Lead score case study

Submitted by :

Suchit gurala

Alexy Lalas T A

Problem statement

- X education sells online course to industrial professionals.
- Company markets and advertises on several websites and engines
- The leads provide their data and information in the website
- The sales contact the leads and get converted while most do not .
- The typical lead conversion rate at X education is around 30%

Business goal

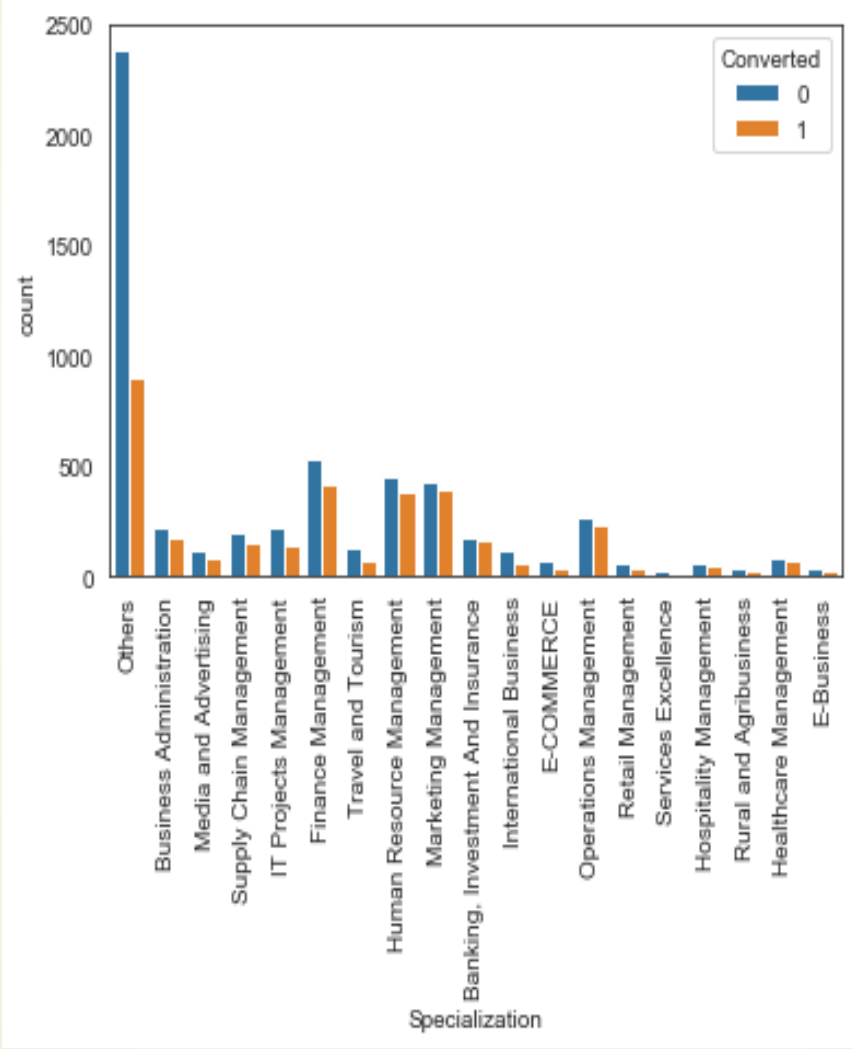
- X education needs help in selecting the most promising leads
- The company needs a model where a score is assigned to each of the leads
- The higher lead score have a high rate of conversion chance and the customers with lower score have a lower conversion chance
- The company needs a total conversion rate of around 80%

