

The left side of the slide features an abstract background with a gradient of soft colors (pink, purple, blue) and a series of overlapping, wavy, translucent lines in red, green, and blue that create a sense of depth and movement.

# **Lead scoring case study**

Presented By: Drashti Patel  
Shivam Pandey  
Sohum Suryavanshi

# 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

1. 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.
2. 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.

# 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

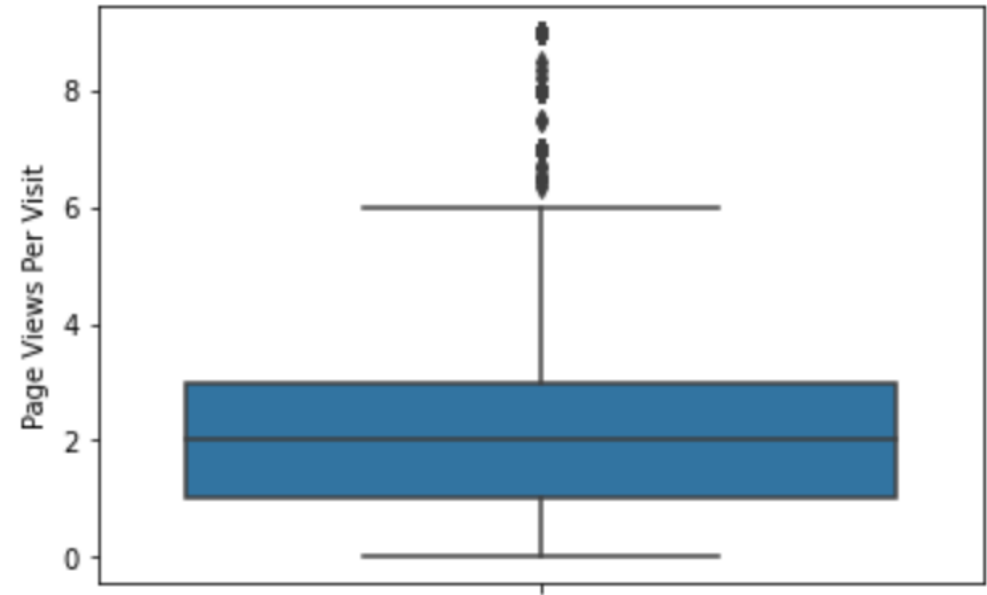
# Data Sourcing, Cleaning & Preparation

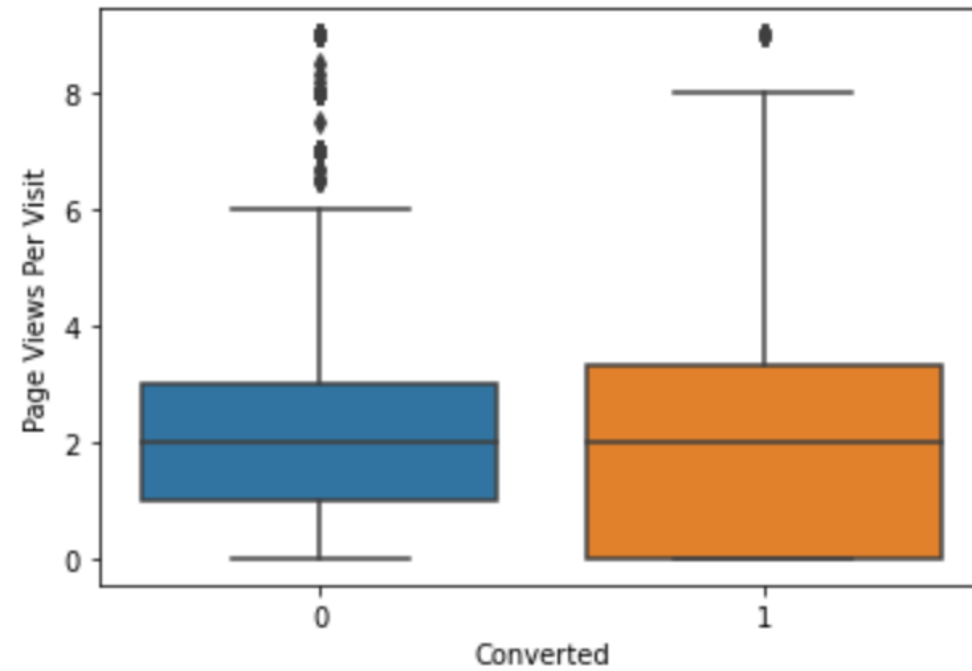
- + Read the data from CSV File
- + Outlier treatment
- + Data cleaning -Handling Null Values & removing higher Null values data
- + Removing Redundant columns in the data
- + Imputing Null Values
- + Exploratory data analysis-approx. Conversion Rate is 38%
- + Feature standardization

