GROUP PROJECT

Lead Score Case Study

By:

- Nguyen Huy Thang
- Bala Subramanyam
- Nairit Dutta

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor.
- 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.

01

Business Objective

- Increase the current lead conversion rate from 30% to a target of 80% by focusing on the most promising leads.
- Build a logistic regression model that assigns a lead score (between 0 and 100) to each lead. This score will help in identifying 'Hot Leads' who are more likely to convert into paying customers.
- Ensure the model is adaptable to accommodate potential changes in the company’s requirements.

Solution Methodology

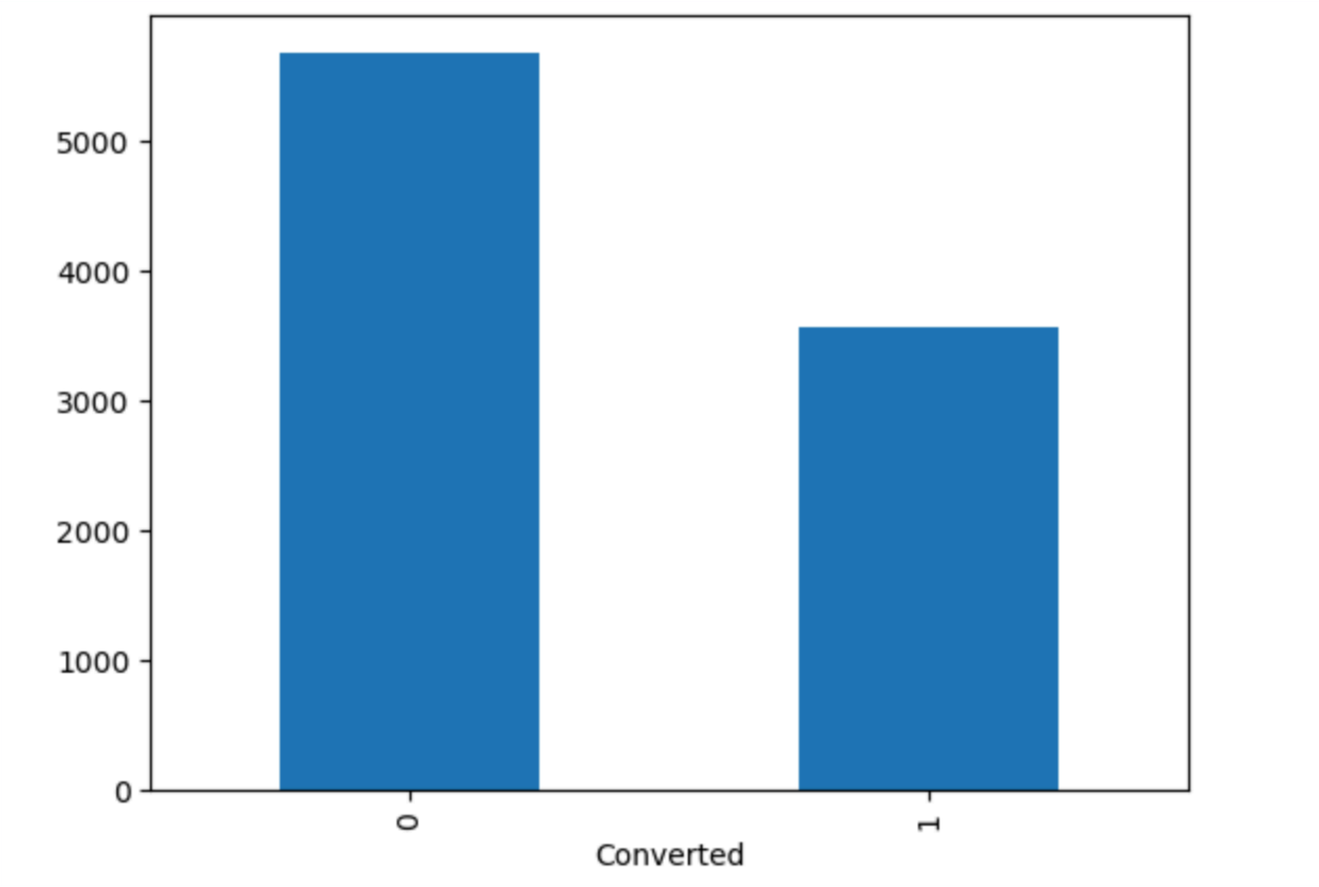
- Exploratory Data Analysis and Data Cleaning
- Data Preparation
- Model Building
- Model Evaluation
- Conclusions and Recommendations for the Company Strategy

