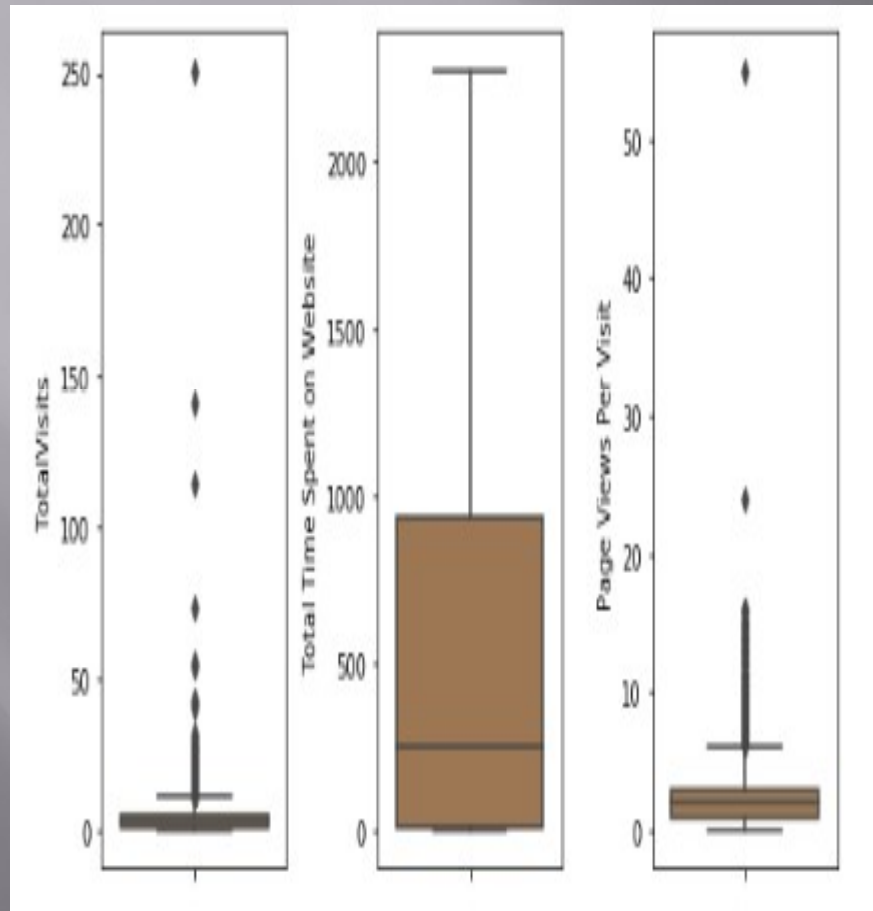


LEAD SCORING CASE STUDY

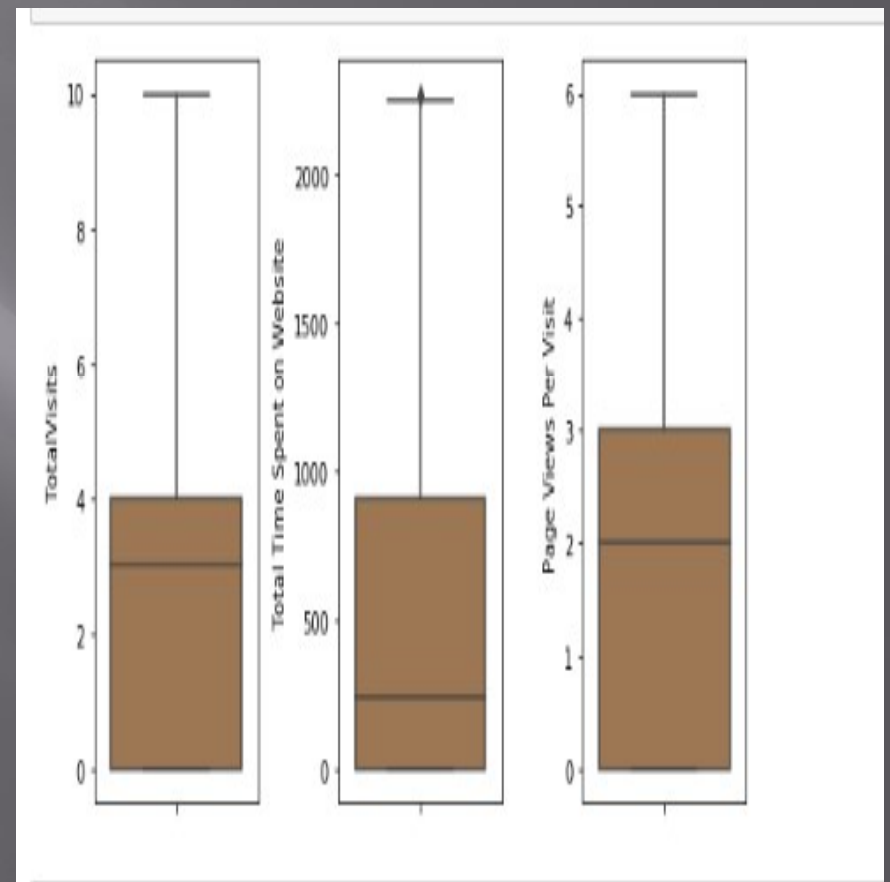
Data cleaned

	Total	Percentage
Last Notable Activity	0	0.0
A free copy of Mastering The Interview	0	0.0
Lead Origin	0	0.0
Lead Source	0	0.0
Do Not Email	0	0.0
Do Not Call	0	0.0
Converted	0	0.0
TotalVisits	0	0.0
Total Time Spent on Website	0	0.0
Page Views Per Visit	0	0.0
Last Activity	0	0.0
Specialization	0	0.0
What is your current occupation	0	0.0
Search	0	0.0
Newspaper Article	0	0.0
X Education Forums	0	0.0
Newspaper	0	0.0
Digital Advertisement	0	0.0
Through Recommendations	0	0.0
Lead Profile	0	0.0
City	0	0.0
Lead Number	0	0.0

Removing Outlier for numeric variables



Before



After

Logistic Regression

Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	6002
Model:	GLM	Df Residuals:	5981
Model Family:	Binomial	Df Model:	20
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2297.3
Date:	Mon, 20 Apr 2020	Deviance:	4594.7
Time:	15:00:38	Pearson chi2:	6.29e+03
No. Iterations:	22		
Covariance Type:	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
const	-0.0544	0.075	-0.729	0.466	-0.201	0.092
Do Not Email	-1.7159	0.185	-9.270	0.000	-2.079	-1.353
Total Time Spent on Website	1.0146	0.042	23.975	0.000	0.932	1.097
Lead Source_Olark Chat	1.3264	0.110	12.059	0.000	1.111	1.542
