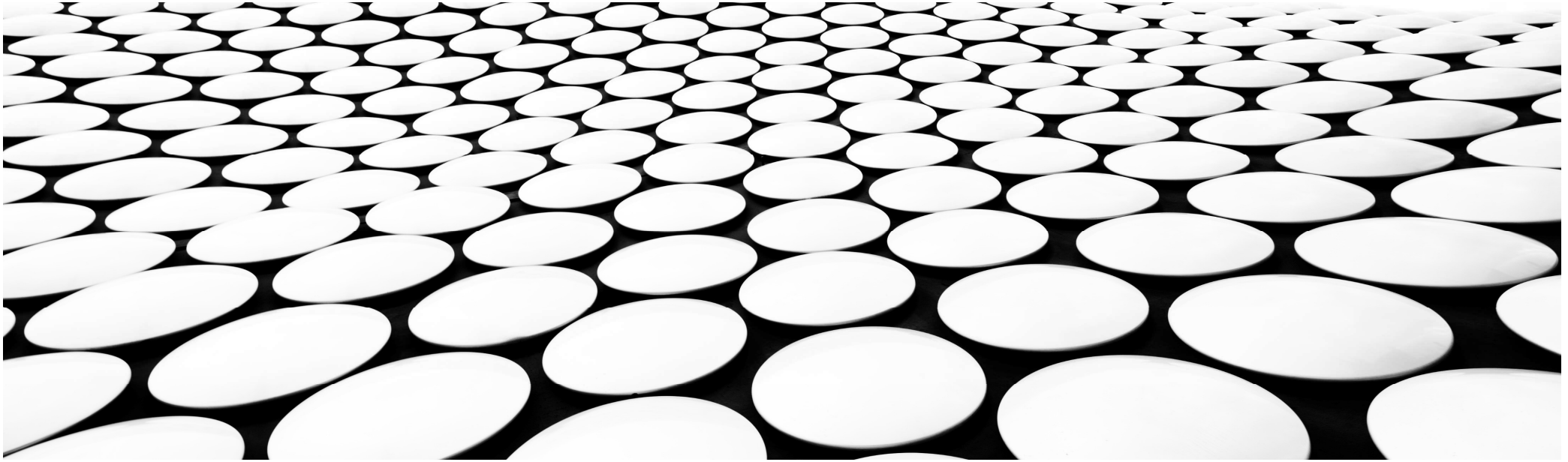
LEAD SCORING CASE STUDAY

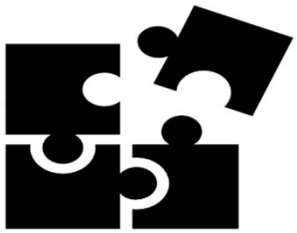
PREPARED BY

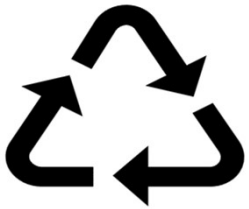
KRUNAL GUJARATHI

UTKARSH RAJ



PROBLEM STATEMENT





- An education company named X Education sells online courses to industry professionals.
- The company markets its courses with several ways to get the leads. But the typical lead conversion rate at X education is around 30%.
- 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.

ASSUMPTIONS IN ANALYSIS

- All the select values considered as not filled and imputed with nulls.
- Wherever missing values are more than 40% are considered as not useful attributes and dropped from the analysis.
- All the Management related specializations are clubbed to the category Management.
- All the unavailable countries are imputed with India as the most of the leads are coming from India only.
- Highly correlated values are considered as non-useful in analysis and dropped.