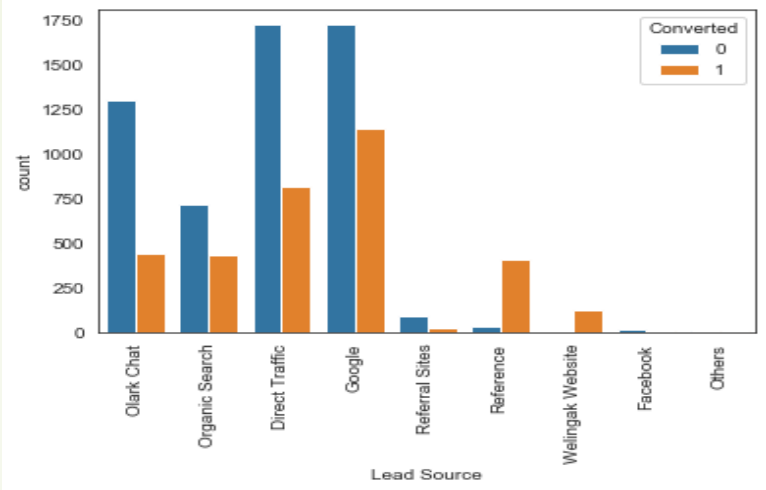
Methodology

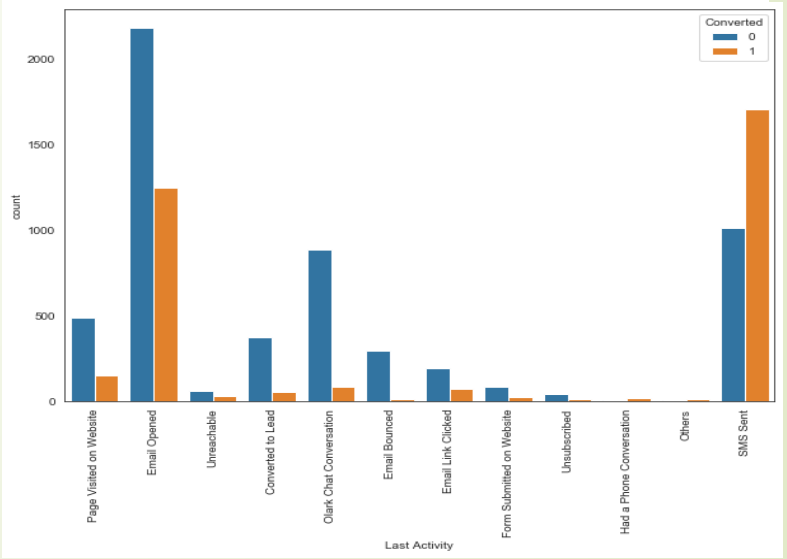
- Source the data
- Cleaning and preparation of data
- Exploratory data analysis
- Feature scaling
- Splitting of data into train and test data set
- Building logistic regression model and calculate Lead Score
- Evaluating the model by using different metrics – Specificity and sensitivity or precision and recall
- Appling the best model in test datta based on the Sensitivity and specificity Metrics

exploratory data analysis

- It clear from the EDA that the conversion rate where higher from google,direct traffic and olark chat
- Email opened are the one in most conversion
- Based on specialization ,marketing field and the OTHERS are the most converted
- We have a conversion rate of around 38%

