# Lead Scoring Case Study Presentation

Problem Statement:

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example,if, say, they acquire 100 leads in a day, only about 30 of them are converted. 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. A typical lead conversion process can be represented using the following funnel: .
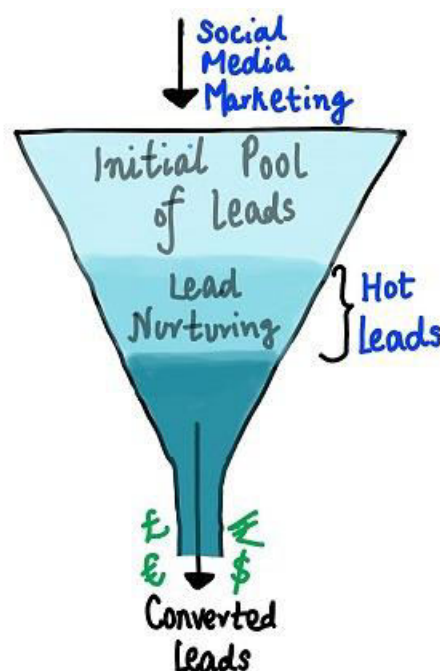
As you can see, there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers

from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to bearound 80%.

Implementation:

1) Weeding out the high null value variables and replacing remaining nulls with modes:-
**After doing the above process we were left with the following 11 columns to work on:**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9240 entries, 0 to 9239
Data