LEAD SCORE ASSIGNMENT PRESENTATION

BATCH - DSC-40
BY:
Ruchi Jain
Sushma SP

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.

BUSINESS GOALS:

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

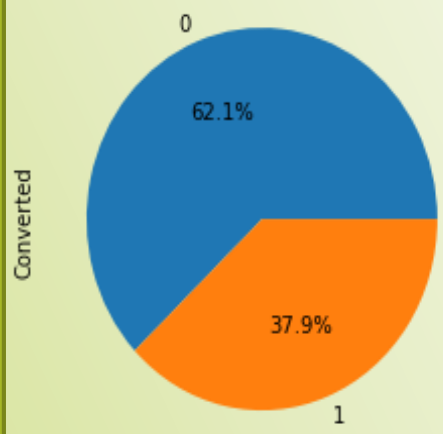
DATASET:

You have been provided with a leads dataset from the past with around 9000 data points. This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not. The target variable, in this case, is the column 'Converted' which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn't converted.

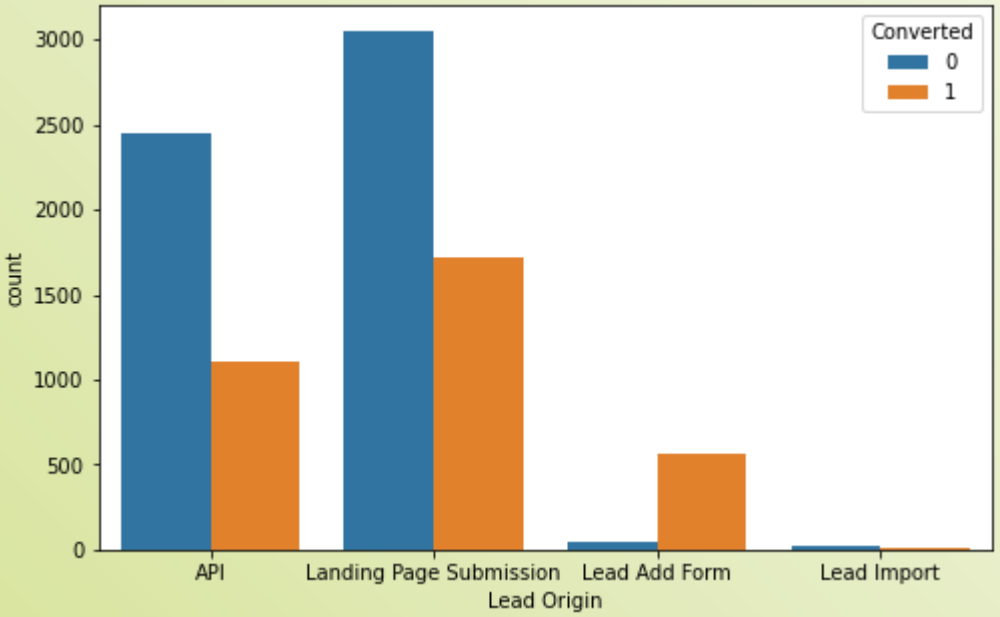
APPROACH:

- Reading & Understanding Data
- Data Cleaning
- EDA
- Creating Dummy Variables
- Splitting Data into Test-Train Dataset
- Feature Scaling
- Model Building
- Model Evaluation with Sensitivity-Specificity or Precision-Recall
- Predictions on Test Set

