LEAD SCORE-CASE STUDY

SUBMITTED BY:-

GARIMA JOSHI

GARIMA MISHRA

GAYATRI BHARADWAJ GADDAM

PROBLEM STATEMENT

- Online company X, offers online courses to professionals so that they can upskills and performs better at their job.
- They got lots of Registration but there are very few who enrols for their course.
- X wishes to identify the professionals who have very high chances of enrolling, they are called hot leads.
- If they successfully identifies the lead who actually enrolls, their business will go up, and they will focus on those leads only rather then focusing on everyone who registers.

BUSINESS OBJECTIVE

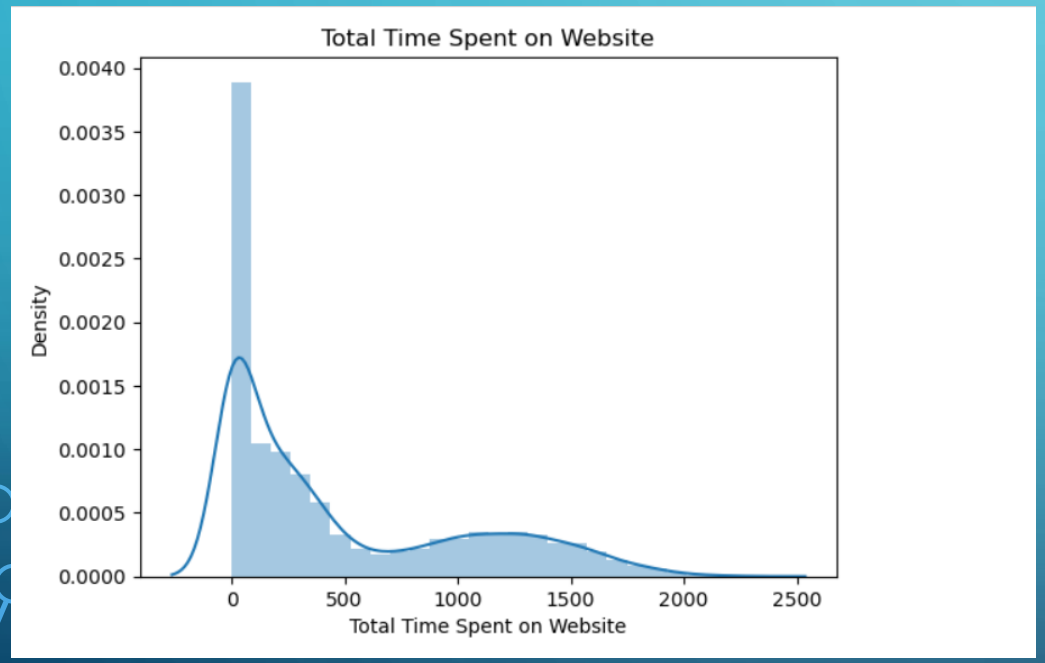
- X wants to know all the possible hot leads.
- X CEO want the conversion rate to be 80%.
- They want to analyse the data and at the end identify the potential leads.
- They want to increase the profit of the company.

