Lead Score Case Study

SUBMITTED BY:

CR LIKITH KUMAR

Objective

Determine the lead score and check if target final predictions amounts to 80% conversion rate

- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

Steps

Data Understanding

- Quality Data Check
- Data Cleaning
- Exploratory Data Analysis
- Data Preparation
- Model Building
- Model Evaluation
- Prediction

Conclusion

Data Cleaning Process

- Check for duplicate rows.
- The percentage of null values in each column.
- Remove columns with null values greater than 45%.
- Missing value Treatment for null values lesser than 45%.
- Outlier Treatment

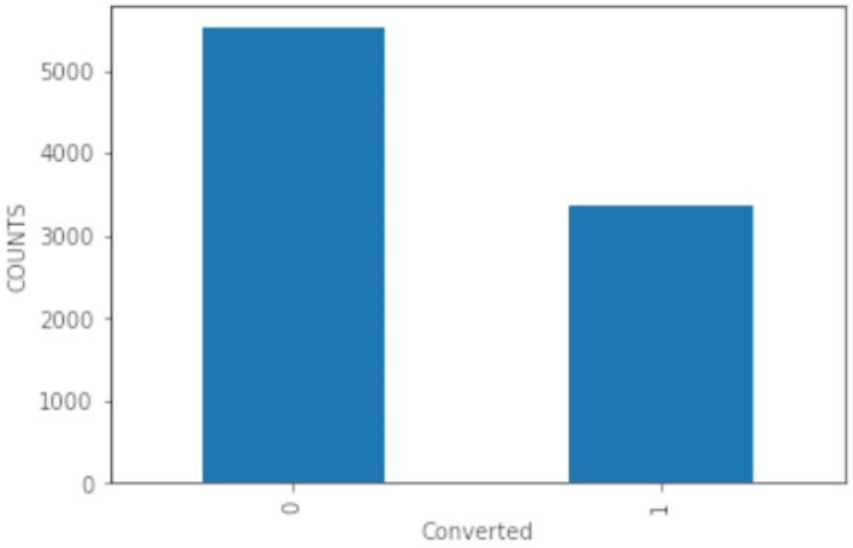
Exploratory Data Analysis

CONVERTED

Lead conversion rate is 38%

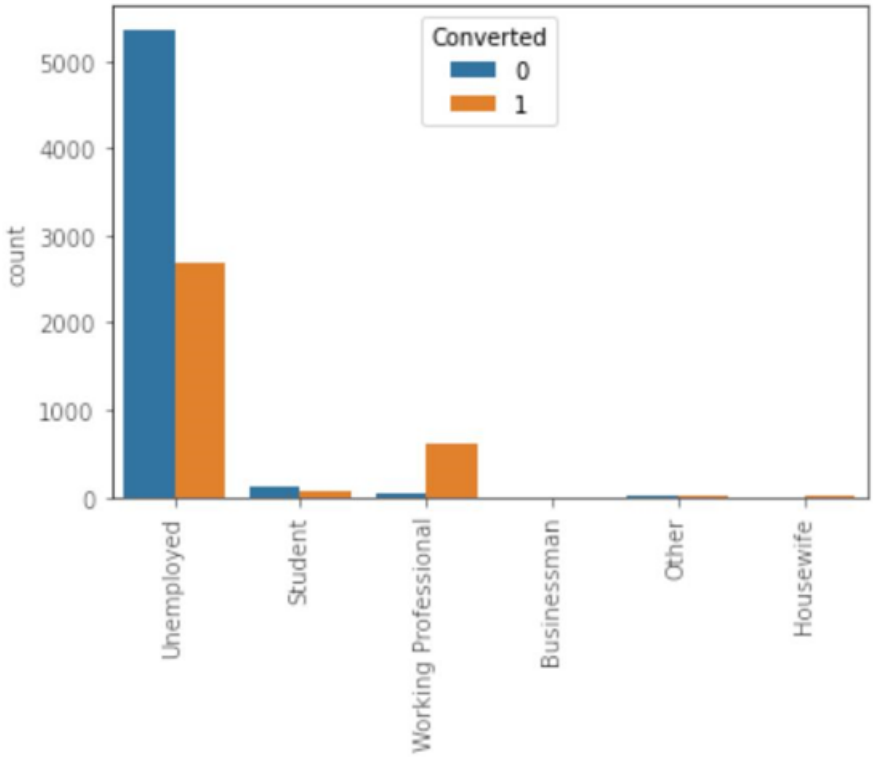
With 3346 rows indicating a conversion (1)

