

Lead Score Assignment

By:-

Sidharth Kumar Singh

Shraddha Choubey

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 or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. 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%.

Business Goal

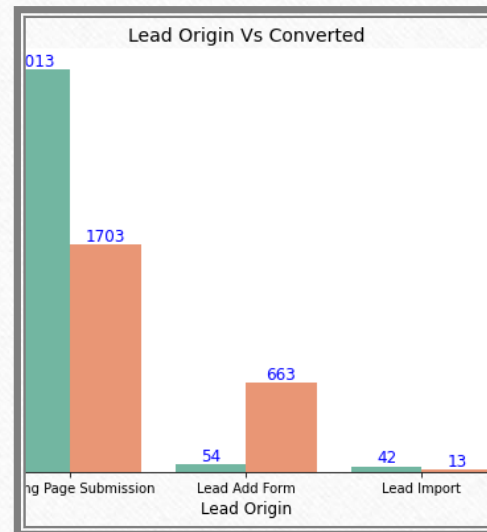
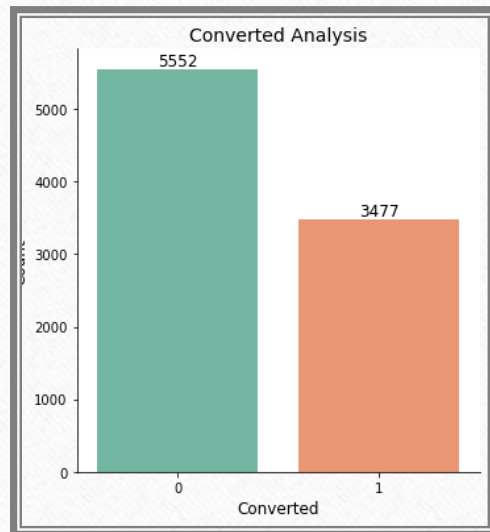
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the 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%.

Solving Methodology

- Reading and Understanding Data
- Data Cleaning and Outliers Treatment
- Data Analysis
- Create Dummy Variables
- Test Train Split
- Feature Rescaling – MinMaxScaler
- Model Building using Feature selection using RFE
- Plotting the ROC Curve
- Finding the Optimal Cutoff Point
- Computing the Precision and Recall metrics
- Making Predictions on Test Set

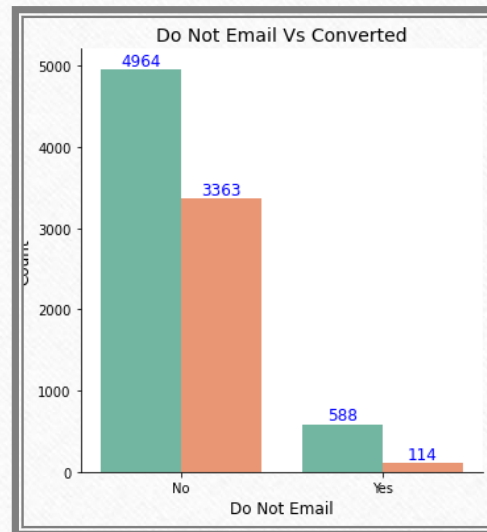
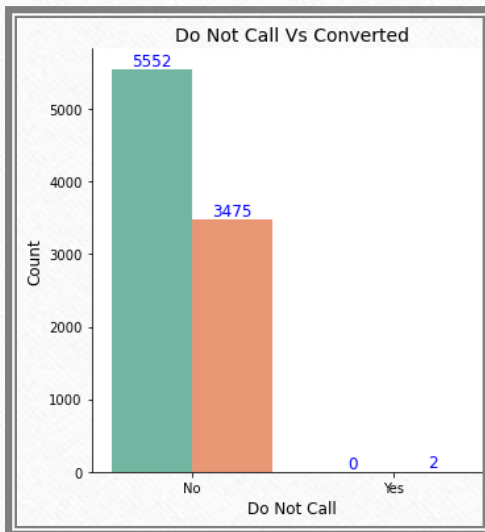
Exploratory Data Analysis

- Graph 1: Shows the total number of Conversion
- Graph 2: Lead Origin Vs Converted



