Lead Scoring Case Study Assignment

By:

Preran ST

Steps

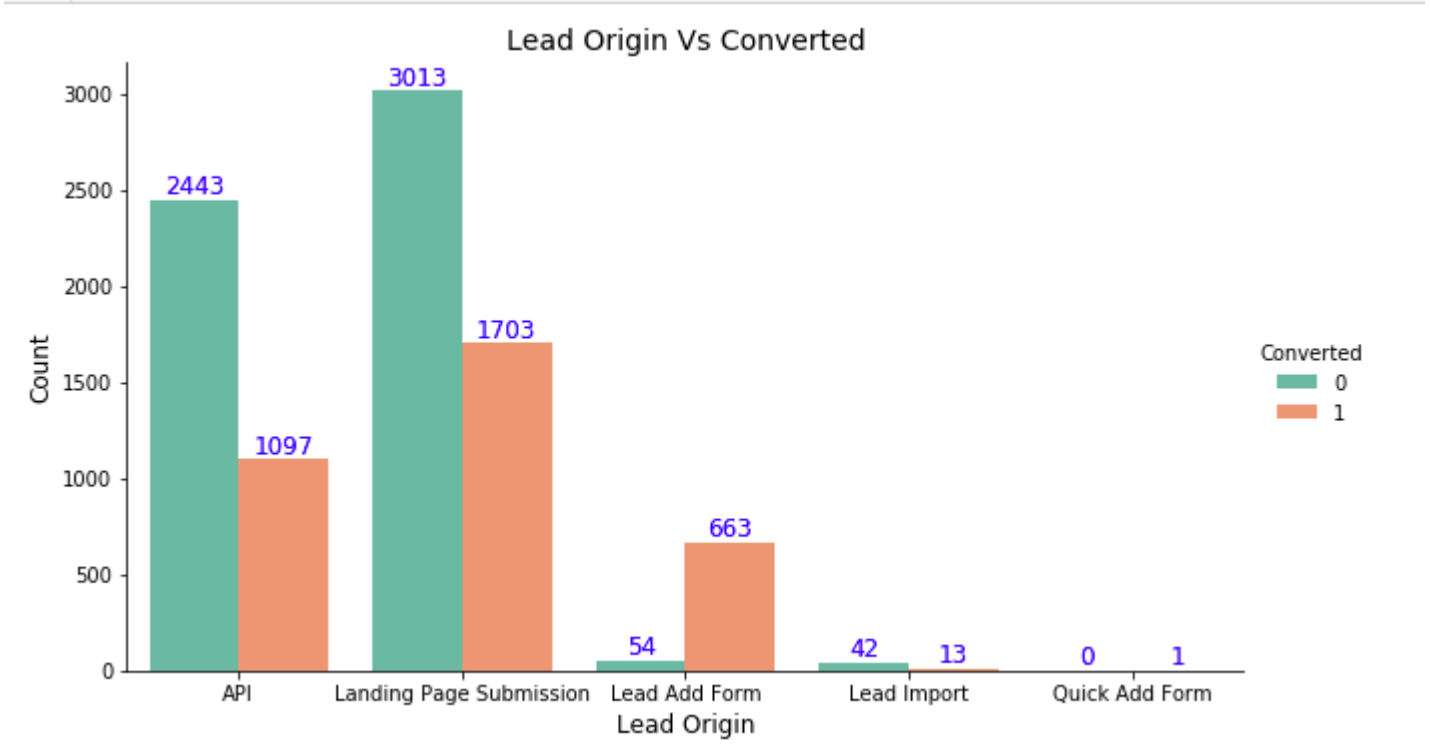
- Problem statement
- Problem approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