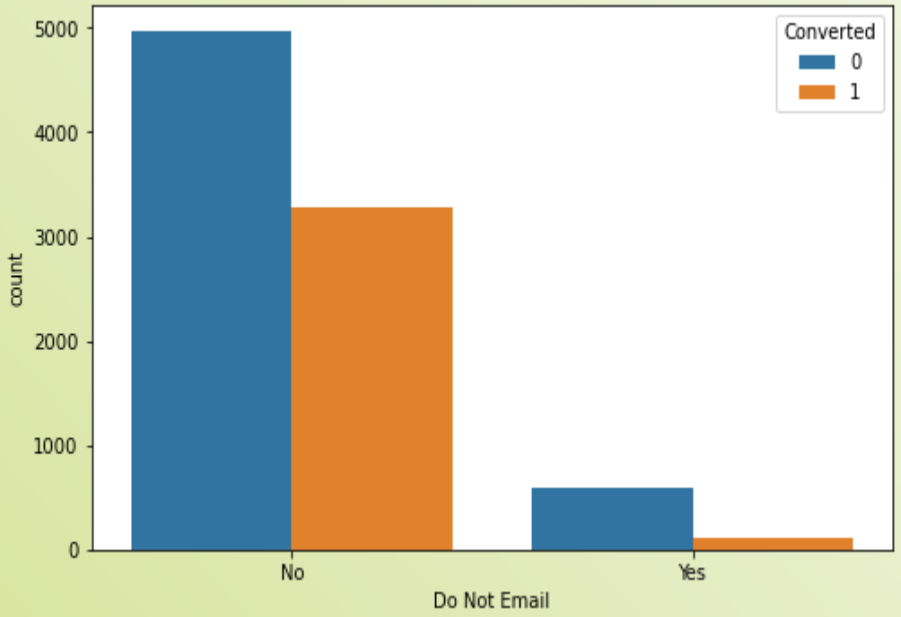
EXPLORATORY DATA ANALYSIS:



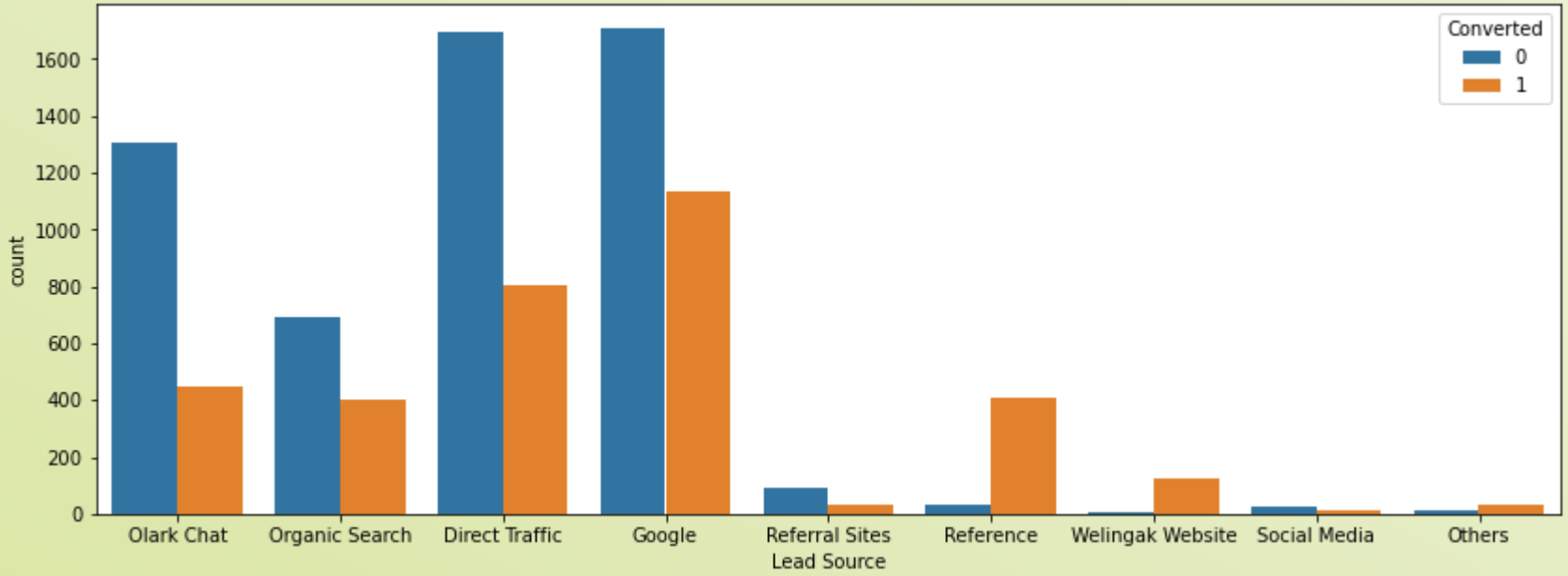
Conversion rate is approx. 38 %



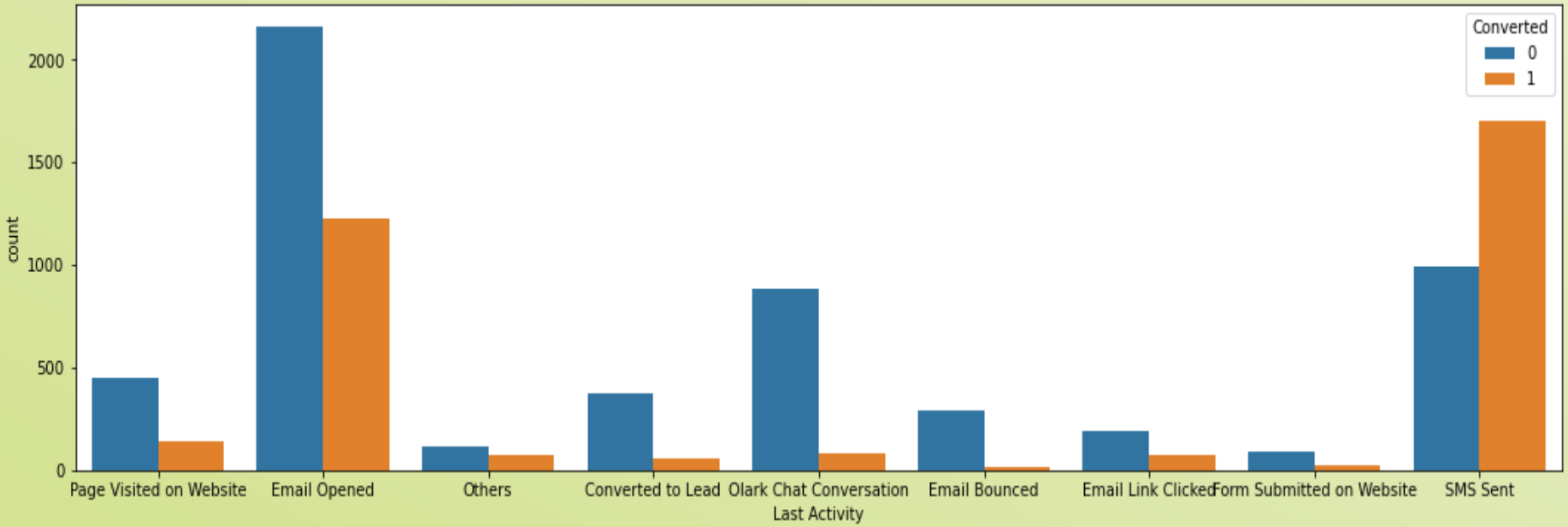
Lead add form have high lead conversion rate.



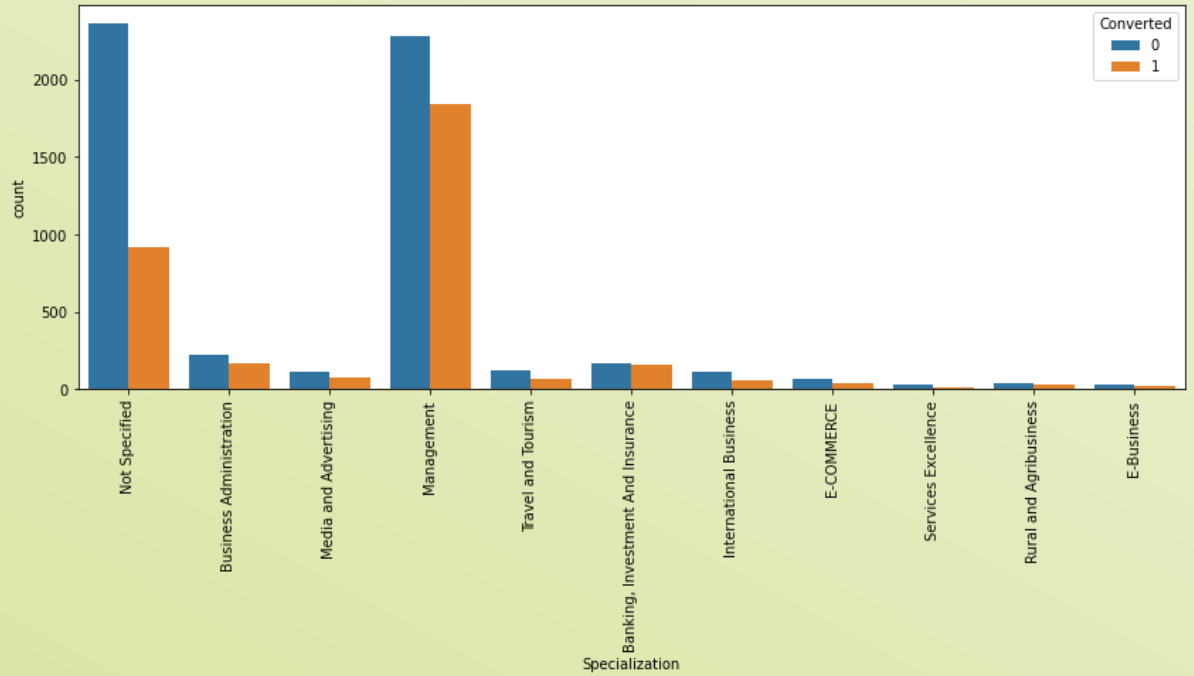
People who opted for email have most number of leads.



