# Lead Score Case Study

Created By :
Aditya Sharma
Naresh Tinnaluri

# Index

1. Business Problem Statement
2. Analysis Approach using EDA
3. Model Building and Evaluation
4. Prediction on test set and Evaluation
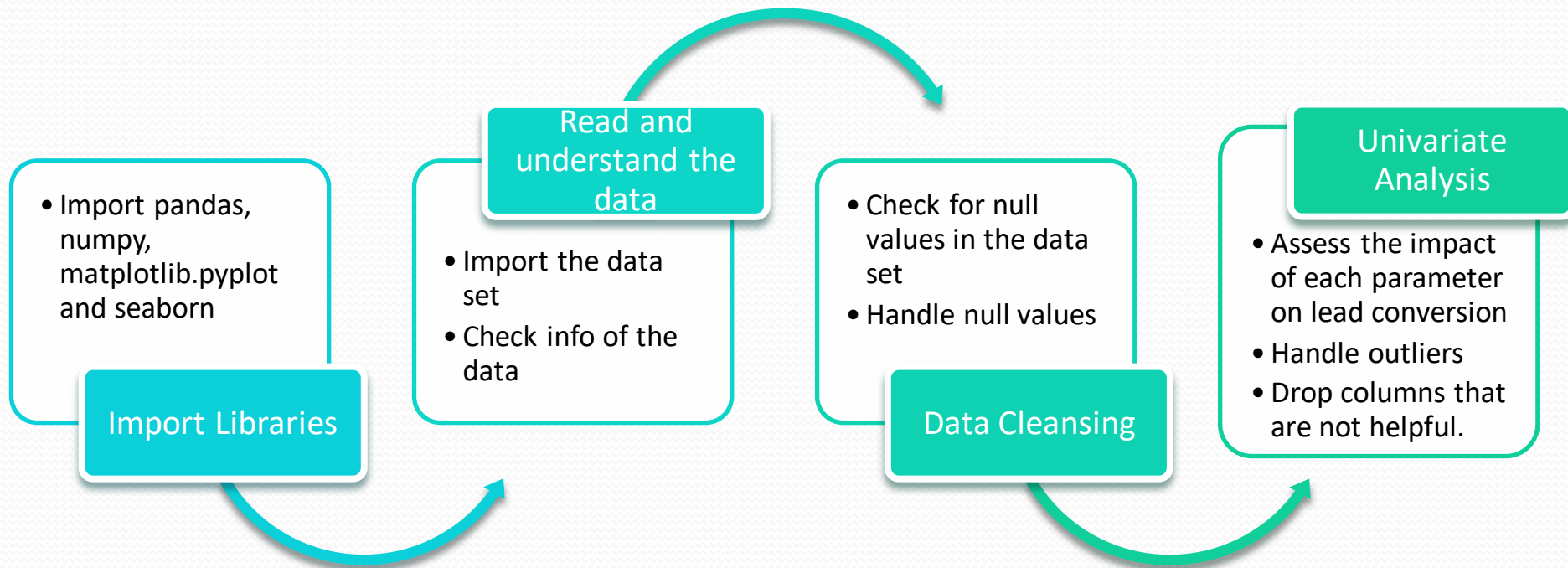5. Inferences from the data set
6. How this model helps?