# Outliers

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

```
sns.boxplot(y=Lead_data['Page Views Per Visit'])  
plt.show()
```





# Data Analysis



# Data Preparation & Feature Scaling

- + Converted Binary variables into 0 & 1
- + Created dummy variables for categorical variables
- + Feature Scaling of Numeric Data
- + Splitting data into Train & Test Set

# Model Building

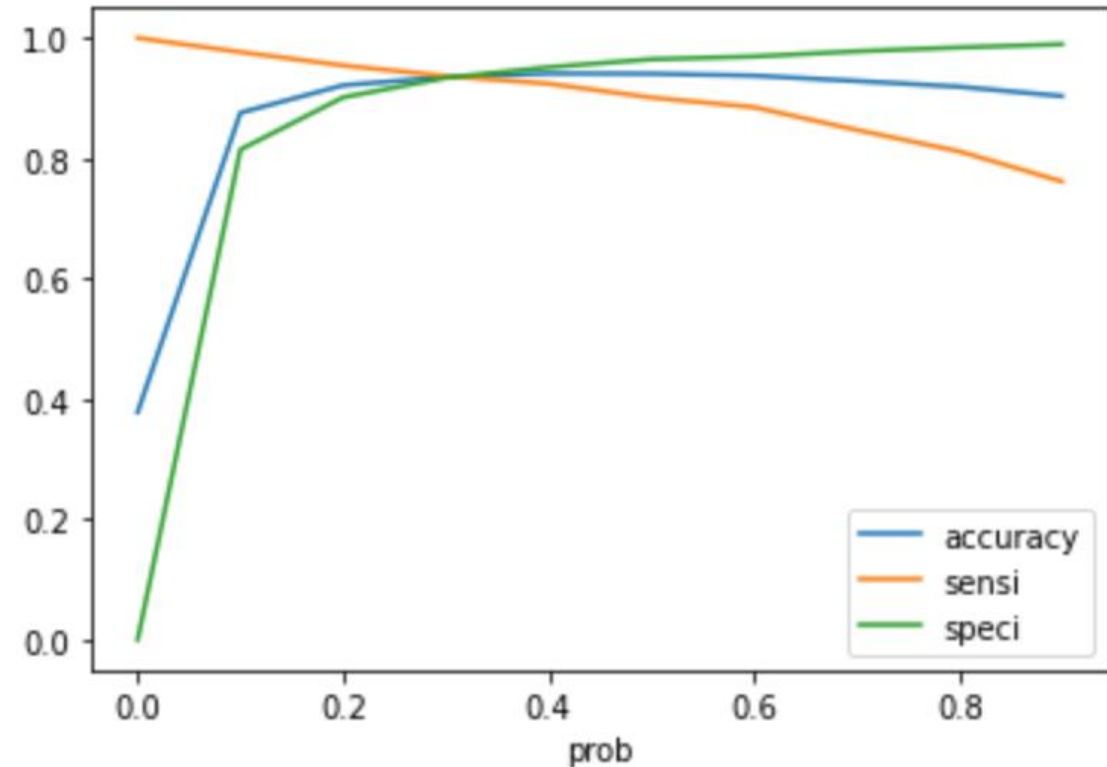
- + Feature Selection using RFE
- + Determined Optimal Model using Logistic Regression
- + Calculated accuracy,sensitivity,specificity,precision & Recall & evaluate model

# Variables Impacting the conversion rate

- + Total Visits
- + Total Time Spent on website
- + Lead Source\_Olark chat
- + Lead Origin\_Lead Add Form
- + Lead Source\_Welingak Website
- + Do Not Email
- + Lead Source \_Referral Sites....etc.

