LEADS SCORING CASE STUDY

SUBMITTED BY :

1. Anadi Mitra
2. Anju Gautam
3. Anmol Bhandari

PROBLEM STATEMENT:

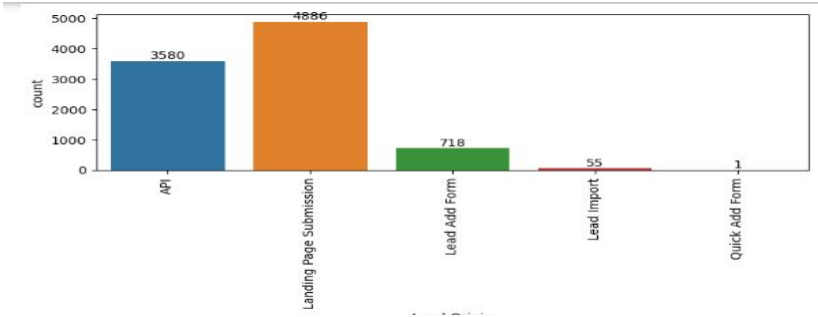
- X Education sells online courses to industry professionals.
- X Education gets a lot of leads on a daily basis, its lead conversion rate is very poor at 30% as every single person is being contacted and very few people are actually buying the courses.
- To make the process more efficient the company wishes to identify most potential leads known as ‘Hot Leads’ and pursue them.
- This will can inturn take the conversion rate to almost 80%.

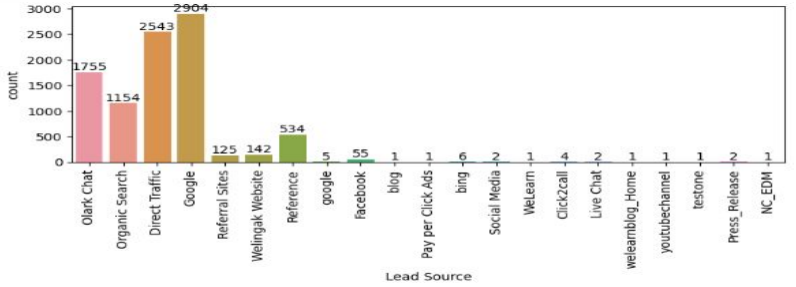
PROBLEM APPROACH:

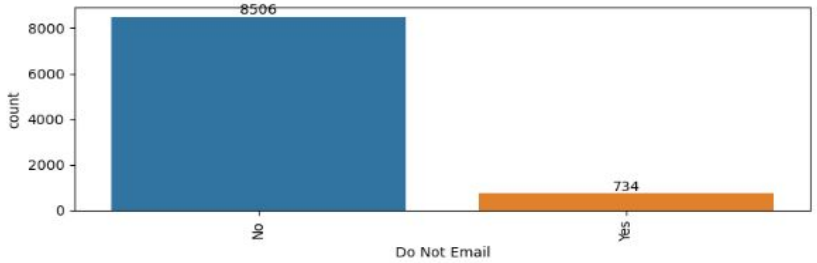
- Importing the data and inspecting the data frame
- Data Understanding and Inspection
- Data Cleaning
- Data Analysis (EDA)
- Data Preparation
- Test-Train Split
- Feature Scaling
- Model Building
- Model Evaluation
- Making Predictions on test set

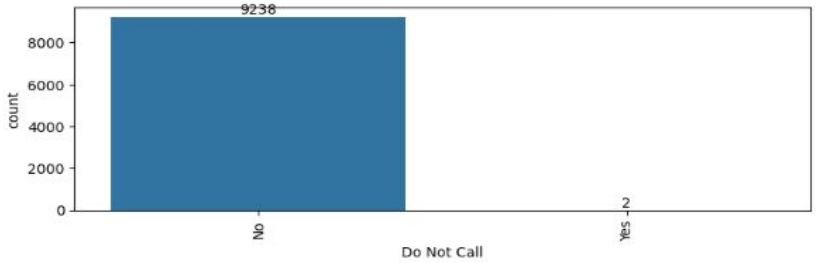
Data Analysis (EDA)

