Lead Scoring Case Study

- Dhairyasheel Jadhav & Radhika Hedau

Problem Statement

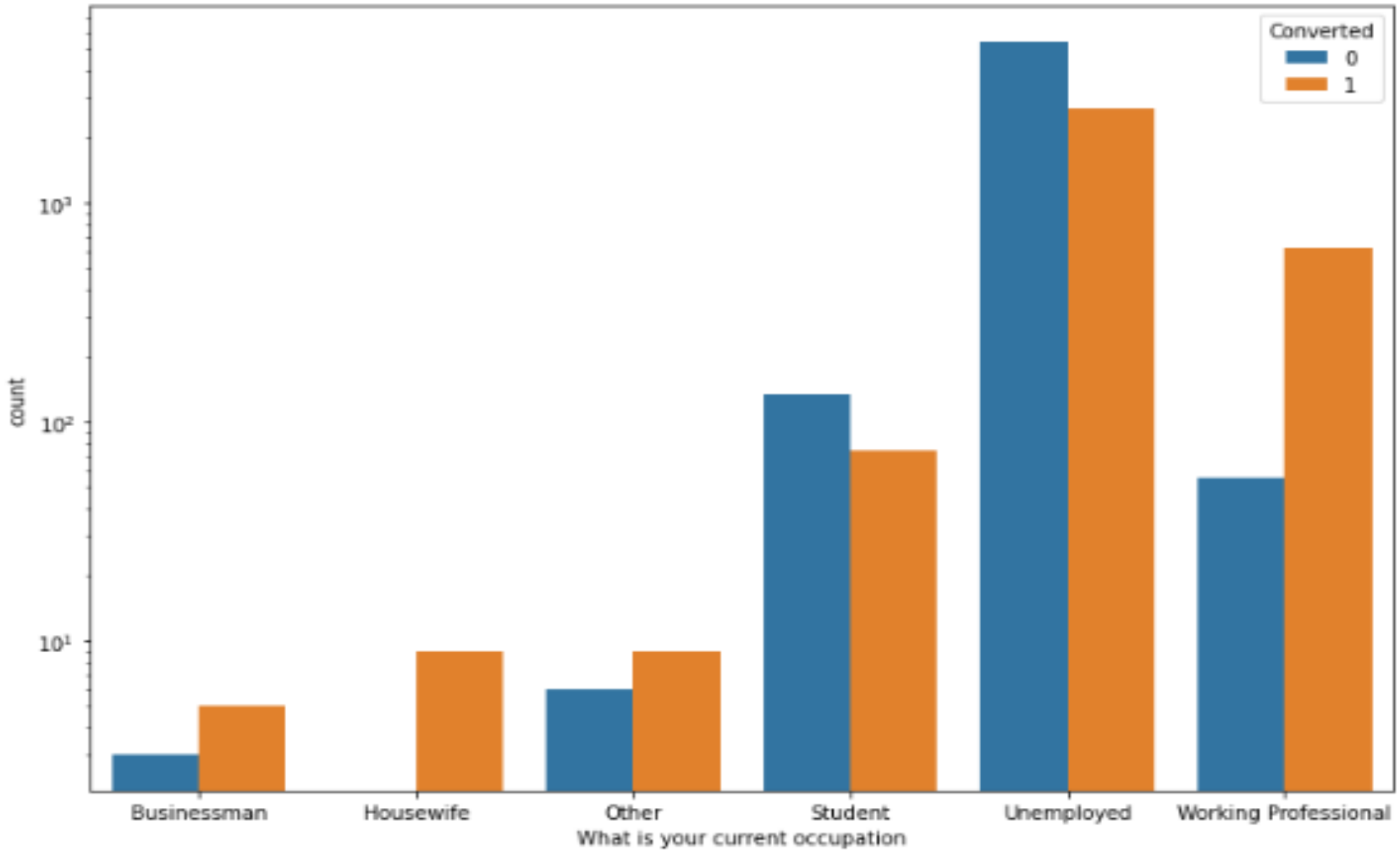
- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with 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%.

