

LEAD SCORING CASE STUDY

-ABHISHEK, PRIYANKA, RHEA(DS C55)

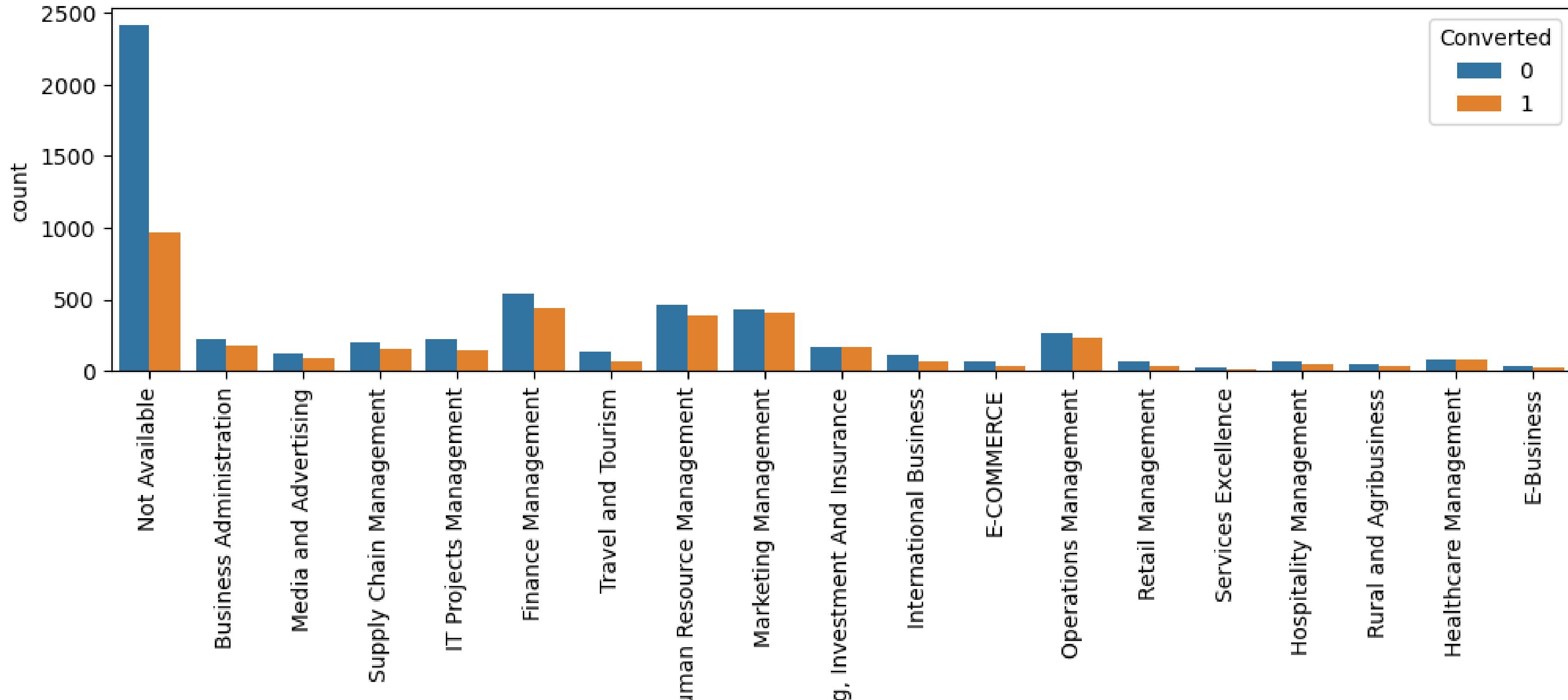
PROBLEM STATEMENT

To Build a logistic regression model for X Education to select the most promising leads i.e. the leads that are most likely to convert into paying customers by assigning a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

