

LEAD SCORING CASE STUDY

GROUP MEMBERS:

BHUVANESH KUMAR J

VINODKUMAR GURUGUBELLI

PROBLEM STATEMENT

- X Education is an education company which sells online courses to industry professionals.
- Though X Education has lot of leads, the conversion rate of the leads is very less. Only around 30% of the leads are getting converted.
- The Leads are categorised into two categories.

Hot Leads – Leads which have high probability of conversion

Cold Leads – Leads which have low probability of conversion

- The aim of this analysis is to identify the potential hot leads, so that the sales team will communicate with the potential leads and make the conversion more effective.

BUSINESS OBJECTIVE

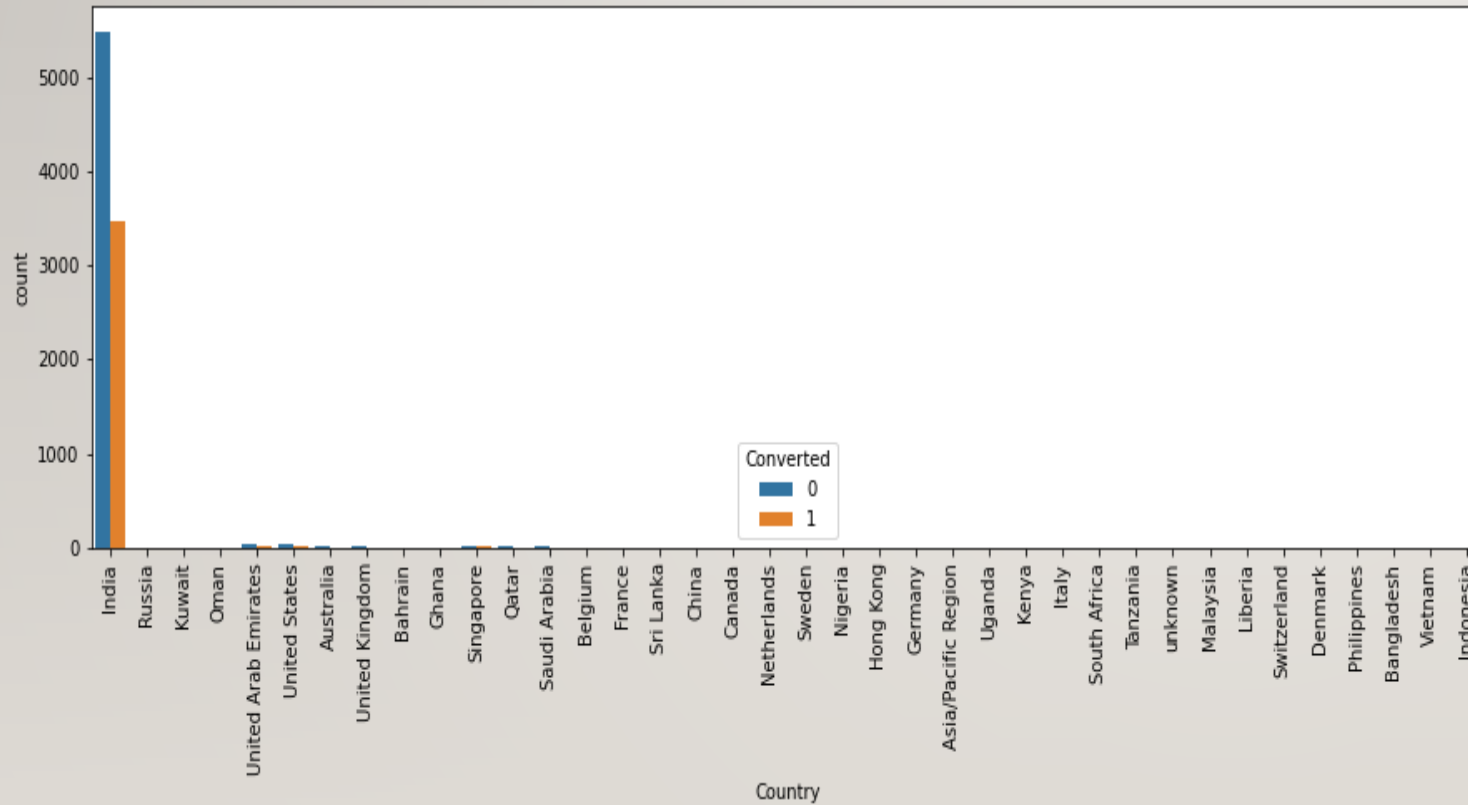
- X education wants to identify the potential/ Hot Leads.
- To identify the potential leads, X education needs to build a Logistic regression model.
- The top features which has to be focused on for maximum conversion has to be identified for sales force to effectively reach out to the leads and yield higher conversion.
- Deploy the model for future use for new set of Data/ modified data.