# Business Problem Statement :

An education company named X Education sell online courses to industory professionals. The company markets these courses through different marketing channels like marketing websites, search engine and referrals and acquire the leads from these channels. 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%.
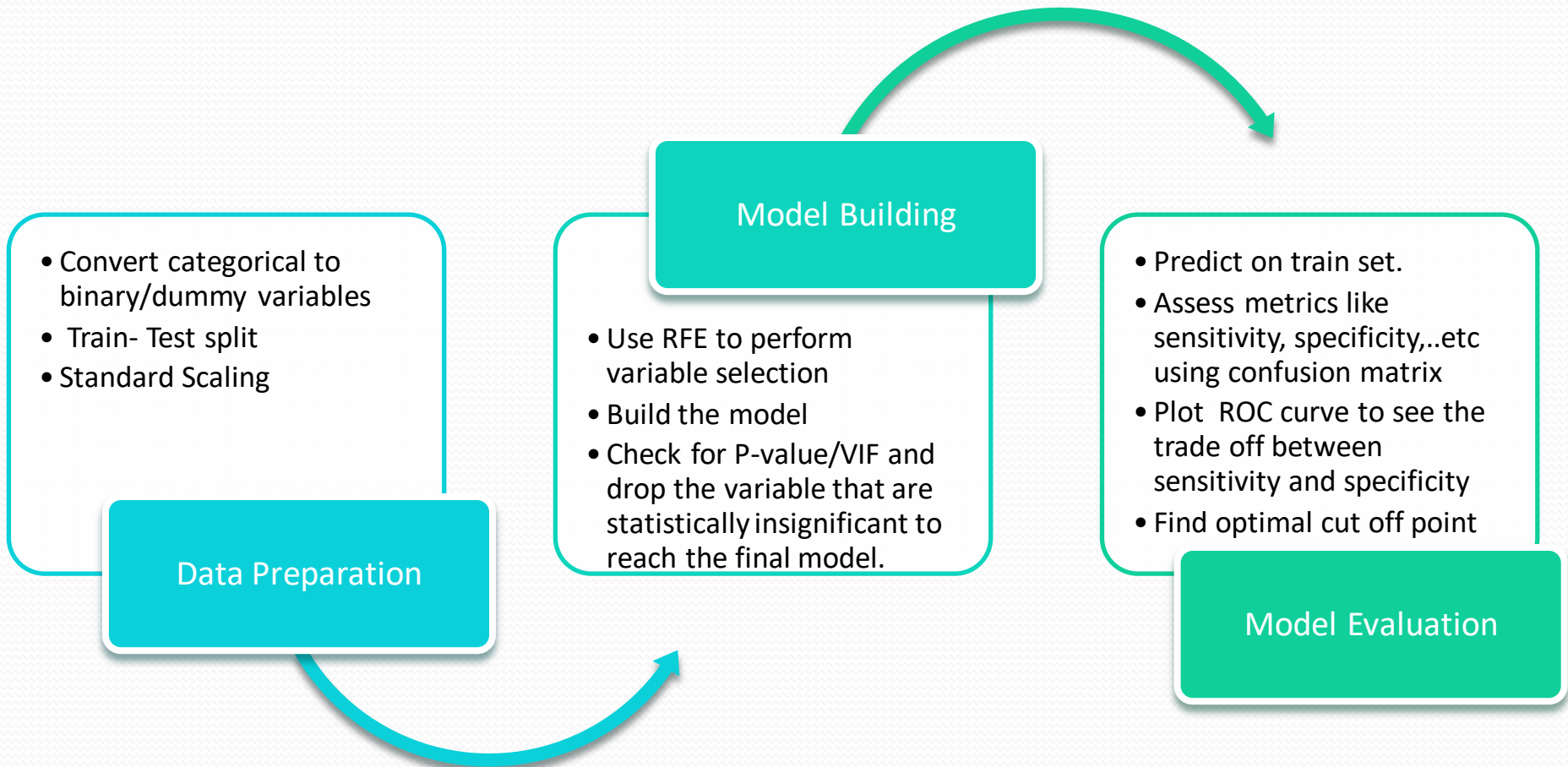
Although the company gets lots of leads but the conversion of the leads are very poor, therefore the company wishes to identify the Hot Leads and focus more on communicating with potential leads rather than making calls to everyone.