Lead Source_Wellingak Website	3.0497	0.754	4.042	0.000	1.571	4.528
Lead Origin_Lead Add Form	3.0081	0.208	14.483	0.000	2.601	3.415
Last Notable Activity_Email Link Clicked	-1.9258	0.282	-6.841	0.000	-2.478	-1.374
Last Notable Activity_Email Opened	-1.5567	0.095	-16.391	0.000	-1.743	-1.371
Last Notable Activity_Modified	-2.0201	0.103	-19.623	0.000	-2.222	-1.818
Last Notable Activity_Olark Chat Conversation	-1.8826	0.419	-4.495	0.000	-2.703	-1.062
Last Notable Activity_Page Visited on Website	-1.7660	0.240	-7.360	0.000	-2.236	-1.296
Lead Profile_Dual Specialization Student	23.6022	2.32e+04	0.001	0.999	-4.55e+04	4.56e+04
Lead Profile_Lateral Student	2.9918	1.098	2.725	0.006	0.840	5.144
Lead Profile_Potential Lead	1.8633	0.105	17.775	0.000	1.658	2.069
Lead Profile_Student of Some School	-2.0913	0.436	-4.800	0.000	-2.945	-1.237
What is your current occupation_Housewife	23.2503	2.42e+04	0.001	0.999	-4.74e+04	4.75e+04
What is your current occupation_Student	0.8791	0.267	3.295	0.001	0.356	1.402
What is your current occupation_Working Professional	2.2479	0.194	11.603	0.000	1.868	2.628
Last Activity_Approached upfront	23.8446	2.47e+04	0.001	0.999	-4.83e+04	4.84e+04
Last Activity_Had a Phone Conversation	-1.6853	0.817	-2.062	0.039	-3.287	-0.084
Last Activity_Olark Chat Conversation	-1.3465	0.212	-6.359	0.000	-1.762	-0.931

