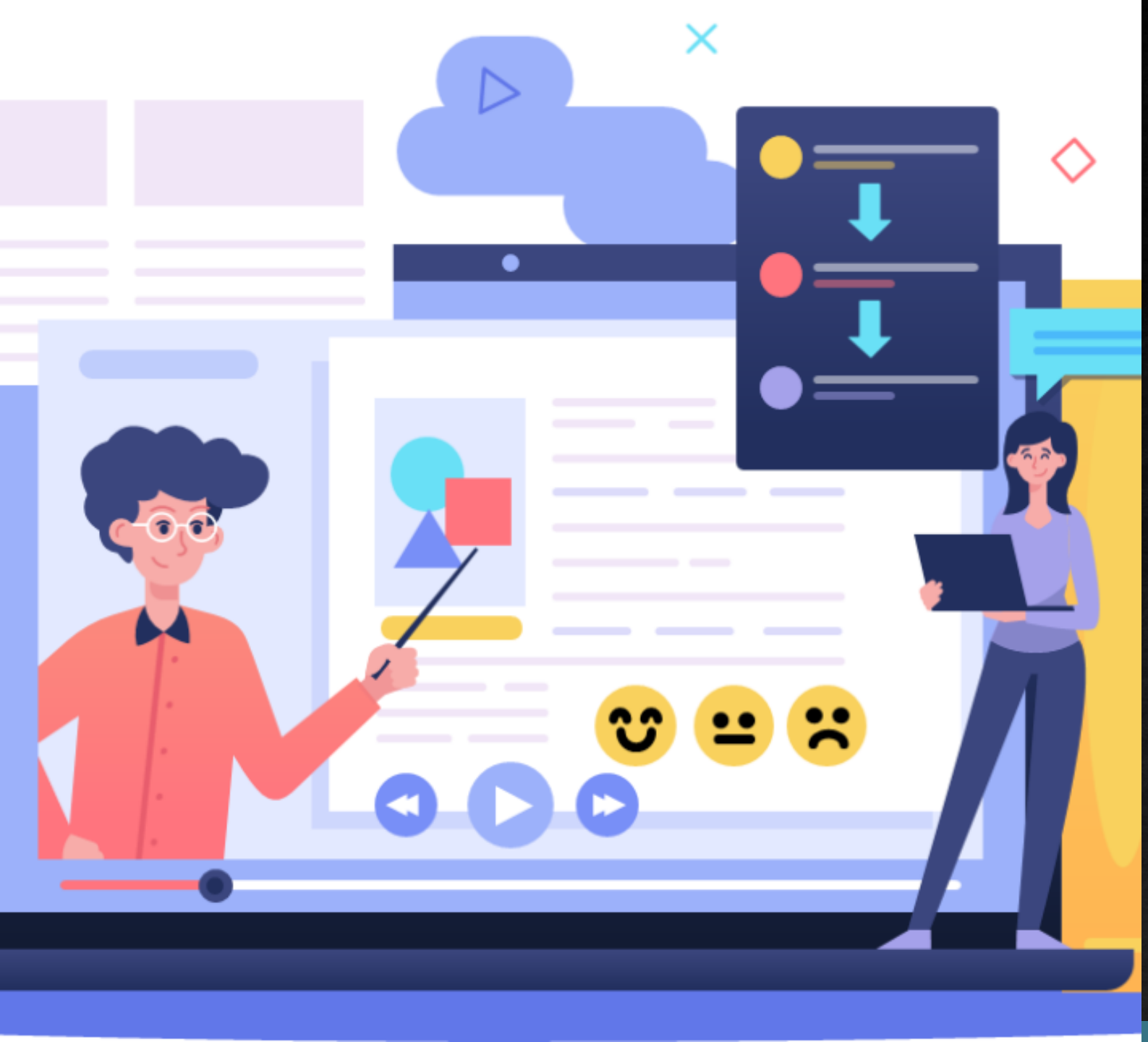


Education Company - Lead Scoring Case Study

By: Himanshu Yadav

Case Study **Contents**

Problem Statement

- X Education Company sells online courses to industry professionals. 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, has given a ballpark of the target lead conversion rate to be around 80%.

Approach

Data Understanding & Data Cleaning :

Handling missing values and 'Select' level.

Removing duplicate data and other redundancies. Outlier analysis & outlier treatment.

Checking skewness of data & Filtering columns as per requirements.

Data imbalance analysis.

Exploratory Data Analysis:

Univariate Analysis.

Bivariate Analysis.

Checking collinearity of data using correlation matrix.

Data Preparation:

Creating dummy variable.

Splitting data into Train and Test Sets.

Feature scaling.

Model Building:

Creating dummy variable.

Splitting data into Train and Test Sets.

Feature scaling.

Model Evaluation:

Predicting on Test set and evaluating the model on different evaluation metrics.

Finding Optimal Probability Threshold & Plotting ROC curve.

Calculating Cross Validation Score

Lead Scores Assigning:

Finalizing the model on basis of the model evaluation.

Using predicted probabilities to Calculate Lead Scores.

