LEAD SCORING- Case Study Assignment

BY:-

1. PRAVIN KOLI

2. AKSHAY MOHAN S

3. DR. MADHURI SONAWALE NAIK

PROBLEM STATEMENT

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.

OBJECTIVES

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. here 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

Data Preparation

1. Impute the raw data, shape-9240-rows, 37-columns.

2. After Analysis dropped the irrelevant variables.

3. Select values were replaced by nan.

4. Imputed null values.

5. Data imbalance was done and dropped the imbalanced columns.

6. Outliers were detected and treated.
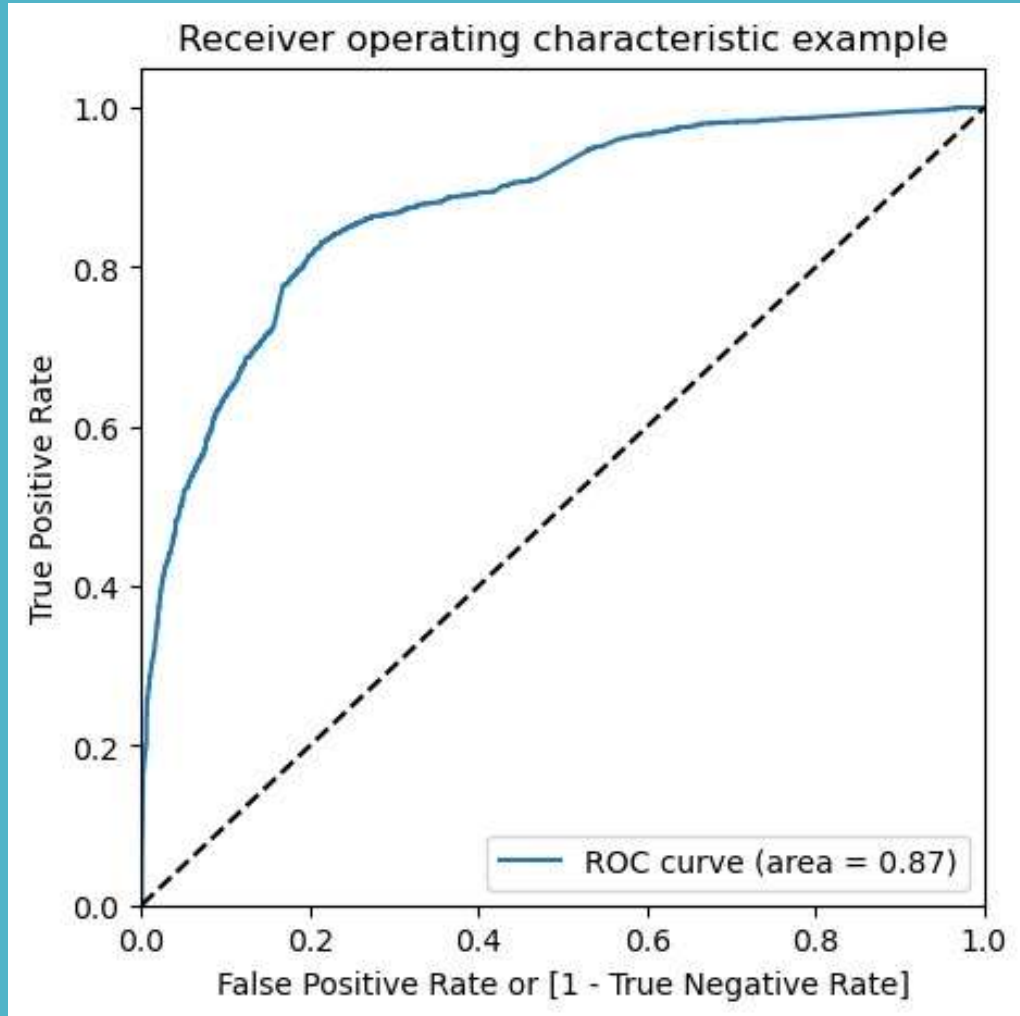
7. EDA

8. Dummy Variable creation

Model Building

1. Train test split-Define X & y

2. Feature Scaling

3. Model Building (total 4 Models built)

4. Checking ViFs

5. Plotting the ROC Curve

6. Check Overall Accuracy, Specificity, Sensitivity.
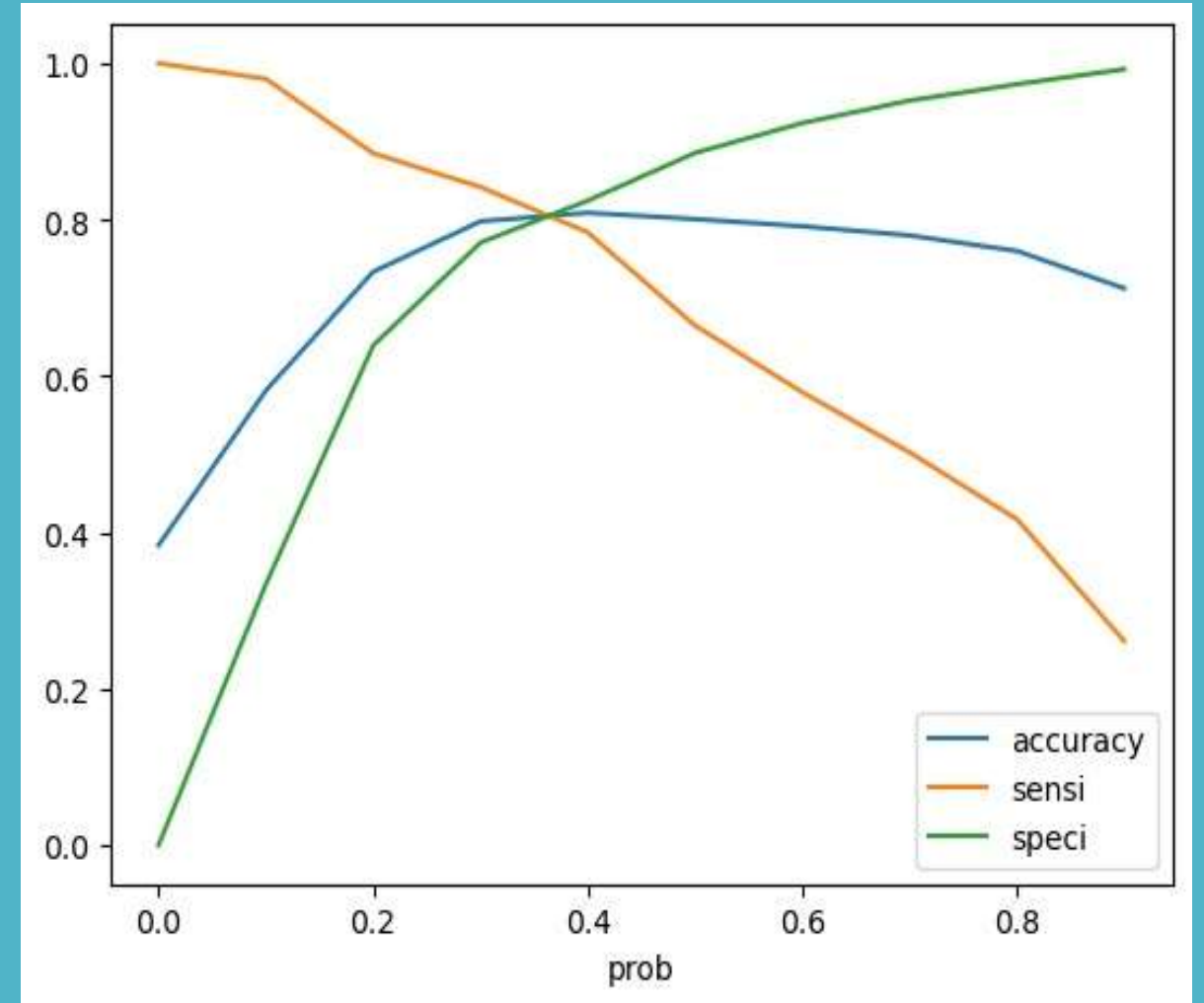
7. Making Prediction on test dataset