# Analysis Approach using EDA :

**Import Libraries**
- Import pandas, numpy, matplotlib.pyplot and seaborn

**Read and understand the data**
- Import the data set
- Check info of the data

**Data Cleansing**
- Check for null values in the data set
- Handle null values

**Univariate Analysis**
- Assess the impact of each parameter on lead conversion
- Handle outliers
- Drop columns that are not helpful.

# Model Building and Evaluation Approach:

**Data Preparation**
- Convert categorical to binary/dummy variables
- Train- Test split
- Standard Scaling

**Model Building**
- Use RFE to perform variable selection
- Build the model
- Check for P-value/VIF and drop the variable that are statistically insignificant to reach the final model.

**Model Evaluation**
- Predict on train set.
- Assess metrics like sensitivity, specificity,..etc using confusion matrix
- Plot ROC curve to see the trade off between sensitivity and specificity
- Find optimal cut off point

# Prediction on Test Set and Evaluation Approach:

**Data Preparation**
- Scale the Numeric variable using the trained scalar
- Keep the variables in test set that were statistically significant on trained set.

**Predictions**
- Predict probabilities using the model that were trained on training set.
- Label each prospect id as converted/not-converted using optimal cut off point

**Evaluation**
- Assess metrics like sensitivity, specificity,..etc using confusion matrix
- Calculate the score for each lead using formula – probability* 100

# Inferences about the data:

Most of the variable (categorical/non-categorical) are impacting the conversion factor of the leads, but most importantly following are the variable that plays a significant role in conversion of leads.

i.    Do not Email
ii.   Total time spent on website
iii.  Lead origin
iv.   Lead Source
v.    What is your current occupation
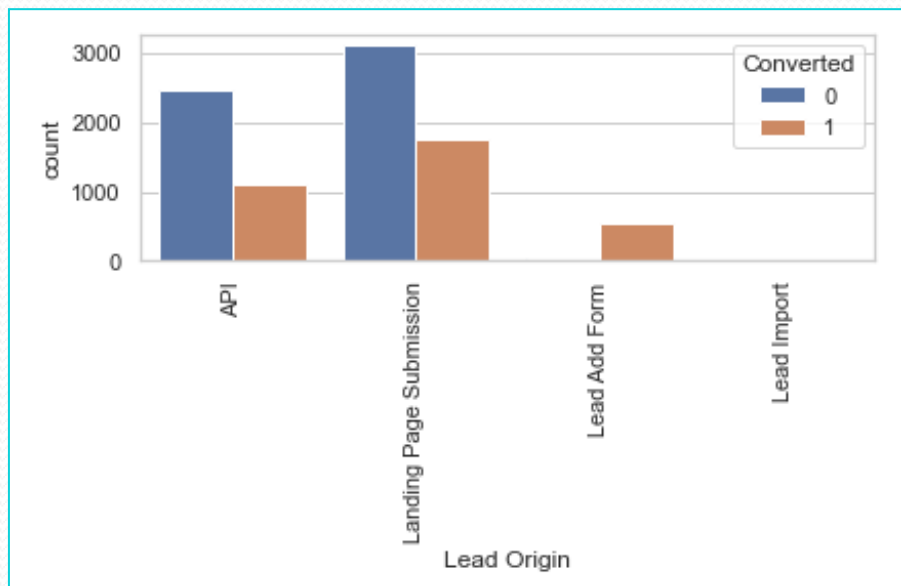
Here, we can see that leads converted more in 'Landing Page Submission' then through 'API' and finally 'Lead Add Form', however leads from lead import is very less in number.

