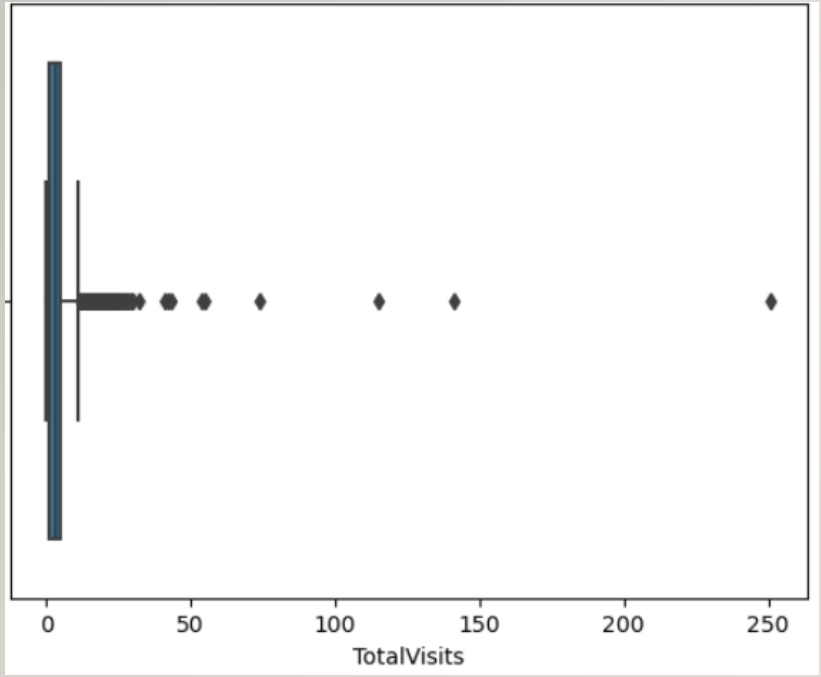
LEAD SCORE CASE STUDY

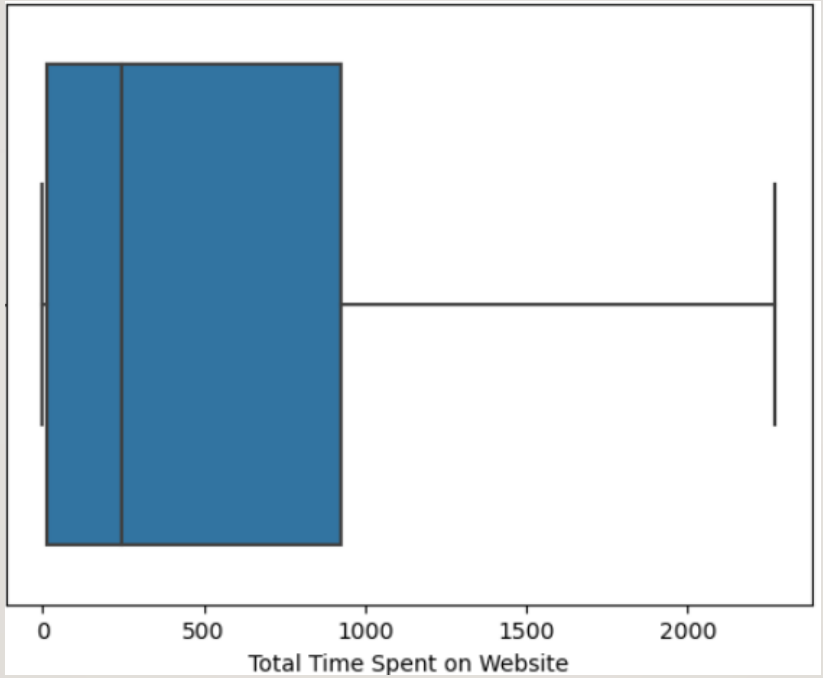
Group Members DS C56:-
1. Sweta Prasad
2. Alok Saxena
3. Navratna Ojha

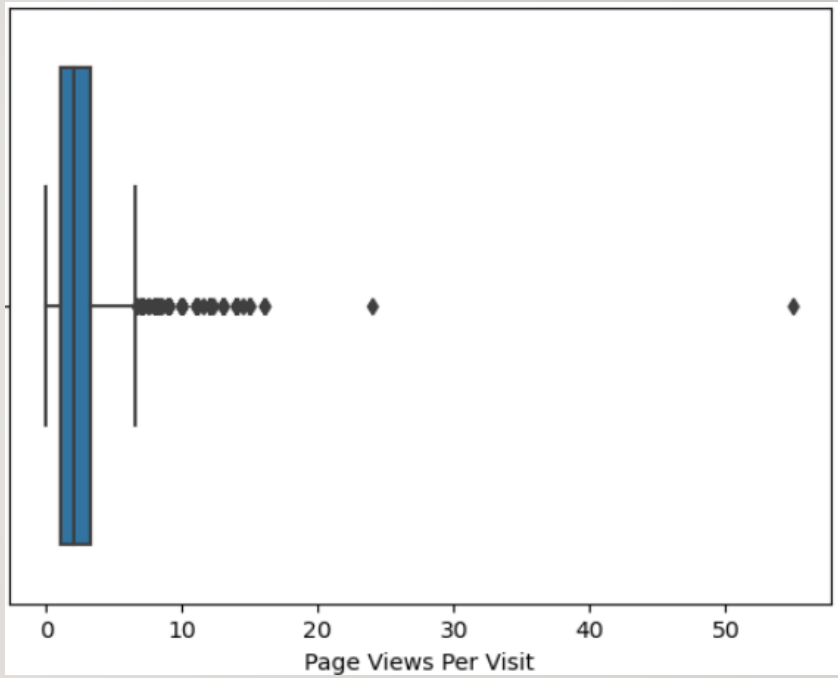
APPROACH

- Data Transformation:
  1. Drop columns with missing values >= 40%
  2. Handling all “Select” Values.
  3. Missing value imputation at individual columns.
  4. Drop irrelevant columns.
  5. Dropping some missing values.
- EDA Data
- Data Scaling & create Dummy Variables
- Classification technique: logistic regression
- Validation of the model.
- Model presentation.
- Conclusions