Correlation between selected features
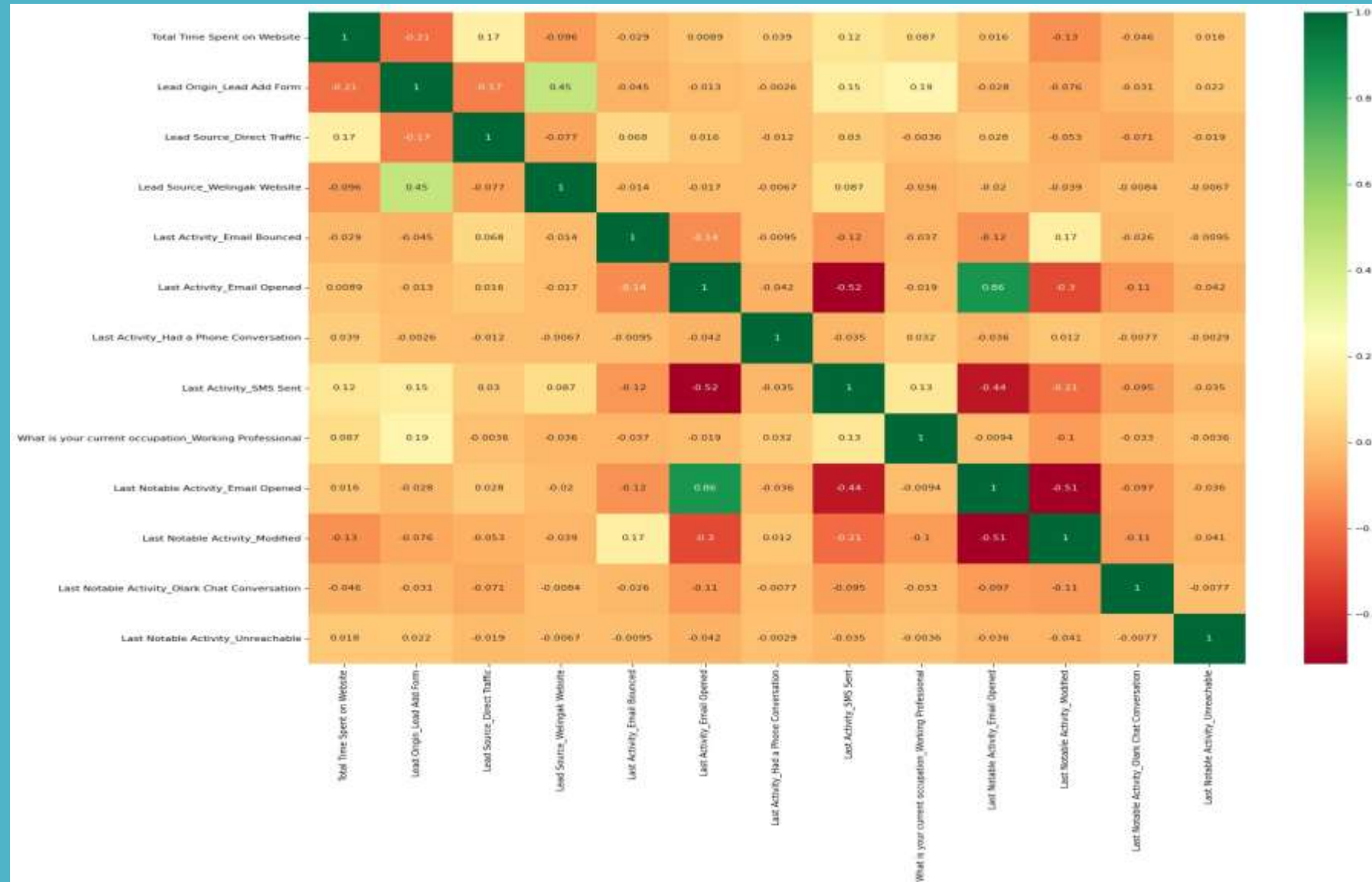
ROC Curve

Cutoff Point

Lead Score

	Prospect ID	Converted	Converted_prob	Lead_Score	final_Predicted
0	7191	0	0.923931	92	1
1	1008	1	0.199180	20	0
2	3494	1	0.737436	74	1
3	3298	1	0.481293	48	1
4	5646	1	0.852627	85	1

Recommendations:-

1. Concentrate on Lead Origin_Add Form
2. More Focus on Working Professionals.
3. Should Pay more attention to phone conversation and SMS sent.
4. Keep an Eye on the Welingak website visitors for more Leads.
5. Make Engageable website so that the visitors spend more time on the website.
6. Traffic on the online sources should be reduced.
7. Provide correct contact details.