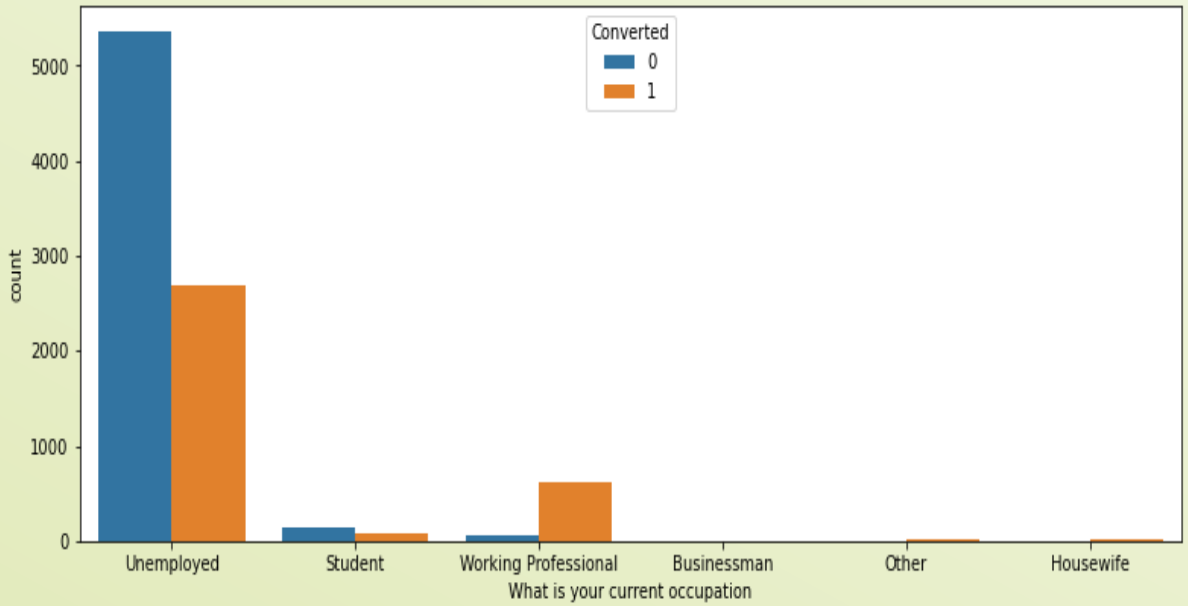
Reference leads and welingak website leads have high chances for conversion.



- Leads whose last activity is SMS sent have higher chances for conversion.
- Leads whose last activity is to have olark chat conversation are more likely to not convert.