OUTLIER ANALYSIS







- Total Visits and Page views per visit have outliers
- We will cap outliers to 95% value for analysis

Conclusion

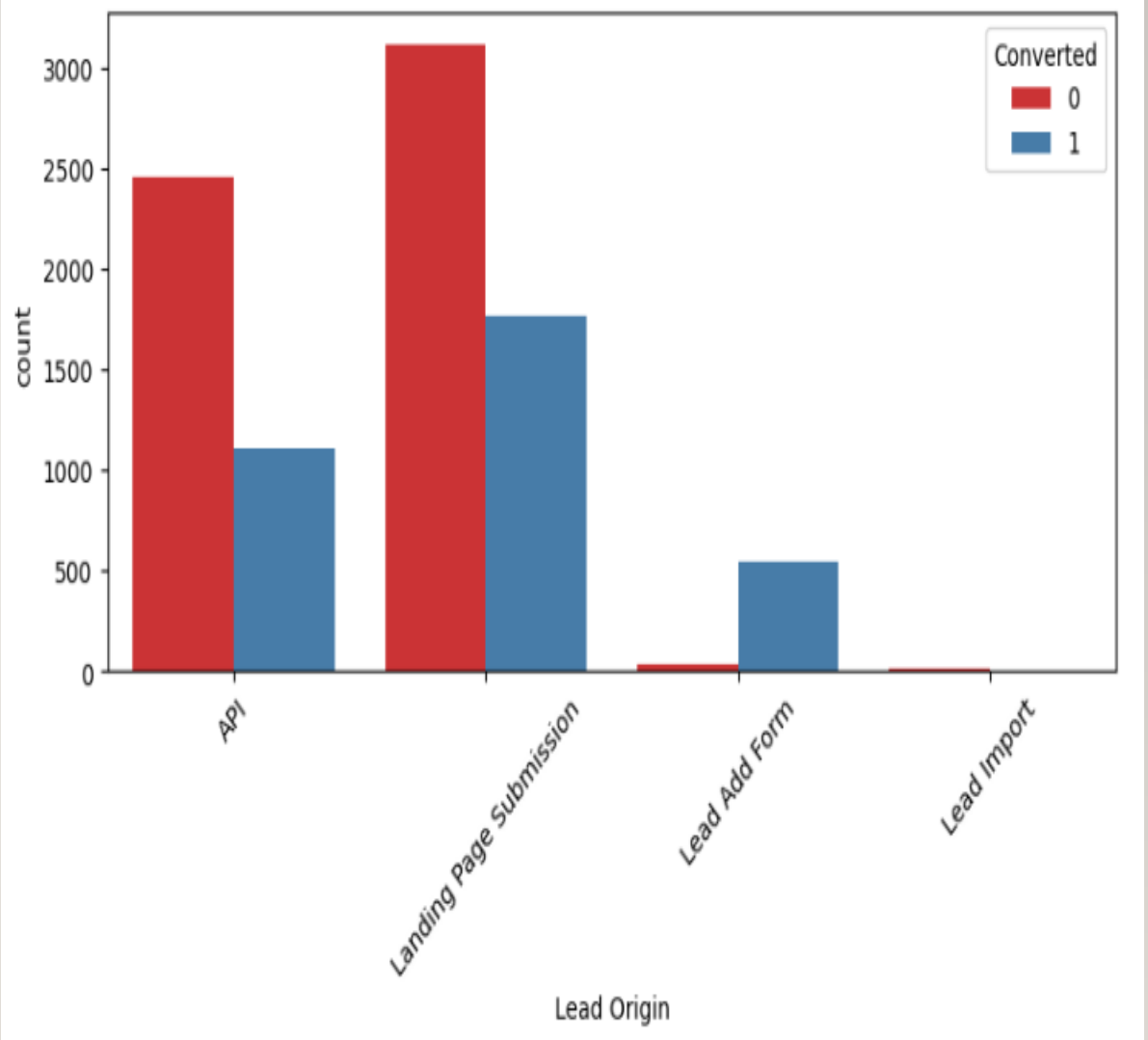
1.API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them seems considerable

2.Lead Add Form has more than 90% conversion rate but count of lead does not seem very high