Leads Determination:

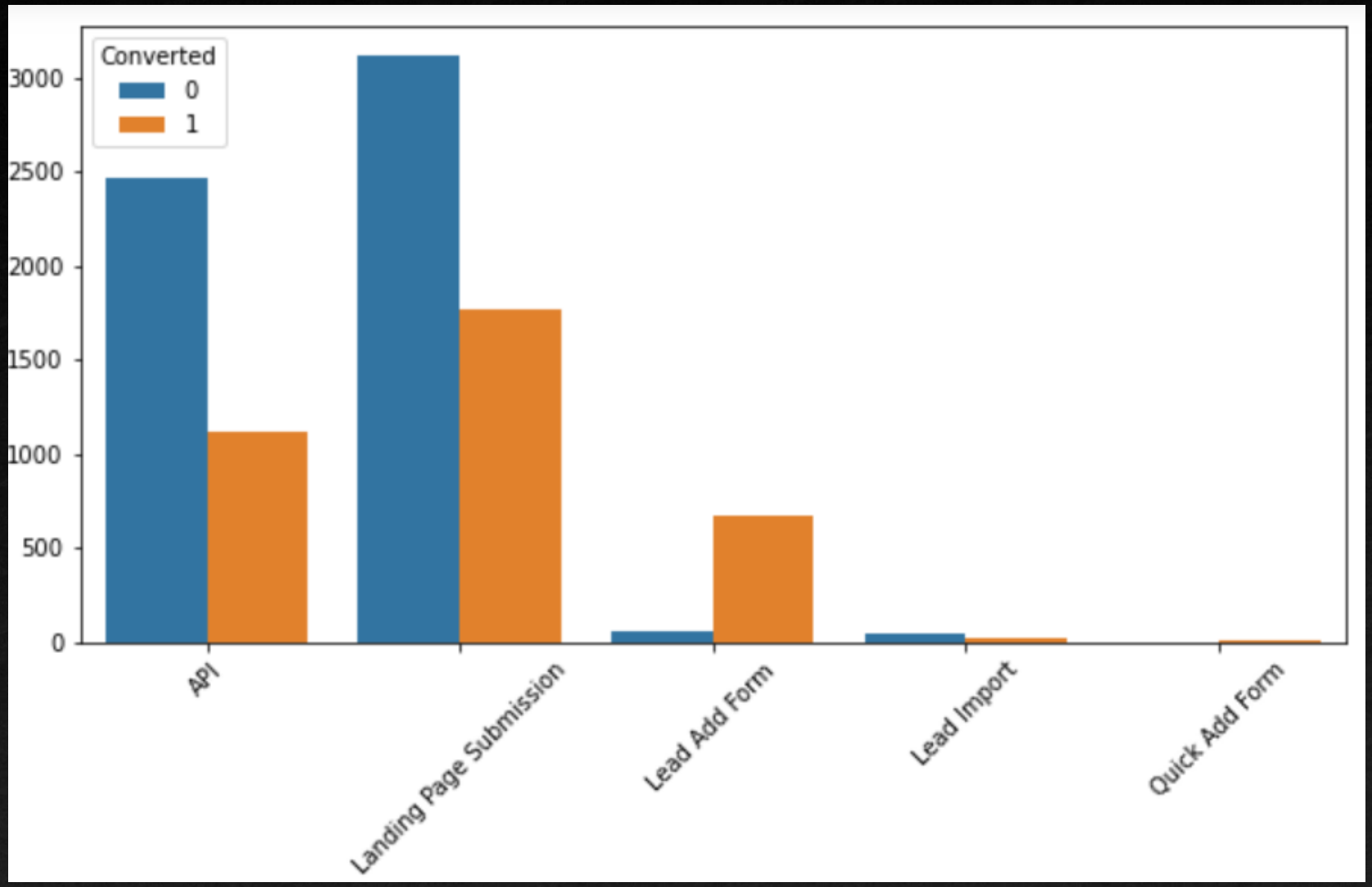
Determining hot & potential leads with more than 80% conversion rate and good accuracy.

Recommendation:

Finally recommending based on final model.