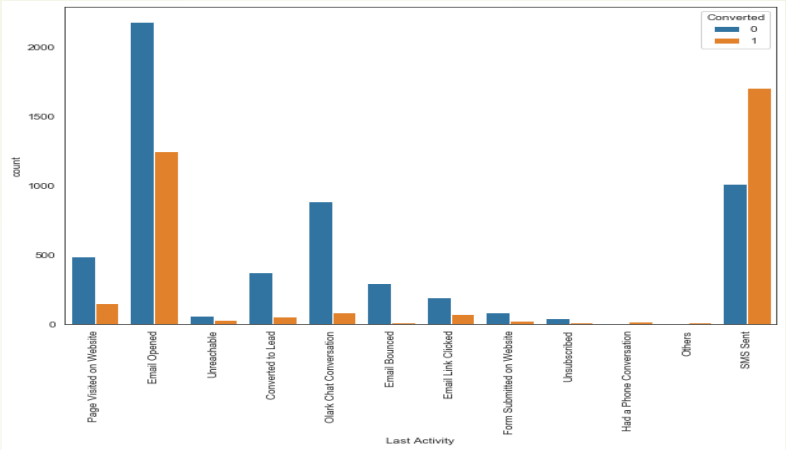


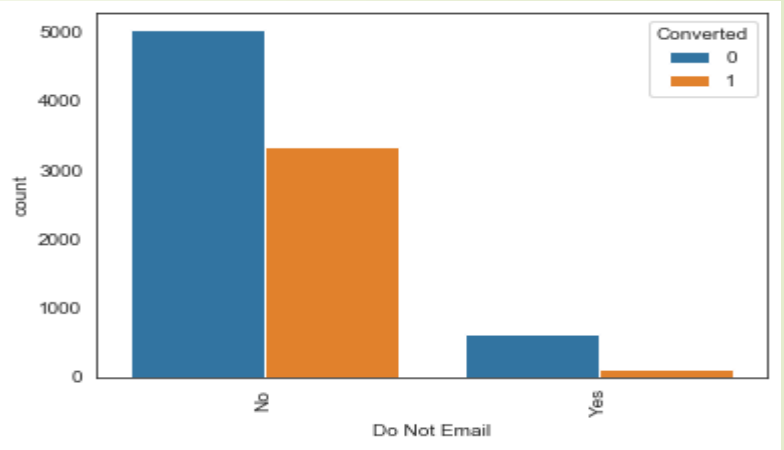


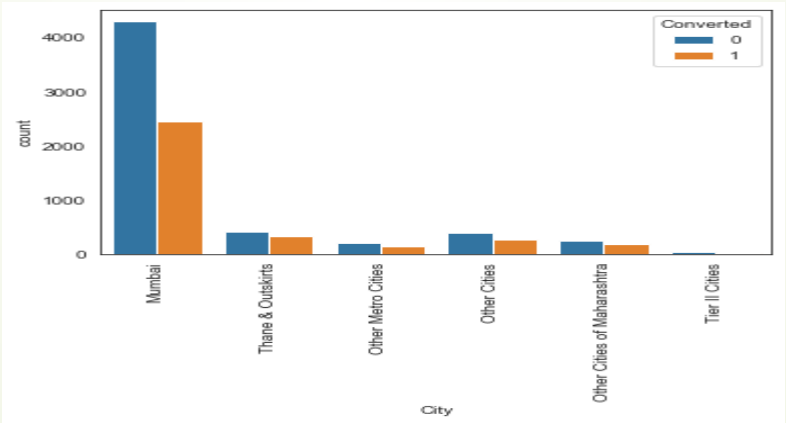


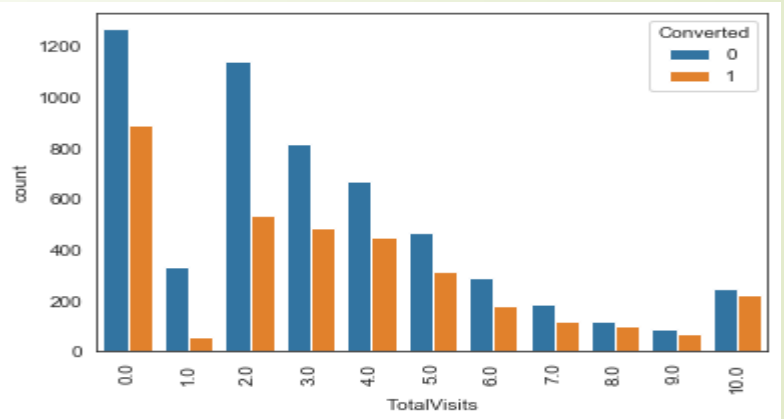
EDA

- Conversion rate is higher for those who have checked their email
- There is a lot of conversion on the bases of there number of visits
- Mumbai is the city with higher amount of leads
- Most of the people are unemployed on the basis of their occupation