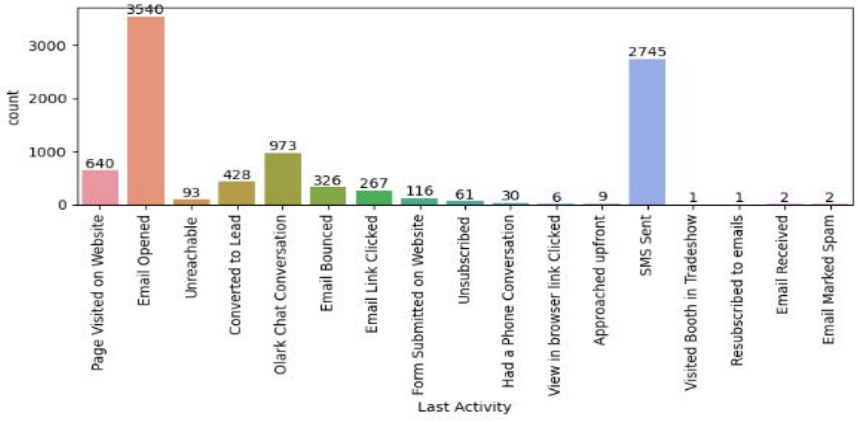
Checking & Dropping Category Columns that are Skewed