EDA

- Exploratory Data Analysis

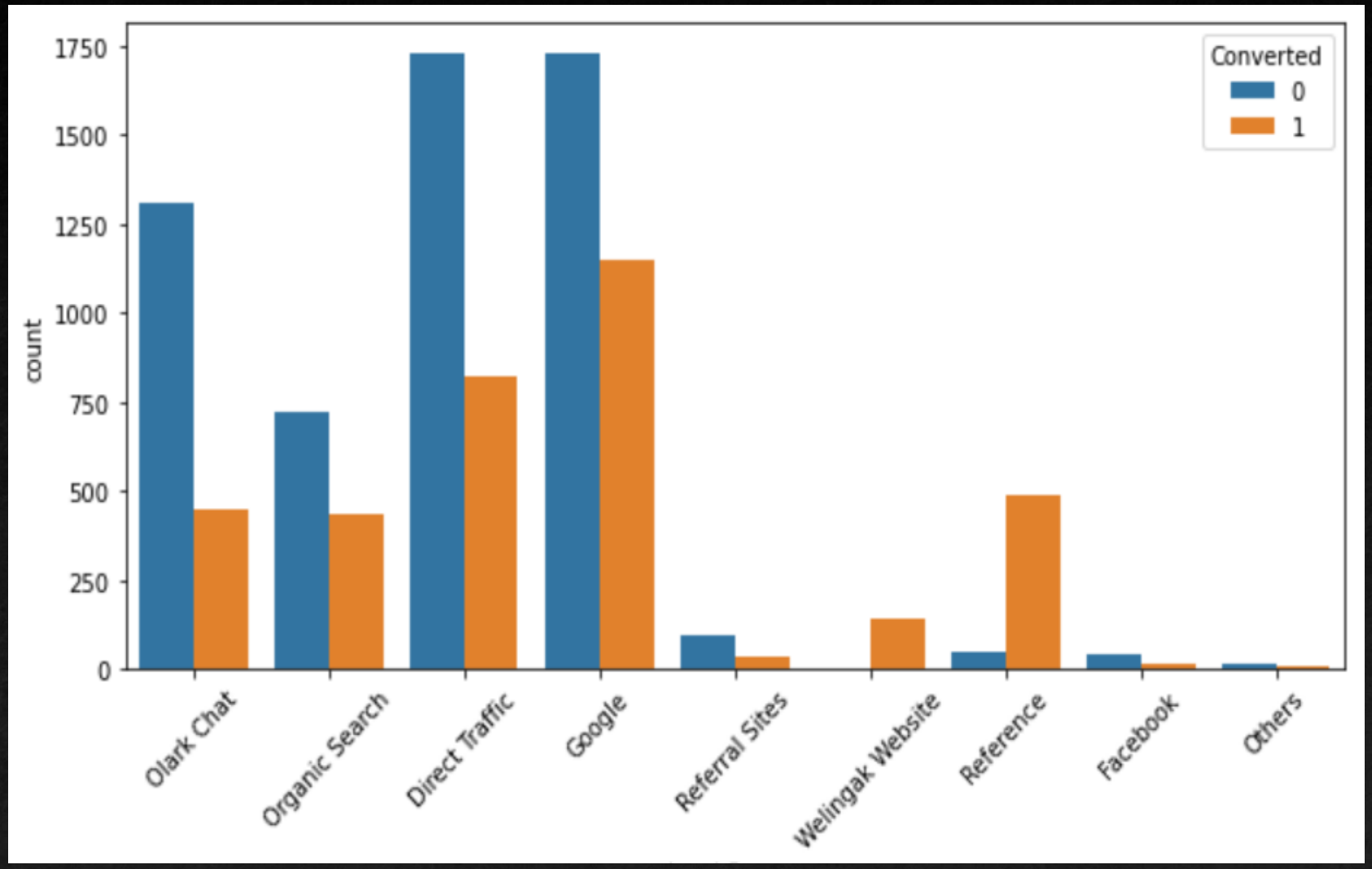


LEAD ORIGIN

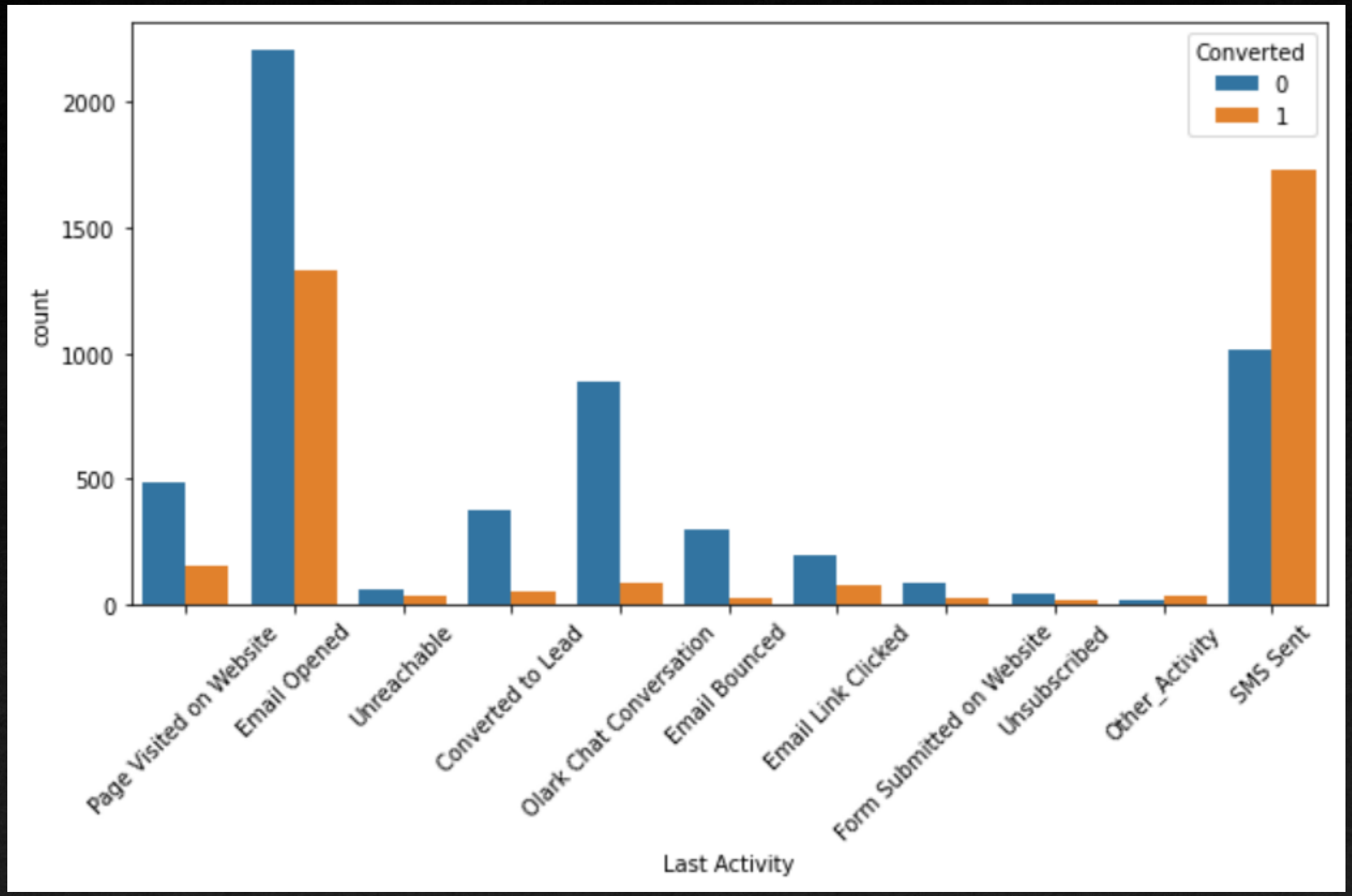
- Here ‘API’ and ‘Landing Page Submission’ generate the Most Leads but have less conversion rates. We need to focus on the increasing conversion rate for ‘API’ and ‘Landing Page Submission.
- ‘Lead Add Form’ generates few leads, but the conversion rate is high. We need to focus on Increasing leads generation using the ‘Lead Add Form’.

