



# Lead Scoring Case Study

# Problem Statement



X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.

X Education is looking to filter out the leads that can be converted to paying customers.

Although X Education gets a lot of leads, its lead conversion rate is very poor. The typical lead conversion rate at X education is around 30%. To make this process more efficient, the company wishes to identify the most potential leads.

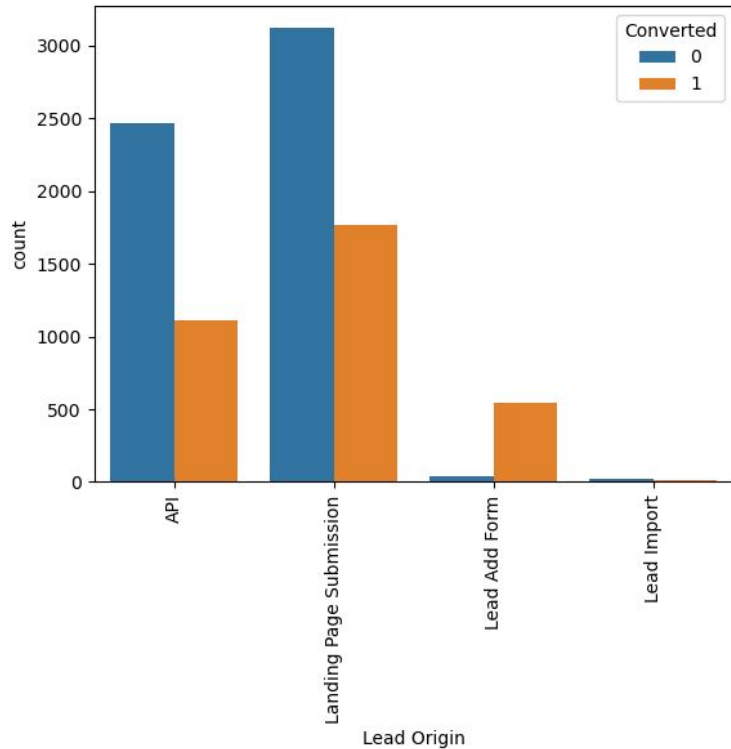
The Company is looking to solve this low conversion issue and needs the model that can help to increase the conversion rate to 80%

# Analysis Approach



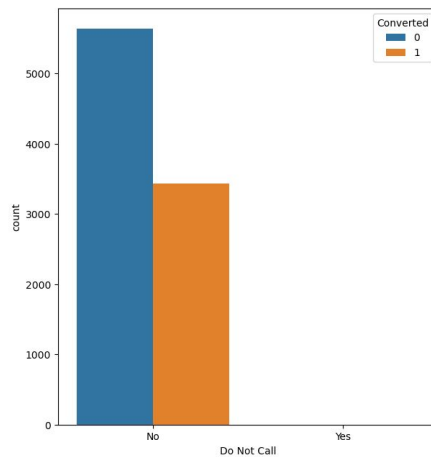
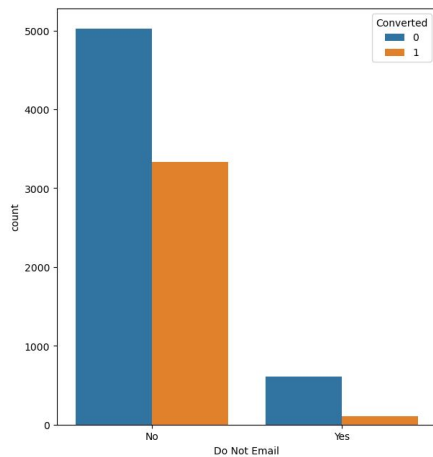
- Understanding the Data, Cleaning it and Preparing Data for exploratory data analytics.
- Performing exploratory data analytics to determine helpful insights.
- Preparing Data for model building like dummy variable and feature scaling.
- Building logistic regression model using RFE and Manual Feature Selection.
- Assigning lead score.
- Training the model on train set.
- Evaluation of model from train set using different metrics.
- Testing the model on test set
- Evaluation of model from test set using different metrics and thus evaluating the accuracy of model.

# Exploratory Data Analytics



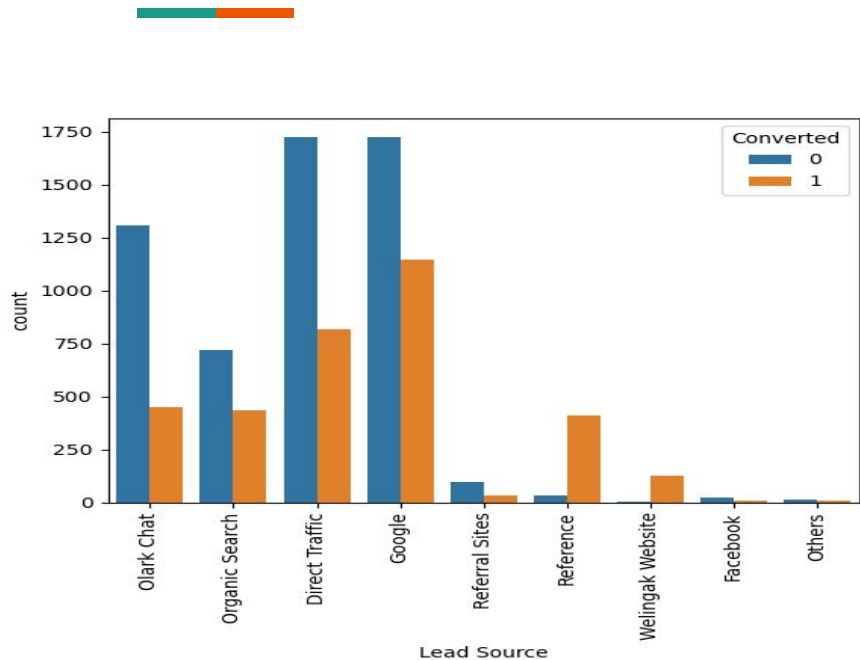
## Lead Origin vs Converted

- Lead Add Form has a very high conversion rate compare to other Lead Origin but a lower Lead origin count which needs to be improved.
- Although API and Landing Page Submission have comparatively lower conversion rate but they have significantly high count of Lead origin, which indicates we need to focus on improving the conversion rate of these two.



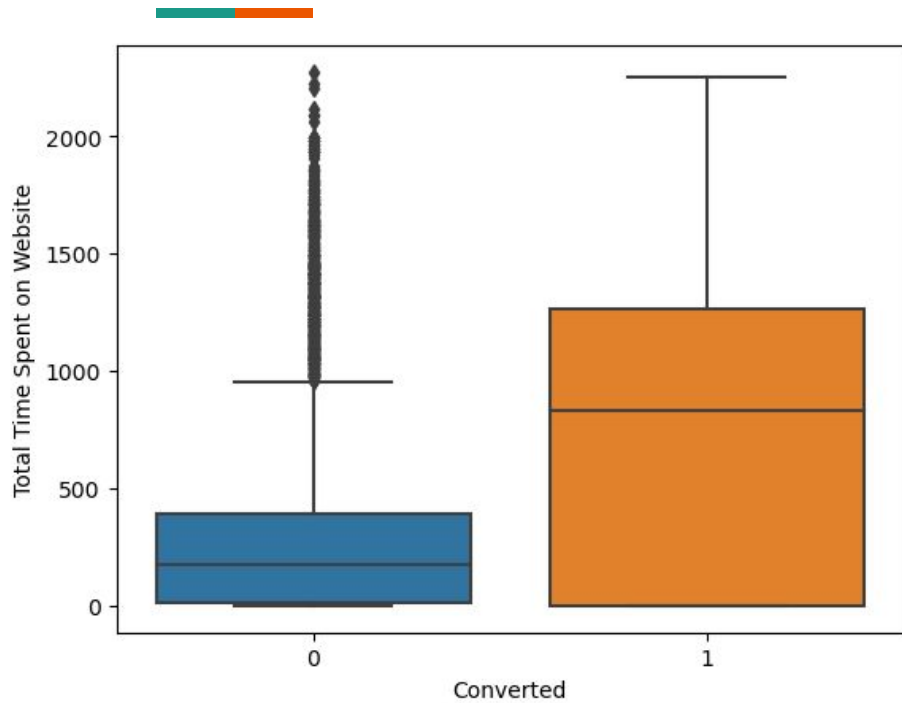
## Do Not Email & Do Not Call

- Both the variables have higher count of 'No' value and conversion rate of both the variables 'No' value is same.



