

A Case Study on Lead Score

Created by:

Bhargav R Kulkarni

Prabhat Kumar

Problem Statement

- ▶ Analysis is done for X education in order to find ways to get more professional to join online courses.
- ▶ The problem is there are multiple leads in a day and their conversion rate is poor , we have to enhance identify most potential leads.

Approach

- ▶ Data Cleaning
- ▶ EDA and visualization
- ▶ Dummy Variable and Splitting Data
- ▶ Model Building
- ▶ Model Evaluation
- ▶ Prediction

Data Cleaning

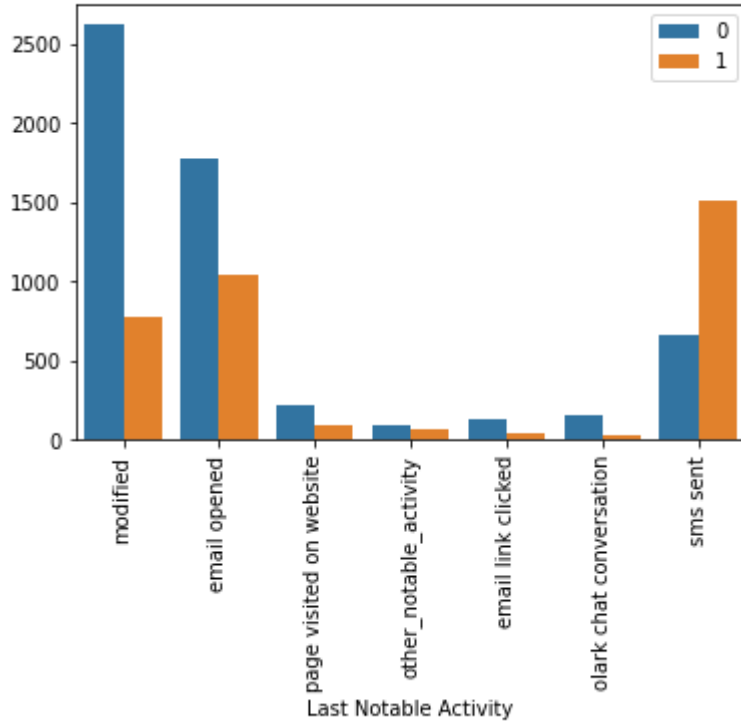
- ▶ Imputation of null values
- ▶ Removal of columns having more than 30% null values
- ▶ Converting some columns data types to numeric from other data types
- ▶ Dropping some redundant variables

EDA

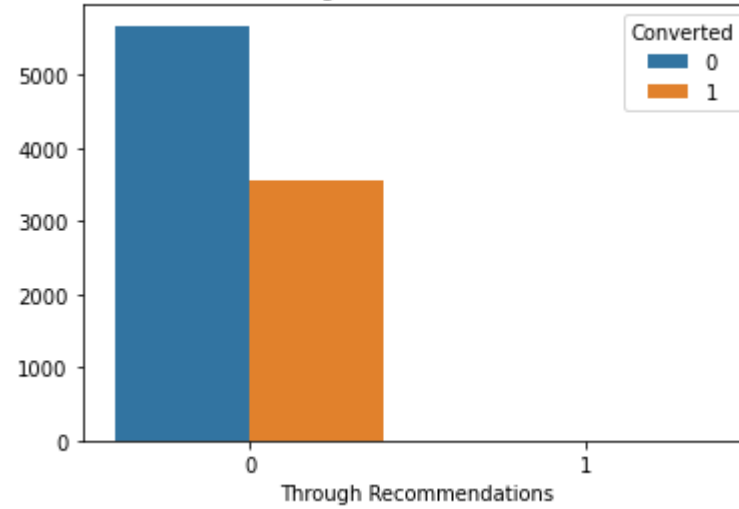
- ▶ Have performed a quick check to visualize condition of Data through different plots .
- ▶ Dropped unnecessary columns.
- ▶ Checking for correlation metrics using annotated heatmap

Visualization

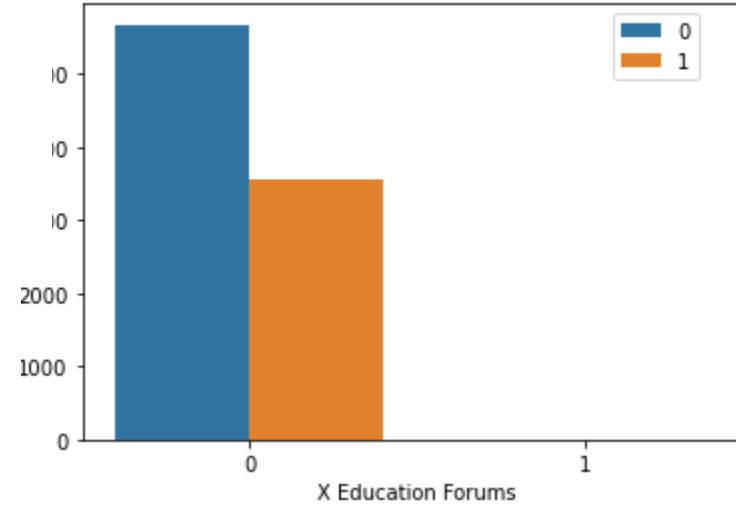
Last Notable Activity



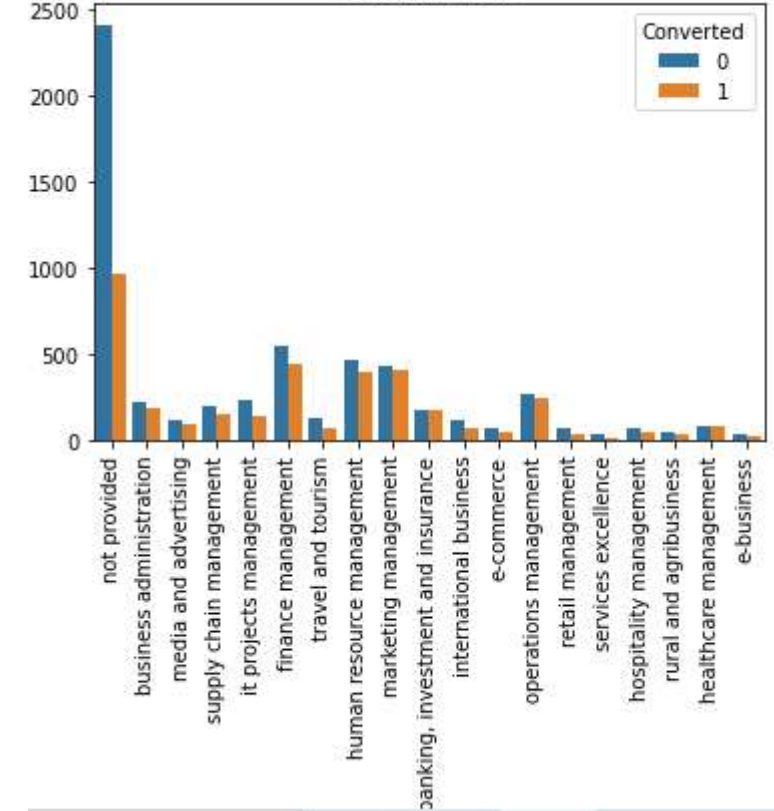
Through Recommendations

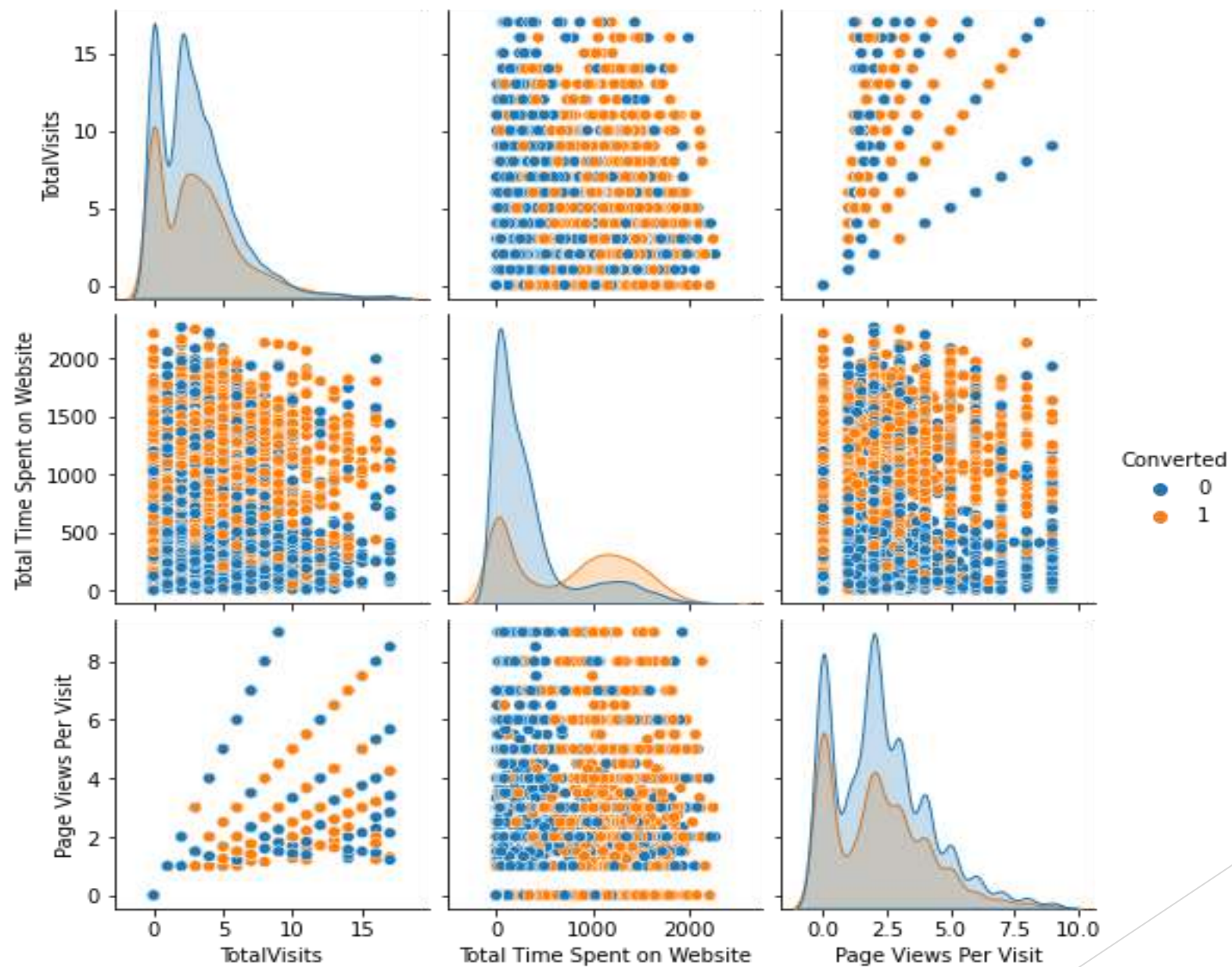


X Education Forums



Specialization





Dummy variables

- ▶ Numerical variables are normalized
- ▶ Dummy variables are created for categorical Type variables.

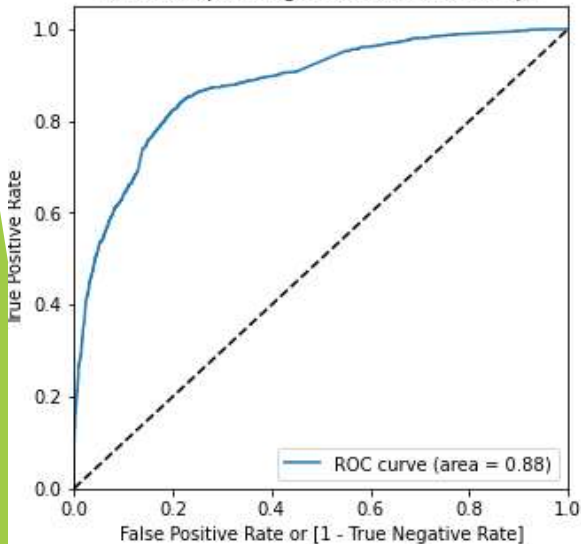
Model Building

- ▶ Splitting the data into Training and Testing sets
- ▶ The first basic step for regression is performing a train-test split in ratio of 70 and 30.
- ▶ Create a first model including all the variables
- ▶ Use RFE for feature selection
- ▶ Running RFE with 15 variables as output
- ▶ Building model by removing the variable whose 'p-value' is greater than 0.05 and 'variance inflation factor' (VIF) value is greater than 5.
- ▶ Prediction on test Data Set

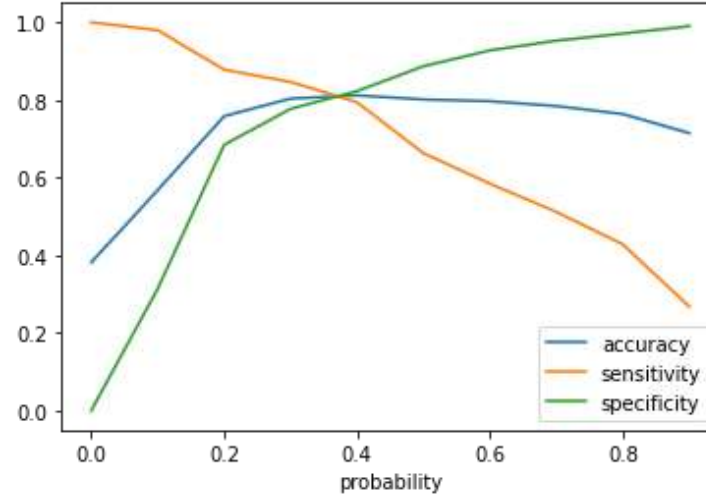
ROC Curve

- ▶ Finding optimal cut-off point
- ▶ Optimal cut off probability is that probability where we get balanced sensitivity and specificity
- ▶ We consider the optimal cutoff from the precision_recall_curve displayed below

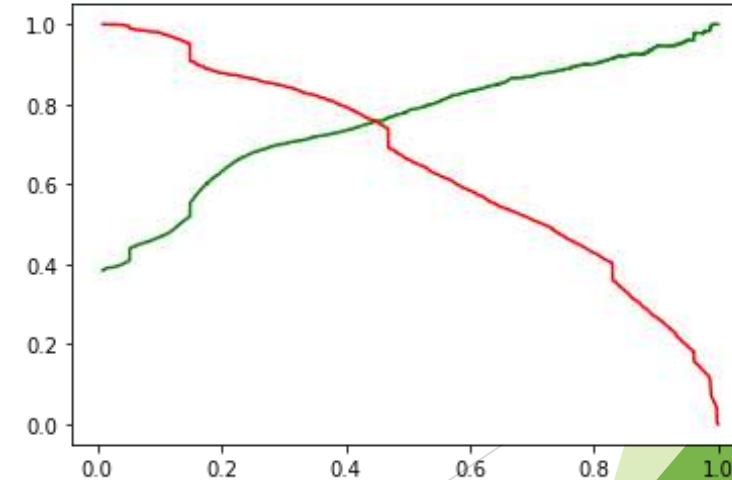
Receiver operating characteristic example



probability vs accuracy,sensitivity,specificity



precision_recall_curve



Conclusion

Valuable Insights -

- ▶ The Accuracy, Precision and Recall score we got from test set in acceptable range.
- ▶ We have high recall score than precision score which we were exactly looking for.
- ▶ In business terms, this model has an ability to adjust with the company's requirements in coming future.
- ▶ This concludes that the model is in stable state.
- ▶ Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :
 - Total Time Spent on Website
 - Lead Origin_Lead Add Form
 - Lead Source_welingak website
 - What is your current occupation_Working Professional
 - Last Notable Activity_other_notable_activity
 - Last Notable Activity_sms sent

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.