01

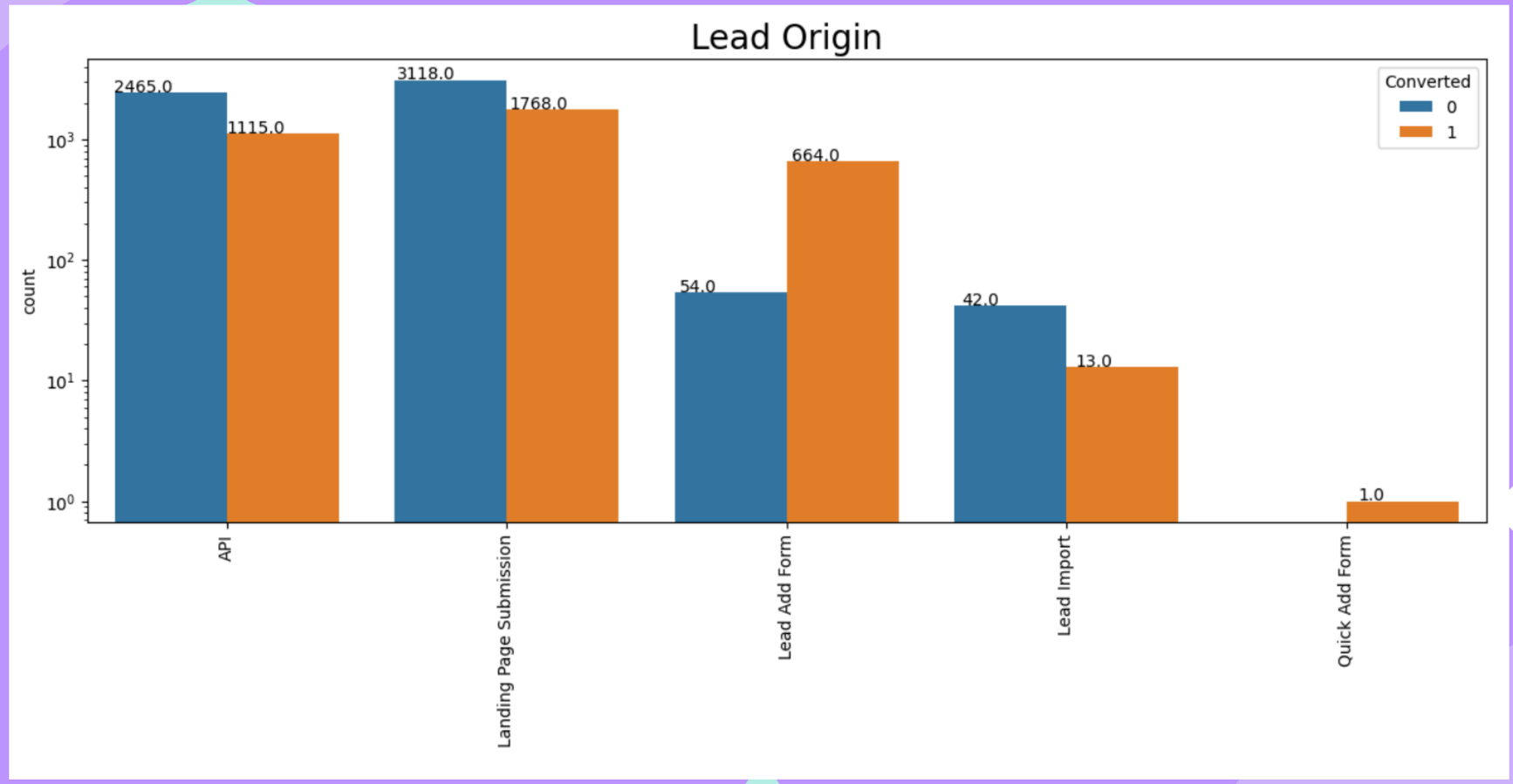
Data Cleaning

- Total number of rows is 37, total number of columns is 9240.
- Eliminate redundant variables.
- Delete columns with null values greater than 40%.
- Remove outlier values.

