# Lead_Case_Study Presentation.

## 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%.

## Steps

- Data Exploration and Understanding
- Data Cleaning
- Exploratory Data Analysis and Data Visualization
- Dropping Irrelevant columns found from Data Visualization
- Splitting the data into Test and Train dataset.
- Scaling
- Building a logistic Regression model and calculate Lead Score.

- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall. ¬
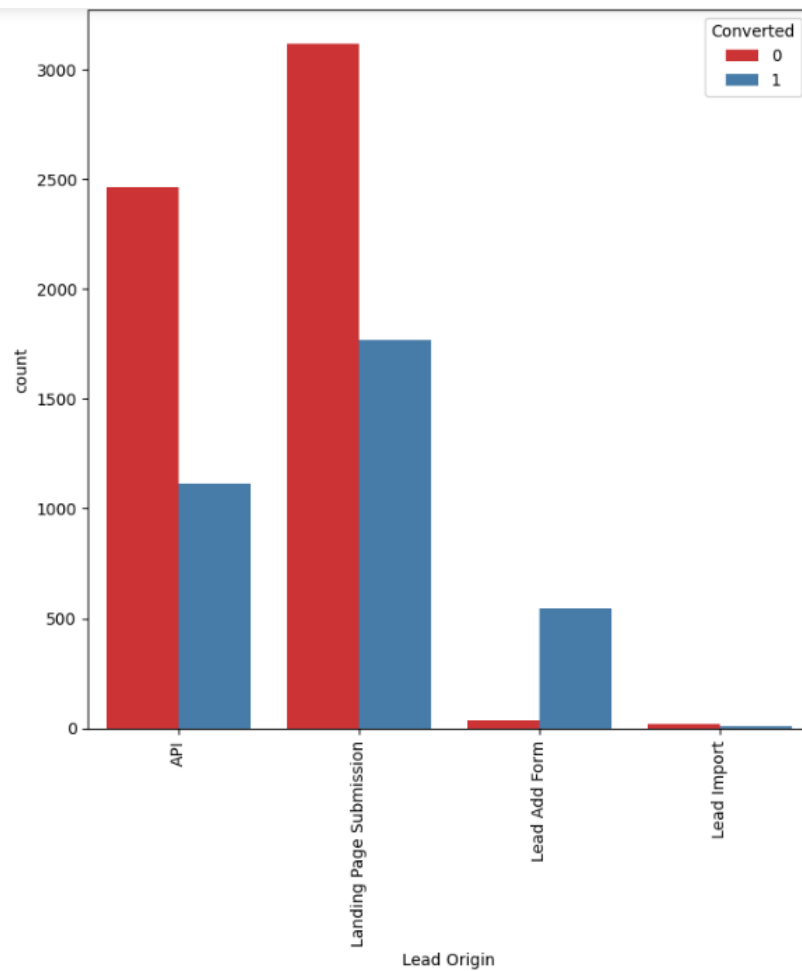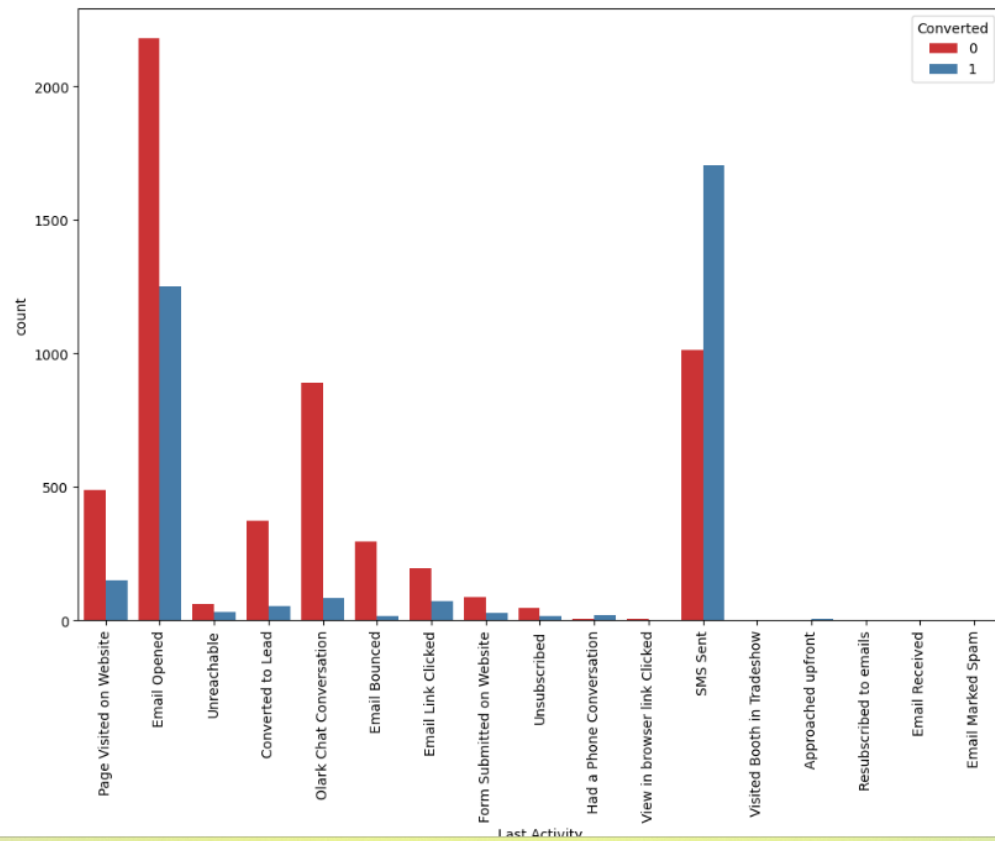- Applying the best model in Test data based on the Sensitivity and Specificity Metrics
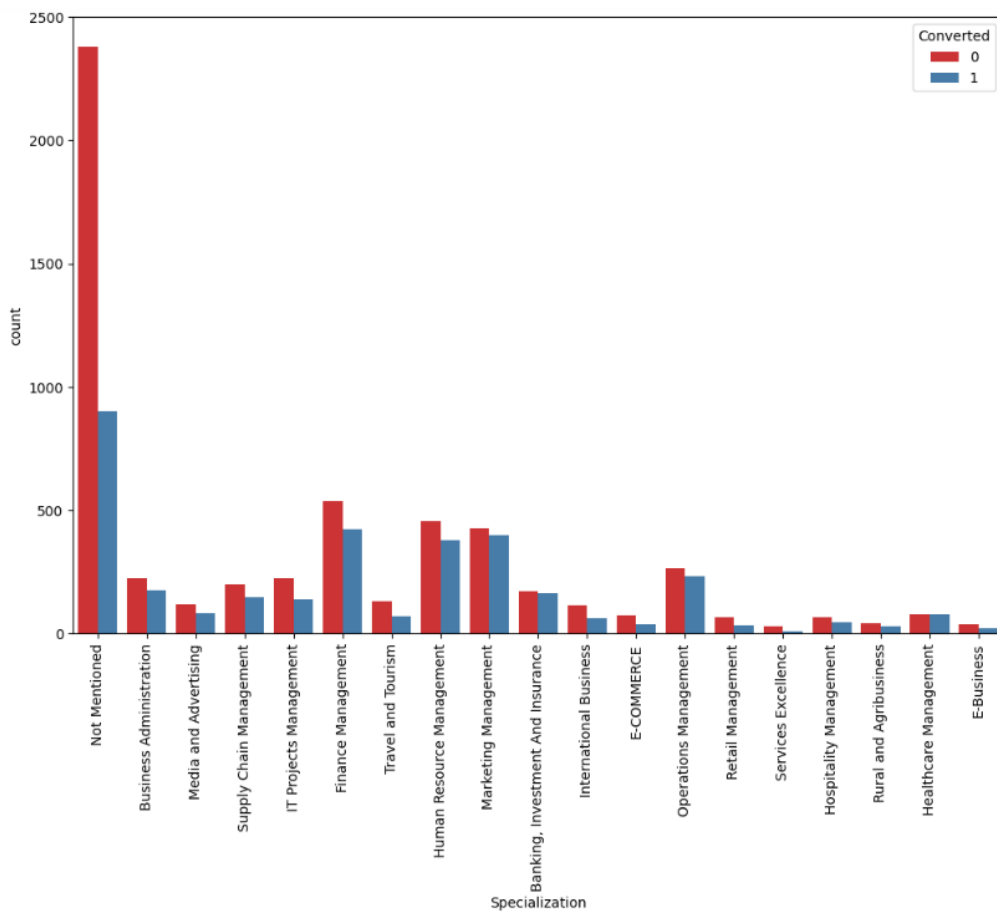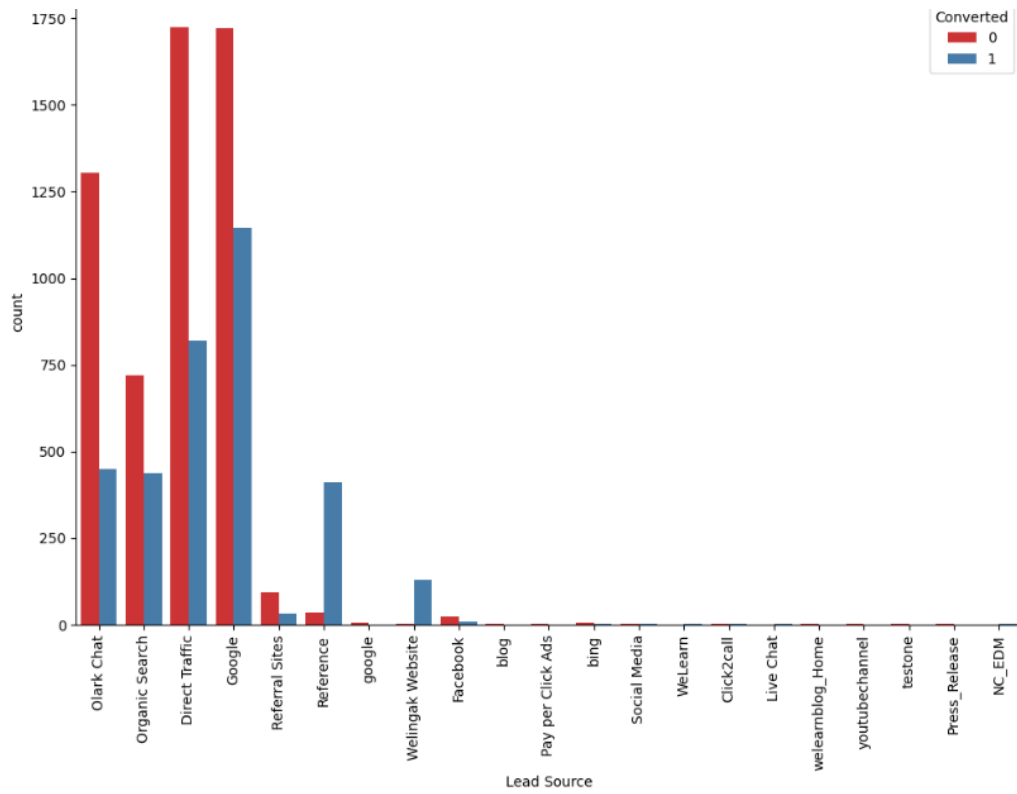
## Early Insights

