LEAD SCORING CASE STUDY

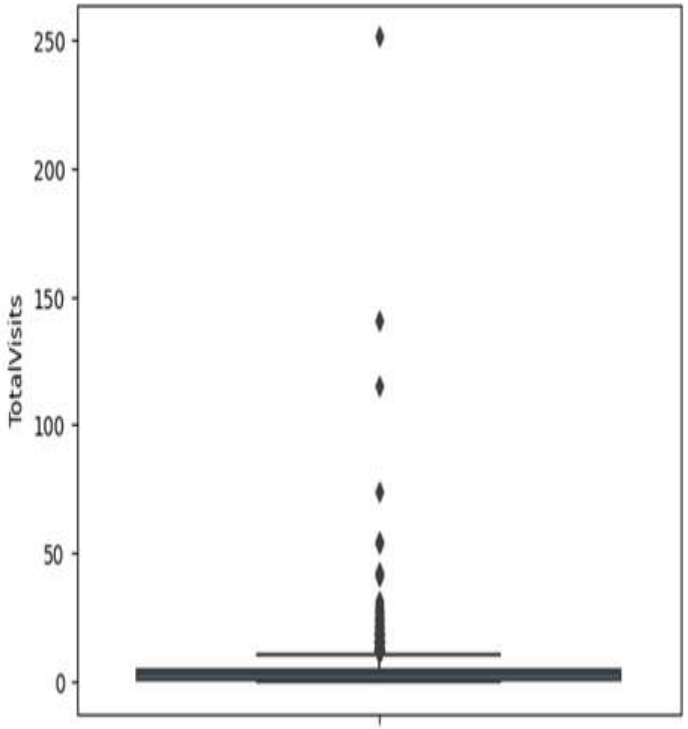
By,

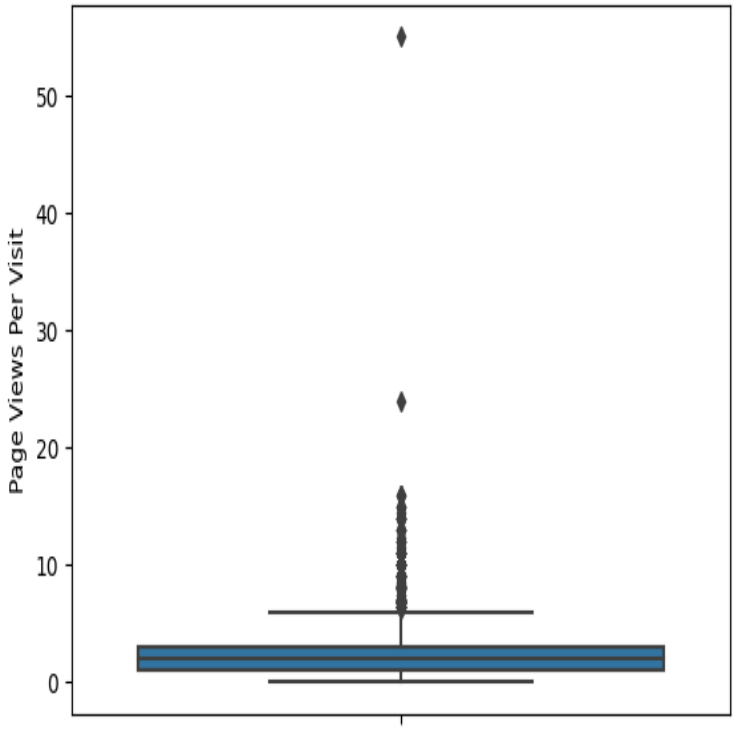
Nisha Ravi

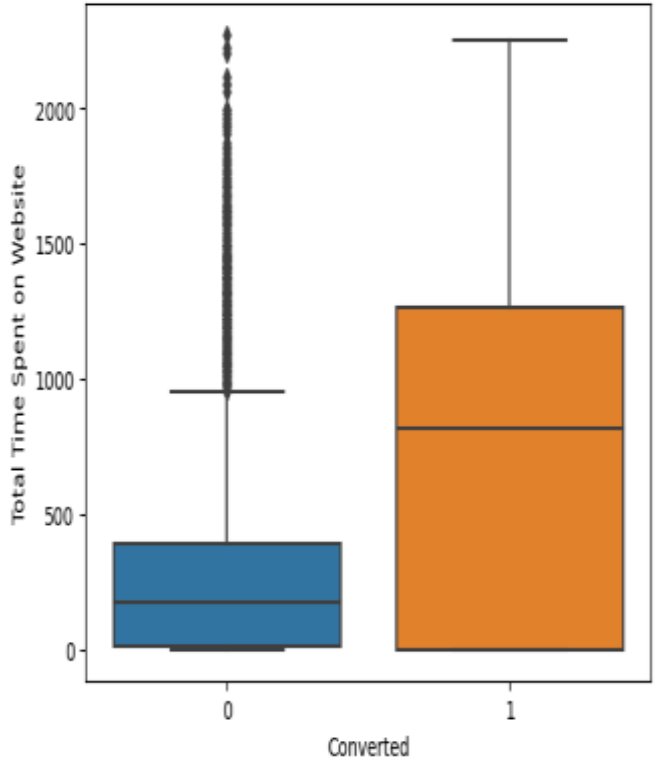