Problem Statement

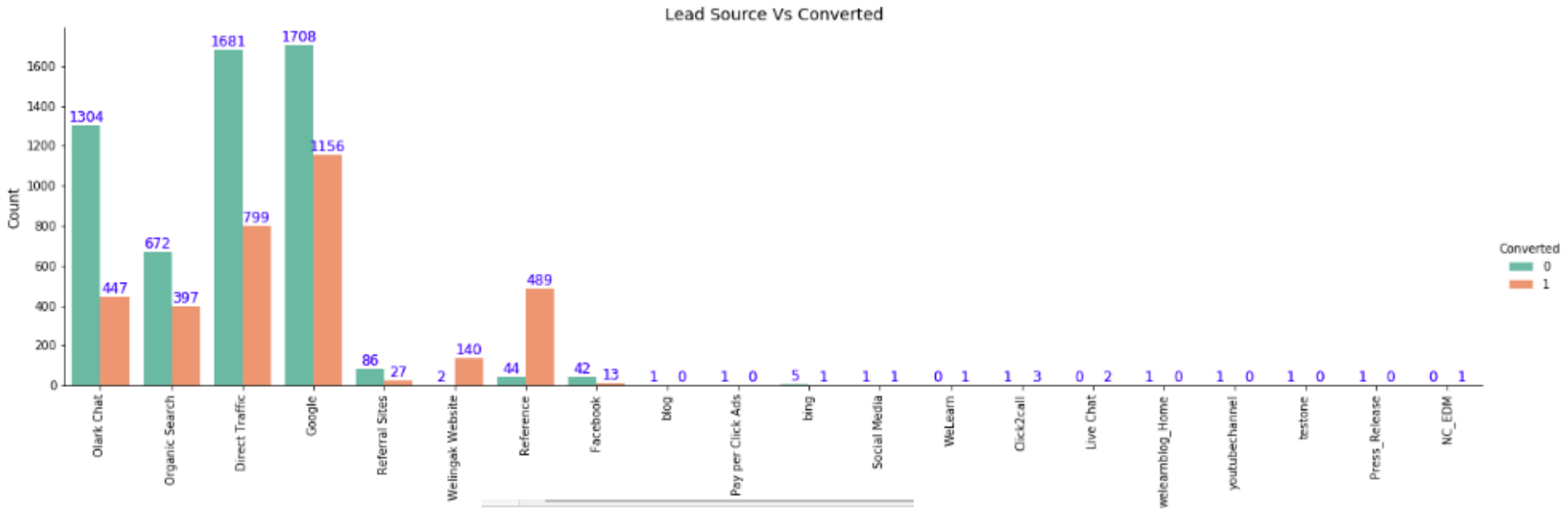
- To help X education to select the most promising leads known as ‘hot leads’ who are most likely to convert into paid customers.
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads where the leads with higher lead score have a higher conversion chance and the leads with lower lead score have a lower conversion chance.
- Identify the driver variables and understand their significance which are strong indicators of lead conversion.
- Identify the outliers, if any, in the dataset and justify the same.
- Consider both technical and business aspects while building the model.
- Summarize the conversion predictions by using evaluation metrics like accuracy, sensitivity, specificity and precision.

Data Cleaning and Preparation, EDA

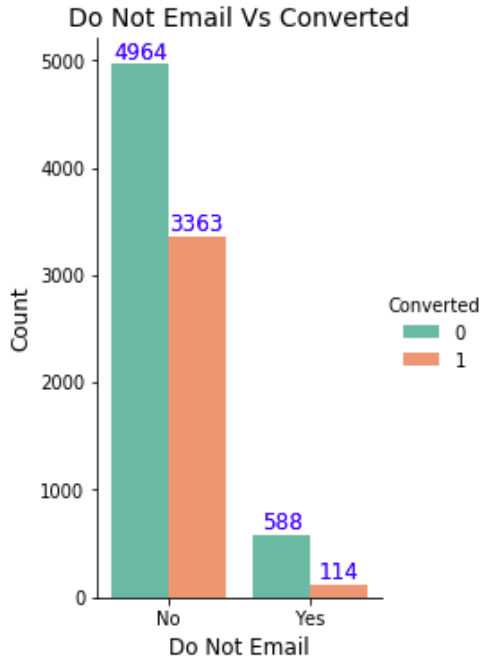
Lead Origin



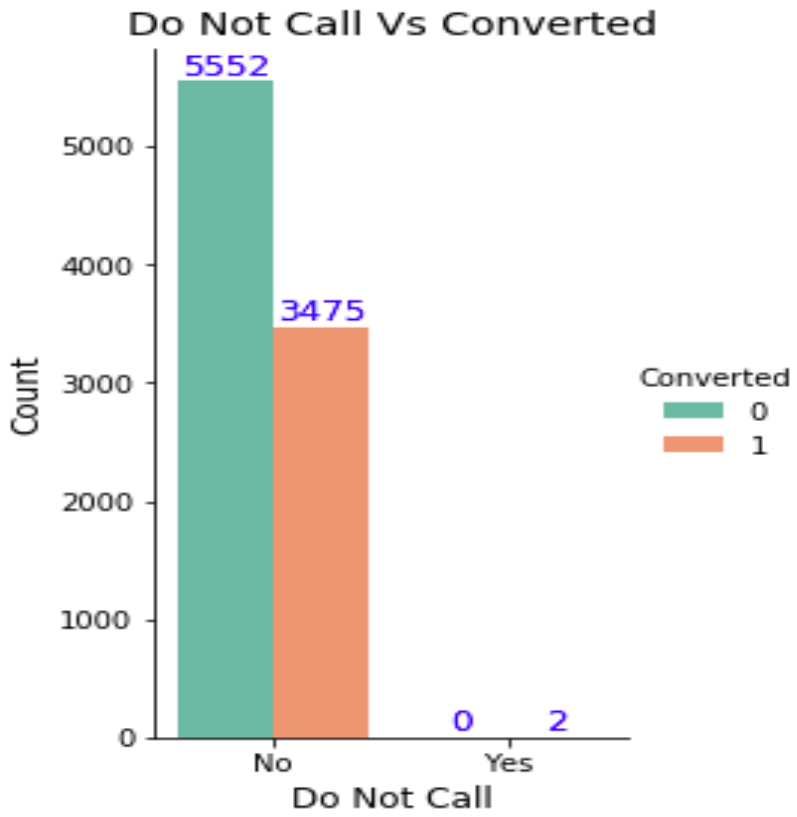
Lead Score



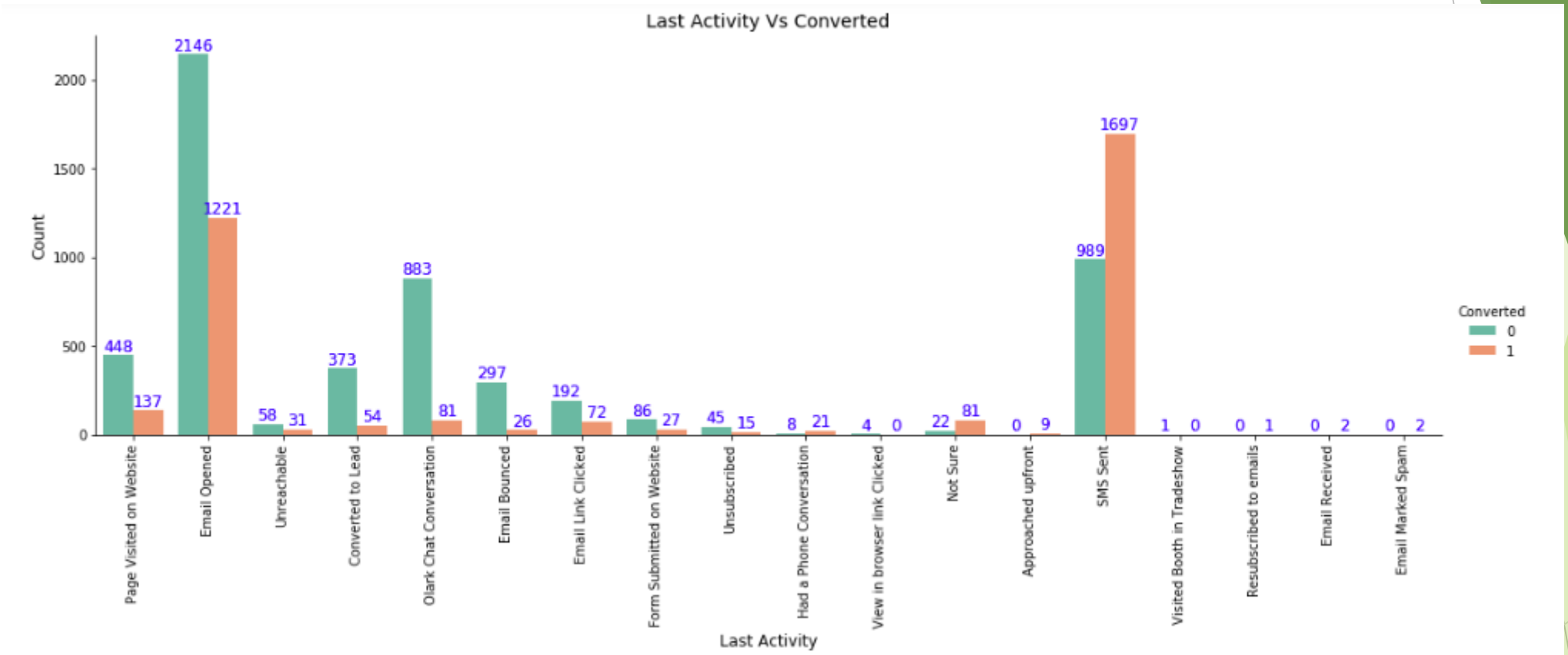
Do not email vs converted



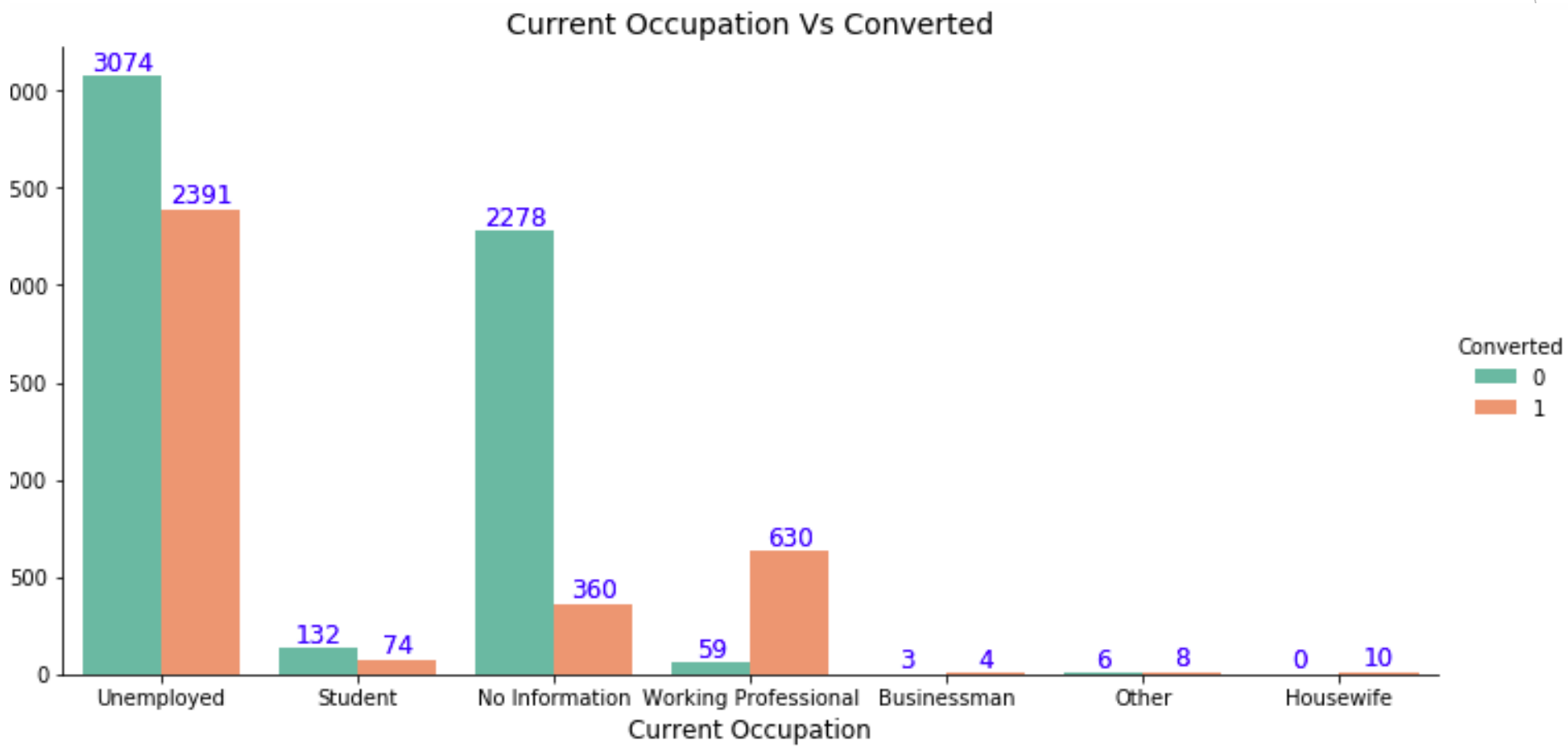
Do not call vs converted



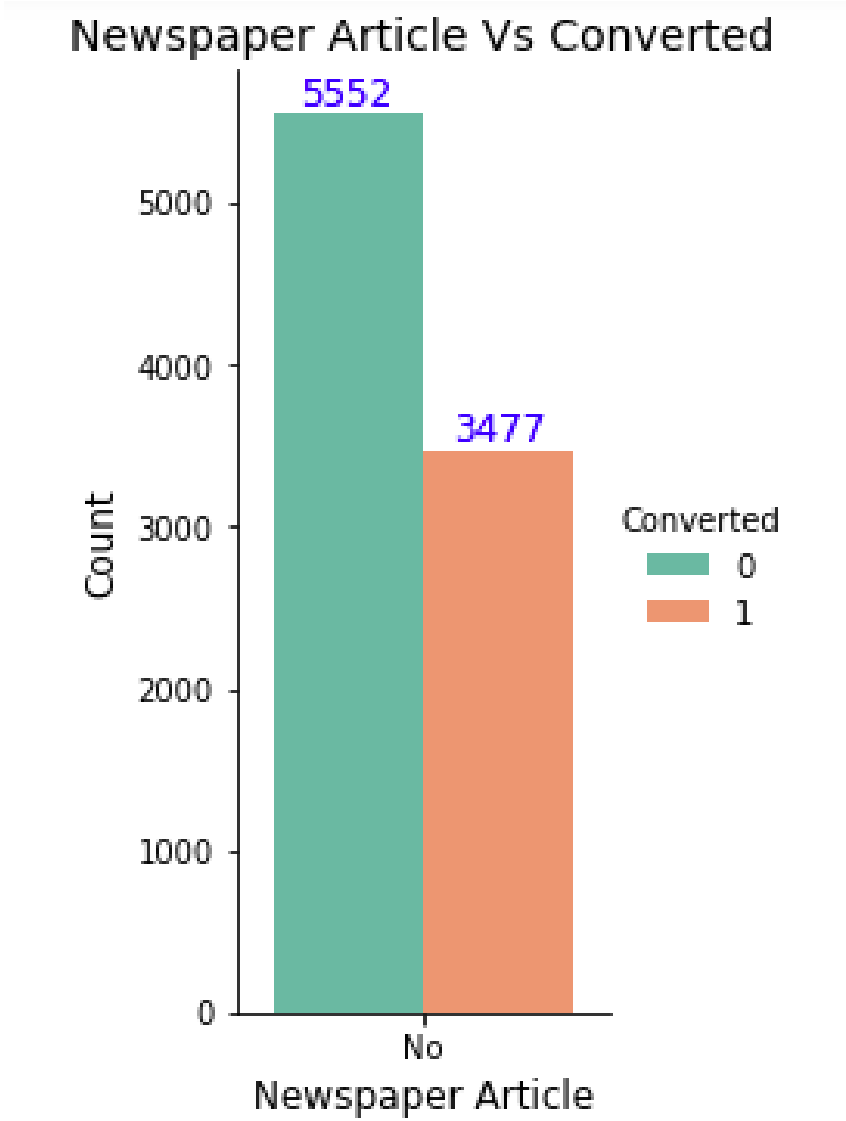
Last Activity



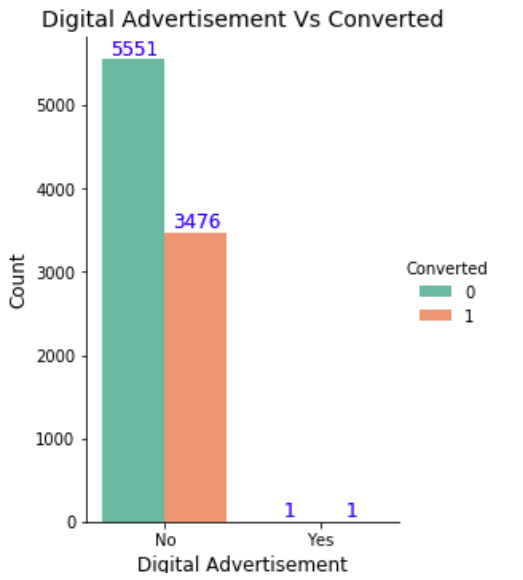
Occupation


Current Occupation Vs Converted
3074
2391
2278
132
74
360
59
630
3
4
6
8
0
10
Unemployed
Student
No Information
Working Professional
Businessman
Other
Housewife
Current Occupation
Converted
0
1