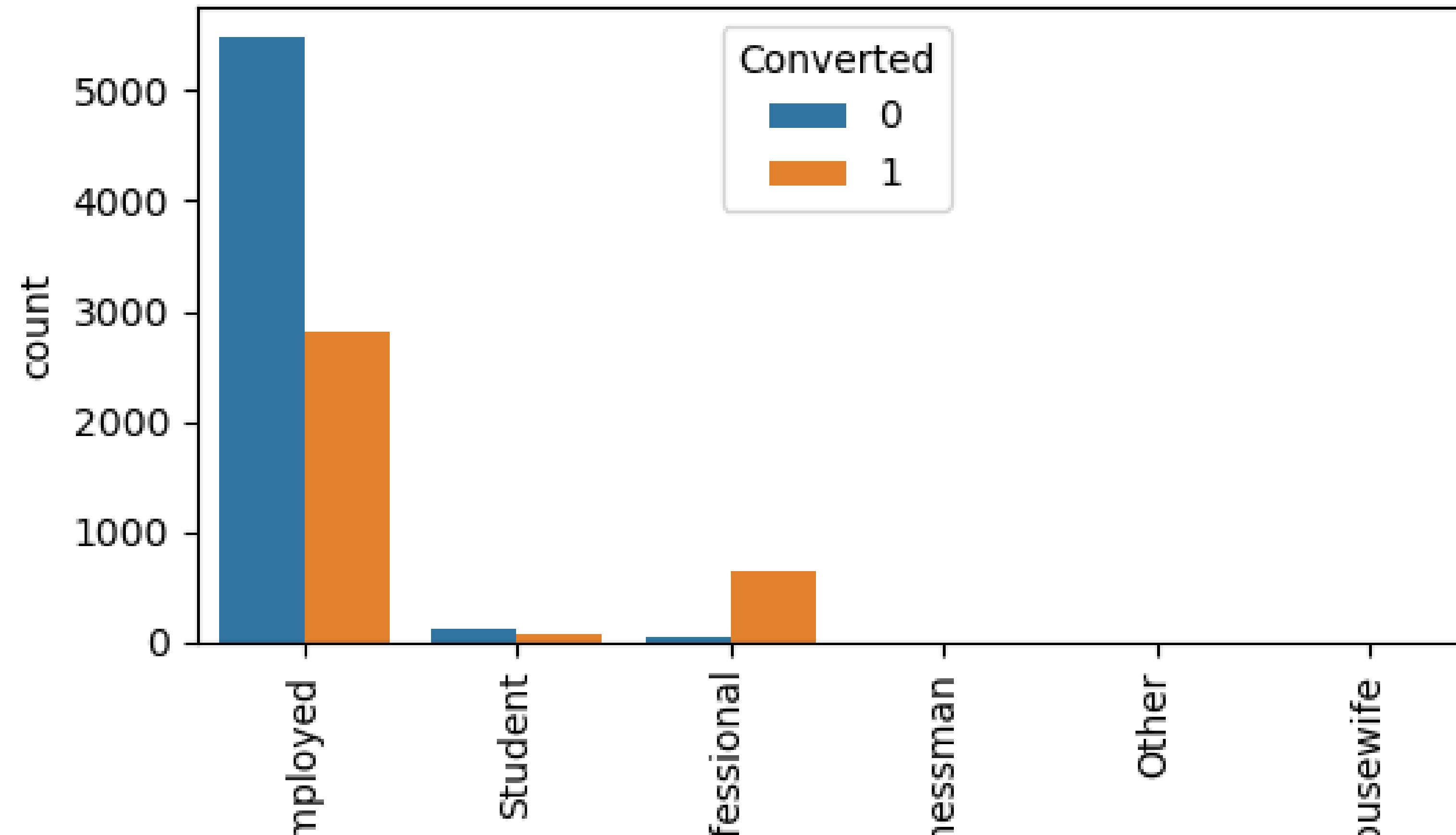
Analysis approach and Model Building-

- **EDA**- Columns with Null values are dropped.
- There are few columns with ‘Select’ value in it. Those values are replaced with Null
- Imputed all missing values in columns with appropriate values
- Performed Outlier analysis and treatment,Univariate analysis
- Dummy variable creation for categorical columns
- **Model Building and evaluation**- Building the model using RFE as Feature Selection method and dropping the columns with high p-value
- check the Variance Inflation Factor (VIF) for any correlation
- Checking the model accuracy
- Model validation using sensitivity, specificity using Confusion matrix
- Plotting ROC Curve