Priyanka Dhannuji Sorate

Anish Parate

OUTLIERS ANALYSIS

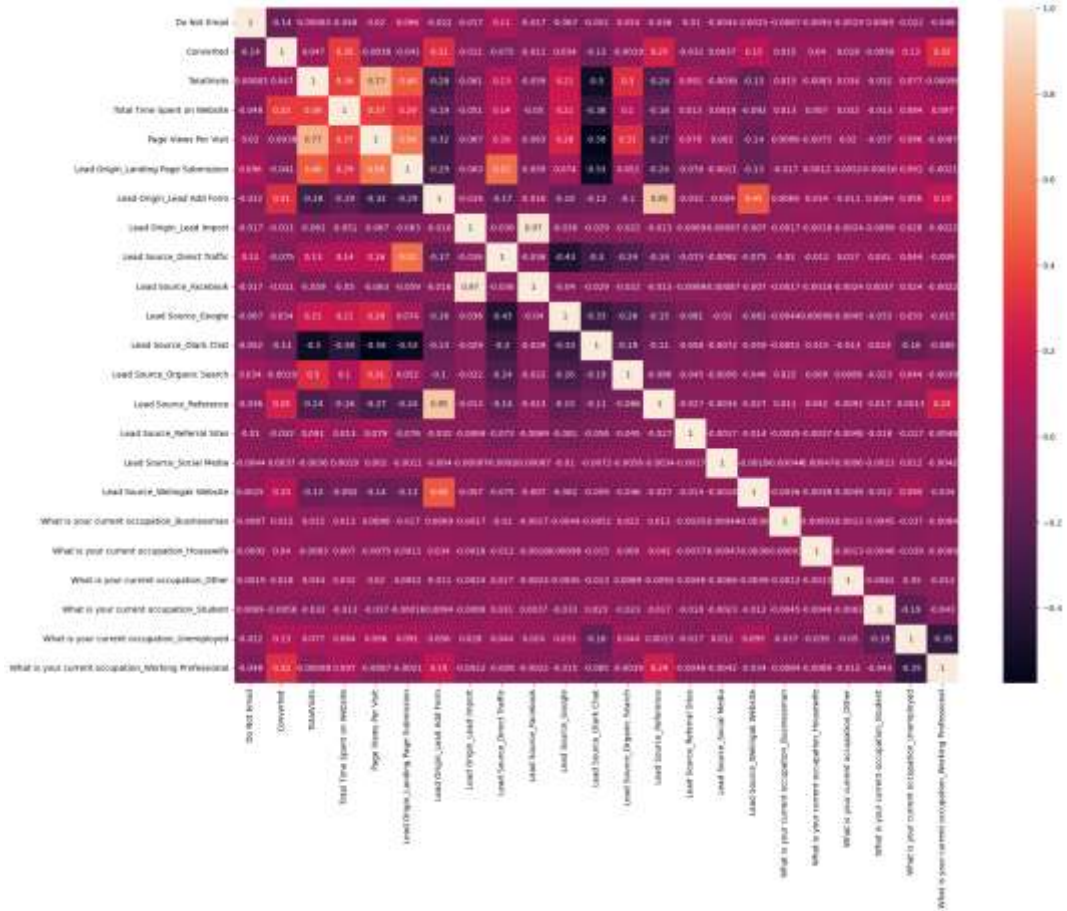






There are outliers in ‘Total Visits’ column and ‘Page Views Per Visit’ column.