Distribution of Converted Column

Lead Converted

37.9%

62.1%

Lead Not Converted
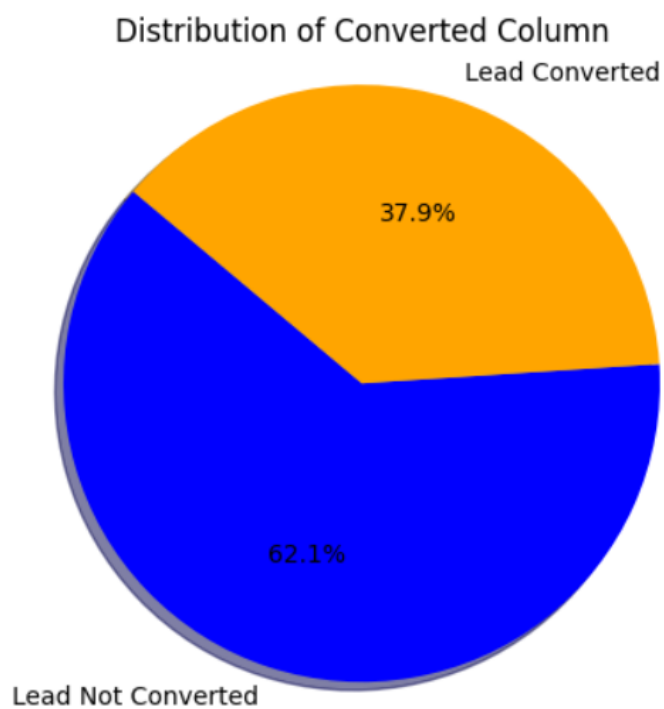
# Initially as we see the Conversion rate is 38%.

# Insights from Data Visualization.
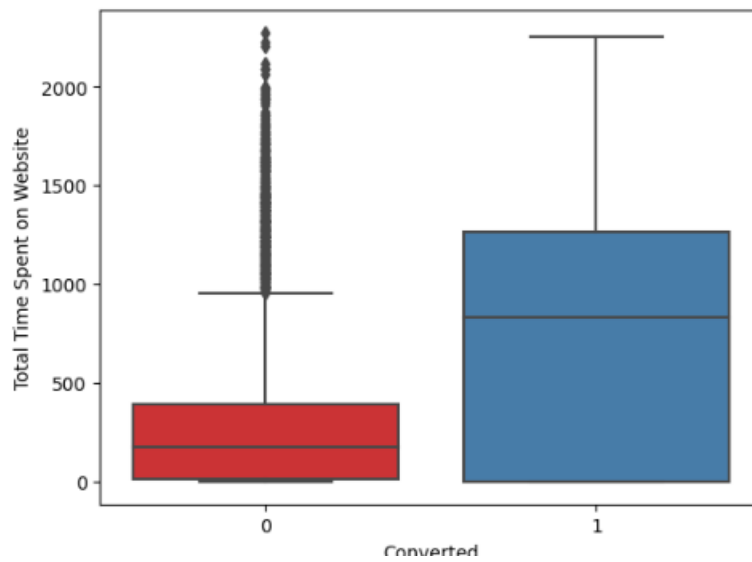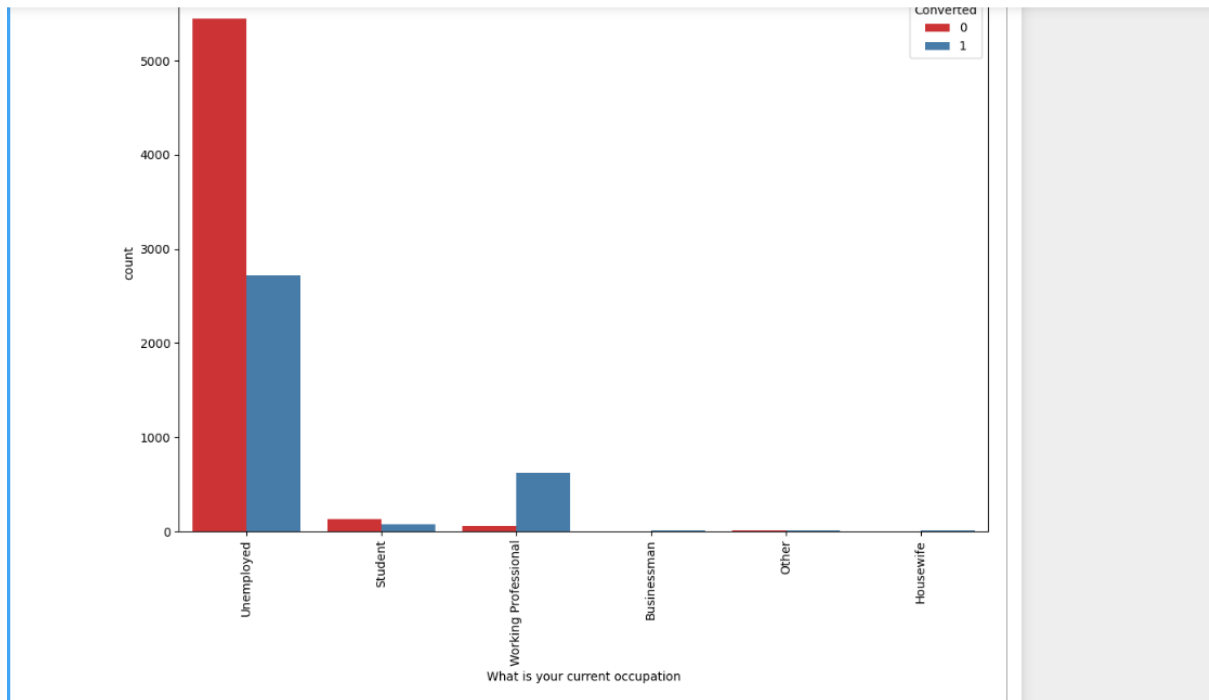
We can clearly see the top variables which have impact the Conversion Rate

