Sales Lead Score Case Study

Submitted by:

- Abhishek Kurri
- Rahul Koley
- Vasanth P

Problem Statement

- X Education is an online course provider that caters to industry professionals. The company promotes its courses through various websites and search engines, including Google.

- When potential customers visit X Education's website, they can peruse the available courses, complete a course inquiry form, or view informative videos. By providing their email address or phone number, these visitors become leads. Additionally, past referrals also serve as a source of leads for the company.

- Upon obtaining these leads, the sales team initiates contact by making phone calls or sending emails. This process results in some leads converting to paying customers, although the majority do not. Typically, the lead conversion rate at X Education hovers around 30%.

Business Goal

- X Education is seeking assistance in identifying the most viable leads, i.e., leads with a higher probability of converting into paying customers. The company requires a model that assigns a lead score to each lead, indicating their likelihood of converting. Higher lead scores correspond to greater conversion chances, while lower scores correspond to lower conversion chances.

- The CEO has set a rough target of an 80% lead conversion rate.

Strategy:

1. Source the data for analysis.
2. Clean and prepare the data.
3. Conduct exploratory data analysis.
4. Perform feature scaling.
5. Split the data into training and test datasets.
6. Build a logistic regression model and calculate the lead score.
7. Evaluate the model using different metrics, such as specificity and sensitivity, or precision and recall.
8. Apply the best model to the test data based on the sensitivity and specificity metrics.

Methodology for Problem Solution: A systematic approach to address the problem.

1

Data Sourcing , Cleaning and Preparation

1. Read the data from the source.
2. Convert the data into a clean format suitable for analysis.
3. Remove duplicate data.
4. Treat outliers.
5. Conduct exploratory data analysis.
6. Perform feature standardization.

2

Feature Scaling & Splitting Train and Test Sets

1. Feature Scaling of Numeric data
2. Splitting data into train and test set.

3

Model Building

1. Feature selection using RFE.
2. Determine the optimal model using logistic regression.
3. Calculate various metrics such as accuracy, sensitivity, specificity, precision, and recall, and evaluate the model.