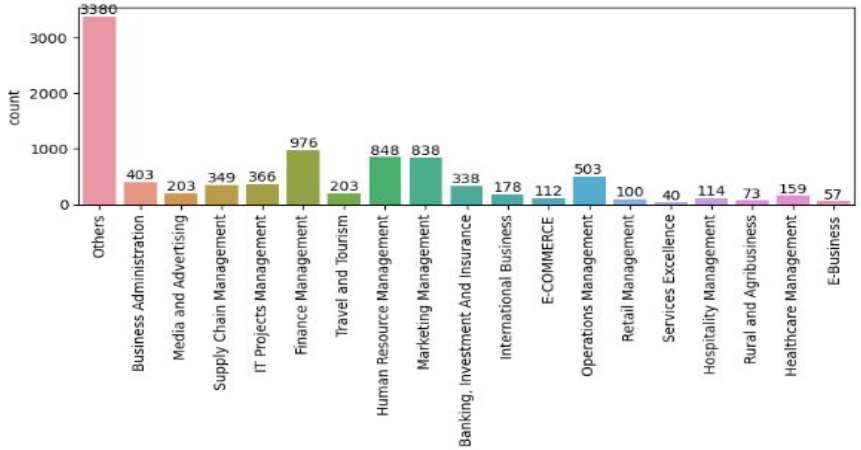


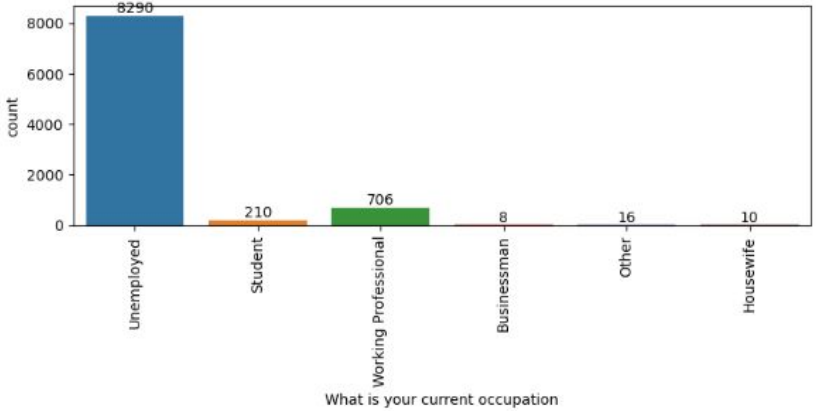


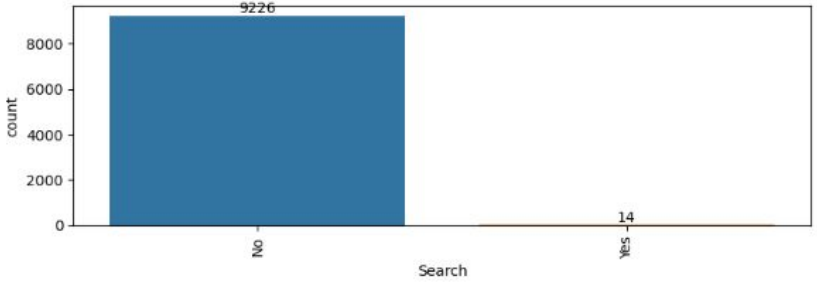


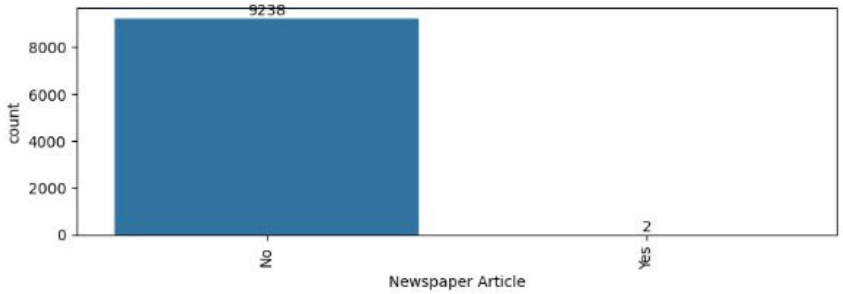


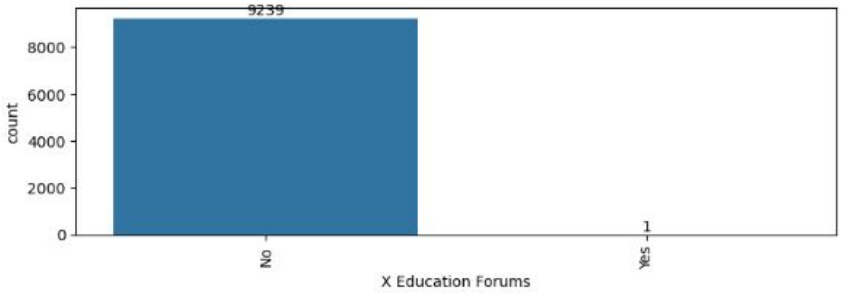

count
3000
2000
1000
0
640
3540
93
428
973
326
267
116
61
30
6
9
2745
1
1
2
2
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
Approached upfront
SMS Sent
Visited Booth in Tradeshow
Resubscribed to emails
Email Received
Email Marked Spam
Last Activity

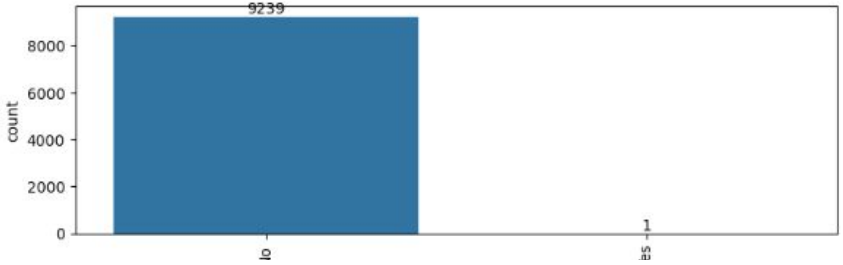

count
3000
2000
1000
0
3380
403
203
349
366
976
203
848
838
338
178
112
503
100
40
114
73
159
57
Others
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business

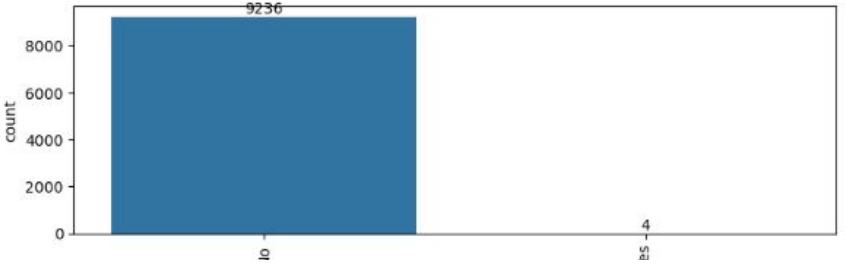

count
8000
6000
4000
2000
0
8290
210
706
8
16
10
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation

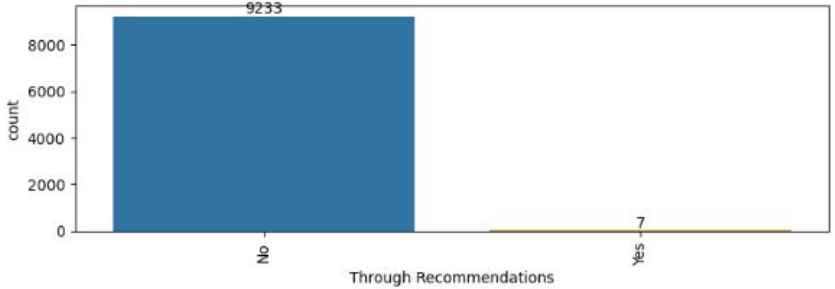

count
8000
6000
4000
2000
0
9226
14
No
Yes
Search

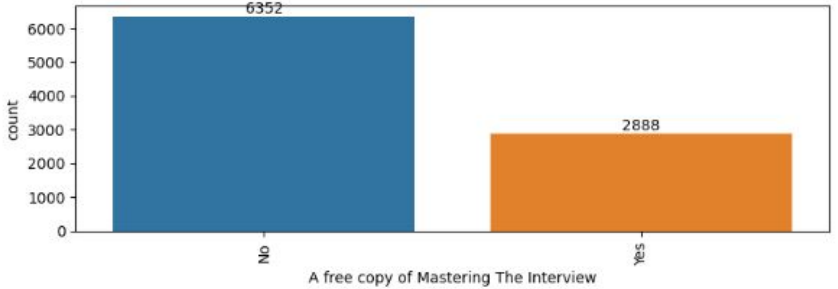

9238
2
count
No
Yes
Newspaper Article


9239
1
count
No
Yes
X Education Forums

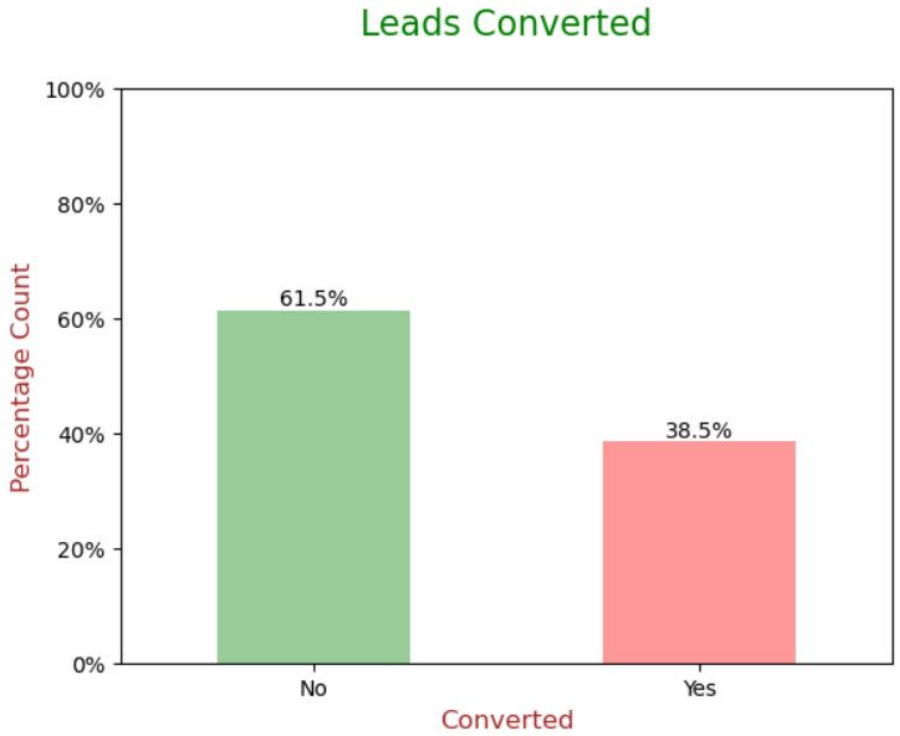

9239
1
count


9236
4
count

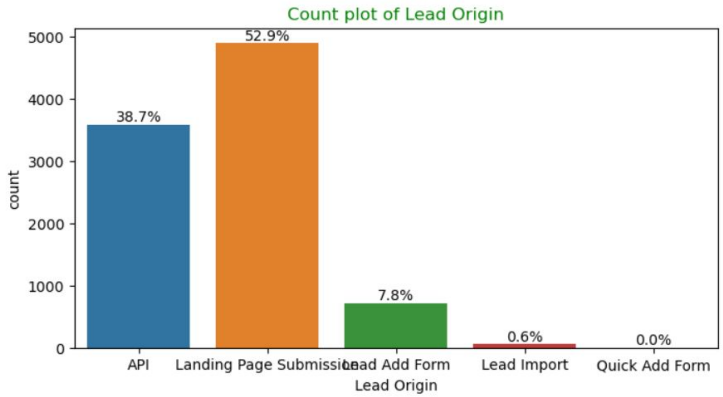

9233
7
count
No
Yes
Through Recommendations

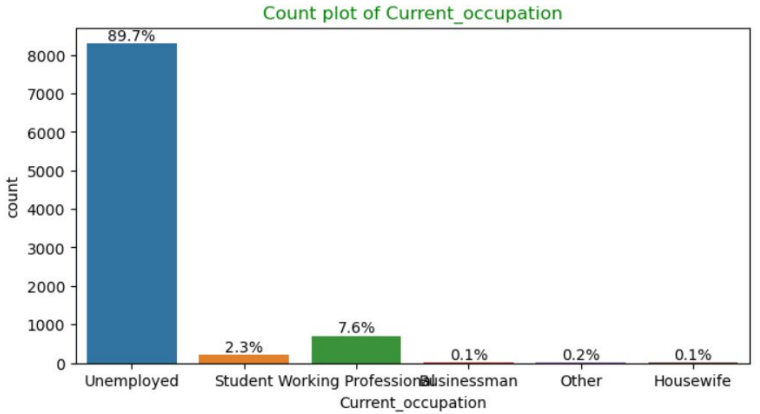

6352
2888
count
No
Yes
A free copy of Mastering The Interview

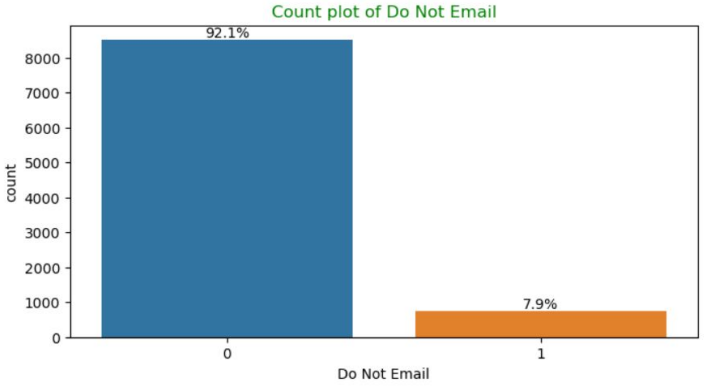
Checking the imbalance in the target variable