3.Lead Import count seems very low

Recommendation

We need to focus on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form



Conclusion

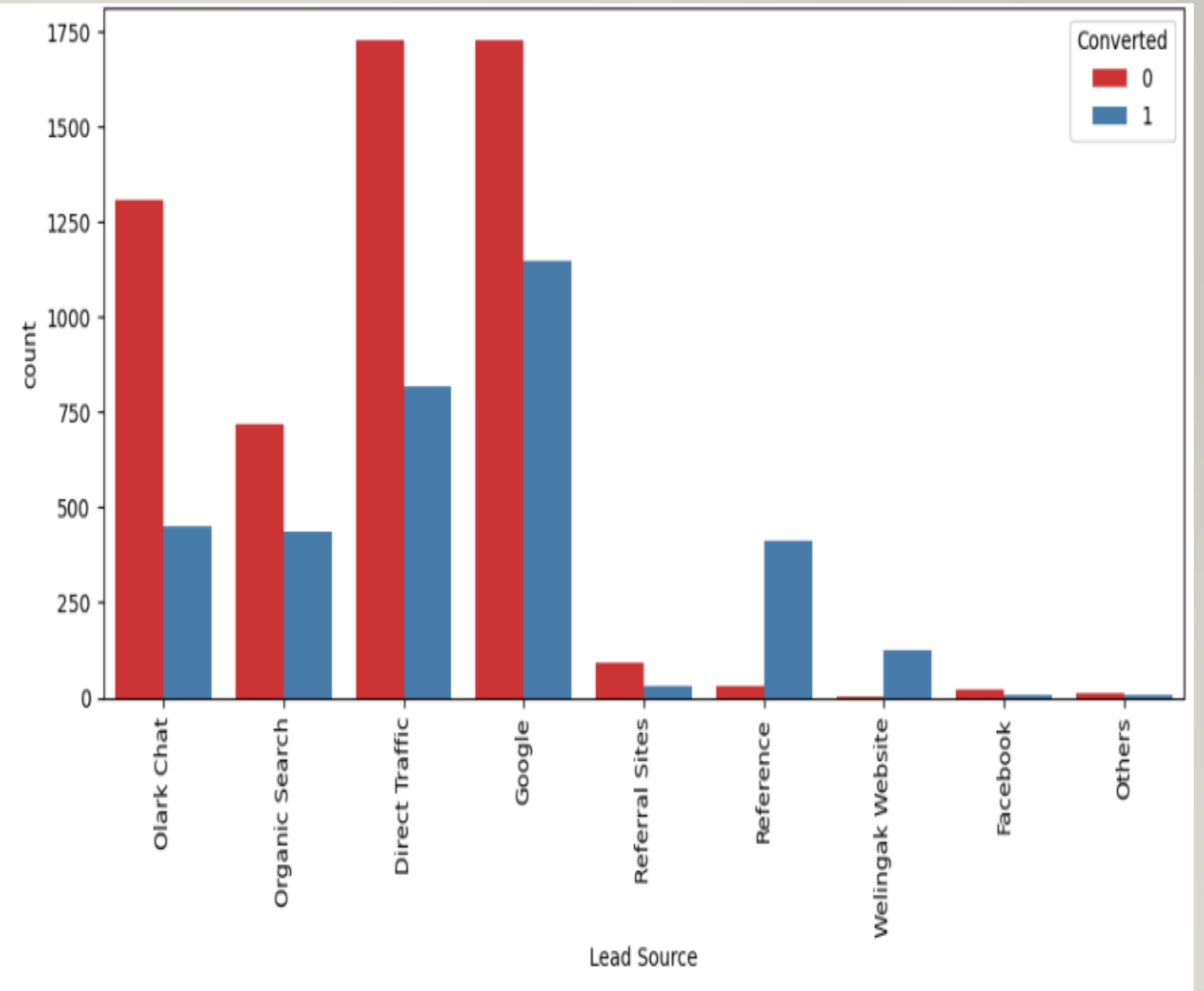
1.Maximum leads are generated by Google and Direct traffic

2.Conversion rate seems high for Reference leads and leads generated through Welingak website

Recommendation

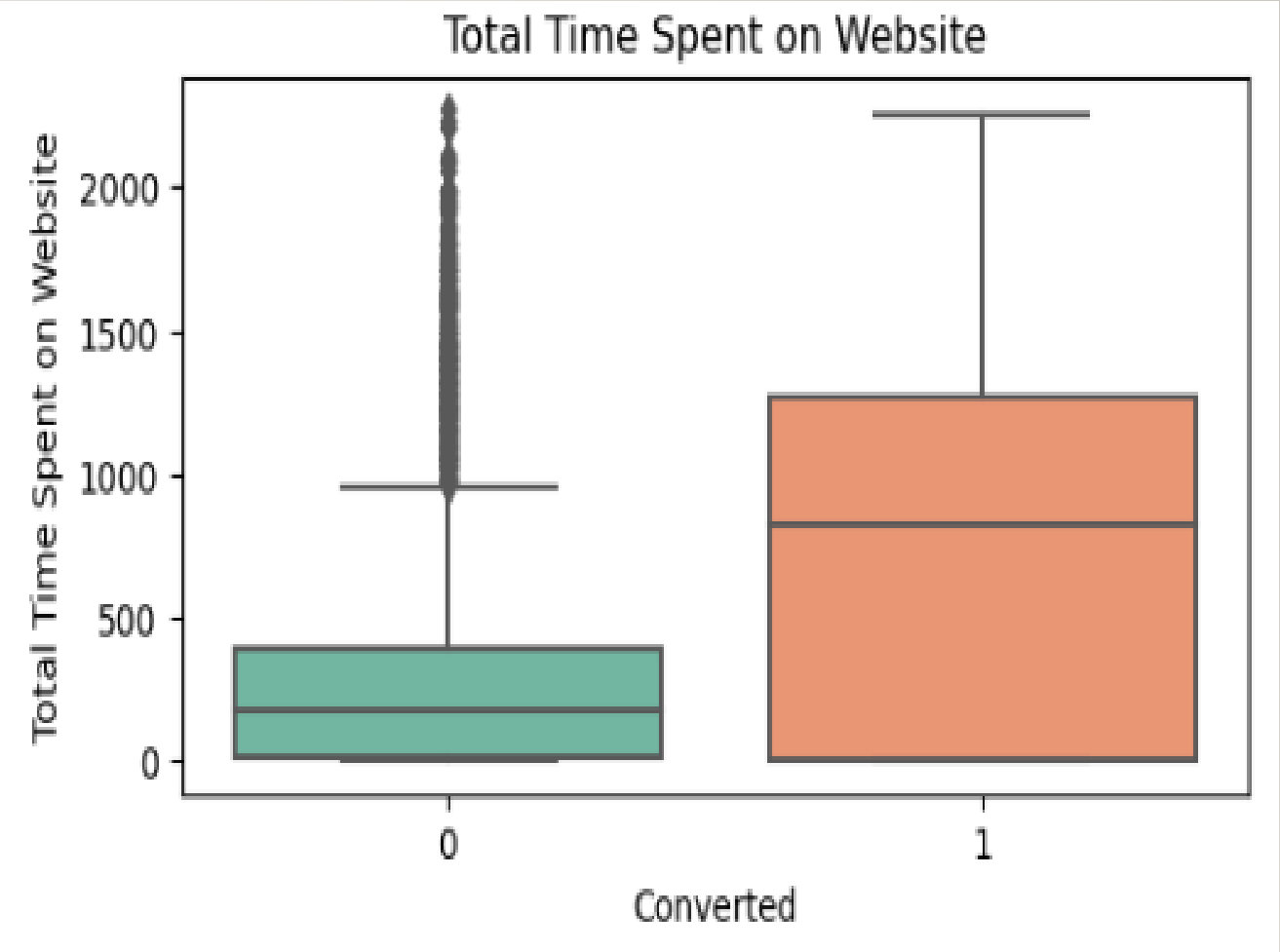
1. We need to focus on improving lead conversion of olark chat, organic search, direct traffic, and google leads.