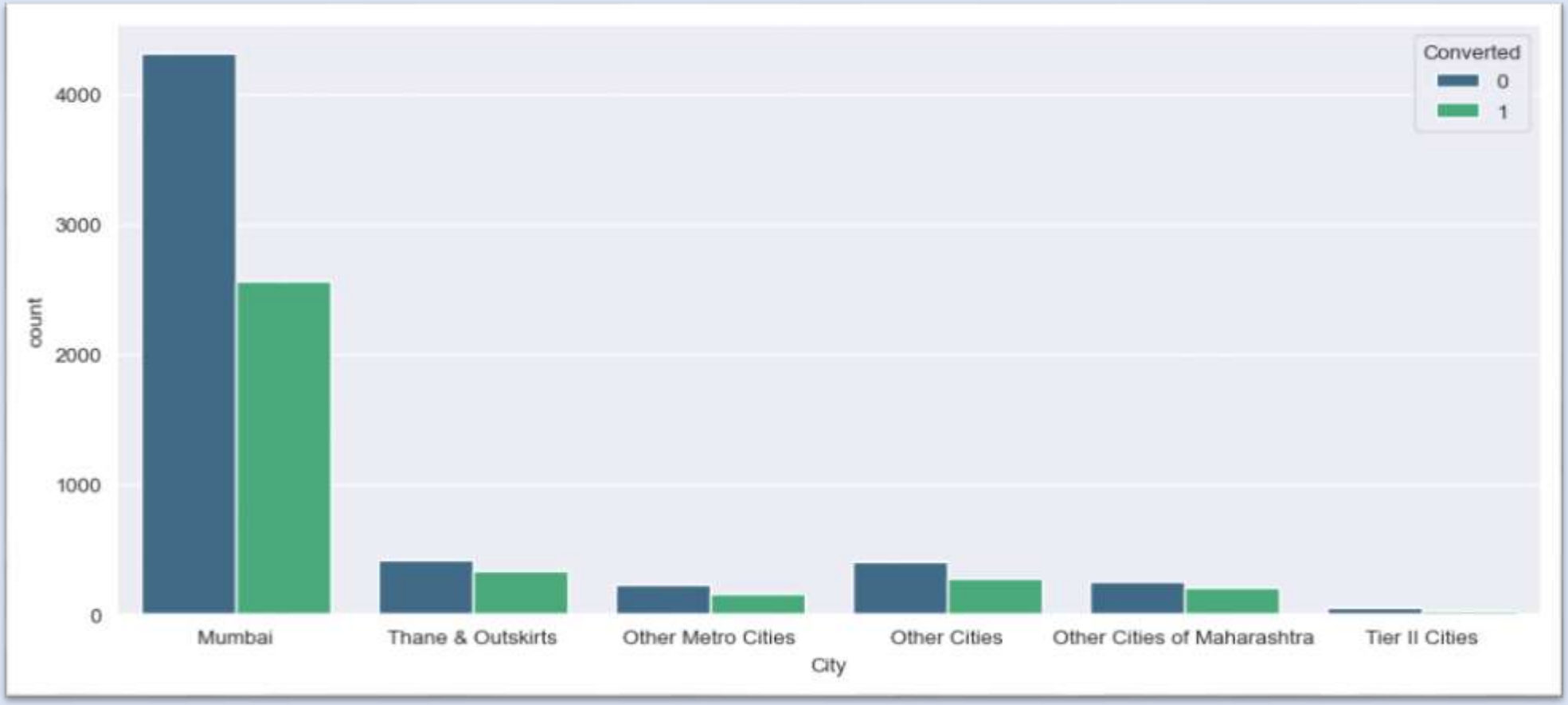
4

Result

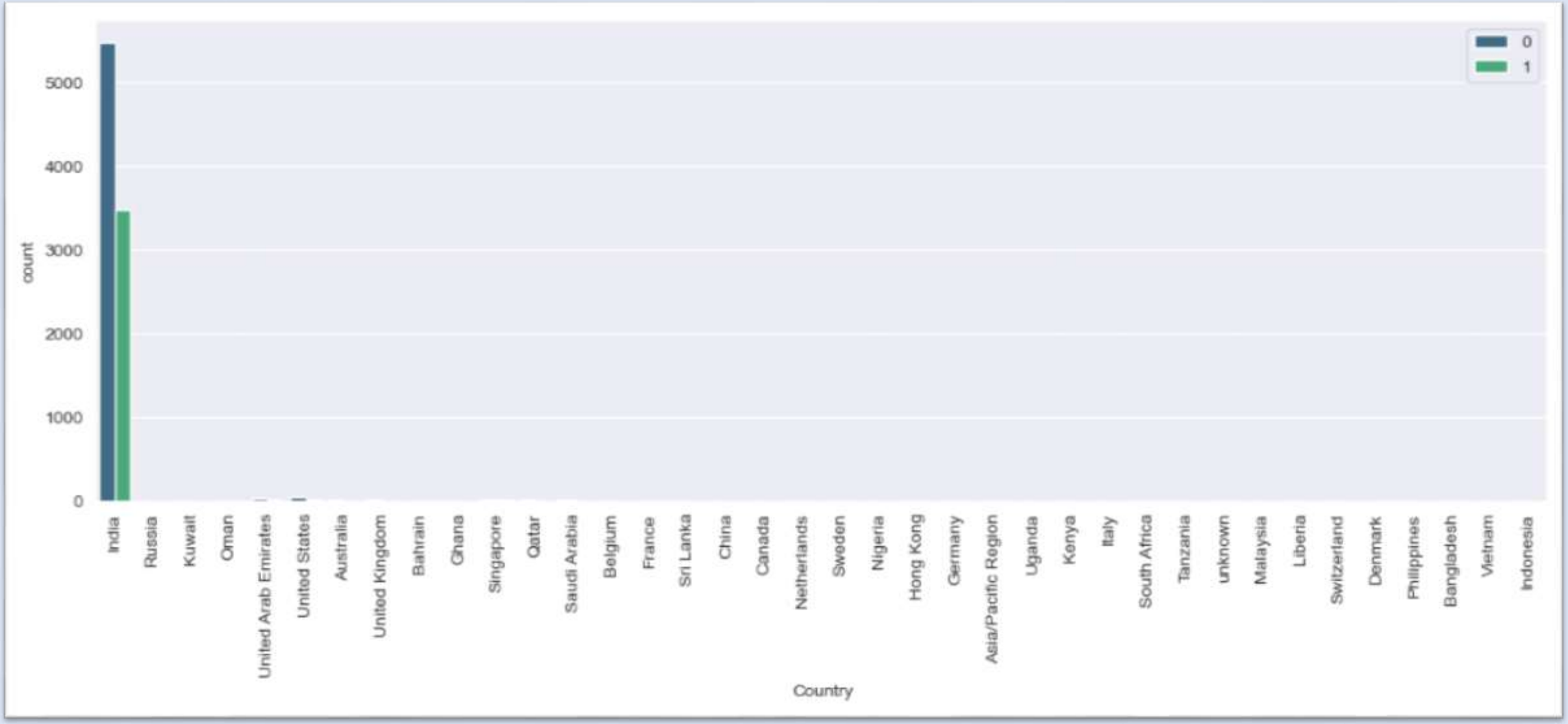
1. Determine the lead score and check if the target final predictions amount to an 80% conversion rate.
2. Evaluate the final prediction on the test set using the cut-off threshold from sensitivity and specificity metrics.