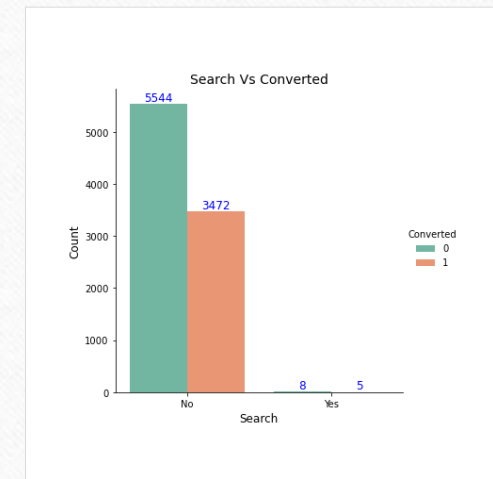
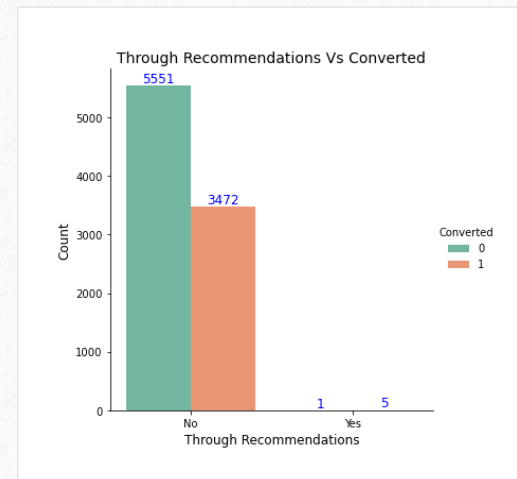
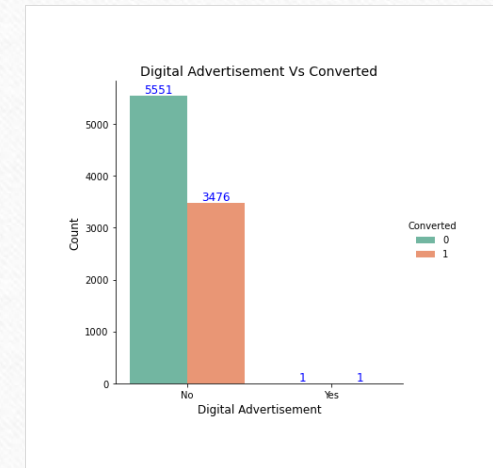
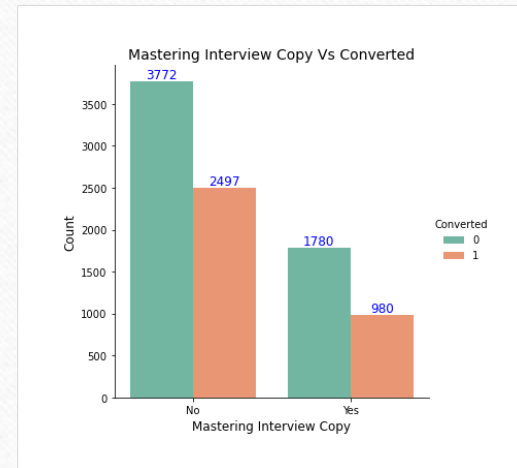
Exploratory Data Analysis

- Graph 1: Do Not Call Vs Converted.
- Graph 2: Do Not Email Vs Converted.

