

# Lead Scoring Case Study

Presented By:

Ankita M.

Hrishabh M.

# Agenda

- The Purpose is to optimize the lead scoring mechanism based on their fit, demographics, behaviors, buying tendency etc. By implementing explicit & Implicit lead scoring modelling with lead point system.

# Problem Statement

- ▶ An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. 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%. Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

# Goals of the Case Study

- There are quite a few goals for this case study. Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

# Approach

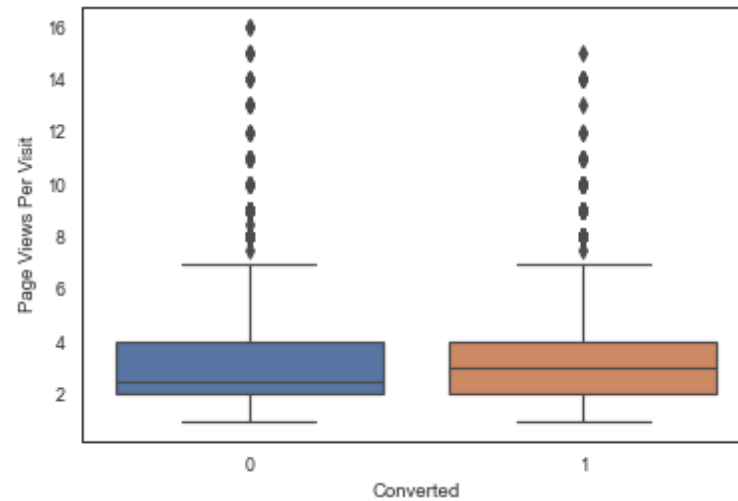
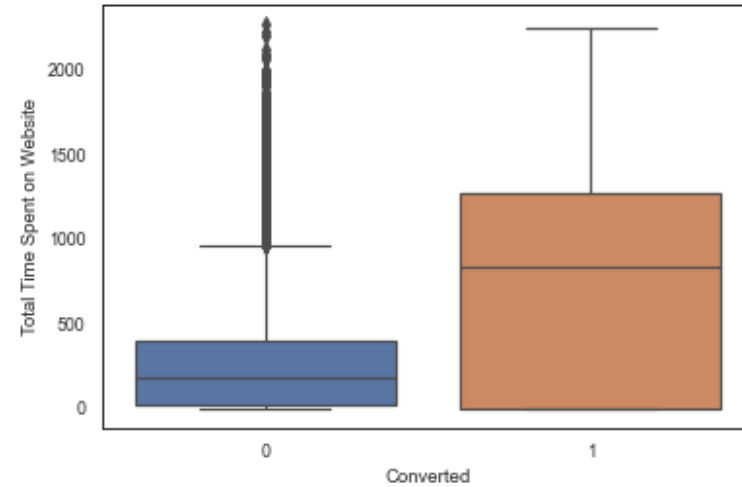
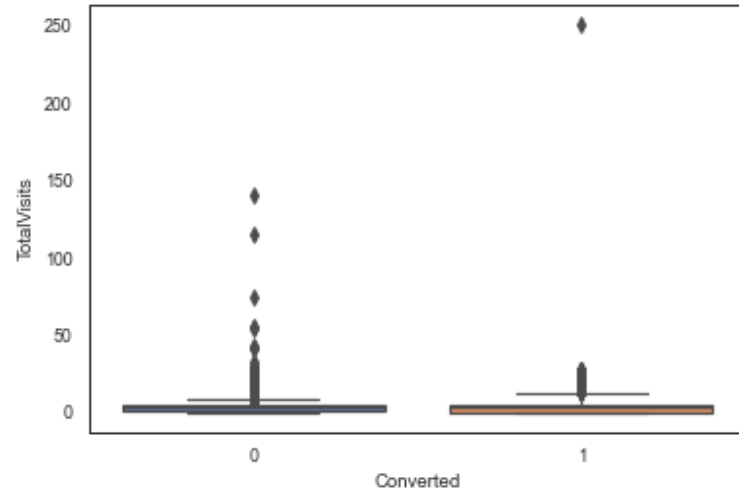
- ▶ Source the data For analysis
- ▶ Reading & Understanding the data
- ▶ Data Cleaning
- ▶ EDA
- ▶ Feature scaling
- ▶ Splitting the data into test & train dataset
- ▶ Prepare the data for modelling
- ▶ Model building
- ▶ Model evaluation-specificity & sensitivity or precision recall
- ▶ Making predictions on the test set

# Data Sourcing, Cleaning & Preparation

- ▶ Read the data from CSV File
- ▶ Outlier treatment
- ▶ Data cleaning -Handling Null Values & removing higher Null values data
- ▶ Removing Redundant columns in the data
- ▶ Imputing Null Values
- ▶ Exploratory data analysis