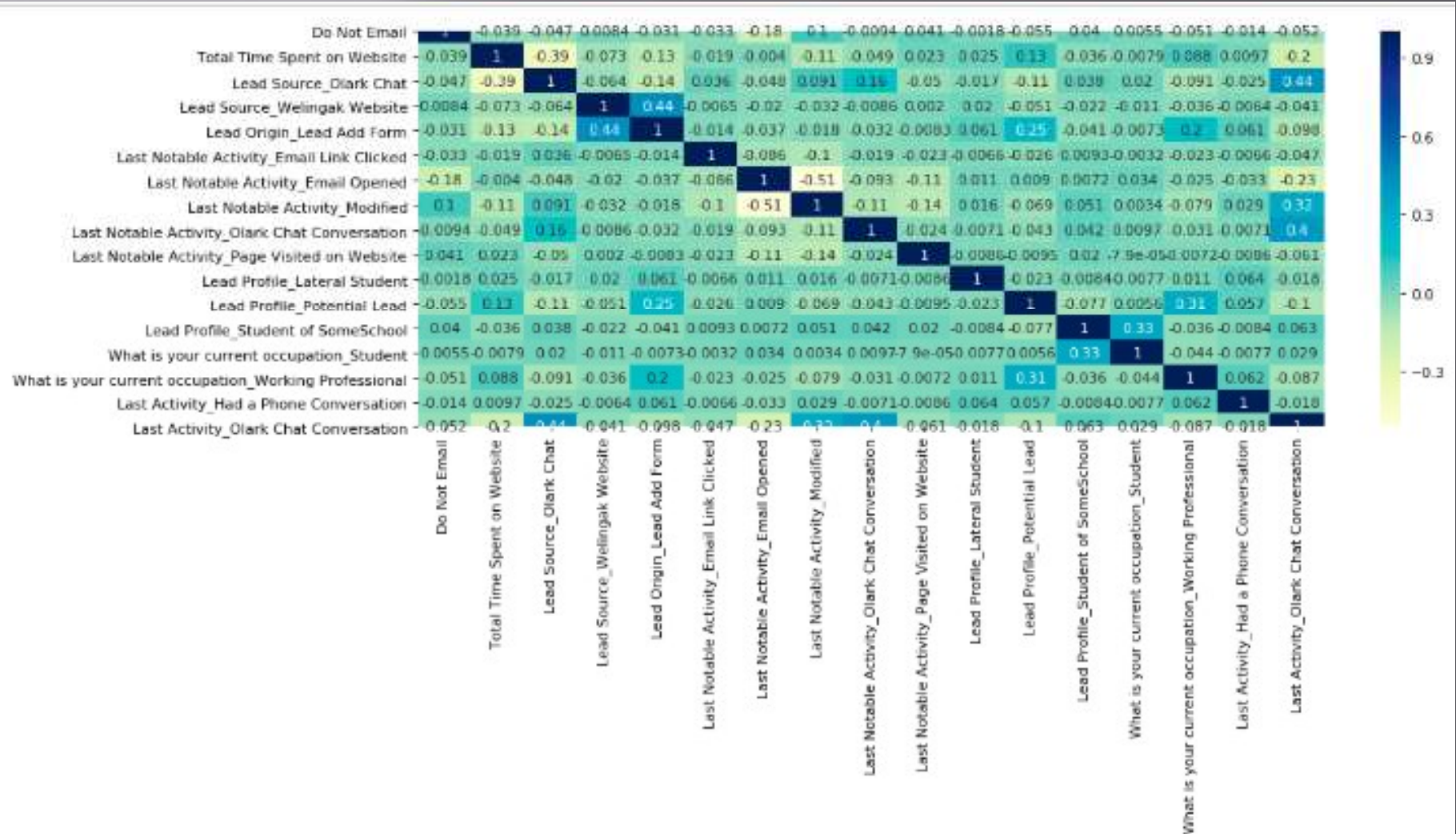
Logistic Regression

Generalized Linear Model Regression Results

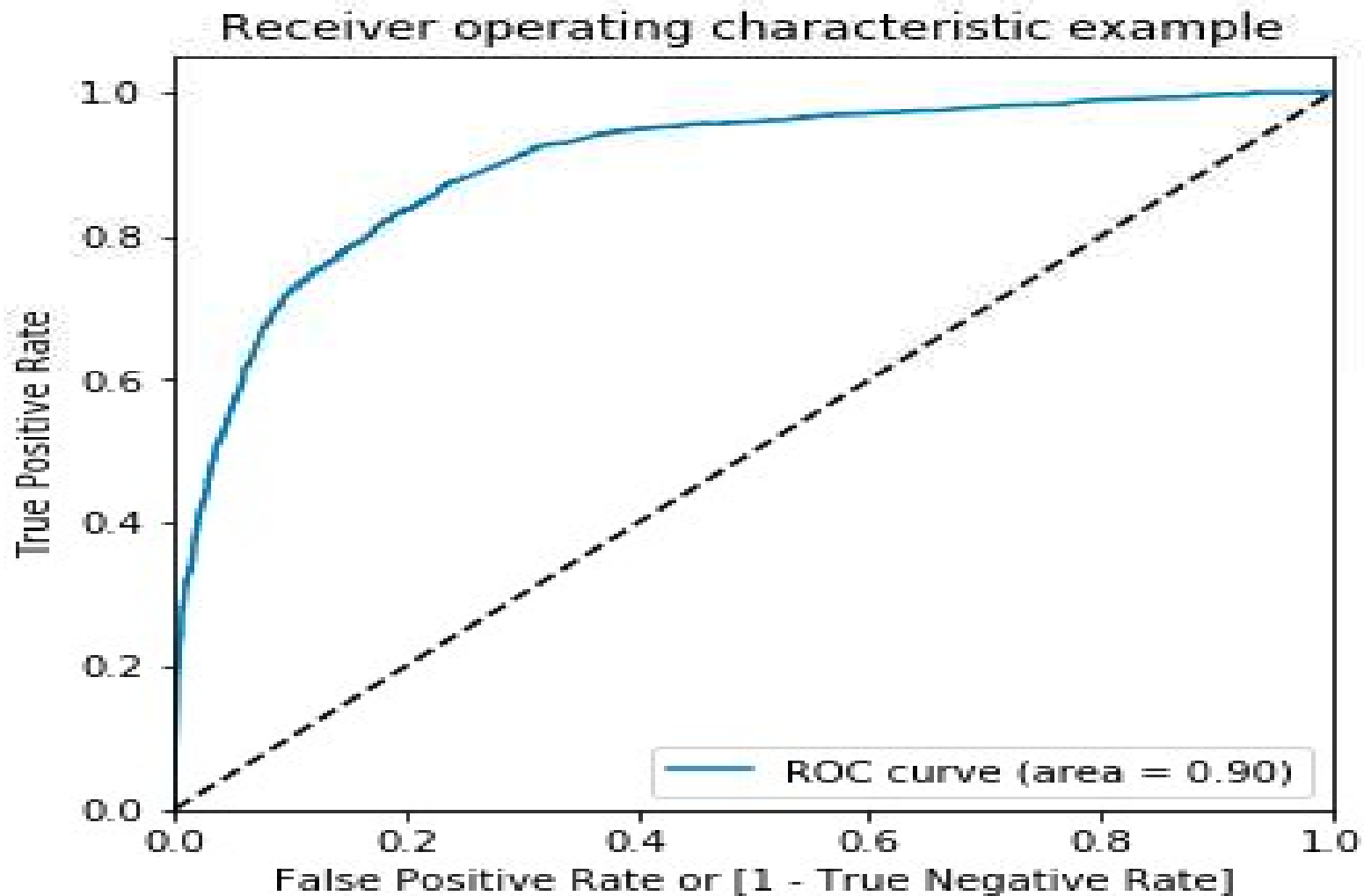
Dep. Variable:	Converted	No. Observations:	6002
Model:	GLM	Df Residuals:	5984
Model Family:	Binomial	Df Model:	17
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-2314.3
Date:	Mon, 20 Apr 2020	Deviance:	4628.6
Time:	15:00:42	Pearson chi2:	6.34e+03
No. iterations:	7		
Covariance Type:	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
const	-0.0516	0.075	-0.692	0.489	-0.198	0.095
Do Not Email	-1.7272	0.185	-9.332	0.000	-2.090	-1.364
Total Time Spent on Website	1.0203	0.042	24.150	0.000	0.937	1.103
Lead Source_Olark Chat	1.3359	0.110	12.192	0.000	1.121	1.551
Lead Source_Wellingak Website	3.0060	0.754	3.985	0.000	1.528	4.484
Lead Origin_Lead Add Form	3.0443	0.207	14.706	0.000	2.639	3.450
Last Notable Activity_Email Link Clicked	-1.8991	0.278	-6.842	0.000	-2.443	-1.355
Last Notable Activity_Email Opened	-1.5439	0.095	-16.305	0.000	-1.729	-1.358
Last Notable Activity_Modified	-1.9993	0.102	-19.512	0.000	-2.200	-1.798
Last Notable Activity_Olark Chat Conversation	-1.8654	0.419	-4.453	0.000	-2.686	-1.044
Last Notable Activity_Page Visited on Website	-1.7701	0.240	-7.371	0.000	-2.241	-1.299
Lead Profile_Lateral Student	2.9645	1.099	2.699	0.007	0.811	5.118
Lead Profile_Potential Lead	1.8513	0.105	17.687	0.000	1.646	2.056
Lead Profile_Student of Some School	-2.1093	0.436	-4.834	0.000	-2.964	-1.254
What is your current occupation_Student	0.8672	0.267	3.249	0.001	0.344	1.390
What is your current occupation_Working Professional	2.2604	0.193	11.692	0.000	1.881	2.639
Last Activity_Had a Phone Conversation	-1.7085	0.819	-2.085	0.037	-3.314	-0.103
Last Activity_Olark Chat Conversation	-1.3723	0.211	-6.491	0.000	-1.787	-0.958