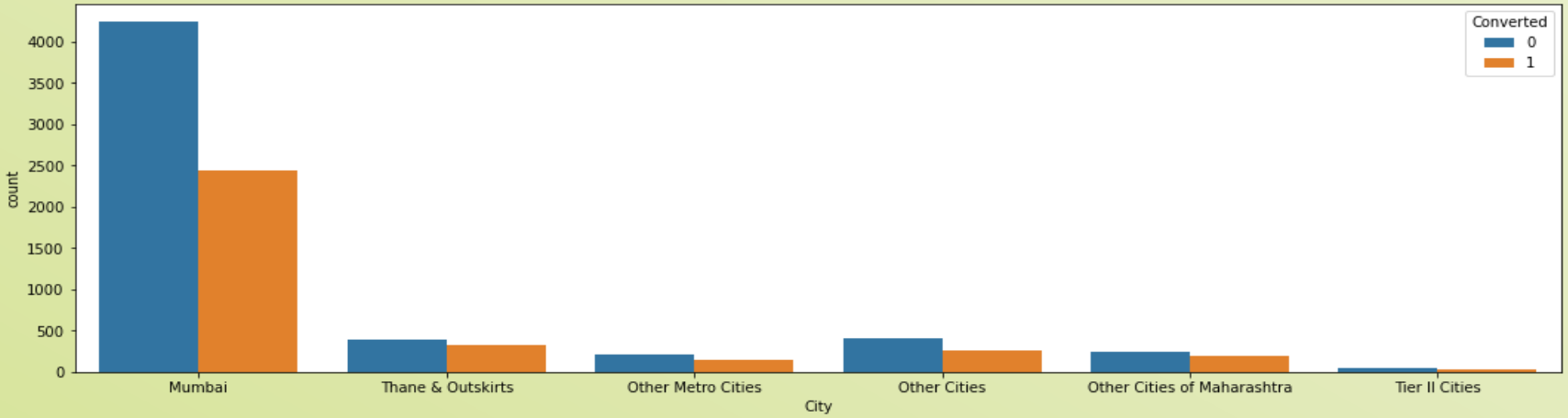


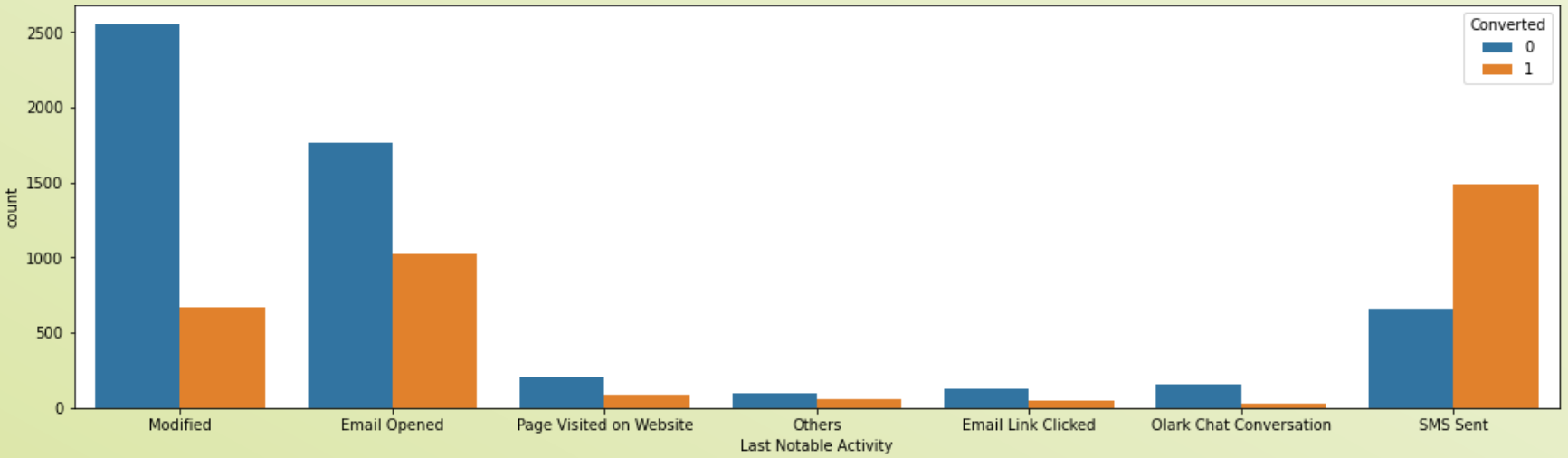
People with Management as specialization tends to have more lead and more likely to convert



Working professional have high chances for conversion.

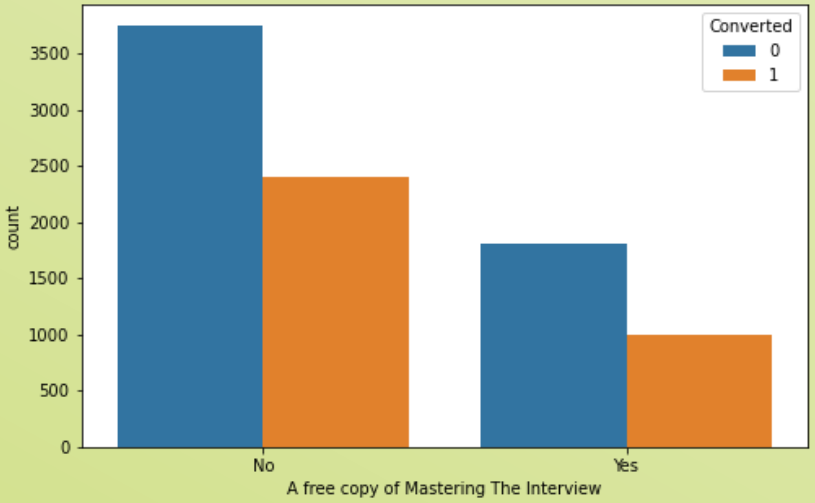
Mostly leads are from Mumbai.



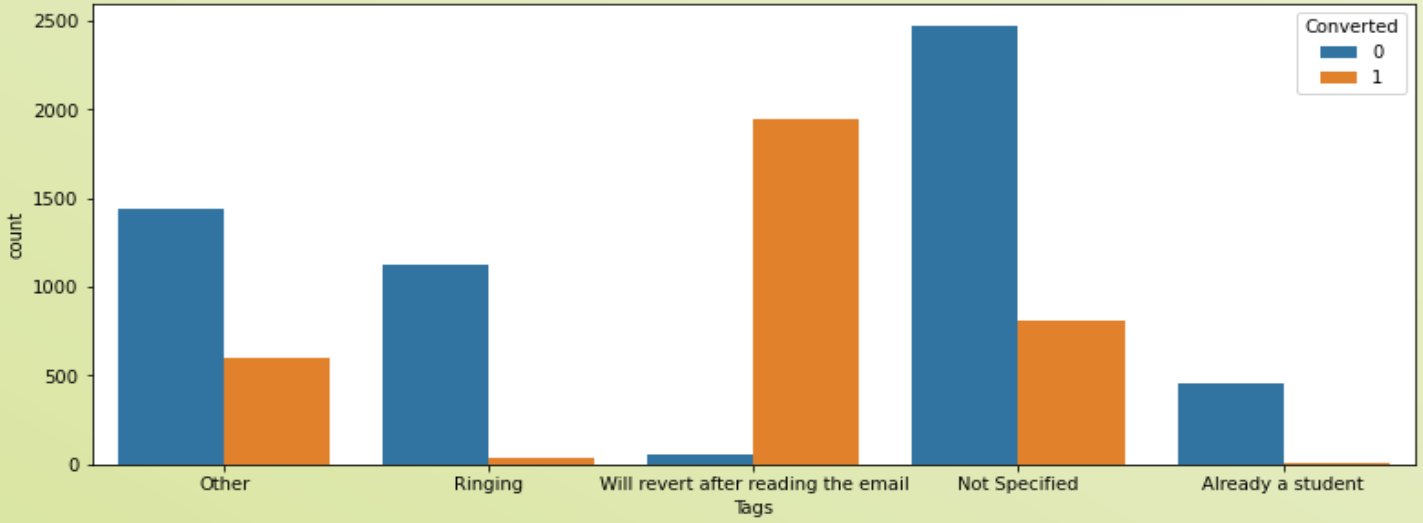


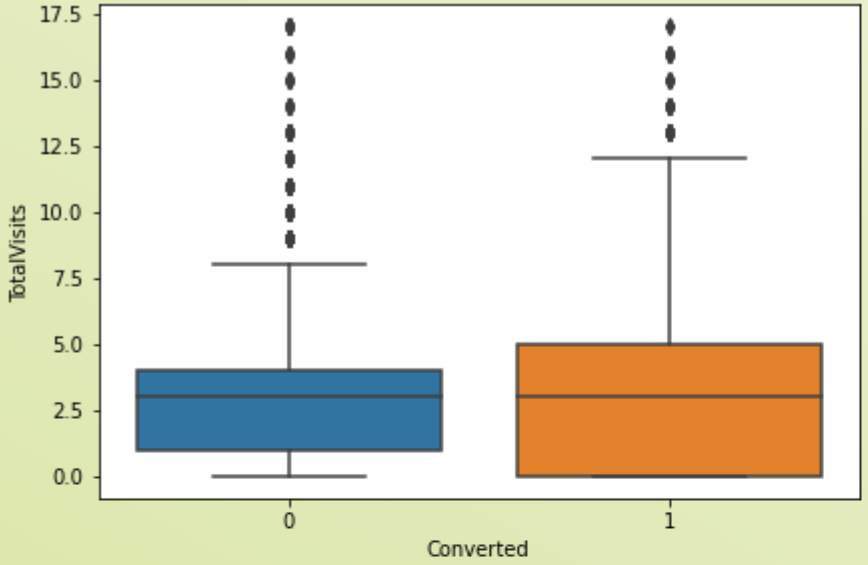
Leads whose last notable activity is SMS sent have higher chances of conversion.

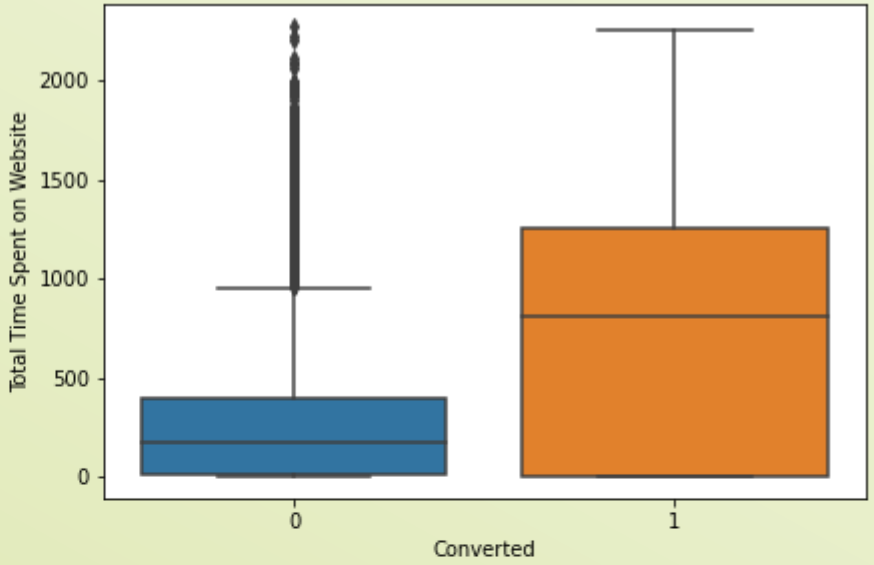
Mostly leads are those who have not opted for a free copy of Mastering The Interview.

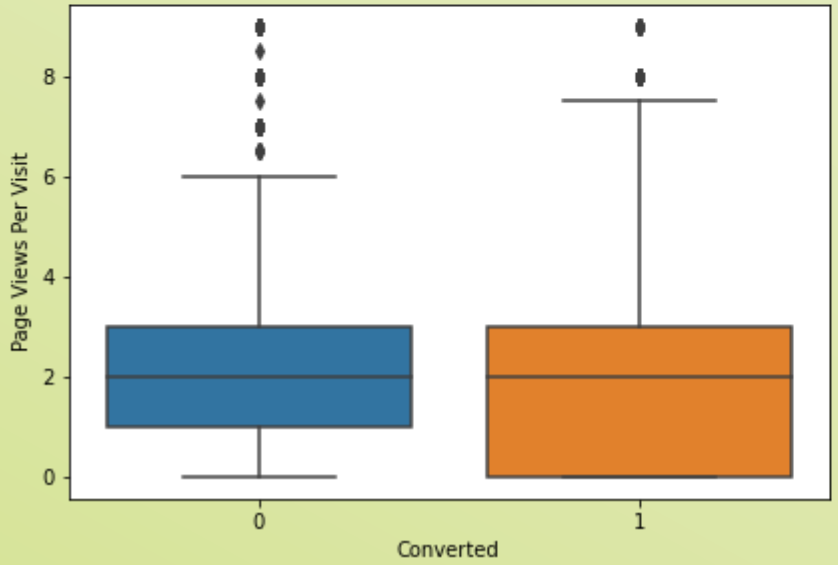


Tags with "will revert after reading mail" are more likely to convert.



