

# *Lead Score Case Study*

*BY-*

*SHUBHNEET ARORA*

*SAMTA MALIK*

# Problem Statement:

---

X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like

Google.

Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these

people fill up a form providing their email address or phone number, they are classified to be a lead.

Moreover, the company also gets leads through past referrals.

Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.

Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

# Business Goal:

---

- *X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.*
- *The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.*
- *The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.*

# Problem solving Methodology

## Data Sourcing, Cleaning and Preparation

- Read the Data from Source
  - Convert data into clean format suitable for analysis
  - Remove duplicate data
  - Outlier Treatment
- Exploratory Data Analysis
- Feature Standardization.

## Feature Scaling and Splitting Train and Test Sets

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

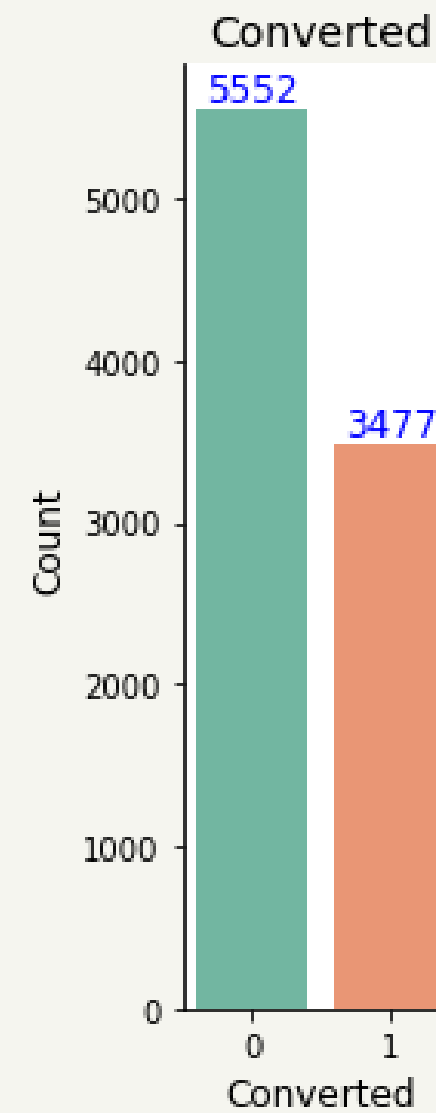
## Model Building

- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

## Conclusion

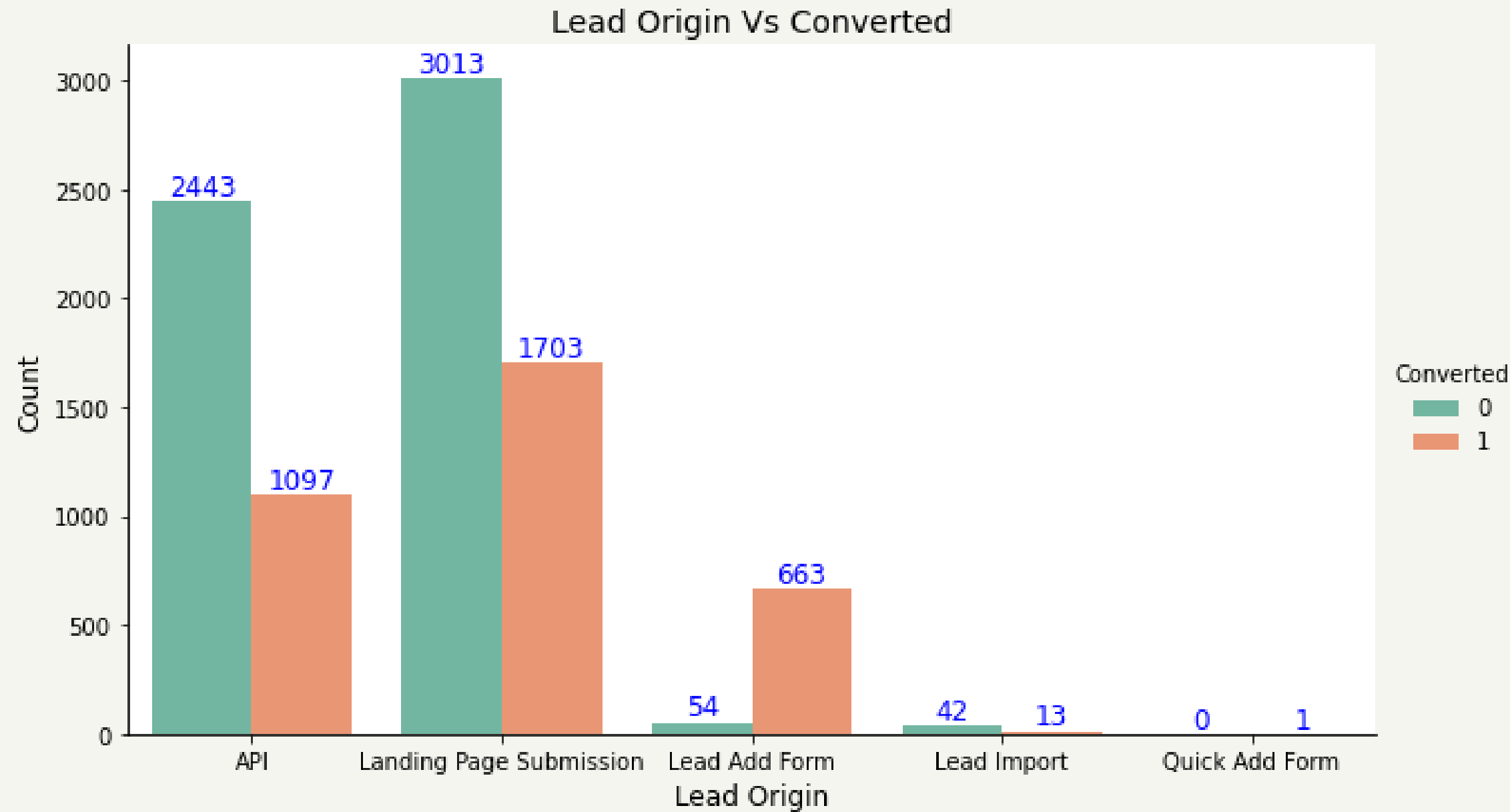
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