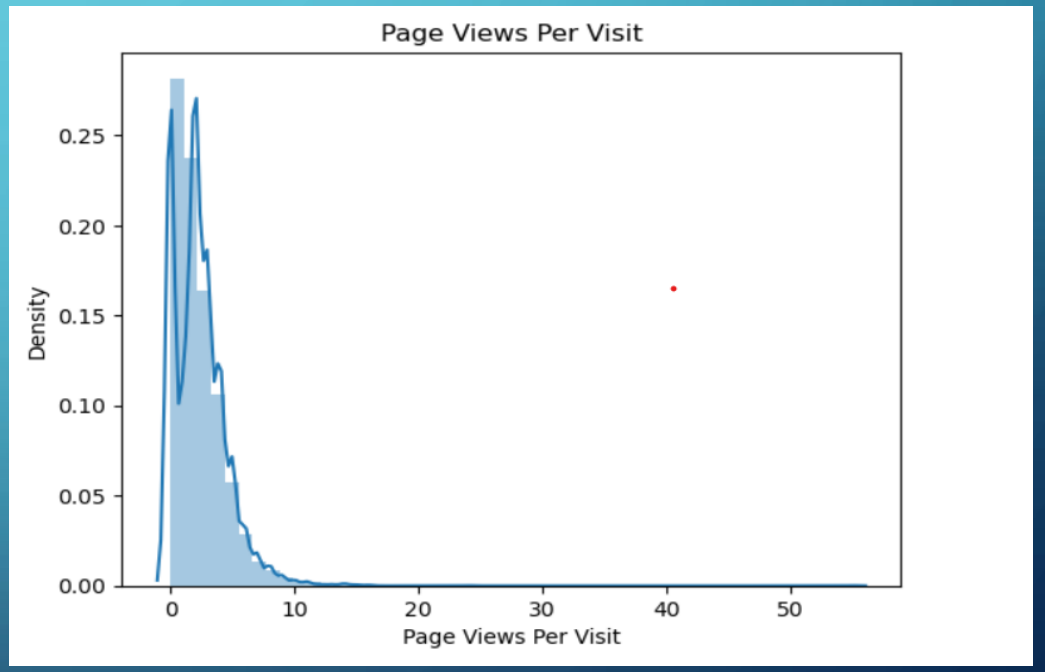
PROBLEM SOLVING APPROACH

- Importing the data and reading and inspecting it.
- Data cleaning
- EDA
- Dummy variable creation
- Test –train split
- Feature Scaling
- Model building
- Model Evalutaion
- Prediction

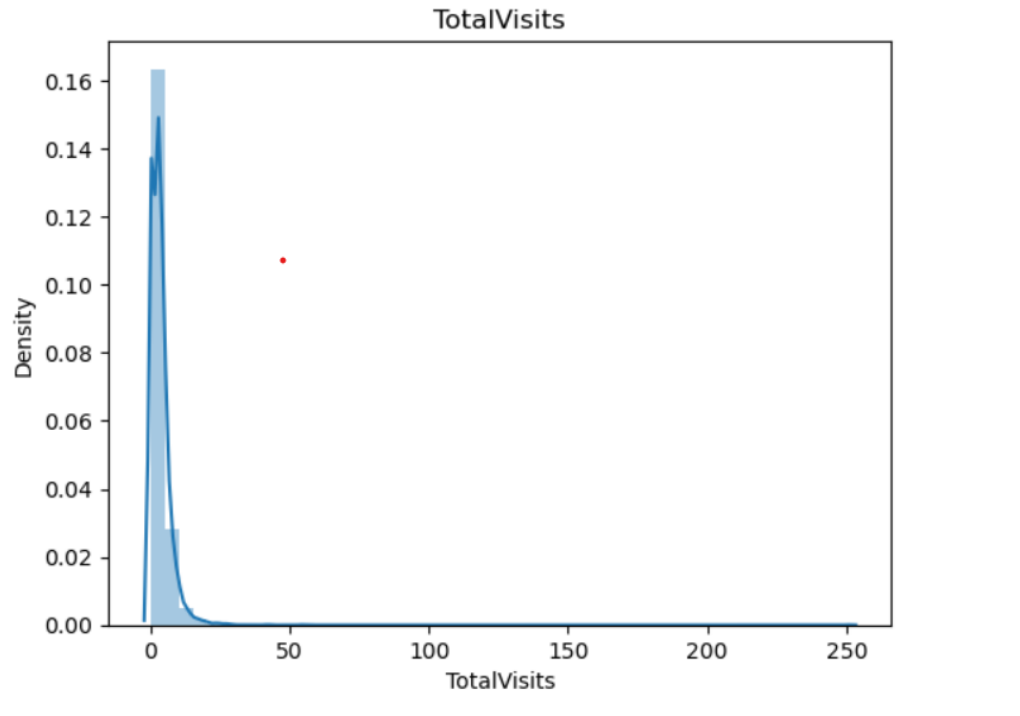
- Rows which are not useful for analysis are deleted like id, tags etc.
- Some columns have “Select” Data which is been taken care of.
- Rows which have less null values are imputed .
- Dummy variables are created for all categorical columns.
- Scaler used is MinMaxScaler().
- Accuracy, precision, recall, sensitivity specificity is calculated for both train and test data set.

