



LEAD SCORING CASE STUDY

VARUNI RAVINDRA

Contents

- ▶ Problem Statement
- ▶ Goal of the Case Study
- ▶ Approach to solve the problem
- ▶ Data Visualization and Observations
- ▶ Prediction
- ▶ Model Evaluation
- ▶ Conclusion

Problem Statement and Goal of the Case Study

- ▶ An education company named X Education sells online courses to industry professionals.
- ▶ Many professionals who are interested in the courses land on their website and browse for courses.
- ▶ A person who provides details like an email id or phone number becomes a lead.
- ▶ The sales team starts pursuing the lead hoping that they enroll in one of the programs by X Education.
- ▶ X Education has very poor conversion of leads which is around 30%.
- ▶ The company wants an efficient process to identify potential leads who will most likely enroll so that the sales team can focus on those leads.

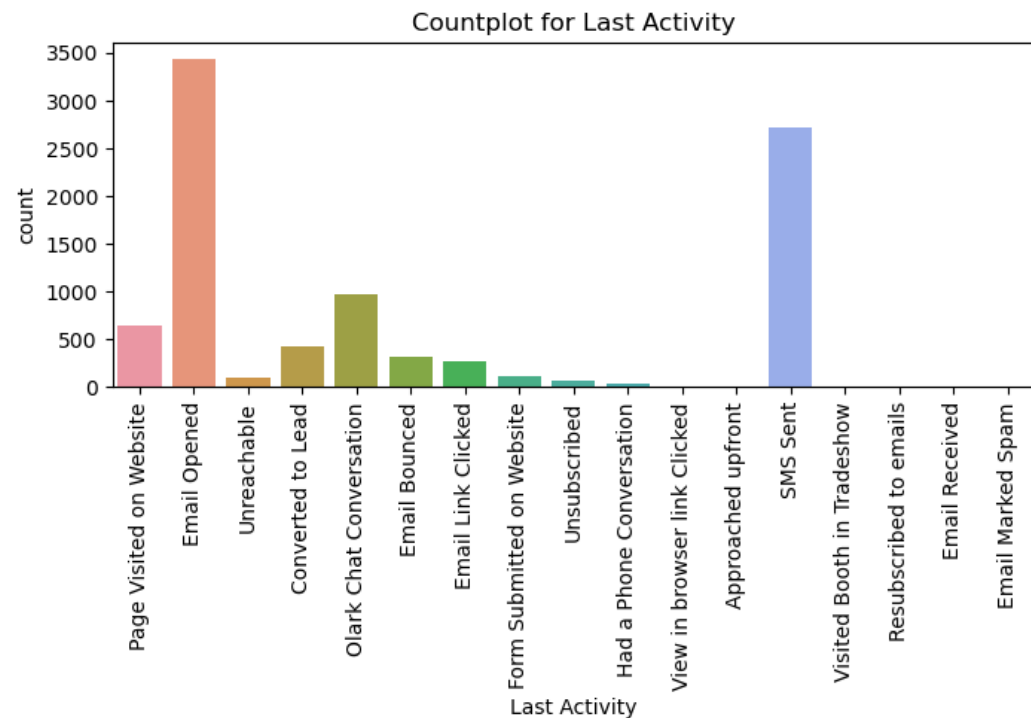
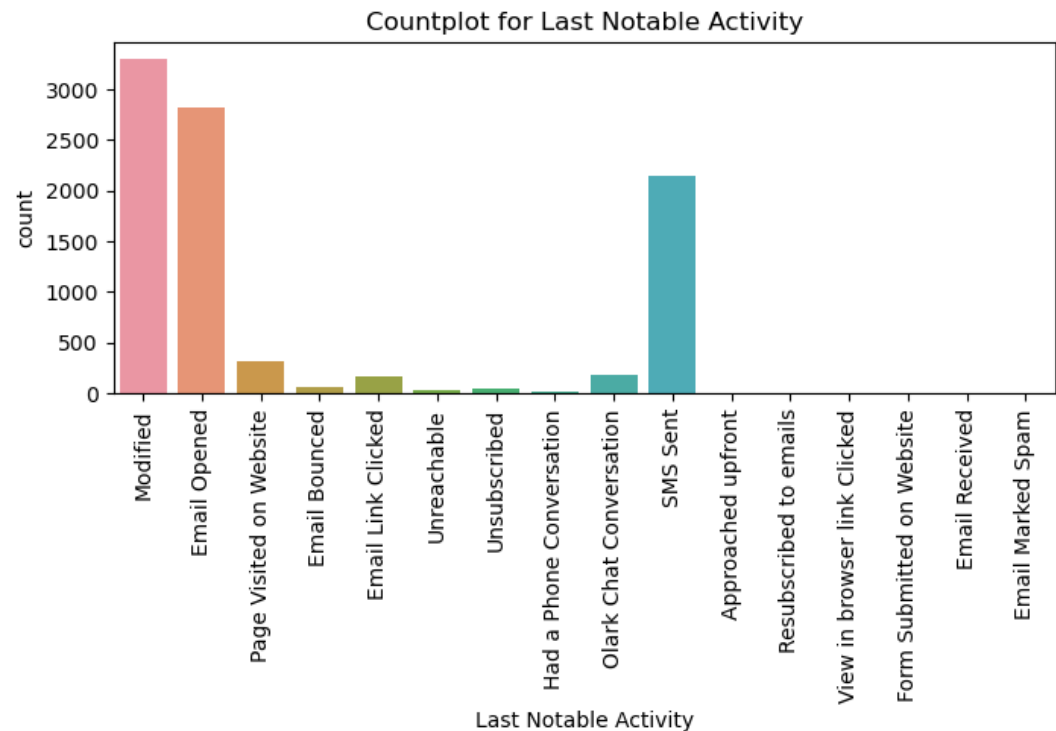
GOAL

