

# **Presentation on LEAD SCORING CASE STUDY**

**Prepared BY:**

**PUJARANI SAMAL**

**Ramakrushna Mohapatra**

**Rohit Bharadwaj Kallakuri**

## Problem Statement

X Education is an online education company that specializes in selling courses to industry professionals. They market their courses on various websites and search engines such as Google. When interested professionals visit their website, they may browse through the courses or fill out a form to express their interest. If they provide their email address or phone number, they are classified as a lead. Additionally, X Education also receives leads through referrals.

Once a lead is acquired, the sales team initiates contact with the individual through calls, emails, and other communication channels. While some leads do get converted into paying customers, the majority do not. The sales team uses various strategies to try and convert leads, but the success rate varies. Overall, X Education relies on lead generation and conversion to drive its business forward. By understanding the behavior of their leads and refining their sales approach, they aim to increase their conversion rate and grow their business.

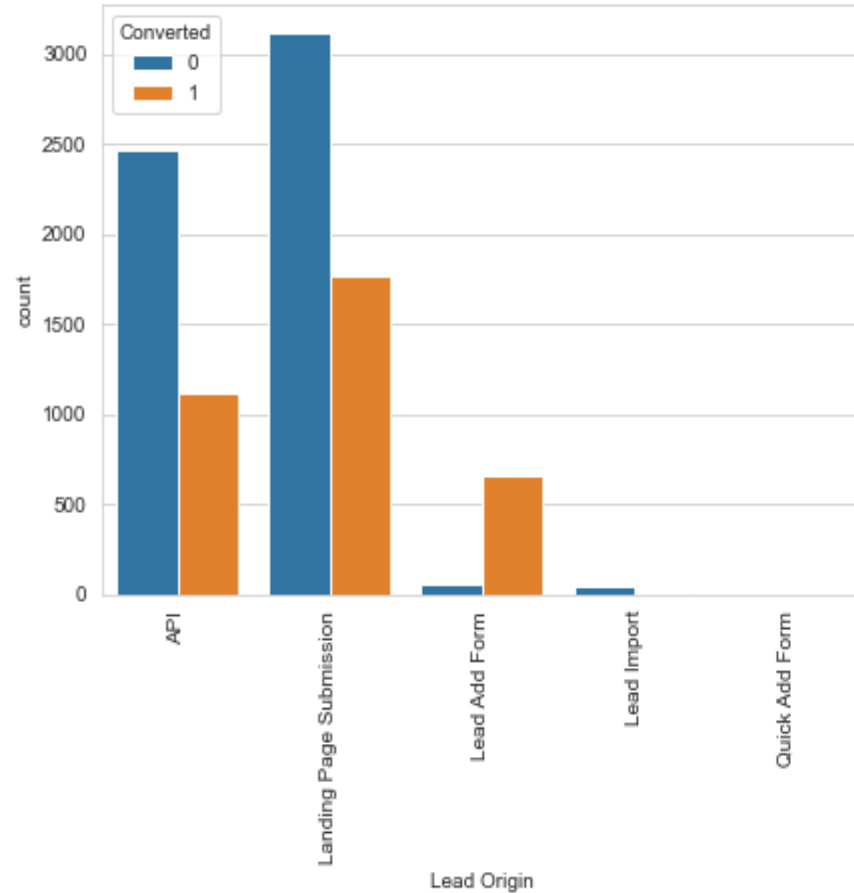
## Steps

- Understanding of Raw Data
- Data Cleaning and Manipulation
- EDA can be used to identify the most significant attributes that impact the target variable.
- Preparing the data for modeling (train-test split, rescaling, etc.)
- Build a Logistic Regression Model
- Evaluate the model by different measures and metrics
- Predictions and evaluation on the test set
- Model evaluation

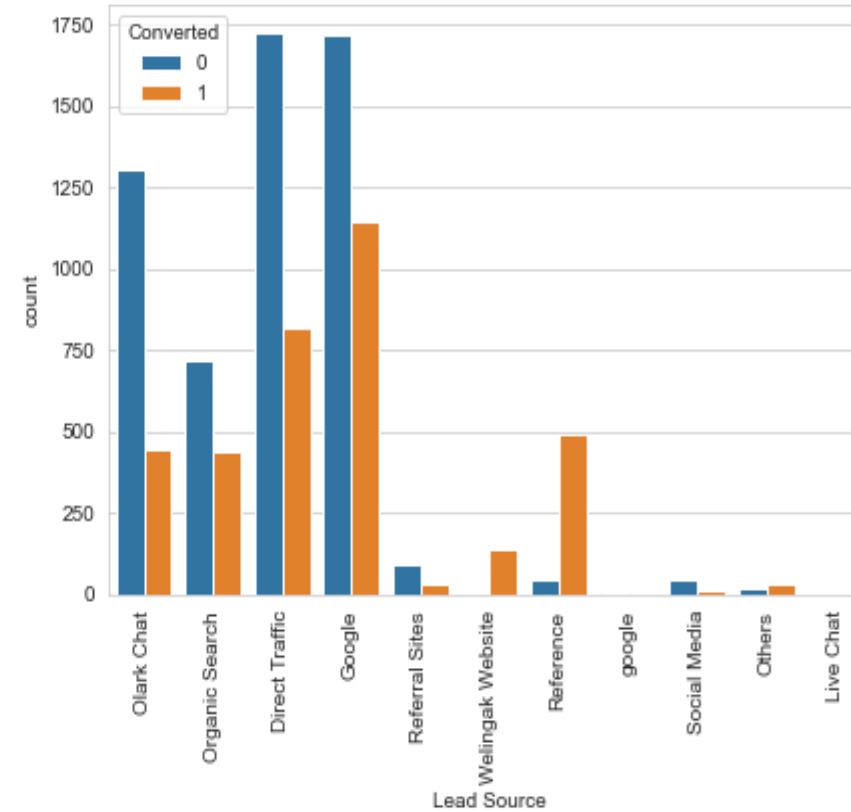
# Preparation of Raw Data

- Preparation of Raw Data using two different files Lead i.e 'leads\_df' and leads\_df1.
- Selected the Required Clean data by removing Null values & Impute null values Value.
- Removing irrelevant columns in the data.
- Based on the values and categorical values.
- Feature Min-Max scaling.

# EXPLORATORY DATA ANALYSIS:

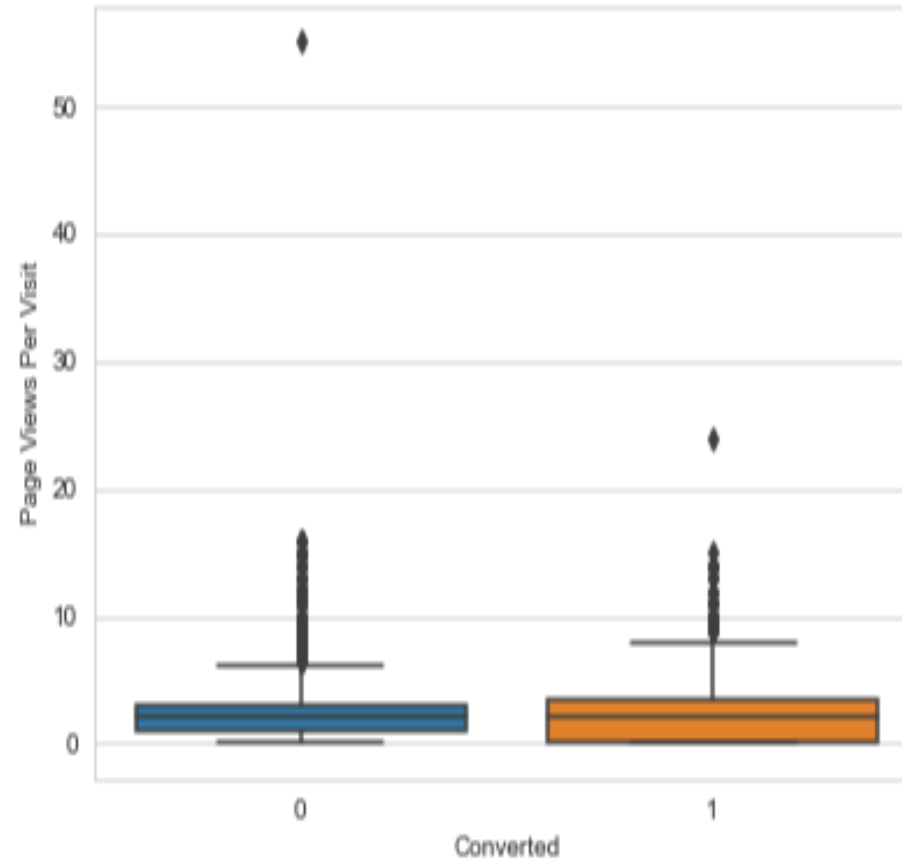
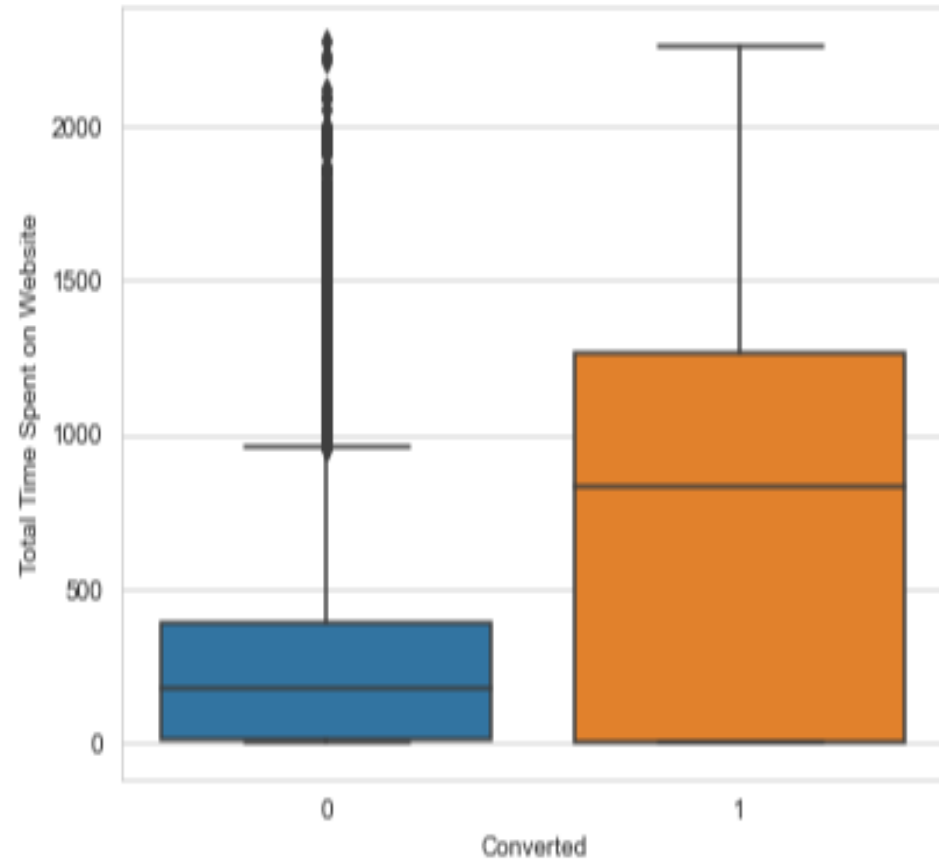


Lead origin vs Converted



Lead Source vs Converted

# Total Time Spent on Website      Page Views Per Visit



## **Data Preparation**

- 1-Converted Binary variables into 0 and 1.
- 2-Created Dummy variables for categorical variables

## **Feature scaling & splitting Train & Test Sets**

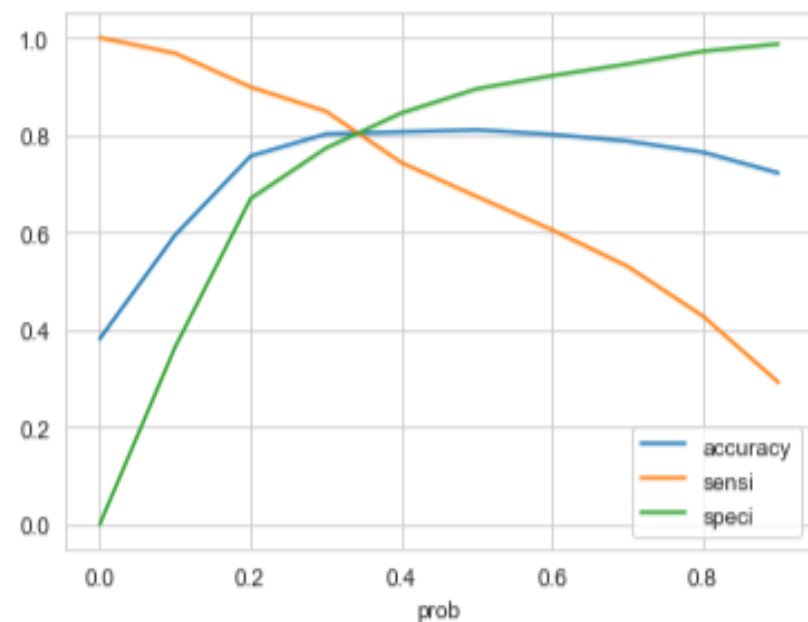
- 1-Feature scaling of Numeric data
- 2-Splitting data into Train & Test Sets