LOGISTIC REGRESSION MODEL

- Cleaning data
- EDA
- Dummy Variables
- Train-Test split
- Model Building
- Model Evaluation
- Prediction
- Conclusion and recommendation

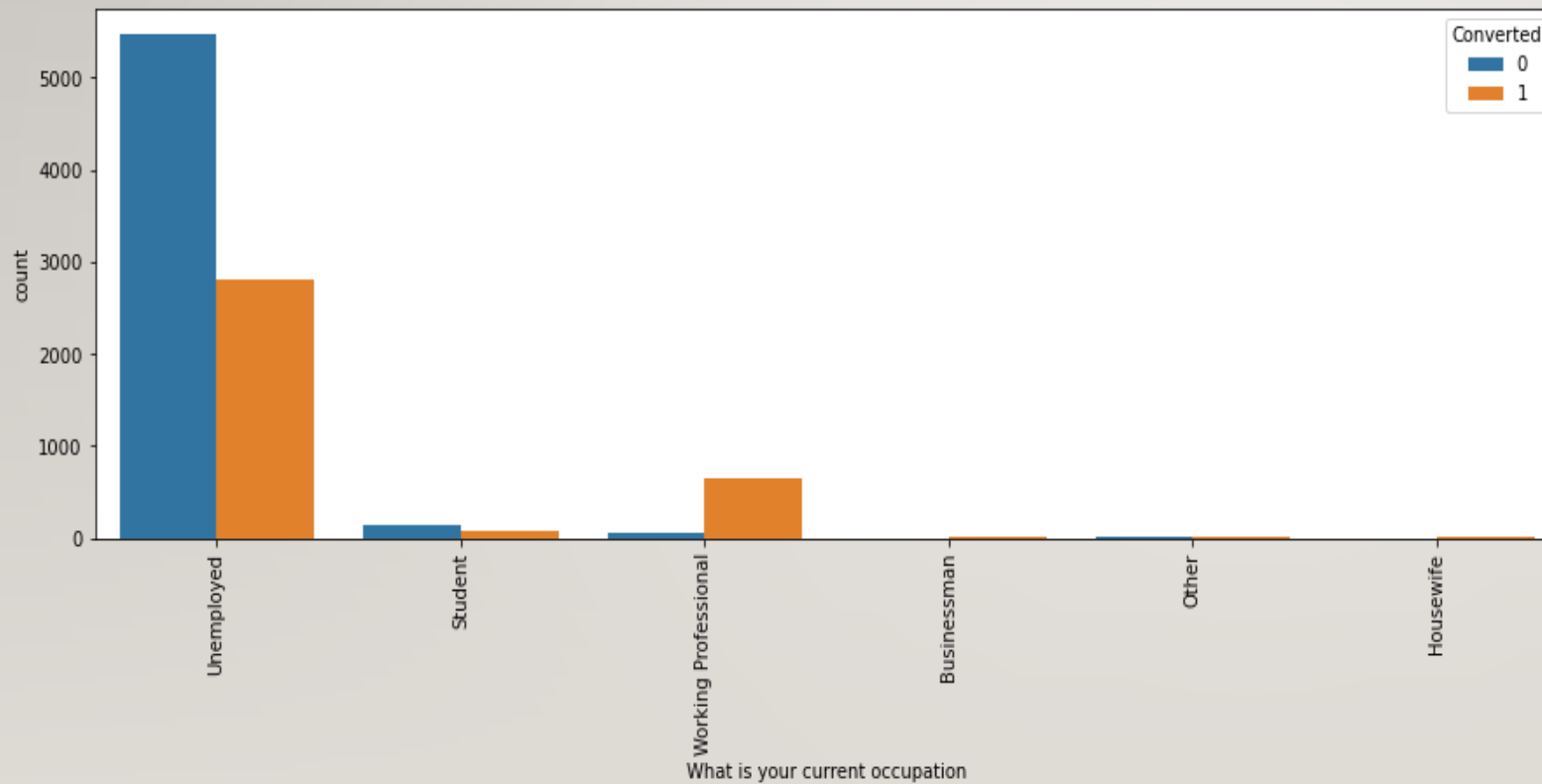
EDA

DISTRIBUTION OF COUNT ACROSS COUNTRY