- ▶ To create an efficient process where a potential lead can be identified who will most likely enroll in one of X Education's courses.
- ▶ Sales team can focus and pursue on the hot leads.
- ▶ Target lead conversion rate to be around 80%.

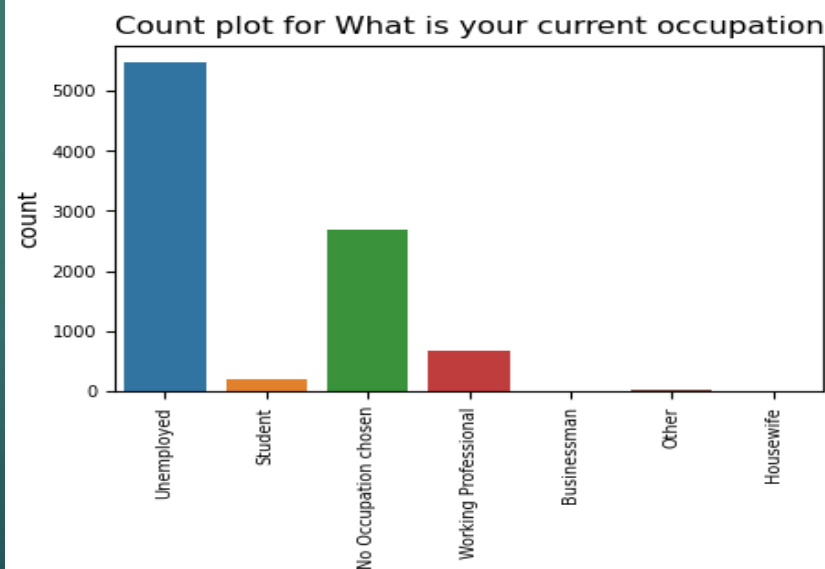
Approach to solve the problem

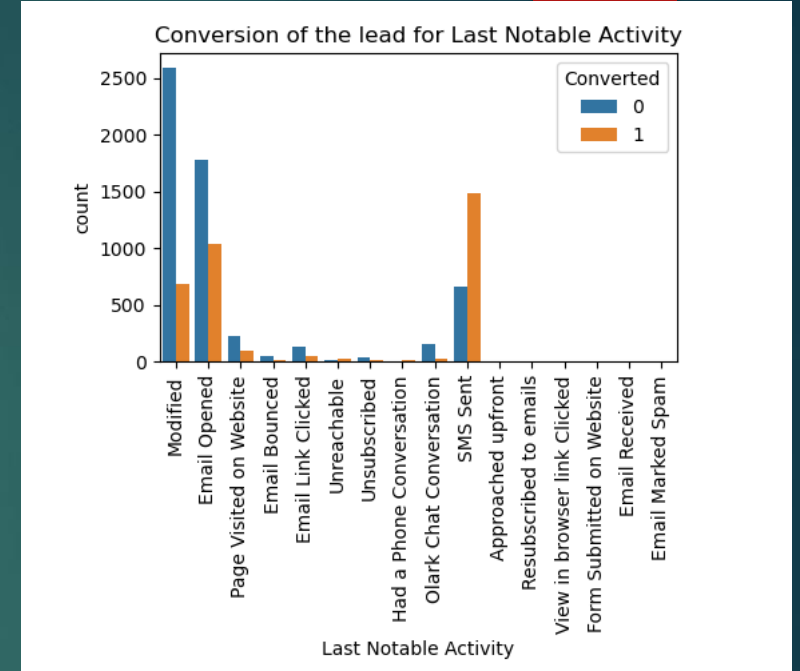
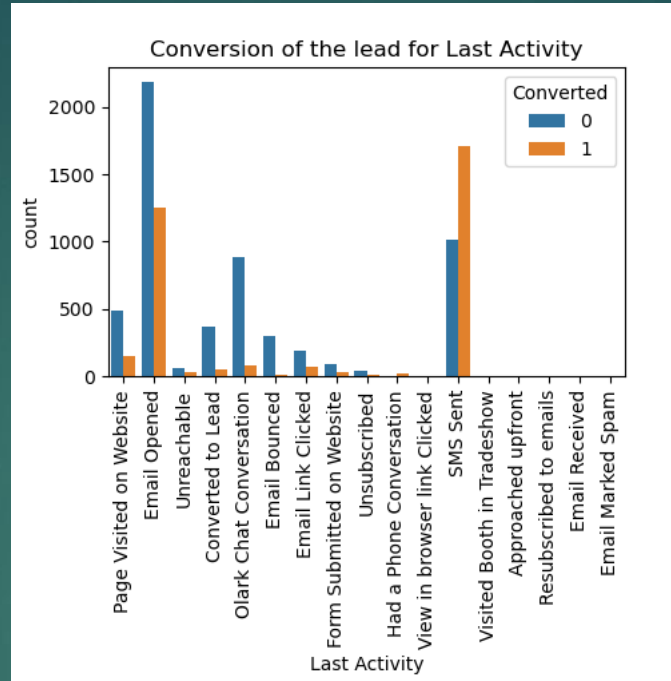
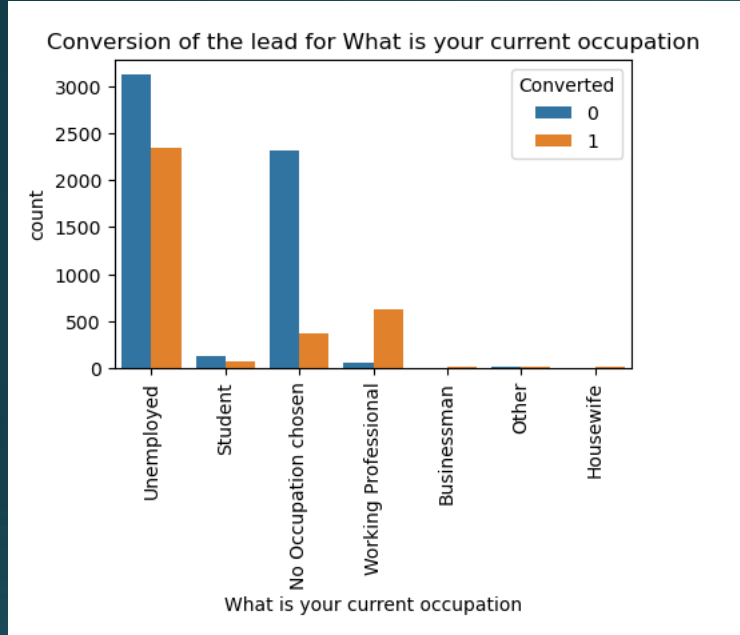
- ▶ Data Reading and Understanding
- ▶ Data Clean up
- ▶ Segmentation of numerical and categorical variables
- ▶ EDA and Data Visualization
- ▶ Creation of Dummy Variables
- ▶ Splitting of target variables and feature variables
- ▶ Train test split
- ▶ Scaling of numerical data
- ▶ Model building
- ▶ Prediction and calculation of various metrics on the train data
- ▶ ROC Curve and finding optimal cut off point
- ▶ Model evaluation using the test data