Google and Direct Traffic creates more Converted Leads.
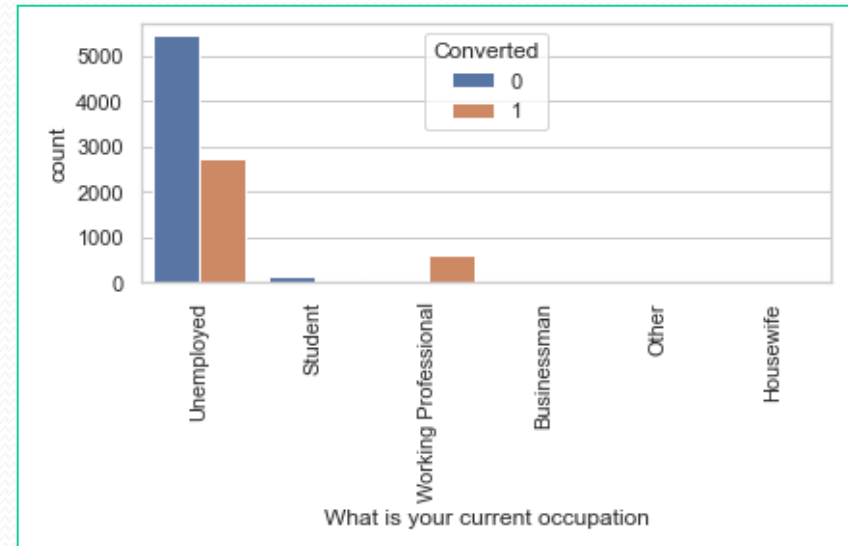
Reference showing significant in converting Leads more
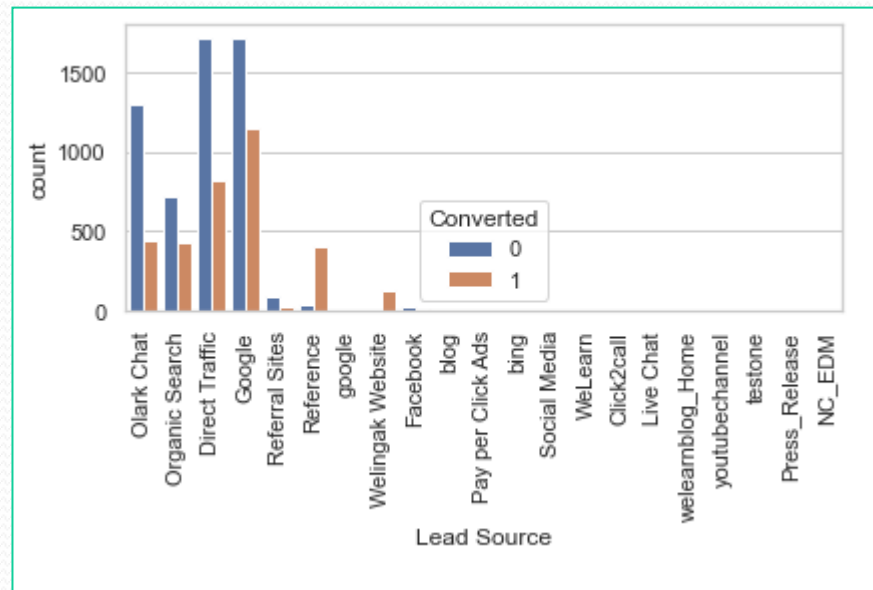
'No' showing significant number in converting Leads when compared to 'Yes'.



working professionals are high rate of joining in the course

Unemployeed people tend to join in, it has 50 percent chance of Leads

# How this model helps?

This model will help sales team to identify the potential lead by looking at the score assigned to the lead.

This way the chances of reaching out to the right prospect would be high and therefore the chances of conversion would be higher.

It will help Sales team and the company to reach their targeted numbers.

| Prospect ID | Converted | Converted_Prob | final_predicted | Score |
|---|---|---|---|---|
| 3271 | 0 | 0.152074 | 0 | 15.21 |
| 1490 | 1 | 0.980336 | 1 | 98.03 |
| 7936 | 0 | 0.131076 | 0 | 13.11 |
| 4216 | 0 | 0.888338 | 1 | 88.83 |
| 3823 | 0 | 0.153965 | 0 | 15.40 |

*Sample records with lead score*

# Thank You