2. We can also generate more leads from Reference leads and Welingak website. This will improve overall lead conversion rate.



Conclusion

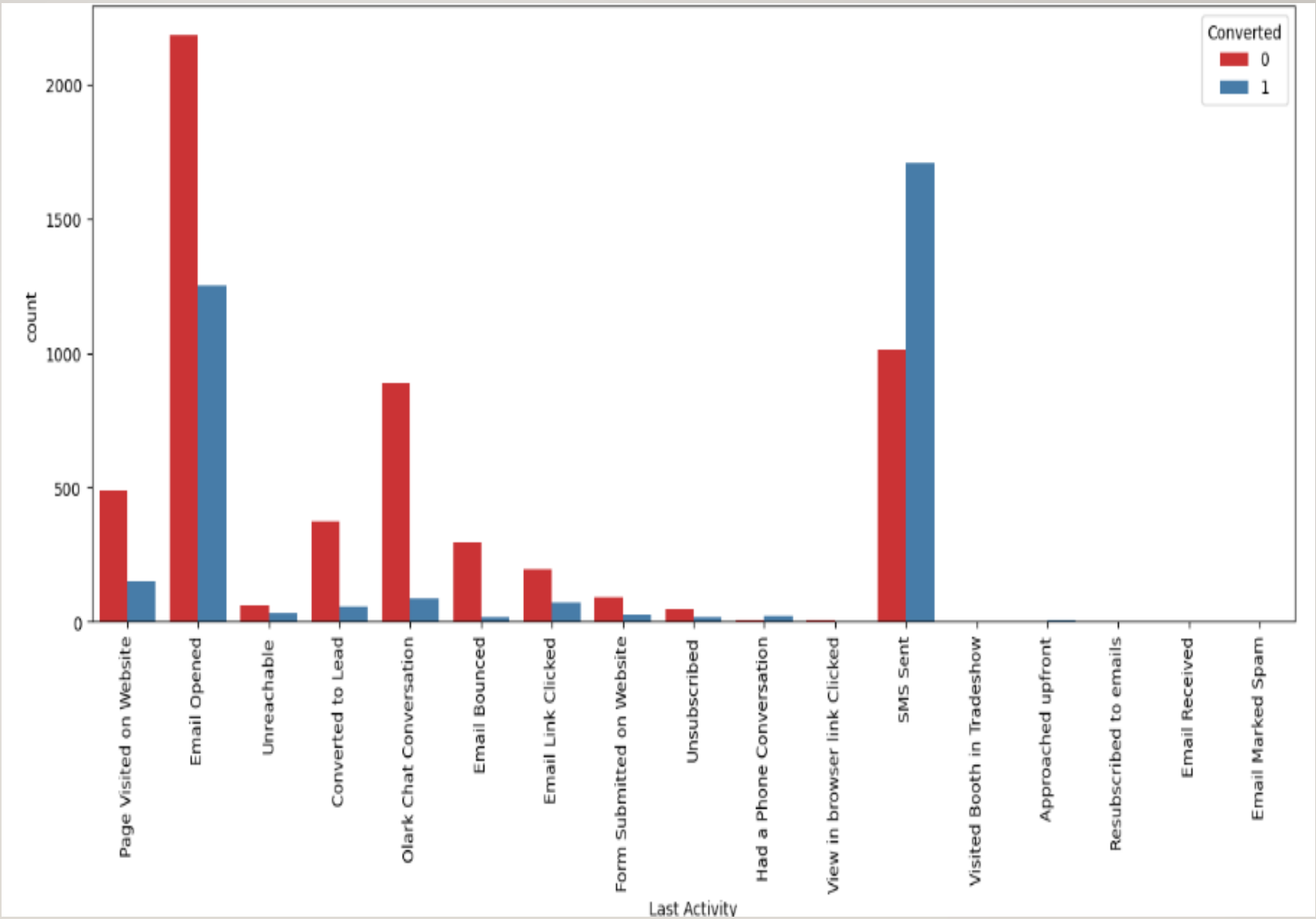
Leads spending more time on website seems to have high conversion rate



Conclusion

1.Conversion rate seems to be high for leads who have "Email Opened" as their last activity

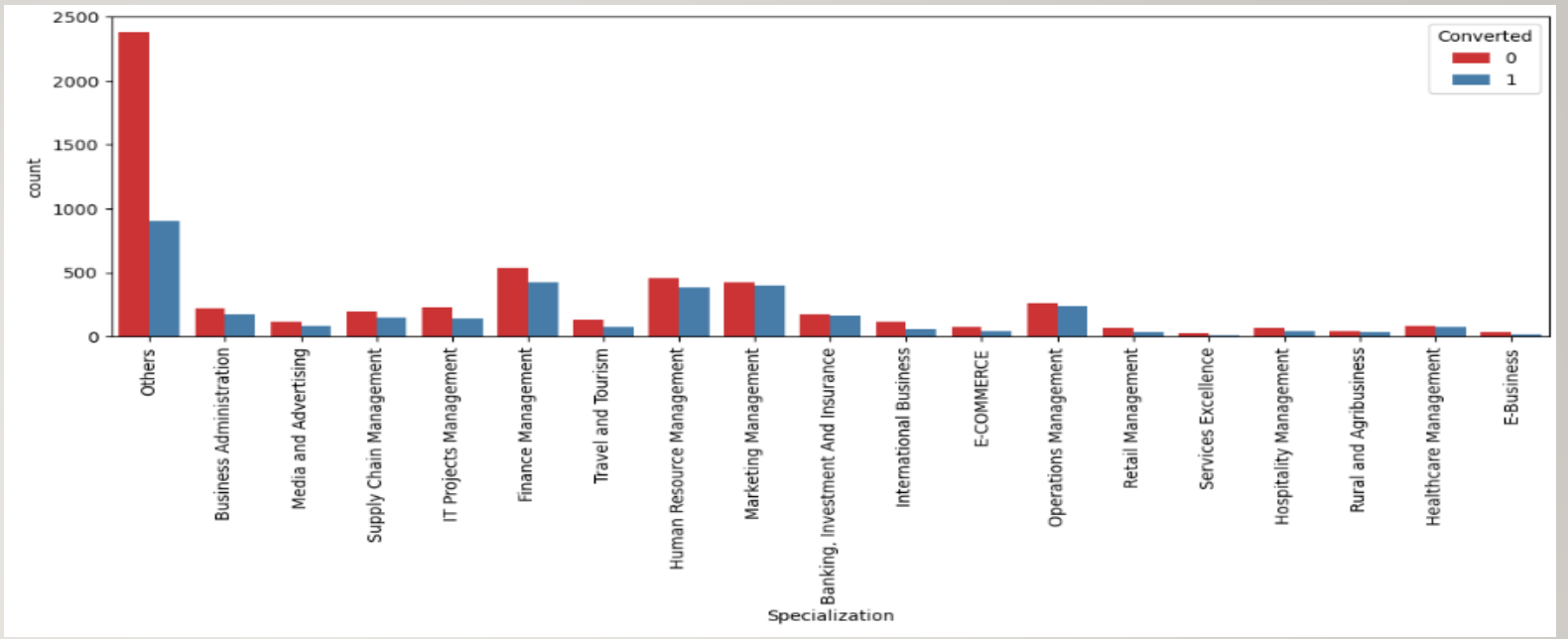
2.Conversion rate for leads with last activity as "SMS Sent" is significantly high compared to remaining activities