# EXPLORATORY DATA ANALYSIS



*there has been a overall conversion rate of around 39%.*

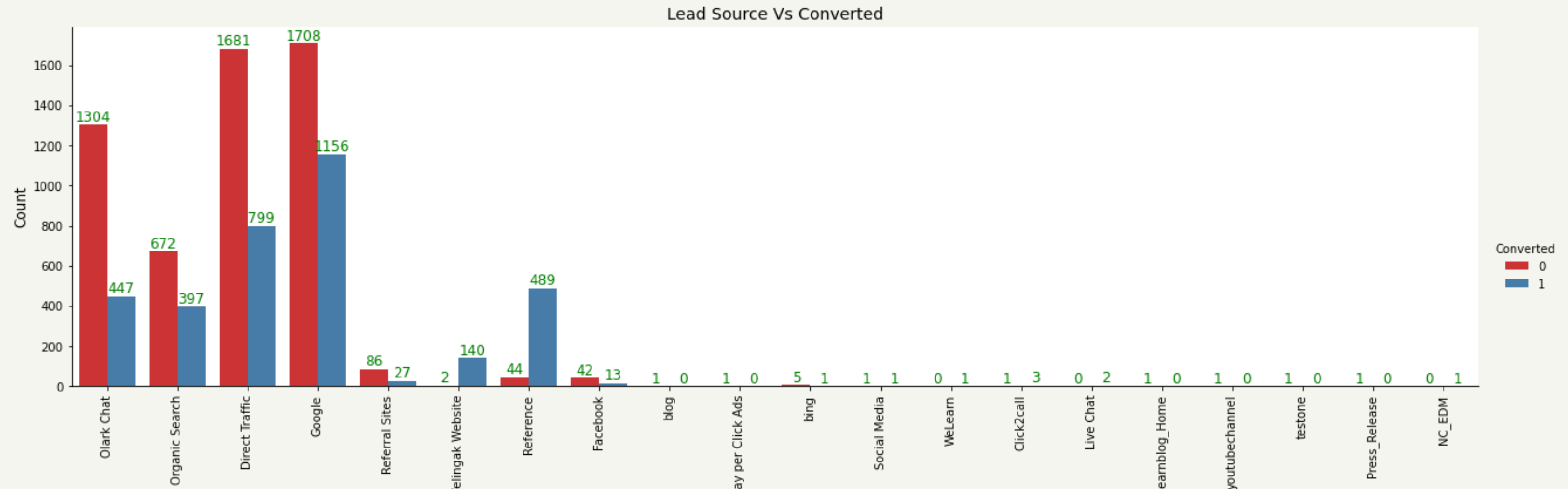
# LEAD ORIGIN VS CONVERTED



*It can be seen that the maximum conversion happened from Landing Page Submission*

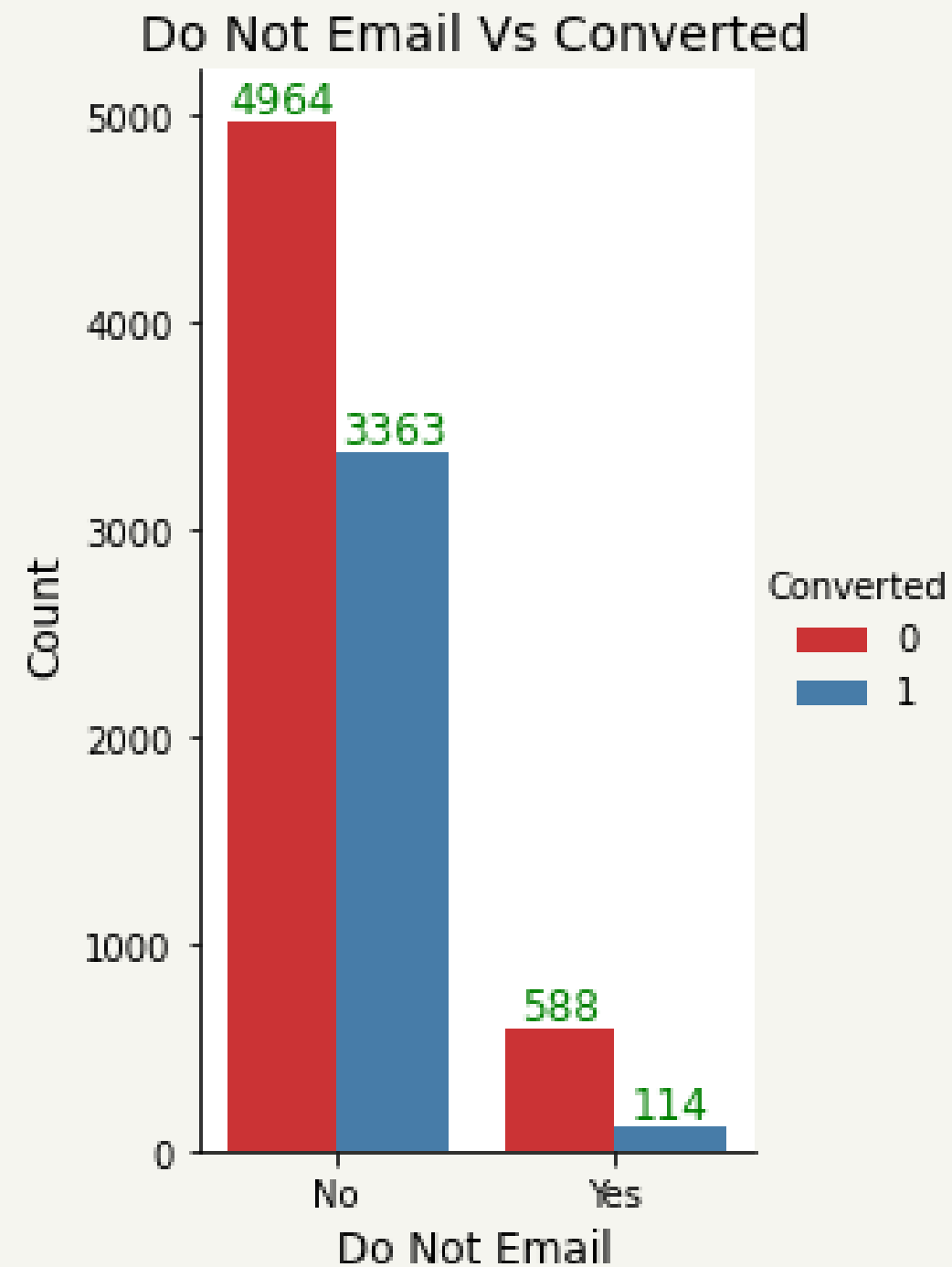
*Also there was only one request from quick add form which got converted.*

