Lead Scoring Case Study

- Name: Prabin N
- Batch: DS C-57 June 2023

Note: - Submission done individually as my group members not co-operative.

Problem Statement:

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.

- 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, this means 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

Business Goal:

- Lead X wants to know the conversion rate to be increased and need to get 80% of conversion rate.

- Lead X wants to know the Hot leads that increase the conversion rate.

- Lead X wants the prediction model to handle the future peak actions and utilize the man power to achieve the target.

<u>Action Taken:</u>

- Imported the data set and required libraries for analysis.

- Data Cleaning and used sweetviz to understand the data set clearly

- Data Modification

- Data Engineering

- Dummy Variables

- Train Set and Test Set

- Scaling the Train set and Test set

- RFE and VIF analysis

- Confusion matrix and Precision & Recall

EDA Data Cleaning:

- There are few columns has 'Select' values where customer not filled any details so they are considered as Null values.

Specialization:

```
Select                              1942
Finance Management                   976
Human Resource Management            848
Marketing Management                 838
Operations Management                503
Business Administration              403
IT Projects Management               366
Supply Chain Management              349
Banking, Investment And Insurance    338
Travel and Tourism                   203
Media and Advertising                203
International Business               178
Healthcare Management                159
Hospitality Management               114
E-COMMERCE                           112
Retail Management                    100
Rural and Agribusiness                73
E-Business                            57
Services Excellence                   40
```

How did you hear about X education:

```
Select                   5043
Online Search             808
Word Of Mouth             348
Student of SomeSchool     310
Other                     186
Multiple Sources          152
Advertisements             70
Social Media               67
Email                      26
SMS                        23
```

EDA Data Cleaning:

- There are few columns has 'Select' values where customer not filled any details so they are considered as Null values.

Lead Profile:

```
Select                        4146
Potential Lead                1613
Other Leads                    487
Student of SomeSchool          241
Lateral Student                 24
Dual Specialization Student     20
```

City:

```
Mumbai                           3222
Select                           2249
Thane & Outskirts                 752
Other Cities                      686
Other Cities of Maharashtra       457
Other Metro Cities                380
Tier II Cities                     74
```
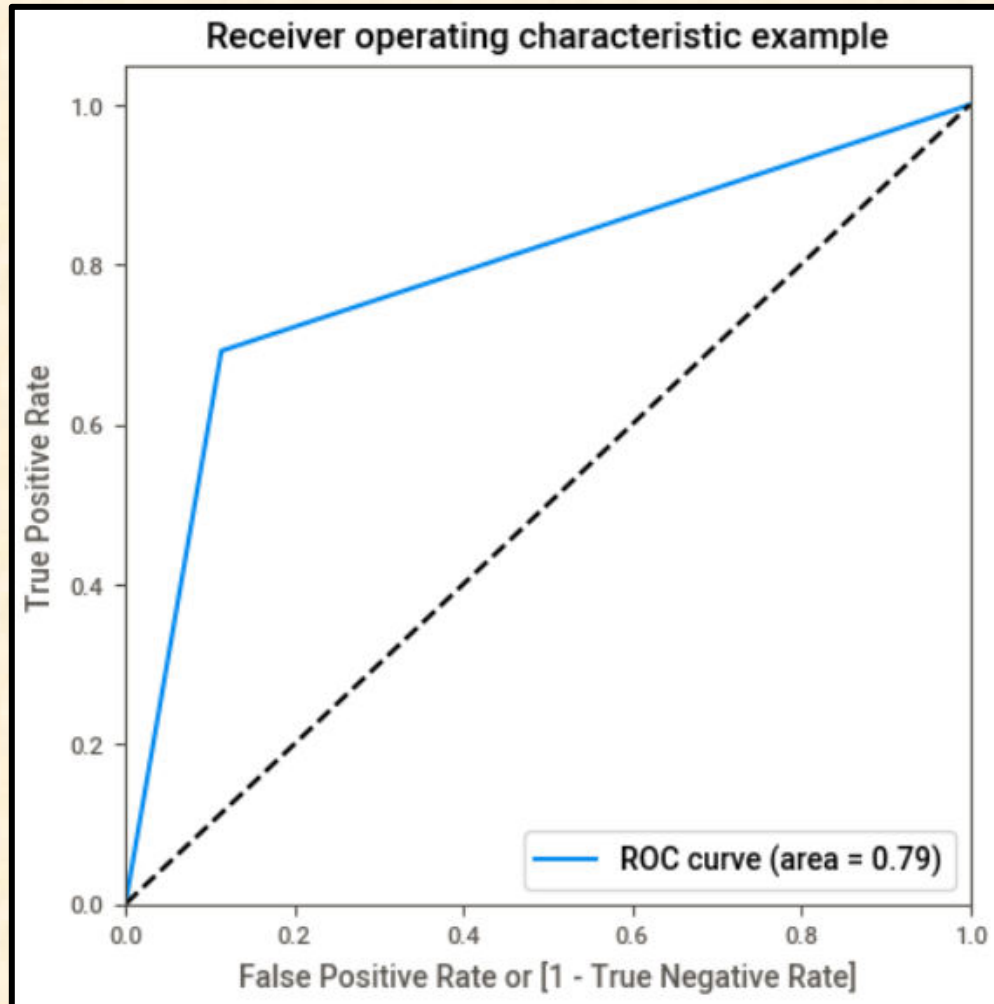
Visualization:

- Used distplot on 'Total Visit' 'Total time spend on website' and 'Page views per visit' to understand the skewness. All three category are right skewness so the outliers are to the right.
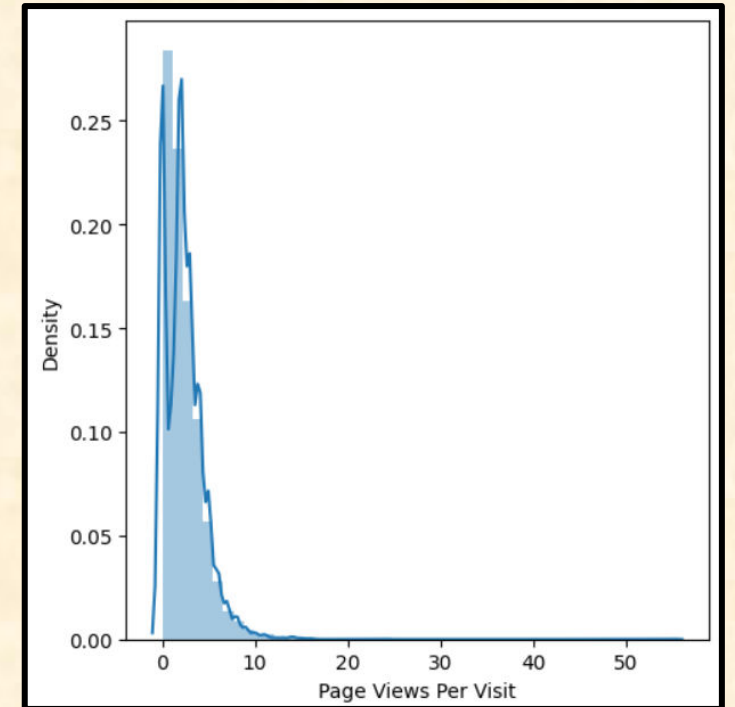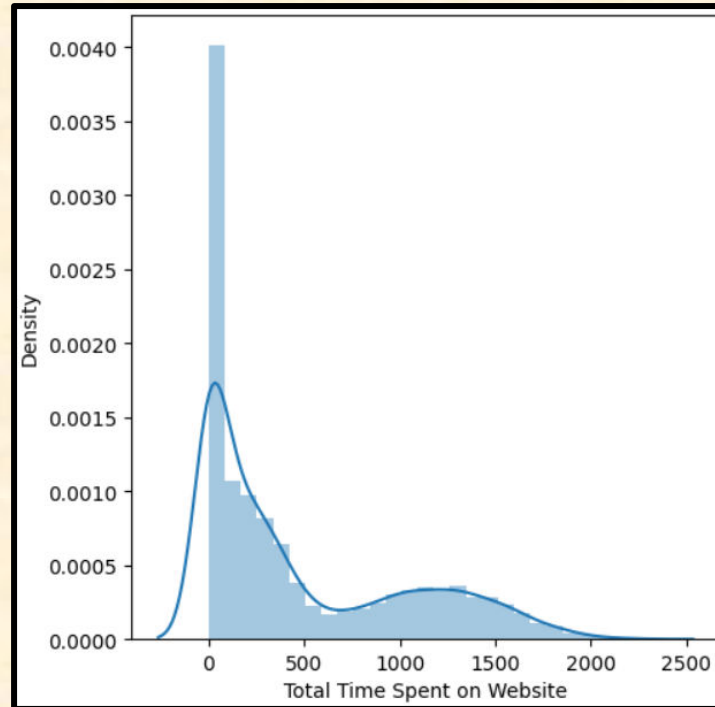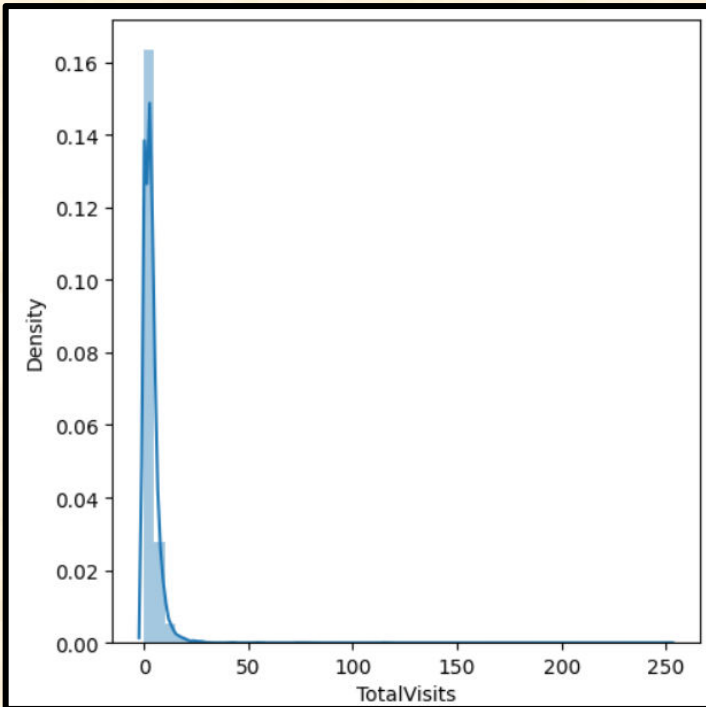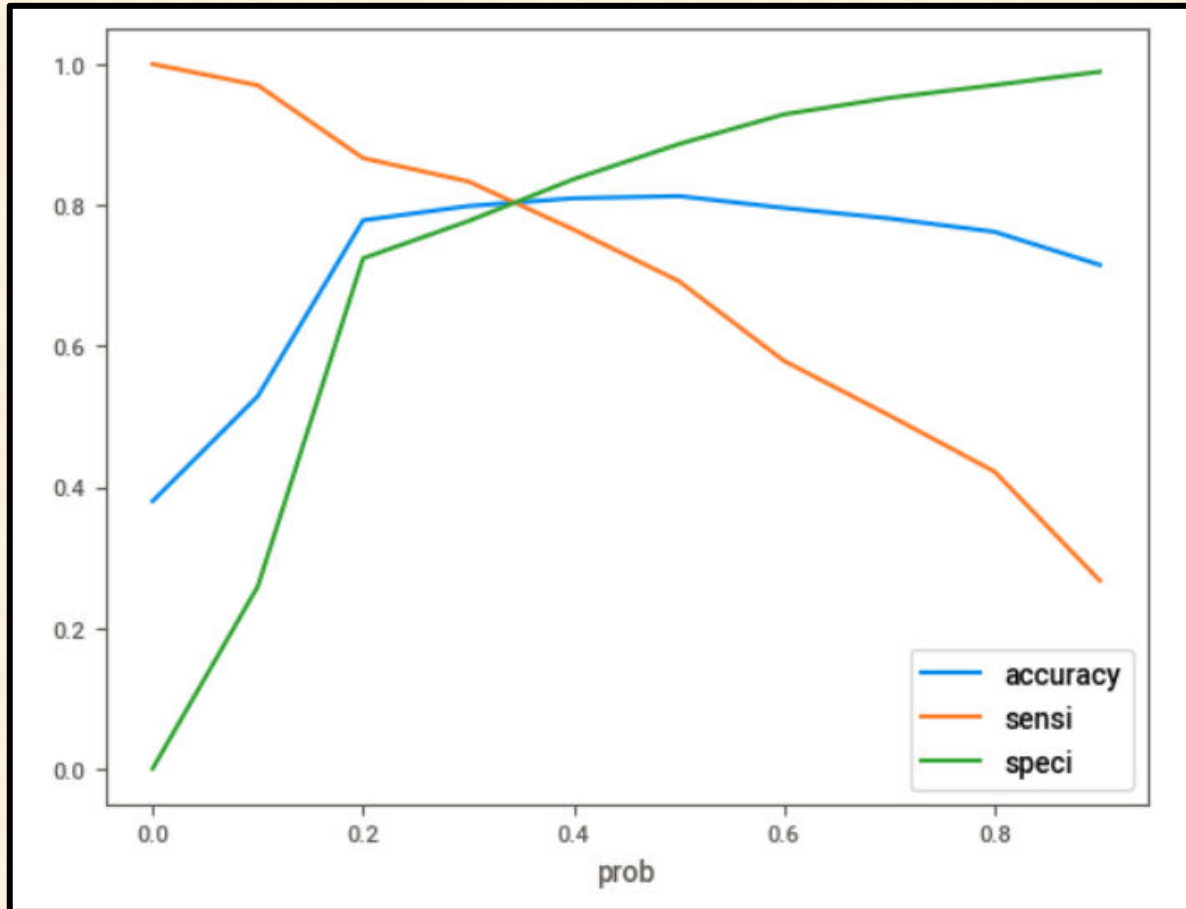
ROC Graph:

- **True positive = confusion[1,1]**

- **False positive = confusion[0,1]**

- **True negative = confusion[0,0]**

- **False Negative = confusion[1,0]**

Sensitivity, Specificity, Accuracy graph:

- **Sensitivity: 69.20%**

- **Specificity: 88.64%**

Conclusion:

The company has to set sales teams to focus on the below to convert maximum leads.

• Leads from unemployed category.

• Leads who spend more time and visited website more frequently.

• Start calling customer with right education counsellor to make understand about the X Education system to customer.

• Focus on the customer who wants get better career prospects.

• Maximum leads are from google platform.