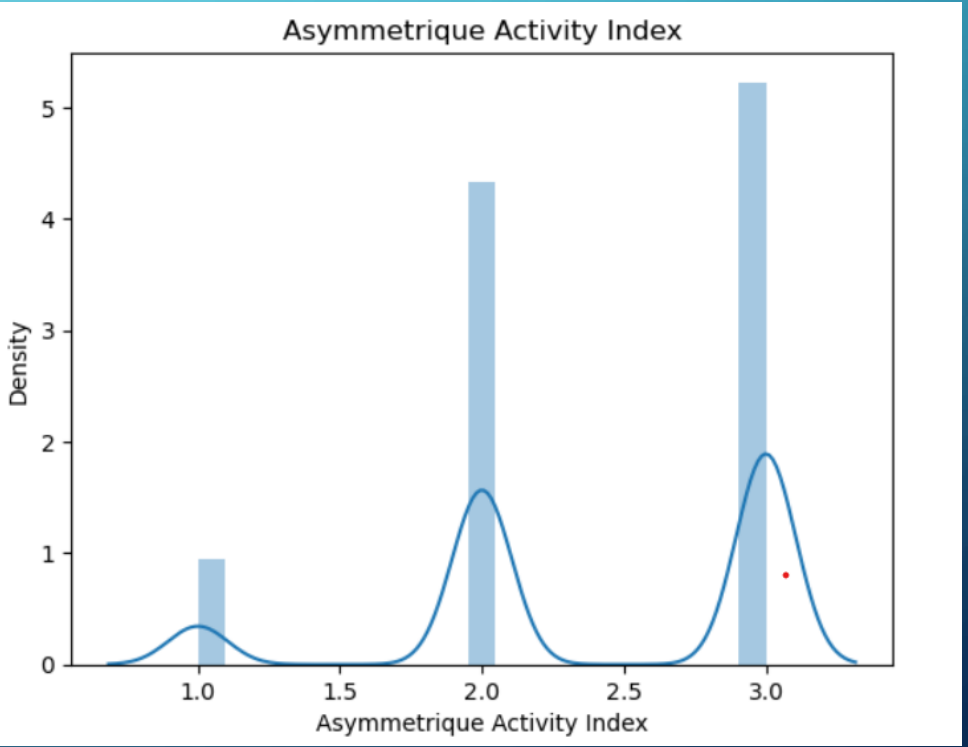
EDA


Total Time Spent on Website
Density
0.0040
0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0
500
1000
1500
2000
2500
Total Time Spent on Website

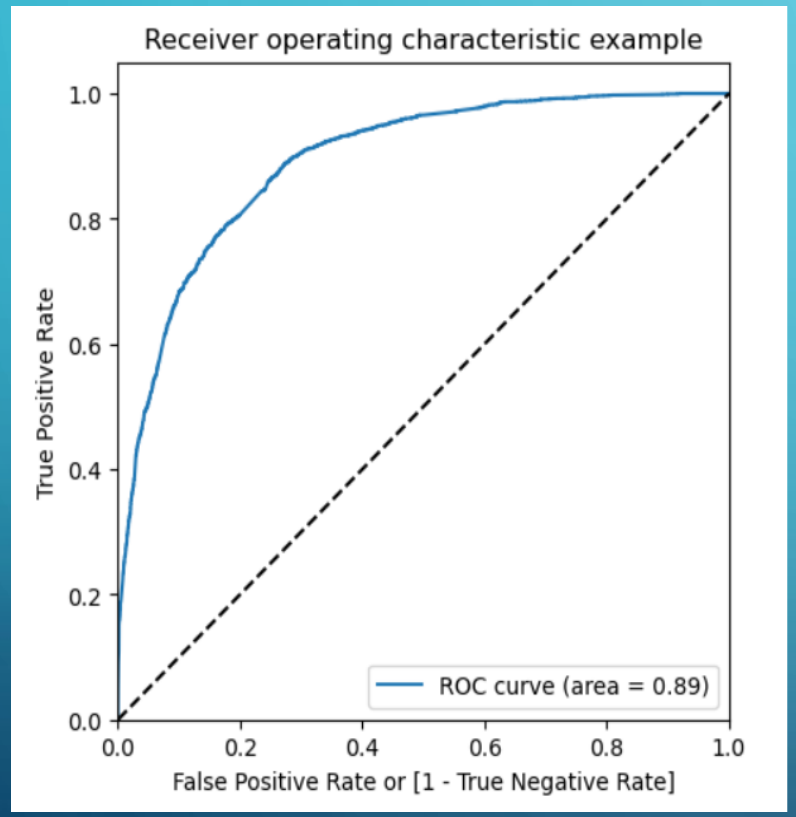

Page Views Per Visit
Density
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
Page Views Per Visit

