LEAD SCORE CASE STUDY

Submitted by:

Anuj Gupta

Indhumathi S

LEAD SCORE CASE STUDY FOR X EDUCATION

Problem Statement :

X Education sells online courses to industry professionals. 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. When these people fill up a form providing their email address or phone number, they are classified to be 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, although X Education gets a lot of leads, its lead conversion rate is very poor. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.

LEAD SCORE CASE STUDY FOR X EDUCATION

Problem Statement : (Contd)

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Goals of the Case Study:

1.To identify the hot leads , so that sales team can focus on the hot leads.
2.Company require to build a model where we need to assign a lead score to each of the lead so that customer which having high score have high conversion chances. The score should be between 0 and 100.
3.To get to know the which columns are most important and which is not important.

APPROACH FOR THE CASE STUDY

➢ Importing the data and inspecting the data frame
➢ Data preparation
➢ EDA
➢ Dummy variable creation
➢ Test-Train split
➢ Feature scaling
➢ Feature Selection
➢ Model Building (RFE ,VIF and pvalues)
➢ AUR-ROC Curve
➢ Accuracy sensitivity and specificity trade off value
➢ Making the Confusion matrix
➢ Making predictions on test set
➢ Conclusion and Recommendations

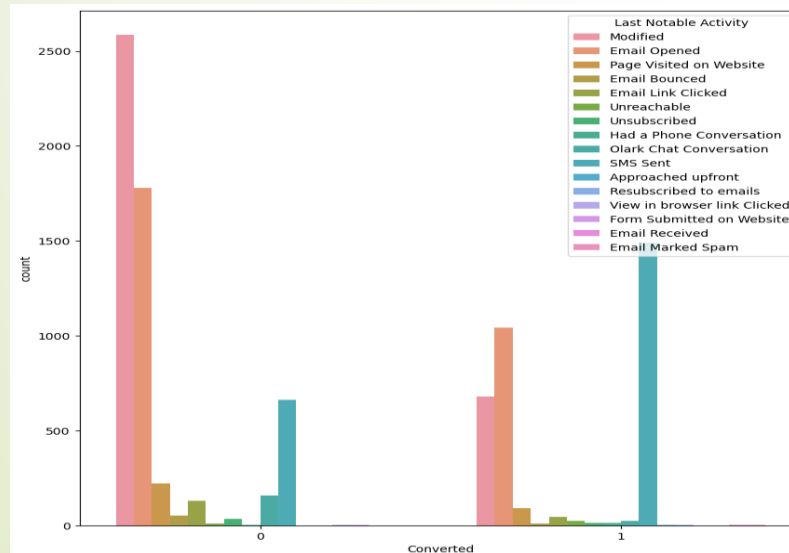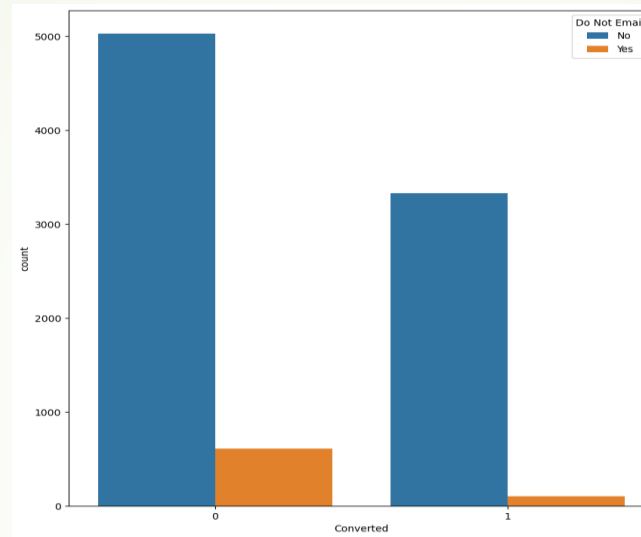# EXPLORATORY DATA ANALYSIS
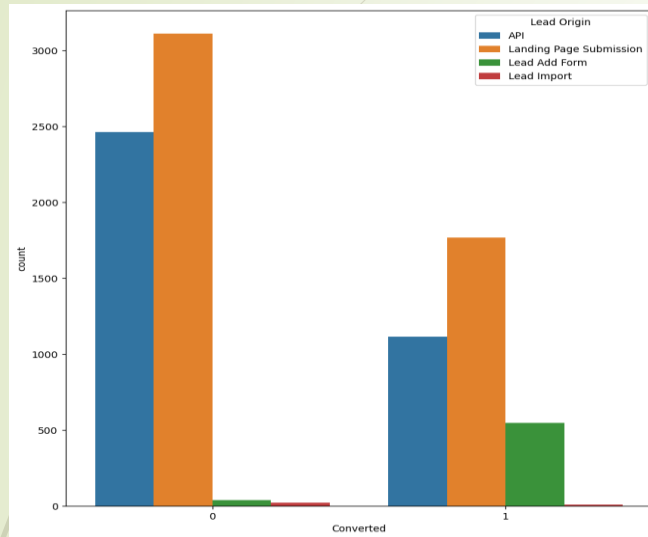
Data Imbalance chart
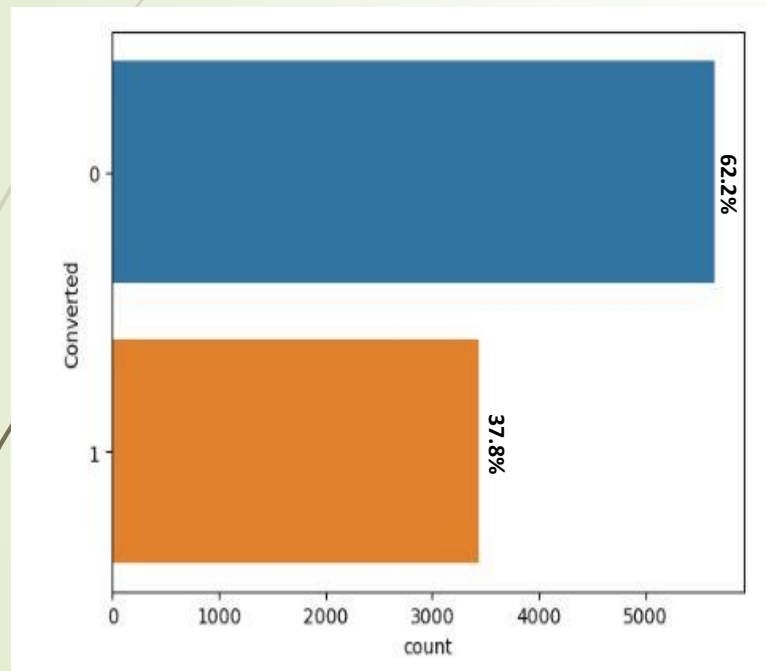
Univariate Analysis
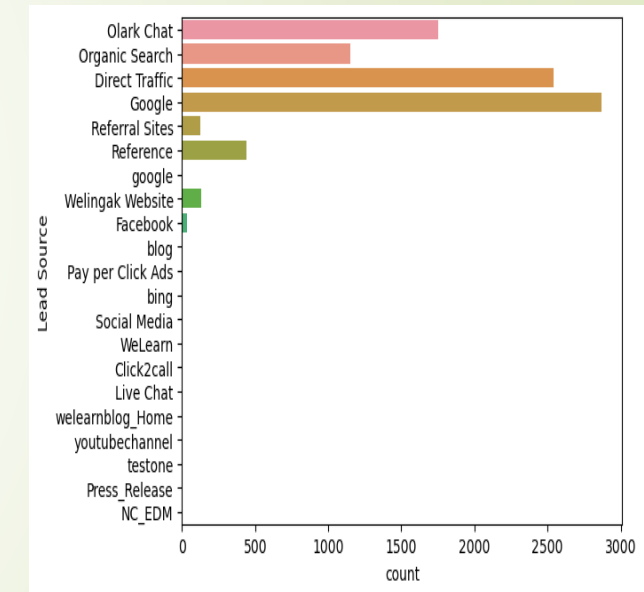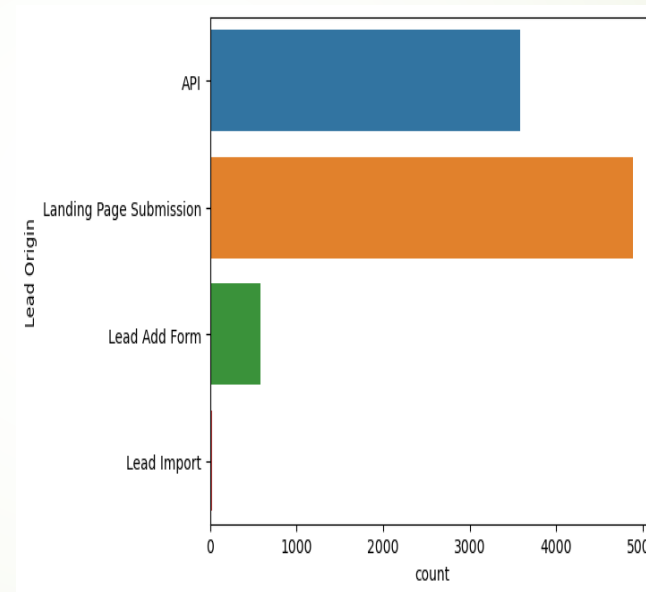
- In Lead Origin, maximum conversion happened from Landing Page Submission
- Major conversion in the lead source is from Google