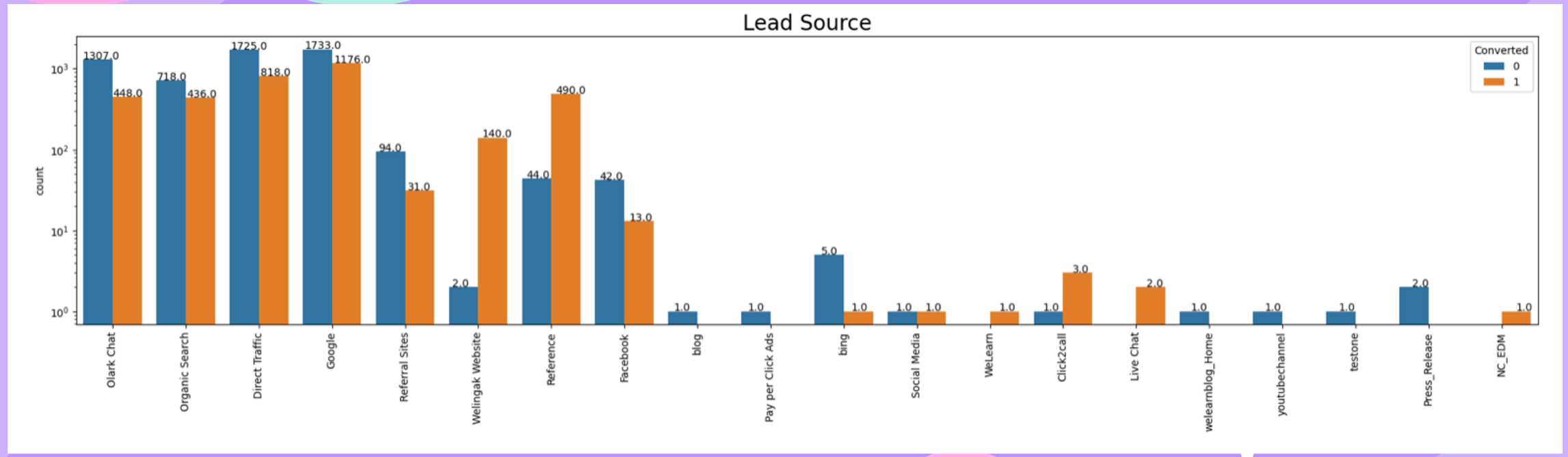
EDA



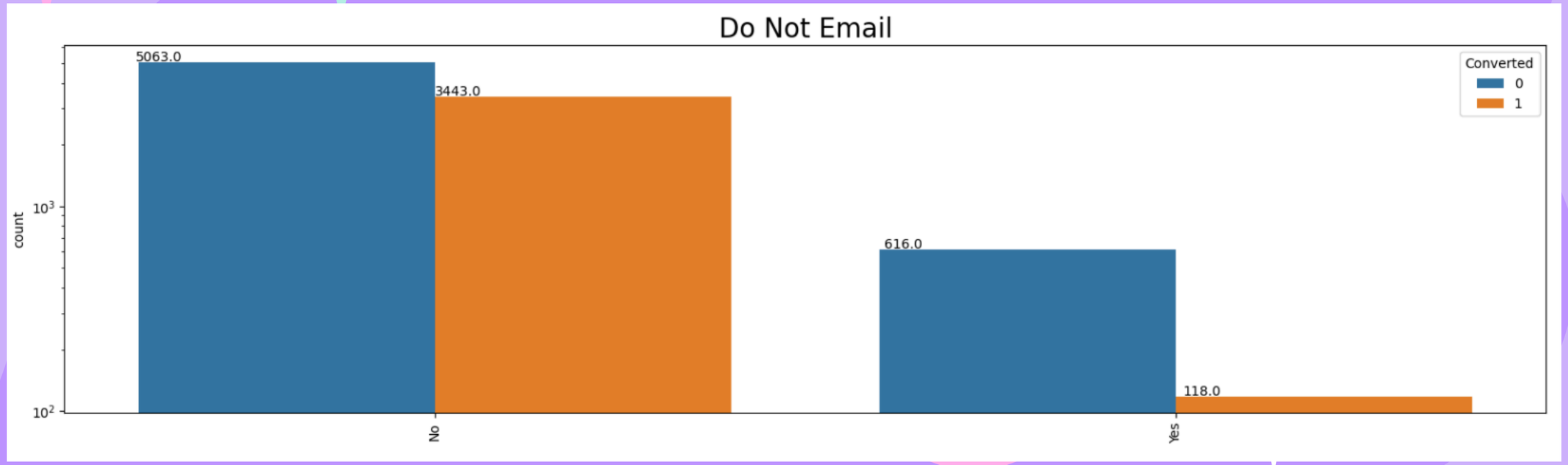
EDA


Lead Origin
Converted
0
1
count
10^3
10^2
10^1
10^0
2465.0
1115.0
3118.0
1768.0
54.0
664.0
42.0
13.0
1.0
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form

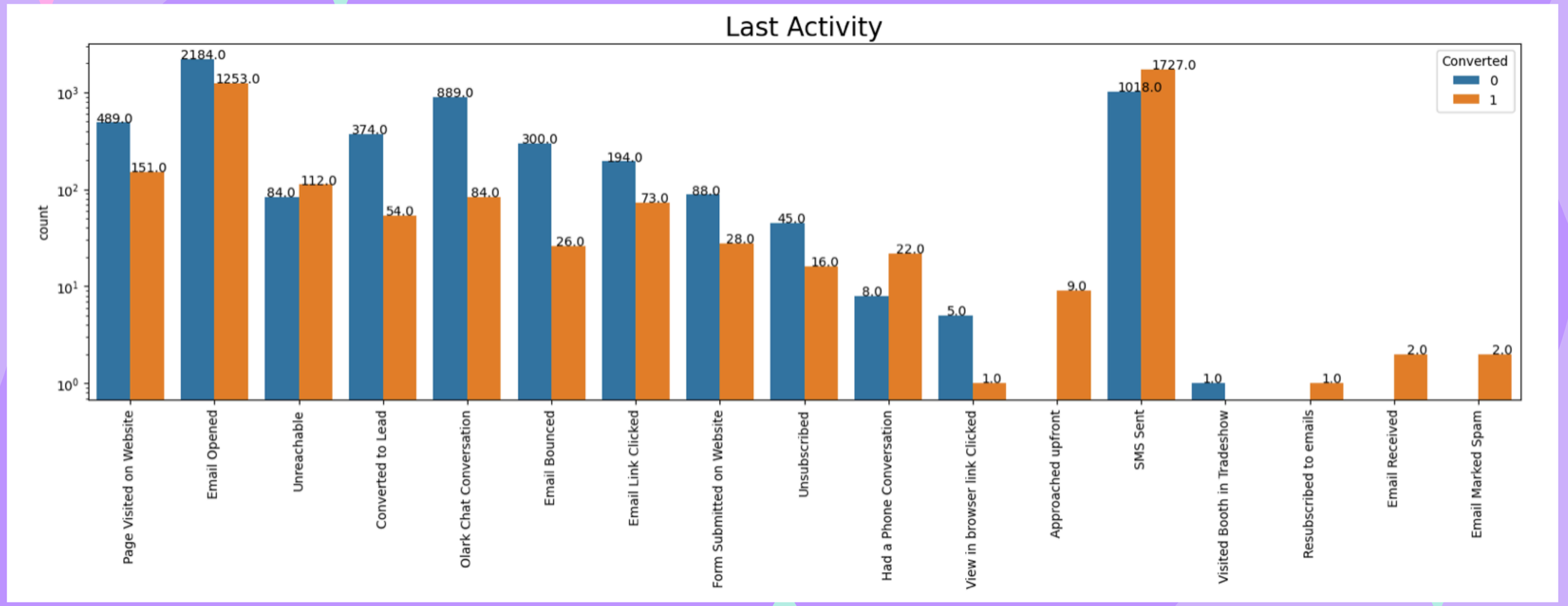
EDA


Lead Source
Converted
0
1
count
10^3
10^2
10^1
10^0
1307.0
448.0
718.0
436.0
1725.0
818.0
1733.0
1176.0
94.0
31.0
2.0
140.0
44.0
490.0
42.0
13.0
1.0
1.0
5.0
1.0
1.0
1.0
1.0
1.0
3.0
2.0
1.0
1.0
1.0
2.0
1.0
Olark Chat
Organic Search
Direct Traffic
Google
Referral Sites
Welingak Website
Reference
Facebook
blog
Pay per Click Ads
bing
Social Media
WeLearn
Click2call
Live Chat
welearnblog_Home
youtubechannel
testone
Press_Release
NC_EDM