LEAD SOURCE

- We should focus on converting the leads of Olark Chat, Organic Search, Direct Traffic, Google. Very High Conversion Rates For Lead Sources ‘Reference’ and ‘Welingak Website’.



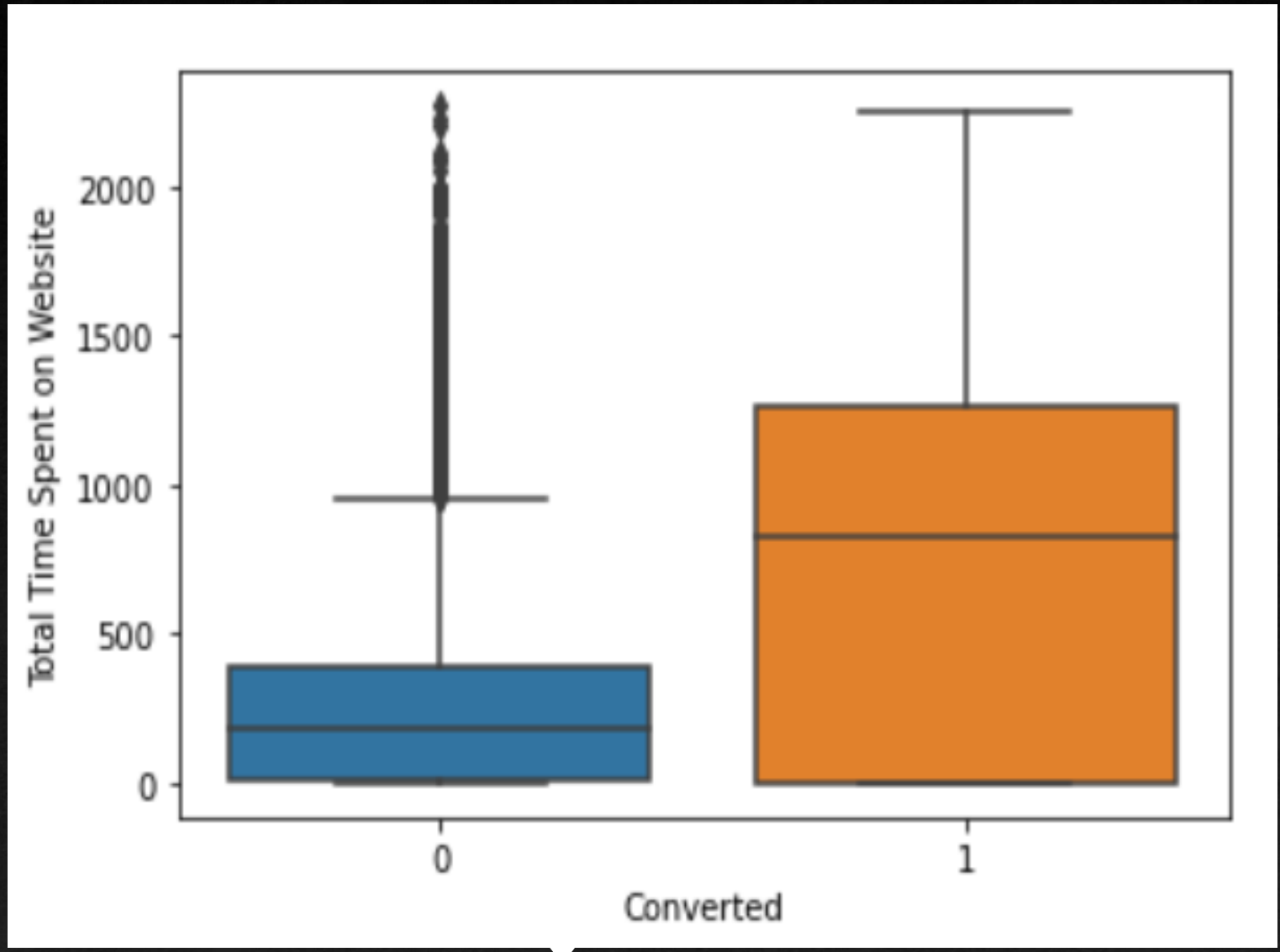
LAST ACTIVITY DONE



- The leads whose Last Activity was SMS sent had the Best Conversion Rate.

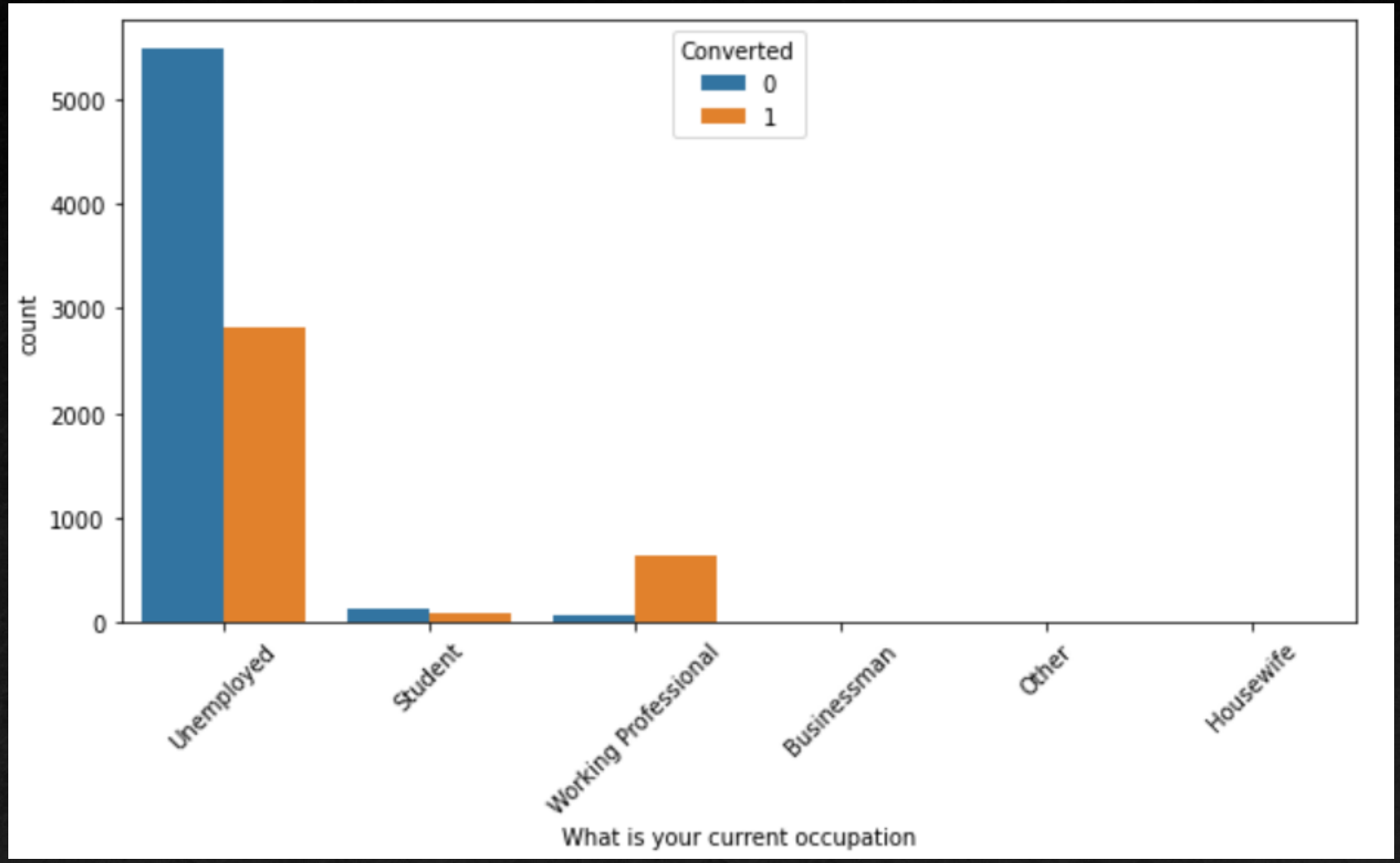
PEOPLE SPENDING MORE TIME ON WEBSITE

- People spending more time on website are more likely to convert. So, we can focus on website advertisement more.



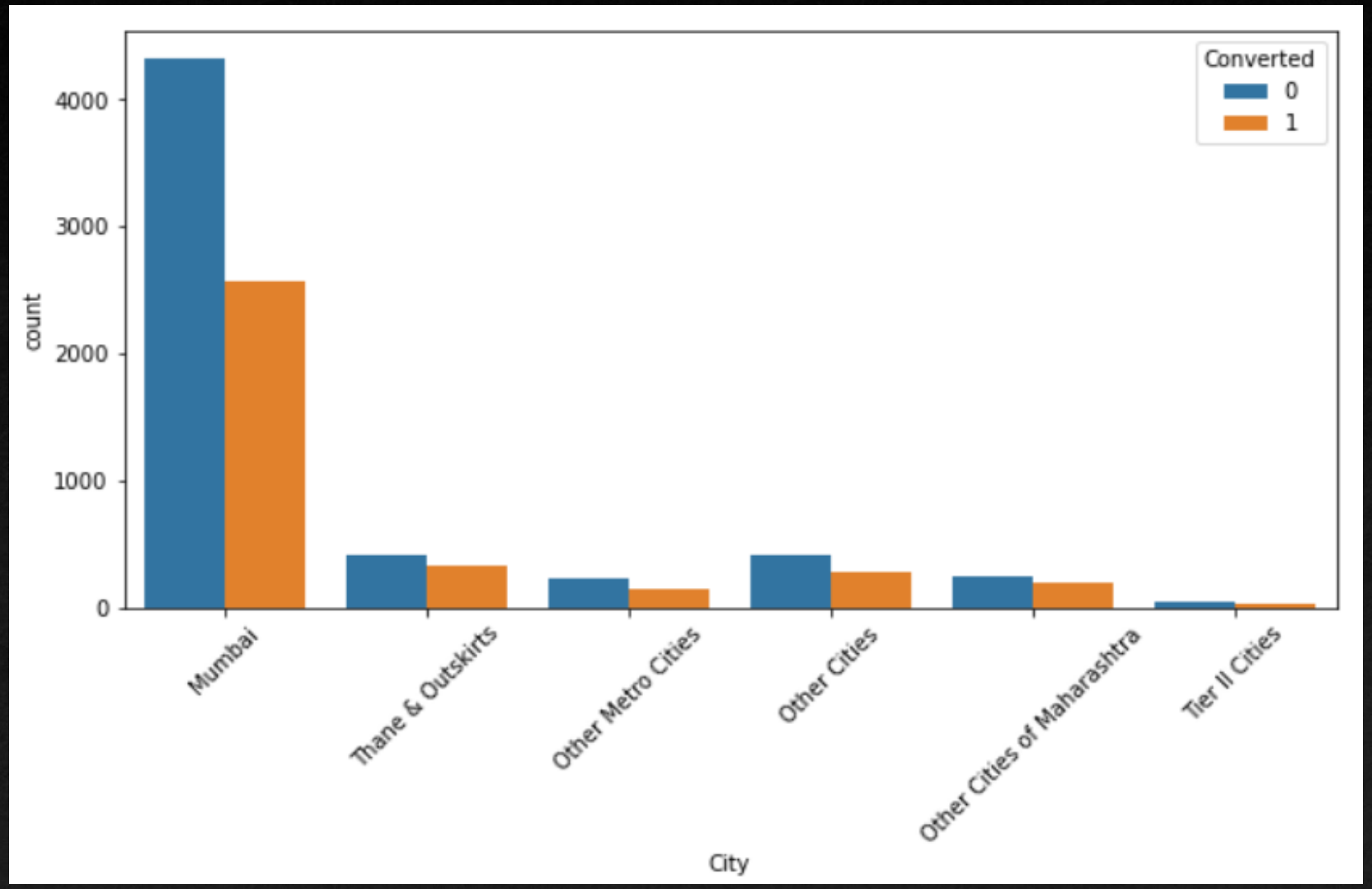
OCCUPATION

- Most of the Leads are Unemployed.
- Working Professionals are more easily converted.



LOCATION

- Mostly the lead conversions are from Mumbai.



MODEL

- Model Summary
- ROC Curve
- Finding Optimal Cut-off Probability
- Model Result

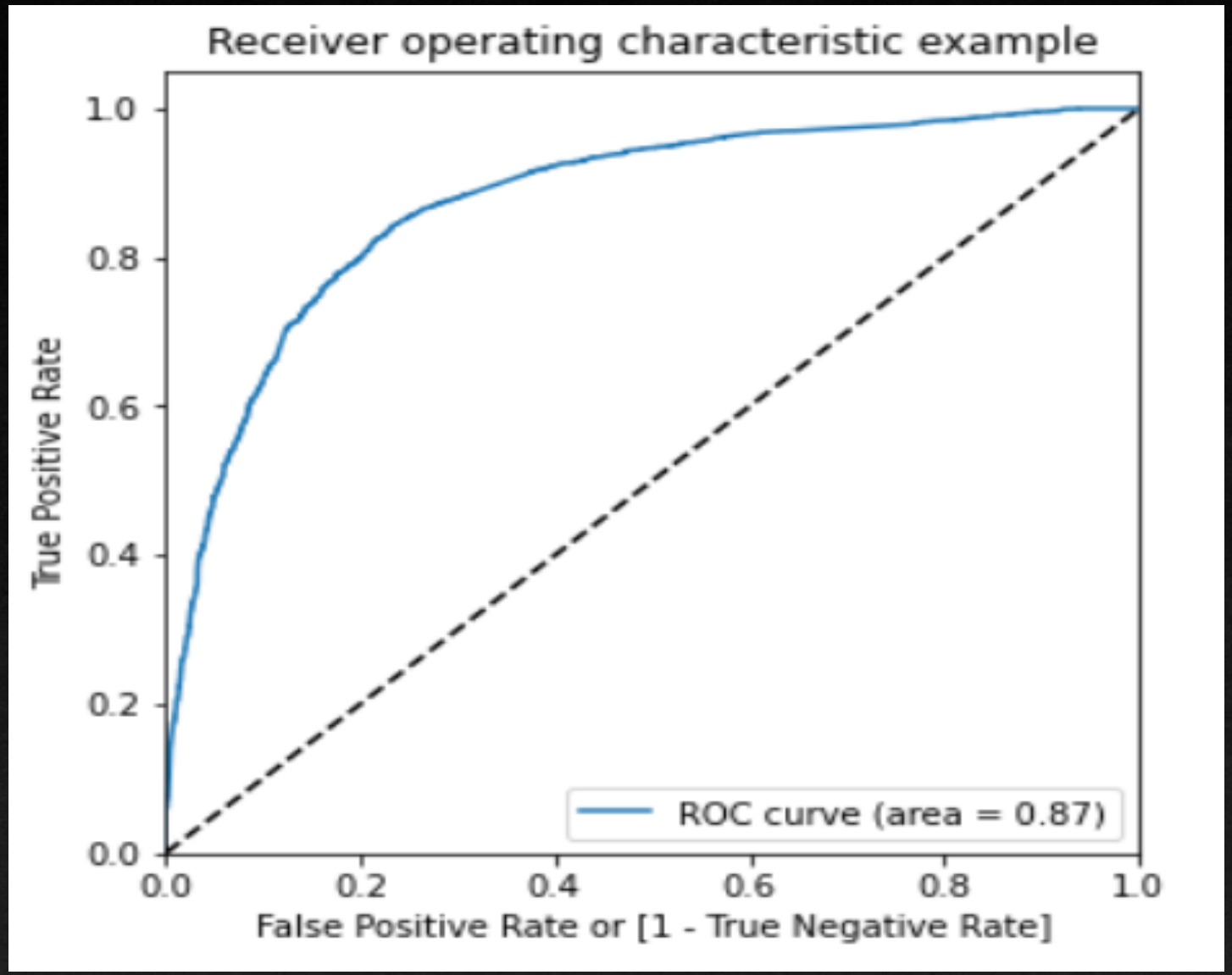
MODEL Summary

- All P-Values Are Zero, showing Significant Features Contributing towards Lead Conversion.

Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	6468
Model:	GLM	Df Residuals:	6449
Model Family:	Binomial	Df Model:	18
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2792.6
Date:	Fri, 26 May 2023	Deviance:	5585.3
Time:	18:16:42	Pearson chi2:	6.74e+03
No. Iterations:	6	Pseudo R-squ. (CS):	0.3724
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0181	0.127	-0.143	0.887	-0.267	0.231
Email	-1.1907	0.162	-7.351	0.000	-1.508	-0.873
Time Spent	1.0960	0.039	28.174	0.000	1.020	1.172
Lead Origin_API	1.1282	0.126	8.932	0.000	0.881	1.376
Lead Origin_Lead Add Form	3.8575	0.216	17.831	0.000	3.434	4.282
Lead Source_Direct Traffic	-1.2103	0.140	-8.660	0.000	-1.484	-0.936
Lead Source_Google	-0.8763	0.119	-7.335	0.000	-1.110	-0.642
Lead Source_Organic Search	-1.0343	0.141	-7.360	0.000	-1.310	-0.759
Lead Source_Referral Sites	-0.9749	0.304	-3.207	0.001	-1.571	-0.379
Last Activity_Other_Activity	2.2730	0.469	4.844	0.000	1.353	3.193
Last Activity_SMS Sent	1.3113	0.072	18.166	0.000	1.170	1.453
Specialization_Hospitality Management	-0.9276	0.315	-2.949	0.003	-1.544	-0.311
Specialization_Others	-1.4474	0.118	-12.224	0.000	-1.679	-1.215
Occupation_Businessman	-0.4888	1.154	-0.424	0.672	-2.751	1.773
Occupation_Other	-0.1089	0.837	-0.130	0.896	-1.749	1.531
Occupation_Student	0.2354	0.252	0.935	0.350	-0.258	0.729
Last Notable Activity_Modified	-1.0557	0.077	-13.737	0.000	-1.206	-0.905
Last Notable Activity_Olark Chat Conversation	-1.2833	0.330	-3.884	0.000	-1.931	-0.636
Last Notable Activity_Unreachable	1.4535	0.513	2.832	0.005	0.447	2.460

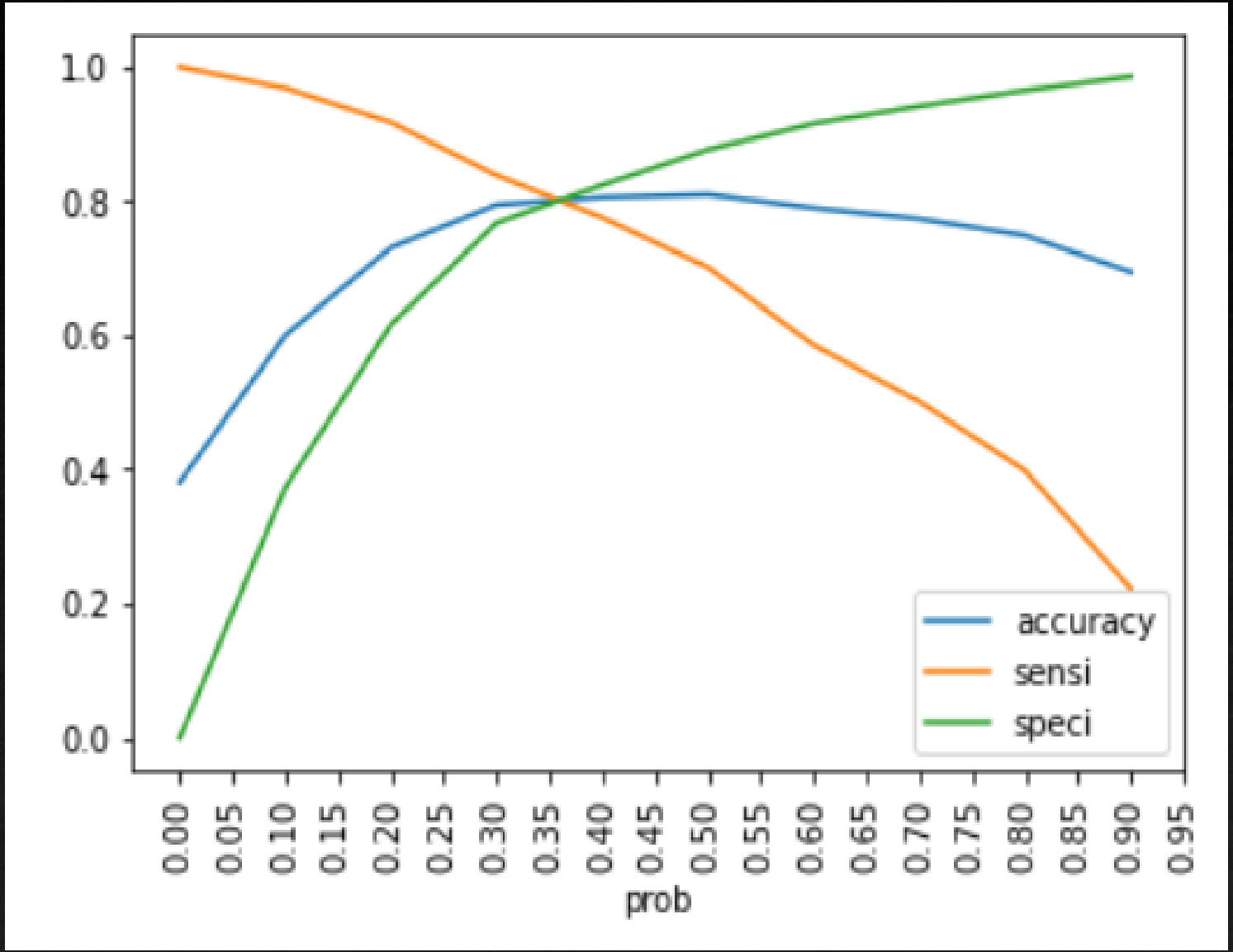


ROC CURVE

- The ROC Curve should be a value close to 1. Curve being on the upper left side of graph, it states that our model is good one.