Business Goals

- 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.
- There 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 these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

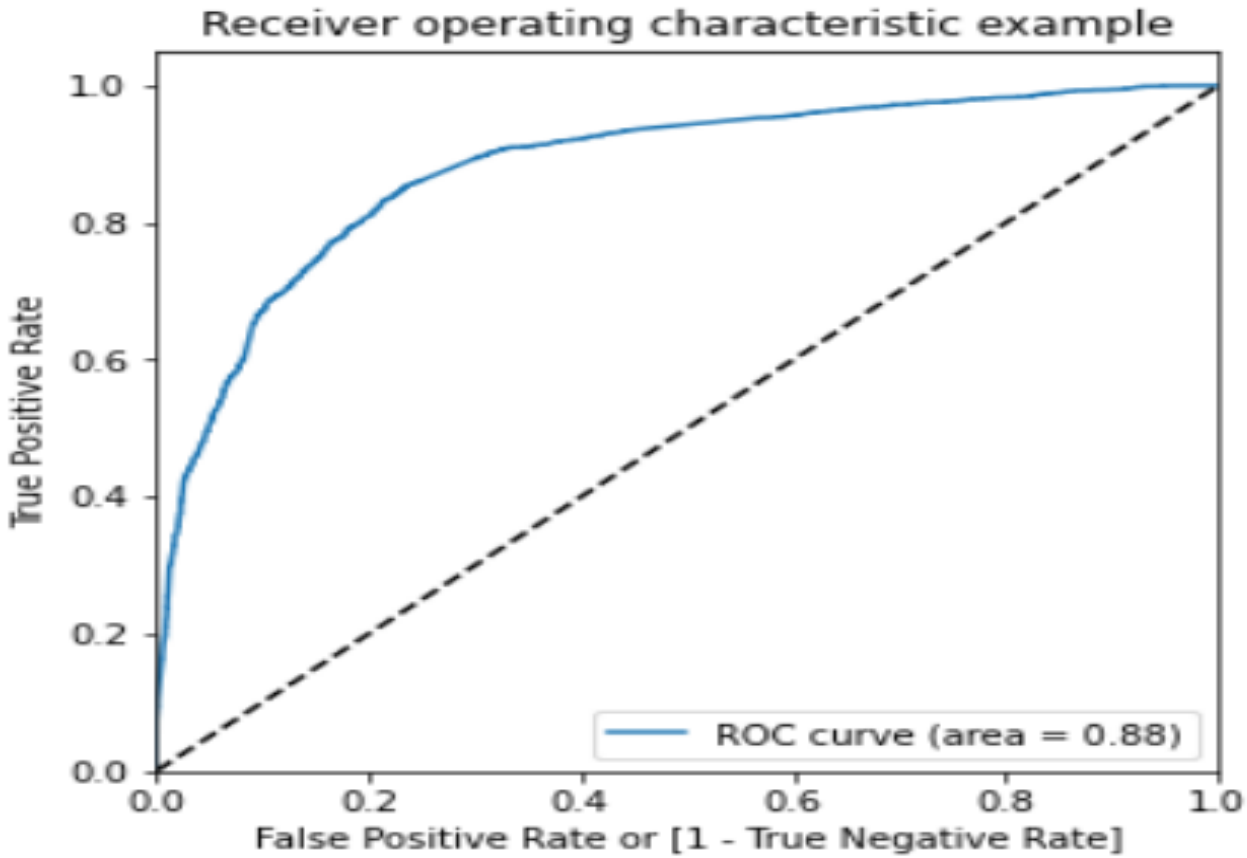
Steps

- Check the Structure of the Data
- Data Quality Check & Missing Values
- Removing Insignificant Columns
- Handling Missing Values
- Data Visualization
- Correlation between Numerical Variables
- Dummy Variables
- Test-Train Split & Scaling
- Correlation among Variables
- Model Building
- Model Evaluation
- Optimize Cut off (ROC Curve)
- Precision-Recall
- Observations & Recommendations
- Conclusion