columns (total 12 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Lead Origin                           9240 non-null   object
 1   Lead Source                           9240 non-null   object
 2   Do Not Email                          9240 non-null   object
 3   Converted                             9240 non-null   int64
 4   TotalVisits                           9240 non-null   object
 5   Total Time Spent on Website           9240 non-null   int64
 6   Page Views Per Visit                  9240 non-null   object
 7   Last Activity                         9240 non-null   object
 8   What is your current occupation       9240 non-null   object
 9   Update me on Supply Chain Content     9240 non-null   object
 10  A free copy of Mastering The Interview 9240 non-null   object
 11  Last Notable Activity                 9240 non-null   object
dtypes: int64(2), object(10)
```

**2)Univariate Analysis on columns with varied responses:-**

A)
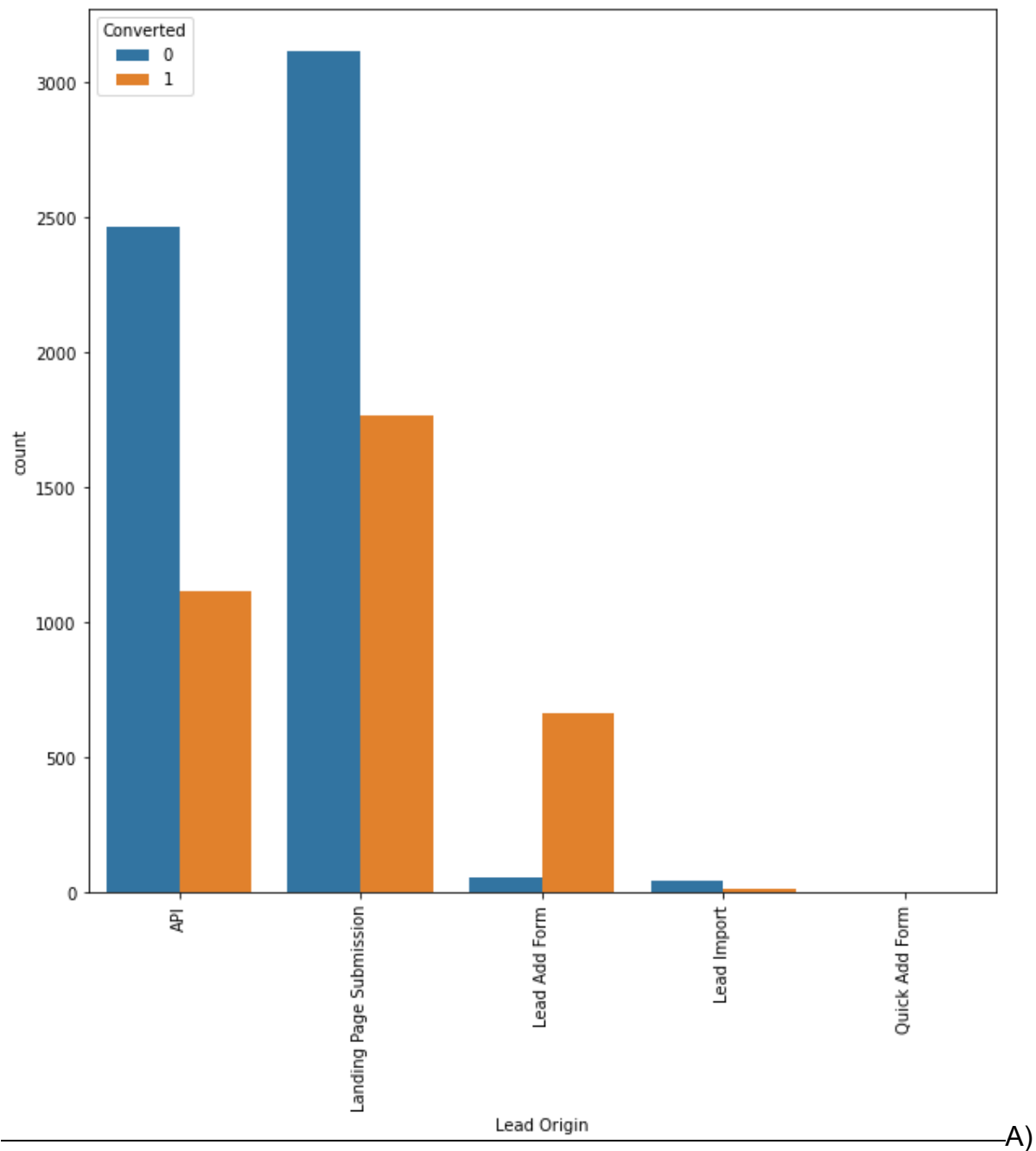
Lead Origin: As shown in the above image for lead origin, those in the Lead Add From category showed a high rate of conversion.
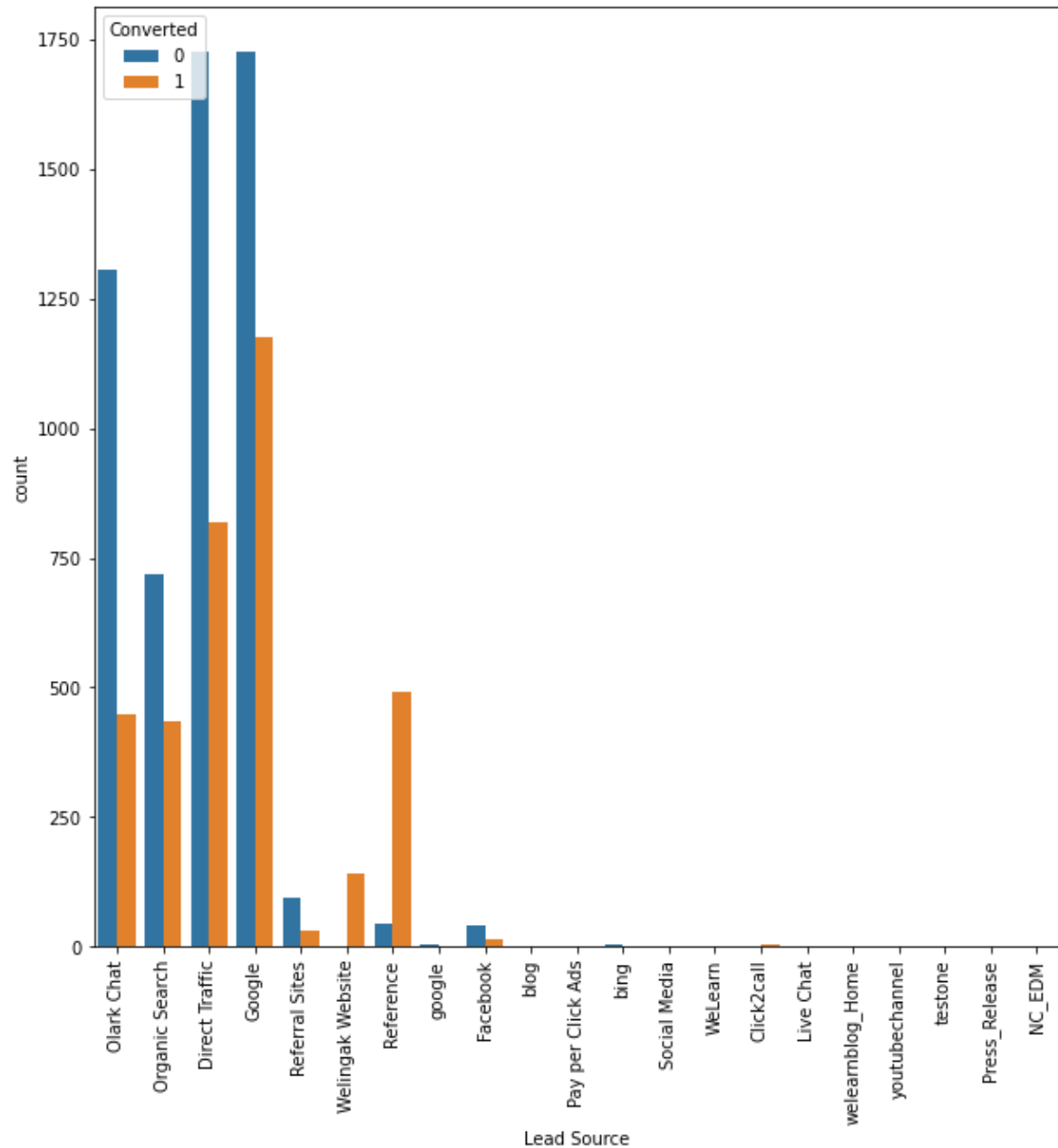
B) Lead Source: As shown in the image above, leads sourced through Reference and Wellingak Website have a very high rate of conversion while those sourced through Google have a higher rate of conversion than through other popular methods like Direct Traffic, Organic Search and Olark Chat.
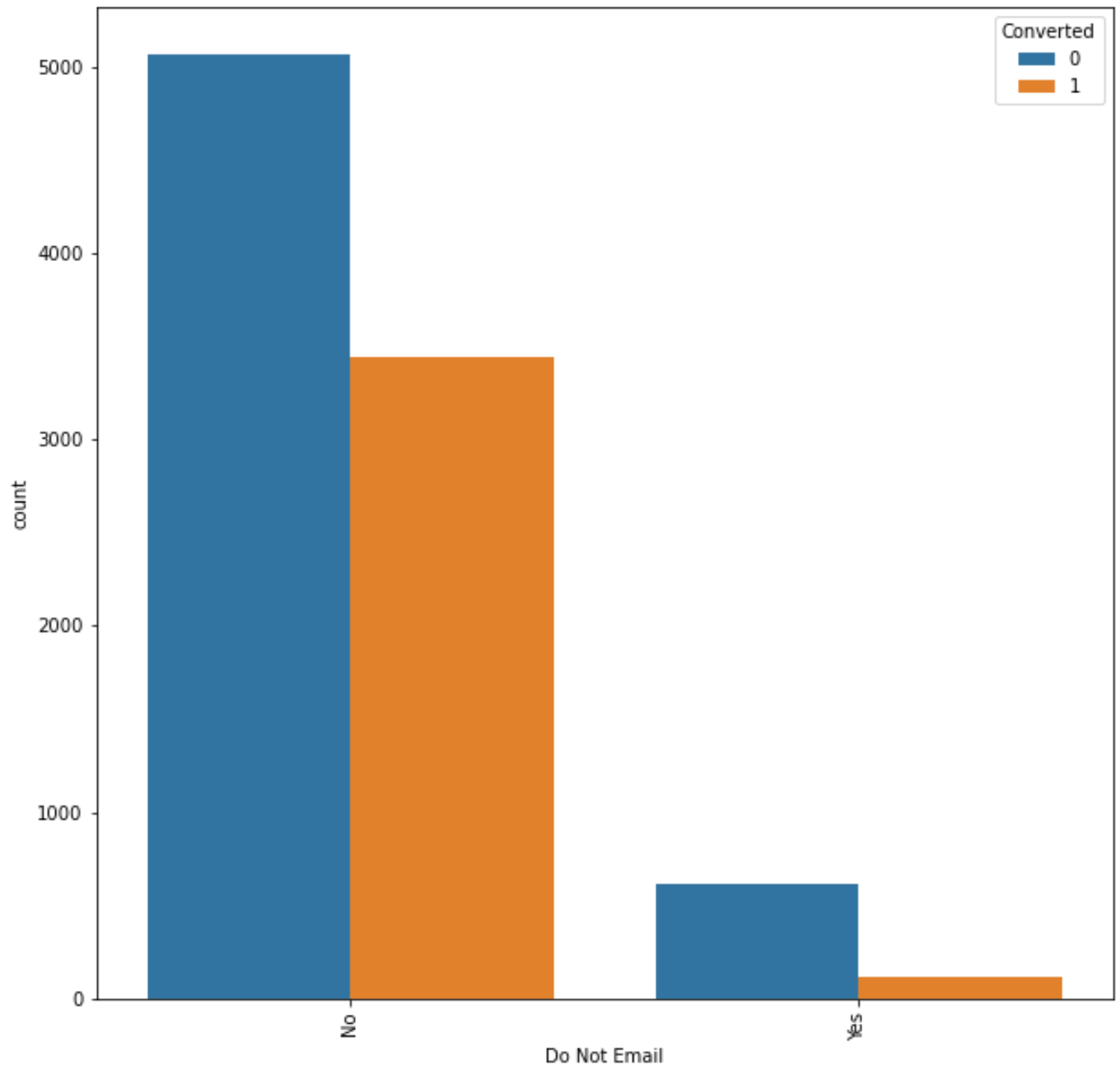
C)Do Not E-Mail: As shown in the graph above, those who chose Do Not Email were were very less likely to get converted.
