Lead score case study

Logistic Regression

Md Taushif Alam

PROBLEM STATEMENT

- X Education is an organization which provides online courses for Industry professional. The company marks its courses are more popular on several website like Google.
- X Education wants to select most promising leads that can be converted to paying customers.
- The company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion. Leads come through numerous modes like advertisement on website ,Google ,searches etc.
- The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interest in taking the course.

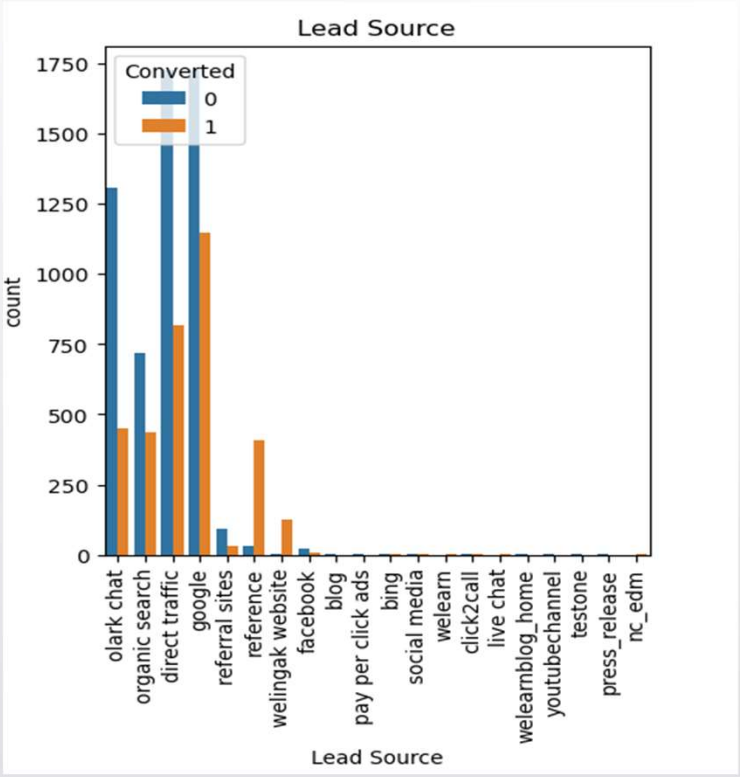
BUISNESS GOALS

- The company requires a model to be built a most promising leads.
- Leads score to be indicates .The higher the lead score the more promising the lead to get converted, the lower it is the lesser the chance of conversion.
- The model to be built in lead conversion rate around 80% or more

STRATEGY

- Import data
- Clean and prepare the data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Create dummy variables
- Scaling the features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each models
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model

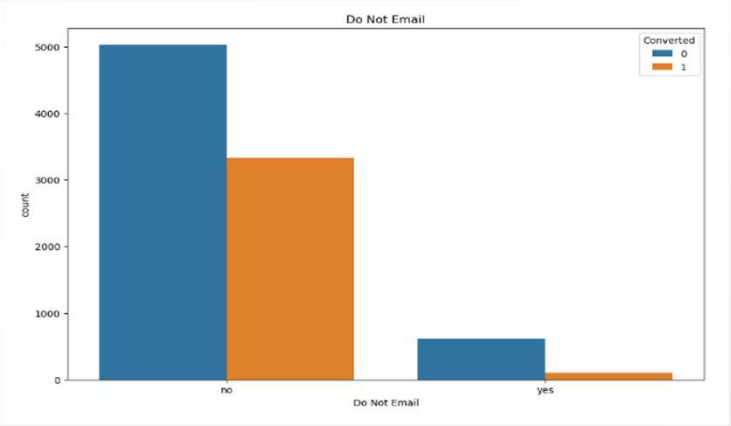
Exploratory data analysis



LEAD SOURCE VS CONVERTED

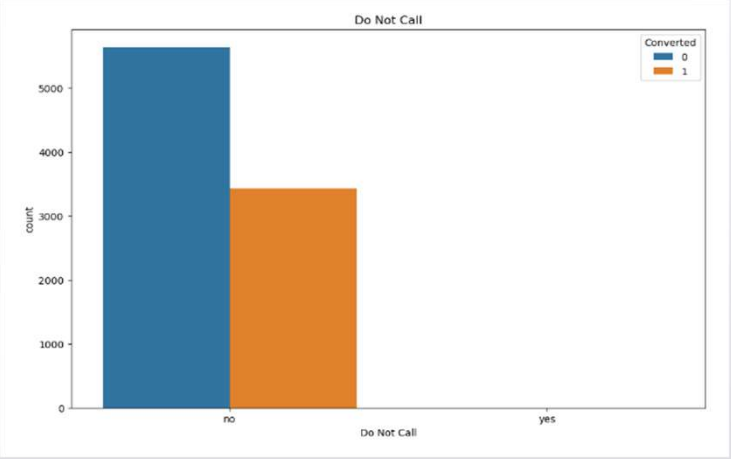
Google searches has high conversion compare to other models, whilst reference has high conversion rate

Exploratory data analysis



DO NOT Email VS CONVERTED

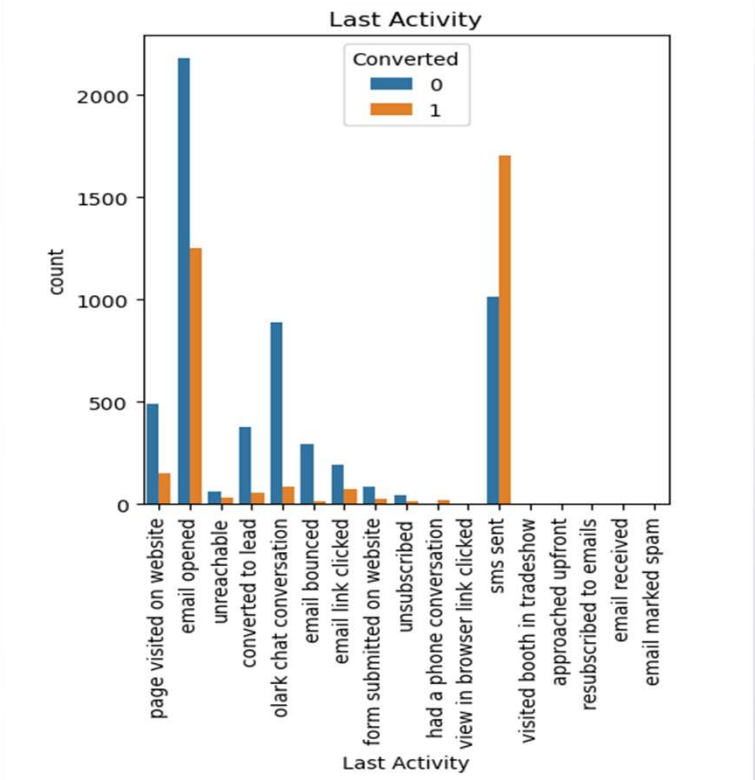
Most leads prefer not to informed through emails



DO NOT CALL VS CONVERTED

Most leads prefer not to informed through Calls

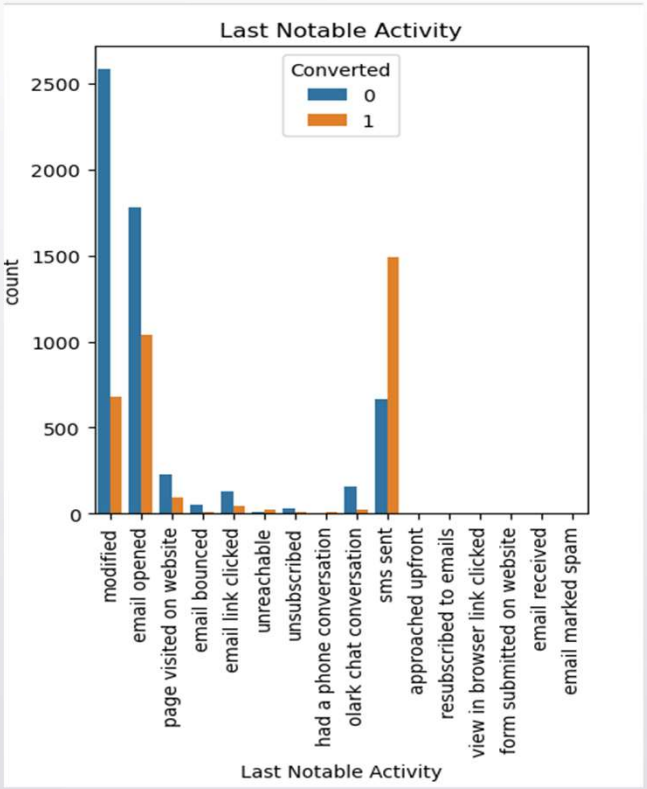
Exploratory data analysis



Last Activity VS CONVERTED

SMS and E-mails has shown promising method to getting higher confirmed leads.

Exploratory data analysis



LAST NOTABLE ACTIVITY VS CONVERTED

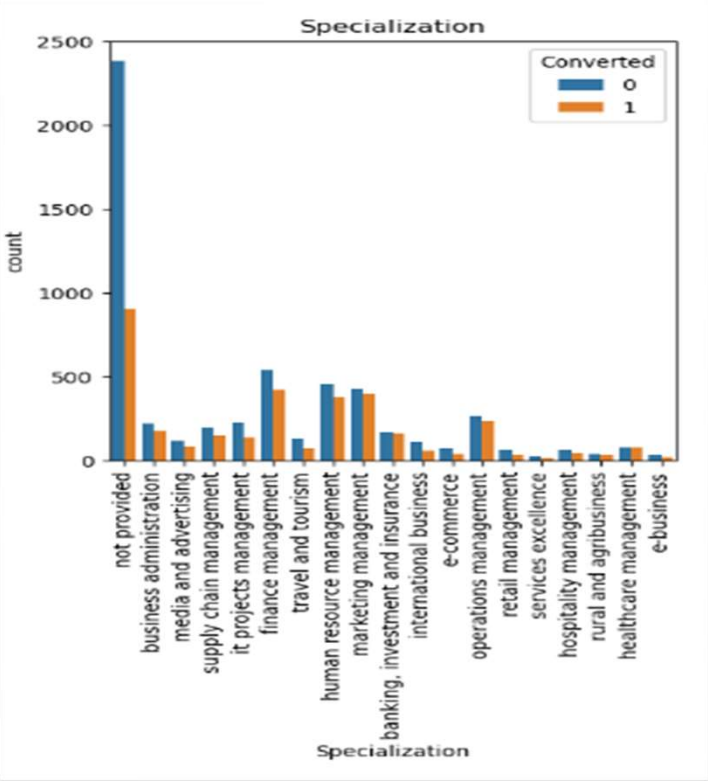
Most leads are converted through emails and SMS.

Exploratory data analysis

A FREE COPY OF MASTERING THE INTERVIEW VS CONVERTED

Leads prefer less copies of interviews.

Exploratory data analysis



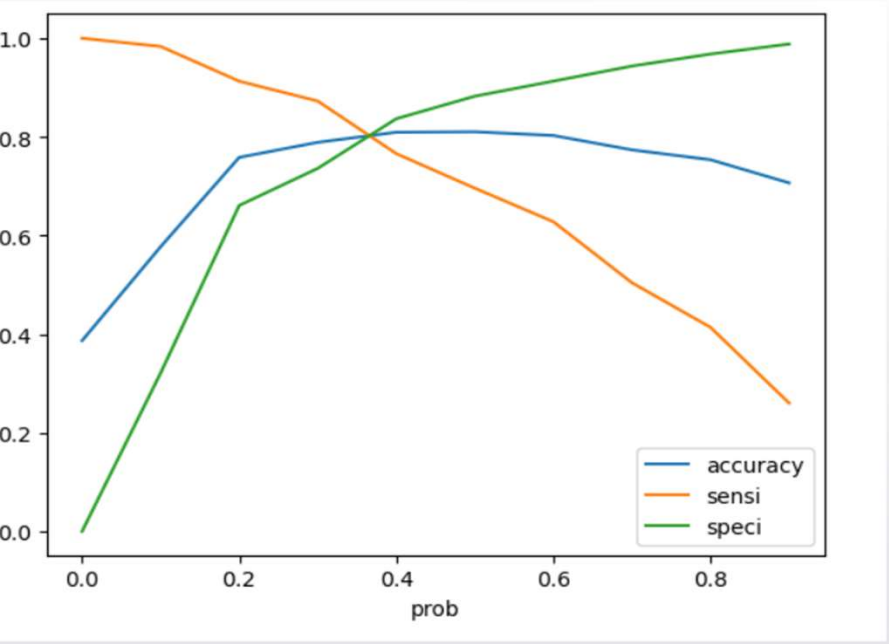
SPECIALIZATION VS CONVERTED

Most of the leads have not information about specialization on the other hand finance, marketing, operations management has high conversion rate peoples from these specialization can be promising leads

MODEL BUILDING

- Splitting into train and test set
- Scale variables in train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate the variables on the basis of high p-value
- Check VIF value for all existing columns
- Predict using train set
- Evaluate the accuracy and other metric
- Predict using test set
- Precision and recall analysis on test prediction

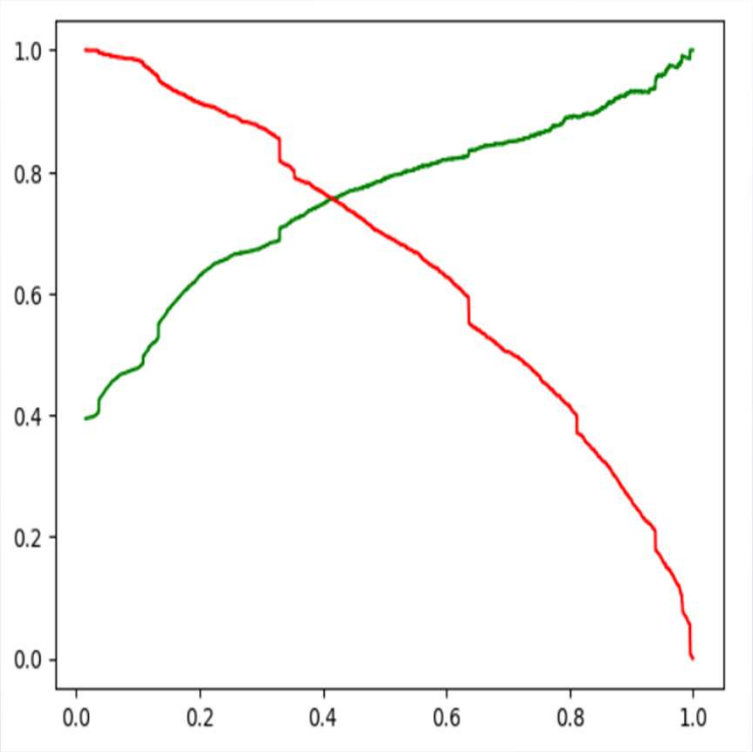
Model evaluation (TRAIN)




```
array([[3127,  768],
       [ 482, 1974]], dtype=int64)
```

Accuracy	80.31%
Sensitivity	80.37%
Specificity	80.28%
Precision	78.86%
Recall	69.58%

Model evaluation (TEST)




```
array([[1470,  274],
       [ 229,  750]], dtype=int64)
```

Accuracy	81.52%
Precession	73.24%
Recall	76.60%

Conclusion

- EDA:

- People spending higher then average time are promising leads, so targeting them and approaching them can be helpful in conversions.
- SMS message can have a high impact on lead conversions
- Landing page submissions can help out more leads
- Marketing , human resources management has high conversion rate. peoples from these specialization can be promising leads
- References and offer for referring a lead can be a good source of higher conversions
- An alert message or information has seen to have high lead conversion rate

Conclusion

- Logistic regression model

- ❑The model shows 81% accuracy
- ❑The threshold has been selected from accuracy, sensitivity and specificity measures and precession, recall curves .
- ❑The model shows 76% sensitivity and 83% specificity
- ❑The model finds correct promising leads and leads that have less chances of getting converted.
- ❑Overall this model proves to be accurate