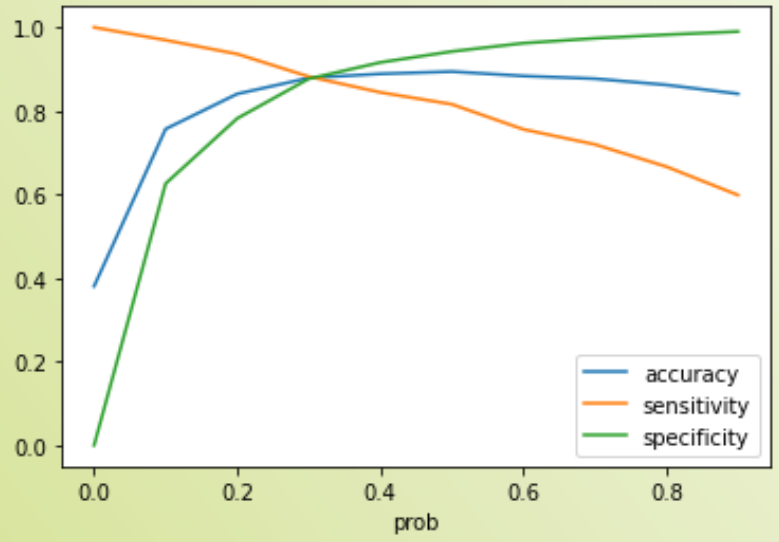






Leads spending more time on website have high chances of conversion.

MODEL EVALUATION PREDICTIONS:



Optimal Threshold = 0.3

Train set:

- Accuracy: 87.88%
- Sensitivity: 88.21%
- Specificity: 87.68%
- Precision: 81.48%
- Recall: 88.21%

Test Set:

- Accuracy: 87.34%
- Sensitivity: 86.13%
- Specificity: 88.06%
- Precision: 81.30%
- Recall: 86.13%

Dep. Variable:	Converted	No. Observations:	6267
Model:	GLM	Df Residuals:	6254
Model Family:	Binomial	Df Model:	12
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1673.6
Date:	Wed, 13 Jul 2022	Deviance:	3347.2
Time:	21:32:39	Pearson chi2:	6.78e+03
No. Iterations:	7		
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-5.0466	0.236	-21.352	0.000	-5.510	-4.583
Total Time Spent on Website	1.0994	0.051	21.607	0.000	1.000	1.199
Lead Origin_Lead Add Form	4.2632	0.298	14.330	0.000	3.680	4.846
Lead Source_Olark Chat	1.4212	0.130	10.943	0.000	1.167	1.676
Lead Source_Welingak Website	2.4180	1.050	2.303	0.021	0.360	4.476
Last Activity_Olark Chat Conversation	-1.2370	0.206	-6.003	0.000	-1.641	-0.833
Last Activity_SMS Sent	1.4927	0.098	15.171	0.000	1.300	1.686
Specialization_Travel and Tourism	-1.3067	0.369	-3.539	0.000	-2.030	-0.583
What is your current occupation_Working Professional	1.4889	0.284	5.247	0.000	0.933	2.045
Tags_Not Specified	3.3026	0.231	14.277	0.000	2.849	3.756
Tags_Other	3.5278	0.236	14.970	0.000	3.066	3.990
Tags_Will revert after reading the email	7.5836	0.287	26.407	0.000	7.021	8.146
Last Notable Activity_Modified	-0.8342	0.099	-8.431	0.000	-1.028	-0.640

RESULT:

- Accuracy, Sensitivity, Specificity, Precision & Recall of train dataset are close to test dataset.
- While we have calculated both sensitivity-specificity & precision-recall, we have considered cutoff threshold from sensitivity-specificity metric.
- Sensitivity is 88.21% in train set and 86.13 in test set which tells that model predicts actual conversion 88.21% and 86.13% correctly.
- Overall model seems to be good.

RECOMMENDATIONS:

Model is able to adjust with company's required changes made in coming future.

Top features which should be keep in mind while targeting customers are:

- Tags_Will revert after reading the email
- Lead Origin_Lead Add Form
- Tags_Other
- Tags_Not Specified
- Lead Source_Welingak Website