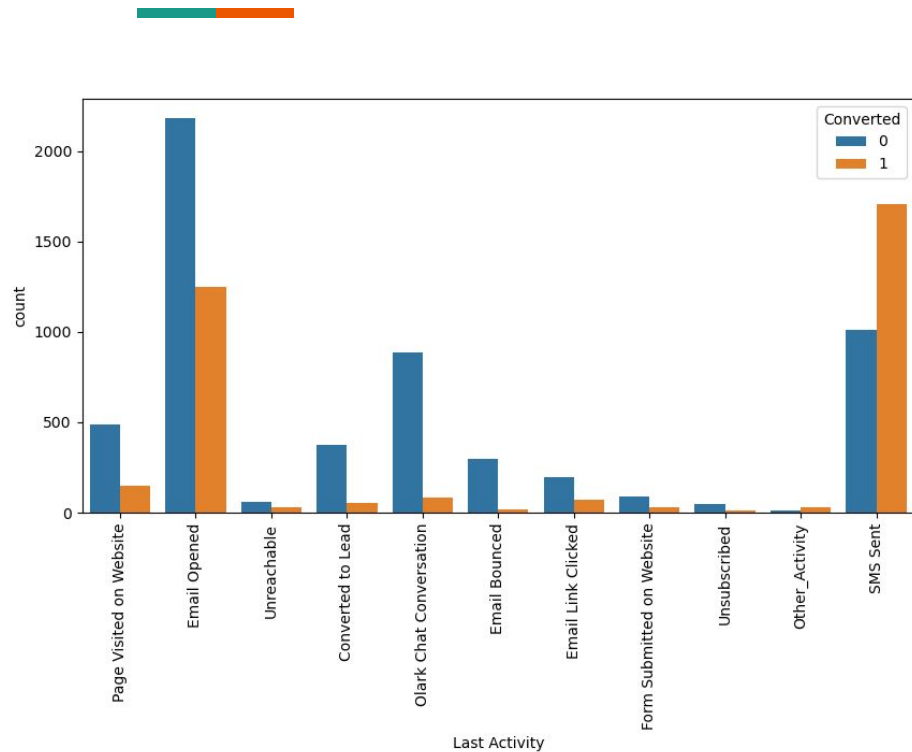
## Lead Source

- Google and Direct Traffic have higher Lead count comparing to others.
- Reference and welingak website have higher conversion rate.



## Total Time Spent on Website

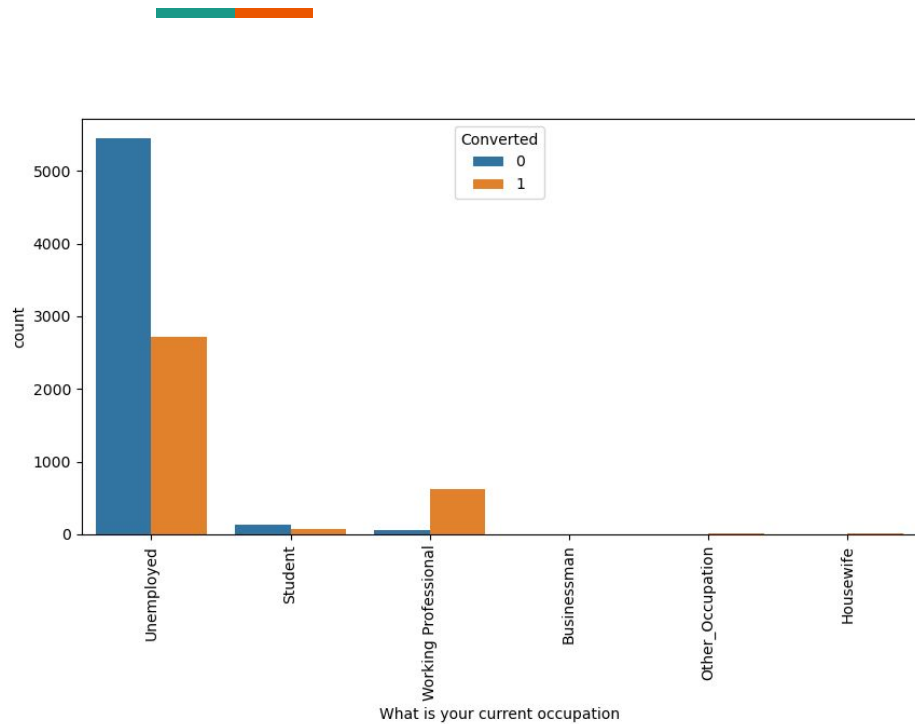
- Conversion rate of lead spending more time on website is high, so website should be made more engaging.



## Last Activity vs Converted

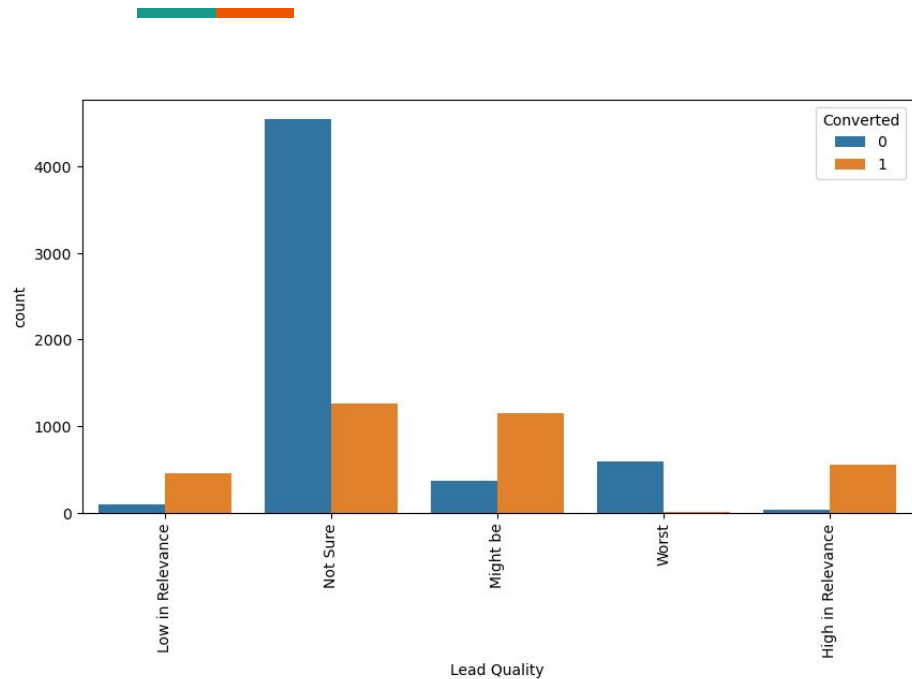
- Email opened has the most number of lead count in Last activity.
- SMS Sent has an high conversion rate.





## Occupation vs Converted

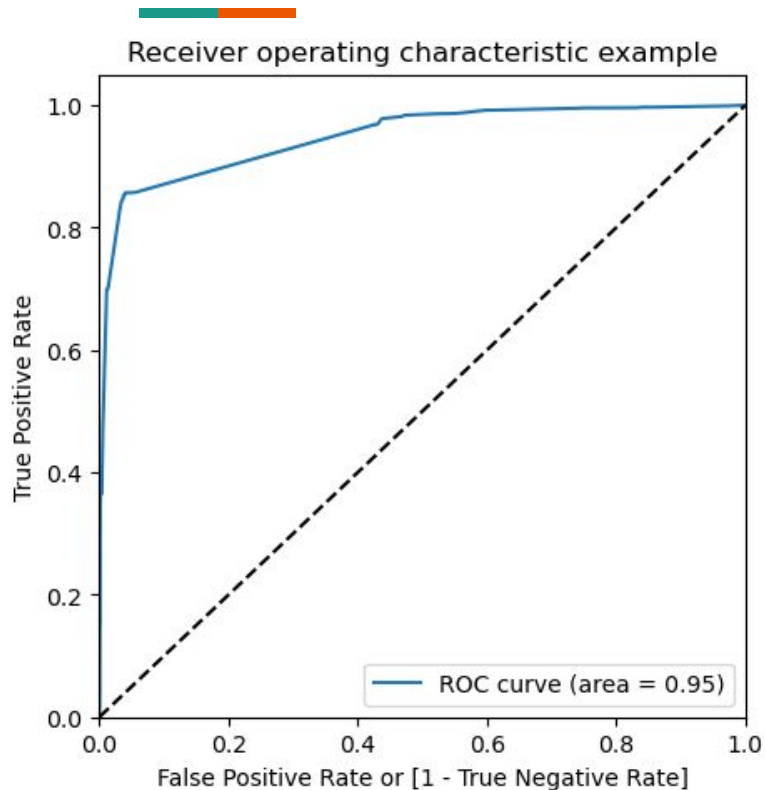
- Working Professionals have higher chances of conversion.
- Unemployed leads have higher counts.



## Lead Quality vs Converted

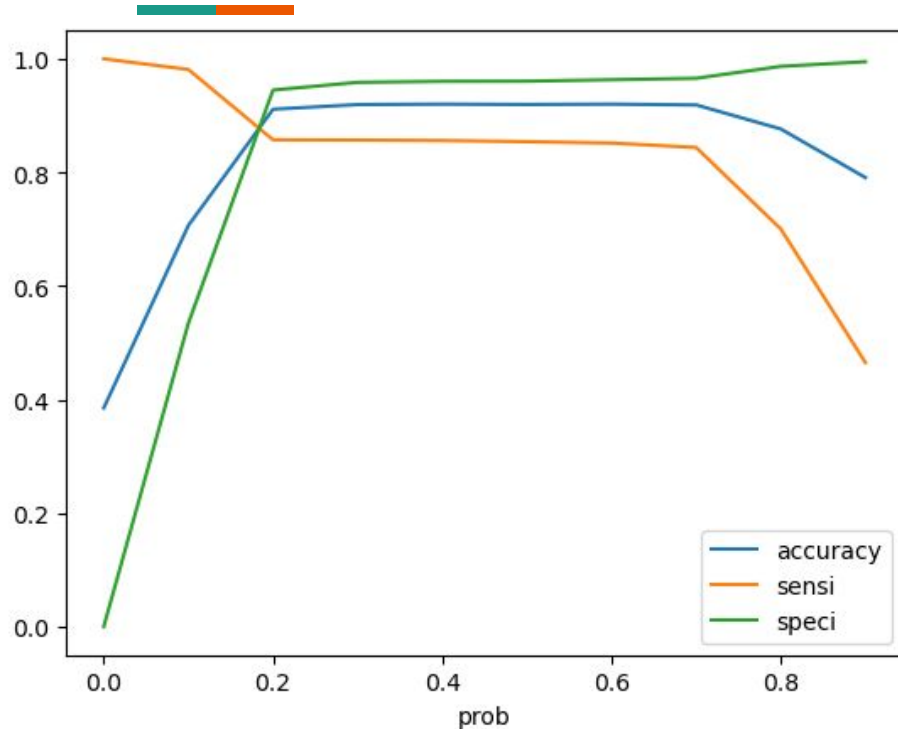
- Lead Quality with Not sure has high lead count but poor conversion rate.

# ROC Curve



From the ROC Curve we can see that it is quite close to the left hand edge of the curve indicating the model to be quite good.

# Model Evaluation (Train)



## Confusion Matrix, Accuracy, Sensitivity and Specificity

3751	154
357	2089

91.1% Accuracy  
85.7% Sensitivity  
94.4% Specificity

## Precision and Recall

93.1% Precision  
85.4% Recall

# Model Evaluation (Test)



## Confusion Matrix, Accuracy, Sensitivity and Specificity

1640	94
157	832

90.7% Accuracy  
84.1% Sensitivity  
94.5% Specificity

## Precision and Recall

89.8% Precision  
84.1% Recall

# Conclusion



## EDA

- Leads spending high times on site are likely to buy the course.
- SMS sent to leads seems to make an great impact.
- Unemployed leads can be a great leads.
- Lead from reference have high chance of conversion.

## Model

- Model has an accuracy of 91.1% on training set and 90.7% on test set
- Model shows 85.7% sensitivity and 94.4% specificity on train set and 84.1% sensitivity and 94.5% specificity on test.
- The precision and recall of model is 93.1% and 85.4% on train set and 89.8% and 84.1% on test set respectively.
- From the above data we can conclude that the model is pretty accurate