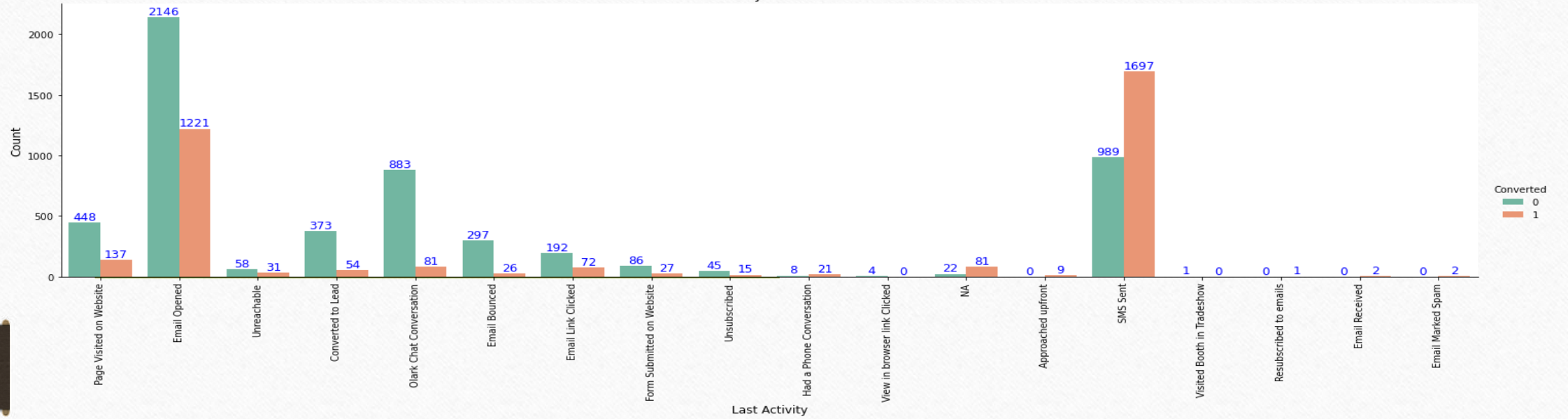


Exploratory Data Analysis

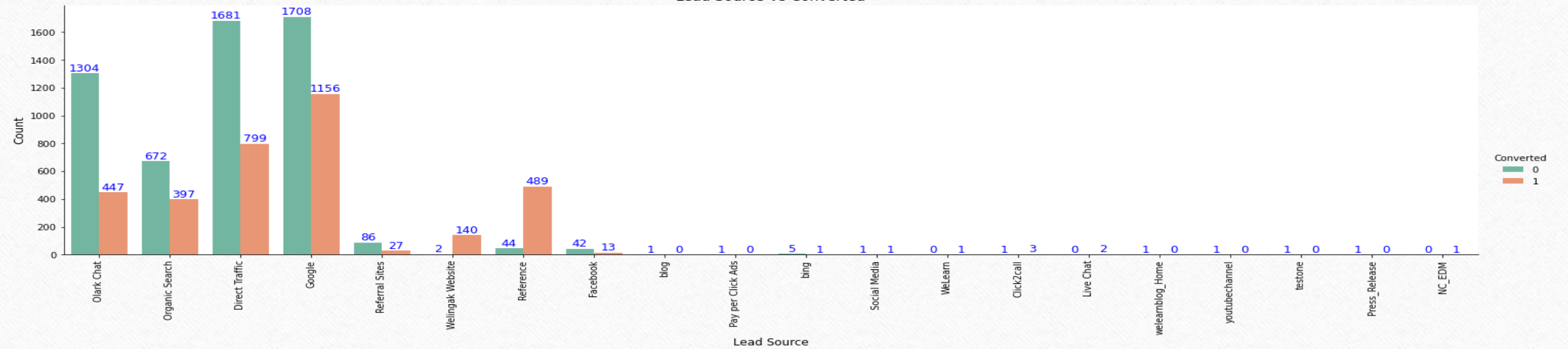
- Graph 1: Mastering Interview Vs Converted
- Graph 2: Digital Advertisement Vs Converted
- Graph 3: Through Recommendation Vs Converted
- Graph 4: Search Vs Converted



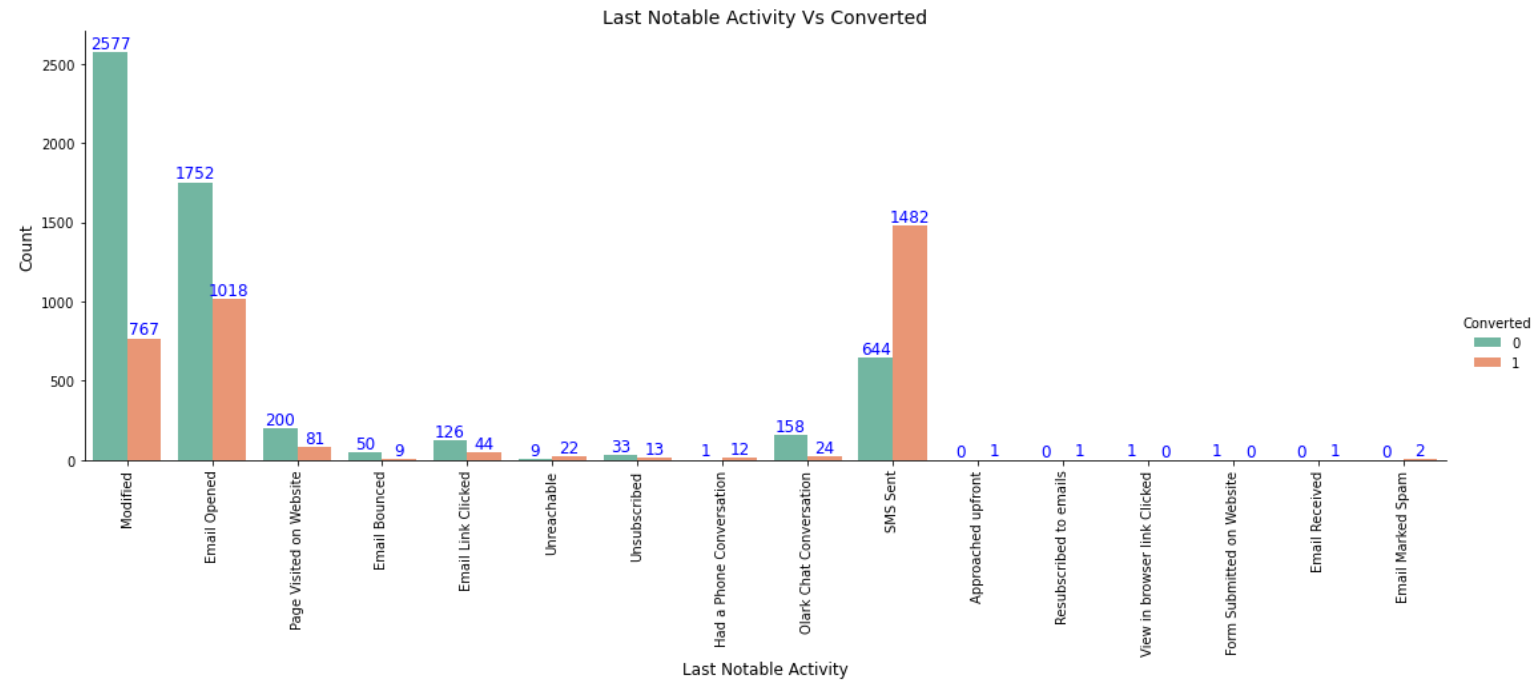
Last Activity Vs Converted



Last Activity
Lead Source Vs Converted

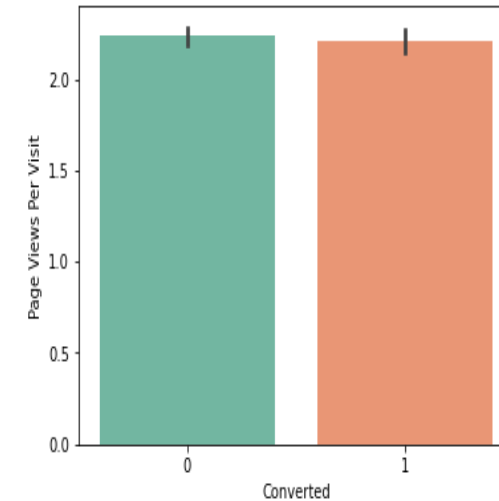
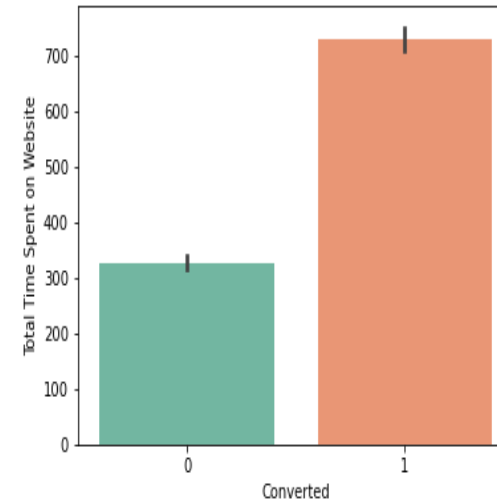
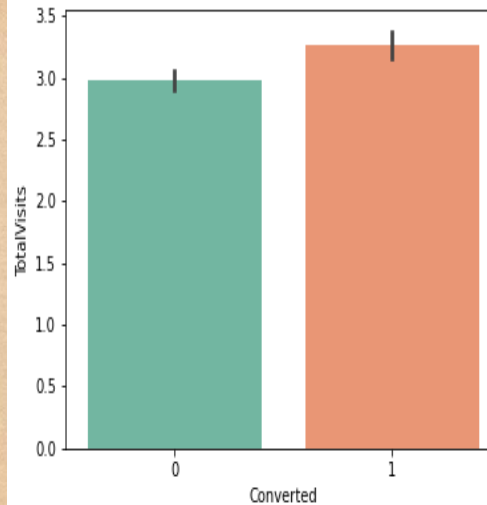


- This is Last Notable Activity vs Converted Graph.



Exploratory Data Analysis

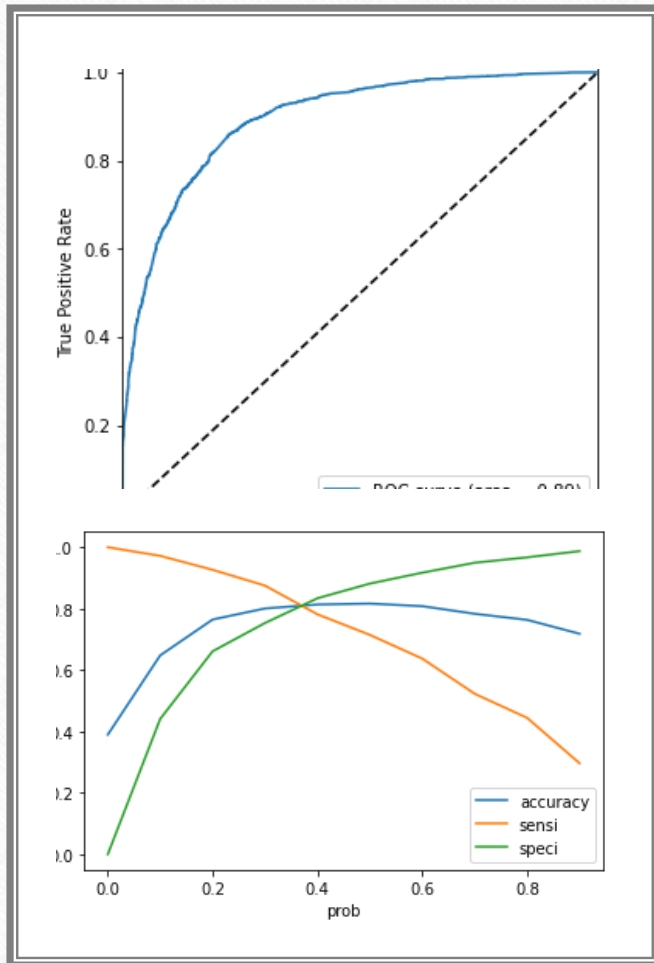
- Graph 1: Total Visits Vs Converted
- Graph 2: Total Time Spent on Website Vs Converted
- Graph 3: Page Views Per Visit Vs Converted.



Model Building

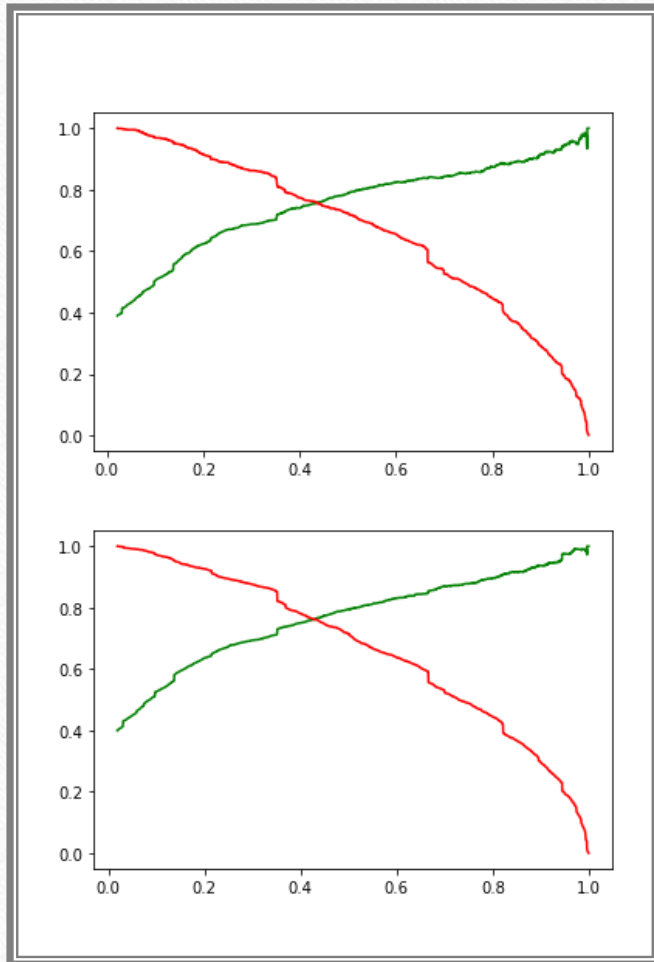
- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set
- Overall accuracy 81% and F1-score to be 81% too.

ROC Curve



- Finding Optimal Cut off Point
- Optimal cut off probability is that probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.37.

Precision-Recall Curve



- Graph 1: It show the optimal point at which Precision-Recall Intersect for training data.
- Graph 2: It show the optimal point at which Precision-Recall Intersect for test data.

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal and cut off based on Sensitivity and Specificity for calculating the final prediction.
- Also, the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%
- Accuracy, Precision and Recall values of test set are around 81%, 82% and 81% which are approximately closer to the respective values calculated using trained set.
- Hence overall this model seems to be good.