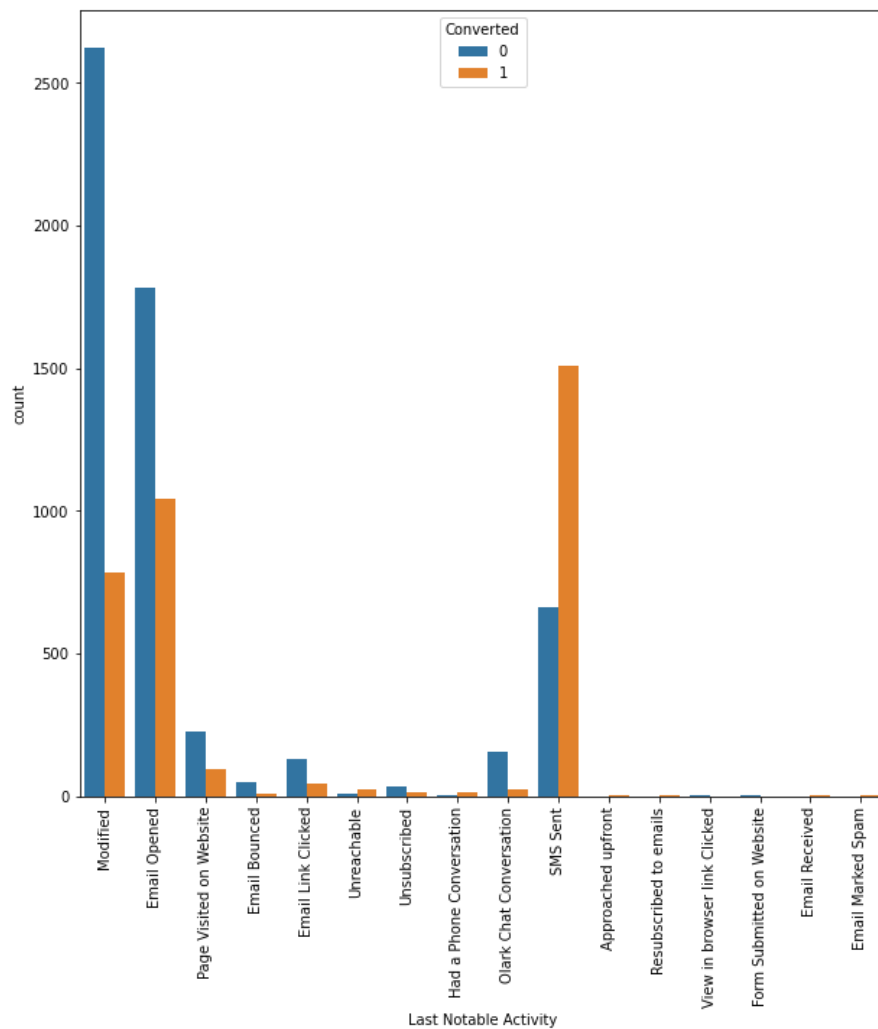
D) Last Notable Activity: As shown in the image above, those whose last notable activity was SMS sent had very high rates of conversion followed by email opened. All others had regular low rates of conversion.