and 5487 rows indicating non-conversion (0)



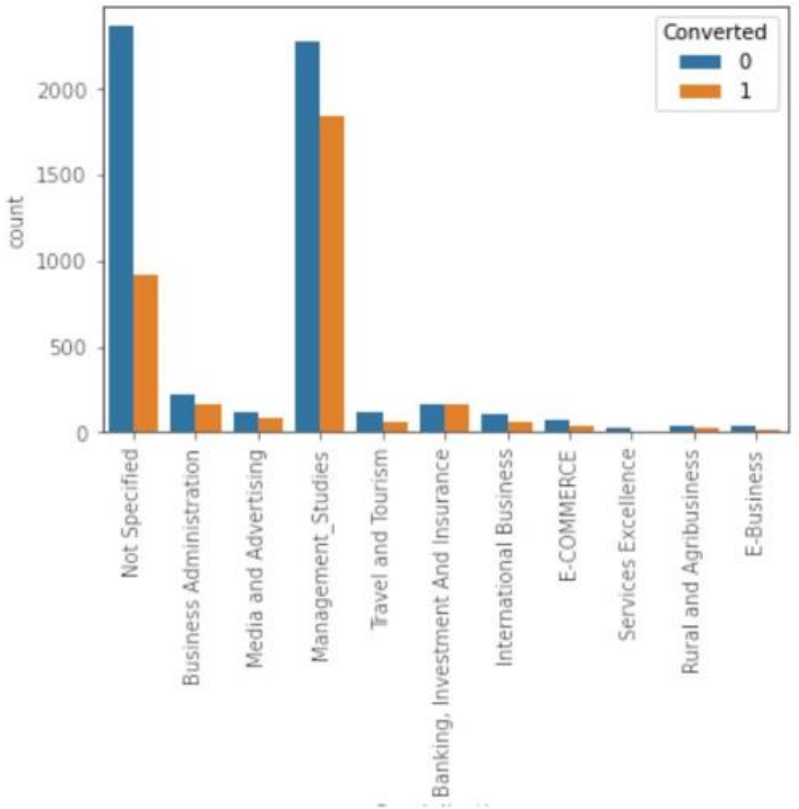
Current Occupation

Working Professionals have higher chances of getting converted or hot lead.
While, Unemployed have higher chances of not getting converted or cold