Correlation:



MODEL BUILDING

	Features	VIF
2	Lead Origin_Lead Add Form	3.81
6	Lead Source_Reference	3.64
10	What is your current occupation_Unemployed	2.59
4	Lead Source_Google	1.71
3	Lead Source_Direct Traffic	1.67
5	Lead Source_Organic Search	1.31
11	What is your current occupation_Working Profes...	1.29
1	Total Time Spent on Website	1.12
9	What is your current occupation_Student	1.05
0	Do Not Email	1.03
7	Lead Source_Referral Sites	1.02
8	What is your current occupation_Businessman	1.00

Generalized Linear Model Regression Results

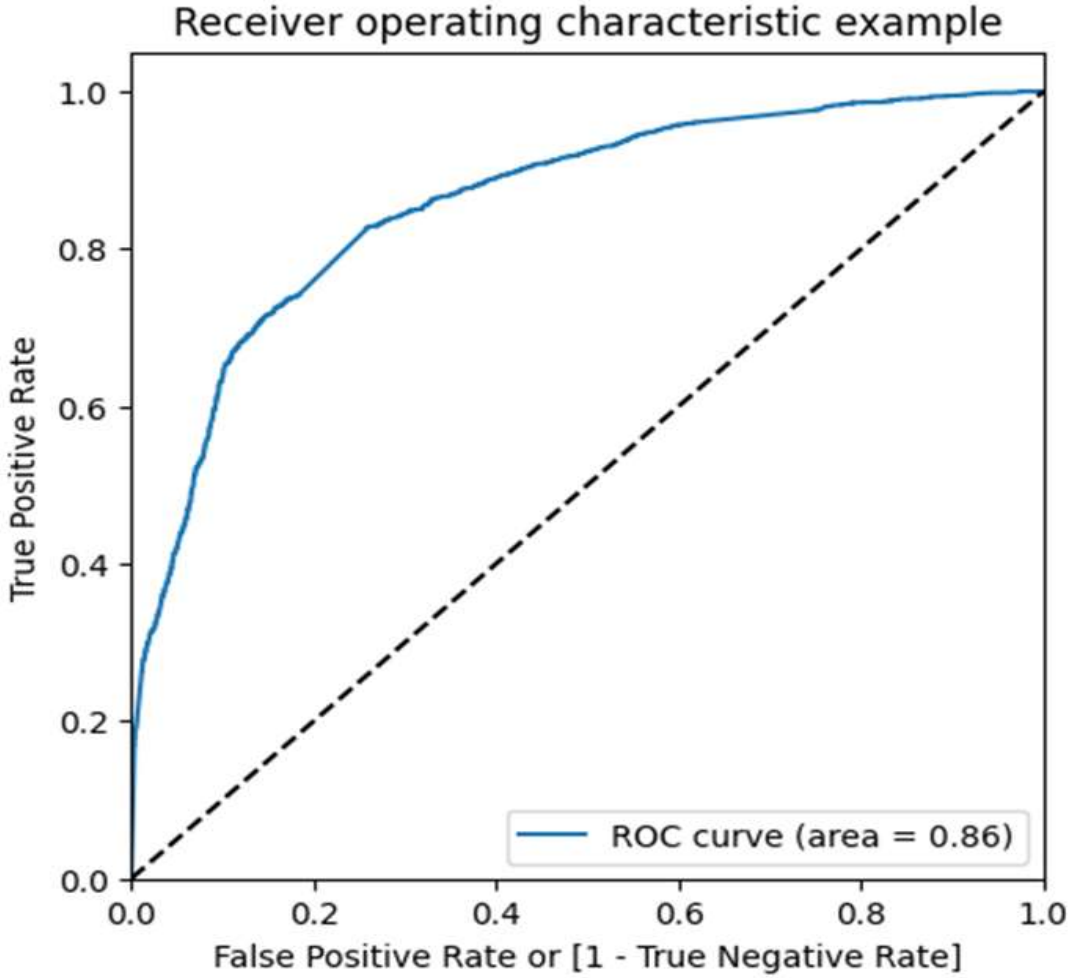
Dep. Variable:	Converted	**No. Observations:**	6372
Model:	GLM	**Df Residuals:**	6358
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2872.5
Date:	Tue, 23 May 2023	**Deviance:**	5745.1
Time:	06:19:45	**Pearson chi2:**	6.42e+03
No. Iterations:	6	**Pseudo R-squ. (CS):**	0.3470
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.2215	0.095	-12.847	0.000	-1.408	-1.035
Do Not Email	-0.3606	0.043	-8.350	0.000	-0.445	-0.276
Total Time Spent on Website	1.1006	0.038	28.654	0.000	1.025	1.176
Lead Origin_Lead Add Form	4.6079	0.523	8.807	0.000	3.582	5.633
Lead Source_Direct Traffic	-1.0559	0.107	-9.832	0.000	-1.266	-0.845
Lead Source_Google	-0.7818	0.103	-7.623	0.000	-0.983	-0.581
Lead Source_Organic Search	-0.8687	0.124	-7.026	0.000	-1.111	-0.626
Lead Source_Reference	-1.7536	0.564	-3.109	0.002	-2.859	-0.648
Lead Source_Referral Sites	-1.3724	0.336	-4.085	0.000	-2.031	-0.714
What is your current occupation_Businessman	1.4745	1.000	1.475	0.140	-0.485	3.434
What is your current occupation_Other	1.3324	0.641	2.080	0.038	0.077	2.588
What is your current occupation_Student	1.1571	0.225	5.154	0.000	0.717	1.597
What is your current occupation_Unemployed	1.2843	0.083	15.505	0.000	1.122	1.447
What is your current occupation_Working Professional	3.7806	0.189	20.002	0.000	3.410	4.151