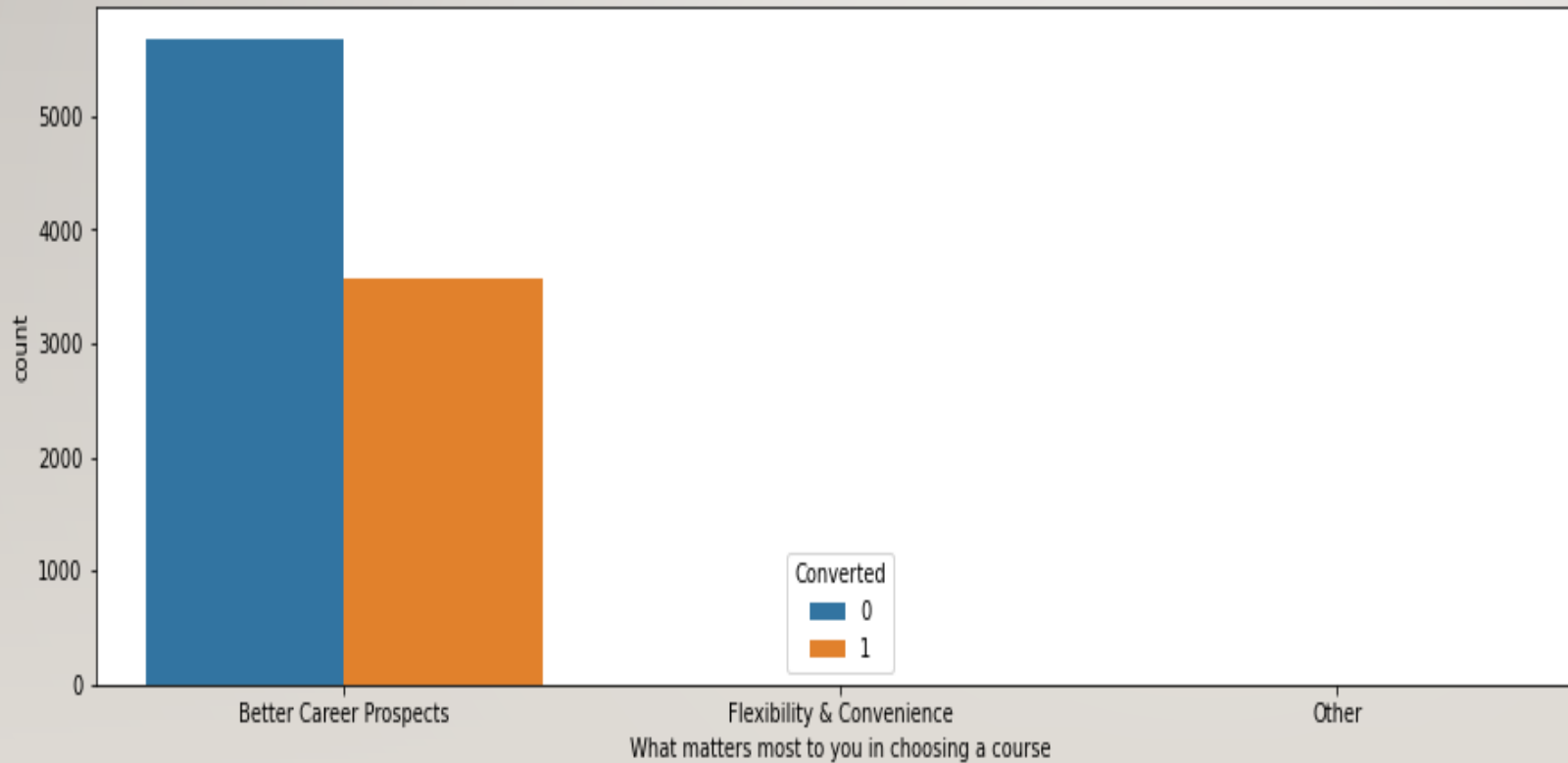
- From the graph, we are able to see that almost most of the people who view the details or get converted are only from India.
- Hence we can remove the Country category from analysis

DISTRIBUTION OF COUNT ACROSS OCCUPATION



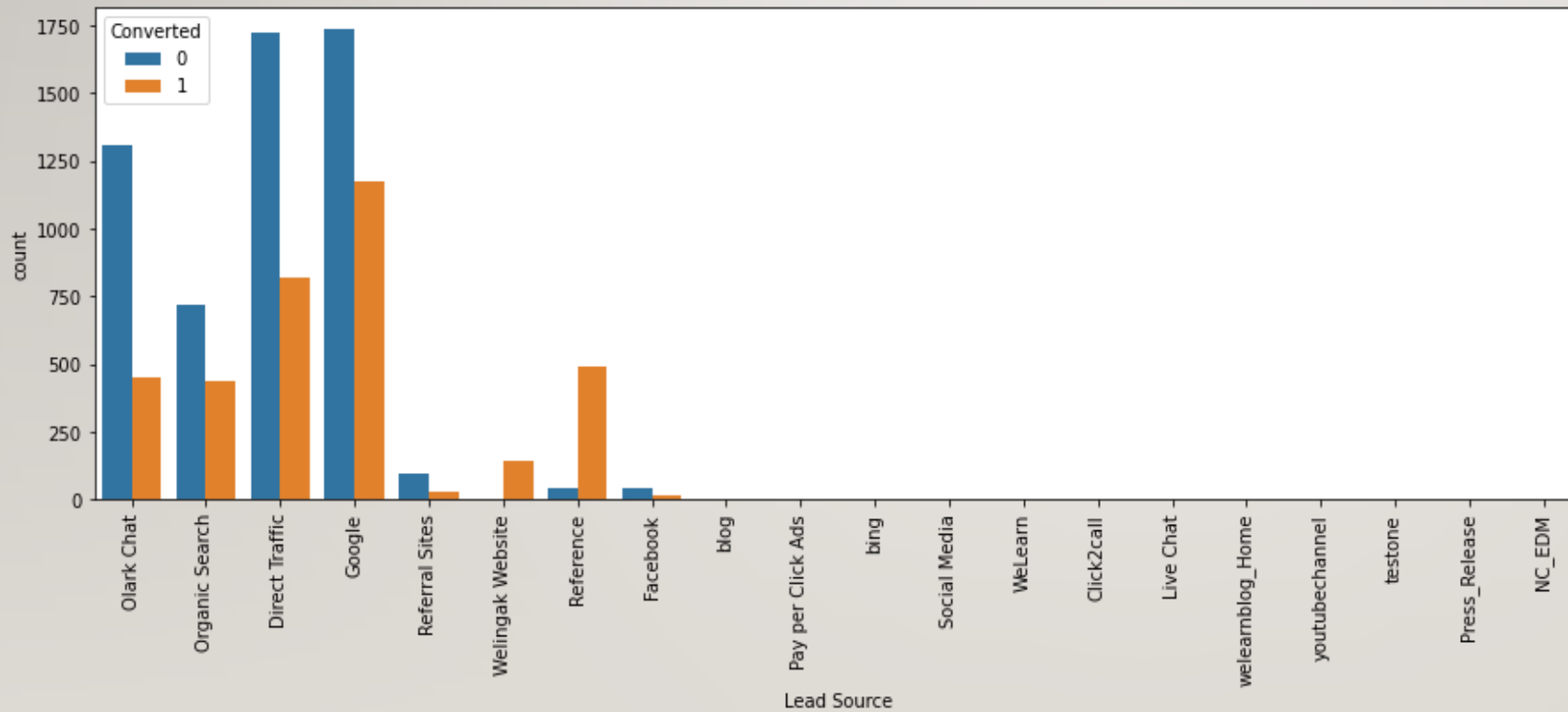
- From the graph, we can see that the people who are unemployed contribute to the majority of the conversions.
- However, the working people also prefer the course. In fact the conversion rate of the Working Professional is very high than unemployed people.

