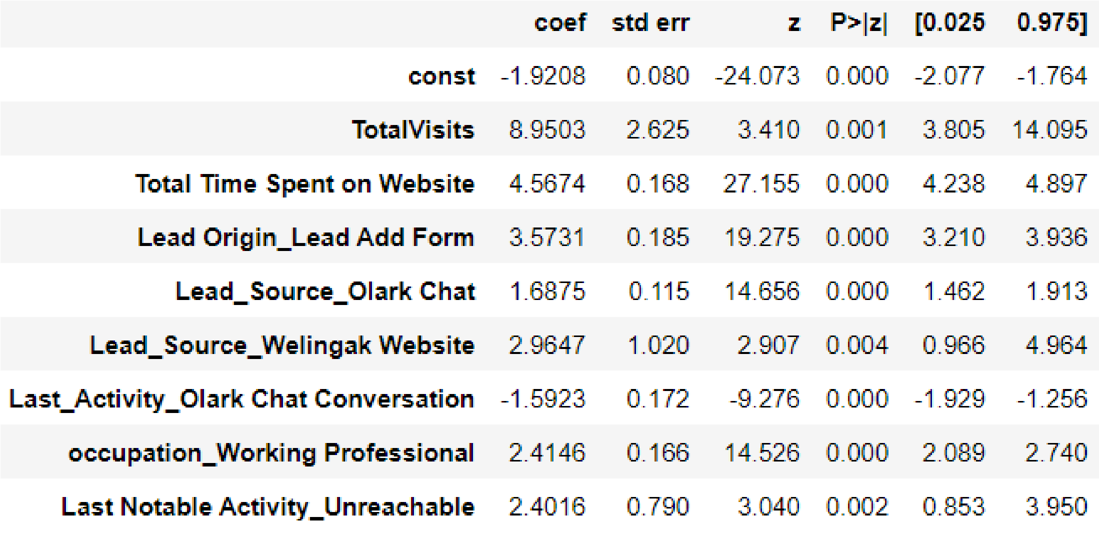
X education company got a good number of leads, and the lead conversion rate is poor. They want to know the potential lead to convert them into paying customers.

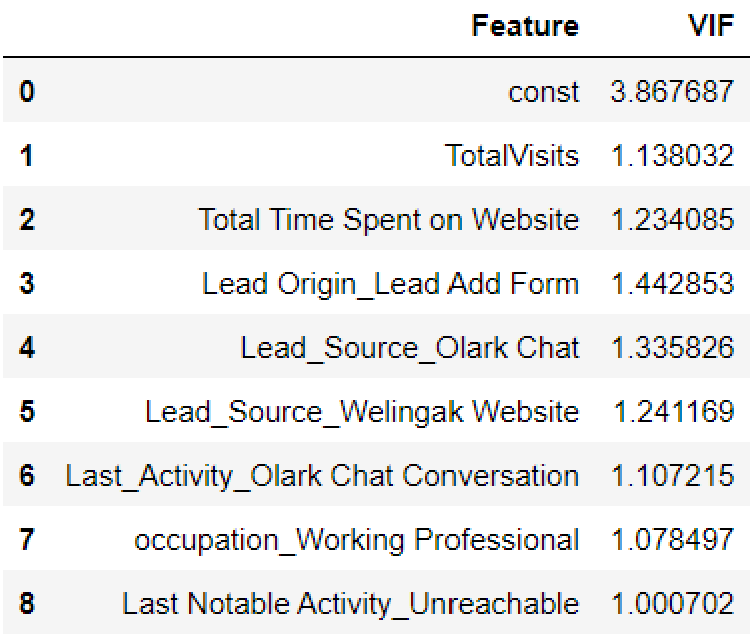
- Importing the data sets and validating the data sets
- Checking and imputing missing data using different techniques. Deleting the data having higher number of missing values.
- Creating dummy variables for categorical variables with 0 and 1 and dropping the variables which are needed for model building.
- Splitting data into train and test by 80, 20 percentage
- Using MinMax scaler transforming the numeric variables to standard values.
- Model building using logistic regression and checking the p values, p values should be less than 0.05 if not delete the particular variable and re run the model again .



	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.9208	0.080	-24.073	0.000	-2.077	-1.764
TotalVisits	8.9503	2.625	3.410	0.001	3.805	14.095
Total Time Spent on Website	4.5674	0.168	27.155	0.000	4.238	4.897
Lead Origin_Lead Add Form	3.5731	0.185	19.275	0.000	3.210	3.936
Lead_Source_Olark Chat	1.6875	0.115	14.656	0.000	1.462	1.913
Lead_Source_Welingak Website	2.9647	1.020	2.907	0.004	0.966	4.964
Last_Activity_Olark Chat Conversation	-1.5923	0.172	-9.276	0.000	-1.929	-1.256
occupation_Working Professional	2.4146	0.166	14.526	0.000	2.089	2.740
Last Notable Activity_Unreachable	2.4016	0.790	3.040	0.002	0.853	3.950

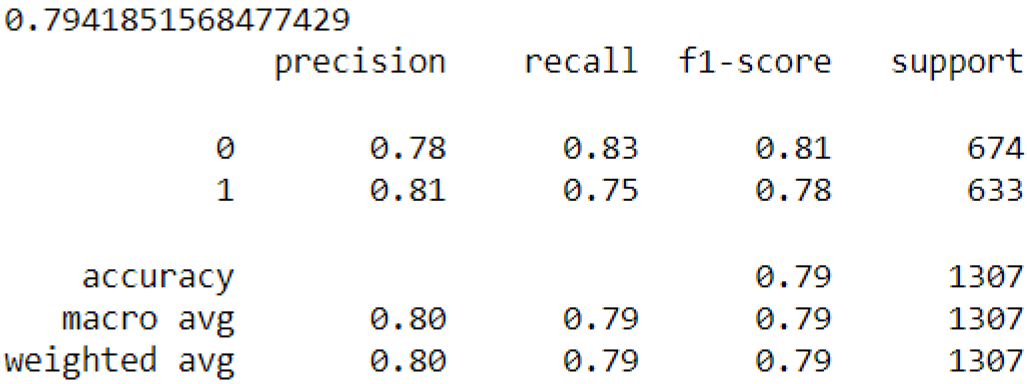
-
- Getting the predicted values on the train set
- Giving the each customer converted_prob using these values gave predicted converted values

- 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.
- Converted prob having the score
- REF values should be less than 5 basing on that build the model



	Feature	VIF
0	const	3.867687
1	TotalVisits	1.138032
2	Total Time Spent on Website	1.234085
3	Lead Origin_Lead Add Form	1.442853
4	Lead_Source_Olark Chat	1.335826
5	Lead_Source_Welingak Website	1.241169
6	Last_Activity_Olark Chat Conversation	1.107215
7	occupation_Working Professional	1.078497
8	Last Notable Activity_Unreachable	1.000702

-
- Important accuracy and precision, recall, f1 score

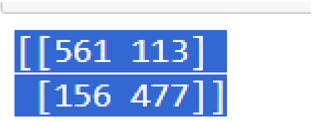


```
0.7941851568477429
              precision    recall  f1-score   support

           0       0.78      0.83      0.81       674
           1       0.81      0.75      0.78       633

    accuracy                           0.79      1307
   macro avg       0.80      0.79      0.79      1307
weighted avg       0.80      0.79      0.79      1307
```

-
- Confusion matrix



```
[[561 113]
 [156 477]]
```