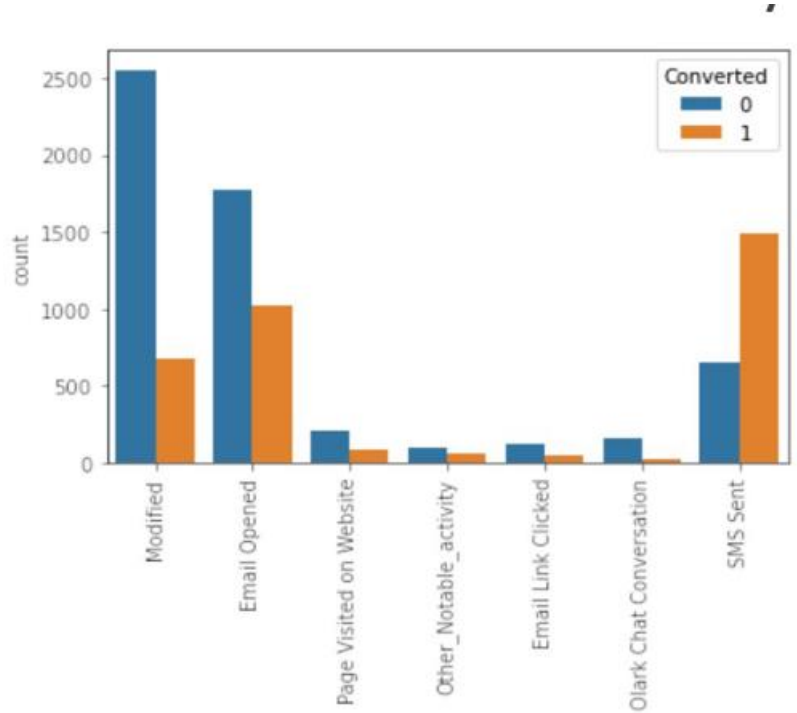


Specialization

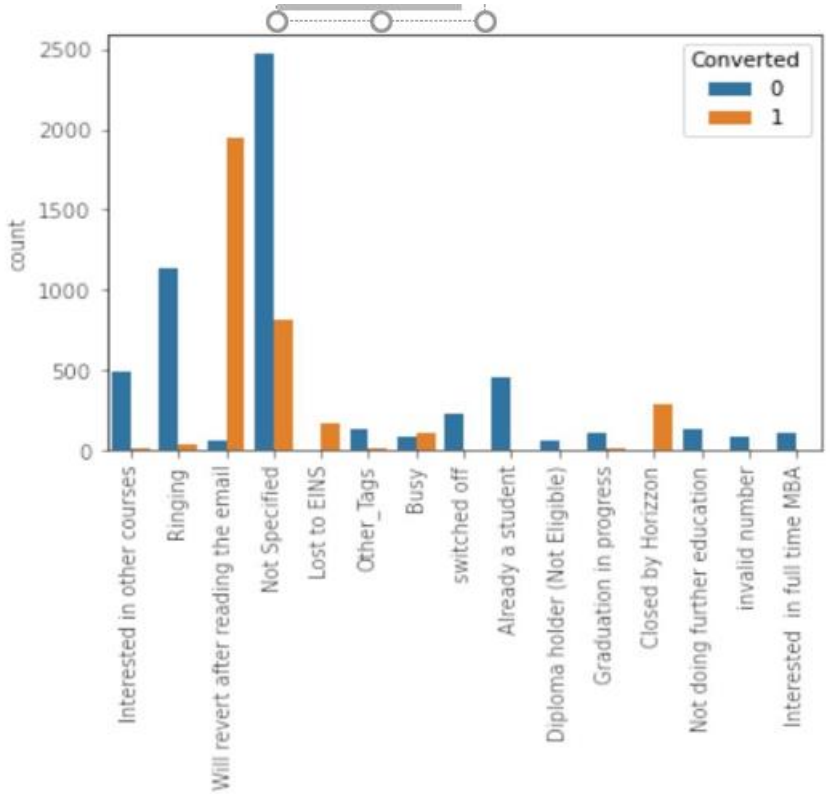
Majority of 'hot' leads have specialization related to Management