## Model Evaluation- Sensitivity & Specificity on Train Data Set

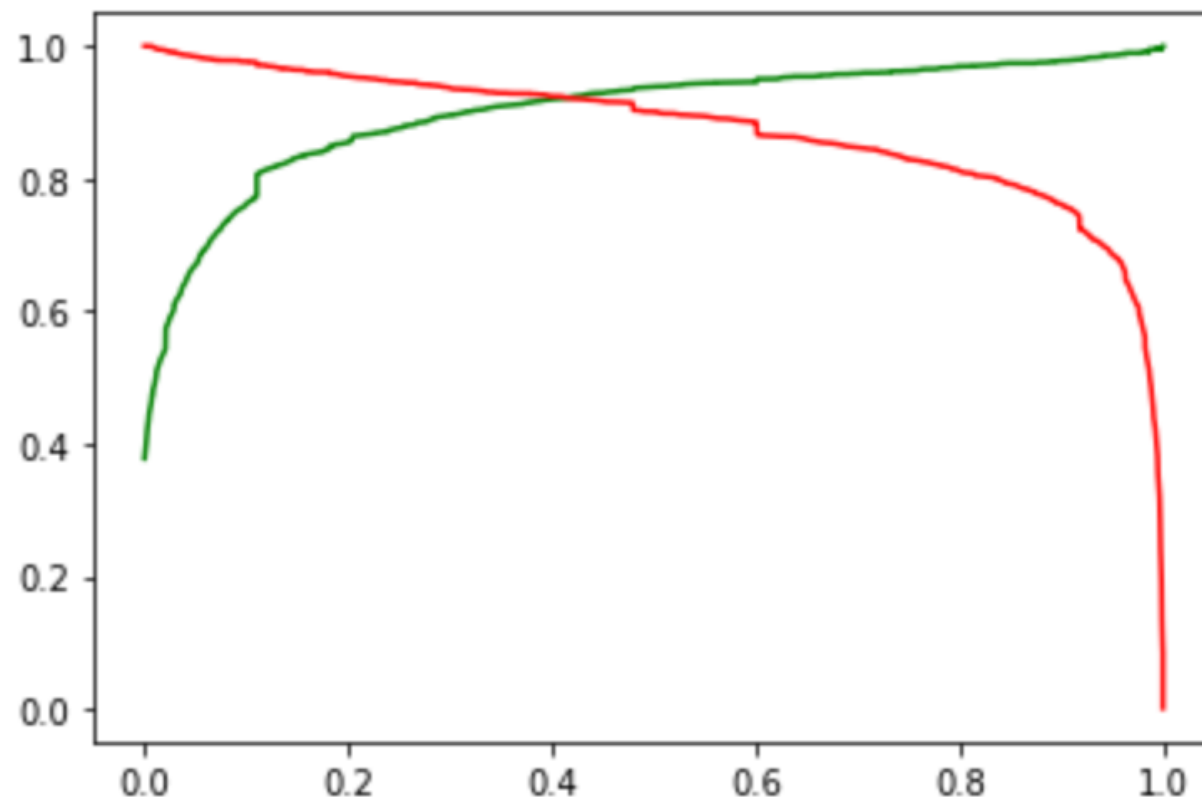
- + Graph depicts an optimal cutoff of 0.37 bases on Accuracy, Sensitivity, Specificity
- + Accuracy = 78%
- + Sensitivity = 82%
- + Specificity = 76%



## Model Evaluation

### Precision & Recall on Train dataset

- + The graph depicts optimal cutoff of 0.42 based on precision & Recall
- + Precision = 79%
- + Recall = 65%



# Model Evaluation

- + Sensitivity & Specificity on Test Dataset
- + Accuracy = 78%
- + Sensitivity = 80.8%
- + Specificity = 76.5%

# Result

- + Accuracy, Sensitivity and Specificity values of training and test set are close to training set
- + Accuracy, Sensitivity and Specificity values of training set are 79%, 82%, 76% Respectively
- + Accuracy, sensitivity & Specificity values of test are 78%, 81%, 76% Respectively
- + Conversion rate for Train & Test Dataset is 82.7% & 80.8% Respectively
- + We have done the prediction on the test set using cut off threshold from sensitivity & specificity metrics

# 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 78%,81%,76% 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 82.7% &80.8% respectively.
- + 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, constantly communicating etc.) 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. A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads. Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.

**Thank you**