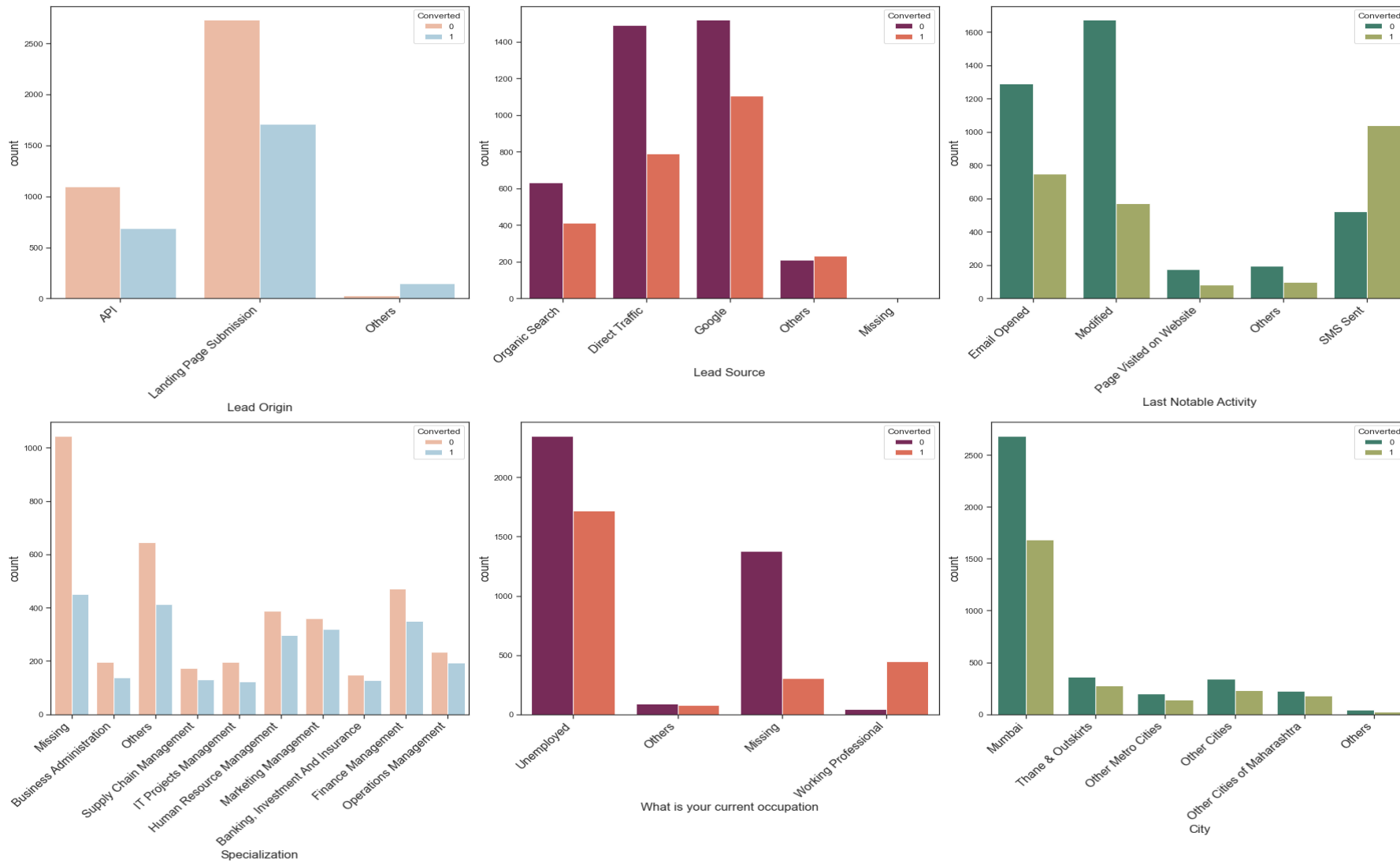
# Outliers

- ▶ Total Visits, Total Time Spent on Website, Page Views Per Visit have outliers



# Data Analysis

## ► Data Analysis of Categorical variables





# Data Preparation

- ▶ Converted Binary variables 'Yes' and 'No' into 0 & 1
- ▶ Created dummy variables for categorical variables.

# Feature Scaling & Splitting Train & Test Sets

- ▶ Feature Scaling of Numeric Data using MinMax Scaler
- ▶ Splitting data into Train & Test Set in 70% and 30%

# Model building

- ▶ Feature Selection using RFE
- ▶ Determined Optimal Model using Logistic Regression
- ▶ Calculated Accuracy ,Sensitivity ,Specificity ,Precision & Recall & evaluate model

