# LEAD SCORING CASE STUDY

BY

VIJAY AJ

JANMEJAY

RINI

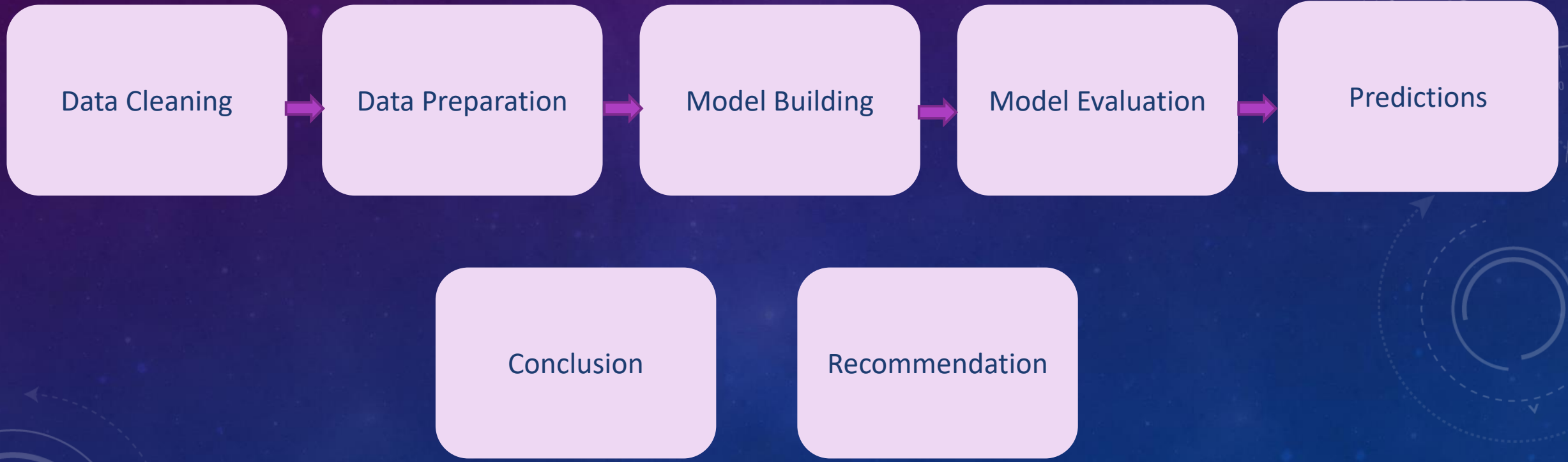
# PROBLEM STATEMENT

- 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%.

# OBJECTIVE

- X Education needs help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a 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%
- 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 lead

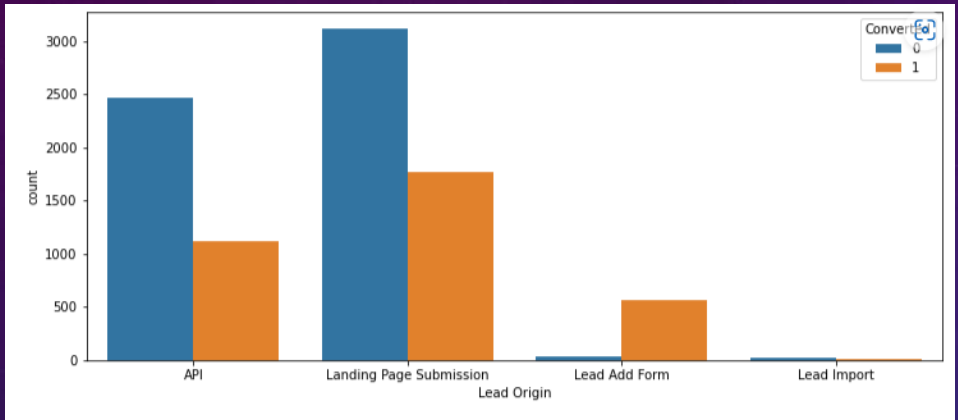
# APPROACH


Data Cleaning
Data Preparation
Model Building
Model Evaluation
Predictions
Conclusion
Recommendation

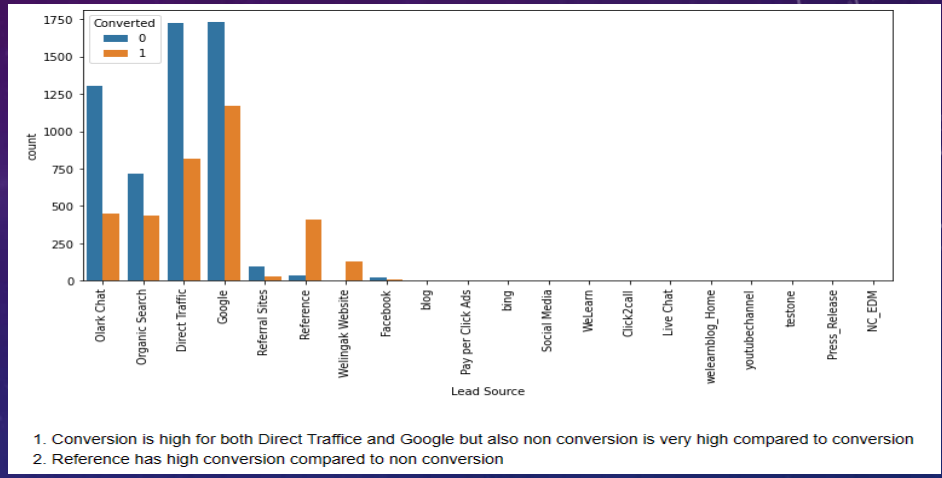
# DATA CLEANING

- Handling Select values in some columns
- Dropping all the columns having missing value greater than 30%
- Imputing null values with respective median/mean/mode values
- Converting negative values to positive values
- Skewed values in the columns were check and droped
- Removing of Unwanted columns

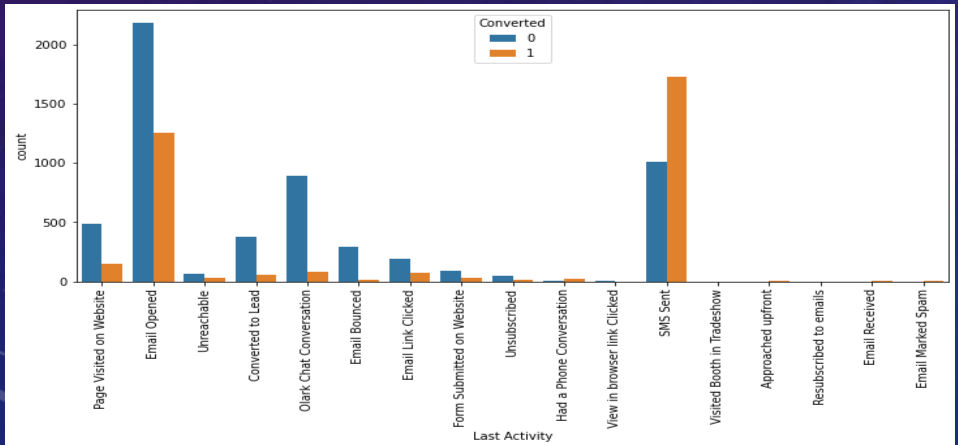
# DATA ANALYSIS



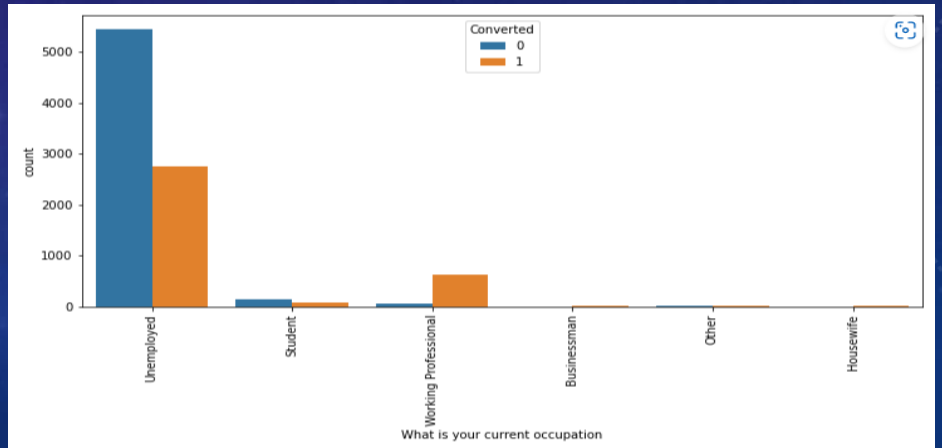

Convertion from Landing page submission is high, But Lead Add From has more conversion rate




1. Conversion is high for both Direct Traffice and Google but also non conversion is very high compared to conversion
2. Reference has high conversion compared to non conversion

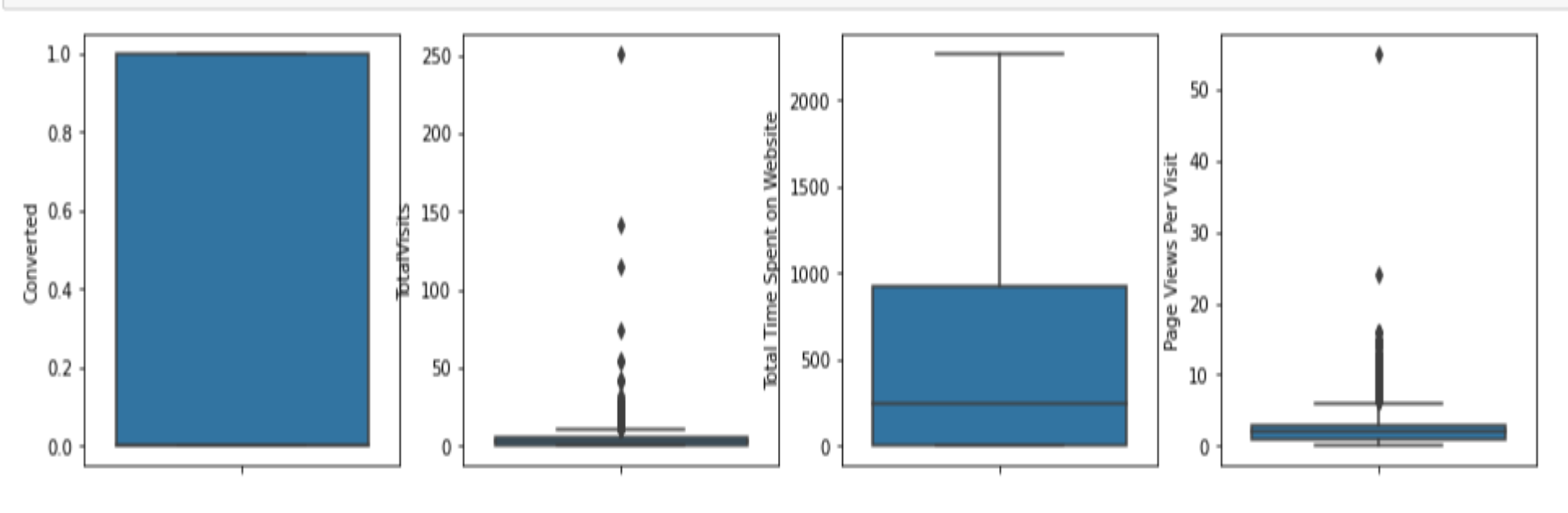



SMS sent has highest conversion




Unemplyed conversion is High, But conversion rate of working professional is high compared to non conversion

# OUTLIER ANALYSIS/TREATMENT




Outliers are observed in Total visits and Page Views per Visit

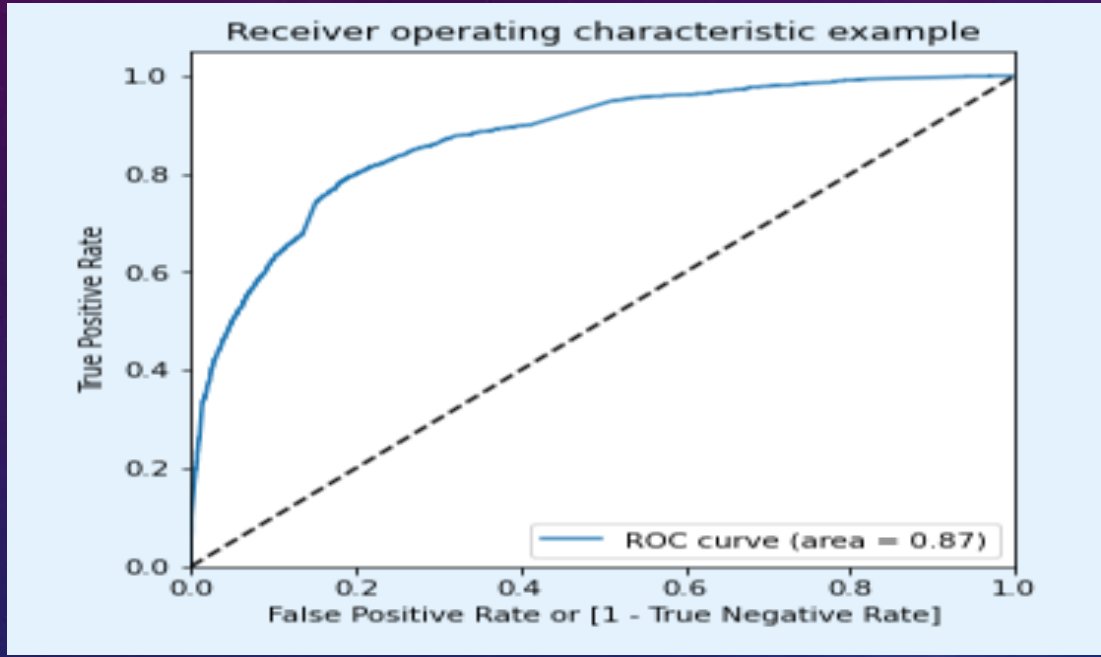
# DATA PREPARATION

- Binary columns are mapped from (Yes/No) to 0/1
- Dummy Variables are created for columns 'Lead Origin', 'Lead Source', 'Last Activity',  'What is your current occupation‘