APPROACH OVERALL

1) Data Cleaning

- Handle the “Select” values and other null values
  - Drop high % missing values columns
  - Check unique categories in the categorical column and remove high % unique value columns
  - Check the rows having 70% missing values in them and remove

2) EDA

  - Perform univariate analysis to check the trends of the conversion rate
  - Check for the outliers and removed the rows which are having outliers in them

APPROACH OVERALL

3) Data Preparation

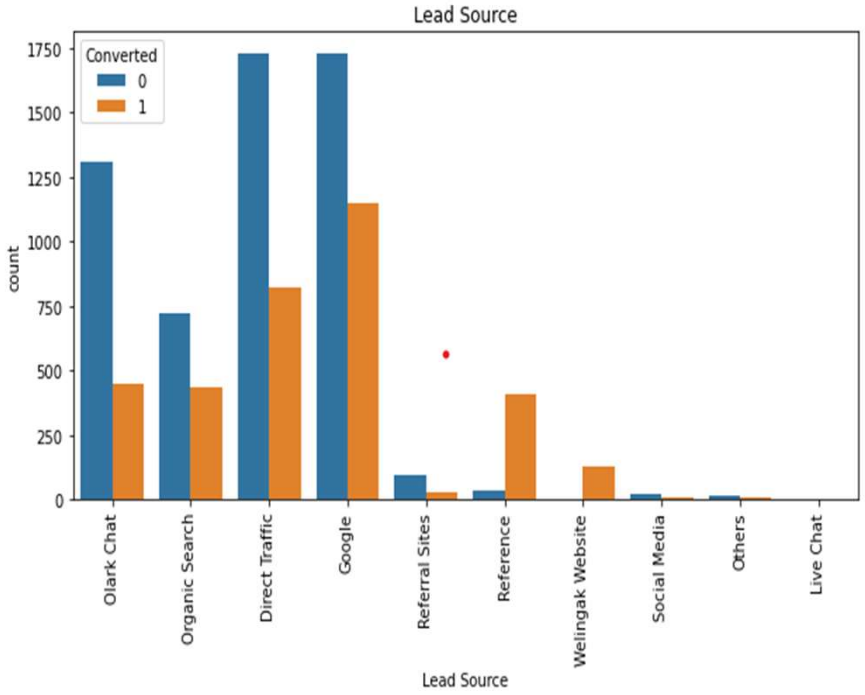
- Dummy variables creation for all the categorical attributes
- Performed test-train split
- Performed Scaling for easier analysis

2) Modelling

- Used RFE to select the most authentic variables
- Checked p-value and VIF for variable removal
- Checked the model performance over train and test data using different matrixes(Confusion matrix, Sensitivity, Specificity, Recall etc.)
- Generated the score variable for sales team to filter out “Hot Lead” data

SOME ANALYSIS BASED ON THE UNIVARIATE ANALYSIS

- - Most of the leads are coming from Google, Olark chat and Direct traffic.
- Lead conversion rate from Olark chat is very low.
- Lead conversion rate for Direct Traffic and Google are Average.
- Lead conversion rates for Reference and Welingak website are very high.
- Other categories from the above mentioned are low in both traffic and conversion.