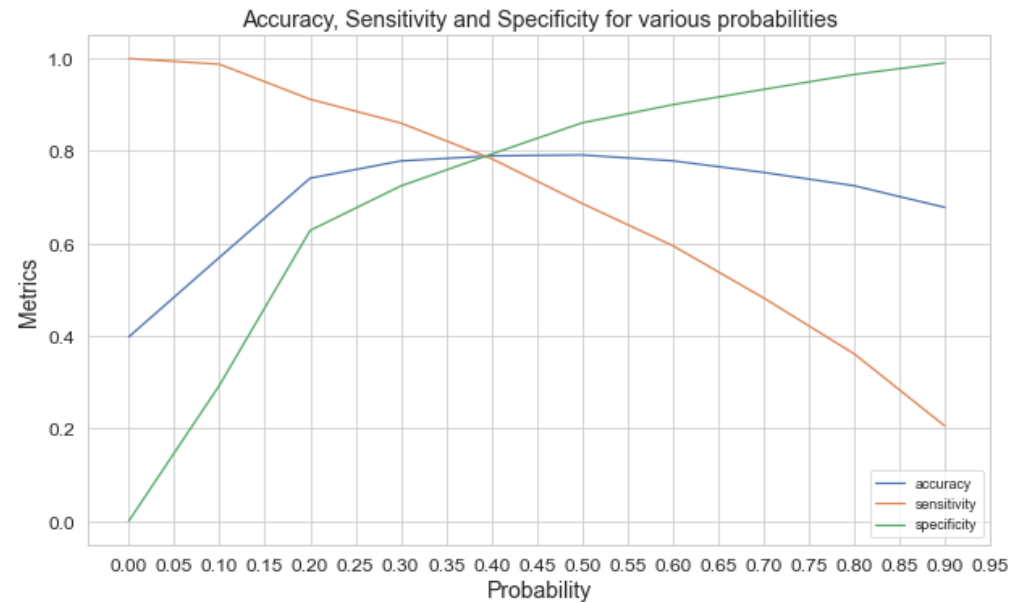
# Variables Impacting the conversion rate

- ▶ Do Not Email
- ▶ TotalVisits
- ▶ Total Time Spent on Website
- ▶ Page Views Per Visit
- ▶ Last Notable Activity\_SMS Sent
- ▶ What is your current occupation\_Others
- ▶ What is your current occupation\_Unemployed
- ▶ What is your current occupation\_Working Professional

# Model Evaluation Accuracy & Sensitivity, Specificity on Train Dataset

Graph depicts an optimal cutoff of 0.39 bases on Accuracy, Sensitivity, Specificity

- ▶ The Accuracy is : 89%
- ▶ The Sensitivity is : 69%
- ▶ The Specificity is : 86%

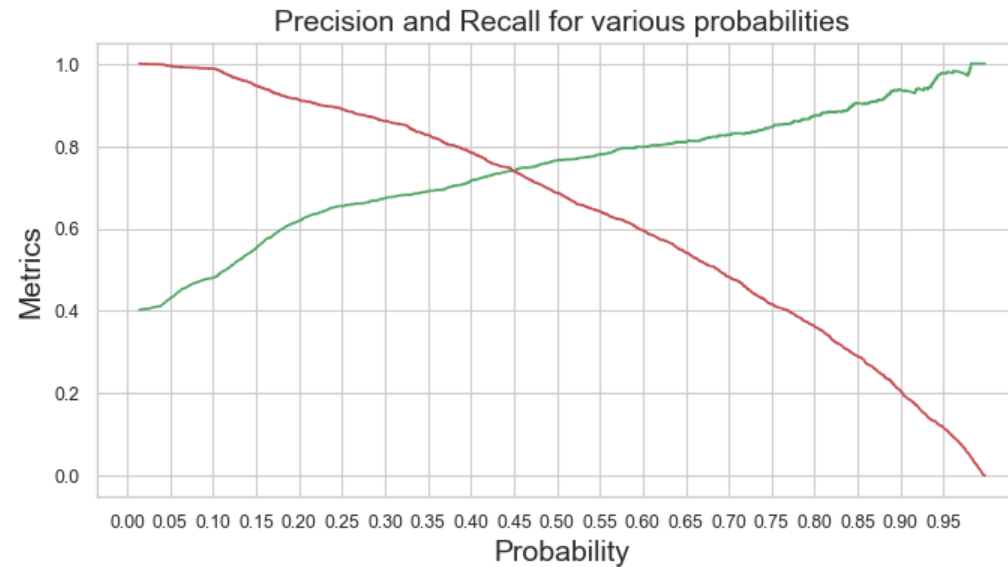


# Model Evaluation Precision & Recall on Train dataset

The graph depicts optimal cutoff of 0.45 based on precision & Recall

The Precision is : 77%

The Recall is : 86%



# Model Evaluation- Sensitivity & Specificity on Test Dataset

- ▶ The Accuracy is : 80%
- ▶ The Sensitivity is : 79%
- ▶ The Specificity is : 81%
- ▶ The Precision is : 73%
- ▶ The Recall is : 81%

# Conclusion

- ▶ While we have checked both sensitivity-specificity as well as Precision & recall metrics, we have considered the optimal cut off based on sensitivity & specificity for calculating the final prediction
- ▶ Accuracy, Sensitivity & specificity values of test set are around 80%,79%,81% which are approximately closer to Values calculated using Trained Data Set
- ▶ Lead Score Calculated for the conversion rate final model on Train & Test dataset is 79% & 80% respectively.
- ▶ Hence, Overall Model seems to be Good



# Summary

- ▶ In order to get a higher lead conversion, first, sort out the best prospects from the leads you have generated. 'Total Visits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.
- ▶ Then, You must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads.