

Lead Scoring Case Study

Submitted By:

Akash Kalita

Jyothi Punya Prasad Bale

Yogeswaaran Panneerselvam

Business Objective

- To help X Education to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- To build a logistic regression model to assign a lead score value between 0 to 100 to each of the leads which can be used by the company to target potential leads.



Problem Statement

Company

X Education is an online education company that sells online courses to industry professionals.

Context

The company does marketing of its courses on various websites and search engines such as Google. Once the visitors land on the website, they browse through different courses or fill up a form for the course or watch related videos. When the visitors fill up a form and provide their contact details, they are classified as leads.

Problem statement

X Education gets a lot of leads, the lead conversion rate is low i.e., about 38%. The company wants to identify the best leads, i.e. the leads that have high conversion probability. The company wants to assign a lead score to each lead. The customers with a higher lead score have a higher conversion probability and the customers with a lower lead score have a lower conversion probability.

Data Cleaning

Handling Null values

The columns with more than 40% of null values were dropped

- How did you hear about X Education (78.46%)
- Lead Profile(74.18%)
- Lead Quality (51.60%)
- Asymmetrique Profile Score, Asymmetrique Activity Score, Asymmetrique Activity Index, Asymmetrique Profile Index (45% for all four)

Dropping redundant columns

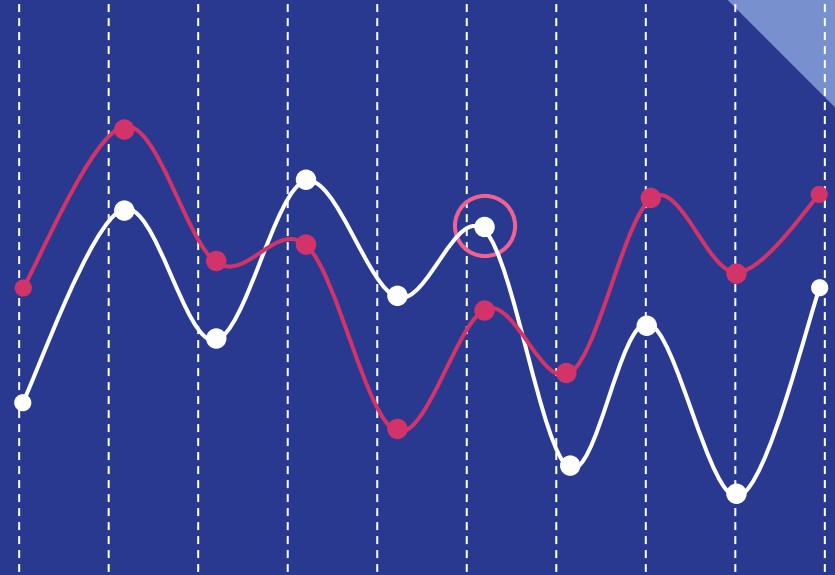
The columns with redundant informations were dropped