E)
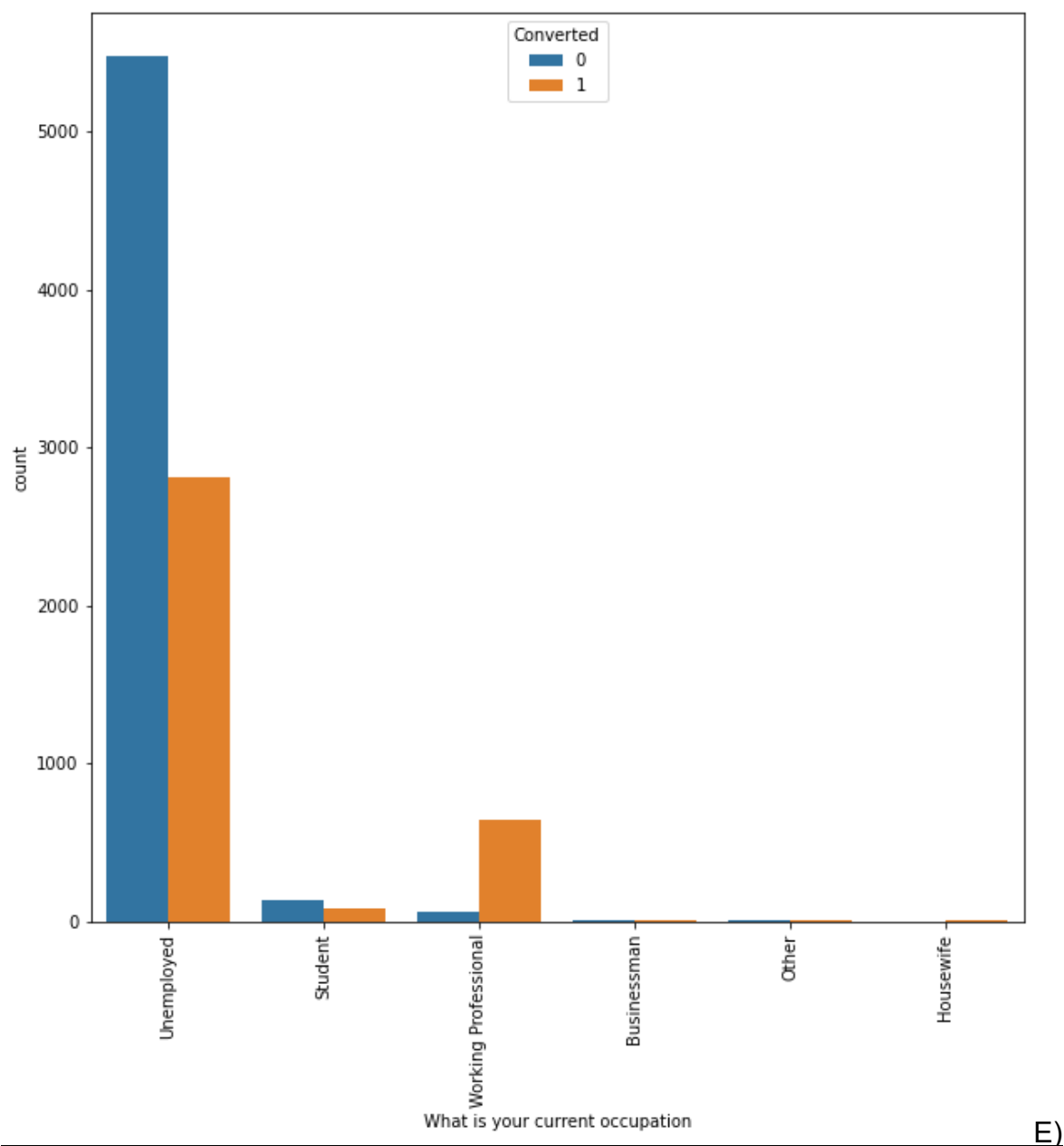
Current Occupation: As shown in the image above, working professionals had much higher rates of conversion compared to students, unemployed and others.