EDA – Exploratory Data Analysis

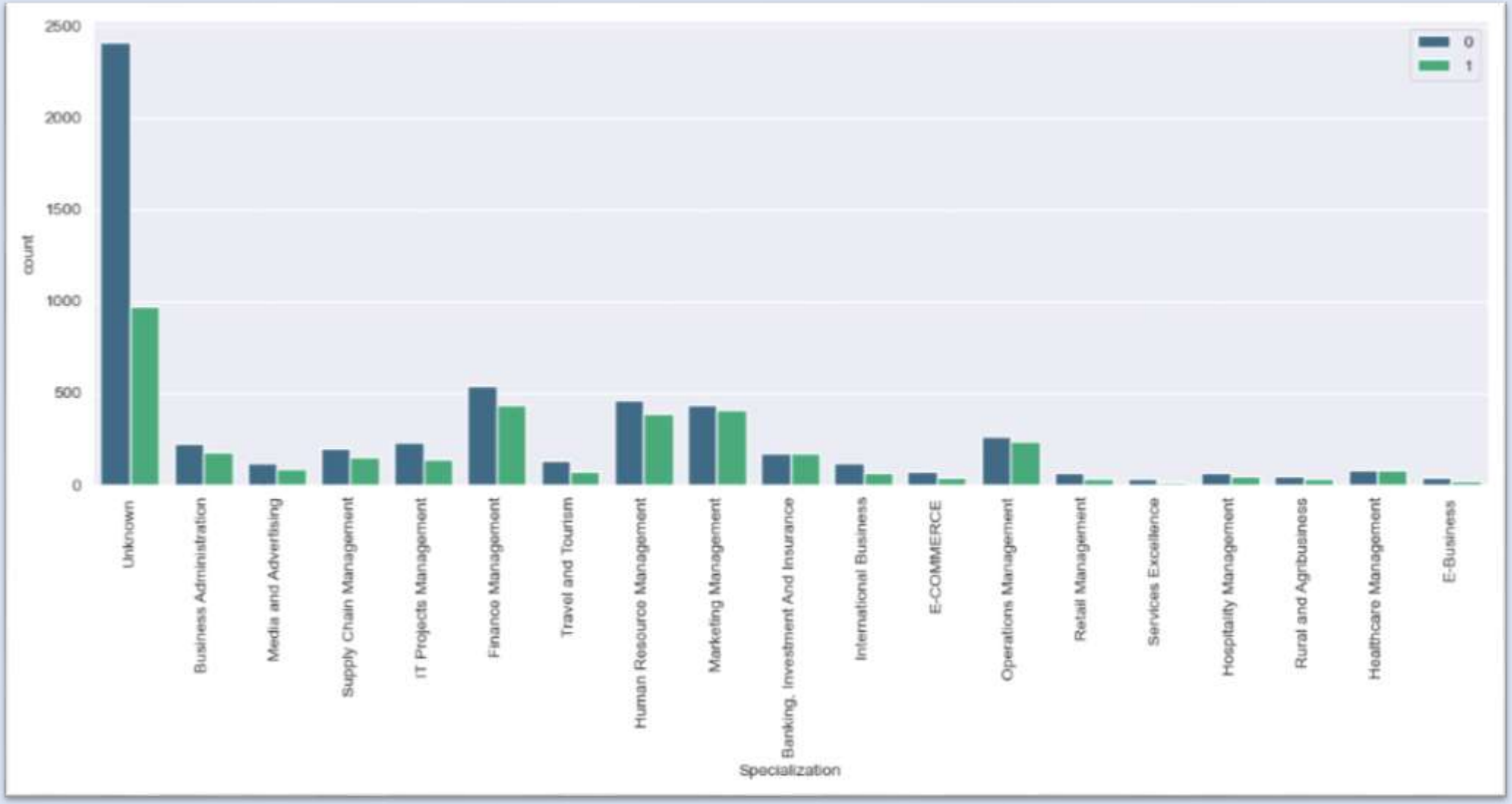
majority of target audience is from Mumbai city



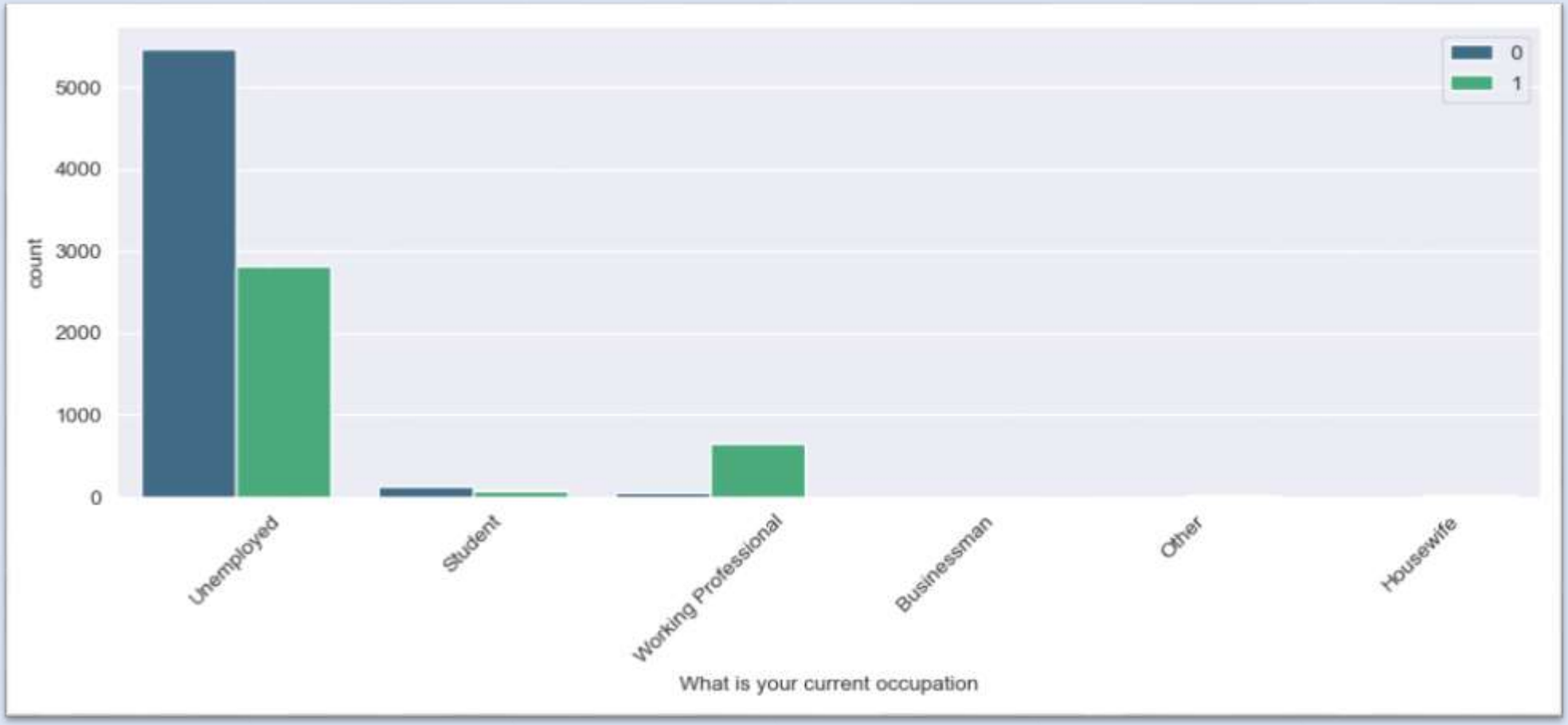
Country India shows biasness in the data, Hence we can remove the column