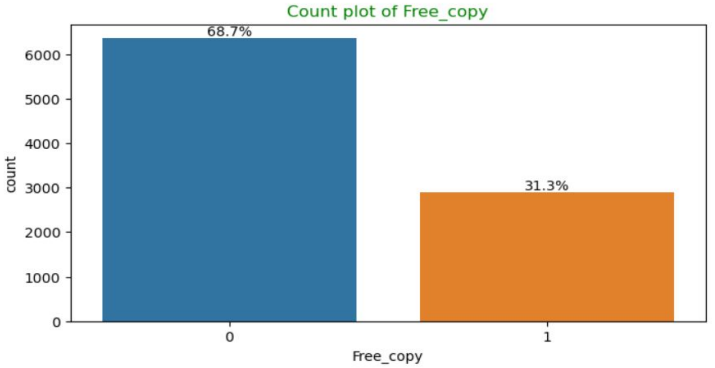


Univariate Analysis for Categorical Variables

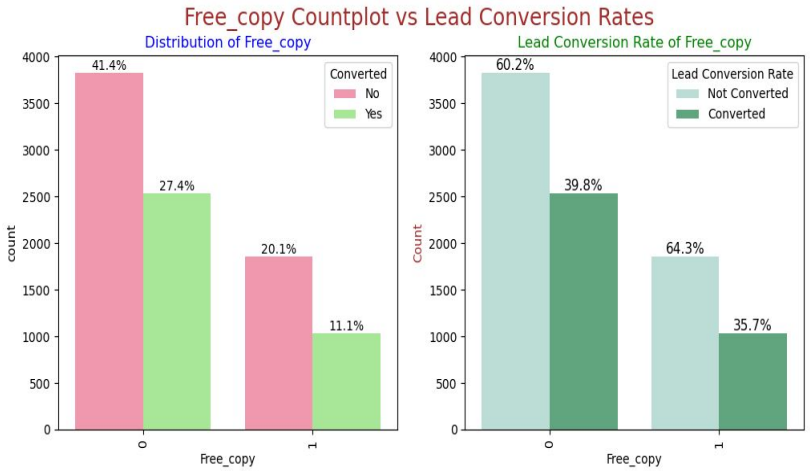


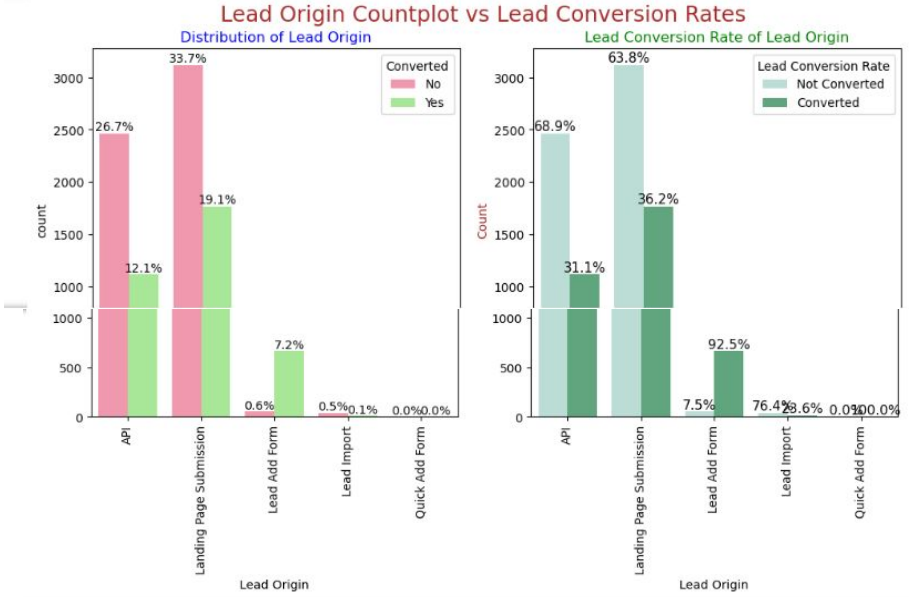




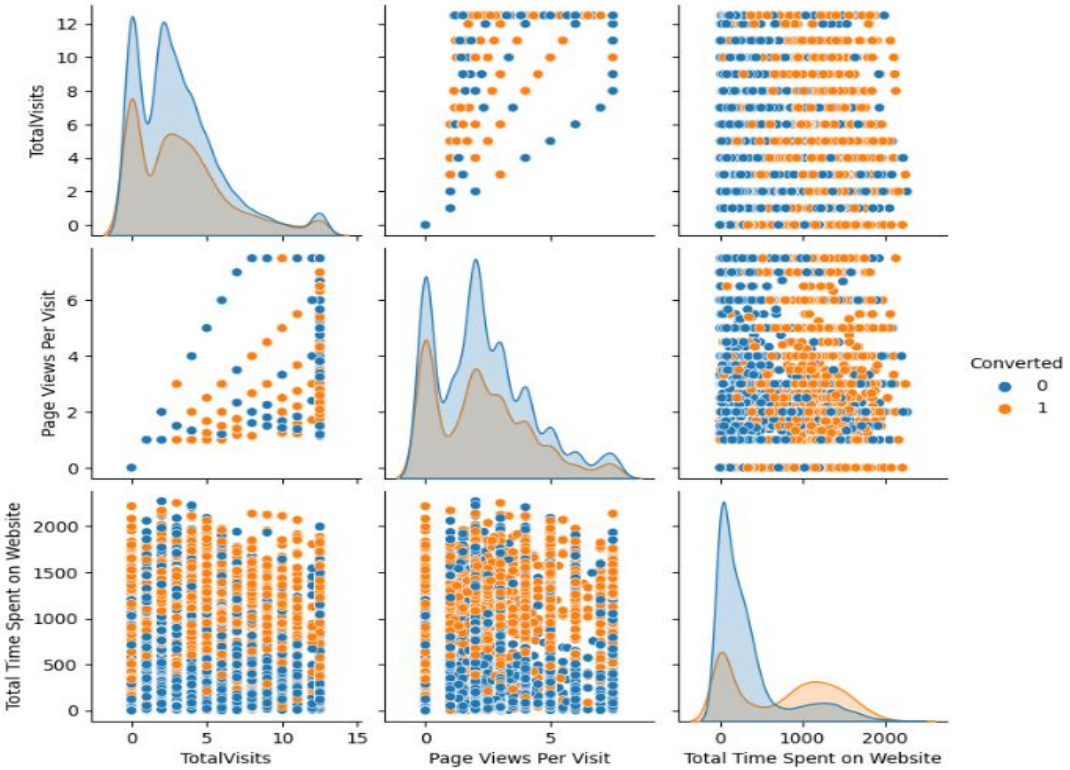


Bivariate Analysis

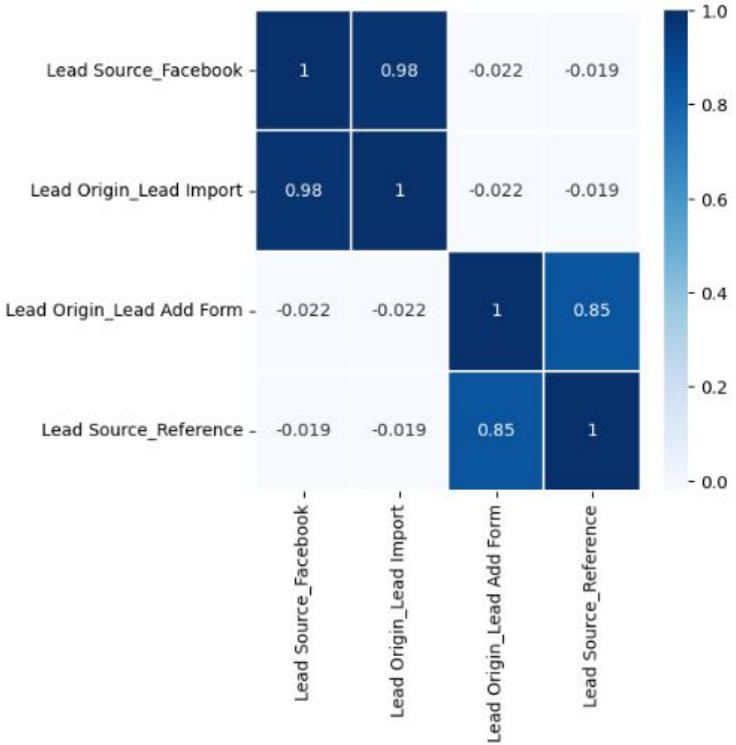