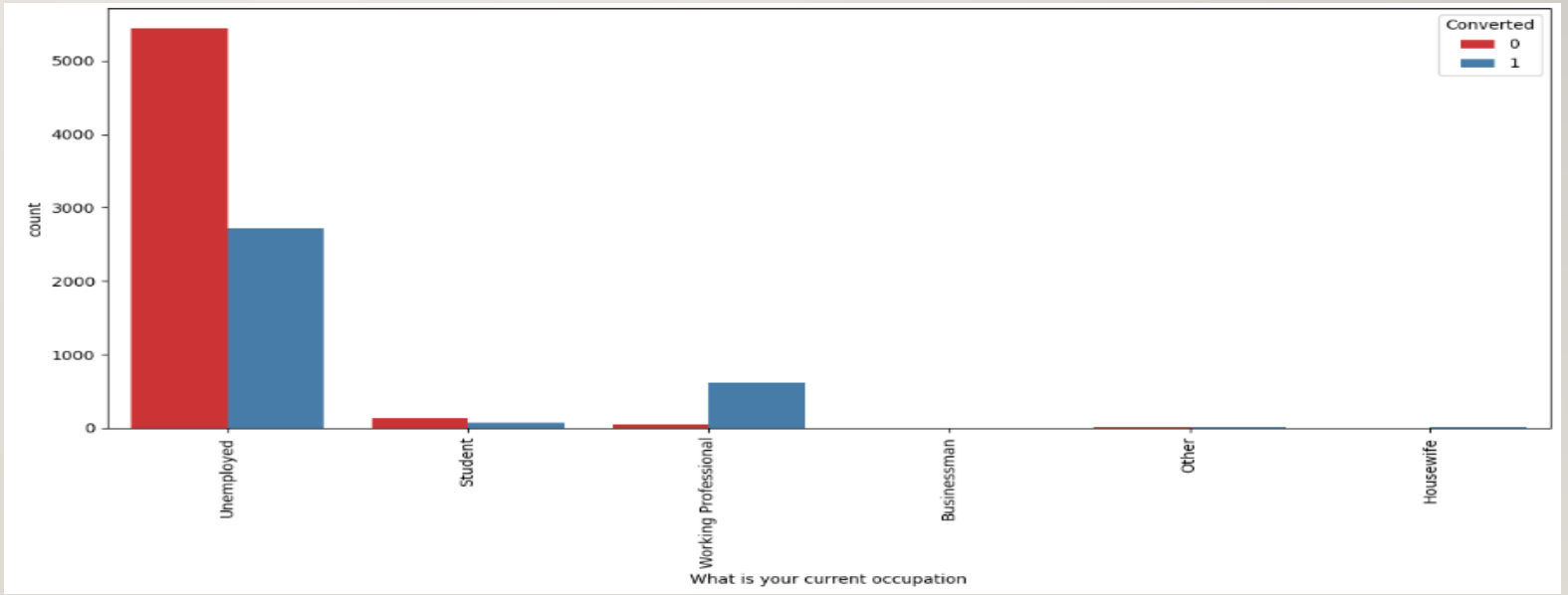


Conclusion

1.Working Professionals are more likely to join the course

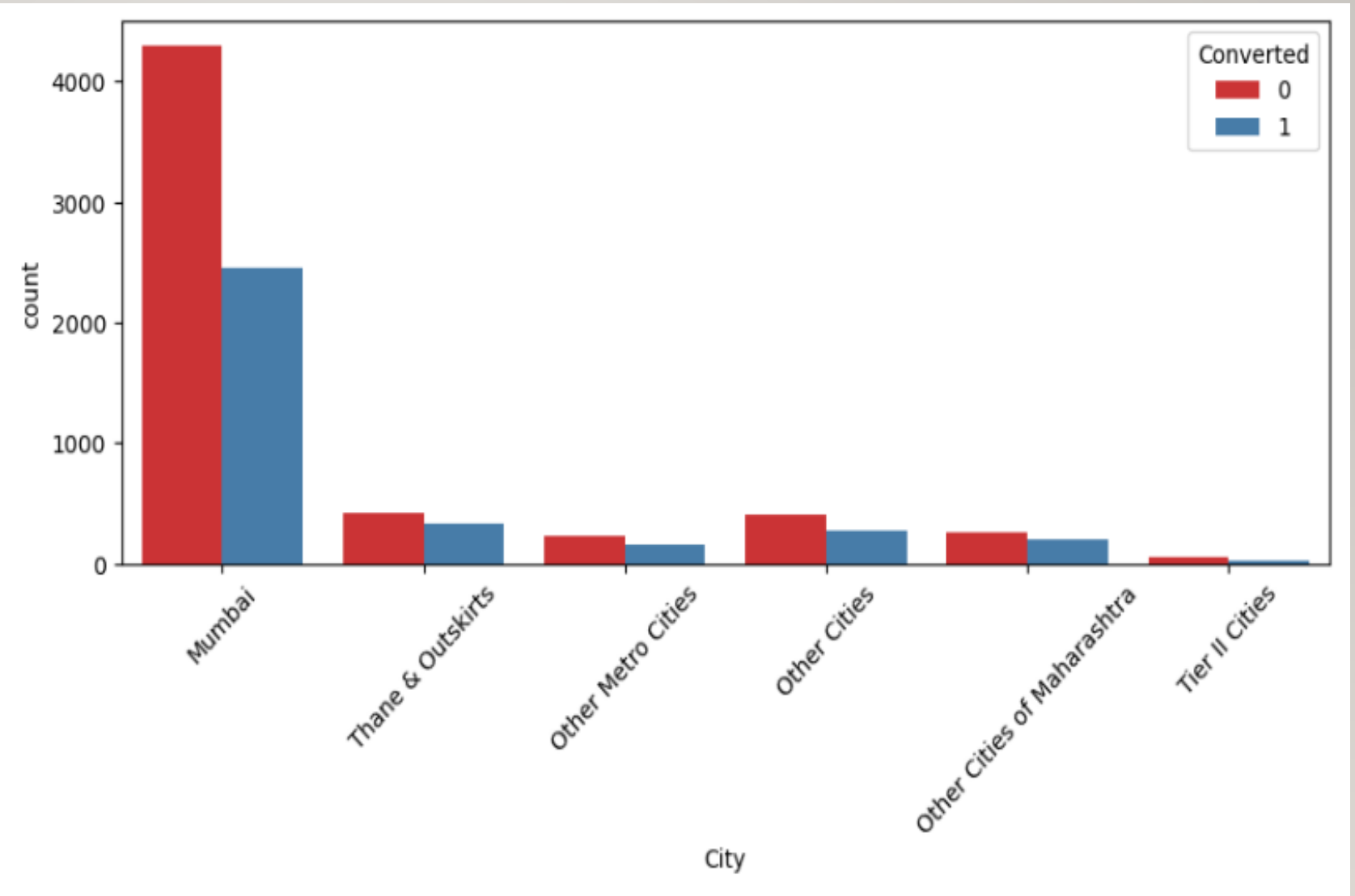
2.Unemployed users are maximum in number but have 2nd highest conversion rate.





Conclusion

1.Most Leads are from 'Mumbai' and have approx. 50% conversion rate



MODEL SELECTION

This is the model finalized. We can see the p-values of all variables is 0 and VIF values are low for all the variables, model-6 is our final model. We have 15 variables in our final model.

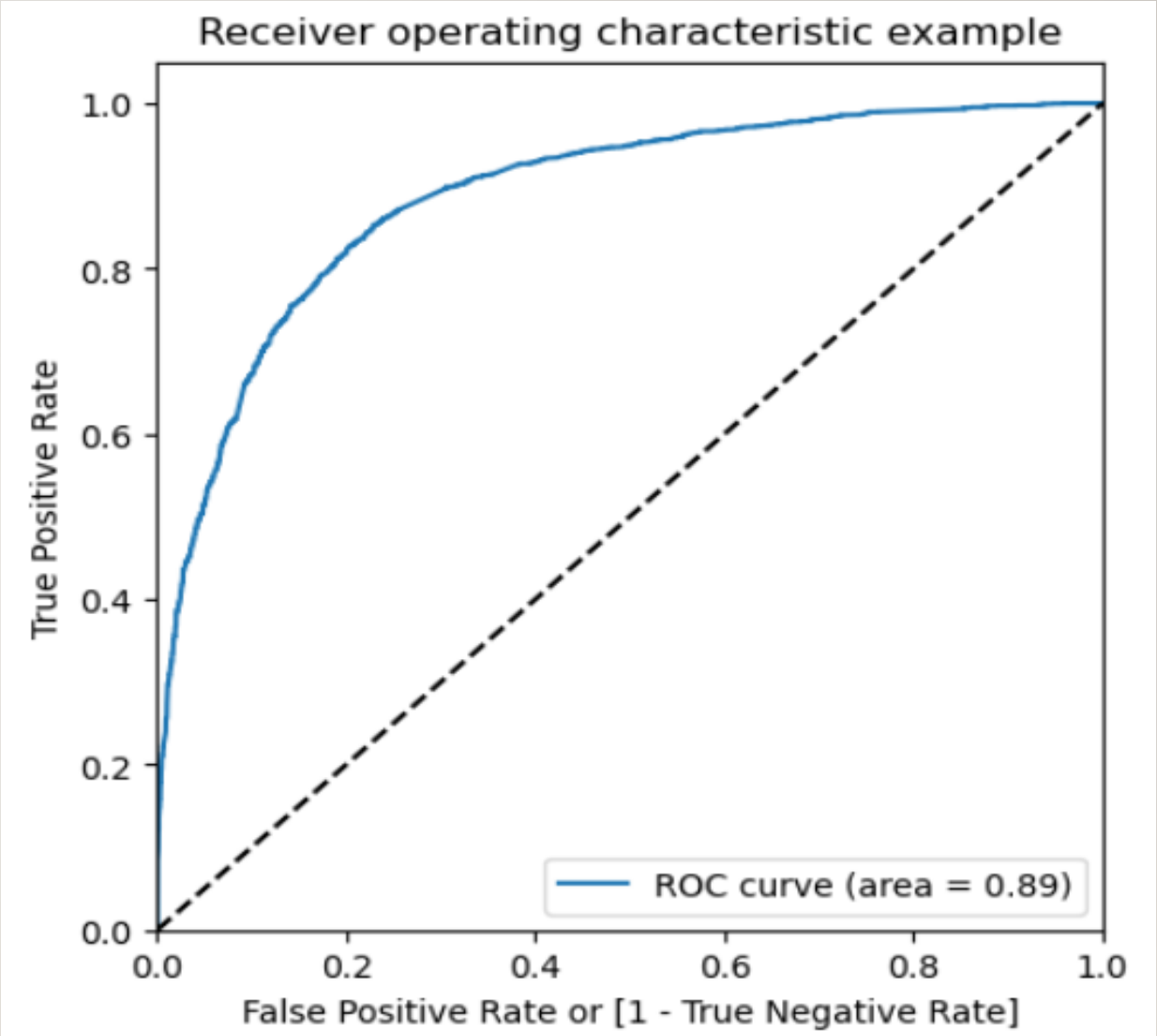
	Features	VIF
12	Last Notable Activity_Modified	2.67
8	Specialization_Others	2.46
2	Lead Origin_Landing Page Submission	2.36
3	Lead Source_Olark Chat	2.08
7	Last Activity_Olark Chat Conversation	2.01
11	Last Notable Activity_Email Opened	1.87
13	Last Notable Activity_Olark Chat Conversation	1.37
1	Total Time Spent on Website	1.27
6	Last Activity_Converted to Lead	1.23
4	Lead Source_Reference	1.21
0	Do Not Email	1.19
9	What is your current occupation_Working Profes...	1.17
14	Last Notable Activity_Page Visited on Website	1.10
5	Lead Source_Welingak Website	1.08
10	Last Notable Activity_Email Link Clicked	1.06

Dep. Variable:	Converted	**No. Observations:**	6351
Model:	GLM	**Df Residuals:**	6335
Model Family:	Binomial	**Df Model:**	15
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2596.6
Date:	Mon, 20 Nov 2023	**Deviance:**	5193.2
Time:	11:57:56	**Pearson chi2:**	6.47e+03
No. Iterations:	7	**Pseudo R-squ. (CS):**	0.4027
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.3480	0.136	9.876	0.000	1.080	1.616
Do Not Email	-1.9286	0.178	-10.834	0.000	-2.277	-1.580
Total Time Spent on Website	1.0972	0.040	27.133	0.000	1.018	1.176
Lead Origin_Landing Page Submission	-1.1685	0.127	-9.177	0.000	-1.418	-0.919
Lead Source_Olark Chat	1.0576	0.123	8.628	0.000	0.817	1.298
Lead Source_Reference	3.3544	0.243	13.803	0.000	2.878	3.831
Lead Source_Welingak Website	5.9651	0.736	8.101	0.000	4.522	7.408
Last Activity_Converted to Lead	-1.0493	0.229	-4.592	0.000	-1.497	-0.601
Last Activity_Olark Chat Conversation	-1.3281	0.196	-6.791	0.000	-1.711	-0.945
Specialization_Others	-1.1560	0.125	-9.234	0.000	-1.401	-0.911
What is your current occupation_Working Professional	2.6087	0.195	13.379	0.000	2.227	2.991
Last Notable Activity_Email Link Clicked	-1.9002	0.256	-7.433	0.000	-2.401	-1.399
Last Notable Activity_Email Opened	-1.4512	0.088	-16.399	0.000	-1.625	-1.278
Last Notable Activity_Modified	-1.7762	0.100	-17.702	0.000	-1.973	-1.580
Last Notable Activity_Olark Chat Conversation	-1.4942	0.378	-3.949	0.000	-2.236	-0.753
Last Notable Activity_Page Visited on Website	-1.7108	0.203	-8.427	0.000	-2.109	-1.313

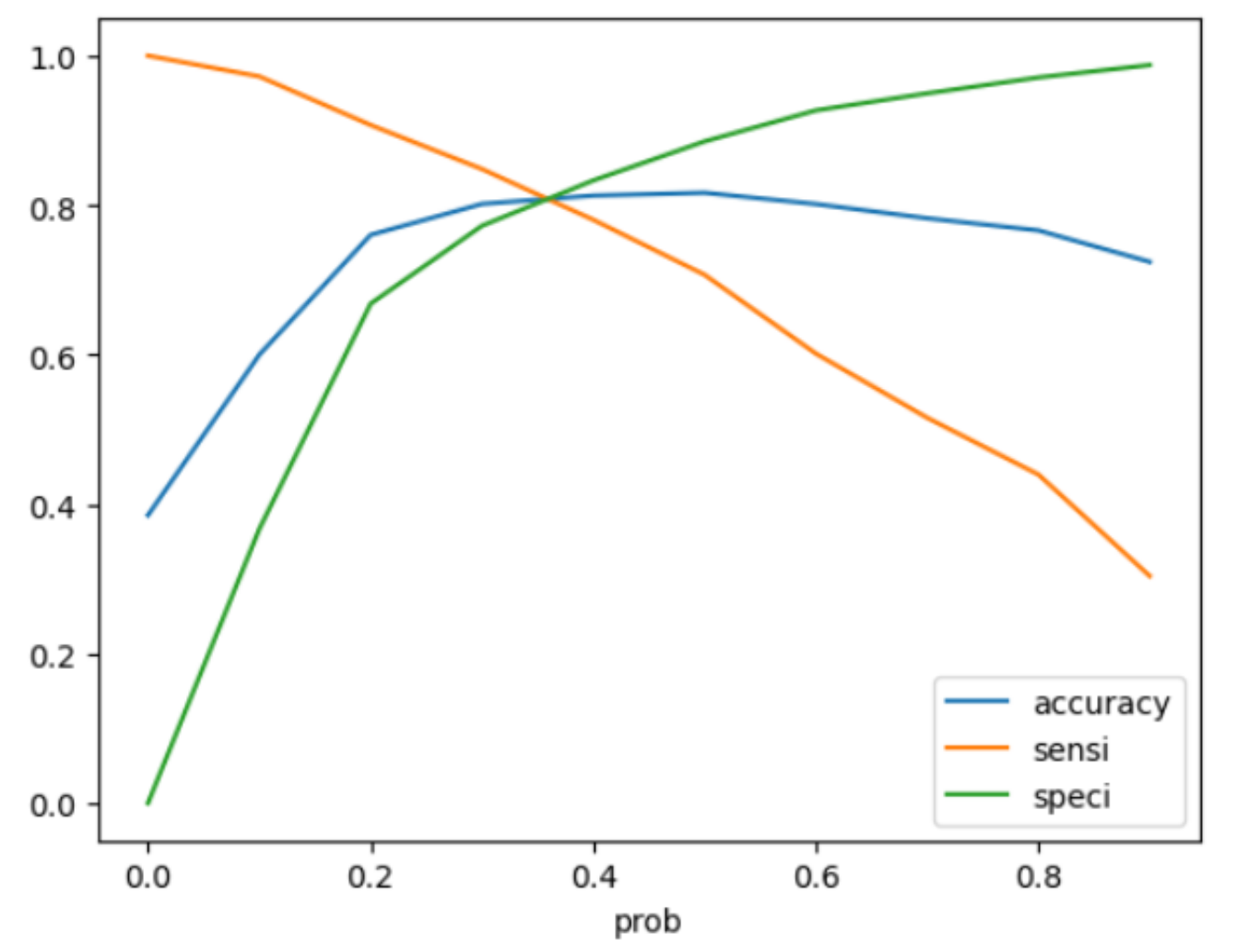
MODEL EVALUATION

From the plot , we can see that we have higher (0.89) area under the ROC curve , therefore our model is a good one.



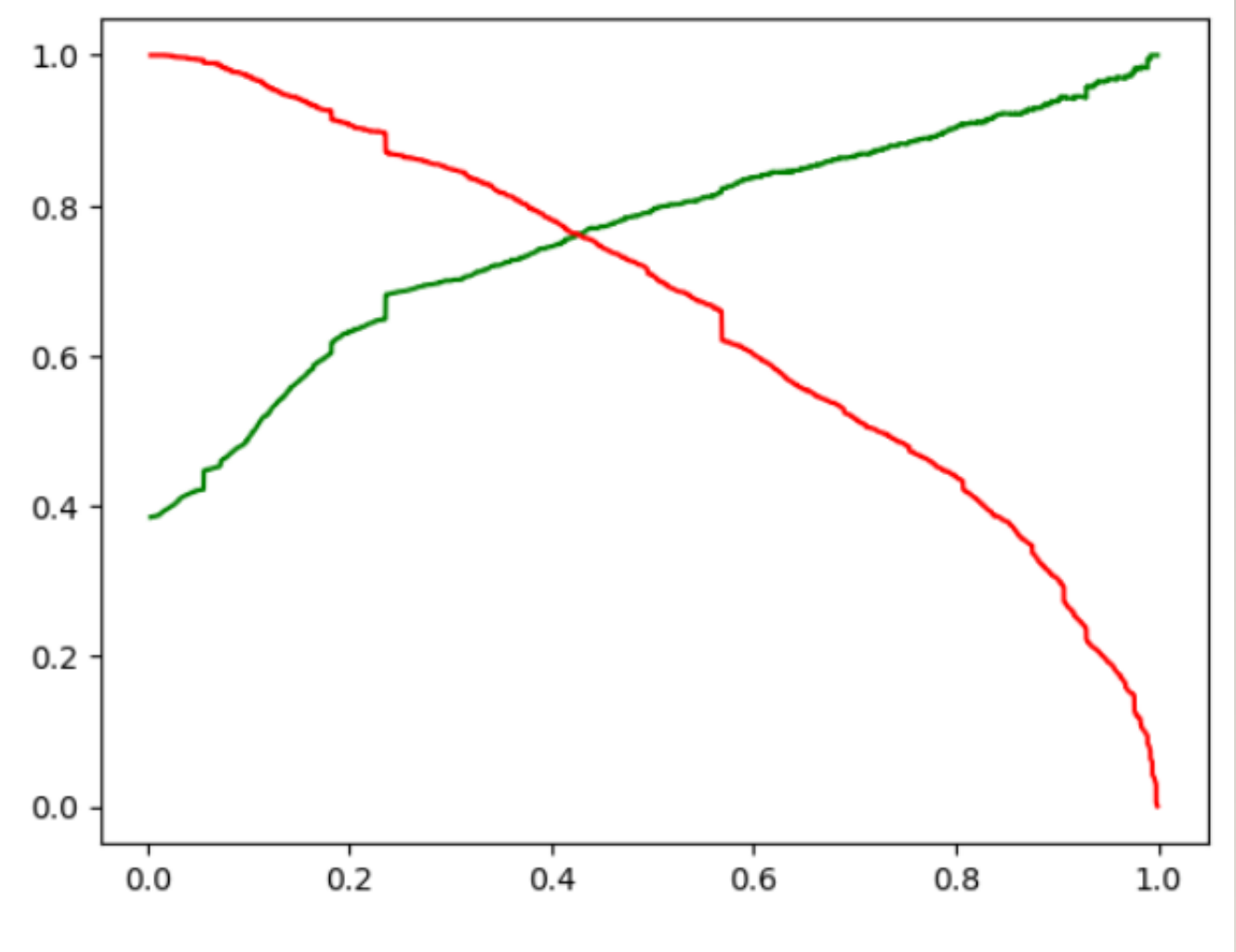
OPTIMAL CUT OFF POINT

From the above curve, 0.34 is the optimum point to take it as a cutoff probability



PRECISION & RECALL TRADE OFF

There is a trade off between precision and recall. The intersection point is approximately 0.4



CONCLUSION

Data cleaning and Exploratory data analysis was conducted and on preliminary analysis of the data, a conversion rate of 37.85% was observed

Majority of the leads came from API and Landing Page submission while the conversion rate was only 30-35%.

Looking at the Lead sources it was discovered that majority of are generated by Google and Direct Traffic while the high conversion rate of leads came from Reference Leads and Welingak website

Based on customer behavior spending more time on the website leads to higher conversion rate. People who have last activity as ‘Email opened’ or ‘sms sent’ seemed to have a higher conversion rate, with ‘sms sent’ having a significantly higher conversion as compared to no conversions.

Based on occupation working professionals are more likely to be converted as compared to unemployed. The unemployed have the second highest conversion rate

The selected model has cut off probability of 0.34 and 0.4 as the precision and recall intersection point.

Final accuracy, sensitivity, and specificity observed is 80.84%, 82.62% and 79.72 respectively on the training model and 80.05%, 81.39% and 72.29% respectively on Test model

THANK YOU