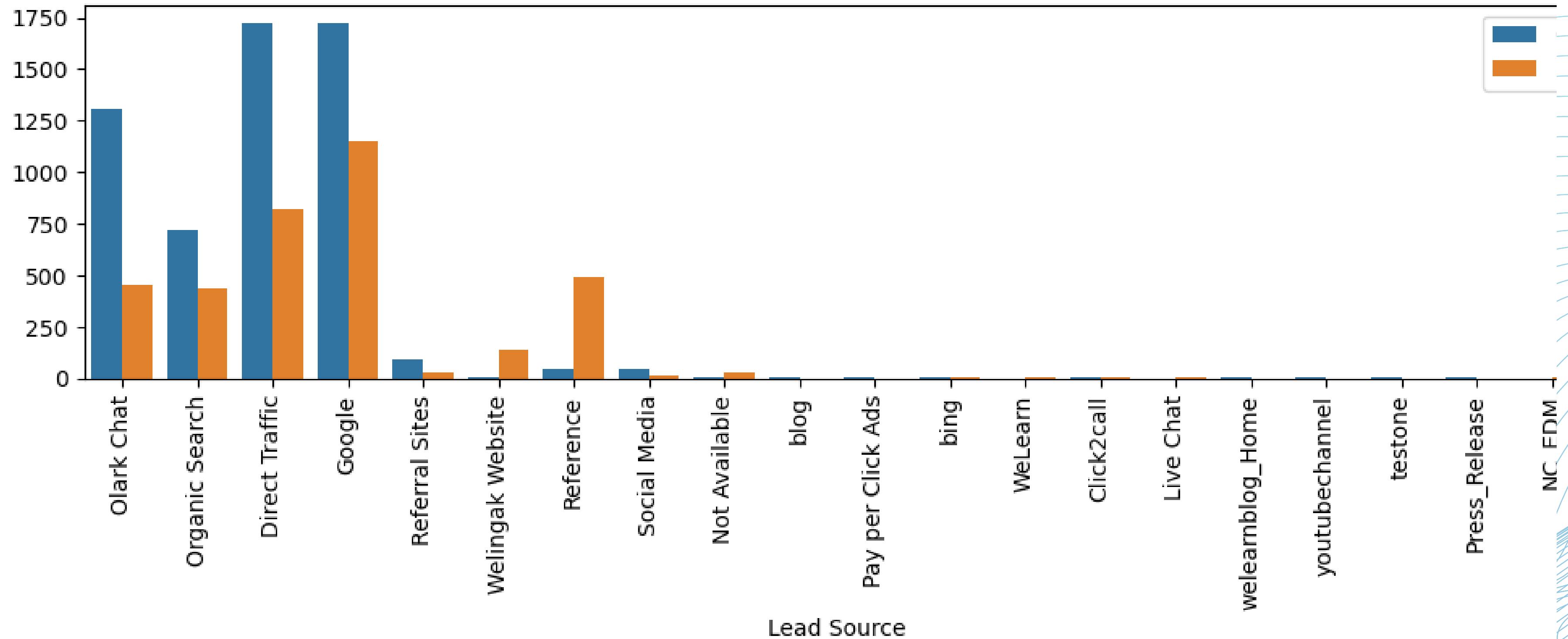
EDA-Management Columns- Conversion rate is high for Management specializations.



Occupation- Conversion rate is higher for Working professionals while more number of Unemployed applicants apply for course even though Conversion rate is not high.

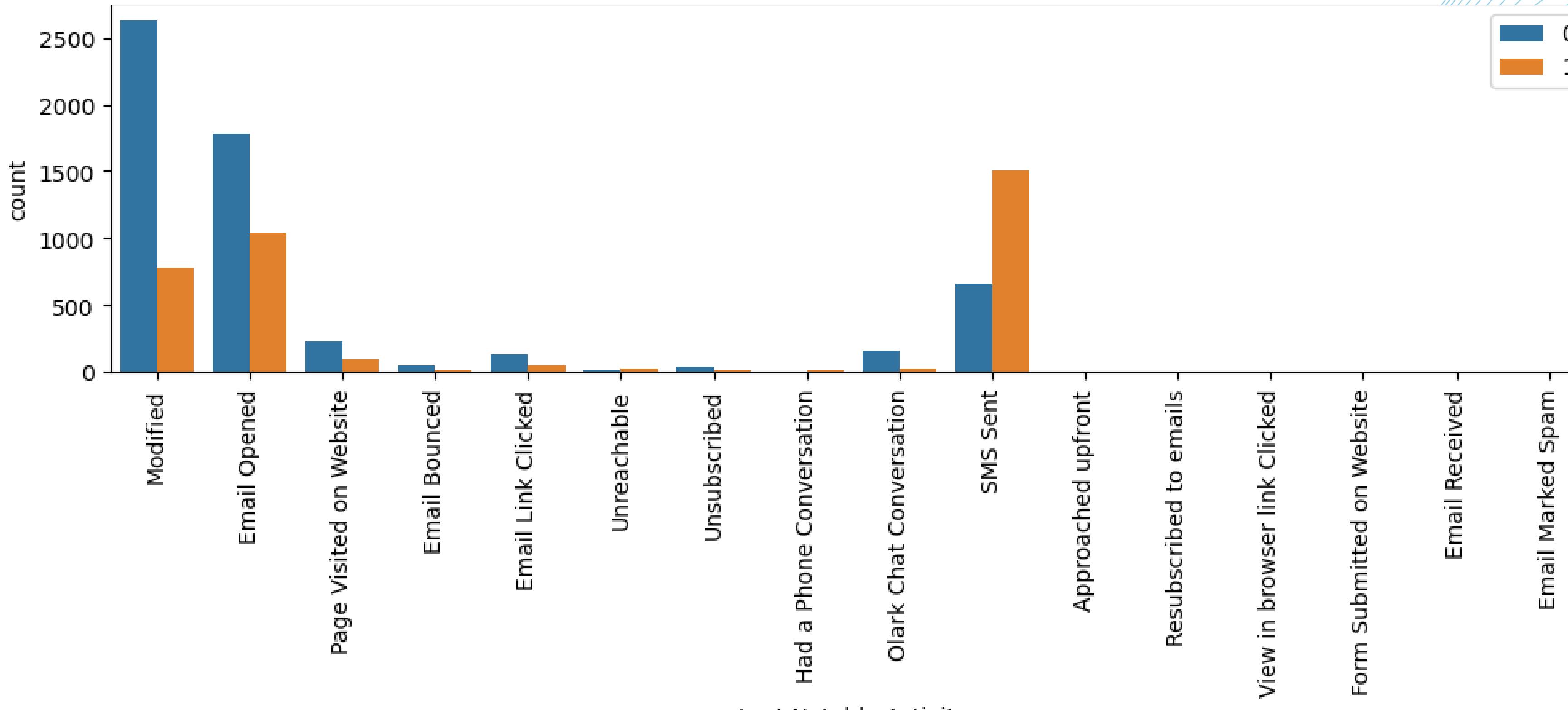


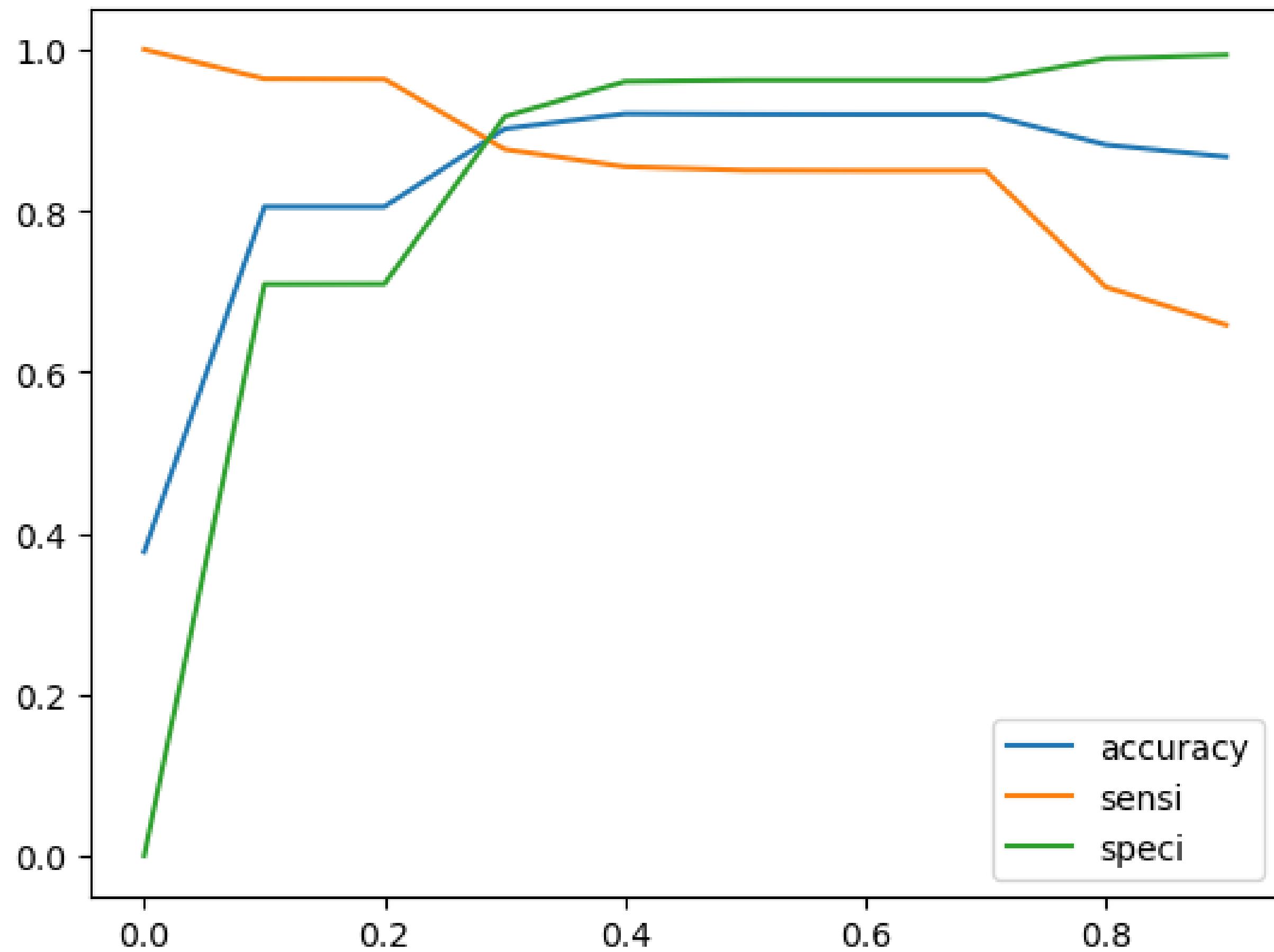
Lead Source- 1. Google and Direct traffic have the highest leads.
2. Conversion Rate of reference and welingak website is high.



Last Notable Activity- 1. API and Landing Landing Page Submission bring higher number of leads.

2. Lead Add Form has a high conversion rate even though count of leads is not high





ROC Curve- From the ROC curve, 0.3 is the optimum point to take it as a cutoff probability.

CONCLUSION-

Below are the variables that can increase leads:

- 1. Last Activity_SMS Sent**
- 2. Last Notable Activity_Modified**
- 3. Tags_Busy**
- 4. Tags_Closed by Horizzon**
- 5. Tags_Lost to EINS**
- 6. Tags_Not Available**
- 7. Tags_Ringing**
- 8. Tags_Will revert after reading the email**
- 9. Tags_in touch with EINS**
- 10. Tags_switched off**

The model is able to predict the Conversion Rate and can be used by X Education to target potential leads.

Thank
you!