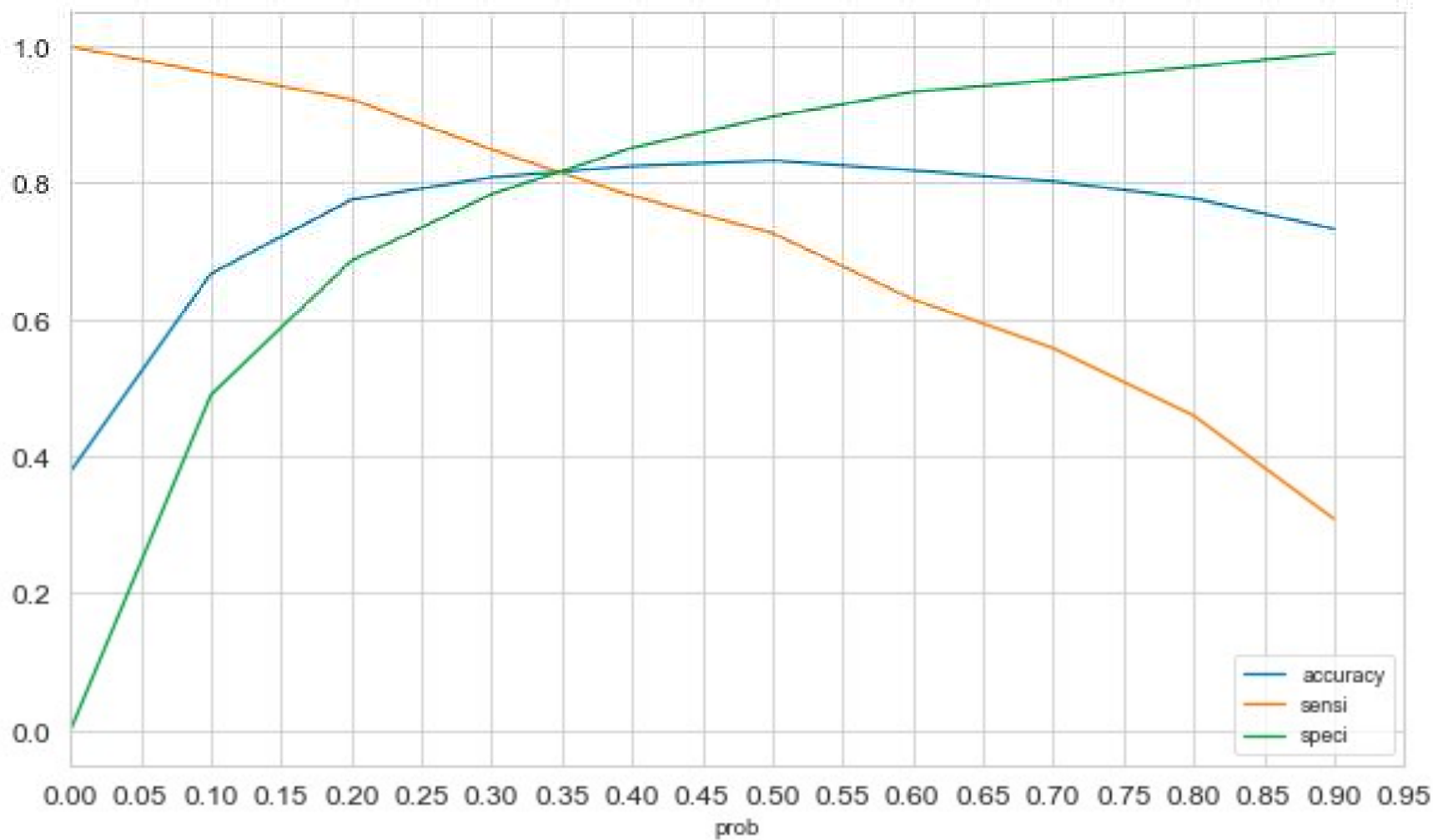
Correlation between the variables



ROC Curve



Optimal Cutoff Point



variables based on their relative coefficient values