BIVARIATE ANALYSIS

Inference :

1. Most of the leads are coming from landing page Submission

2. Most of the people say "no" in do not email but ratio of no vs yes is more in not conversion lead.

3. If a person marked the email as a spam , it generally means he is not intrested in the course , it is the waste of time to run behind him.
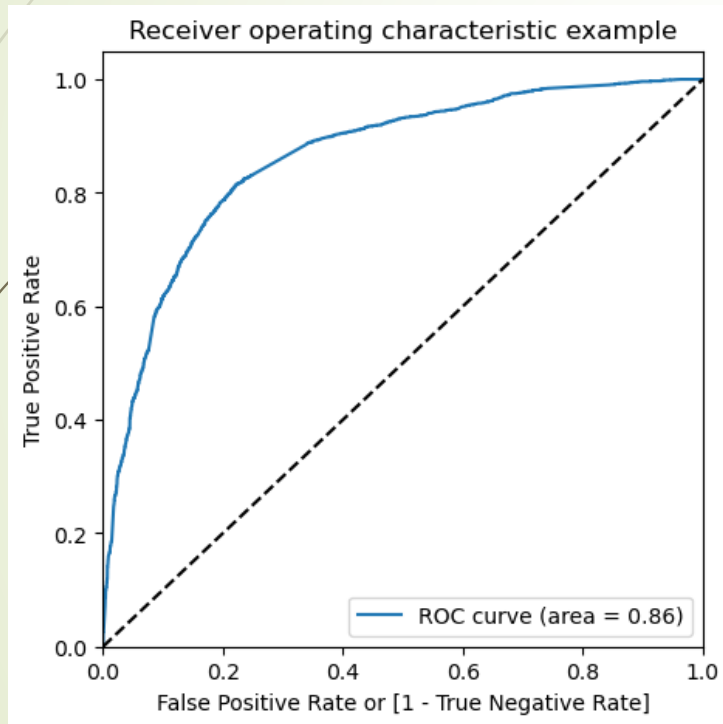
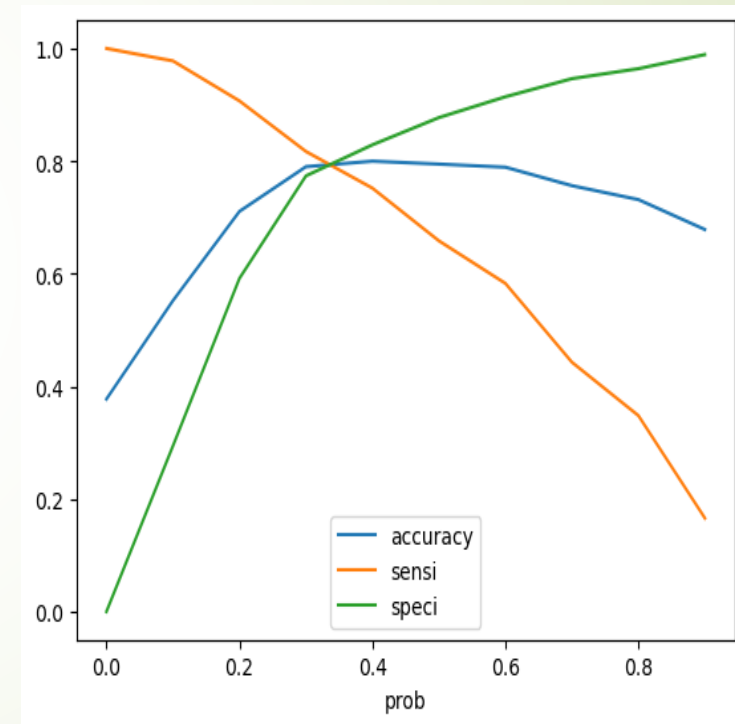# BIVARIATE ANALYSIS
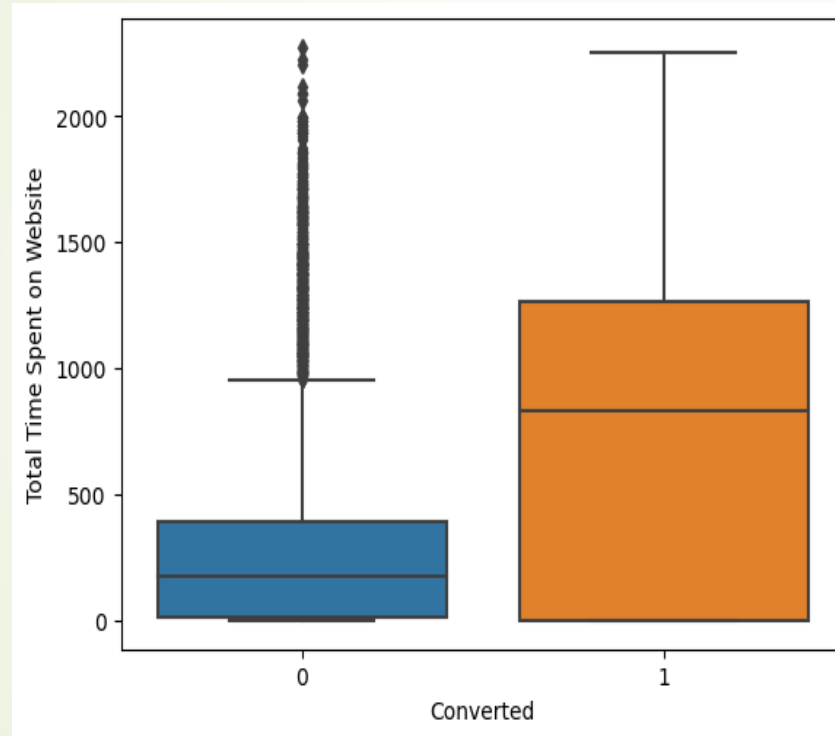
In "Total Time Spent on website" mean is more in converted lead than non- converted lead.