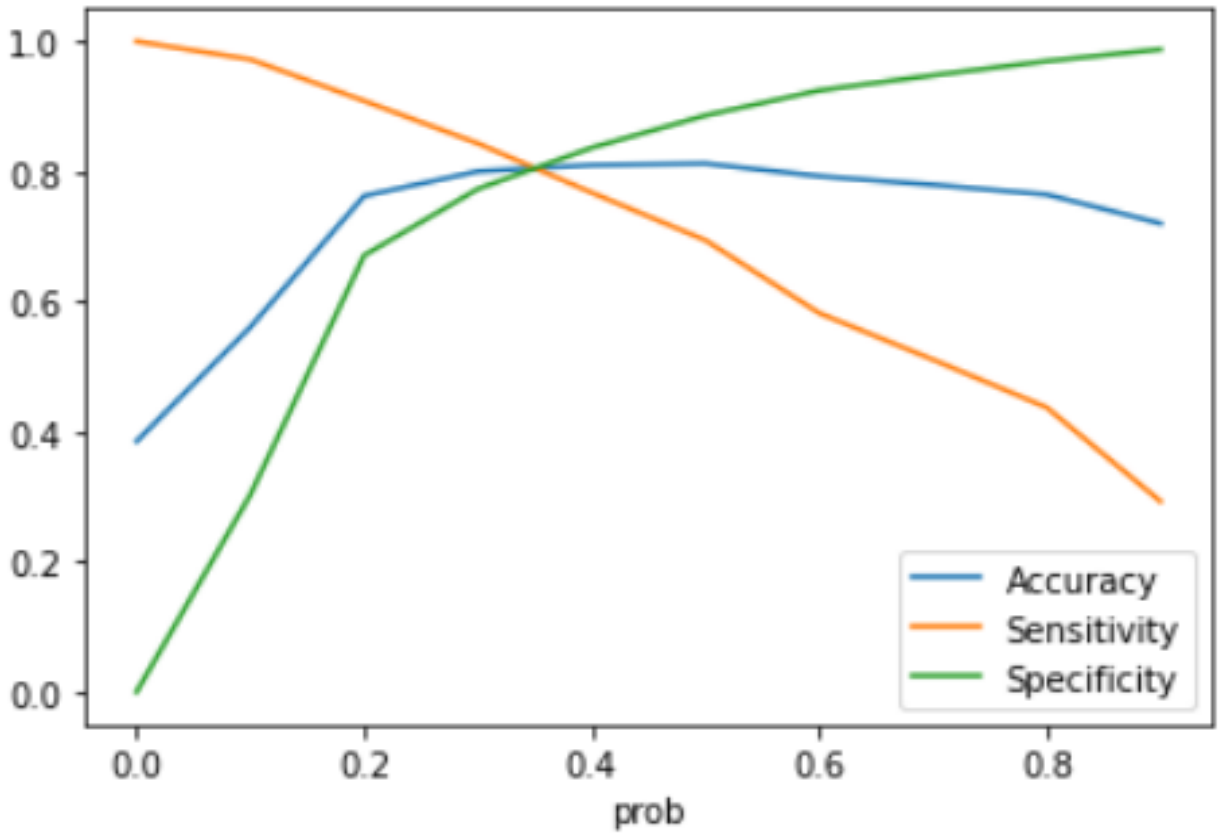

Converted
0
1
count
10^3
10^2
10^1
Businessman
Housewife
Other
Student
Unemployed
Working Professional
What is your current occupation

Insights :