- Updates
- Last_activity
- Prospect_ID

Imputing missing values

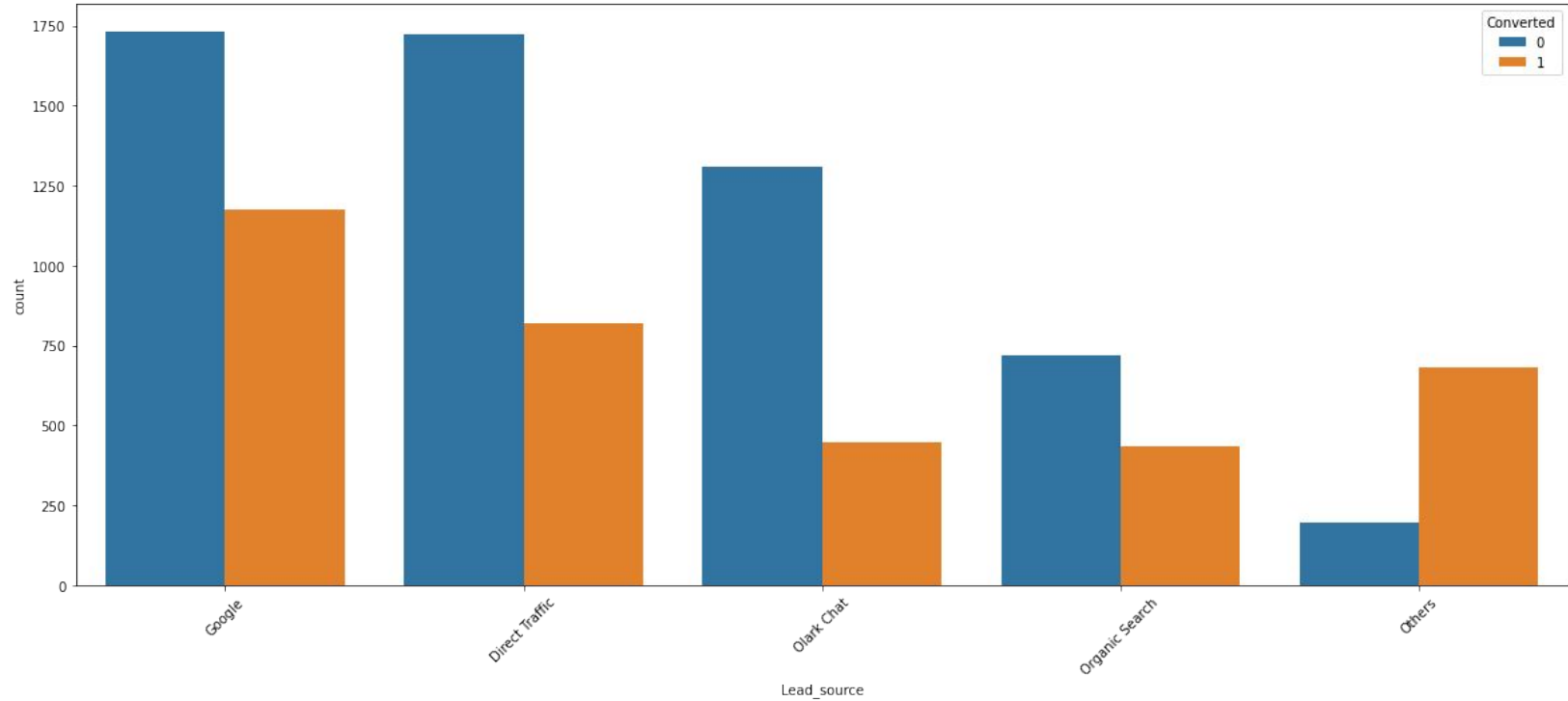
The numerical columns are imputed with median and categorical columns are imputed with mode