EDA

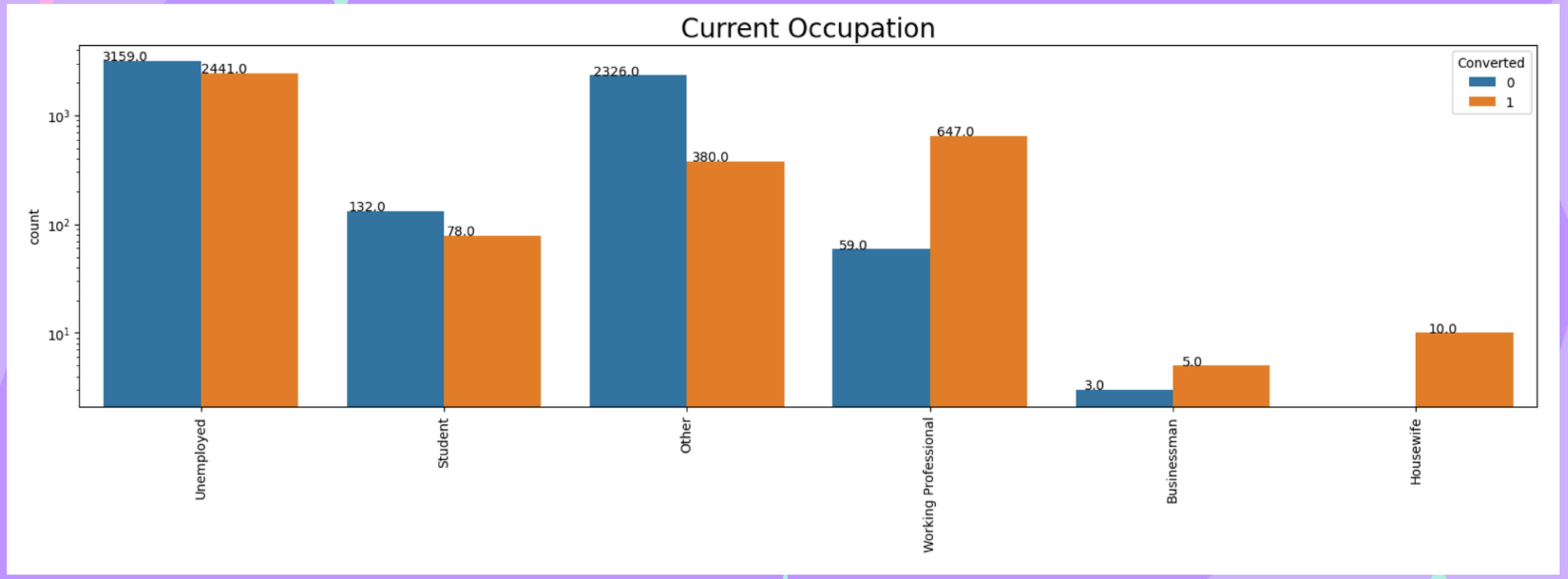

Do Not Email
Converted
0
1
count
10^3
10^2
5063.0
3443.0
616.0
118.0
No
Yes

EDA

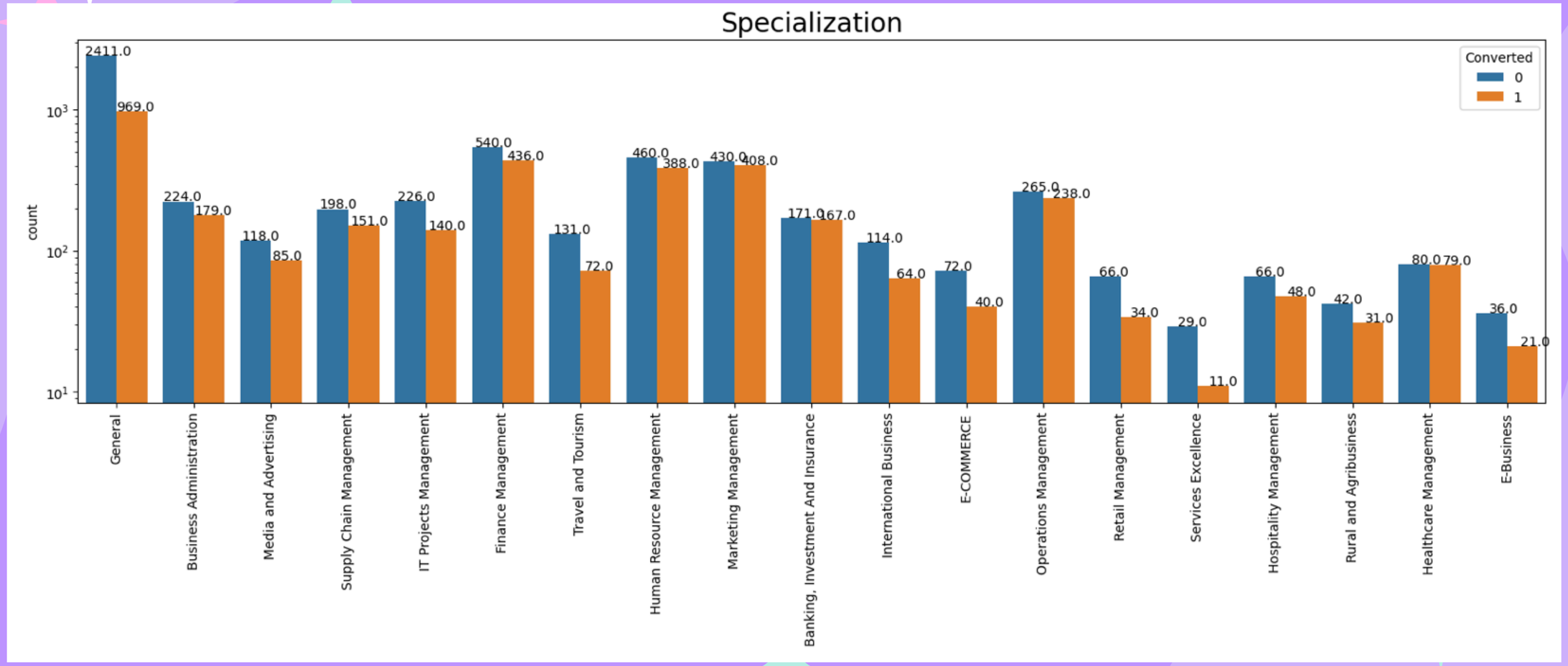

Last Activity
Converted
0
1
count
10^3
10^2
10^1
10^0
489.0
151.0
2184.0
1253.0
84.0
112.0
374.0
54.0
889.0
84.0
300.0
26.0
194.0
73.0
88.0
28.0
45.0
16.0
8.0
22.0
5.0
1.0
9.0
1018.0
1727.0
1.0
1.0
2.0
2.0
Page Visited on Website
Email Opened
Unreachable
Converted to Lead
Olark Chat Conversation