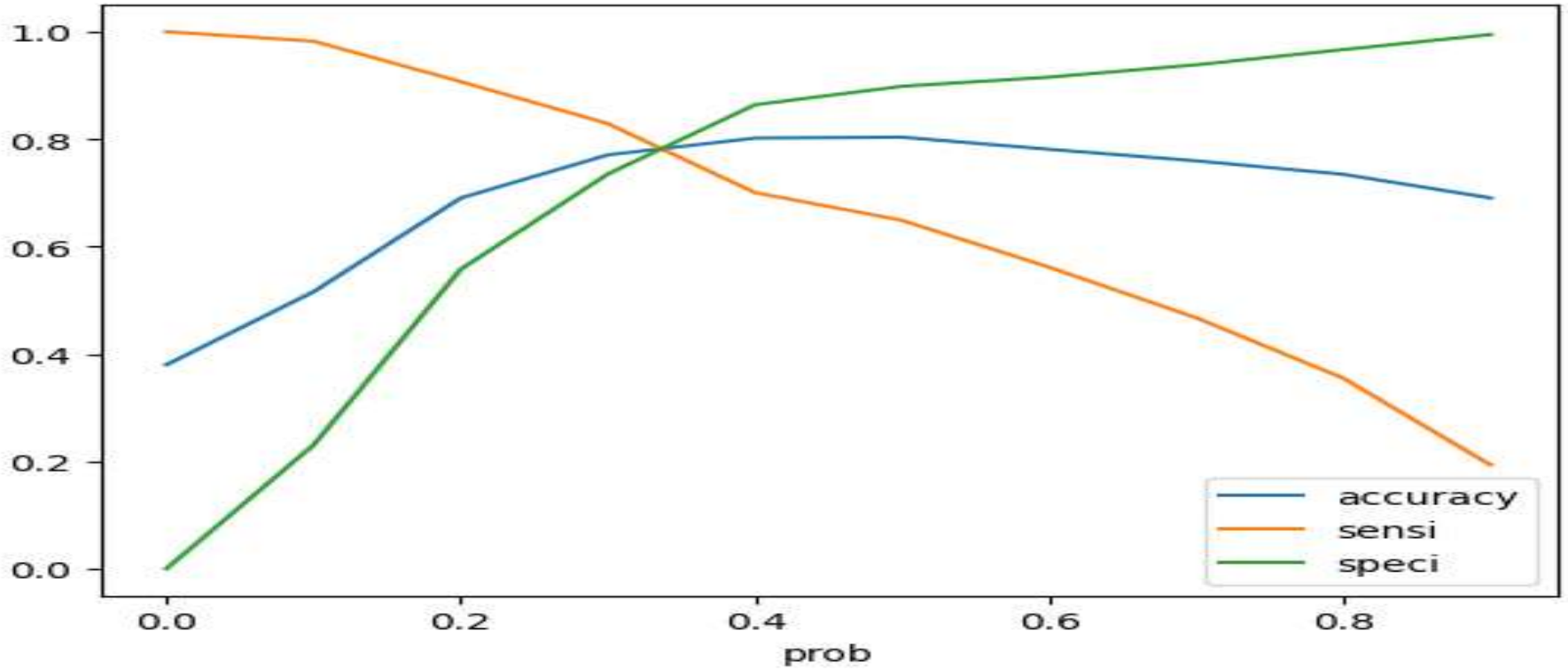
With the help of RFE, we can identify the insignificant variables present in our model.

EVALUATING THE MODEL

- After building the final model making prediction on it(on train set), we created ROC curve to find the model stability with AUC score(area under the curve) As we can see from the graph plotted on the right side, the area score is 0.88 which is a great score.
- And our graph is leaned towards the left side of the border which means we have good accuracy.