FINDING OPTIMAL CUT-OFF POINT

- Optimal cut-off probability is the probability where we get the balanced sensitivity & specificity.
- From the curve 0.36 on x-axis, is the optimum point to take it as a cut-off probability.



RESULTS OF MODEL

- Thus, we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80%. The model seems to be predicted the conversion rate very well and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 80%

Train Data

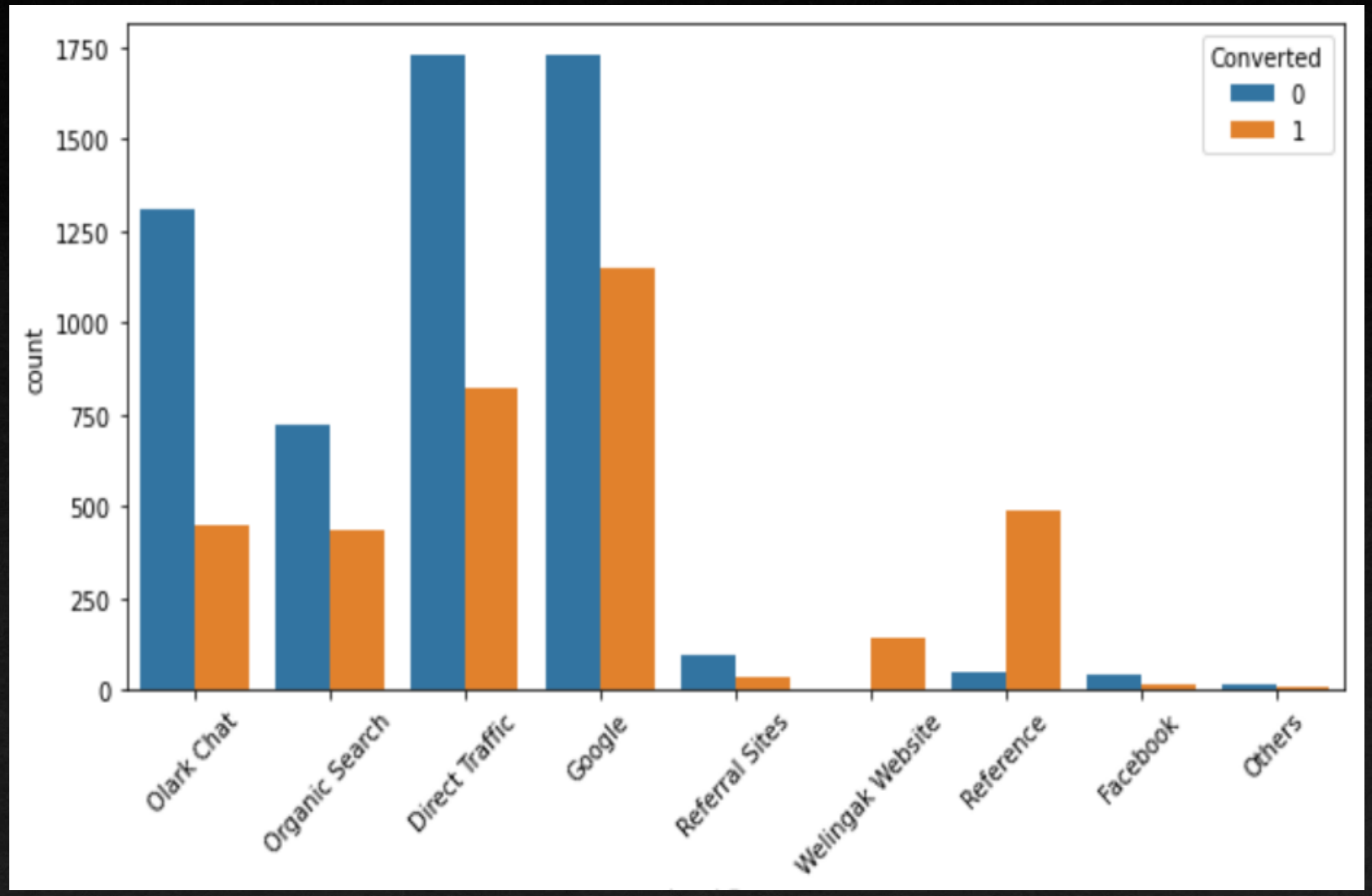
```
Accuracy : 80.1 %
Sensitivity : 79.8 %
Specificity : 80.2 %
```

Test Data

```
Accuracy : 80.8 %
Sensitivity : 74.8 %
Specificity : 84.7 %
```

LEAD SOURCE

- We should focus on converting the leads of Olark Chat, Organic Search, Direct Traffic, Google. Very High Conversion Rates For Lead Sources Reference and Welingak Website.



RECOMMENDATION

- The company should make calls to Working Professionals because they are more likely to get converted as leads.
- The website can be easy and reactive, and more informative to attract the people who are visiting website repeatedly or who spend much time on website.
- Through SMS and Emails the people can be targeted for conversion.
- The last and most important activity that can be done is conversation on Phone call.

THANK YOU