Email Bounced
Email Link Clicked
Form Submitted on Website
Unsubscribed
Had a Phone Conversation
View in browser link Clicked
Approached upfront
SMS Sent
Visited Booth in Tradeshow
Resubscribed to emails
Email Received
Email Marked Spam

06

EDA

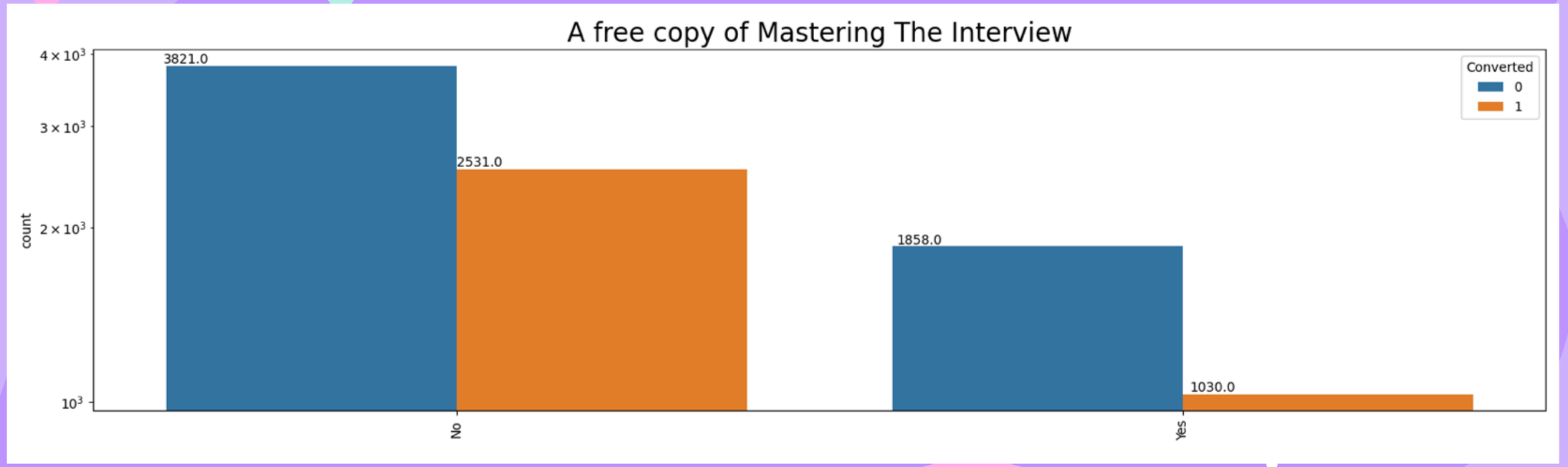

Current Occupation
Converted
0
1
count
10^3
10^2
10^1
3159.0
2441.0
132.0
78.0
2326.0
380.0
59.0
647.0
3.0
5.0
10.0
Unemployed
Student
Other
Working Professional
Businessman
Housewife

EDA

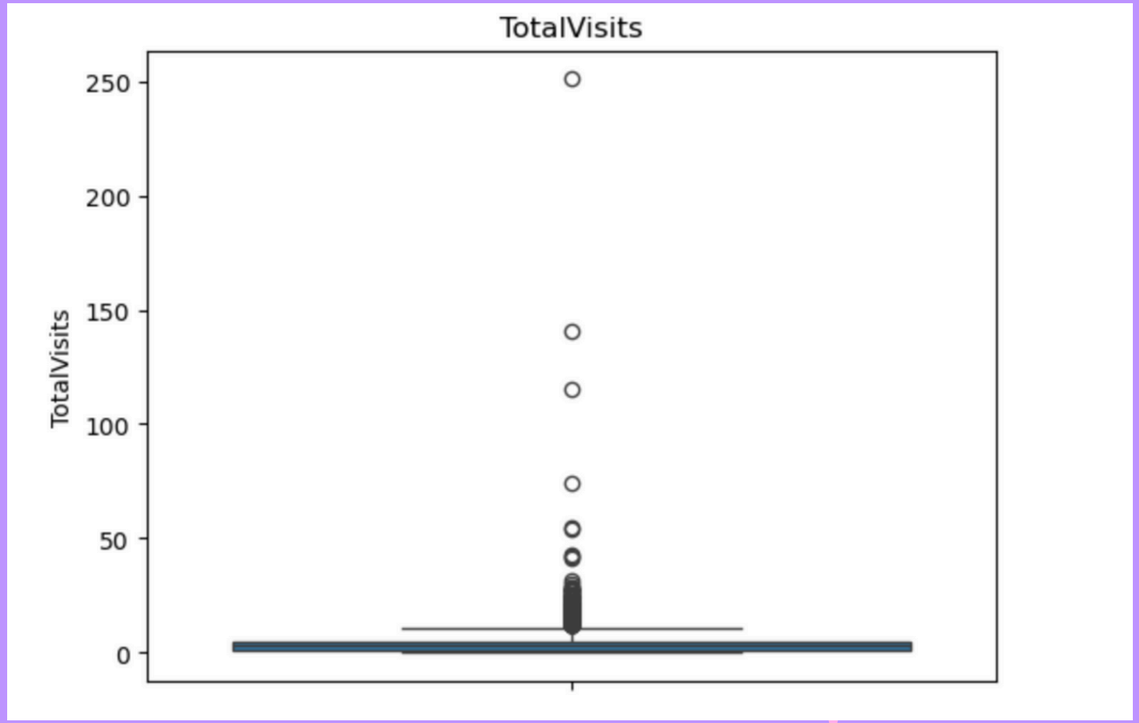

Specialization
Converted
0
1
count
10^3
10^2
10^1
2411.0
969.0
224.0
179.0
118.0
85.0
198.0
151.0
226.0
140.0
540.0
436.0
131.0
72.0
460.0
388.0
430.0
408.0
171.0
167.0
114.0
64.0
72.0
40.0
265.0
238.0
66.0
34.0
29.0
11.0
66.0
48.0
42.0
31.0
80.0
79.0
36.0
21.0
General
Business Administration
Media and 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

06

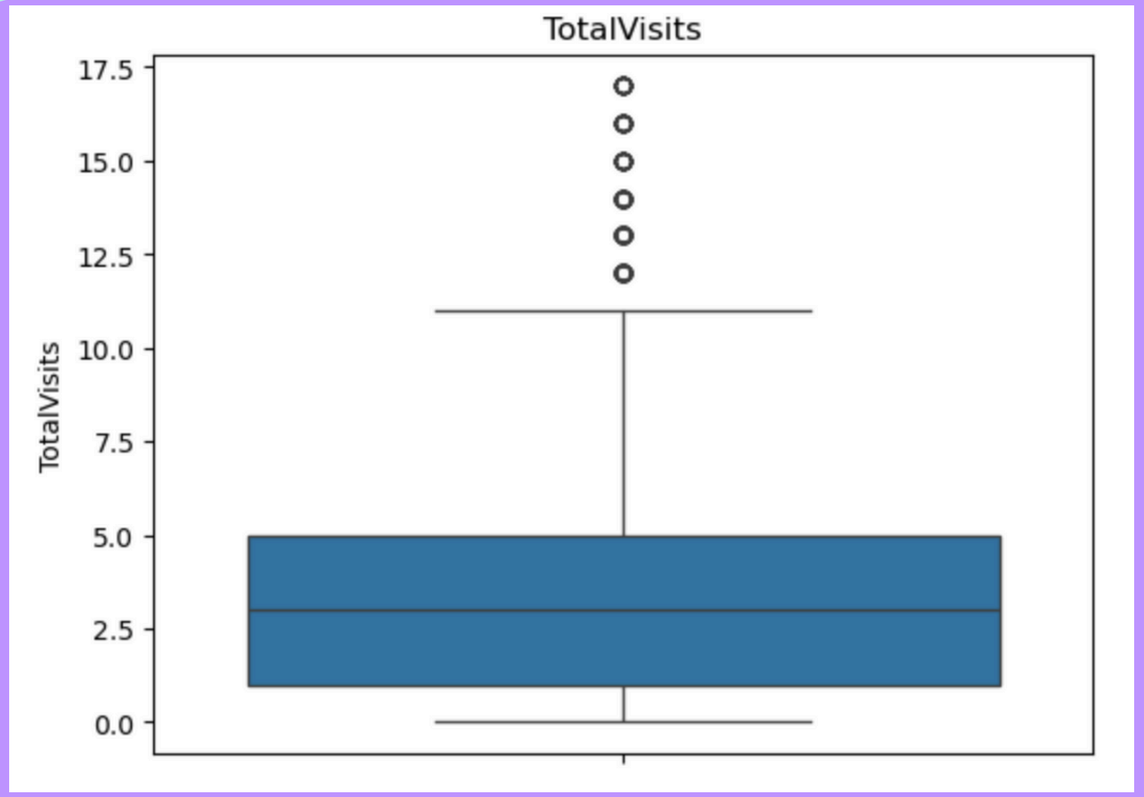
EDA


A free copy of Mastering The Interview
count
4×10^3
3×10^3
2×10^3
10^3
3821.0
2531.0
1858.0
1030.0
Converted
0
1
No
Yes