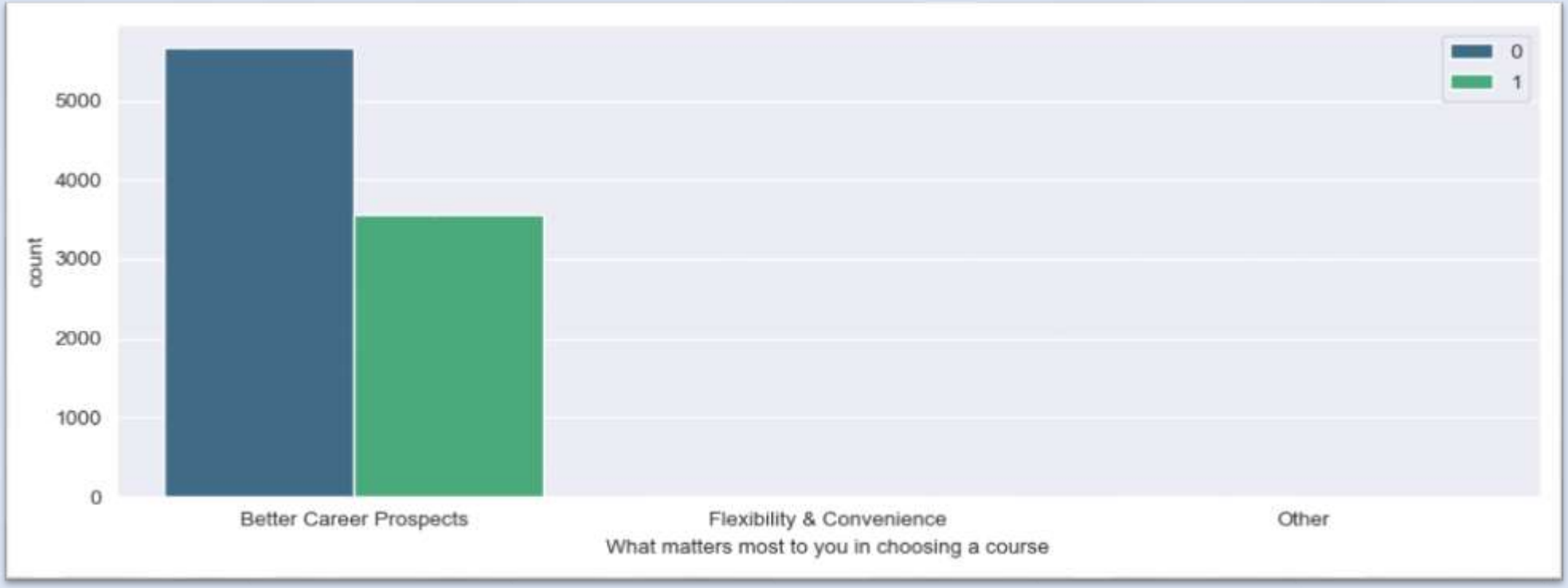
Profiles with different specializations: As you can see, the number of profiles listed as "unknown" is quite high.
To address this issue, we combined many management positions into a single category called "Management Specializations."



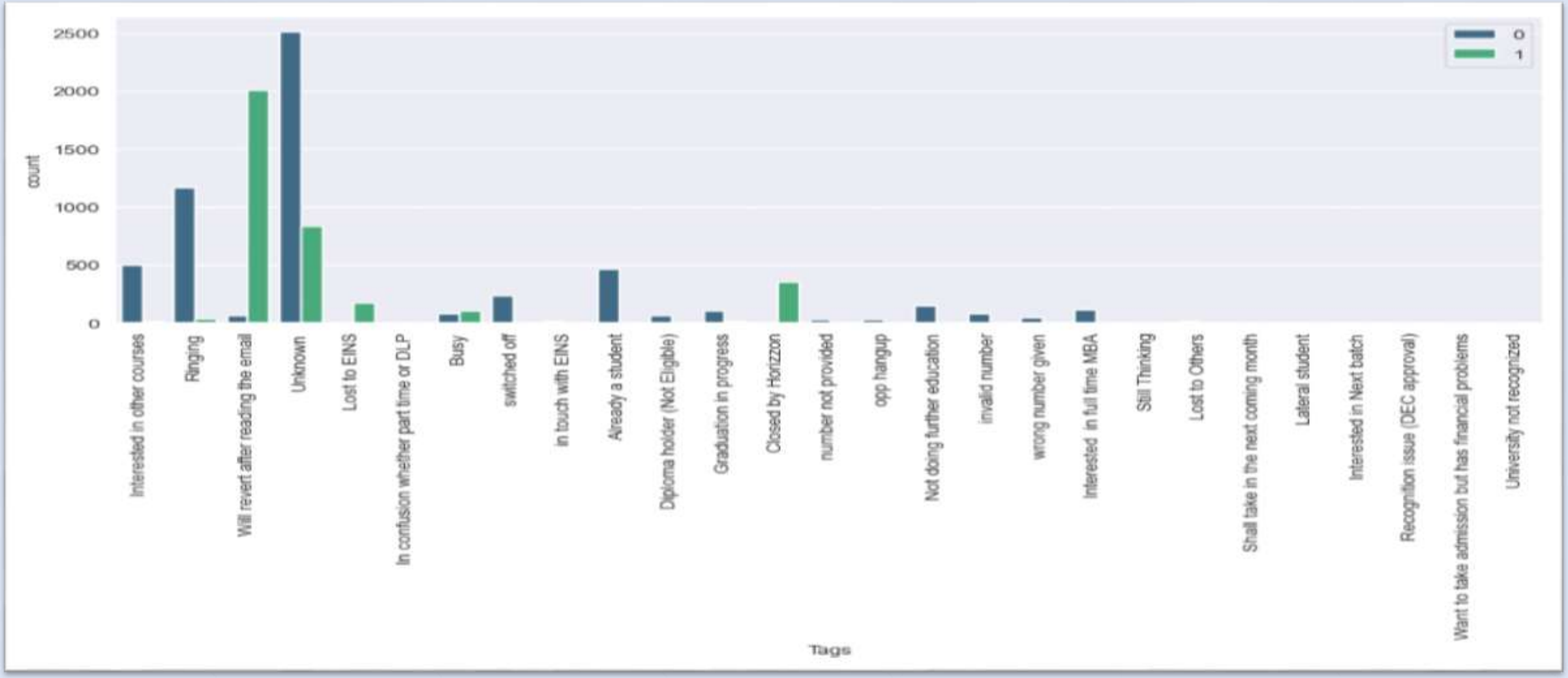
Analysis by occupation: As you can see, the conversion rate is high for unemployed profiles.



Below are some reasons for choosing online courses

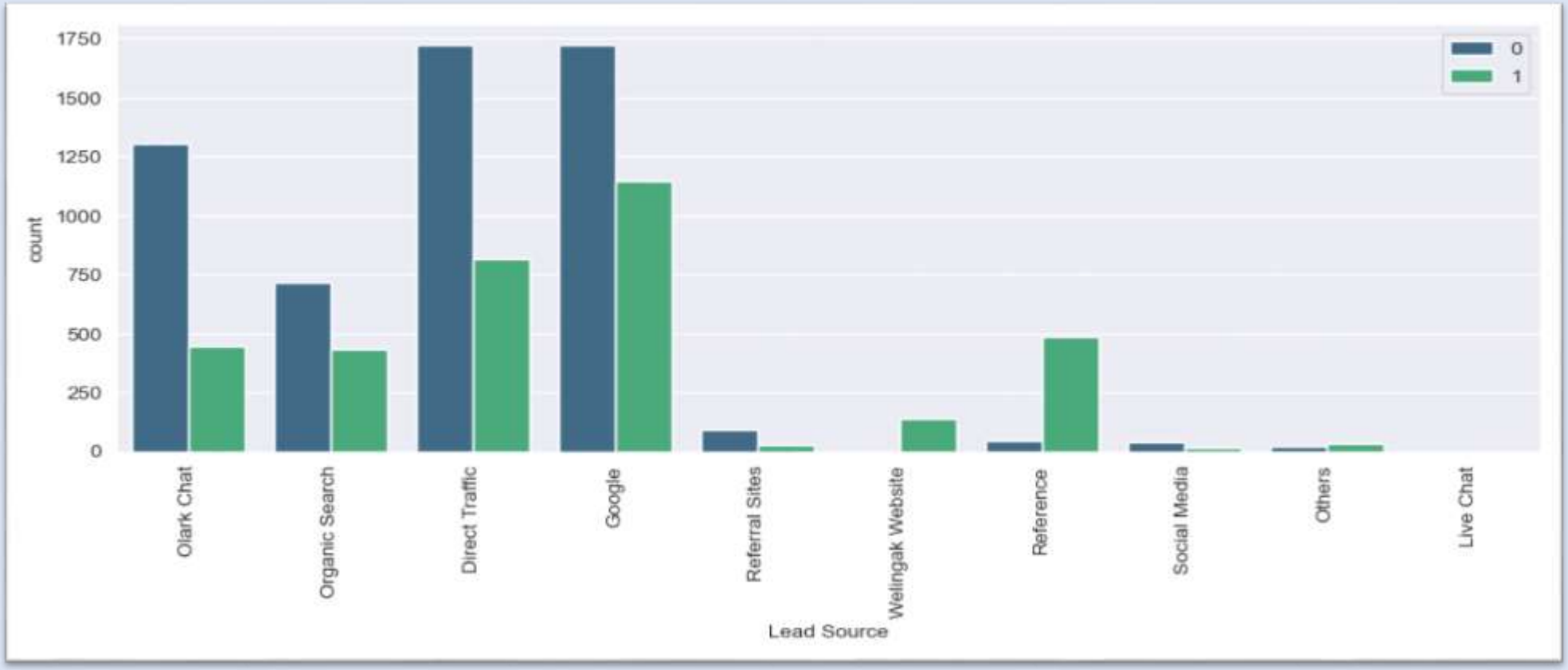


1. Analysis based on tags revealed that some tags provided no useful insights.
2. To address this, we categorized those tags as "Other Tags."