DATA VISUALIZATION

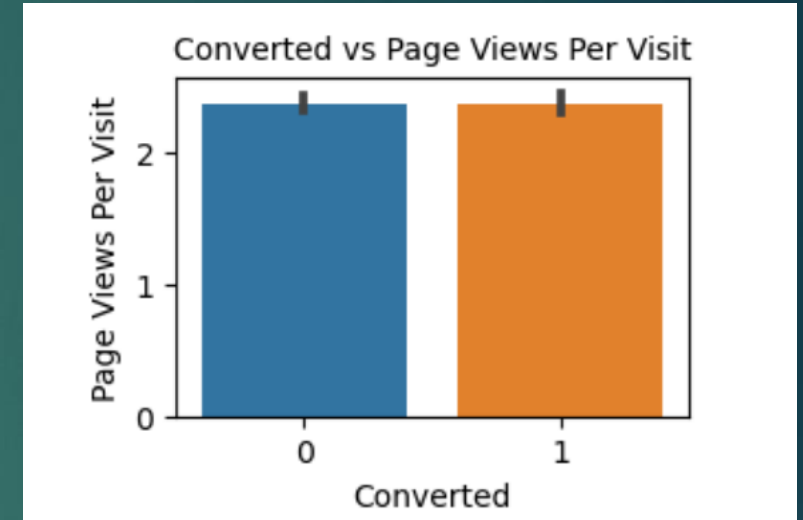
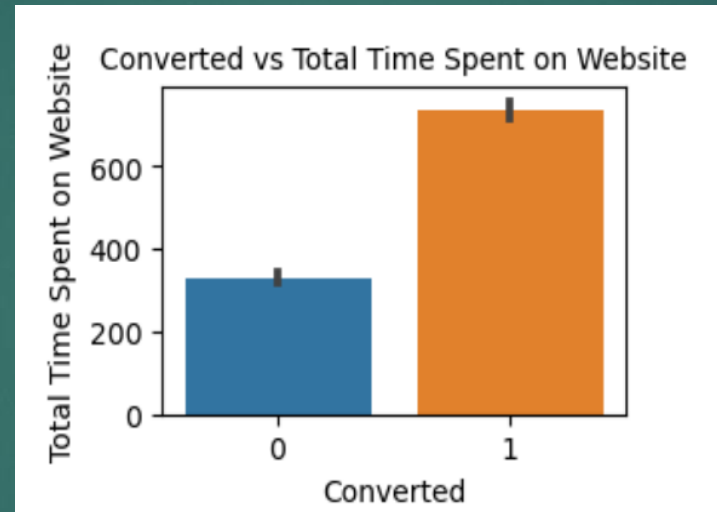
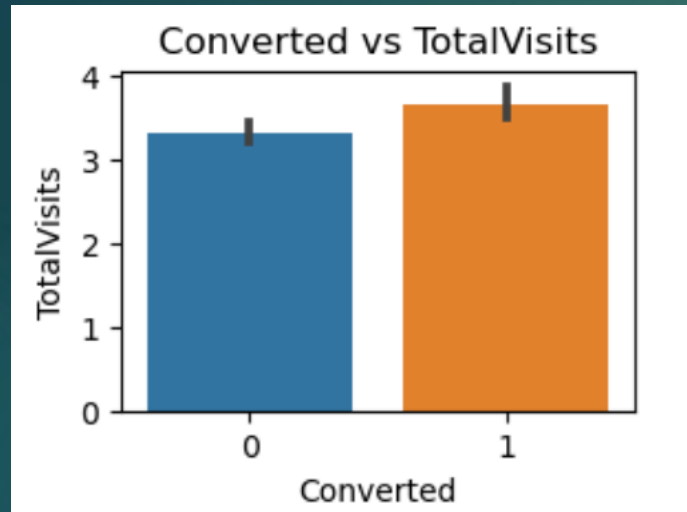


Univariate Analysis of Categorical columns chosen
by the Final Model
Count of different options in:
What is your current occupation
Last Activity
Last Notable Activity





Count of Converted and non converted users for the options in the columns:
What is your current occupation
Last activity
Last notable activity

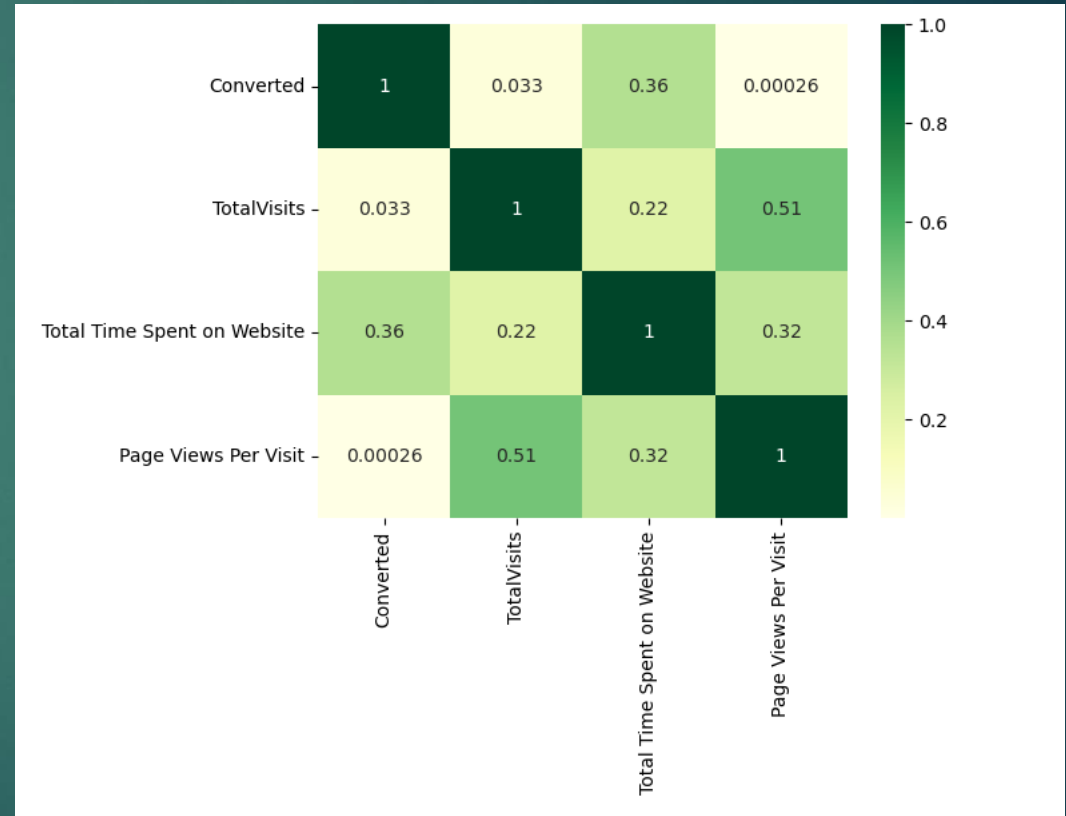


Analysis of numerical variables

1. Comparison of converted and non converted in the numerical columns

Correlation Matrix

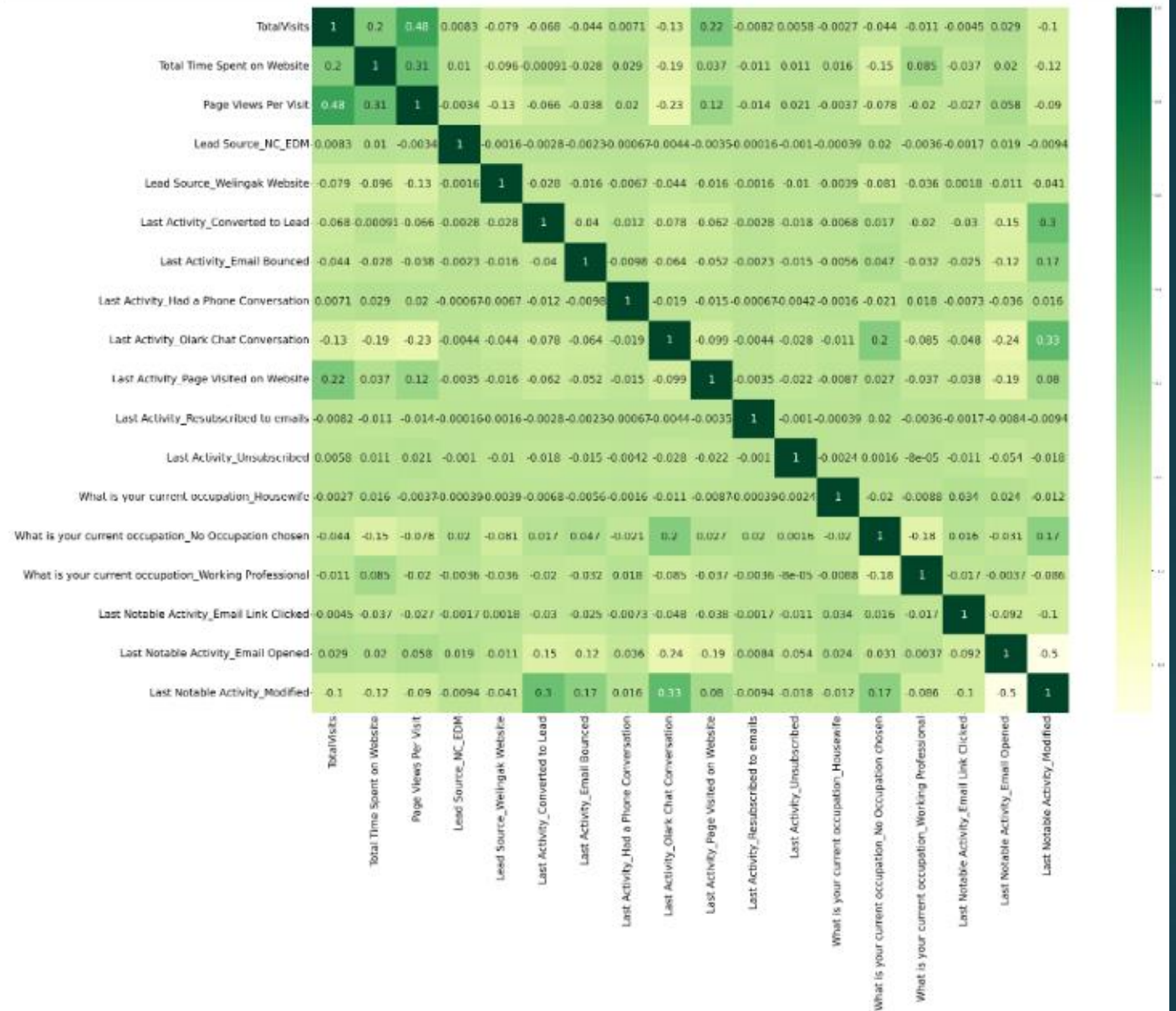
- ▶ The correlation matrix for all the numerical variables in the data set.
- ▶ There is no high correlation between any of the numerical variables.