EDA- CONTINUED

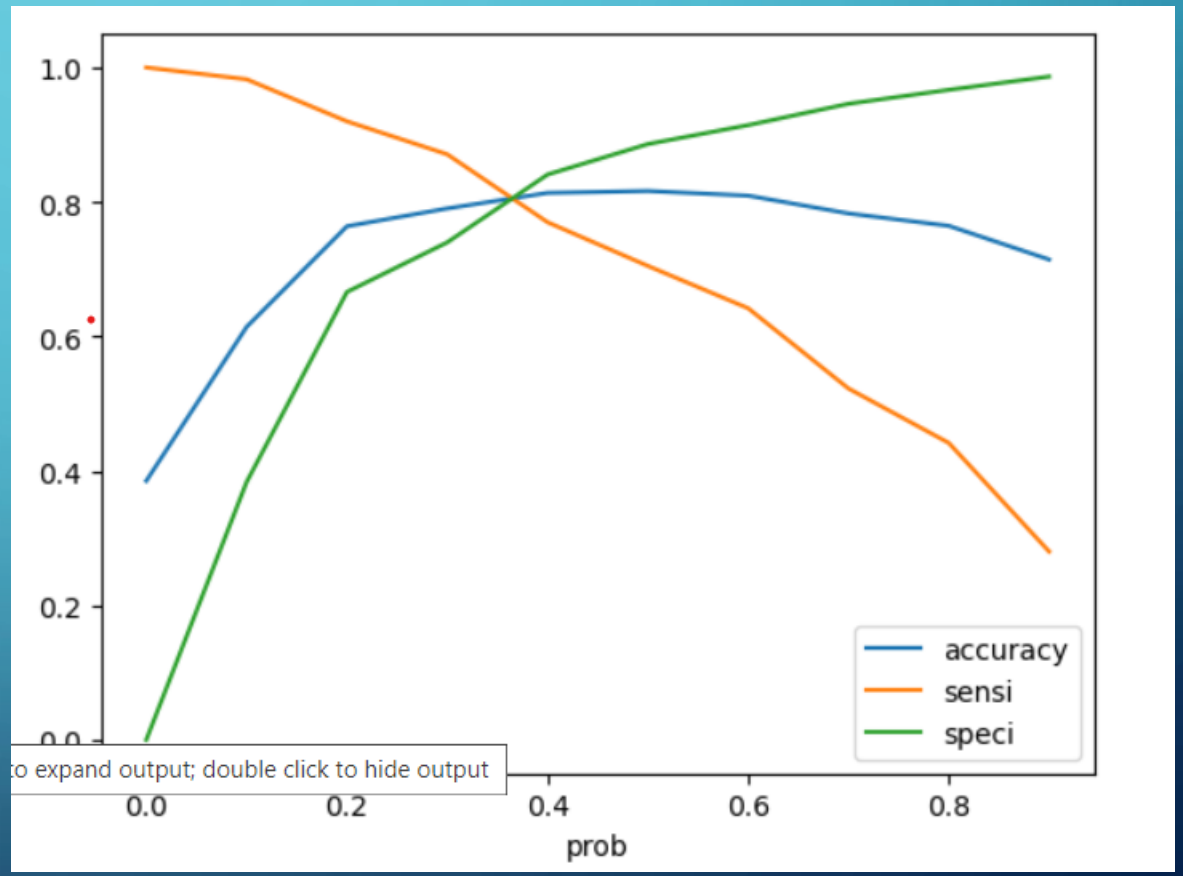




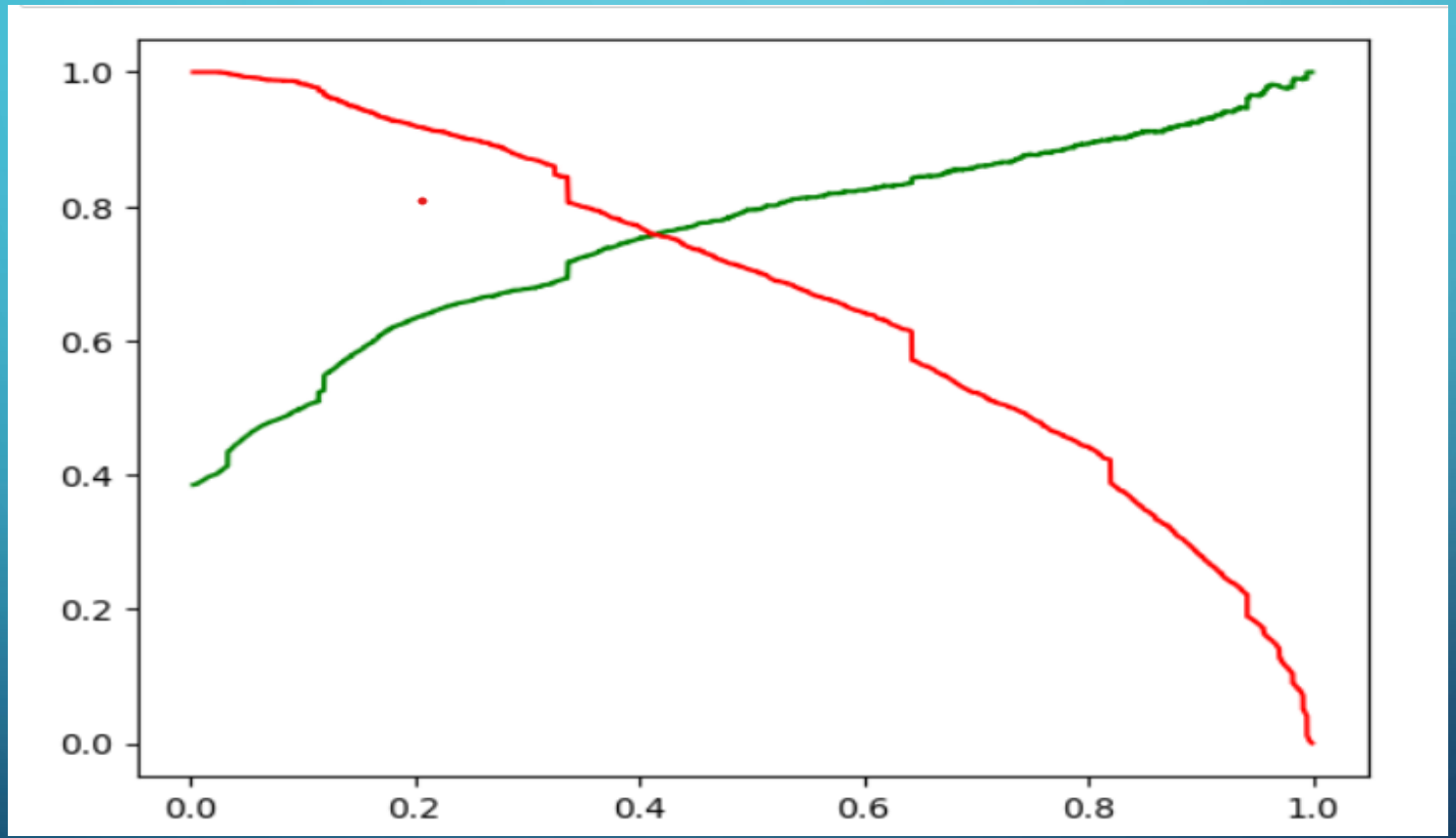
MODEL EVALUATION (ROC CURVE)



Roc=0.89, which is quite good



Cutoff= 0.4 approx



Green color line: precision, red clolor line : Recall

CONCLUSION

- Those who spend more time on X, likely to get converted.
- High number of leads are generated through google and direct Traffic generally.
- Max conversions are working professionals.
- Least conversions are unemployed.
- Those who visits X more, are potential lead, say hot leads