DISTRIBUTION OF COUNT ACROSS REASON FOR CHOOSING



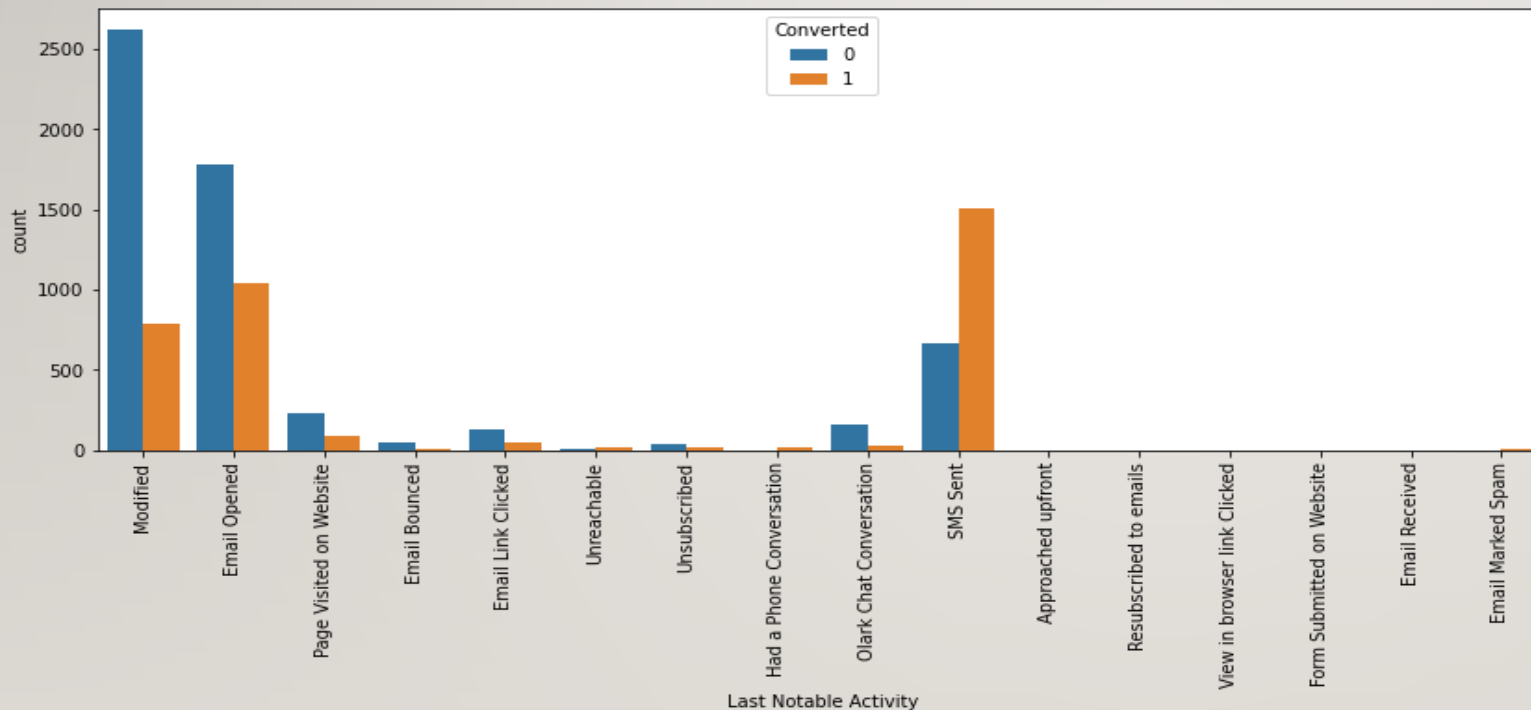
- It is Clear that almost everyone who views the details or getting converted are checking for Better Career Prospects.