Last Notable Activity

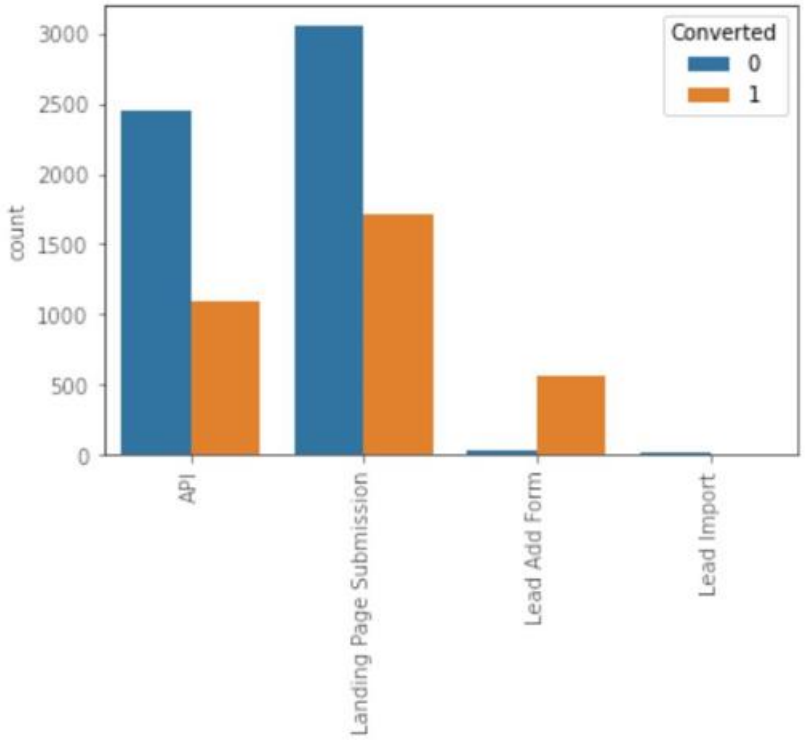


Tags



Lead Origin

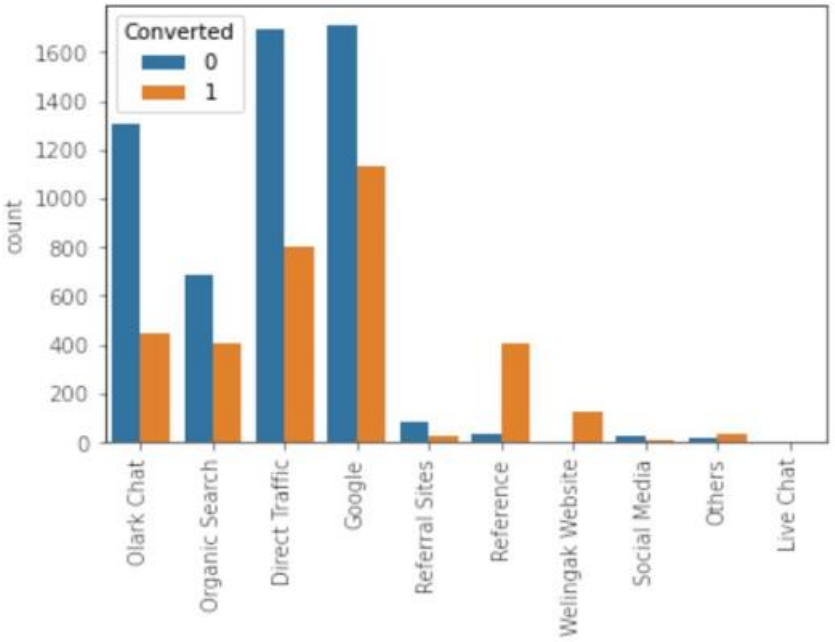
Lead add form showed a higher conversion, while, API showed lower conversion



Leads Source

Maximum lead come from ‘Google’, ‘Reference’, 'Welingak Website' and others showed higher conversion rates.

Maximum cold leads from ‘Olark Chat’ , ‘Organic Search’, ‘Direct Traffic’, and ‘Google’