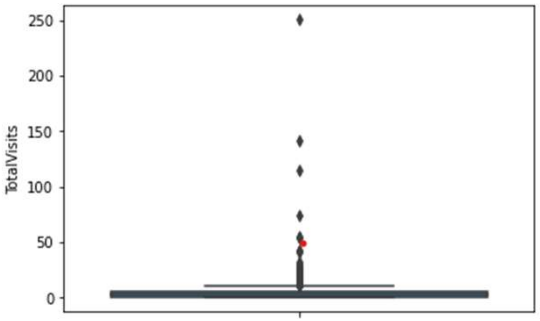


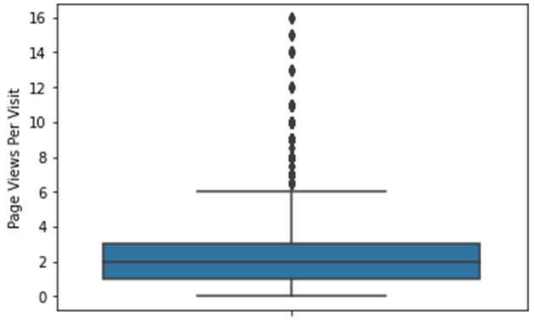
OUTLIER TREATMENT

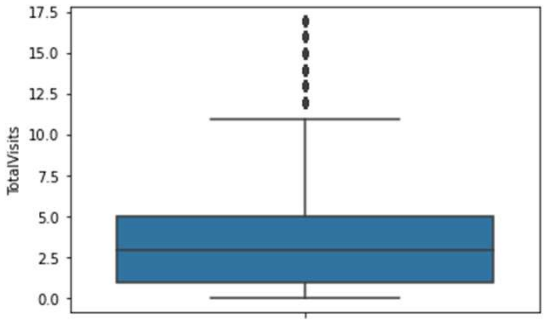
• There were outlier available for numeric variables Total Visits and Page views per Visit.

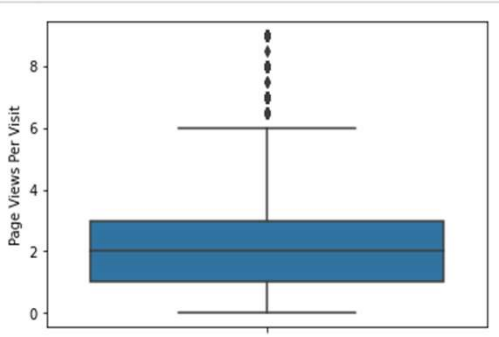
• Check the same using box plot and perform quartile for both the variables.

• Based on the same removed 1% data as an outliers from both the attributes.







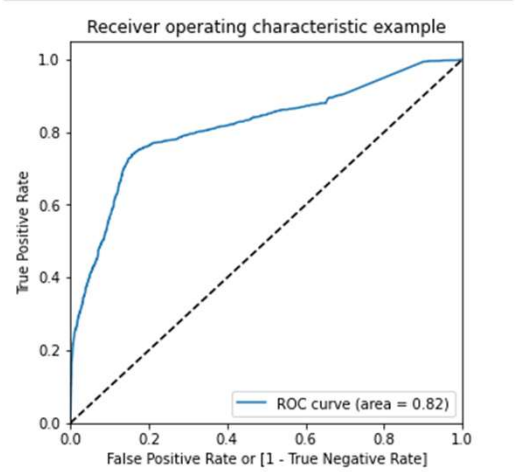


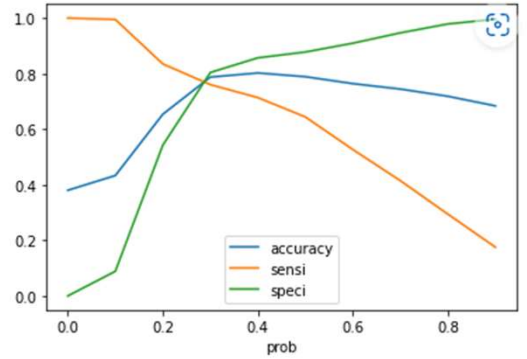
ROC CURVE AND ACCURACY-SENSITIVITY-SPECIFICITY CUT-OFF

• For Logistic regression ROC curve value should be maximum near to 1.

• In our model it is obtained as 0.82, which is good ROC value for the model.

• Probability cutoff curve for Accuracy vs Sensitivity vs Specificity obtained in well format and in optimal range. (Figure-2)





RESULTS USING THE MODEL AND OBSERVATIONS

• Accuracy score on train and test data are in optimal range 79% and 77% respectively.

• Sensitivity of the model on train data and test data is 68% and 76%. Need some improvement but still it is in the acceptance range.

• Other matrix are well in proper range.

• Lead score has been created for all the leads, sales team can filter 3 types of data for doing phone calls and convert the leads.
1) >=65% = “Hot Leads”,
2) >=40% and <65% = “Medium Leads”
3) <40% = “Cold Leads”

→ So, overall it is a good model and sales team can use it further to make phone calls to the highest scoring leads to increase the sales and improve the business goal.

THANK YOU