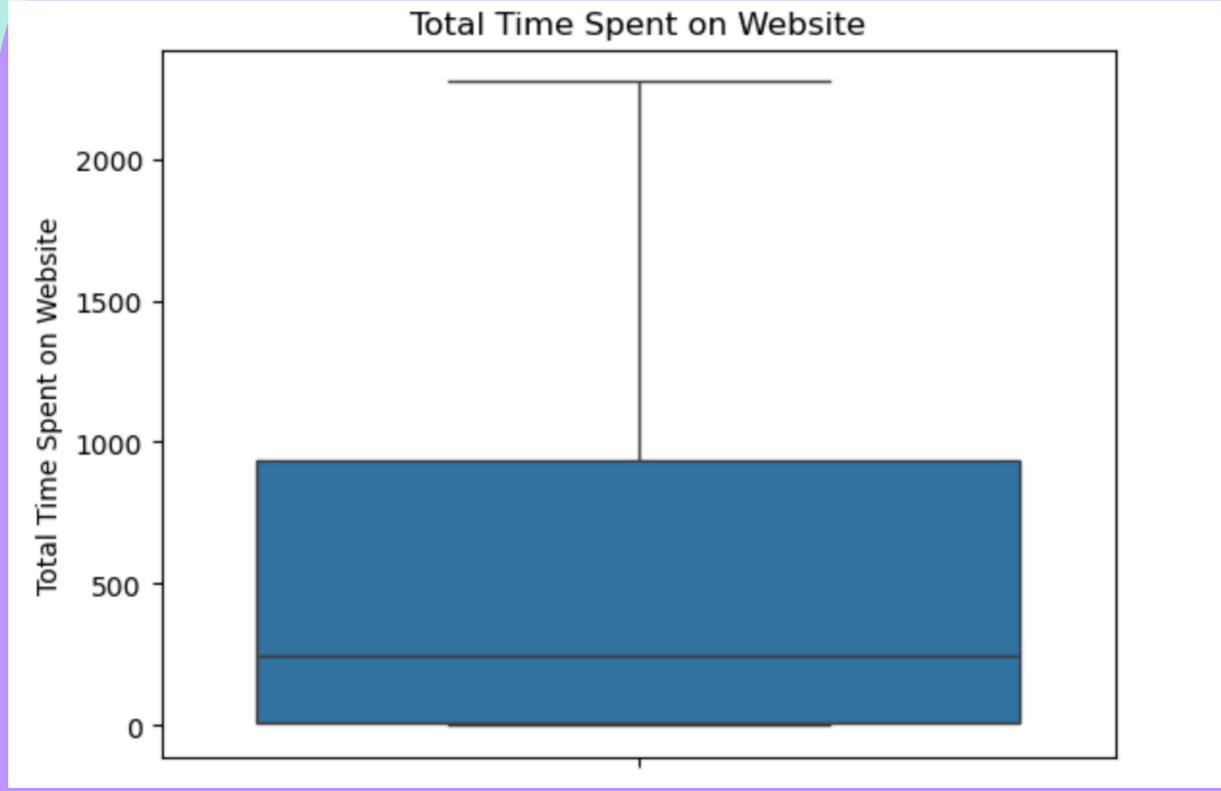
EDA



EDA

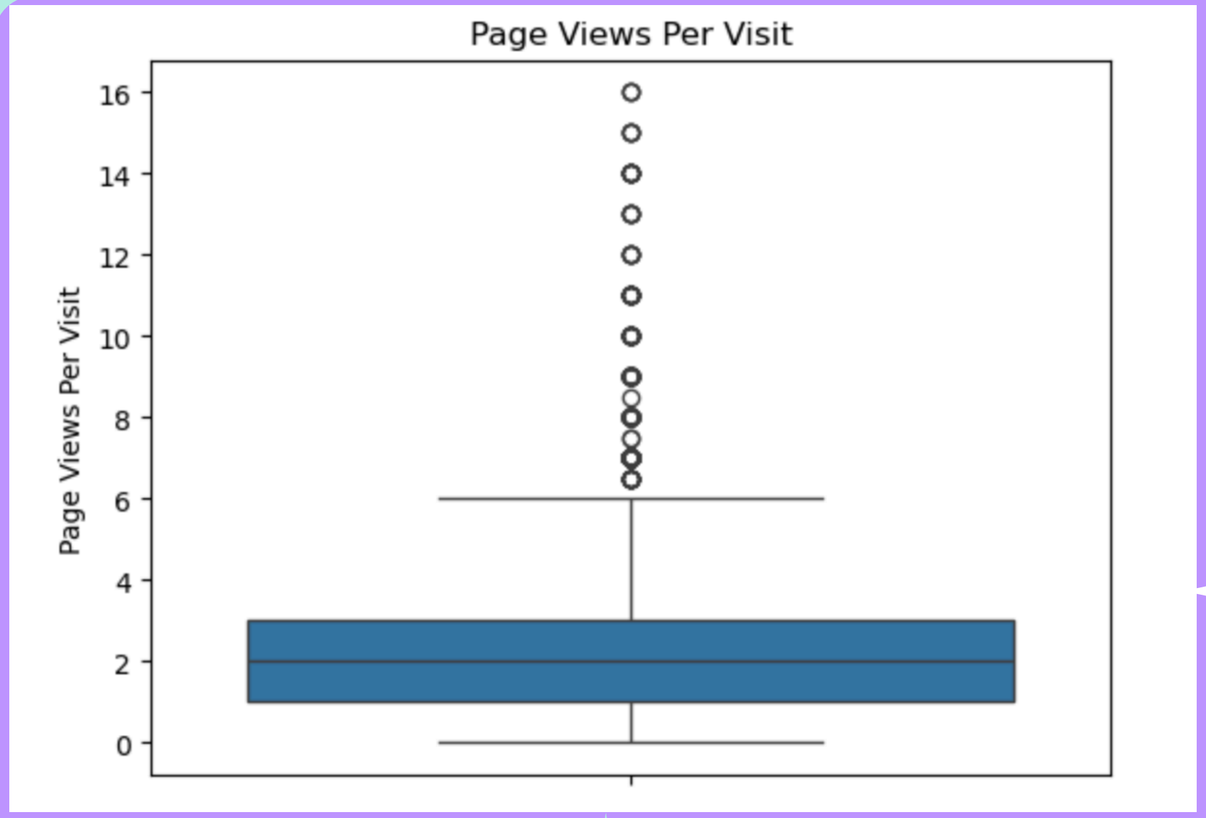


EDA

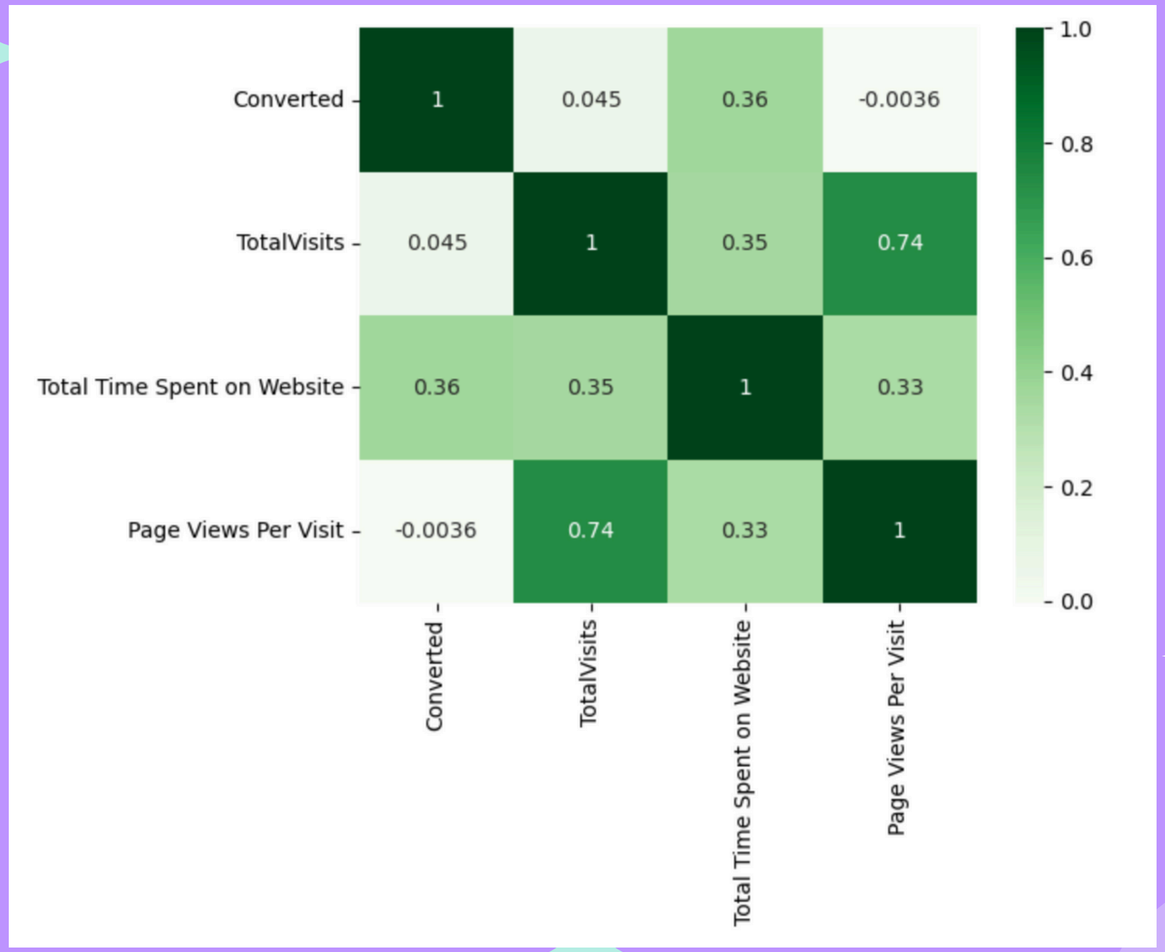

Total Time Spent on Website
Total Time Spent on Website
2000
1500
1000
500
0

06

EDA



EDA


Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
1
0.045
0.36
-0.0036
0.045
1
0.35
0.74
0.36
0.35
1
0.33
-0.0036
0.74
0.33
1
1.0
0.8
0.6
0.4
0.2
0.0

01

Data Preparation

- Categorical variables are converted to 0 or 1.
- Split the dataset into train and test sets with a 70:30 ratio.
- Data points are standardized using the StandardScaler for numerical features to avoid bias from variables with higher scales in the model.

01

Model Building

- The initial model is created by selecting 15 features using the RFE technique.
- Insignificant variables have been identified and removed to optimize the model for better performance. VIF is also checked to assess multicollinearity among the model's features.
- An important step in logistic regression is to find the optimal threshold for probability that aligns with business needs and ensures good accuracy, sensitivity, and specificity. We obtained 0.398 as the cutoff point from the chart that balances accuracy, sensitivity, specificity, and probability range.
- Precision and recall views were then examined to refine thresholds for predictions in the test dataset. The cutoff point obtained from the Recall-Precision chart was 0.445. Model evaluation was completed by checking all performance metrics, including the ROC curve, accuracy, sensitivity, specificity, recall, and precision, all of which were found to be at acceptable levels. Optimal point 0.398 was selected as it had higher sensitivity and recall of 81% .

01

Model Evaluation

Predicted values are calculated using the model. The lead score is 100 times the predicted log odds, ranging from 0 to 100.

01

Conclusions and Recommendations

- The model evaluation steps indicate that the metrics for accuracy, precision, and recall are all at acceptable levels. The recall score is also slightly higher than the precision score, which aligns well with future business needs.
- Focusing on leads from the "Add" form is crucial, as they contribute the most to conversions. Prioritizing these leads can enhance your conversion rates significantly.
- Professionals should be the ideal target.
- Leads from the Welingak website have potential for conversion, so advertising revenue should be focused on this website.
- Individuals who do not wish to receive emails should be avoided.
- Avoid individuals without specific expertise.

Thank You