Correlation Matrix

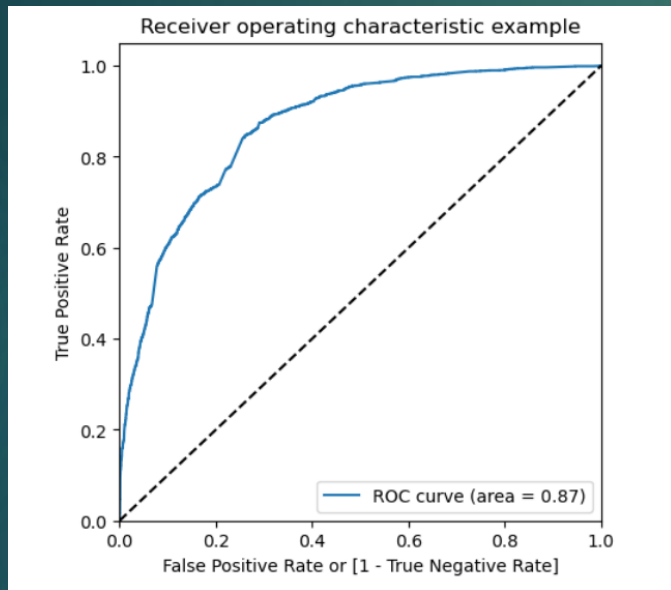
Correlation matrix
after RFE

Observation: No
variables are highly
correlated

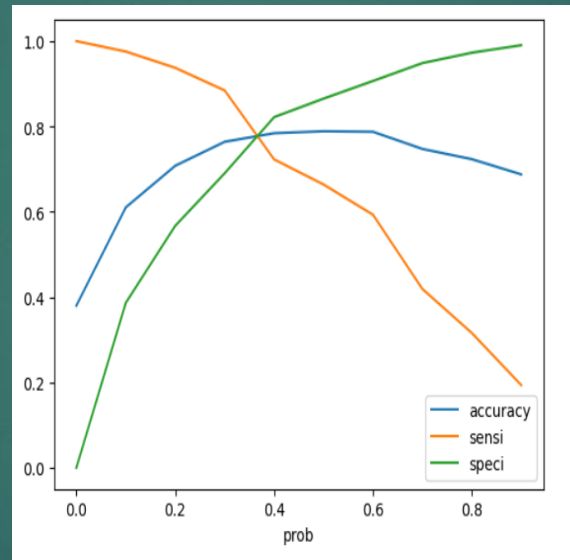


ROC, Optimal cutoff point and Precision and recall trade off

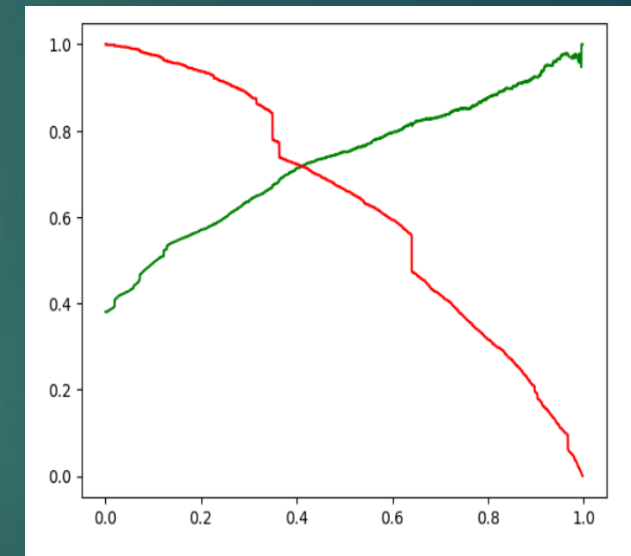
- ▶ ROC curve – area is 0.87 which is pretty good.



- ▶ The optimal cut off point is 0.34. The point where all the three meet.



Precision and Recall trade off



We can consider the probability greater than 0.34 for a lead as a hot lead to get a better conversion rate

Model Evaluation

Test data

- ▶ Accuracy of Test data: 77.85%
- ▶ Sensitivity of Test data: 83.75%
- ▶ Specificity of Test data: 74.30%

Train data

- ▶ Accuracy of Train data: 77.81%
- ▶ Sensitivity of Train data: 84.66%
- ▶ Specificity of Train data: 73.61%

Final set of features chosen by the model

- TotalVisits
- Total Time Spent on Website
- Page Views Per Visit
- Last Activity_Converted to Lead
- Last Activity_Email Bounced
- Last Activity_Olark Chat Conversation
- Last Activity_Page Visited on Website
- Last Activity_Unsubscribed
- What is your current occupation_No Occupation chosen
- What is your current occupation_Working Professional
- Last Notable Activity_Email Link Clicked
- Last Notable Activity_Email Opened
- Last Notable Activity_Modified

Conclusion

- ▶ A probability of 0.34 or more can be considered a hot lead to get a better conversion rate.
- ▶ Working Professionals are more likely to convert.
- ▶ Users who have spent more time on the website are likely to convert.
- ▶ Users with more number of visits to the website are more likely to convert.
- ▶ Users who have not mentioned their current occupation are not likely to convert.
- ▶ Users who have unsubscribed and whose emails are bouncing are not likely to convert.

Recommendation:

The sales team should focus on working professionals who visit the website often and spend more time on the website.

THANK YOU!