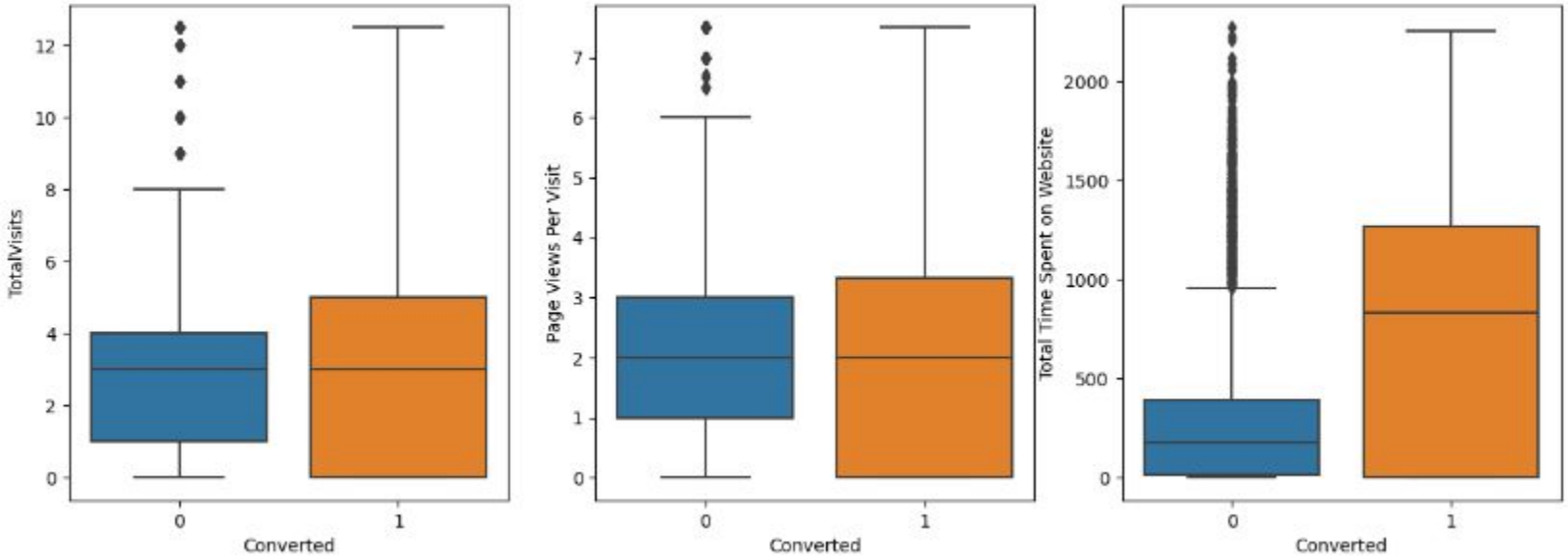


Bivariate Analysis for Numerical Variables

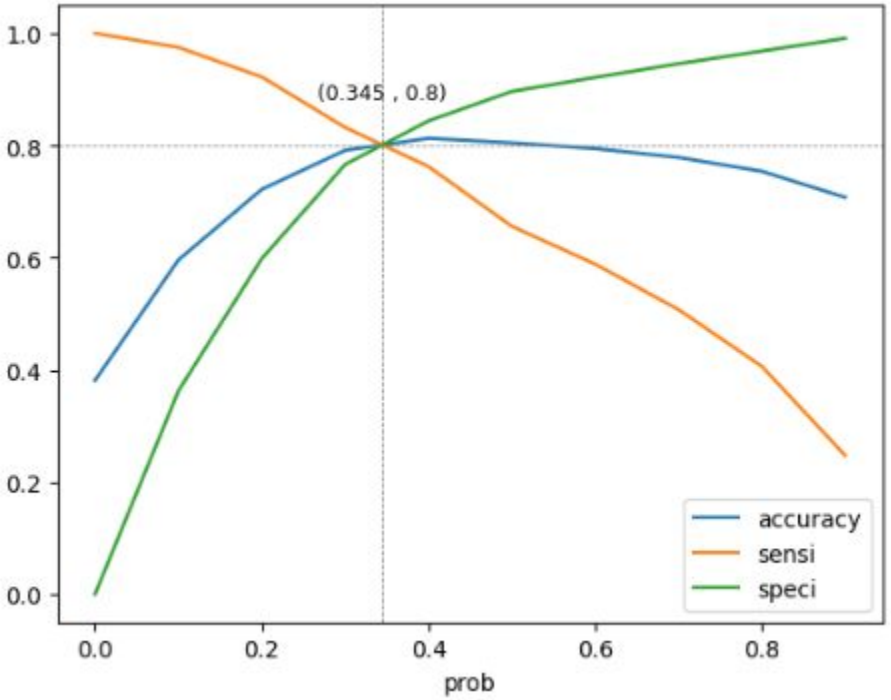


Correlation Matrix

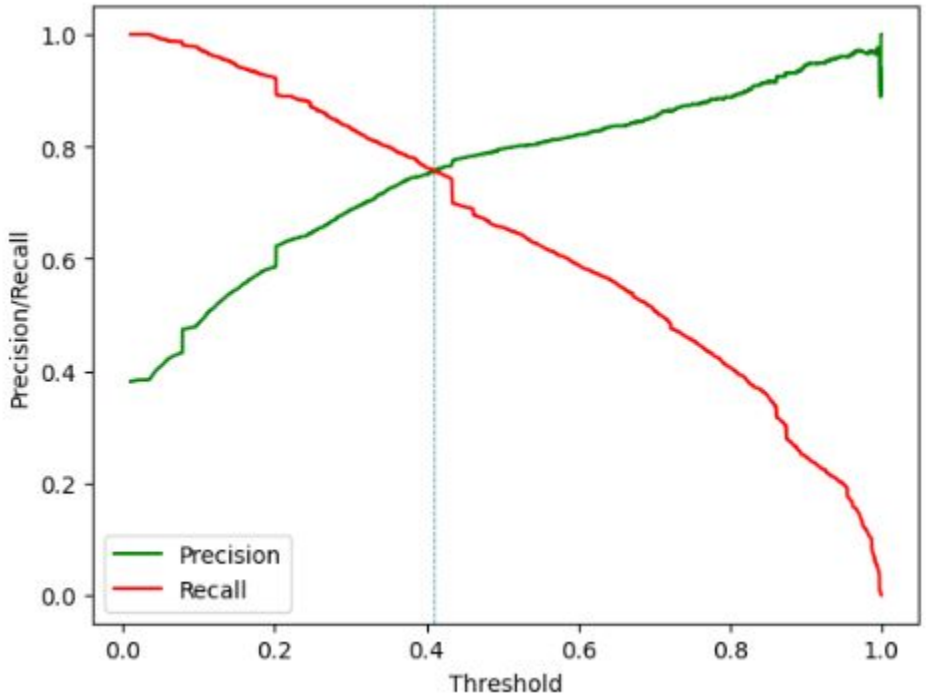



TotalVisits
Converted
Page Views Per Visit
Converted
Total Time Spent on Website
Converted

Finding Optimal Cutoff Point/ Probability



Precision and recall tradeoff



Conclusion

Train Data Set:

- Accuracy: 80.46%
- Sensitivity: 80.05%
- Specificity: 80.71%

Test Data Set:

- Accuracy: 80.34%
- Sensitivity: 79.82% ≈ 80%
- Specificity: 80.68%

Model Parameters:

- The Optimal cutoff probability point is 0.345. - Converted probability greater than 0.345 will be predicted as Converted lead (Hot lead) & probability smaller than 0.345 will be predicted as not Converted lead (Cold lead).
- Top 3 features that contribute positively are

1. Lead Source_Welingak Website
2. Lead Source_Reference
3. Current_occupation_Working Professional