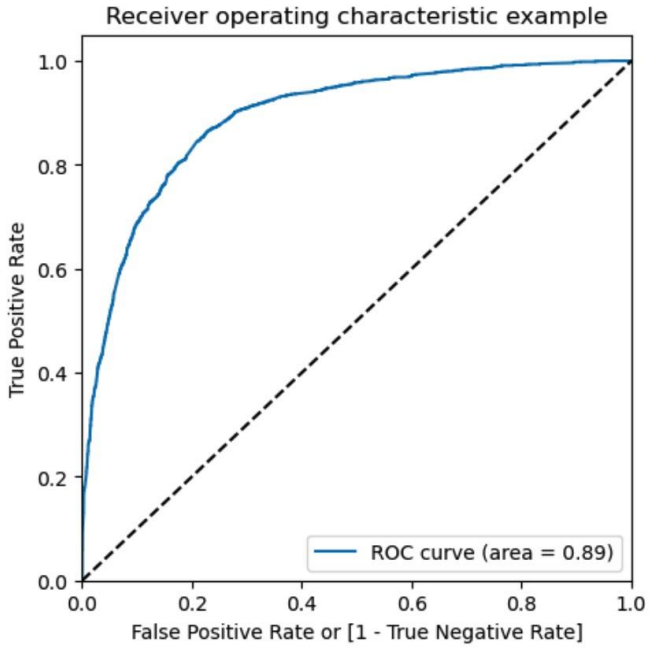
# Model Building

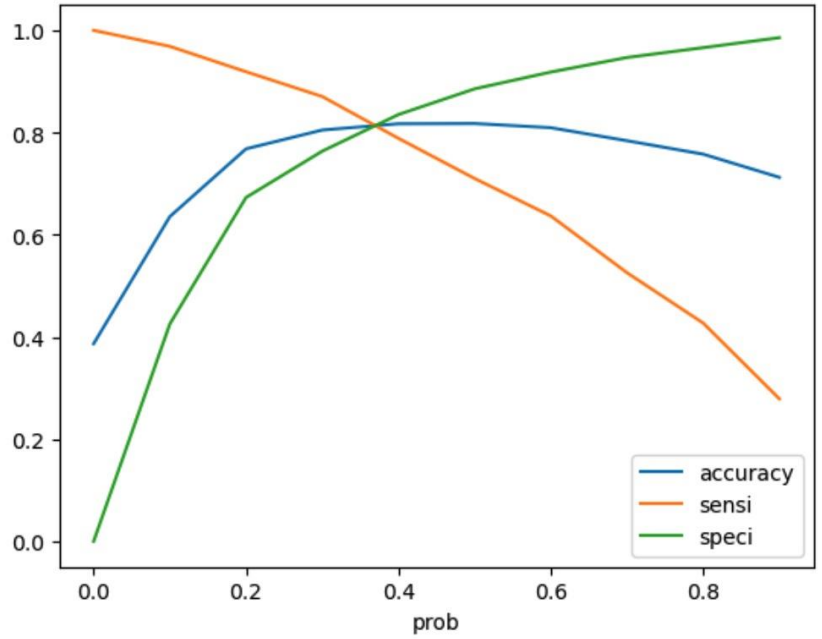
- The data has been divided into training and testing sets using a 70:30 ratio, which is a common practice in regression analysis.
- Recursive Feature Elimination (RFE) has been employed for feature selection. This technique selects the top 15 variables as the output.
- A model has been built by removing variables with a p-value greater than 0.05 and a VIF (Variance Inflation Factor) value greater than 5.
- Predictions have been made on the test dataset.
- The overall accuracy of the predictions on the test dataset is 81%.


Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
1.0
0.8
0.6
0.4
0.2
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit

# ROC Curve

- The optimal cut-off point refers to the probability threshold where there is a balance between sensitivity and specificity.
- By analyzing the second graph, it is evident that the optimal cut-off point occurs at 0.35.



# Conclusion

The variables that have been identified as the most significant factors in influencing potential buyers, in descending order, are:

1. Total time spent on the website.
2. Total number of visits.
3. Lead source:
    3. Google
    4. Direct traffic
    5. Organic search
    6. Welingak website
4. Last activity:
    4. SMS
    5. Olark chat conversation
5. Lead origin as Lead add format.
6. Current occupation as a working professional.

Taking these factors into consideration, X Education has a high likelihood of success in persuading almost all potential buyers to reconsider and make a purchase of their courses.