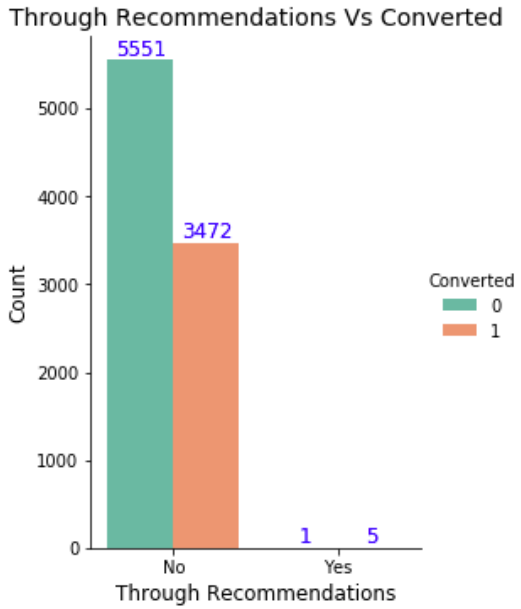
Newspaper Article



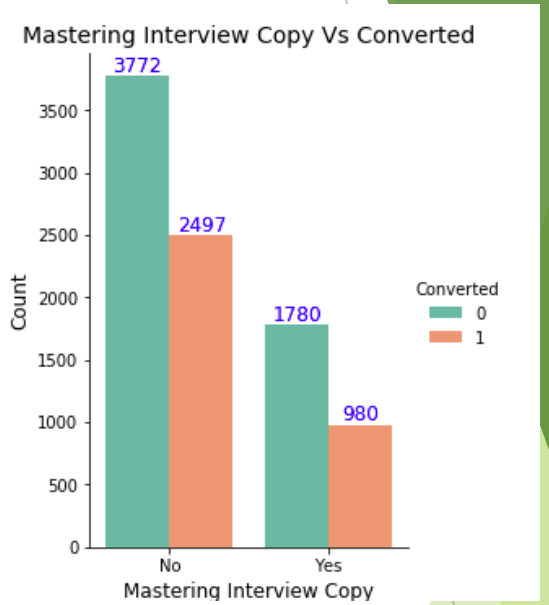
Digital Advertisement



Recommendations



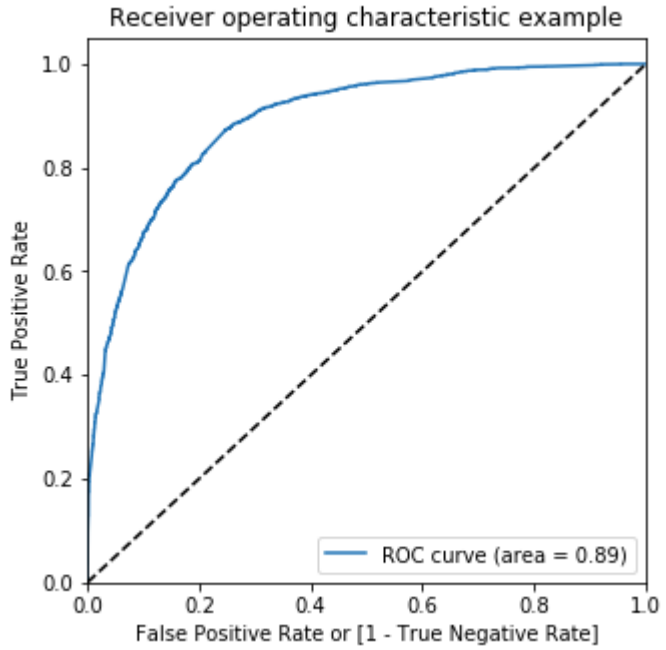
Interview copy

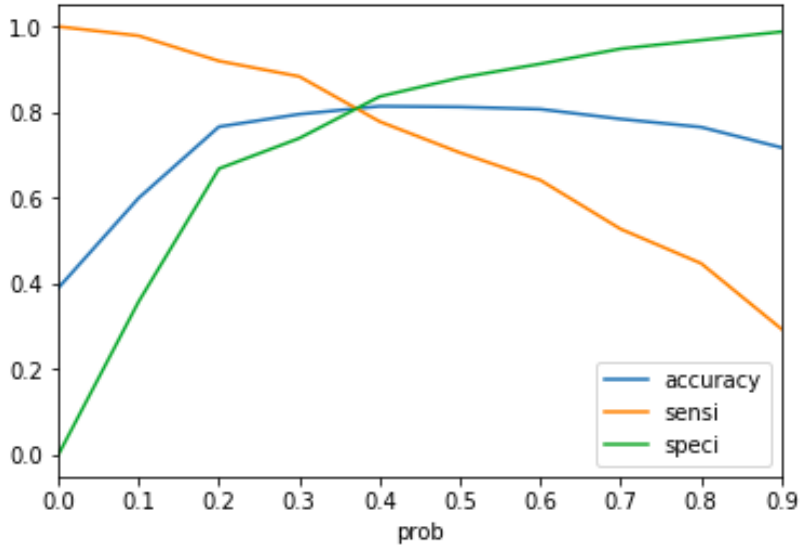


Model Evaluation

- ROC Curve

0.37 is the tradeoff between Precision and Recall - Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 37 % to be a hot Lead





Observations

Train Data

Accuracy : 81%

Sensitivity : 73%

Specificity : 79%

Test Data

Accuracy : 78%

Sensitivity : 80%

Specificity : 82%

Conclusion

- We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore we can intervene that we need to focus more on the leads originated from API and Landing page submission.
- We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.