

# Lead Scoring Case Study

**Presented by**

**Bhushan Daga  
Mahua Purkait  
Ranjana Bajpayee**



# Objective

The purpose is to optimize the lead scoring mechanism based on their fit, demographics, buying tendency etc. By implementing explicit & implicit lead scoring modeling with lead point system

# Problem Statement

- ◆ An education company named 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.
- ◆ 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%**. 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

- ◆ 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.
- ◆ There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

# Steps followed

- Source the data for Analysis
- Reading & Understanding the Data and problem statement
- Data Cleaning
- EDA
- Feature scaling
- Splitting data into train and test dataset
- Preparing the data for Modeling
- Model building
- Model evaluation- Specificity & Sensitivity or precision recall
- Making Predictions on the test set
- Check the accuracy and validity of model

# 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 data analysis-approx. Conversion rate is 38%
- ◆ Feature standardization

# Data Preparation

- ◆ Convert Binary variables into 0 & 1
- ◆ Created dummy variables for categorical variables

# Feature Scaling & Splitting Train & Test Sets

- Feature Scaling of Numeric Data
- Splitting data into Train & Test Set

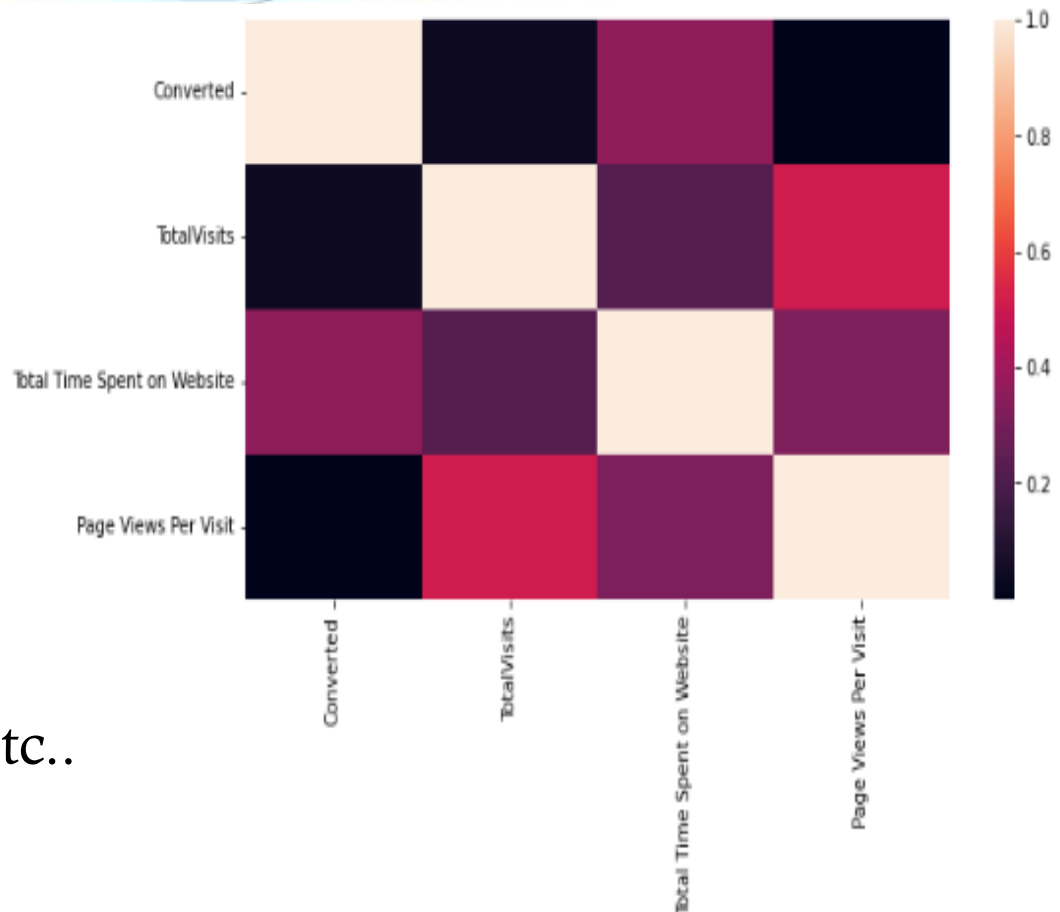


# Model Building

- Feature Selection using RFE
- Determined Optimal Model Using Logistic Regression
- Calculated accuracy, sensitivity, specificity, precision and Recall & evaluate model

# Variables Impacting the Conversion rate

- Total Visits
- Total Time Spent on Website
- Lead Source
- Lead Orgin\_Lead Add Form
- Lead Source
- Do not Email
- Lead Source\_Referral Sites..etc..



# Model Evaluation

## Sensitivity & Specificity on Train Dataset

### Sensitivity & Specificity on Train Dataset

- Accuracy =  $\sim 80\%$
- Sensitivity =  $\sim 82\%$
- Specificity =  $\sim 76\%$

# Model Evaluation

## Precision & Recall on Train Dataset

- 💧 Precision =  $\sim 75\%$

- 💧 Recall =  $\sim 75\%$

# Model Evaluation

## Sensitivity & Specificity on Test Dataset

- Accuracy =  $\sim 81\%$
- Sensitivity =  $\sim 81\%$
- Specificity =  $\sim 80\%$

# Result

- ◆ Accuracy, Sensitivity and Specificity values of training and test set are close to training set
- ◆ Accuracy, Sensitivity & Specificity values of training set are ~80%, ~82%, ~76% respectively.
- ◆ Accuracy, Sensitivity & Specificity values of test are ~81%, ~81%, ~80% respectively
- ◆ Conversion rate for Train & Test Dataset is 82.7% & 80.8% respectively.
- ◆ We have done the prediction in the test set using cut off threshold from sensitivity & specificity.

# Conclusion

- While we checked both sensitivity-specificity as well as precisions & recall metrics, we have considered the optimal cut off based on sensitivity & specificity for calculating the final prediction
- Accuracy, sensitivity and specificity value of test set are around ~81%, ~81%, ~80% 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 approx. ~80%.
- Hence Overall Model seems to be Good

# Summary

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom.

In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, communication ) in order for conversion

Sorting out the 'Total visits, Total Time spent on the website, Pages views Per visit' which contribute most towards getting converted.

Informing about new courses, services, job offers and further higher studies options

A proper plan to chart the needs, Focus on converted leads, holds Q&A sessions to extract the right information you need, make further inquiries and appointments with the leads to determine their intention to join the online courses



A decorative blue curved shape at the top of the slide.

Thank you!