EXPLORATORY DATA ANALYSIS



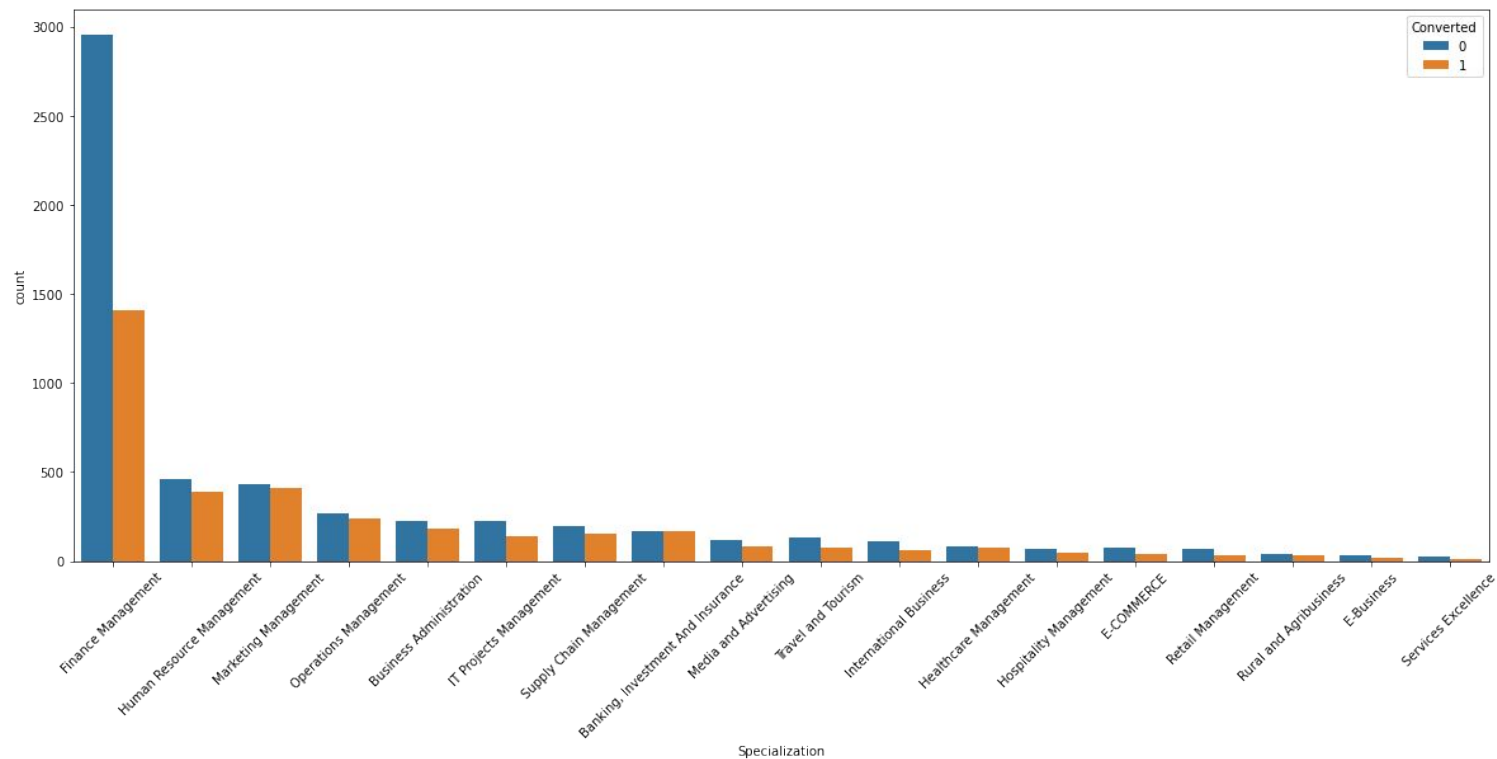
Univariate Analysis

Count plot for Lead_source column



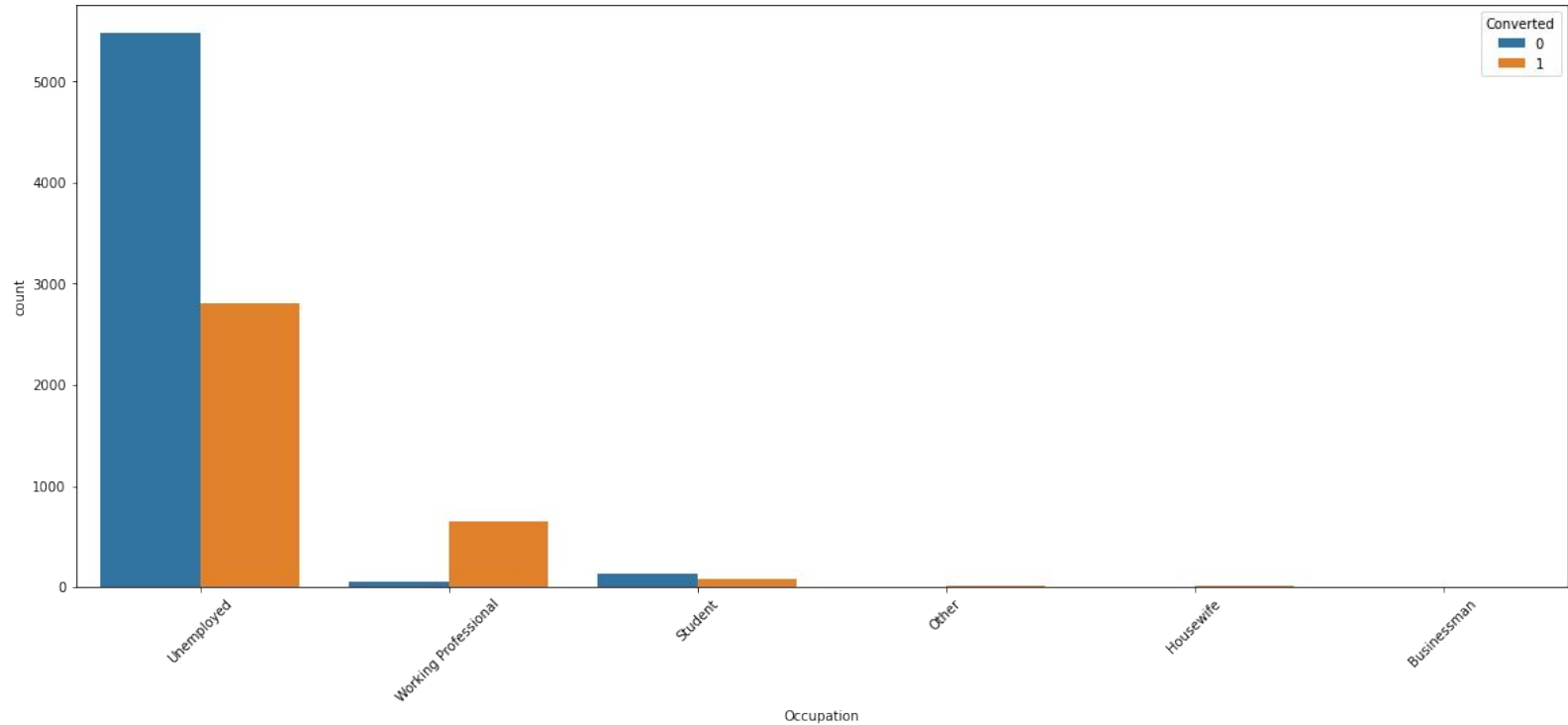
Leads sourced from Google or Direct Traffic will have high conversion rate

Count plot for Specialization column



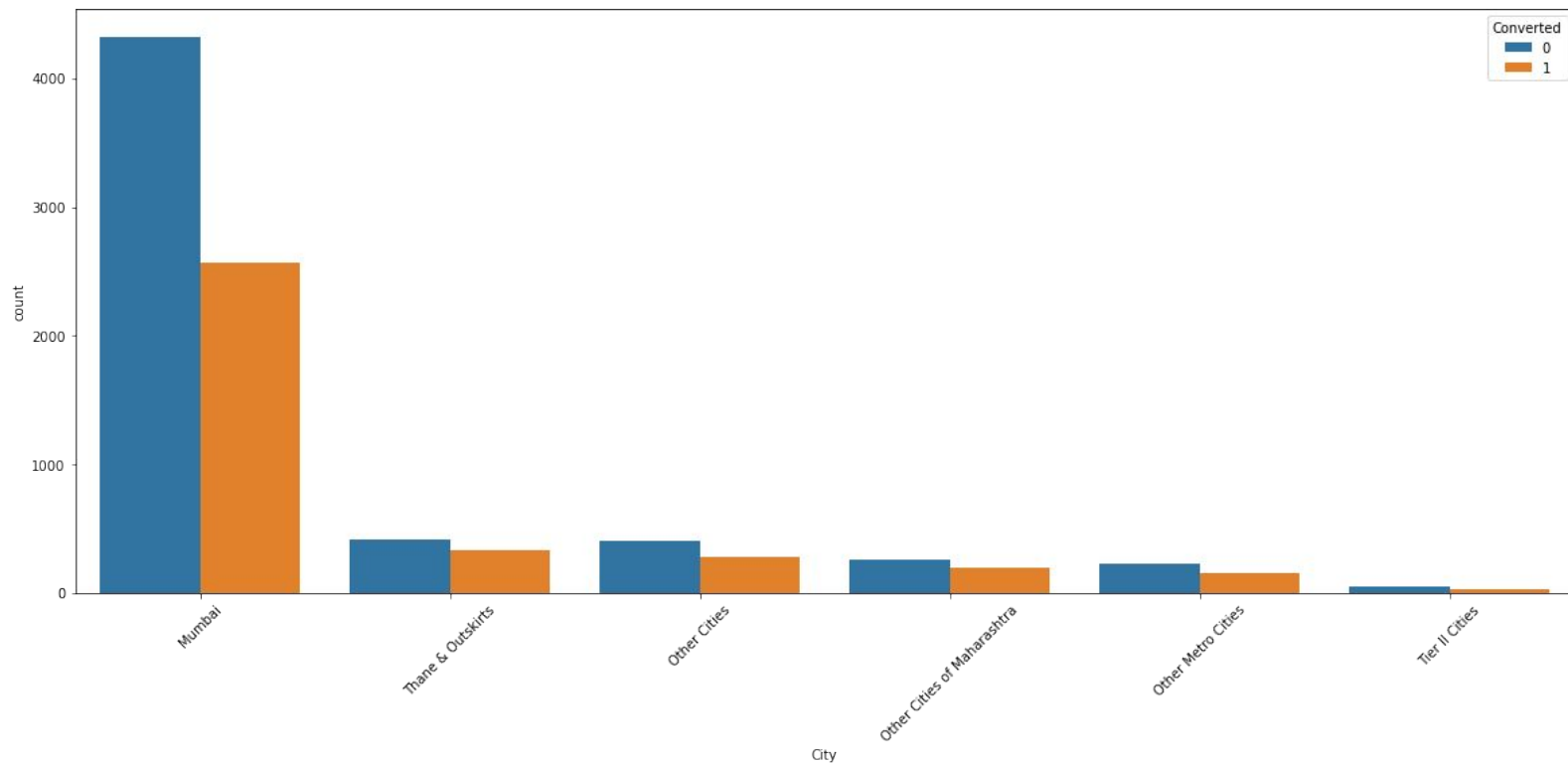
Lead from Finance management specialization having highest conversion rate.

Count plot for Occupation column



- Leads with occupation type as Working professional will have high conversion rate.
- We are having the highest converted leads from the unemployed occupation type but the conversion ratio is less than 50%

Count plot for City column



Most of the leads are from the Mumbai city and having the conversion ratio more than 50%.

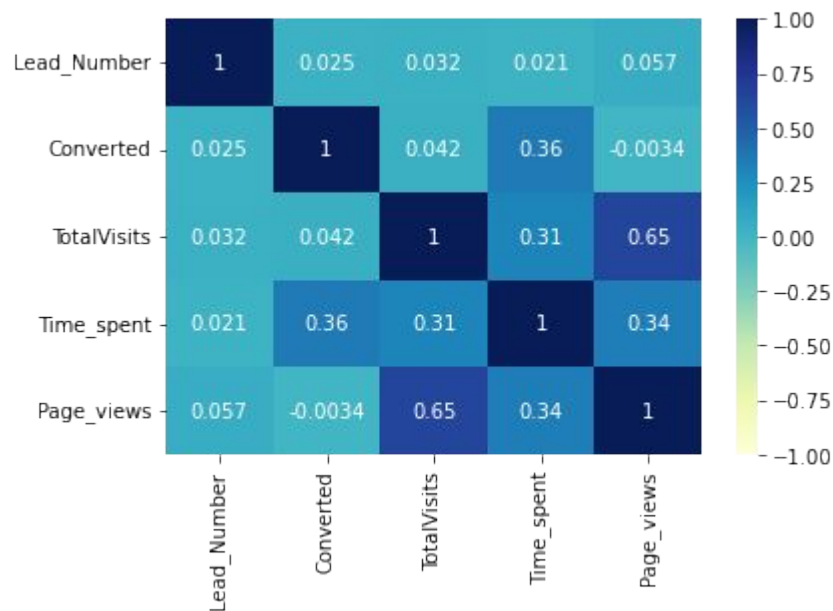
Univariate Analysis - Inferences

- Lead origin
 - Maximum leads are originated from Landing Page Submission and API
 - Conversion rate is also high from these
- Lead Source
 - Maximum leads are sourced from Google and Direct Traffic
 - Conversion rate is also high for these
- Most of the leads have opted for email and phone call communications
- Country
 - ~96% of leads are from India
- Specialization
 - Most of the leads have Finance and Human Resource Management specialization
- Occupation
 - 85% of leads are unemployed with high conversion rate
- Reason_for_Choosing
 - ~99% of leads are choosing the platform for better career prospects
- City
 - Most of the leads are from Mumbai with high conversion rate



Bi-variate Analysis

Correlation



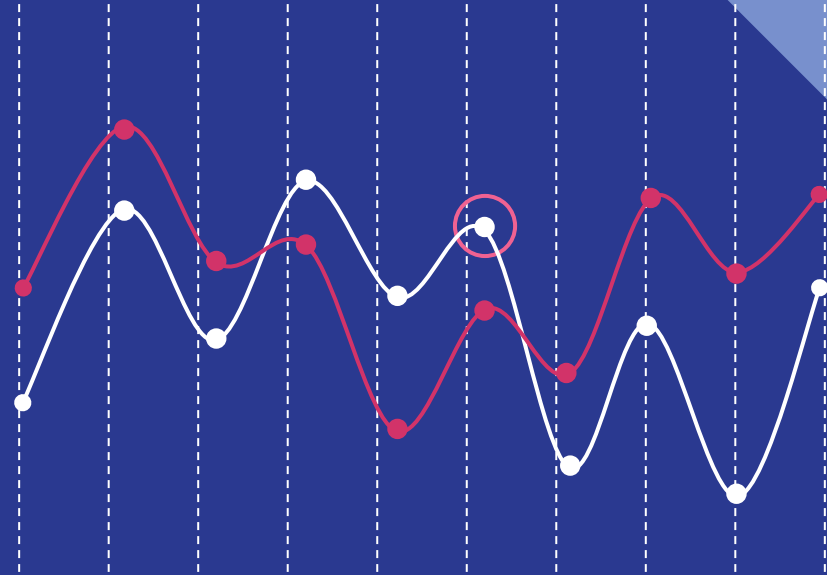
- There is high positive correlation between TotalVisits and Pages viewed per visit
- Time spent is positively correlated with the Converted, Page views and Total Visits.
- Page views is negatively correlated with Converted

Bi-variate Analysis - Inferences

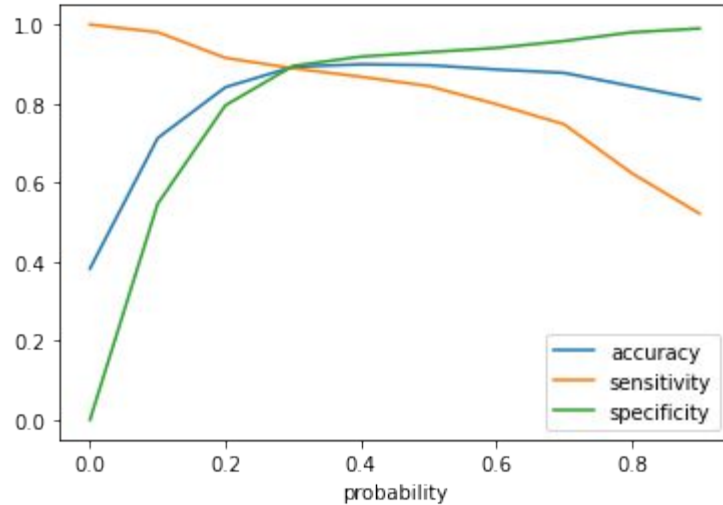
- Numeric-Numeric
 - There seems to be a linear relationship between `TotalVisits` and `Pages viewed per visit`
- Numeric-Categorical
 - Leads who are sourced from organic search have high page views per visit and have the highest conversion, followed by Google and Direct Traffic
 - Leads who are sourced from organic search have spent most of the time on website and have the highest conversion, followed by Google and Direct Traffic
 - Leads who are sourced from organic search have most number of visits on the website and have the highest conversion, followed by Google and Direct Traffic
 - Leads who are choosing the services for better prospects have high conversion



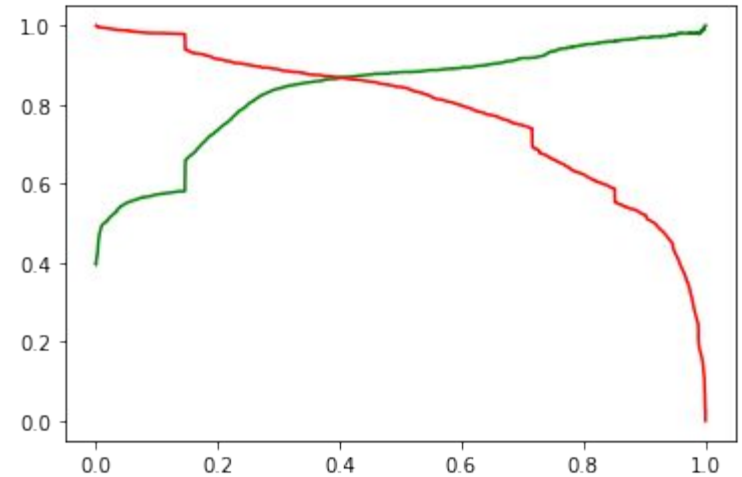
Model Details



Model Evaluation - Train Dataset



The graph shows the optimal cut off of 0.3 based on Accuracy, Sensitivity and Specificity.



The graph shows an optimal cut off of 0.4 based on Precision and Recall

Model Metrics - Test Dataset

Confusion Matrix	
3692 (TN)	278 (FP)
383 (FN)	2077 (TP)

Accuracy
89%

Sensitivity
86%

Specificity
91%

Precision
86%

Recall
86%

Conclusion

- The leads having high 'Lead Score' can be focused on more for better conversion rate.
 - Marketing on Google, since the conversion rate from the traffic from Google is high.
 - Encouraging existing converted leads for referrals by providing some incentives for the referrals.
 - Since the number of leads is high in Mumbai as compared to other major cities, the company can increase marketing in the other cities as well to achieve more leads.
 - The unemployed category can be focused on more and also individuals having Finance Management as specialization.
 - Focus on the students can be minimized since the conversion rate is significantly low.
-