- **Reference**
- **Welingak Website**
- **Lead Add form**
- **SMS SENT**
- **Working Professional**
- **Total Time spend on website.**

In most of them it is important to note that number of leads are not that significant, so the recommendation would be to increase number of leads in this variables.

We are seeing also seeing some variables which are having very poor Conversion rate but are having higher leads.

So, there are two options

- **Improve the Conversion rate of these variables:** This might be a costly method because already data is showing lower Conversion rate and increasing efforts in it might not result in expected outcome.
- **Improve number of leads in variables that have good Conversion rate:** This is more significant approach while we already know from data their promising signs and increasing number of leads might result in expected outcome.

## Model Evaluation –

Accuracy Sensitivity and Specificity on Train Data Set.

The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity.
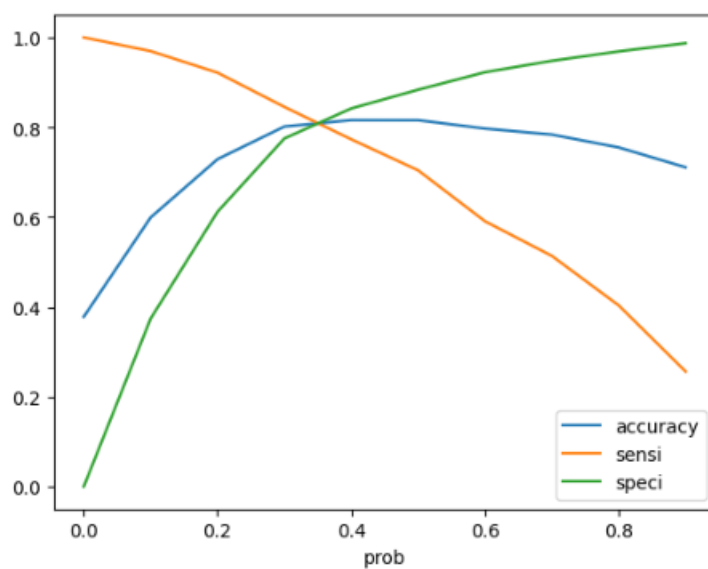


**Accuracy** - 81%

**Sensitivity -** 81 %

 **Specificity -** 82

**False Positive Rate** - 18 %

**Positive Predictive Value** - 73 %

**Positive Predictive Value** – 87%

**Accuracy, Sensitivity and Specificity on Test Data Set**

**Accuracy** - 80%

**Sensitivity** - 79 %

**Specificity** – 80%

## <u>Conversion.</u>

Accuracy, Sensitivity and Specificity values of test model is around 80%, 79% and 80% which are closer to their respective values calculated using trained model. Although it can be improved further

The lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 81% which is above the CEO ballpark of the target lead conversion rate of around 80%.

Hence overall this model seems to be good.