



# Case Study Leads Scoring

Sambhav Jain

# Agenda

---

01 About Business

02 Problem Statement

03 Goals of Case Study

04 EDA Insights

05 Logistic Regression Model

06 Recommendations



# About Business

## Who are we?

We are X Education who provides educational courses for upskilling

## Who are the target customers?

X Education caters to the needs of industrial professionals

## How do we acquire customers

- Marketing courses through Google, referral networks, etc
- Calling the leads generated to convert them

# Problem Statement

---

Our Lead conversion rate is

30%

---

Although loads of leads are generated, but a small fraction is converted into business



# Goals of Case Study



## Build logistic regression model and assign scores from 0-100

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



## Identify variables that affect the conversion

Building logistic regression model to identify variables that affect the chances of conversion of a lead



## Conversion rate of 80%

The target lead conversion rate that the company wants to achieve is 80% or above

# EDA Insights

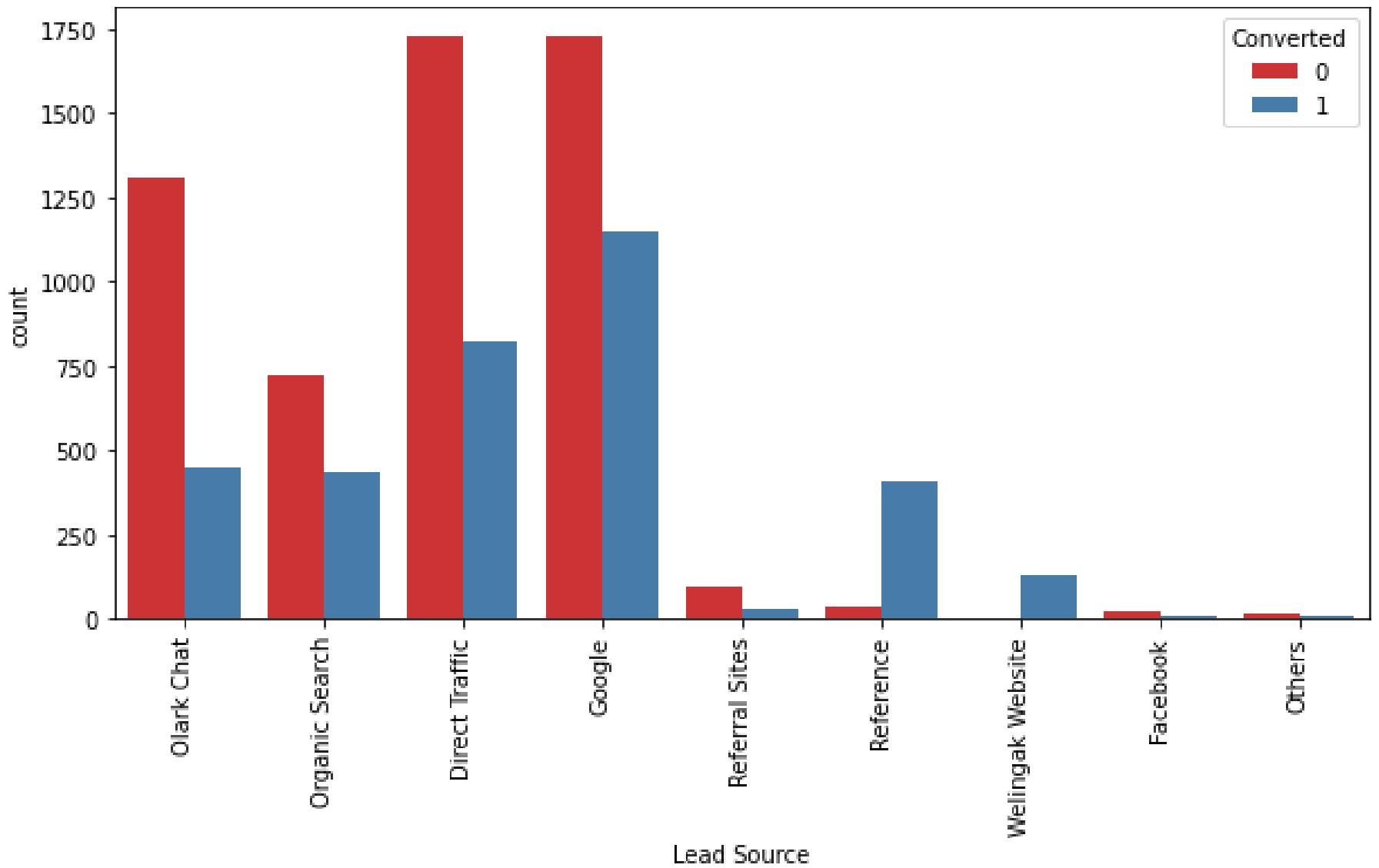
---

To understand impact of each variable on conversion rate, each variable has been plotted against converted tag and draw insights on the importance of variable for the model

---

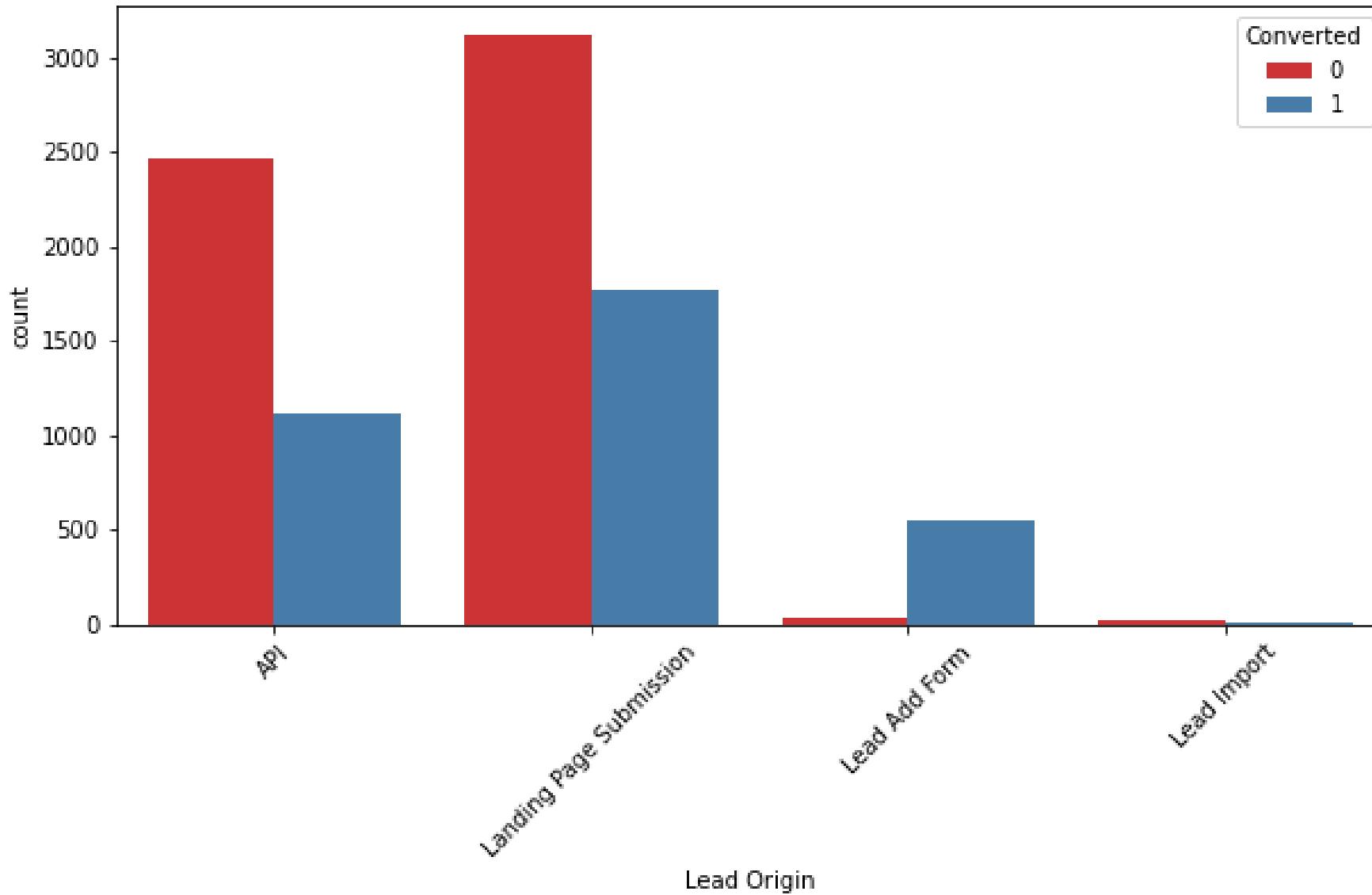


# Lead Source Insights



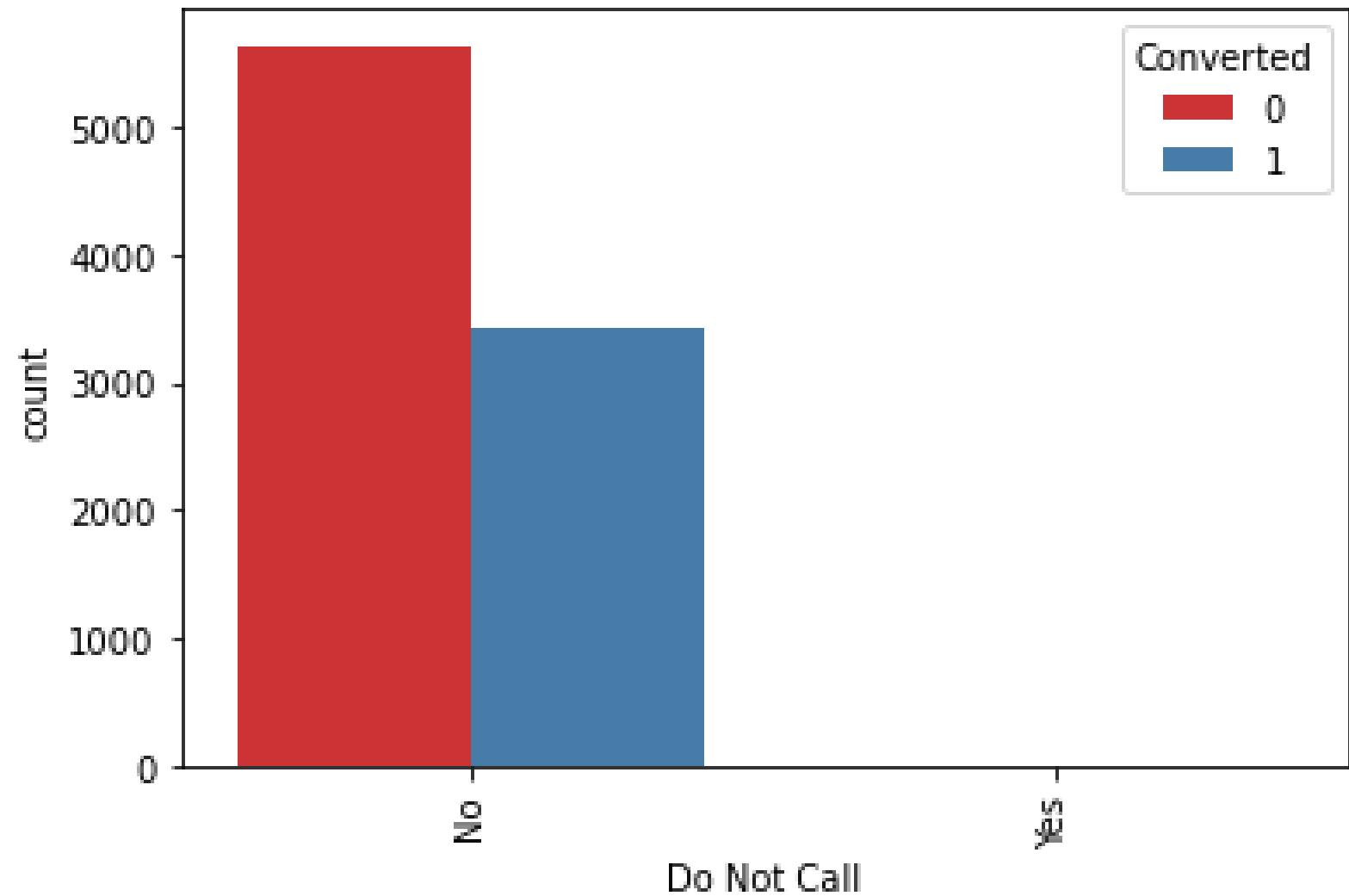
- Google and Direct traffic generates maximum number of leads
- Conversion Rate of reference leads and leads through welingak website is high

# Lead Origin Insights



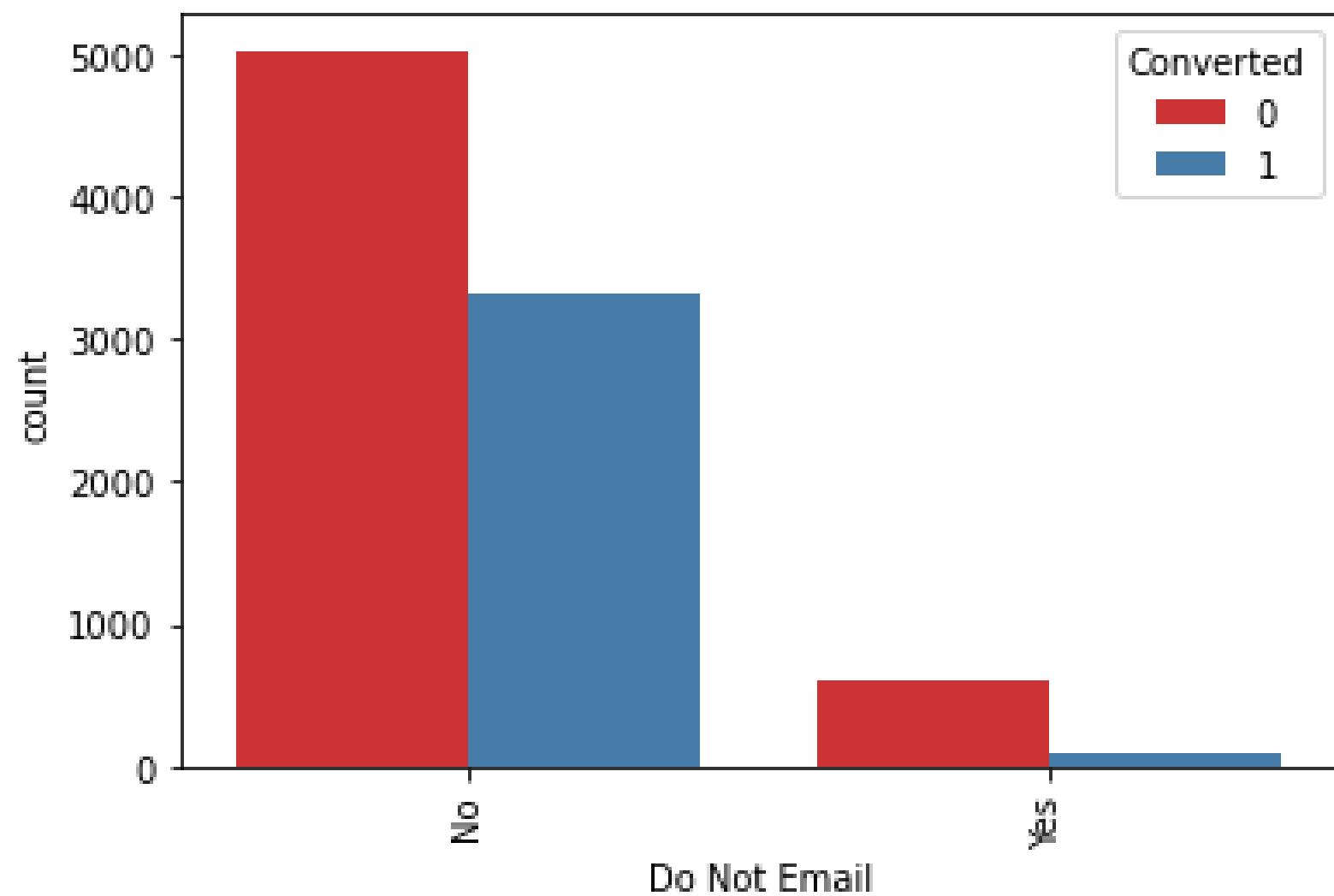
- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable
- Lead Add Form has more than 90% conversion rate but count of lead are not very high
- Lead Import are very less in count

# Do Not Calls Insights



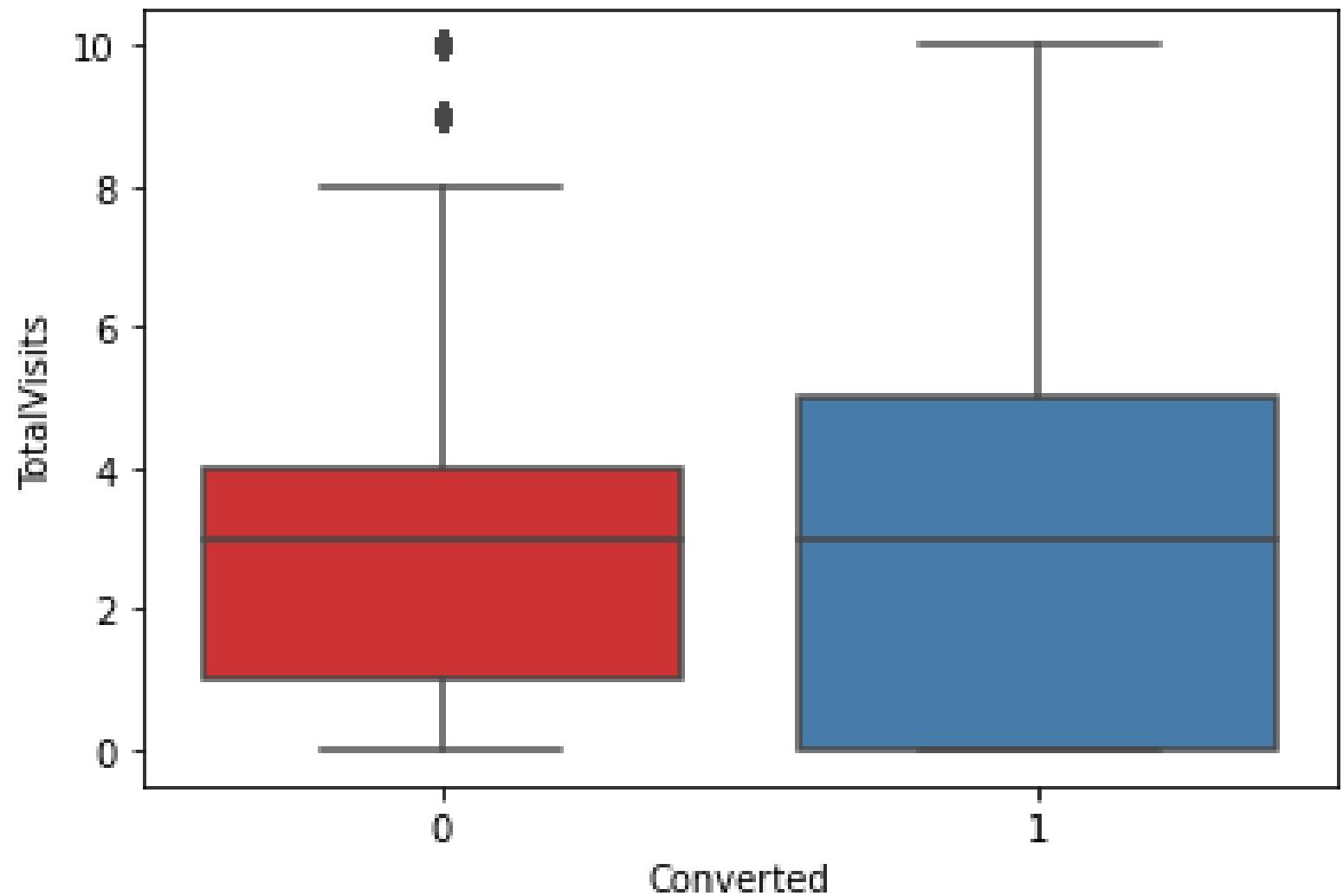
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Do Not Email Insights



Since most of the entries are No, no  
inference can be drawn with this  
parameter

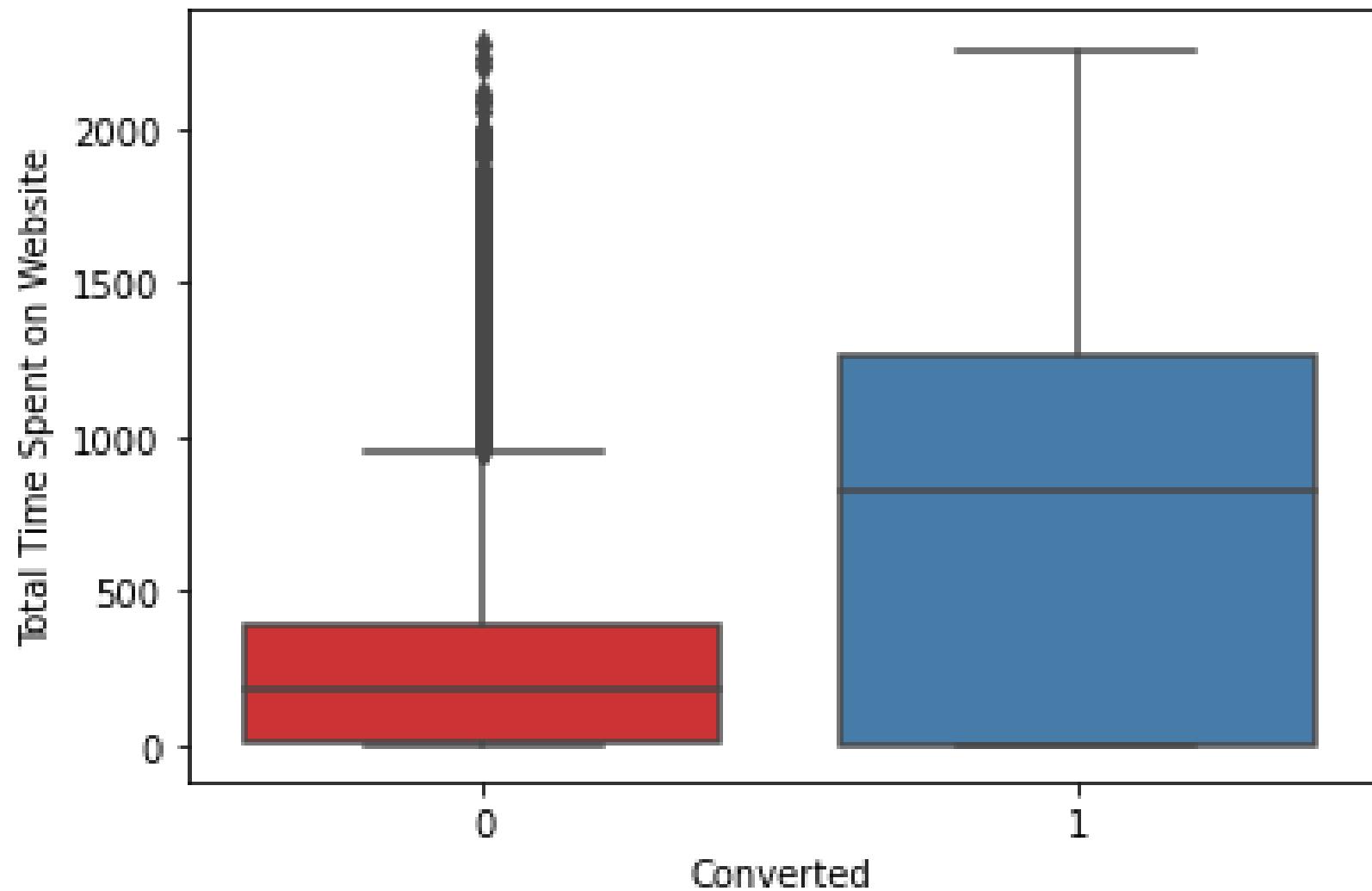
# Total Visits Insights



Median for converted and not converted  
leads are the same

# Total Time on Websites

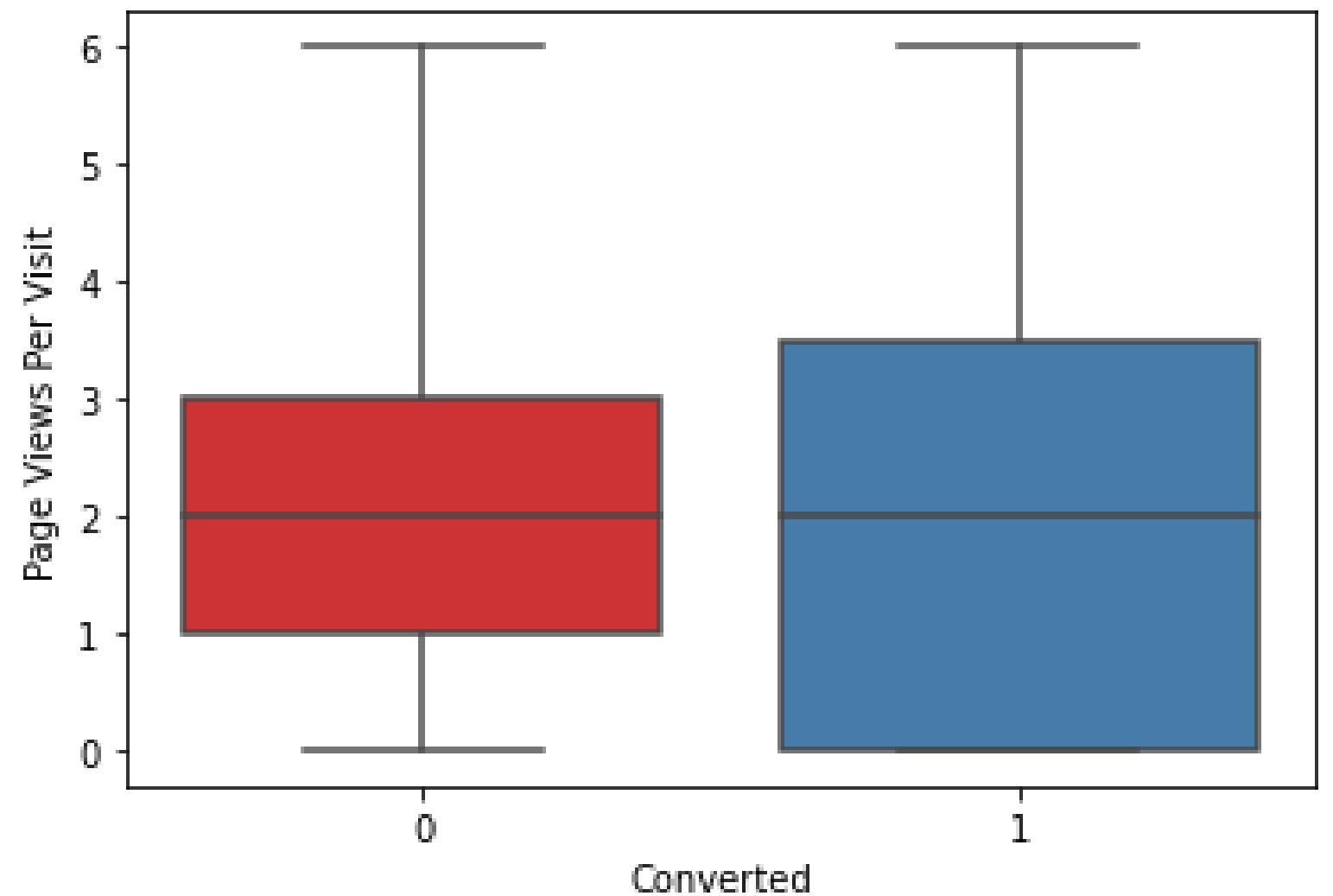
## Insights



Leads spending more time on the website  
are more likely to be converted

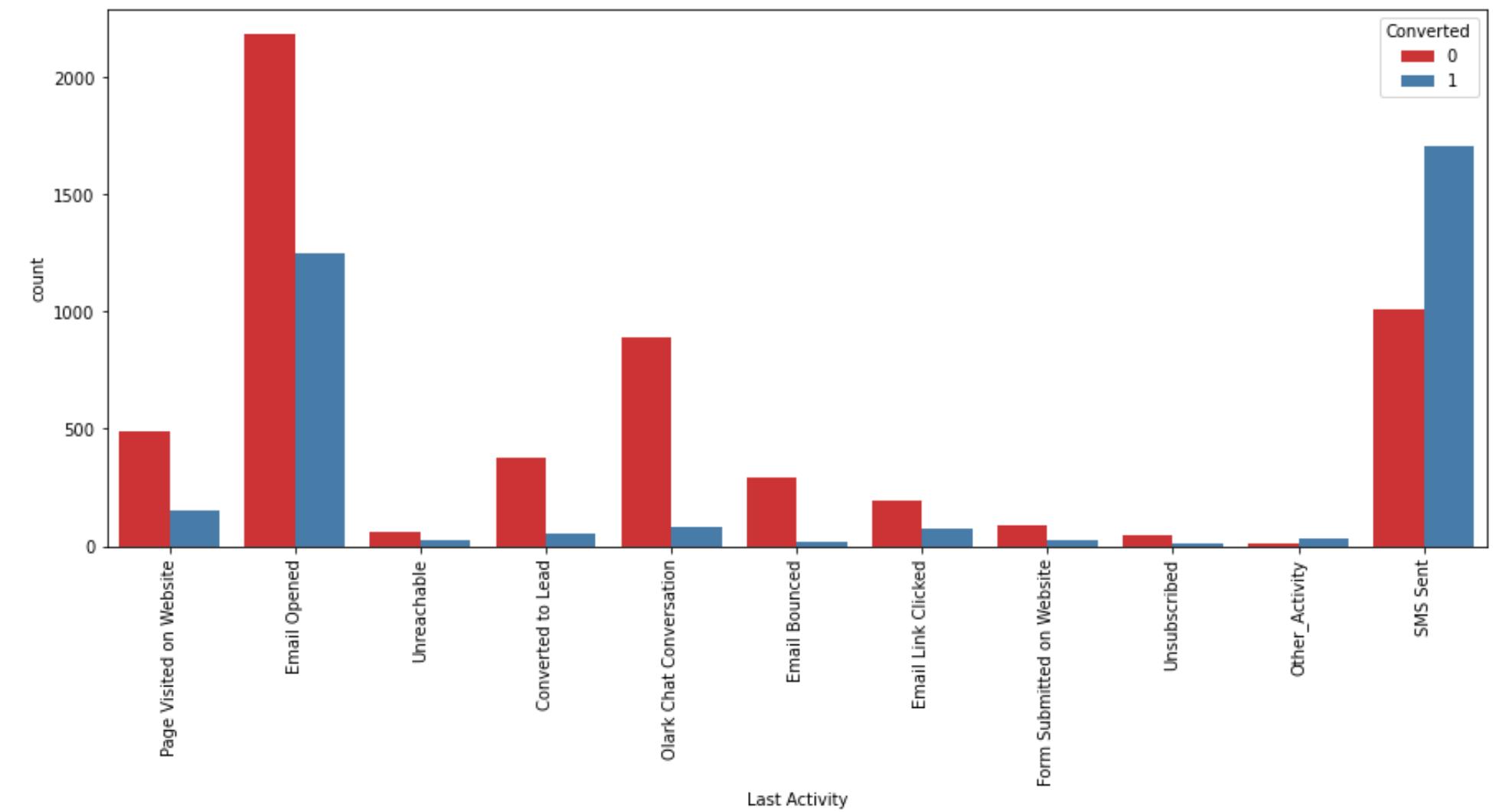
# Page views per Visit

## Insights



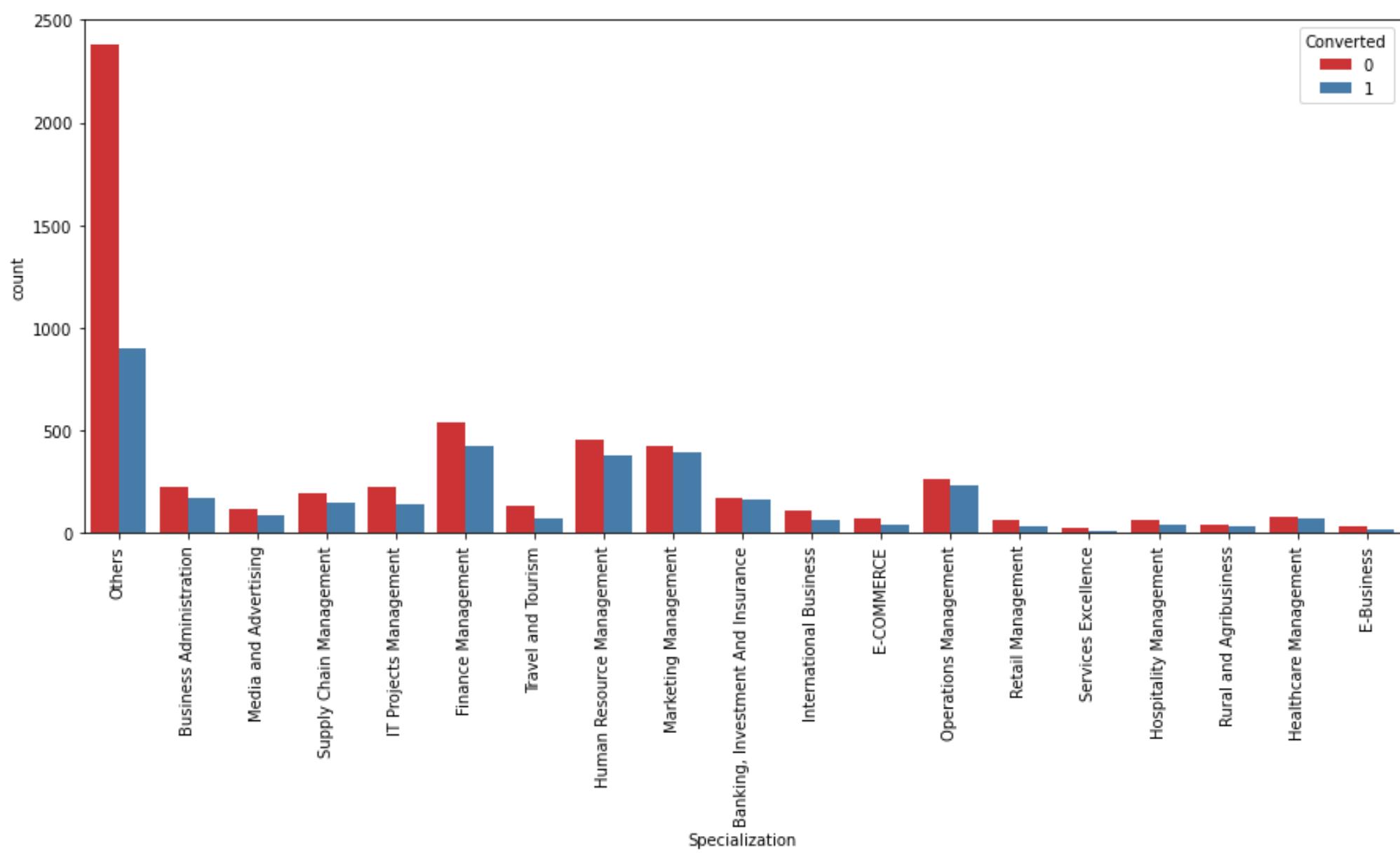
Median for converted and unconverted leads is the same

# Last Activity Insights



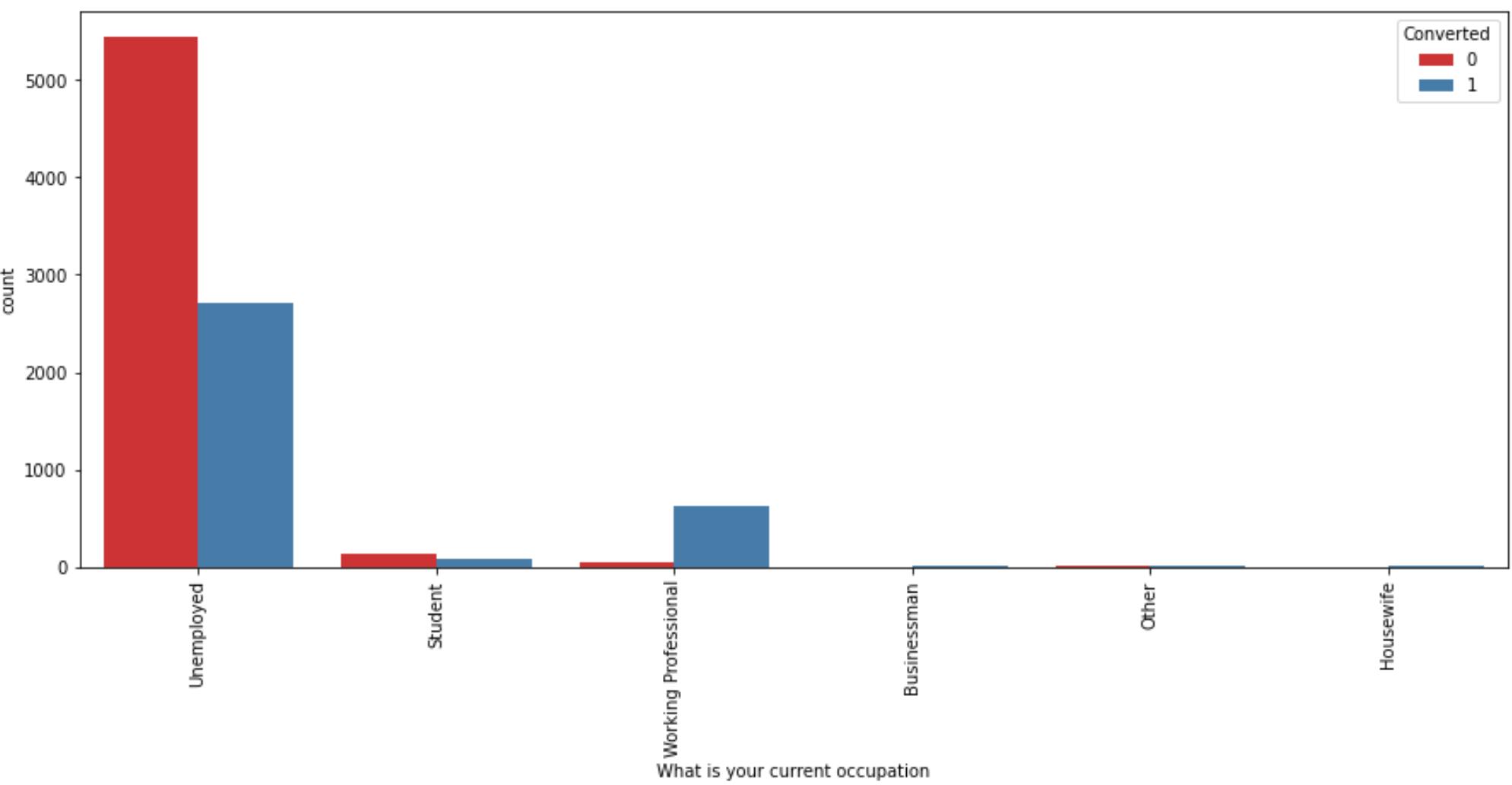
- Most of the lead have their Email opened as their last activity
- Conversion rate for leads with last activity as SMS Sent is almost 60%

# Specialisation Insights



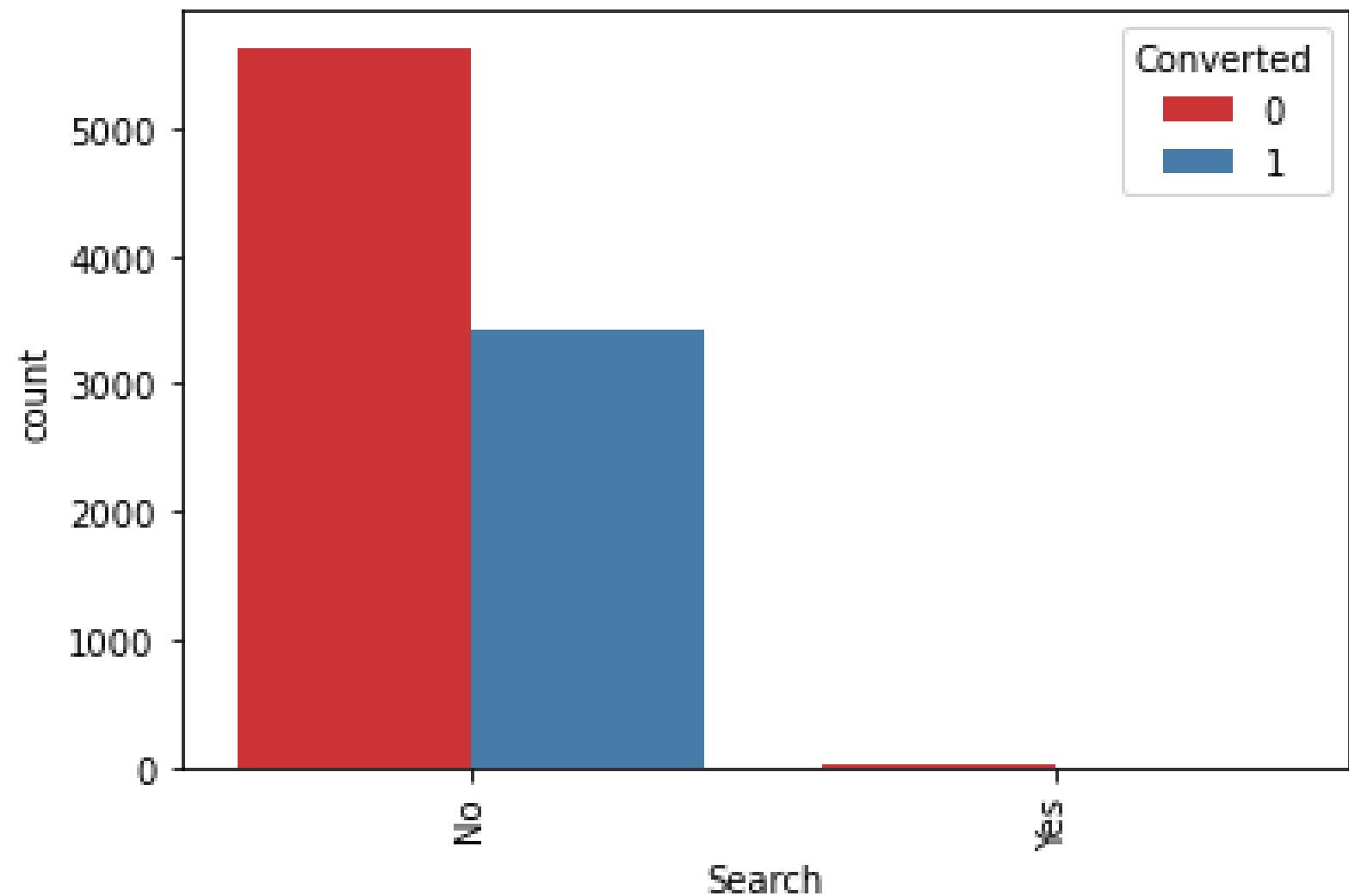
Baring others, Finance Management, HR Management, Marketing Management and Operations Management are specialisation with good conversion rate

# Current Occupation Insights



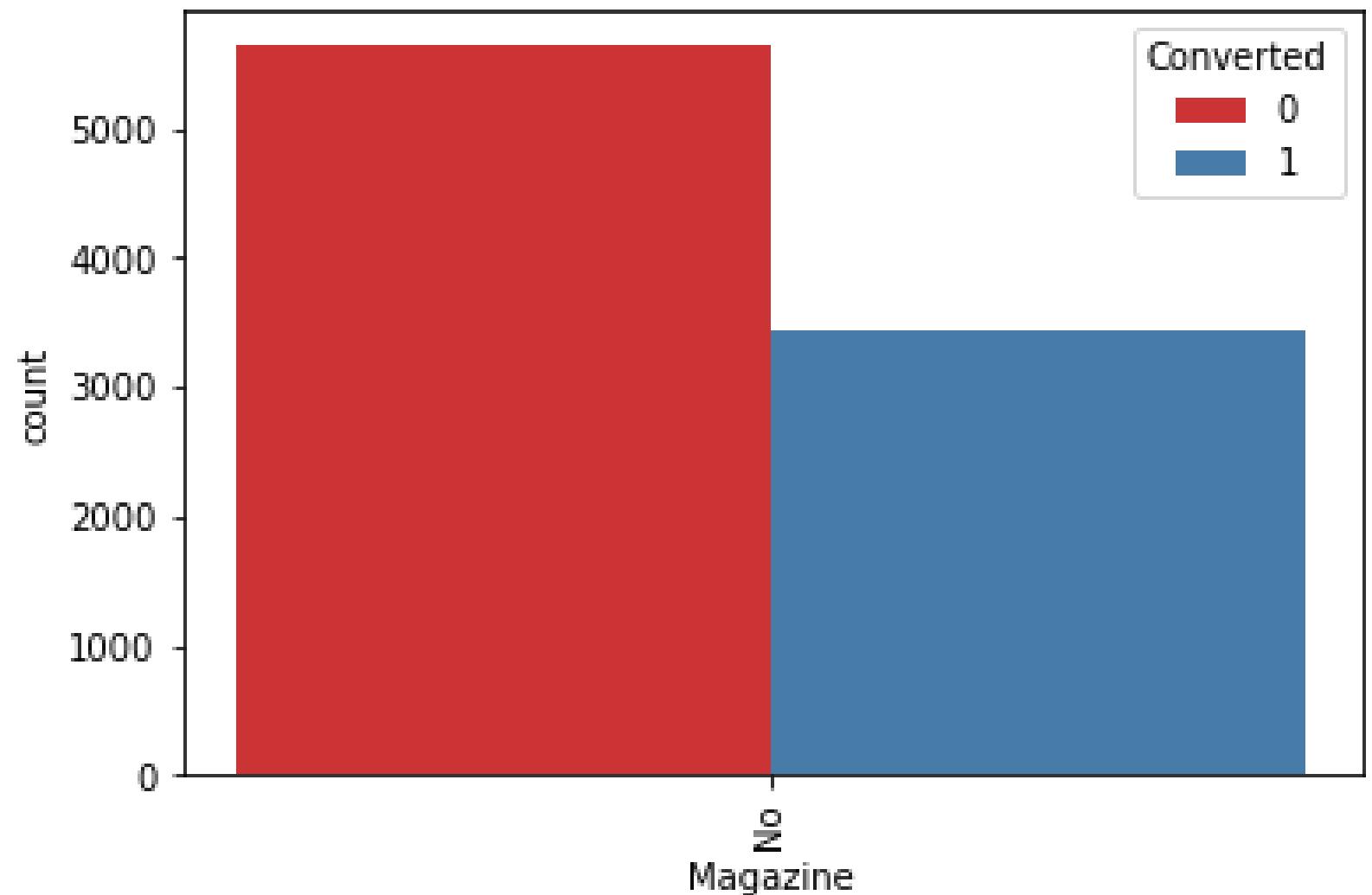
- Working Professionals going for the course have high chances of joining it
- Unemployed leads are the most in numbers but has around 30-35% conversion rate

# Search Insights



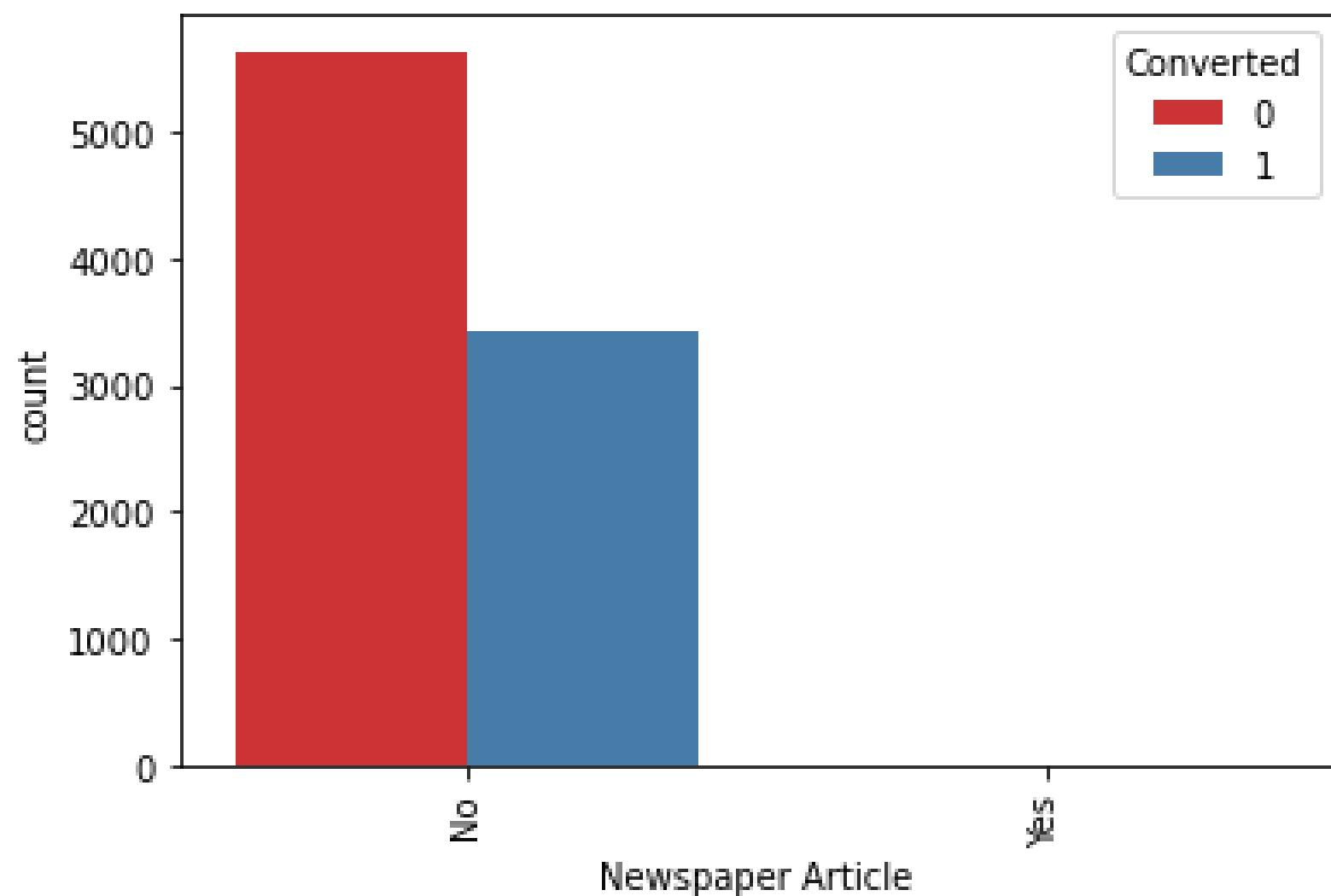
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Magazine Insights



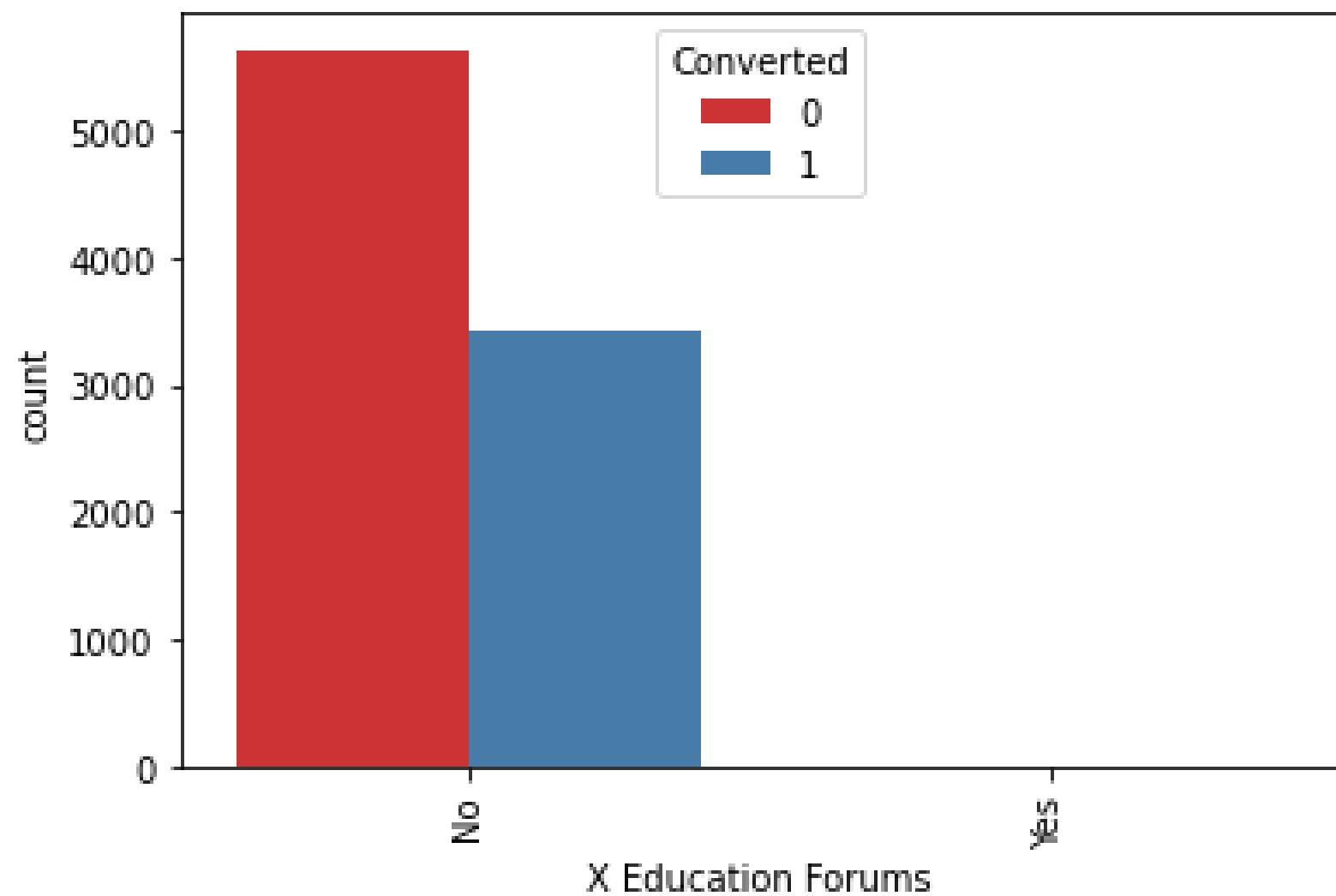
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Newspaper Article Insights



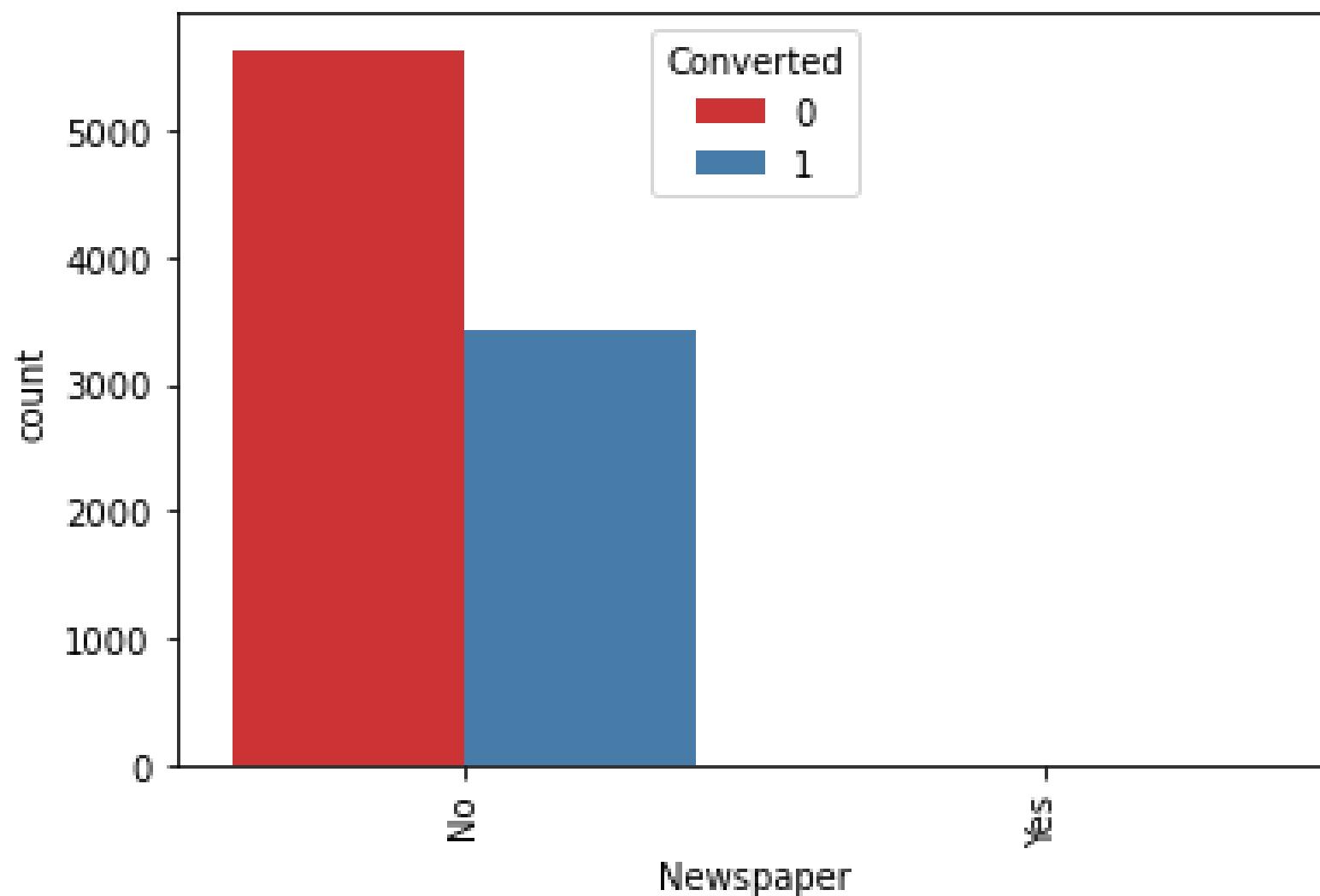
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# X Education Forum Insights



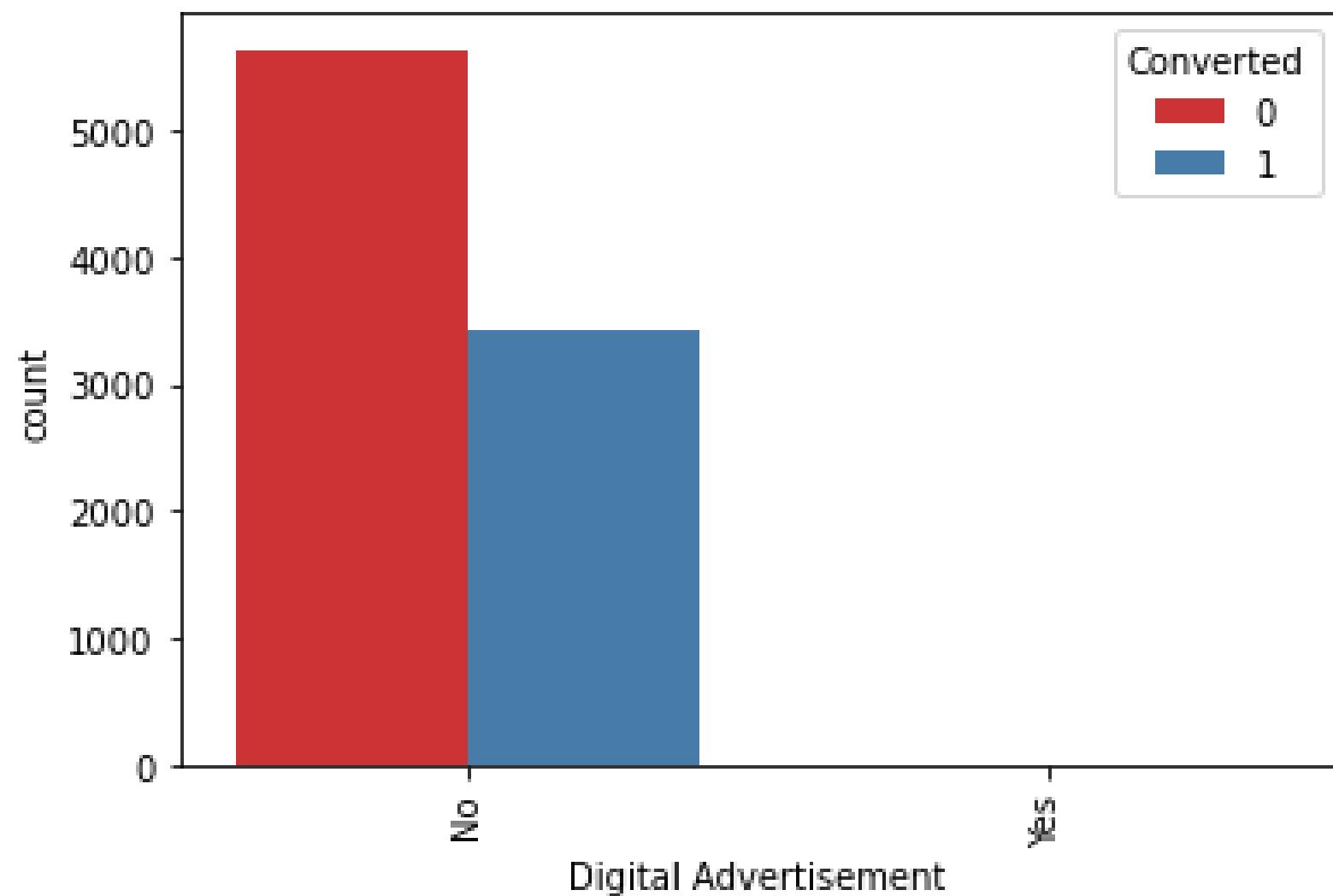
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Newspaper Insights



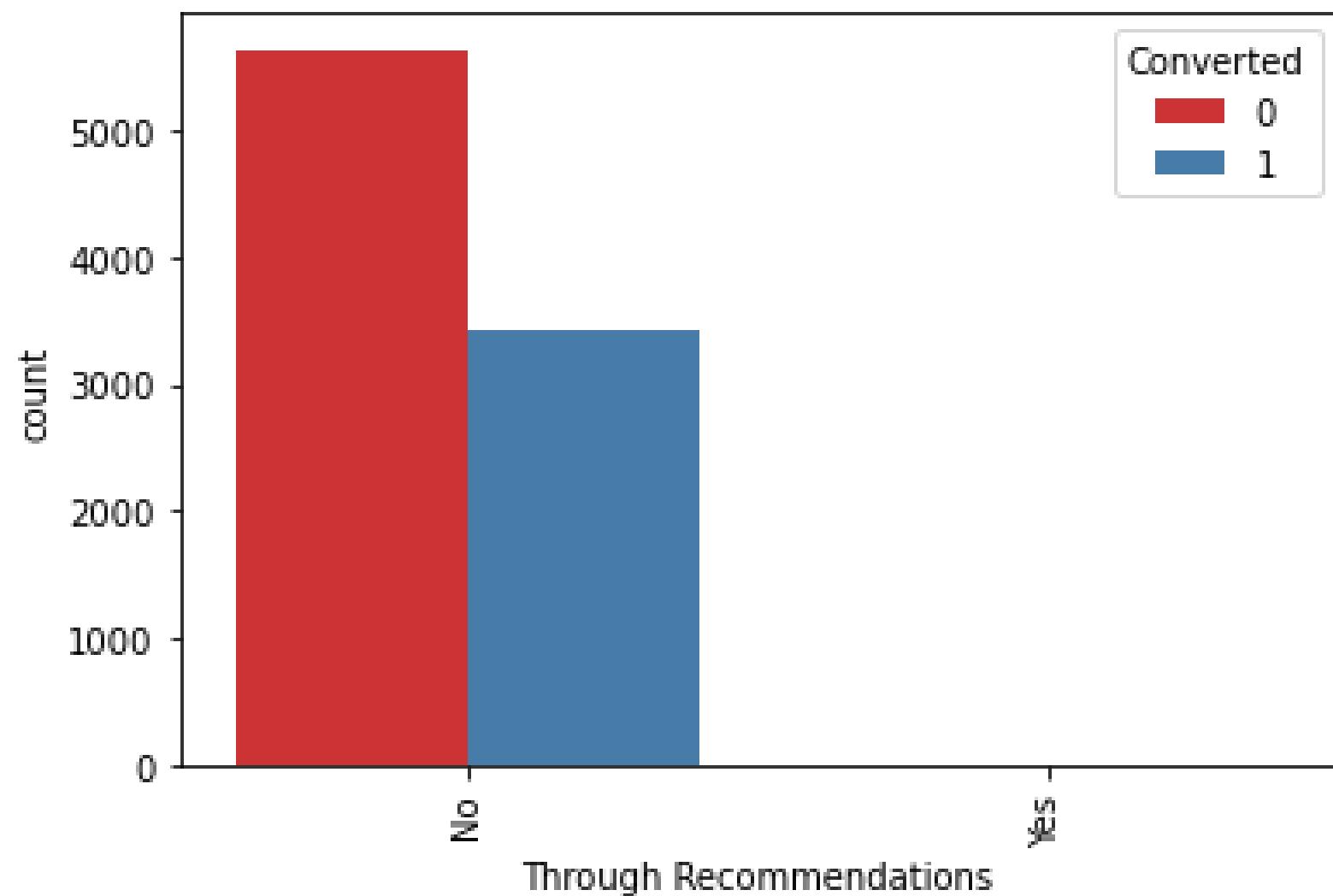
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Digital Advertisement Insights



Since most of the entries are No, no  
Inference can be drawn with this  
parameter

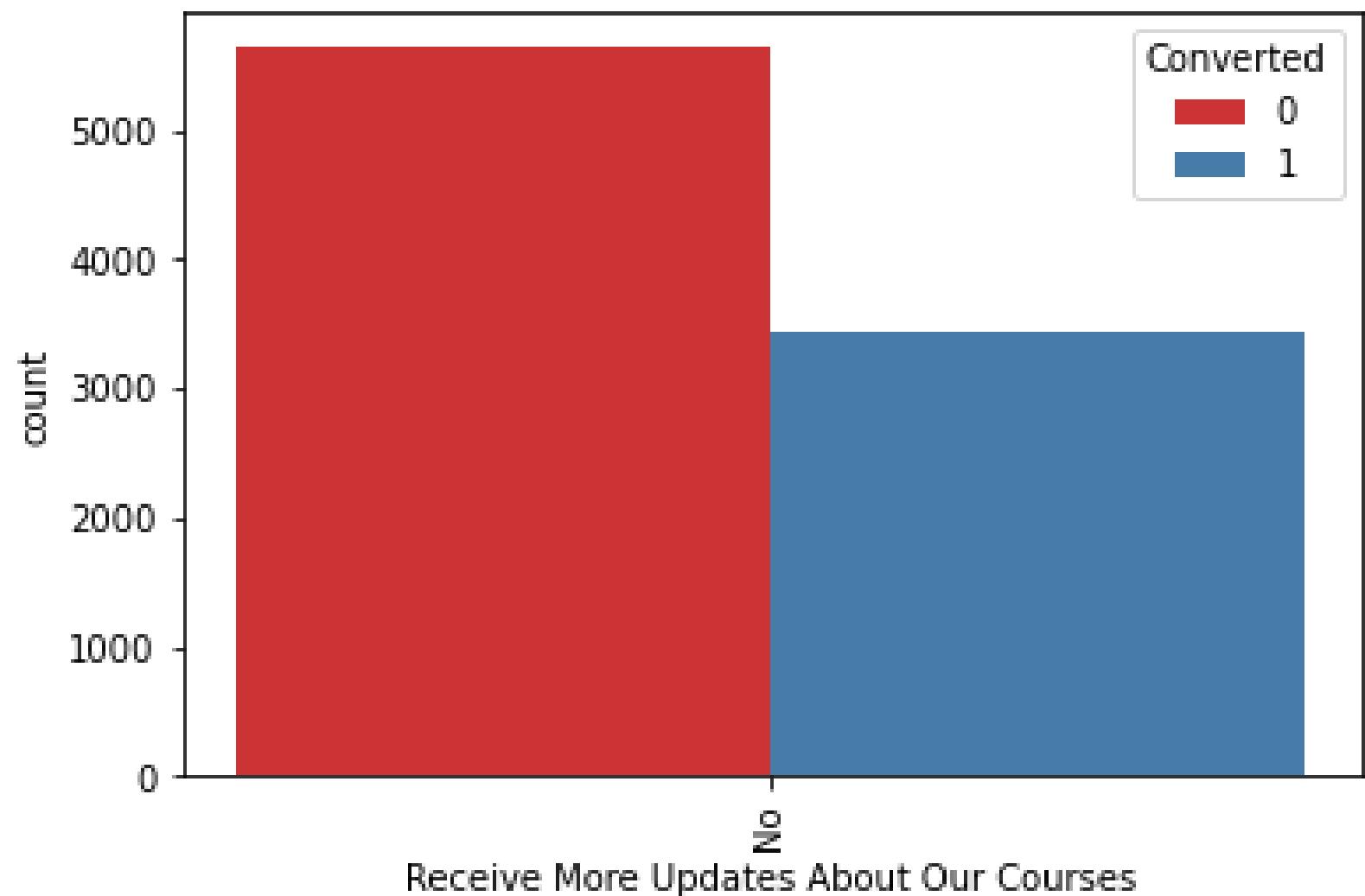
# Through Recommendations Insights



Since most of the entries are No, no  
Inference can be drawn with this  
parameter

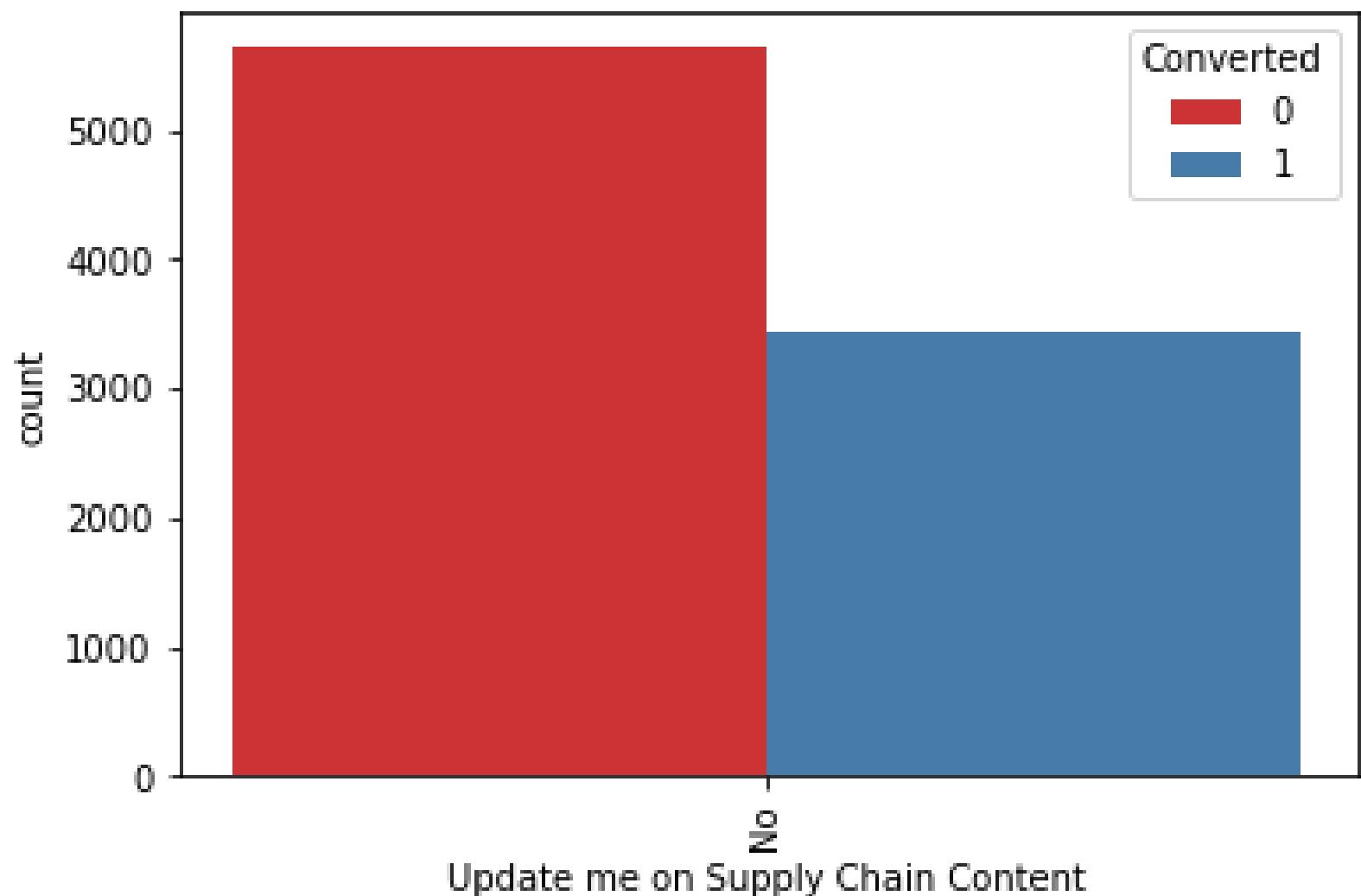
# Update on courses

## Insights



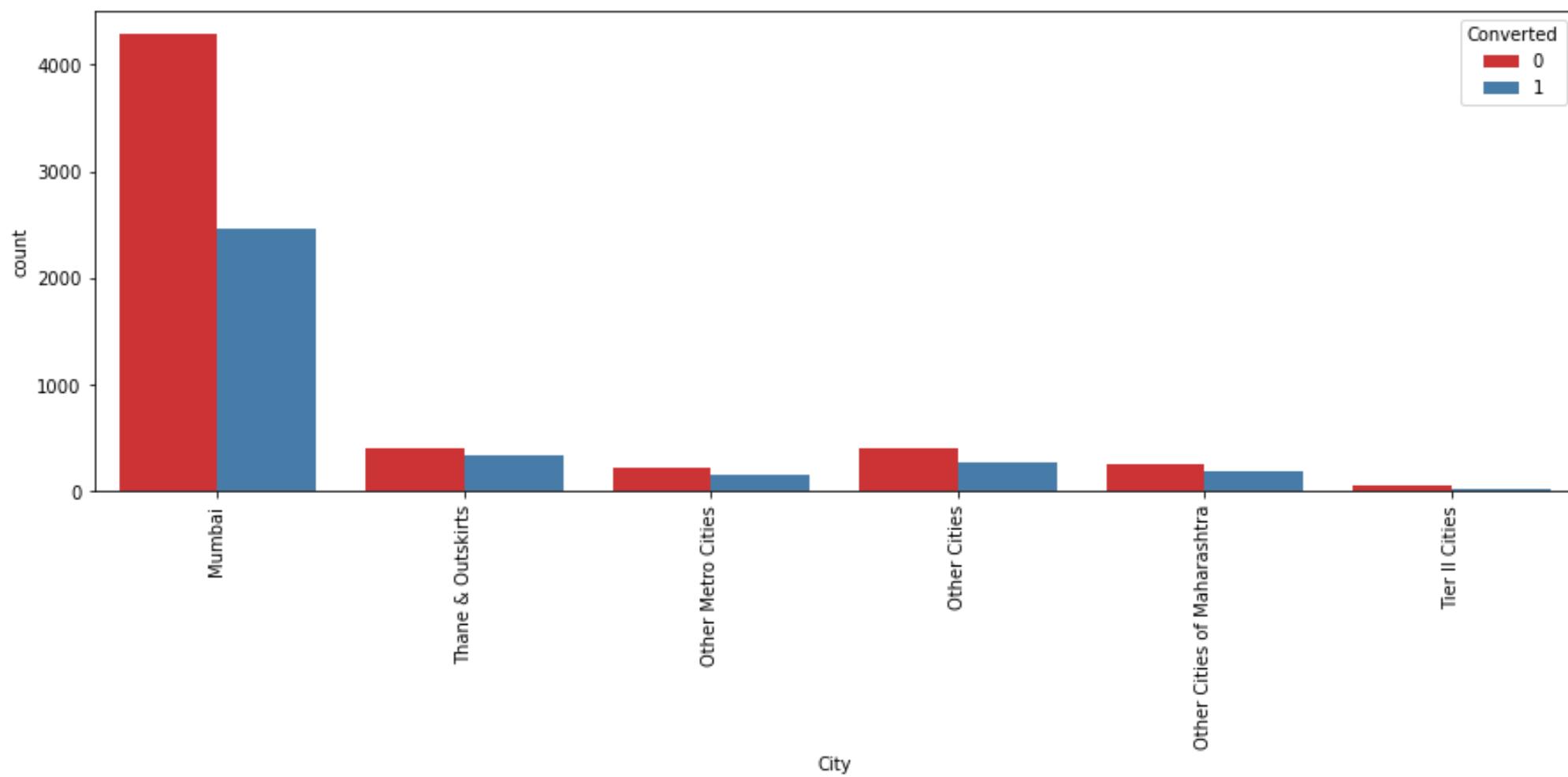
Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Update on SC Content Insights



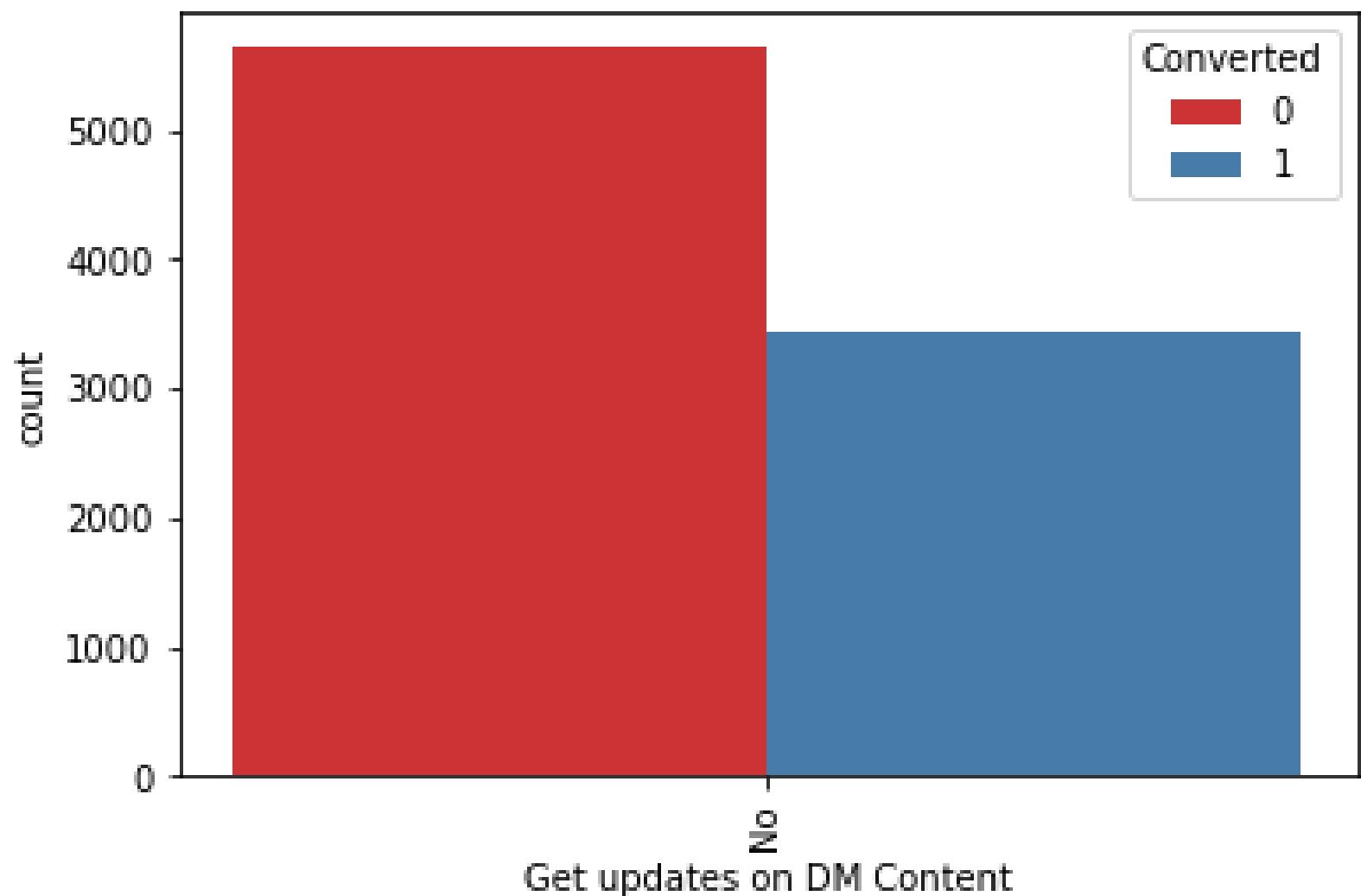
Since most of the entries are No, no inference can be drawn with this parameter

# City Insights



- Mumbai has the highest number of leads with around 38% conversion rate
- Thane and Outskirts along with other cities have a good conversion rate

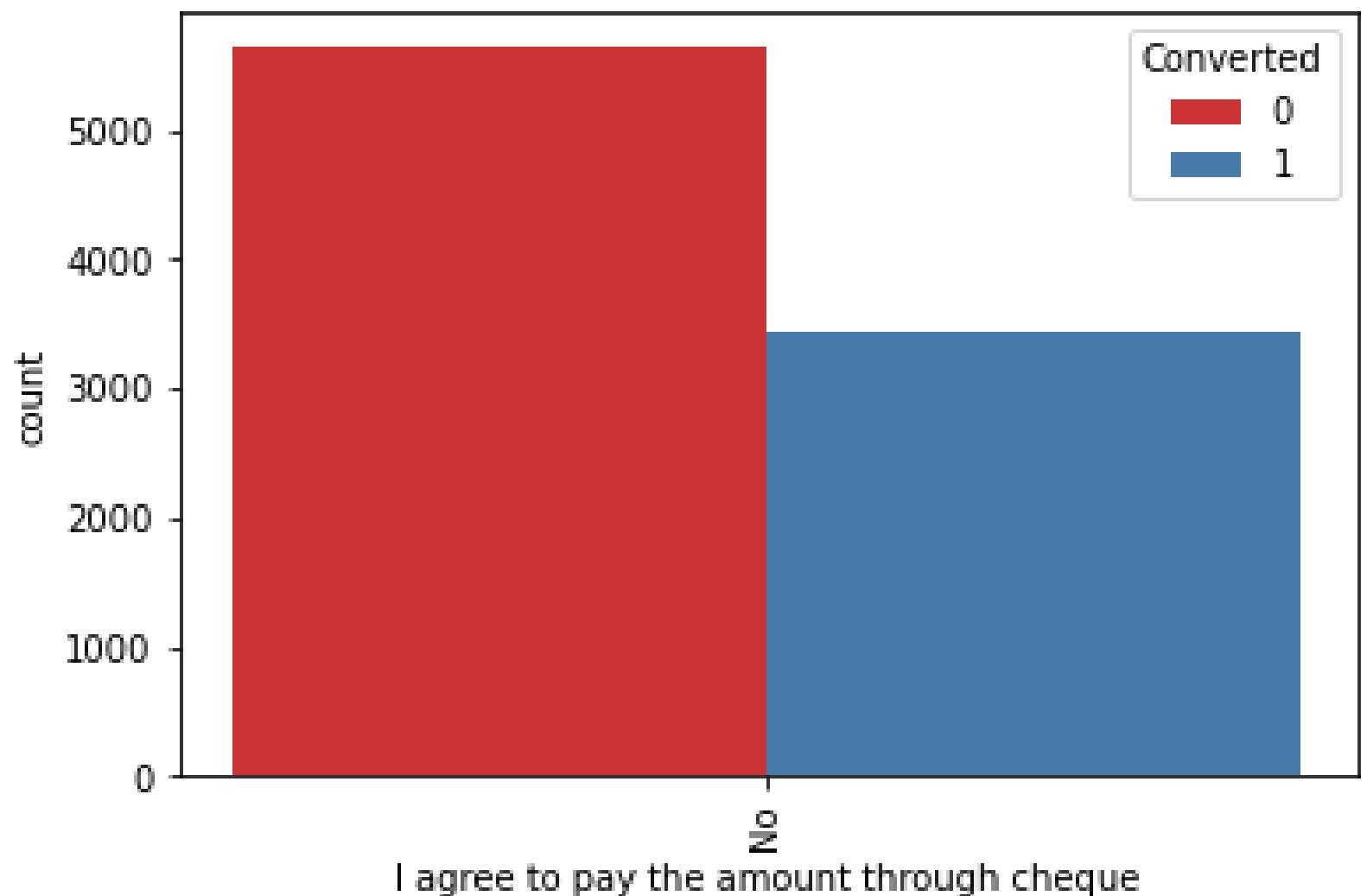
# Update on DM Content Insights



Since most of the entries are No, no  
Inference can be drawn with this  
parameter

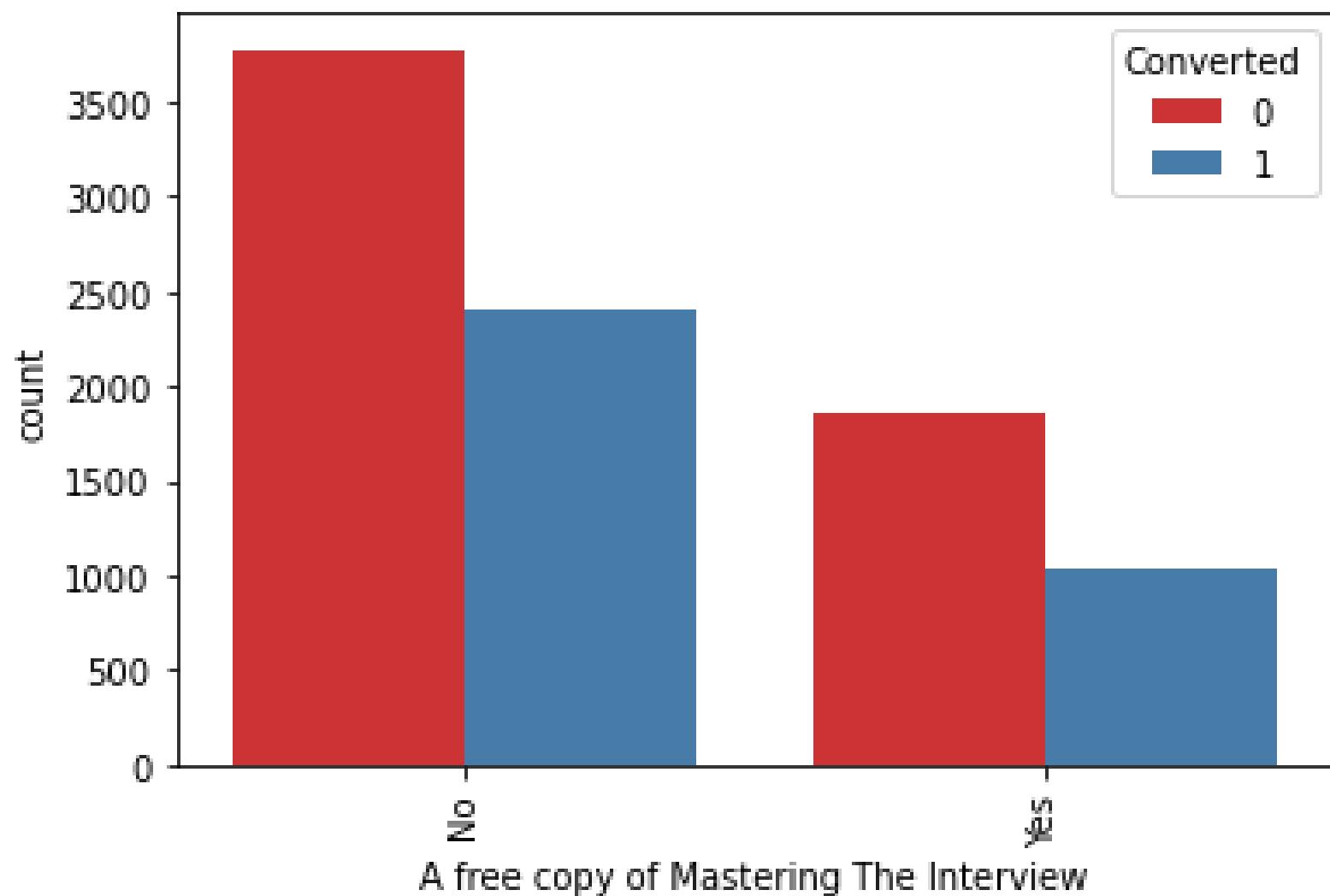
# Pay through cheque

## Insights



Since most of the entries are No, no  
Inference can be drawn with this  
parameter

# Master Interview Insights



Although the number of leads answering no to free copy is higher, there is no much difference in ratio of conversion between 2 categories

# Logistic Regression Model

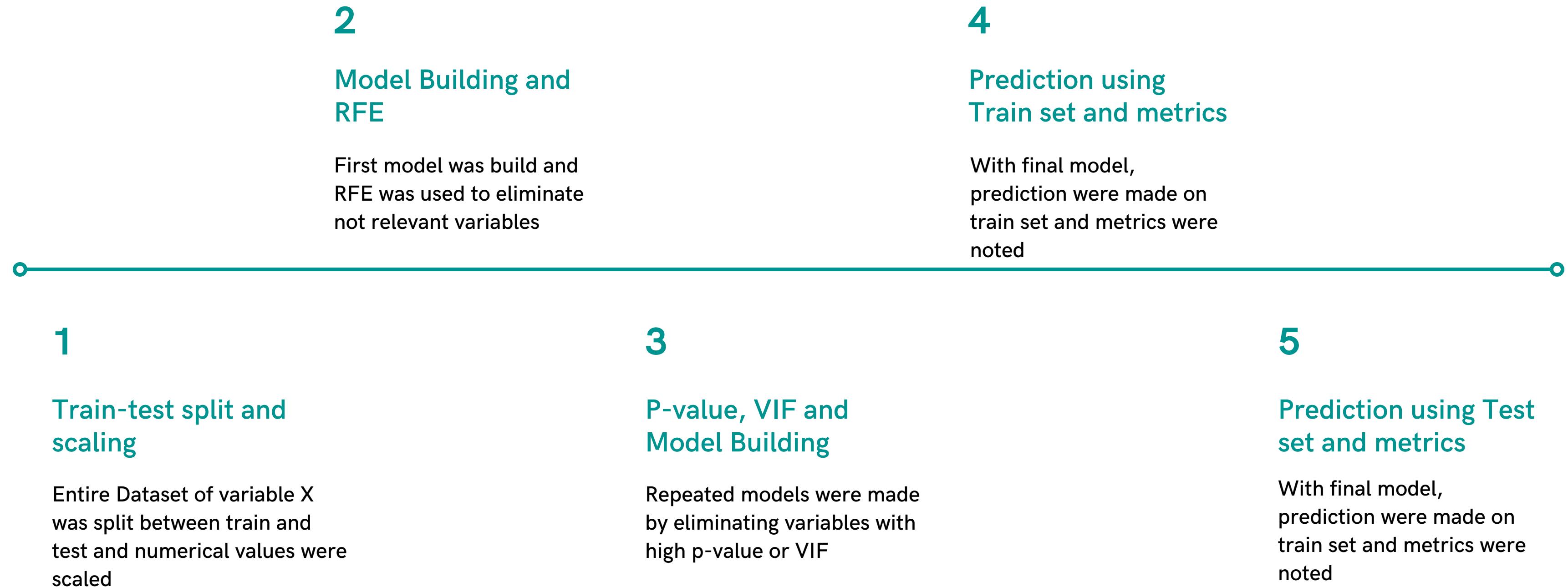
---

**Post EDA, variables that did not have any significant relation with conversion were dropped from dataset and variables other than converted column were treated as Independent Variable while converted column became dependent variable**

---



# Model Building Steps



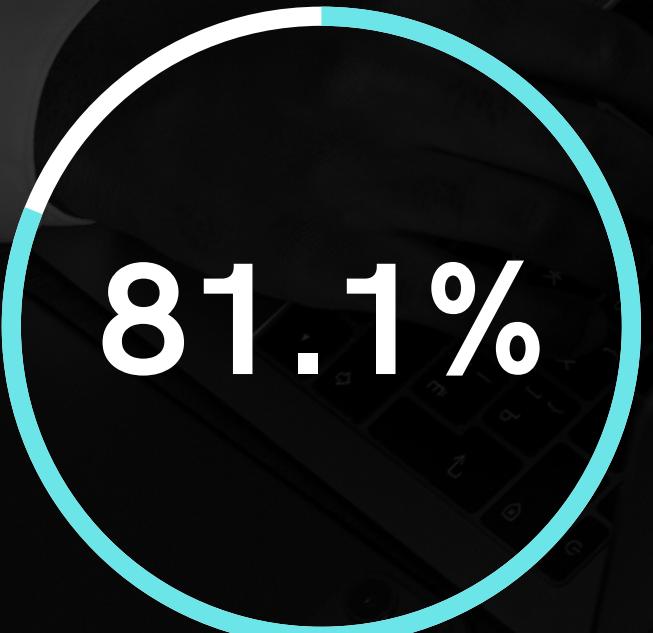
# Model Summary

<b>Lead Source_Welingak Website</b>	5.811465
<b>Lead Source_Reference</b>	3.316598
<b>What is your current occupation_Working Professional</b>	2.608292
<b>Last Activity_Other_Activity</b>	2.175096
<b>Last Activity_SMS Sent</b>	1.294180
<b>Total Time Spent on Website</b>	1.095412
<b>Lead Source_Olark Chat</b>	1.081908
<b>const</b>	-0.037565
<b>Last Notable Activity_Modified</b>	-0.900449
<b>Last Activity_Olark Chat Conversation</b>	-0.961276
<b>Lead Origin_Landing Page Submission</b>	-1.193957
<b>Specialization_Others</b>	-1.202474
<b>Do Not Email</b>	-1.521825

Based on different models prepared, results of final model are given with value of coefficient for the respective variables and constant value being -0.038

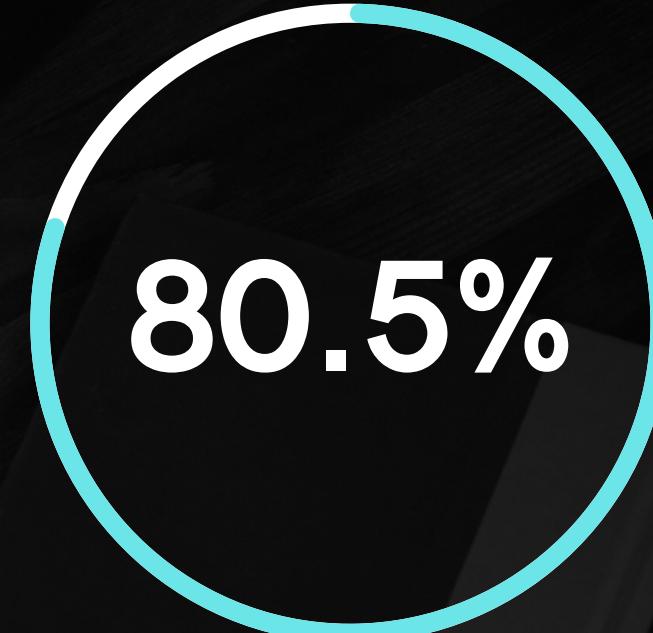
# Train Set Summary

---



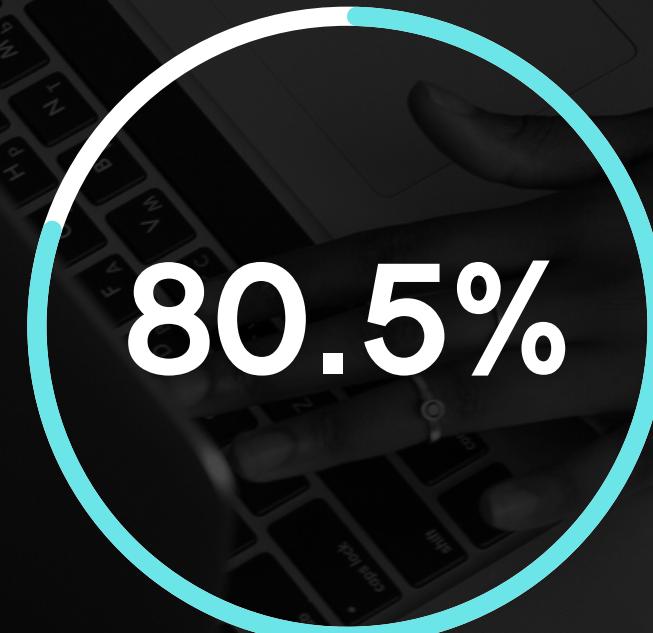
81.1%

Accuracy



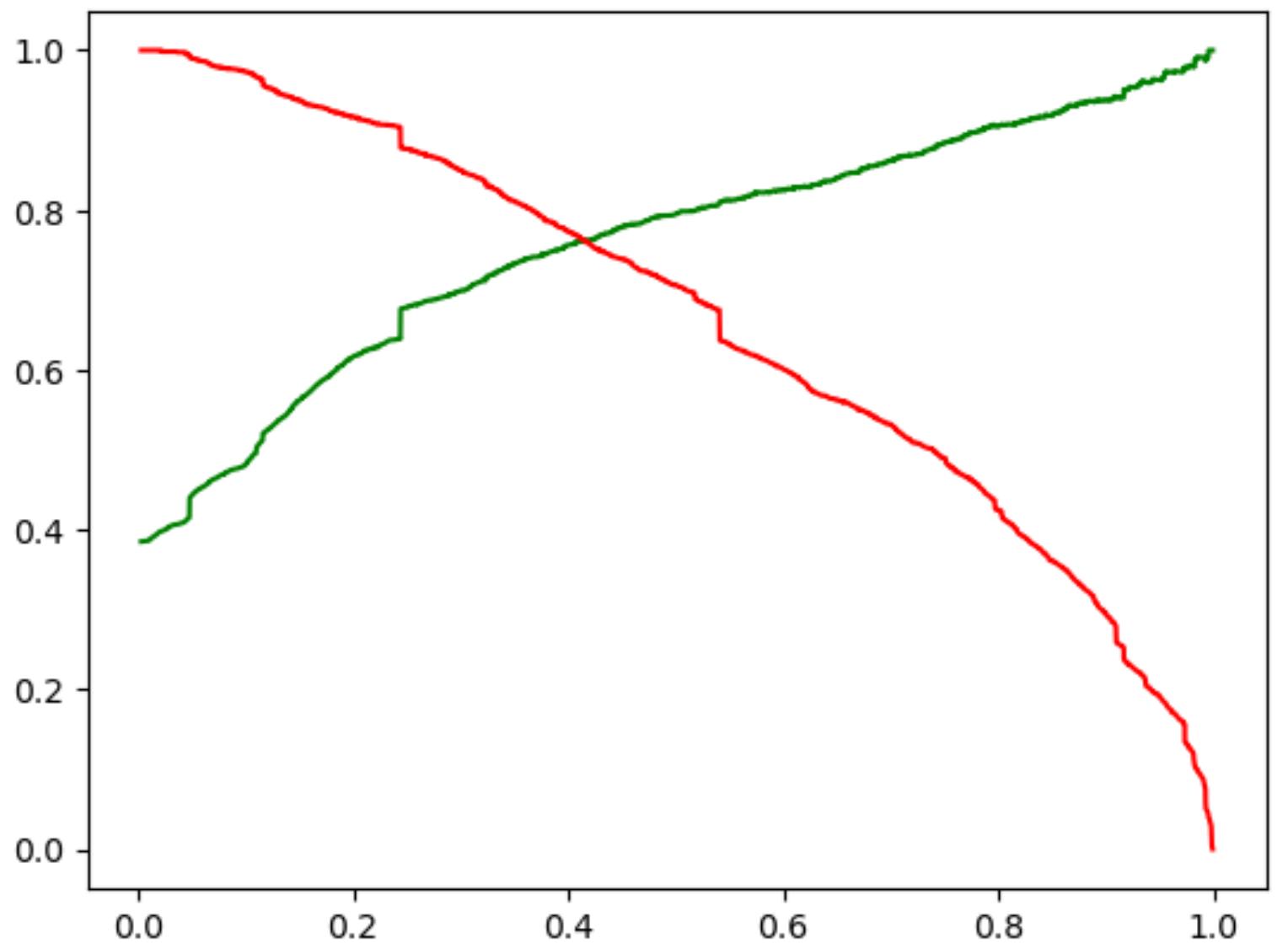
80.5%

Sensitivity



80.5%

Specificity



3151  
754  
444  
1999

# Train Set Summary

79.5%

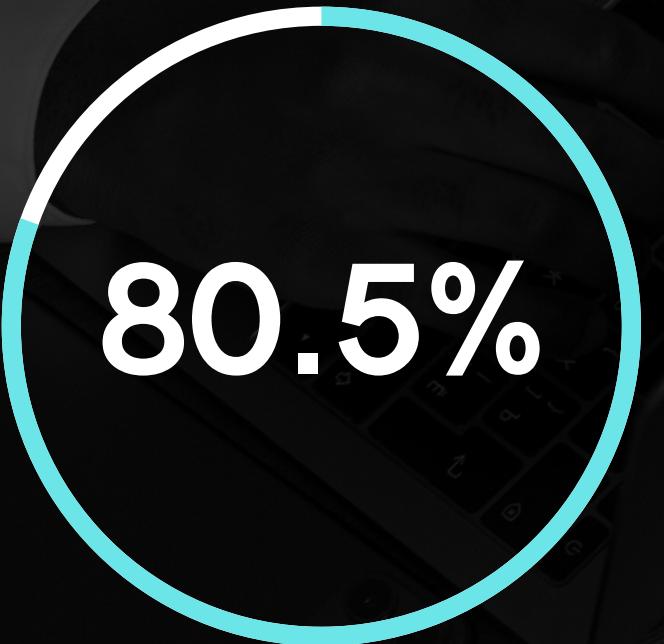
Precision

70.6%

Recall

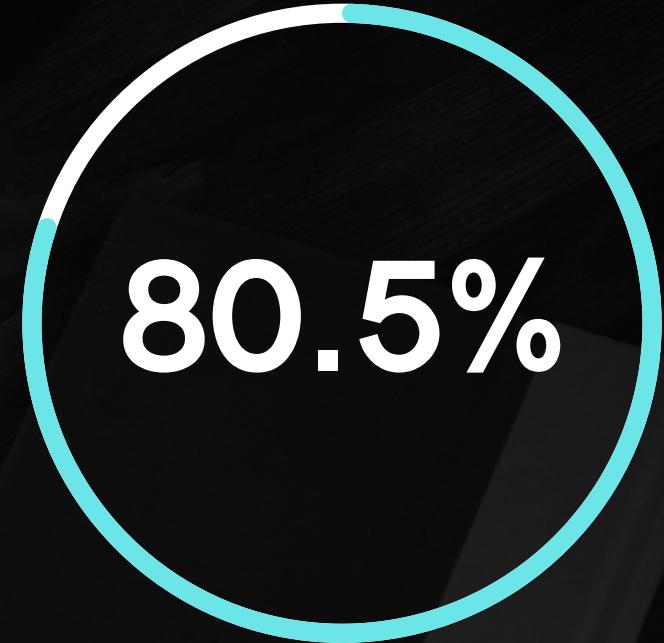
# Test Set Summary

---



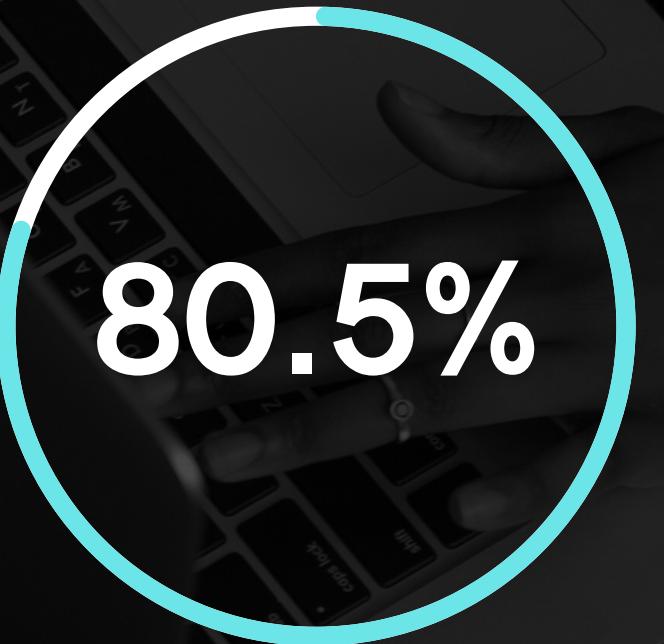
80.5%

Accuracy



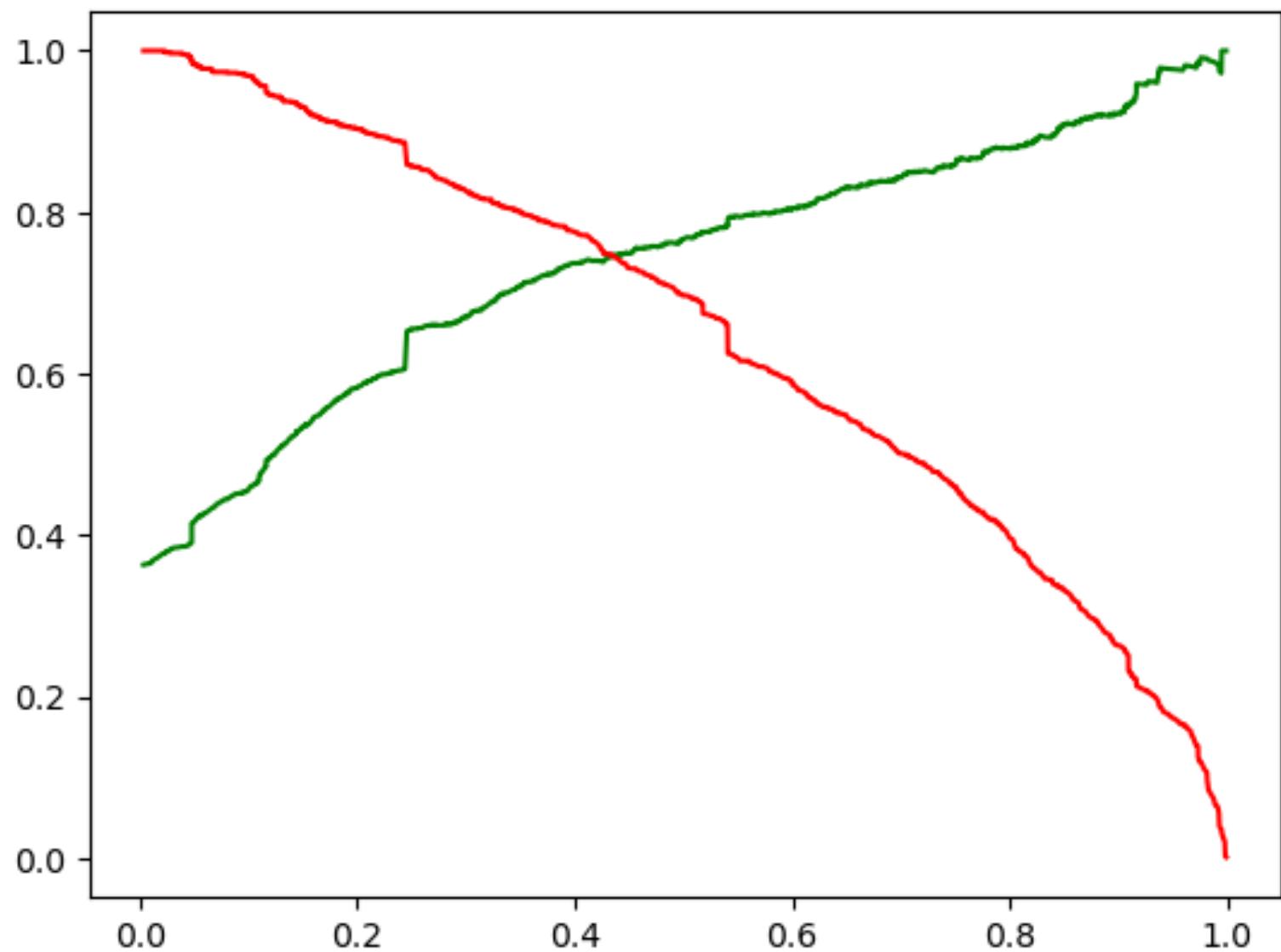
80.5%

Sensitivity



80.5%

Specificity



1396  
338  
193  
796

# Train Set Summary

76.7%

Precision

69.7%

Recall

# Recommendations

---

Based on the regression model on the dataset, following recommendations can be followed by X Education to improve the probability of conversion of leads

---



# X Education should make calls to leads

- 1 Coming from sources Welingak website and reference
- 2 Working Professionals
- 3 Spend more time on website
- 4 Coming from source Olark Chat
- 5 Last Activity was SMS sent

# X Education should make calls to leads

- 
- 1 Last activity was Clark chat conversation
  - 2 originating from Landing Page submission
  - 3 Specialisation is others
  - 4 Responding yes to do not Email



A black and white aerial photograph of a city skyline, likely New York City, showing a high density of skyscrapers and buildings.

Thank you!