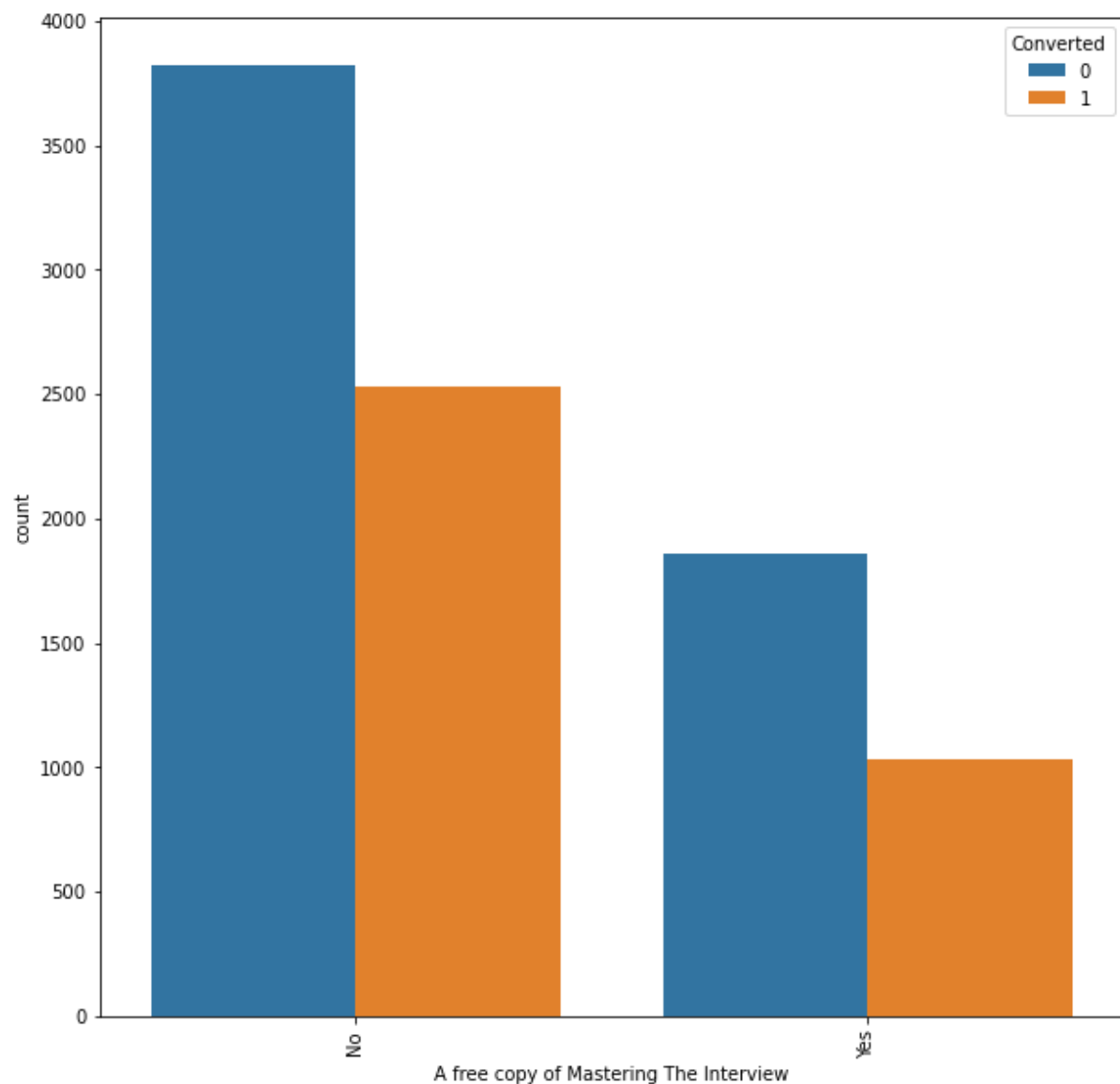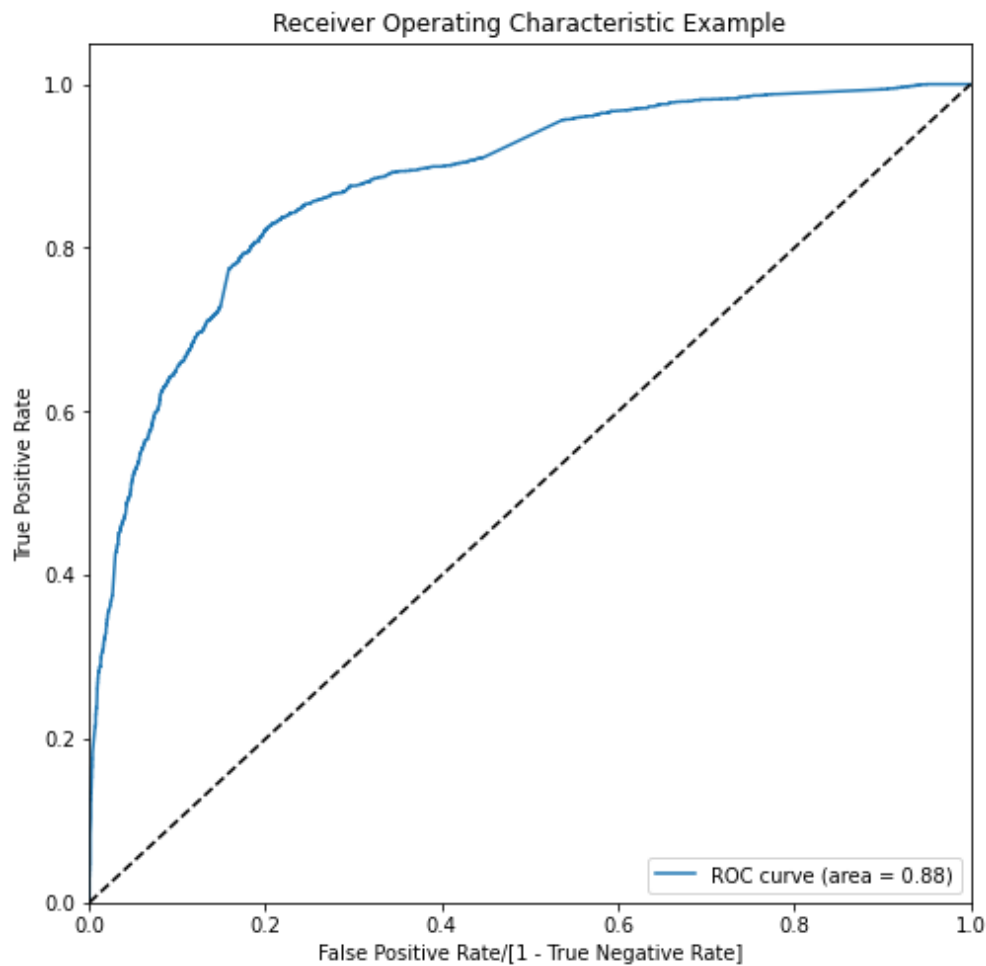
F) A Free Copy of Mastering the interview: As shown in the image above, handing out a free copy of mastering the interview dosent seem to make much of a difference either.

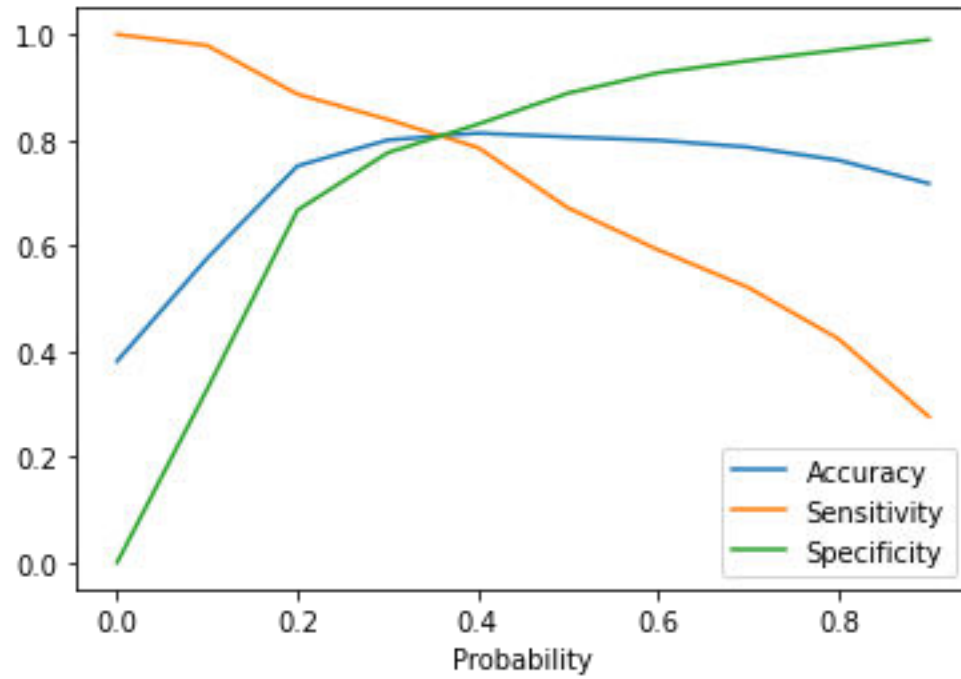## 3) Train-Test Split and Final Analysis:
First we did the Train/Test split 70:30 and here are the results of the final analysis:
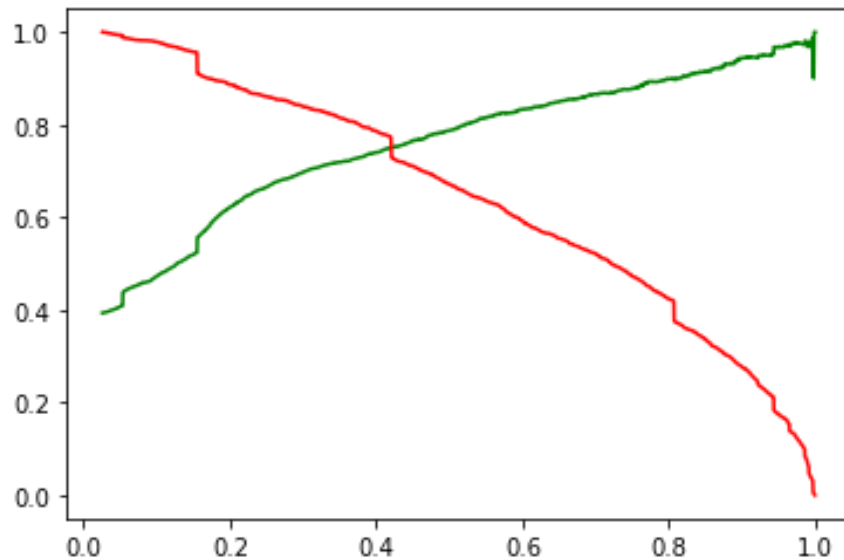
A) ROC Curve:-

As shown in the image above, the ROC Curve area was achieved to be 0.88

B) Optimal Cutoff Point:

As shown in the image above, the optimal cutoff point was around 0.35

C) Precision Recall Tradeoff:

As shown in the Image above the Precision-Recall Tradeoff is roughly around 0.4

D) Accuracy Sensitivity and Specificity on Train/Test Sets:-

On Train Set:-

Accuracy: **0.808**
Sensitivity: **0.815**
Specificity: **0.804**

On Test Set:-

Accuracy: **0.813**
Sensitivity: **0.805**
Specificity: **0.819**

E)Average Lead Score: On Average the lead score of those who got converted was **69** while those who didnt convert was **14** so it is recommended to avoid those below the score of 14 and try to convery those above the score of 69.

F)Final Recommendations: Finally, based on the chart obtained below for the coefficients of each columns in the final regression model, here are the

recommendations for contacting and avoiding which kinds of leads for conversion:

Based on the image below, it would be more useful to convert those who are working professionals, whose last notable activity was having a phone conversation and whose lead origin is lead add form while its best to avoid those whose last activity was email bounced, who were unreachable or who had olark chat conversation.

```
Last Notable Activity_Had a Phone Conversation         3.208052
Last Notable Activity_Unreachable                      3.147343
Lead Origin_Lead Add Form                              3.124055
What is your current occupation_Working Professional   2.779770
Lead Source_Welingak Website                           1.871276
Last Notable Activity_Email Bounced                    1.693999
Last Notable Activity_SMS Sent                         1.374076
Total Time Spent on Website                            0.965467
Last Notable Activity_Page Visited on Website          0.720156
Do Not Email                                          -0.273210
Lead Source_Organic Search                            -0.386521
Lead Source_Direct Traffic                            -0.606431
Last Activity_Form Submitted on Website               -0.684080
const                                                 -0.921823
Last Activity_Page Visited on Website                 -0.956100
Last Activity_Olark Chat Conversation                 -1.182601
Last Activity_Converted to Lead                       -1.414138
Last Activity_Unreachable                             -1.590383
Last Activity_Email Bounced                           -1.605562
dtype: float64
```