# LEAD SOURCE VS CONVERTED



*It can be seen that major conversion in the lead source is from google.*

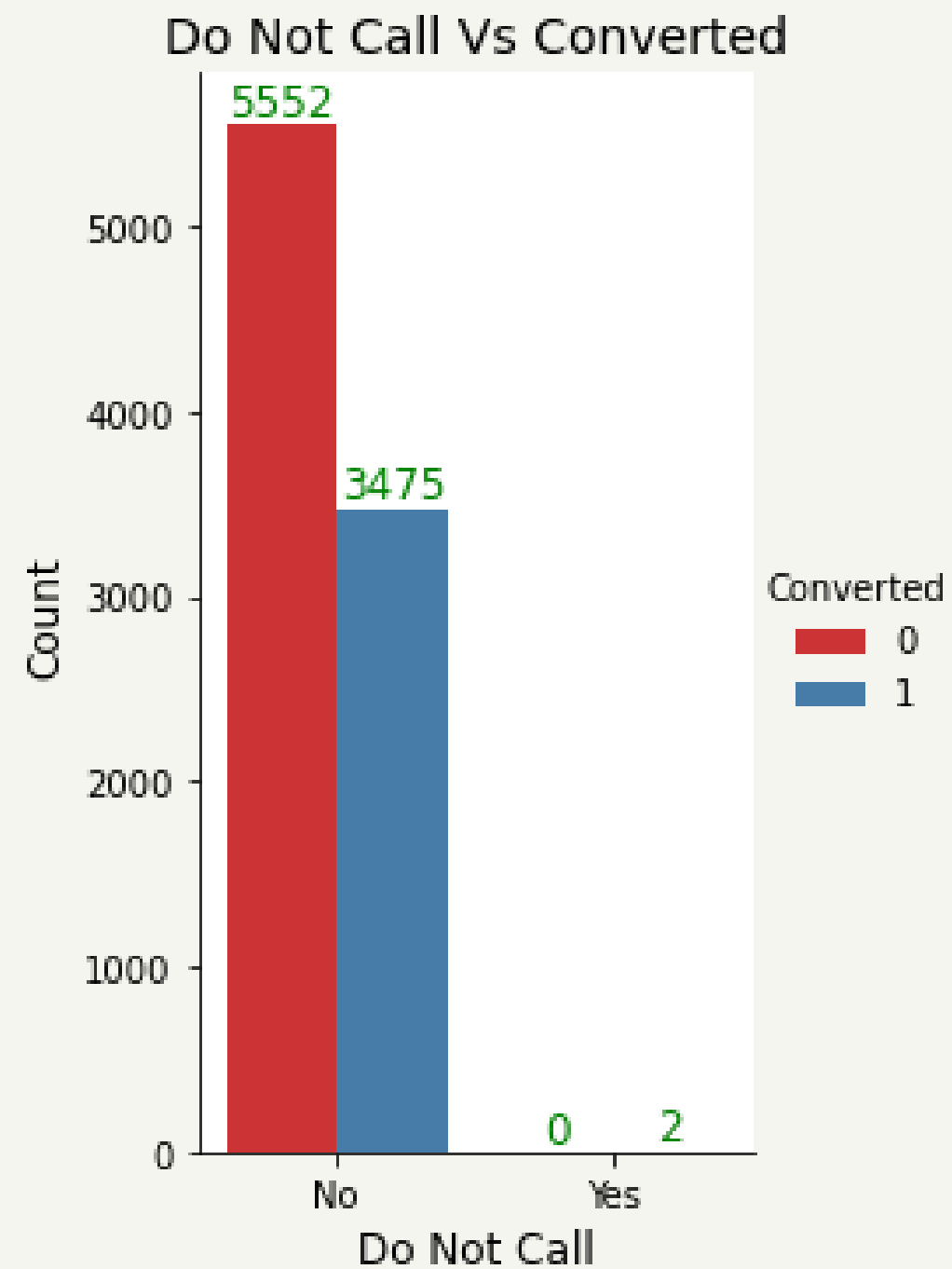
# DO NOT EMAIL VS CONVERTED



*Major conversion has  
happend from  
the emails that have  
been sent.*

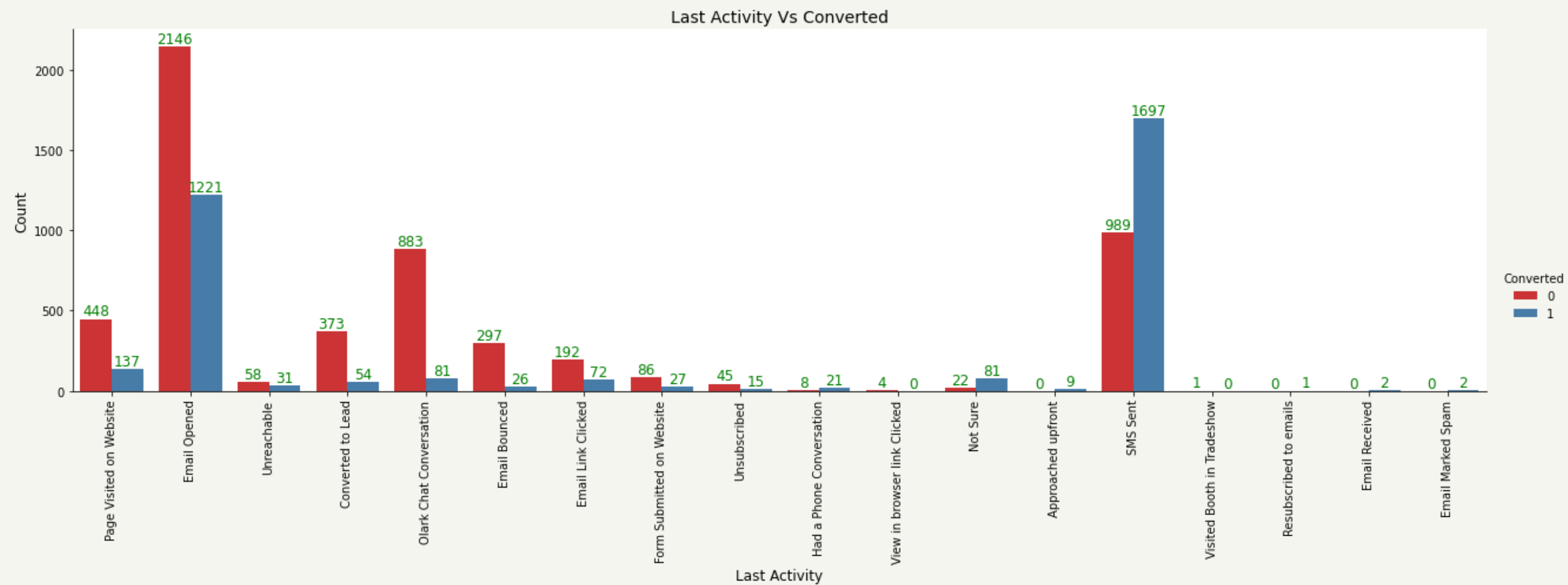


# DO NOT CALL VS CONVERTED



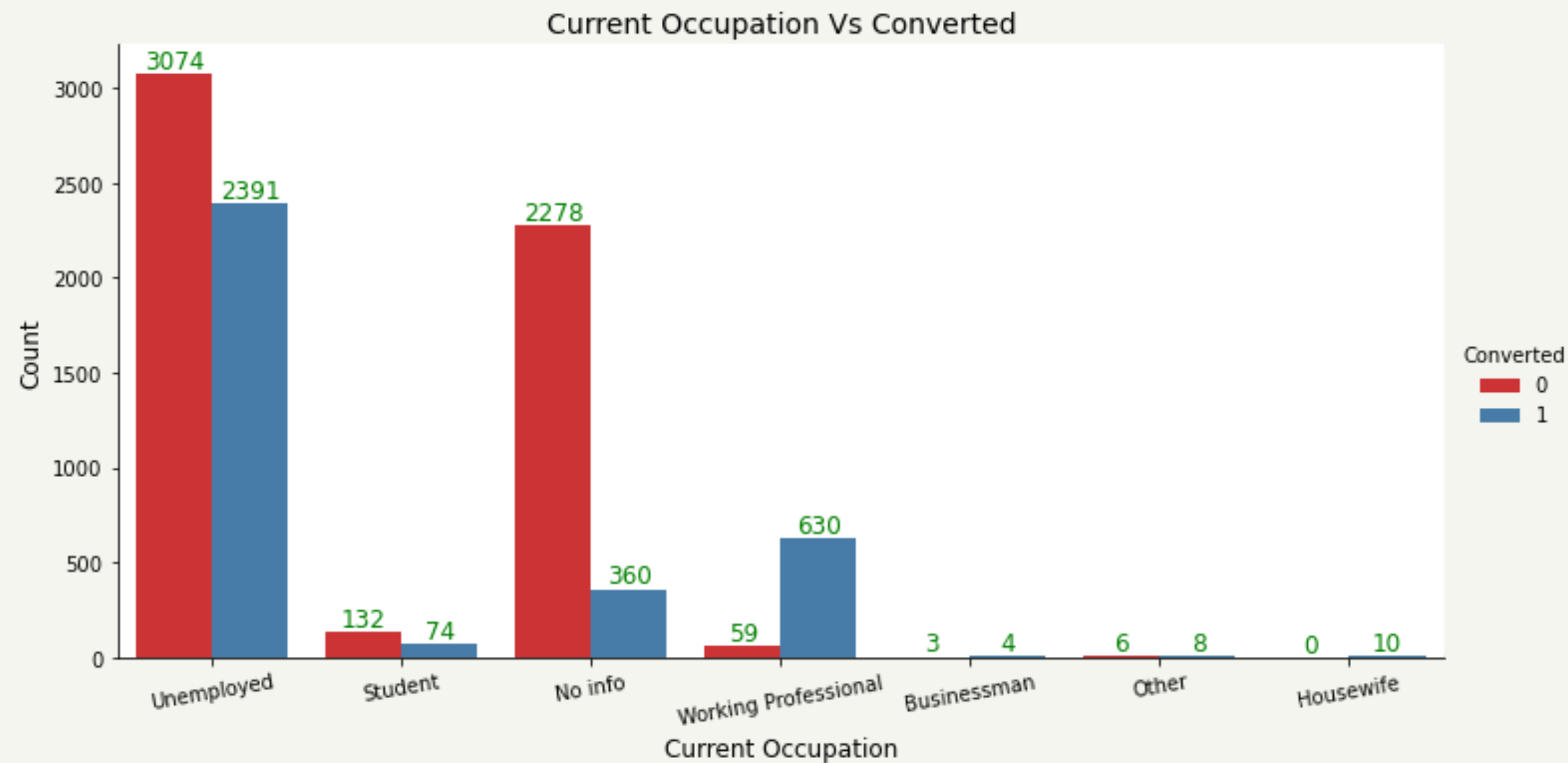
*It can be noticed that major conversions happened when calls were made.*

# LAST ACTIVITY VS CONVERTED



*It can be seen that last activity value of 'SMS Sent' had more conversion*

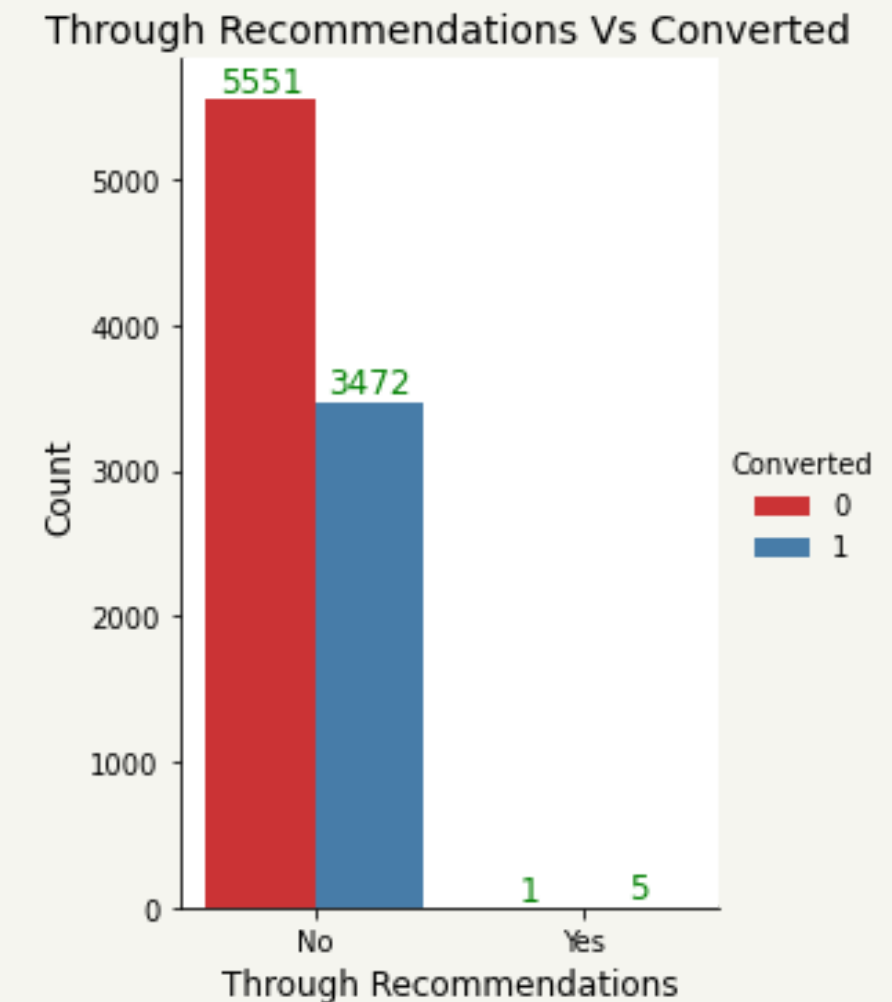
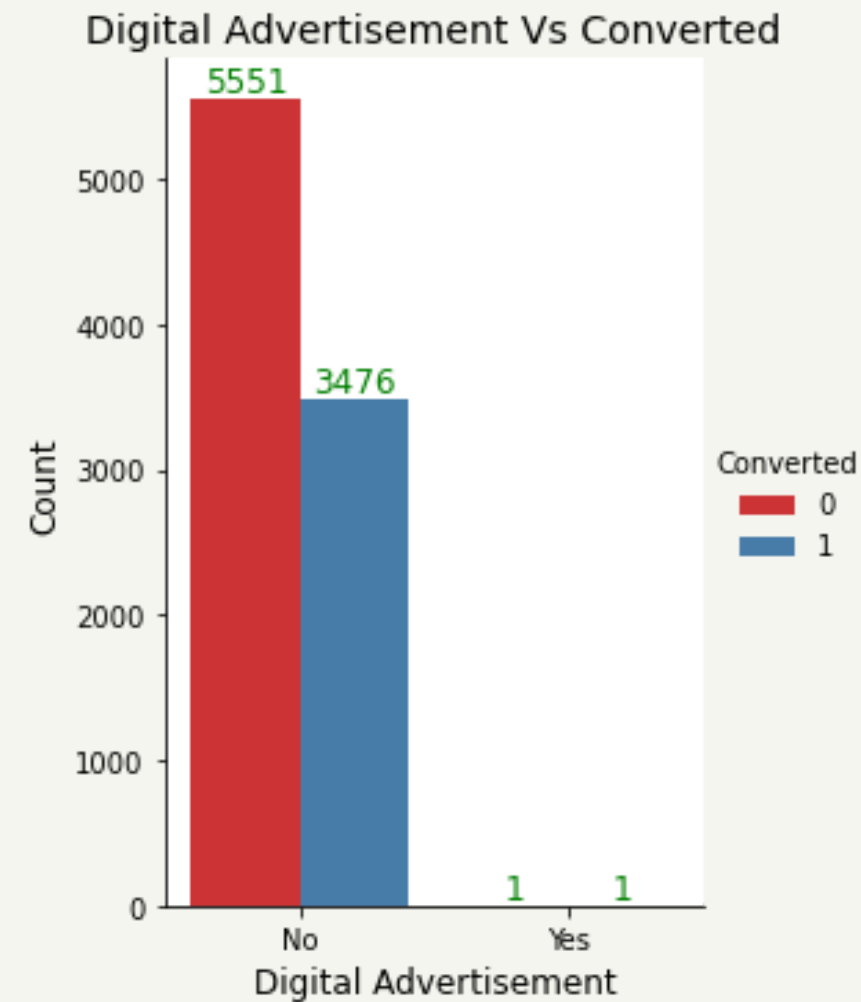
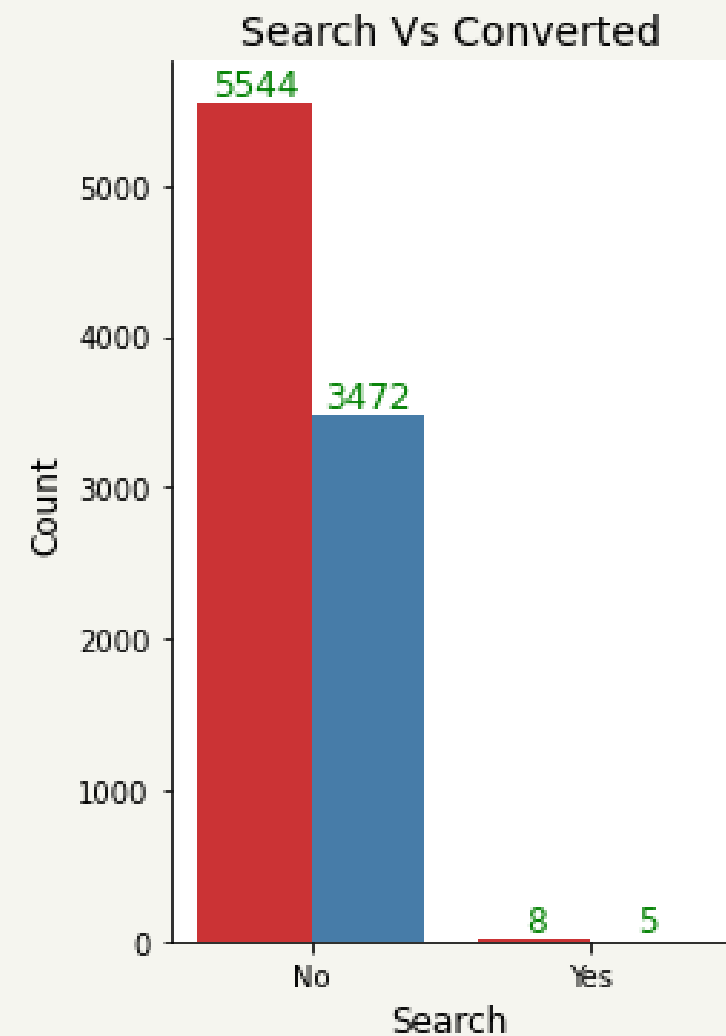
# CURRENT OCCUPATION VS CONVERTED



More conversion happen  
with people who are  
unemployed. It can also be  
noticed from the above data  
that:

1. Out of 7 business men, 4  
got converted.
2. Out 10 housewives, all 10  
leads got converted.

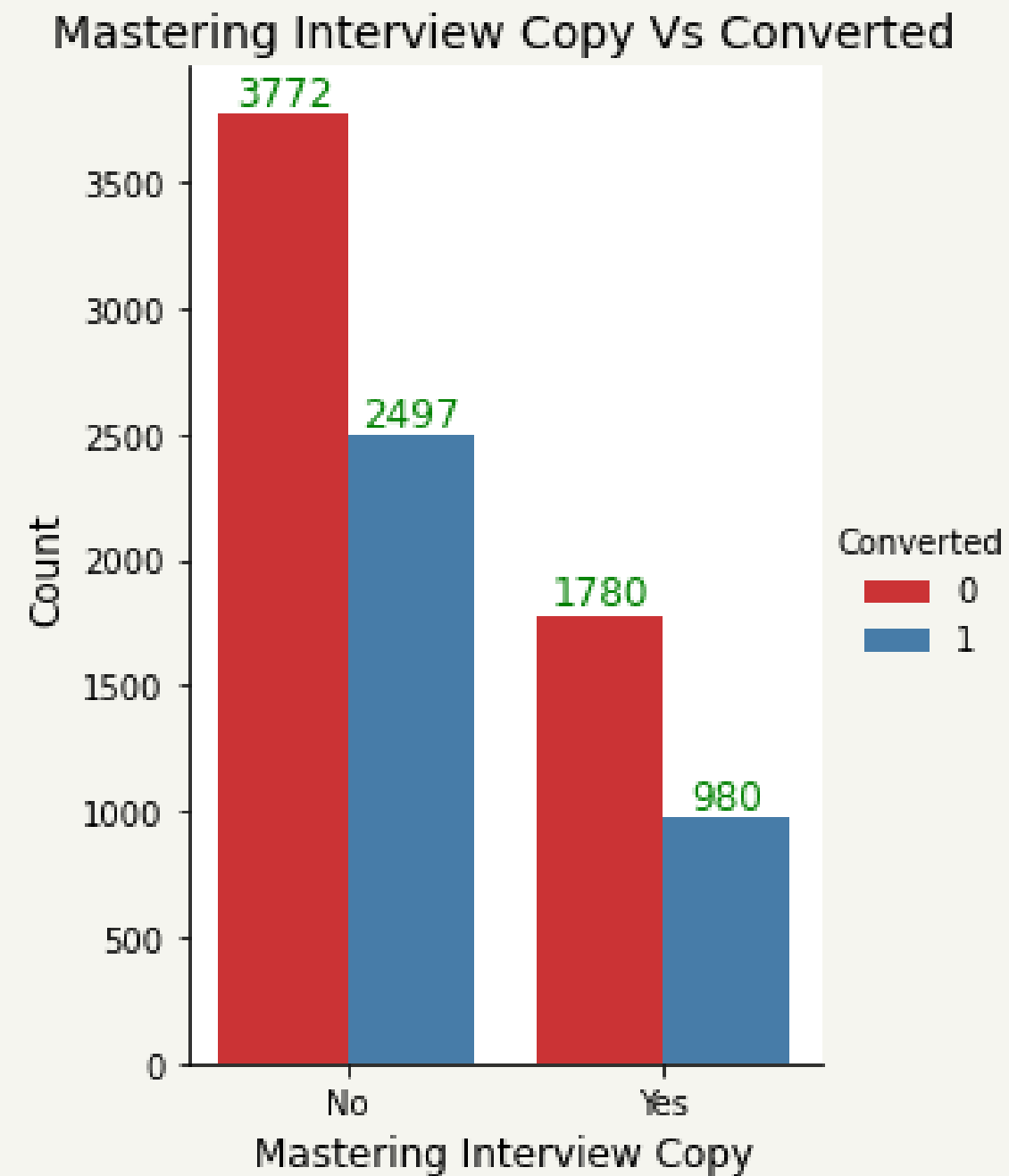
# SEARCH, DIGITAL ADVERTISEMENT AND THROUGH RECOMM.



---

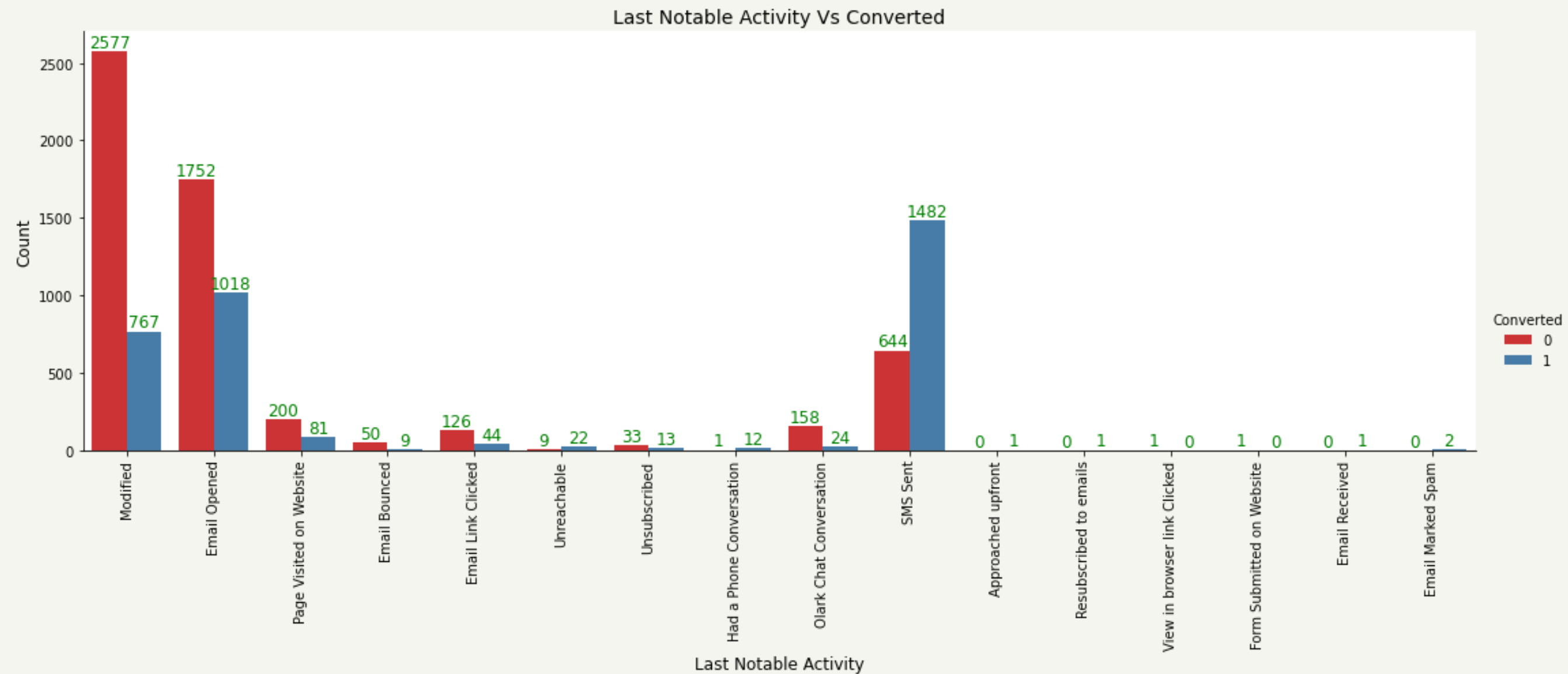
*Not much impact on conversion rates through Search, digital advertisements and through recommendations*

# MASTERING INTERVIEW COPY VS CONVERTED



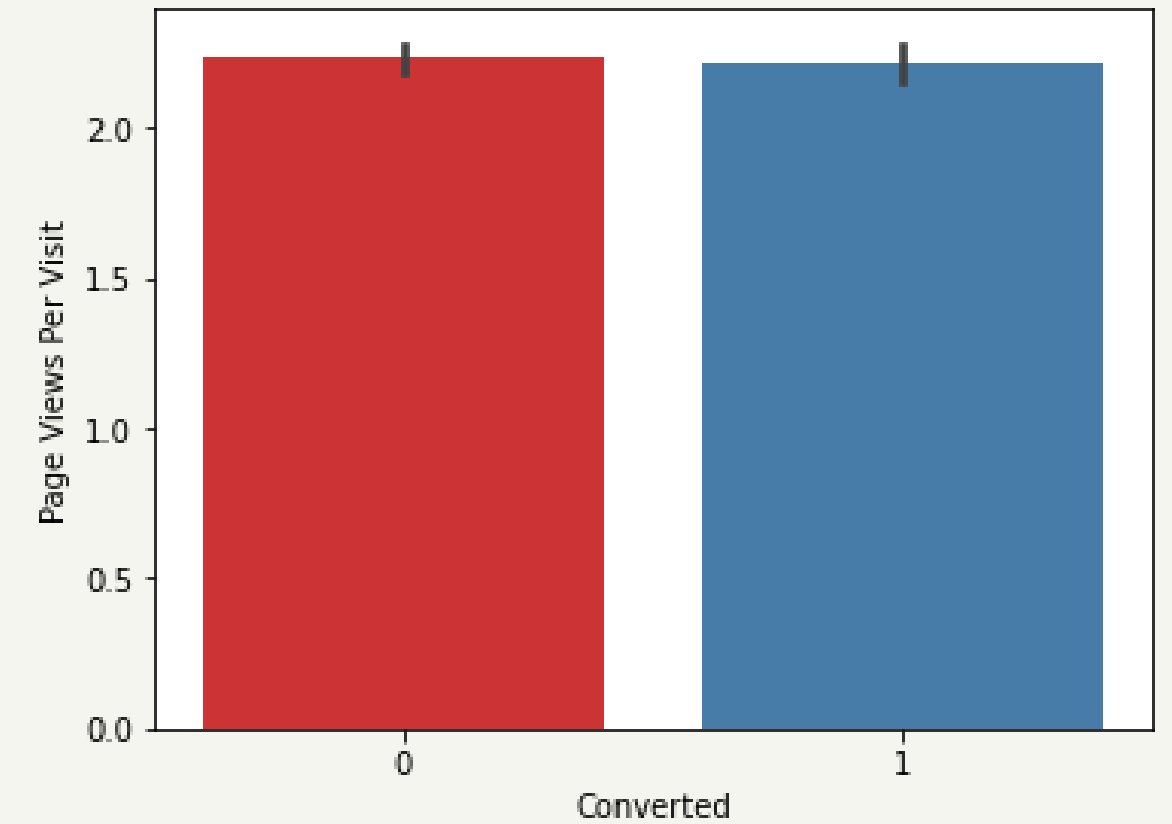
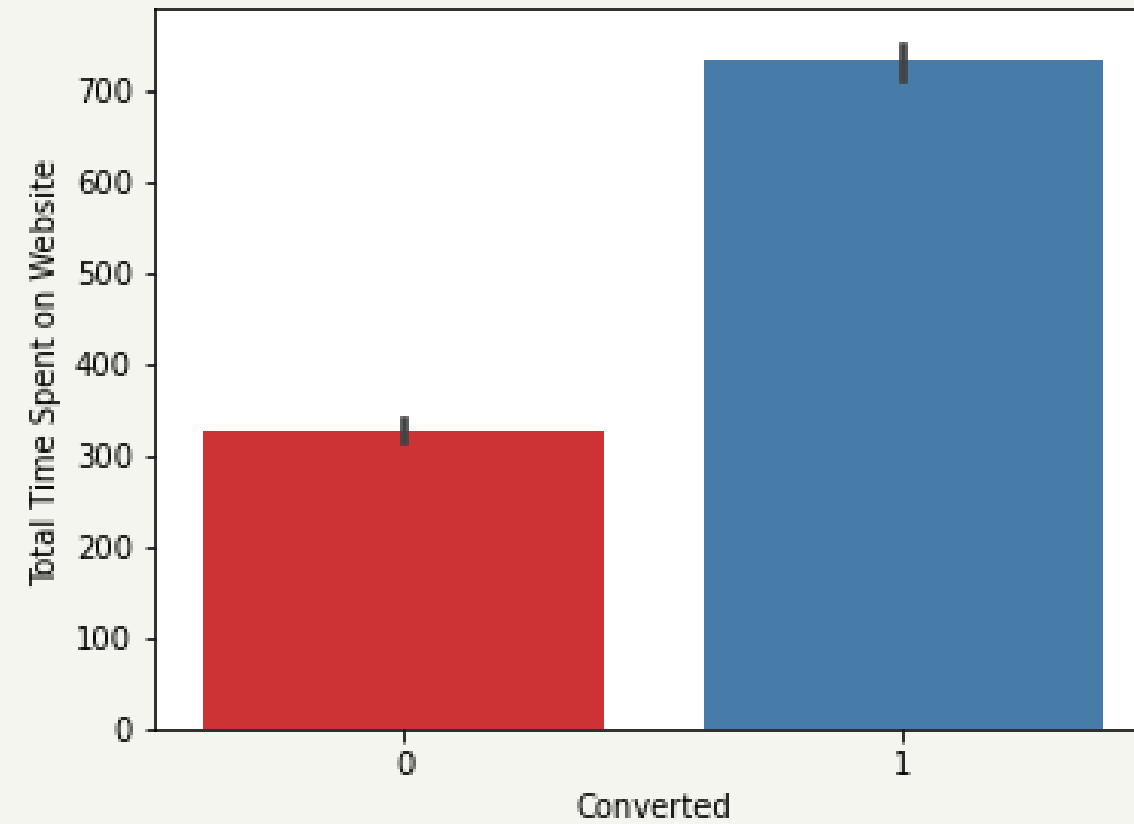
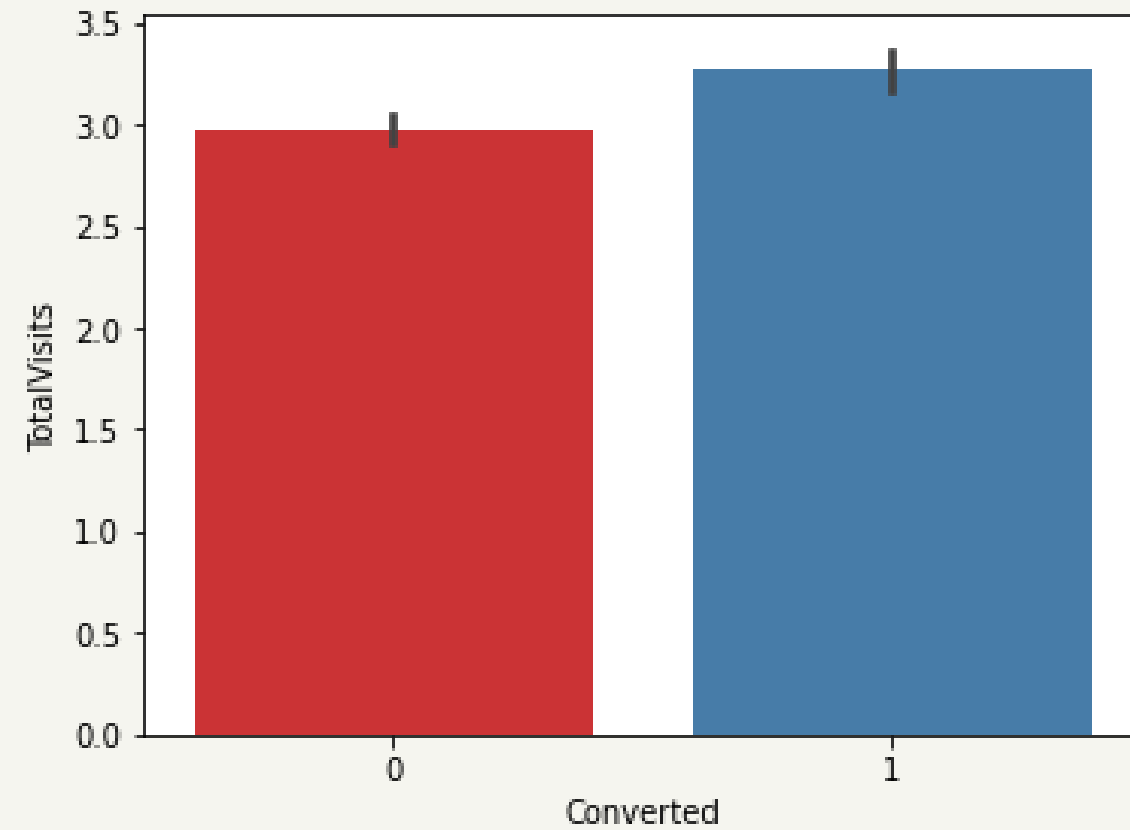
*Conversion rate is high on leads who do  
not want  
a free copy of Mastering Interviews*

# LAST NOTABLE ACTIVITY VS CONVERTED



*It can be noticed that the conversion rate is high for "SMS Sent".*

# NUMERIC COLUMNS



---

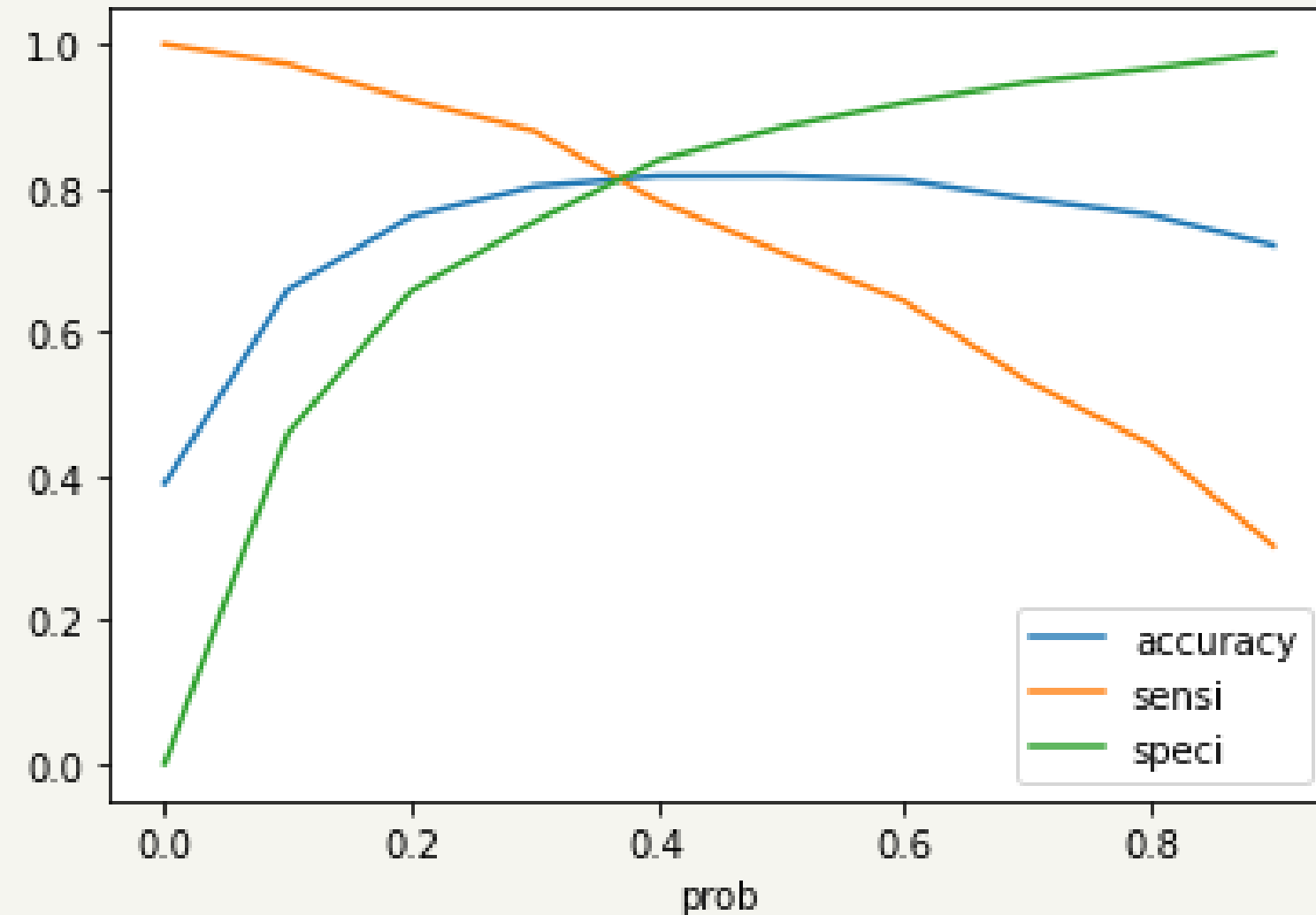
*The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit.*

# Variables Impacting the Conversion Rate

- TotalVisits
- Total Time Spent on Website
- LastActivity\_Email Opened
- LeadSource\_Olark Chat
- LastActivity\_SMS Sent
- Do Not Email
- LastNotableActivity\_Unreachable
- LastActivity\_Olark Chat Conversation
- LastNotableActivity\_Modified
- LeadOrigin\_Lead Add Form
- CurrentOccupation\_No info
- LeadSource\_Welingak Website
- CurrentOccupation\_Working Professional
- LastNotableActivity\_Had a Phone Conversation



# Model Evaluation



## Finding Optimal Cut off Point

Optimal cut off probability is that probability where we get balanced sensitivity and specificity..

The graph depicts an optimal cut off of 0.38 based on Accuracy, Sensitivity and Specificity

# CONCLUSION

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 81%, 79% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- The top 3 variables that contribute for lead getting converted in the model are:
  1. Total time spent on website
  2. Lead Add Form from Lead Origin
  3. Had a Phone Conversation from Last Notable Activity

Hence overall this model seems to be good.

*THANKYOU*