# MODEL BUILDING

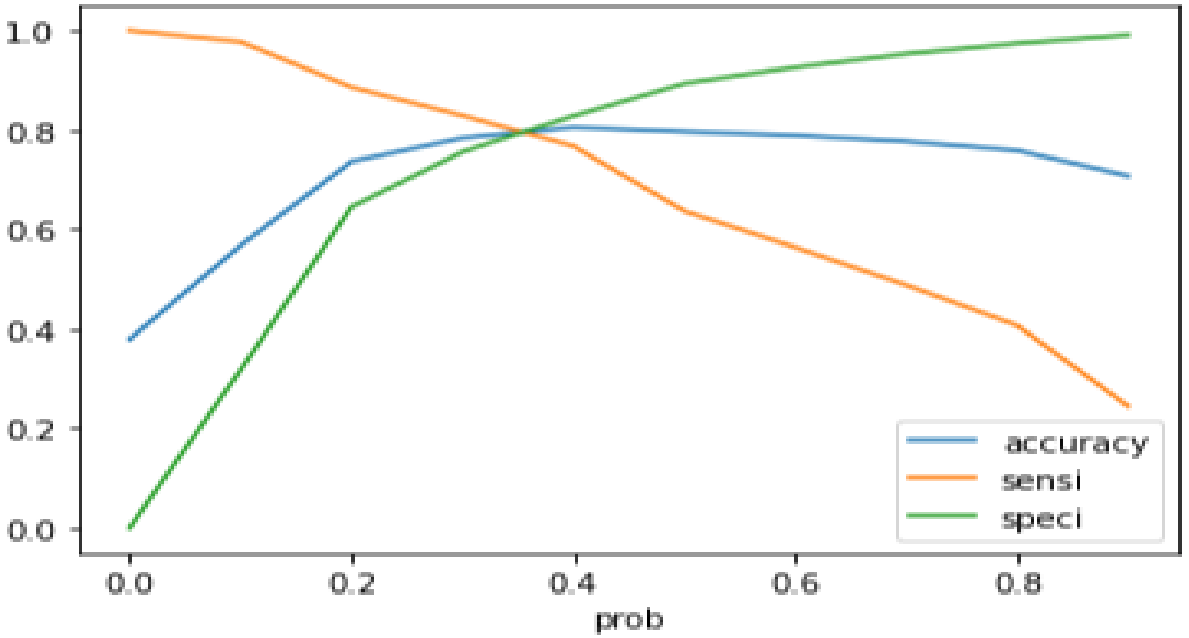
- The data set was divided into Train and Test set (70% - Train set and 30% - Test set)
- Feature scaling is done via Standard Scaler. Check for high correlation variables and remove 2 variables in this step
- Look for Correlations - Lead Source_Reference and Lead Origin_Lead Add Form are highly correlated
- Variable selection is done using **RFE**, in which 15 variables were selected. We check the **p-value** to select the most significant that should be present. In this step we dropped 3 variables with **p-value** greater that threshold i.e. >0.05(4 repetitive steps). Next we checked **VIF** of each variable and found to good.
- Model 4 is the stable model which is achieved after 4 iterations

# PLOTTING ROC CURVE




ROC curve was plotted, where we got ROC curve area coverage around 87%

# OPTIMAL CUTOFF




From the curve, we can take 0.35 as cutoff

# PREDICTION – TRAIN AND TEST DATASET

Train DataSet:

1. Accuracy : 79.93% ≈ 80%
2. Sensitivity : 80.03%
3. Specificity : 79.87% ≈ 80%

Test DataSet:

1. Accuracy : 79.14%
2. Sensitivity : 79.55% ≈ 80%
3. Specificity : 78.90%

# CONCLUSION

- **The target lead conversion rate is around 80%**
- **The Accuracy of the model is around 80%**
- **Accuracy, Sensitivity and Specificity of both Test and Train have closer values**
- **Important variables which positively contributed for lead conversions are:**
    1. What is your current occupation_Working Professional
    2. Last Activity_Had a Phone Conversation
    3. Total Time Spent on Website
    4. Last Activity_SMS Sent

# RECOMMENDATION

**Company should focus on the below points**

- **Working professionals should be targeted.**
- **Target the website as more time is spent on website are more likely to get converted. Allocate some budget on the website.**
- **The company should focus on sending Email, SMS, incentives and on the ads in website.**
- **It’s good to send SMS along with phone calls**