MODEL EVALUATION

ROC Curve

Model is good as the area is high as 0.86

0.33 is the cut off between sensitivity and specificity

CONCLUSION

Train Data Accuracy, Sensitivity and Specificity
1. Accuracy   : 79.27%
2. Sensitivity : 80.61%
3. Specificity : 78.45%

Test Data Accuracy, Sensitivity and Specificity
1. Accuracy   : 78.89%
2. Sensitivity : 83.25%
3. Specificity : 76.20%

Lead Count for Different Lead Score

1. Number of rows with Lead_score > 90: 228 counts

2. Number of rows with Lead_score > 80: 429 counts

3. Number of rows with Lead_score > 70: 578 counts

4. Number of rows with Lead_score > 60: 813 counts

5. Number of rows with Lead_score > 50: 957 counts

RECOMMENDATIONS

Hot Leads Selection

Segmenting leads according to their lead scores provides valuable insights into their likelihood of conversion. Notably, leads with scores surpassing distinct thresholds such as 90, 80, 70, 60 and 50 cluster prominently, signifying different levels of conversion potential. This understanding enables the implementation of focused engagement tactics aligned with individual lead scores, thus enhancing conversion rates by strategically enhancing interactions.

Feature Importance

1. Lead Source_Welingak Website: 5.598763
2. Lead Source_Reference : 4.148916
3. Last Notable Activity_Had a Phone Conversation: 3.070394