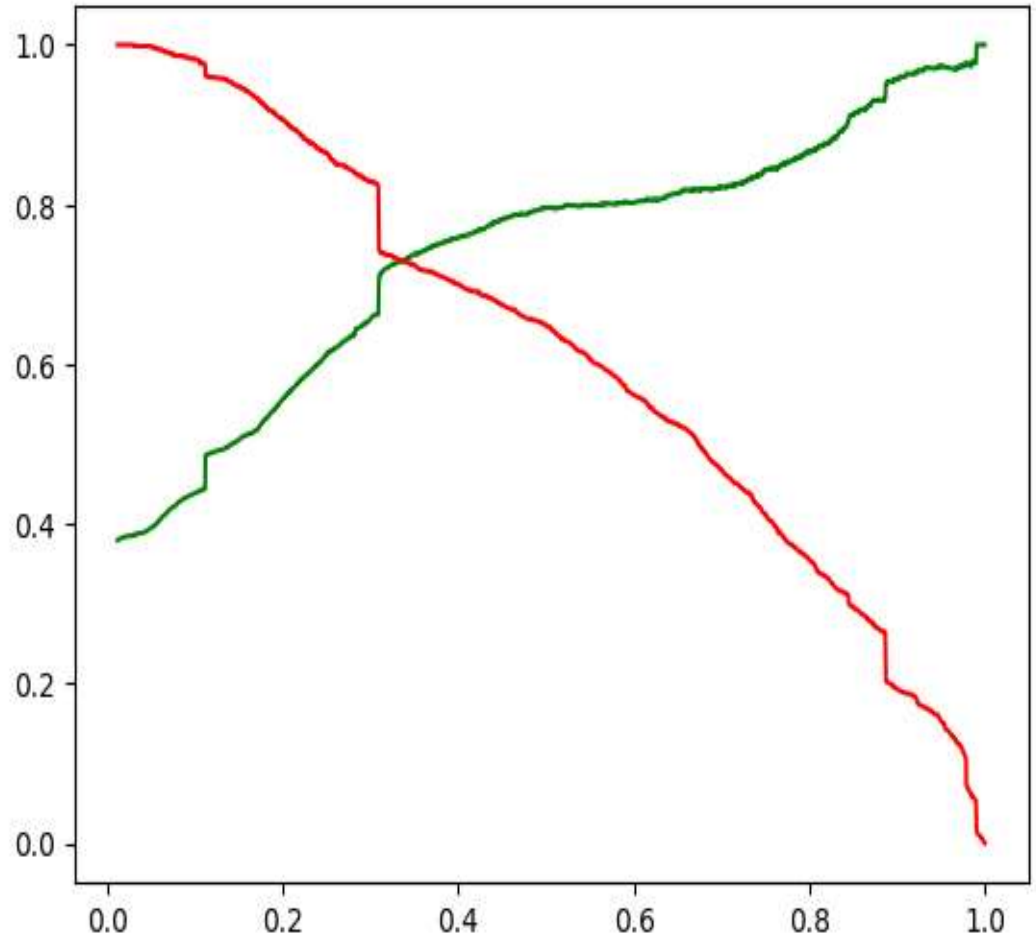
FINDING THE OPTIMAL CUT OFF POINT



We found that on 0.4 point all the score of accuracy, sensitivity and specificity are in a close range which is the ideal point to select and hence it was selected.

PRECISION AND RECALL TRADE OFF POINT

- We created a graph which will show us the trade off between Precision and recall.
- We found that there is a trade off between Precision and Recall and the meeting point is approximately at 0.3.



Conclusion:

- The model was run on the test data and the following results were obtained:
- Accuracy: 77.52% Sensitivity: 83.01% Specificity: 74.13% These results are similar to the results obtained on the training data, which suggests that the model is performing well.
- The important features that contribute to the probability of a lead being converted are:
- Lead Origin: Lead Add Form
- What is your current occupation: Working Professional
- Total Time Spent on Website:
- These features can be used to target marketing campaigns more effectively. For example, if a company is trying to sell a product or service to working professionals, they may want to focus their marketing efforts on leads that have those characteristics.
- Here are some additional insights that can be gained from this information:
- The model is more likely to correctly predict that a lead will convert if the lead came from the Lead Add Form.
- The model is more likely to correctly predict that a lead will convert if the lead is a working professional.
- The model is more likely to correctly predict that a lead will convert if the lead spent a longer time on the website.
- This information can be used to improve the conversion rate of a company's marketing campaigns