DISTRIBUTION OF COUNT ACROSS SOURCE OF LEAD



- The source of Leads are majorly through Google, Direct Traffic, Organic search and Clark Chart.
- The Conversion rate of the leads are very high when they get to know about the course through reference.

DISTRIBUTION OF COUNT ACROSS LAST ACTIVITY OF LEADS

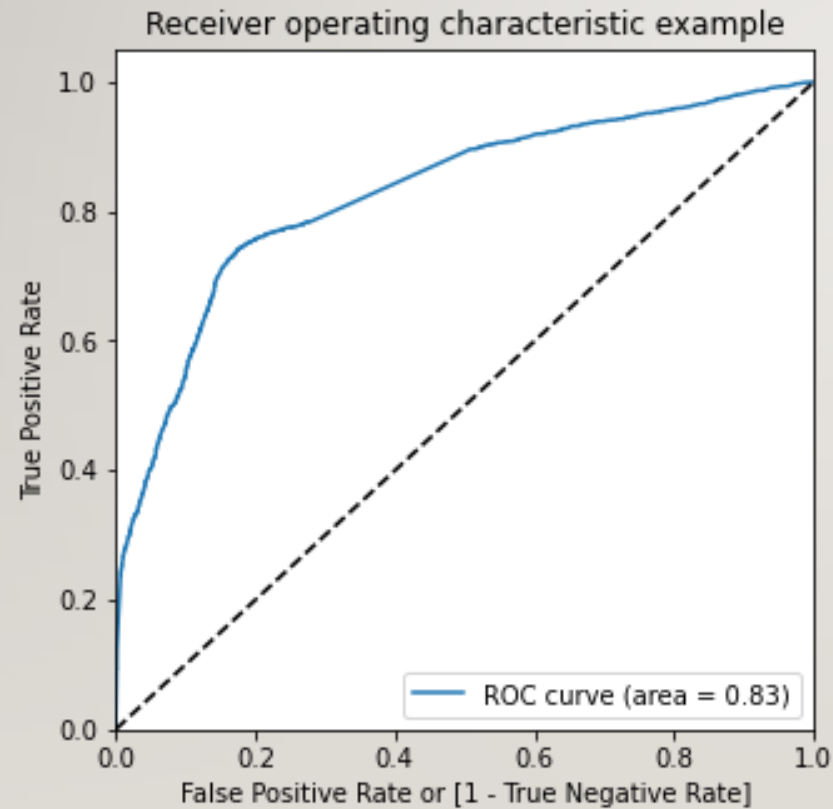


- The Leads who have sent SMS have the highest conversion rates followed by the Leads who have opened emails. The conversion rate is low when the details have been modified.
- The Lead's activity through other sources are less.