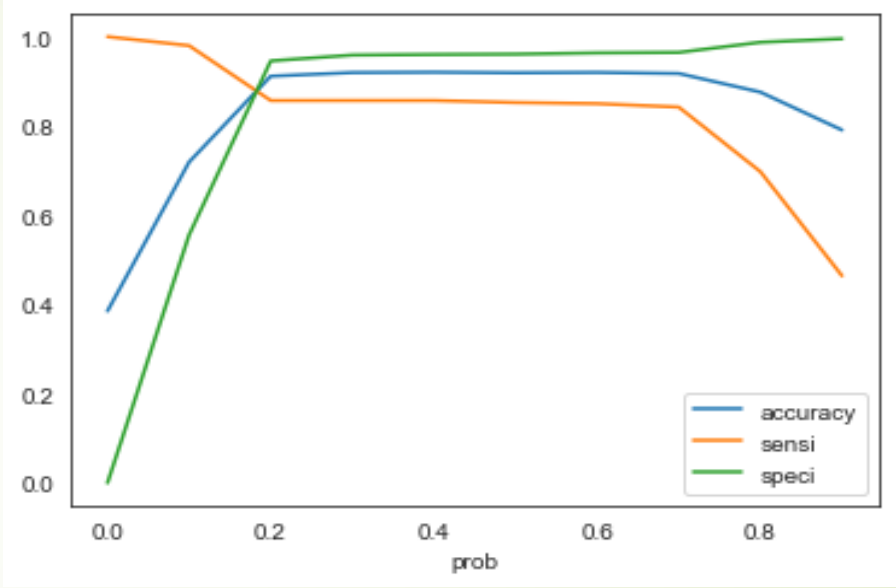


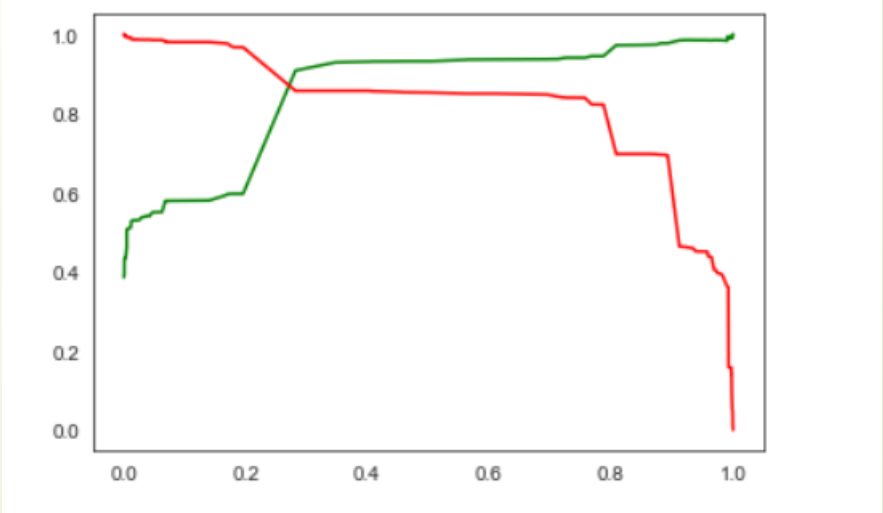


Model evaluation (train set)

- Sensitivity and specificity on train data set
- Accuracy – 91%
- Sensitivity – 85%
- Specificity- 96%
- Precision – 93%
- Recall - 85%
- False positive rate –.039
- Positive predictive value – 93%
- Negative predictive Rate – 91%
- Confusion matrix

3752	153
361	2085



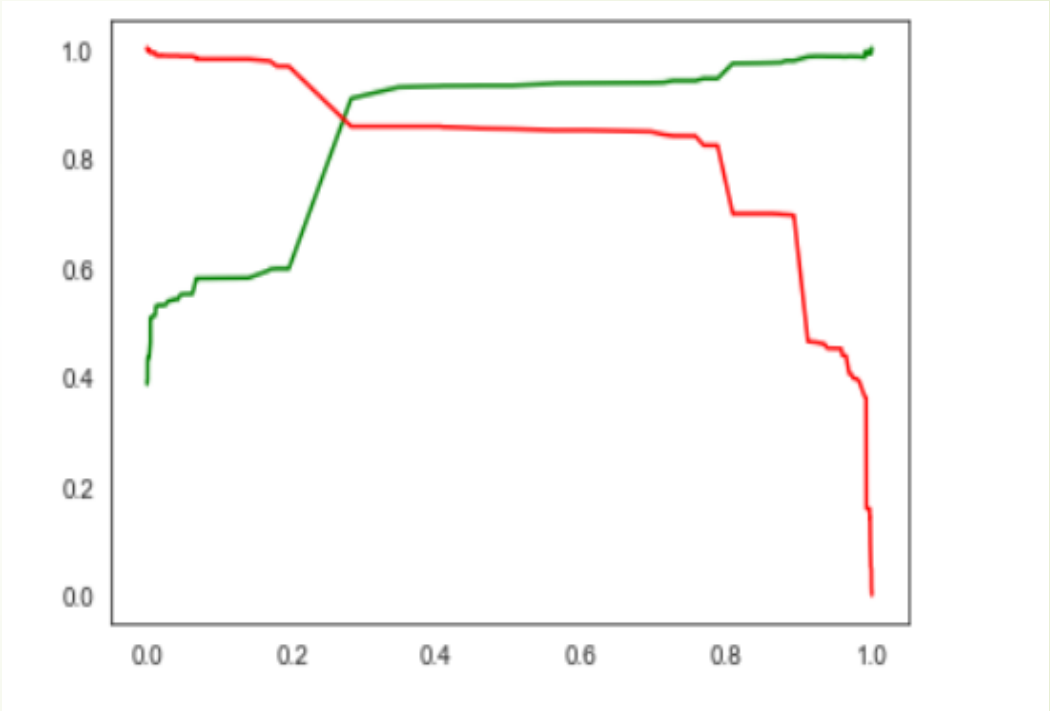


Model evaluation (test set)

- Accuracy – 90%
- Sensitivity– 85%
- Specificity- 94

- Confusion matrix

1642	92
157	832



Conclusion

- We have checked the sensitivity ,specialization, accuracy ,precision and recall
- Optimal cutoff was selected based on the sensitivity and specificity calculated
- The train and the test data set provides a overall same result
- From the lead score calculated we have conversion rate of 85% on both train and test set
- As a business aspect we need to improve the advertising in the digital market
- Overall this model seems to provide a good value