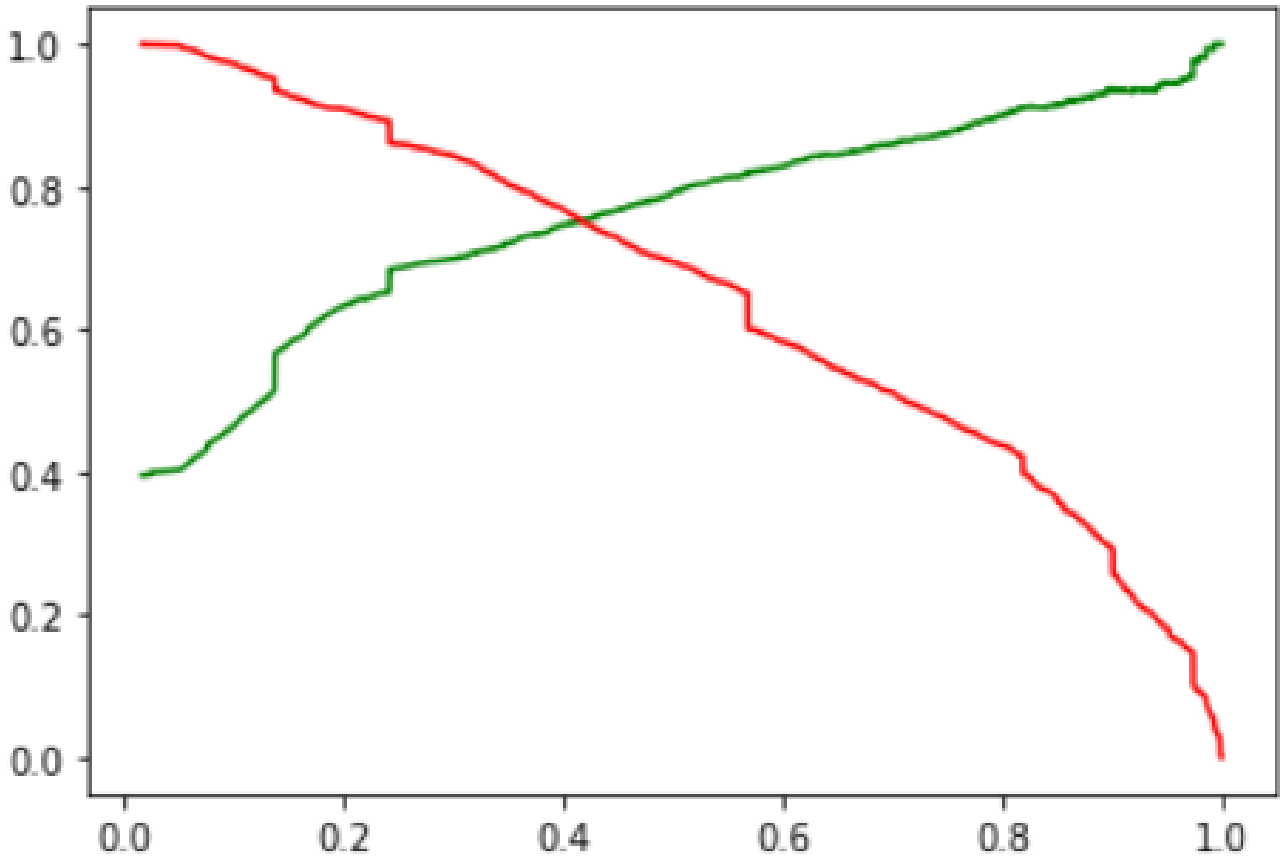
- Lead Origin From Landing Page Submission has Highest Conversion.
- Majority of Customers are from India
- Working Professionals, followed by Unemployed & Students have High Conversion.
- Search Ads have very low Conversion
- Leads are likely to Pay when recommended.
- Leads from Google, Olark Chat & Press Release Ref have high conversion.
- Leads who Resubscribed, Opened Emails or had Phone Conversation are very likely to buy Courses.



- Area under ROC Curve is 88%



- Graph concludes that the optimal cut off is at 0.35.



Intersection of Precision-Recall at 0.43

Observations

- Let us compare the values obtained for Train & Test:

Training DataSet

- Accuracy : 80.42%
- Sensitivity : 80.30%
- Specificity : 80.51%

Test Dataset

- Accuracy : 79.72%
- Sensitivity : 78.86
- Specificity : 80.21

Conclusion

- The top three variables in model which contribute most towards the probability of a lead getting converted are as follows:-

Current Occupation :

> Working Professionals have Highest Lead Conversion followed by Unemployed & Students Business Owners have lowest conversion

Lead Origin :

> Leads from Google, Olark Chat & Press Release Ref have high conversion.

Last Activity :

> Leads who Resubscribed, Opened Emails or had Phone Conversation are very likely to buy Courses.

> I Recommend to Select above Leads to CEO to make target lead conversion rate to be around 80%

Thank You