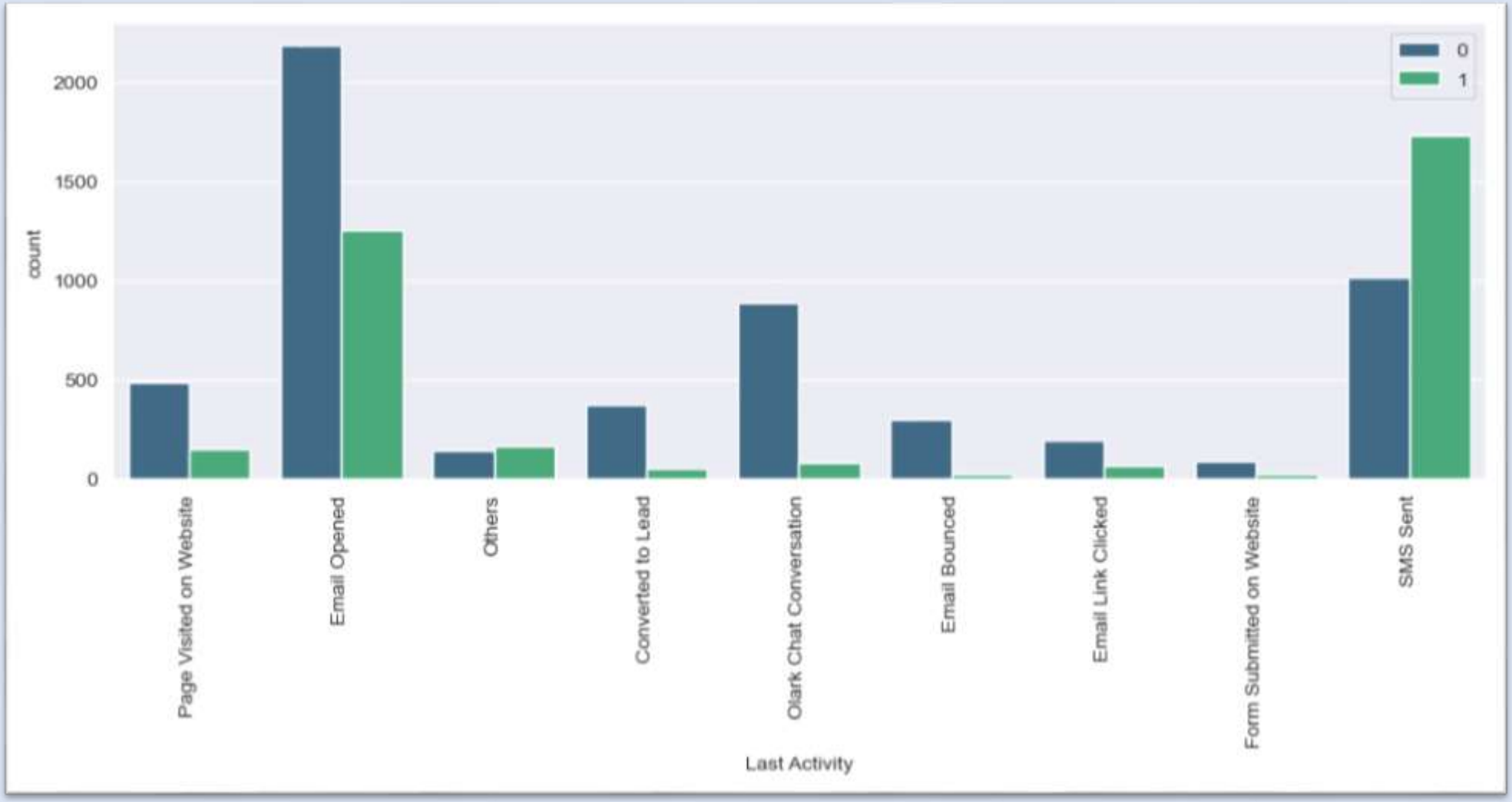


Analysis based on Lead source:

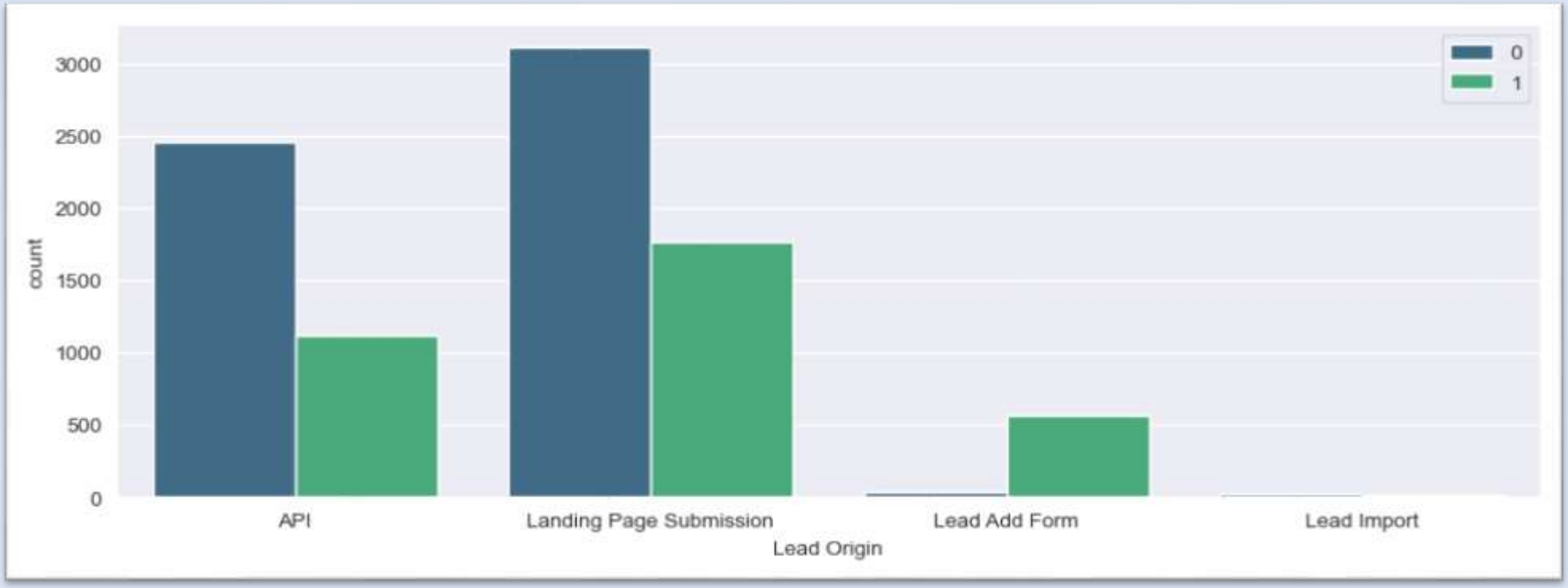
1. As you can see, the majority of leads come from Olark Chat, Google, Direct Traffic, and Organic Search.
2. Among these sources, Welingak and Reference have the highest conversion rates.
3. Based on this analysis, we recommend that businesses effectively utilize these channels to increase their lead generation and conversion rates.



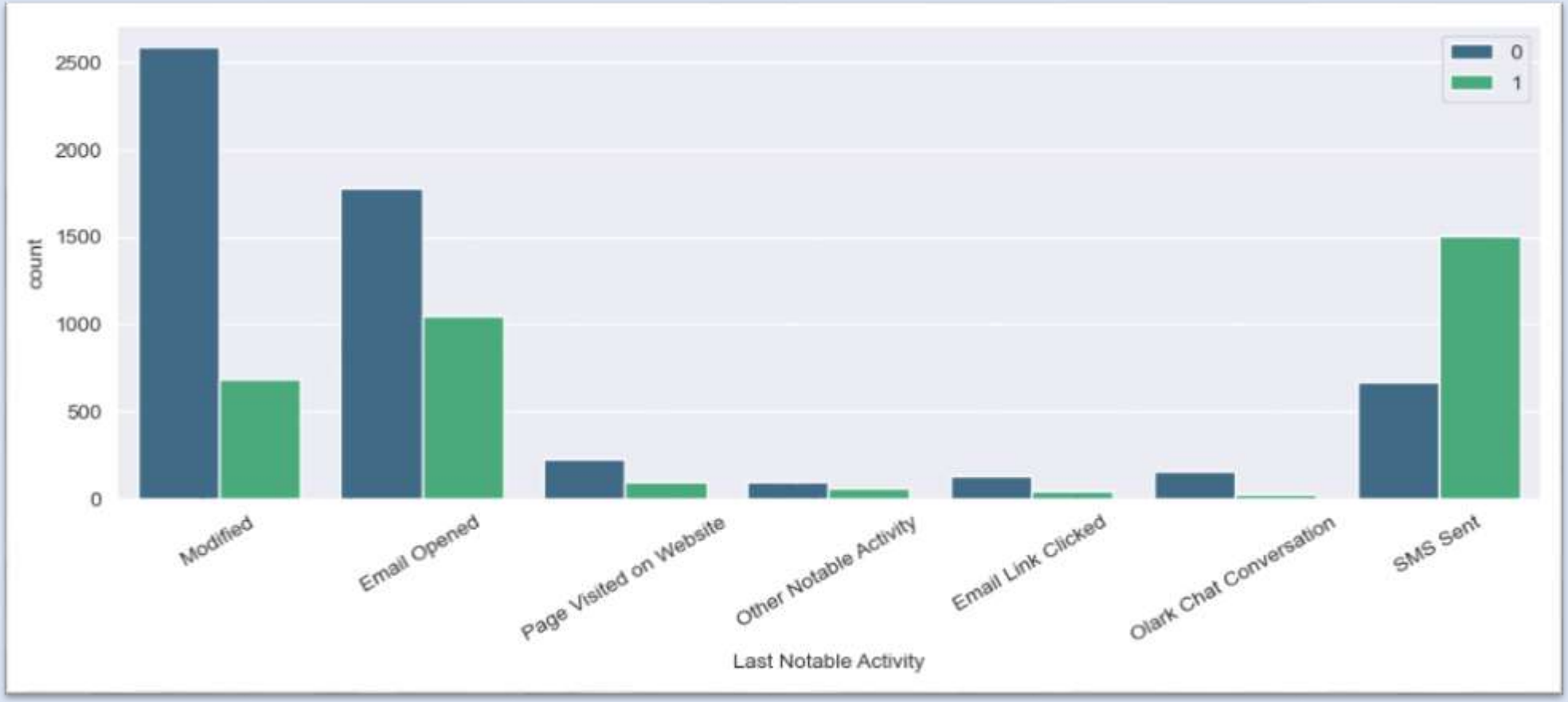
Based on our analysis of Last Activity, we found that the conversion rate for leads who received an SMS message is quite high compared to other activities. This suggests that businesses should consider incorporating SMS as a part of their lead generation and follow-up strategy



Based on our analysis of **Lead Origin**, we found that API and Landing Page submissions had the highest number of leads and conversion rates. However, we also found that the Lead Add Form had a high conversion rate, indicating a significant opportunity for generating more leads. We recommend leveraging the Lead Add Form and optimizing the lead generation process for maximum conversion rates

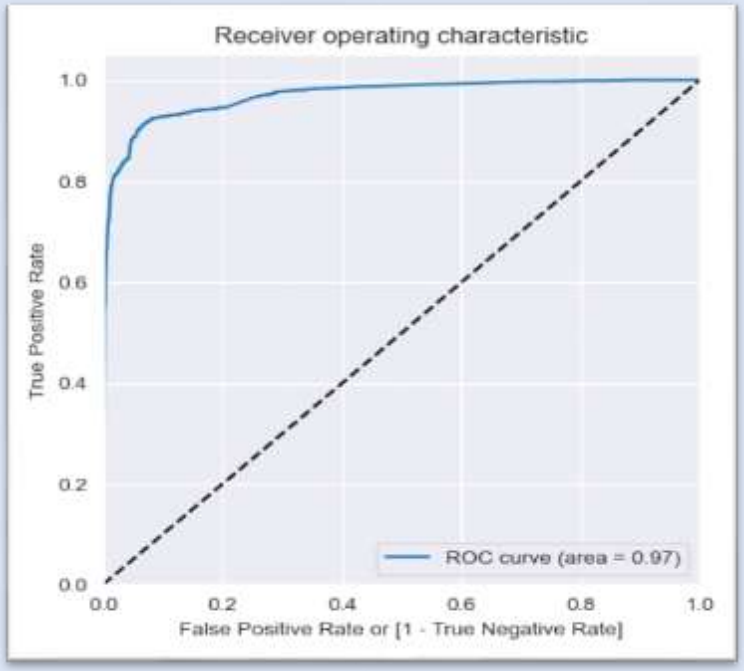


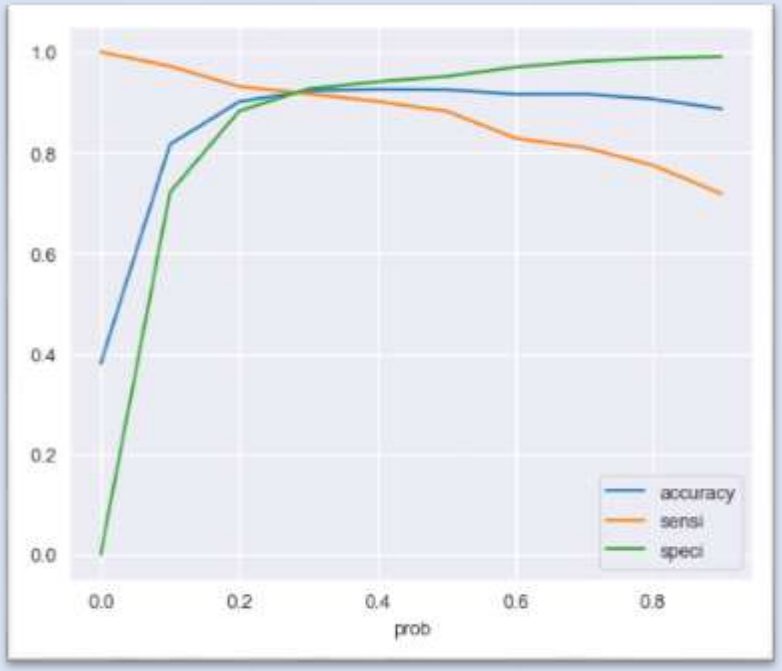
Based on our analysis of Last Notable Activity, we found that SMS Sends had the highest conversion rate among all activities.
This highlights the importance of incorporating SMS into the lead generation and follow-up process.
While Modified had a high number of leads, it did not result in a significantly higher conversion rate.
Email Opened also had a good conversion rate and should be considered as an effective communication channel for lead engagement and follow-up.

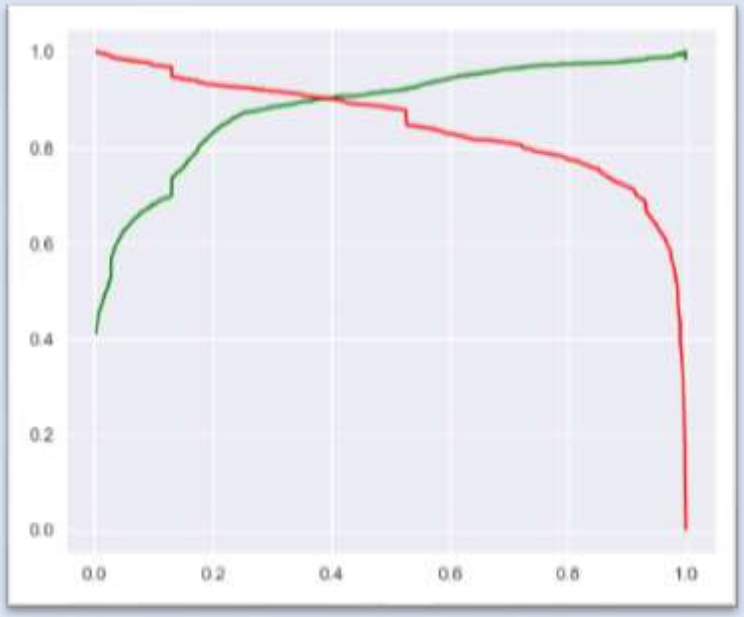


1. After conducting our analysis, we identified several variables that were significant predictors of lead conversion.
2. By focusing on these variables, businesses can increase the probability of lead conversion and optimize their lead generation and follow-up strategies.

```
------------------------------------------------
const
Total Time Spent on Website
Lead Origin_Lead Add Form
Lead Source_Direct Traffic
Lead Source_Welingak Website
Last Activity_SMS Sent
Last Notable Activity_Modified
Last Notable Activity_Olark Chat Conversation
Tags_Closed by Horizzon
Tags_Interested in other courses
Tags_Lost to EINS
Tags_Other Tags
Tags_Ringing
Tags_Will revert after reading the email
================================================
```







Metrics on train dataset

1. Accuracy on new model: 92.29%
2. Sensitivity : 91.70%
3. Specificity : 92.66%
4. Precision Rate: 88.47
5. Recall Rate: 91.7

1. Our predictive model had a good ROC curve value of 0.97, indicating a strong level of predictive accuracy.

2. After calculating accuracy, sensitivity, and specificity for various probability cutoffs, we found that 0.3 was the optimal cutoff probability for predicting lead conversion.

3. This means that leads with a probability of 0.3 or higher have a higher likelihood of conversion, and businesses should prioritize their lead follow-up efforts accordingly.

4. By using this cutoff probability and our predictive model, businesses can increase the efficiency and effectiveness of their lead generation and follow-up strategies.

Final Observation:

Let us compare the values obtained for Train & Test:

Train Data:

- Accuracy : 92.29%
- Sensitivity : 91.70%
- Specificity : 92.66%

Test Data:

- Accuracy : 92.78%
- Sensitivity : 91.98%
- Specificity : 93.26%

- ✓ Our model has demonstrated a high level of accuracy in predicting lead conversion, and we can confidently say that it provides an effective tool for decision-making. With a strong ROC curve value of 0.97 and an optimal cutoff probability of 0.3, the model is able to identify leads with a high likelihood of conversion, giving businesses the ability to focus their resources on the most promising leads. By leveraging the insights from our model, the CEO can make data-driven decisions that are grounded in statistical analysis, ultimately leading to a more effective and efficient lead generation and follow-up strategy

Thank You !!!