Data Preparation & Model Building

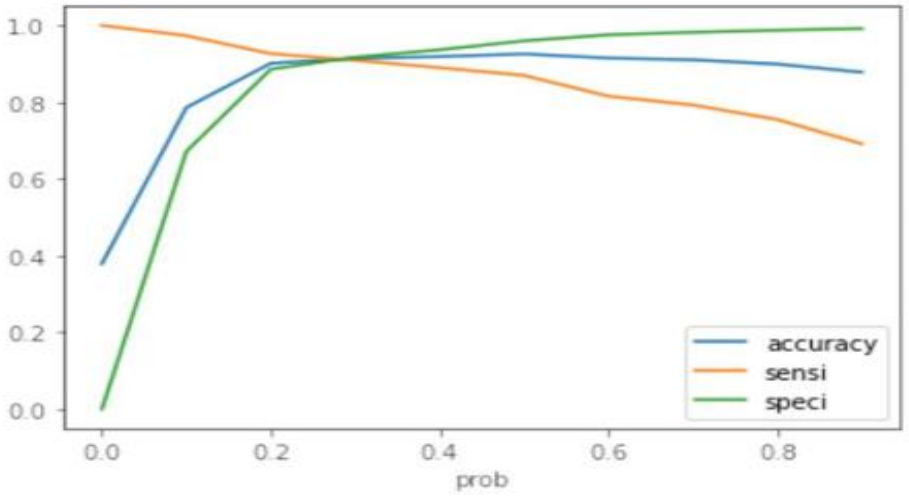
- Creating Dummy Variables for Categorical variables
- Train Test Split - 80% for train and 20% for test data
- Feature Scaling
- Feature Selection using RFE
- Model Selection
- Check for Multicollinearity

Important Variable through RFE

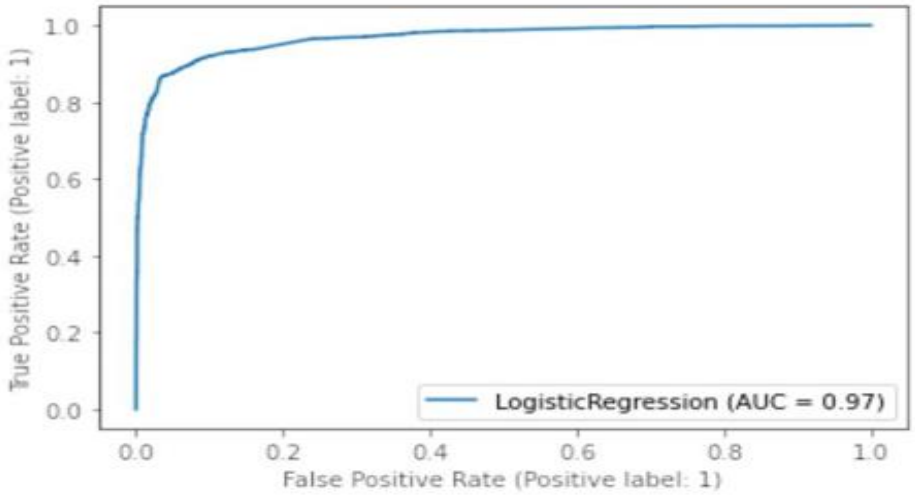
1. Total Time Spent on Website
2. Page Views Per Visit
3. Last Activity_SMS Sent
4. Last Notable Activity_Modified
5. Lead Source_Welingak Website
6. Tags_Already a student
7. Tags_Closed by Horizzon
8. Tags_Diploma holder (Not Eligible)
9. Tags_Interested in other courses
10. Tags_Lost to EINS
11. Tags_Not doing further education
12. Tags_Ringing
13. Tags_Will revert after reading the email
14. Tags_invalid number
15. Tags_switched off

Model Evaluation

Plot for Best Cut-off



ROC Curve



Prediction

- Cut-off at 0.3

Metrics	Train Data	Test Data
Accuracy	0.9141	0.9241
Sensitivity	0.9011	0.8621
Specificity	0.9170	0.9621

➢ These results indicate that the model can correctly identify hot leads
which should have a positive impact on the business by helping the
company to prioritize their efforts on the most promising leads and
ultimately increase their chances of closing a sale.

Conclusion

The Top 5 variables that mattered the most in the potential:

- The total time spent on the Website.
- Total number of visits.
- Lead source: Google, Direct traffic, Organic search, Welingak website.
- Lead origin: Lead add format.
- Current occupation: Working professional.

THANK YOU