# Model Building

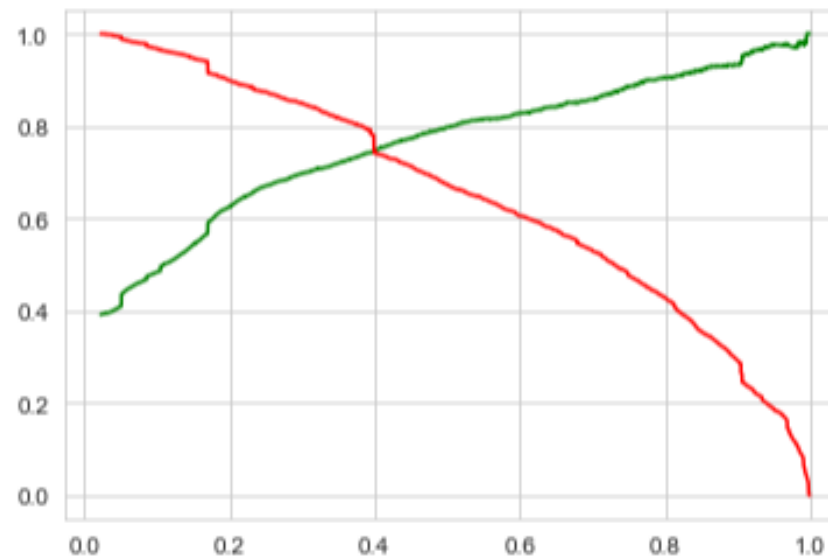
- 1-Feature selection using RFE
- 2-Determined Optimal model using Logistic Regression
- 3-Calculate accuracy, sensitivity, specificity, precision & Recall & Evaluate model



## Model Evaluation(Train)

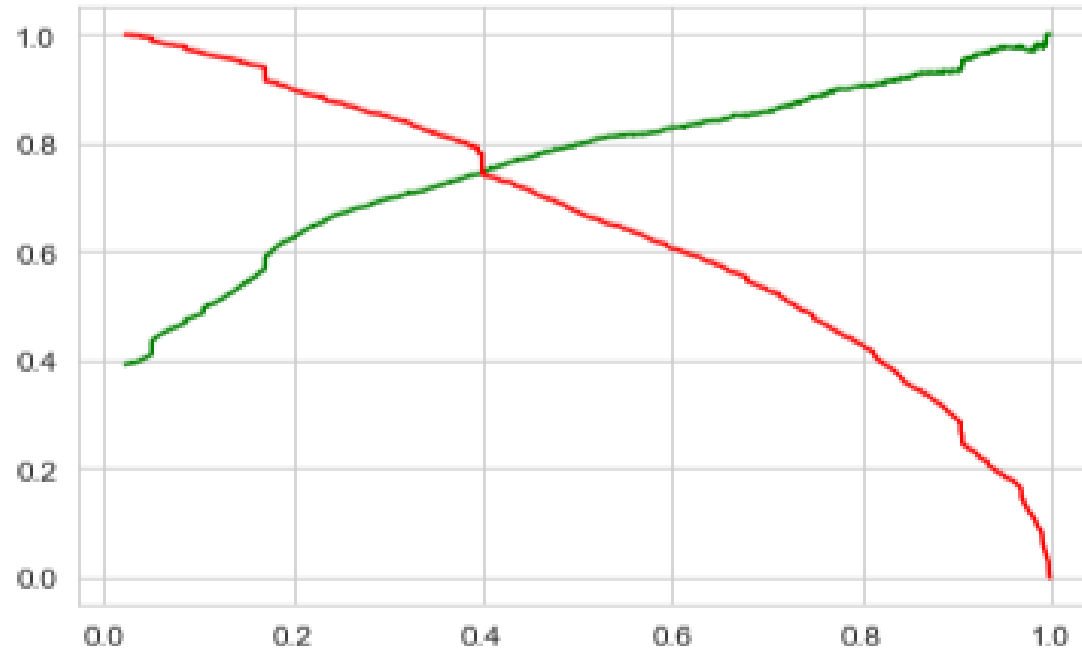


81.0 -Accuracy  
67.4 -Sensitivity  
89.5-Specificity



71.5- Precision  
82.2 -Recall

## Model Evaluation(Test)



Accuracy: 0.8152958152958153, Sensitivity 0.7990867579908676, specificity 0.8258795468097794

## CONCLUSION

- The best cut-off based on sensitivity and specificity was taken into consideration when calculating the final prediction, even though we had already evaluated both sensitivity and specificity as well as Presence and Recall metrics.
- The final model's lead score for conversion rate on Train and Test is 89% and 82%.
- Therefore, the overall model appears to be good.