MODEL BUILDING

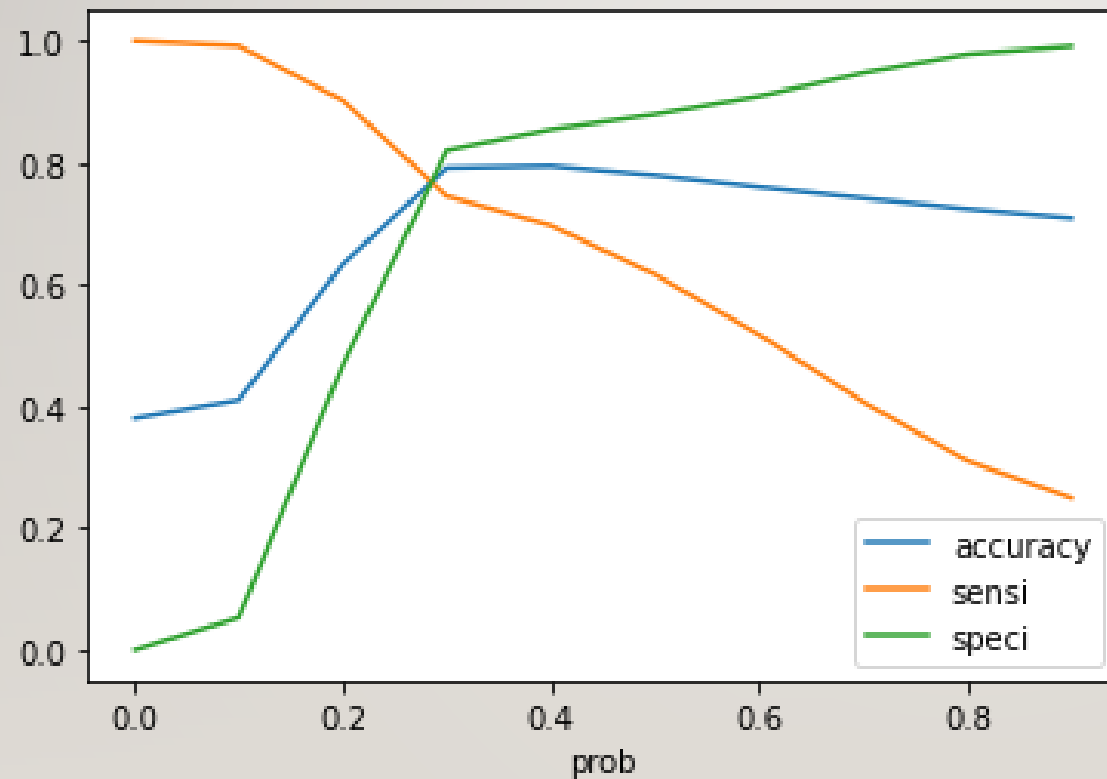
- The Dataset has been divided into Train and Test Datasets with the ratio of 70:30 respectively, which is the basic step of the logistic regression.
- The top 15 features have been selected from the model using RFE (Recursive Feature Elimination)
- Then the variables which were irrelevant was removed one after another based on their VIF values and P-values. (The variables with $VIF < 5$ and $p\text{-value} < 0.05$ were kept).
- The accuracy, sensitivity and specificity were also found using ROC curve.

ROC CURVE



- An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at classification thresholds.
- The generated model seems to be performing well. The ROC curve has a value of 0.83, which is good.

OPTIMAL CUTOFF



- The intersection point at which there is a balance between sensitivity and specificity; it corresponds to the optimal cutoff on logistic regression model
- From the graph, 0.3 is the optimal cutoff point.

RECOMMENDATION AND CONCLUSION

- It was identified that the variables that mattered the most in the potential lead conversion are:
 - Total Time Spent on Website
 - Lead Origin_Lead Add Form
 - current occupation_Working Professional
 - Lead Source_Direct Traffic
 - Lead Source_Google
 - Lead Source_Organic Search
- Considering this, the X Education can identify their Hot Leads and Cold Leads and focus on the hot leads for effective conversion of leads.

THANK YOU

