LEAD SCORE CASE STUDY

LOGISTIC REGRESSION

VIDYALAKSHMI SURESH

PROBLEM STATEMENT

- X Education is an organization which provides online courses to industry professionals. The company marks its courses on several popular websites like google.
- X Education wants to select most promising leads that can be converted to paying customers.
- Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher lead conversion. Leads comes through numerous modes like email, advertisements on websites ,google searches etc.
- The company has had 30% conversion rate through the whole process of tuning leads into customers by approaching those leads which are to be found having interest in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

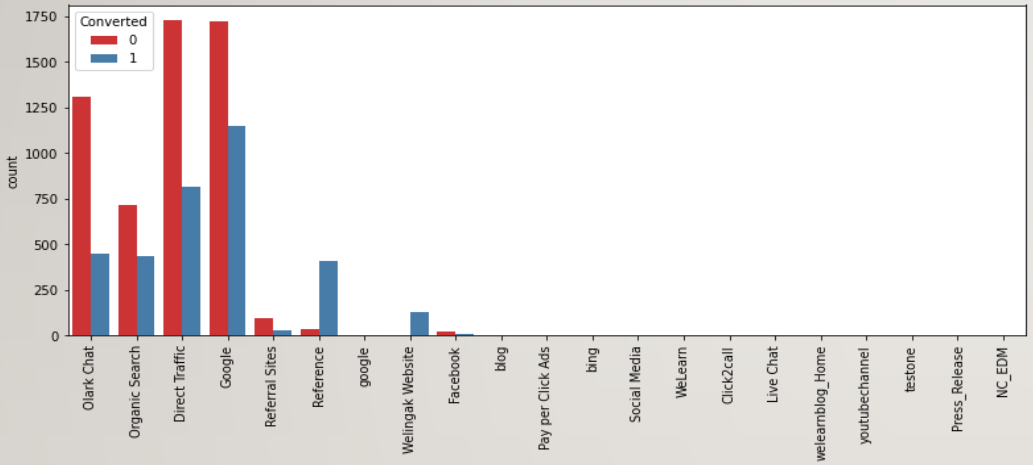
BUSINESS GOAL

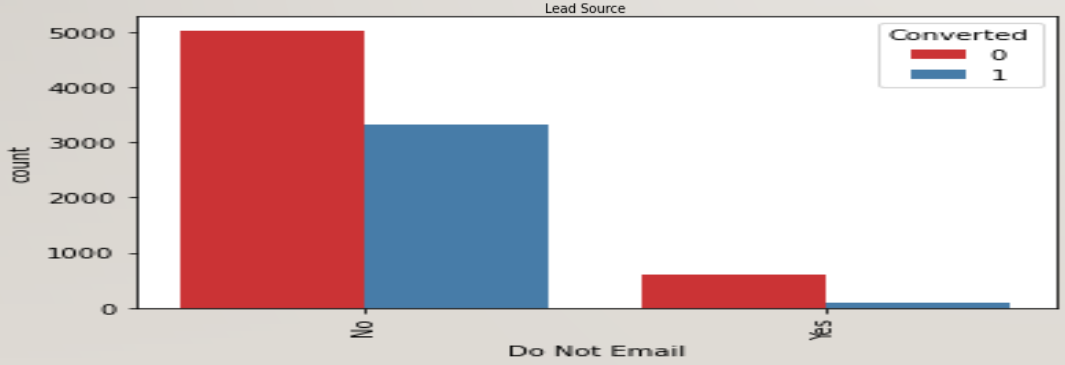
- The company requires a model to be built for selecting most promising leads.
- Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted, the lower it is the lesser the chances of conversion.
- The model to be built in lead conversion rate around 80% or more.

STRATEGY

- Import Data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- 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 and other metrics for evaluation

EXPLORATORY DATA ANALYSIS





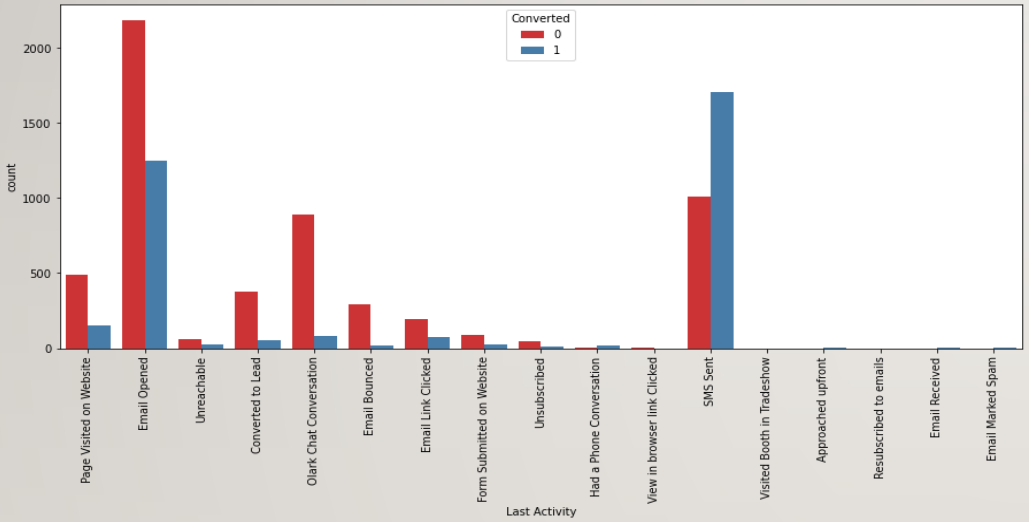
- **Lead Source Vs Converted**

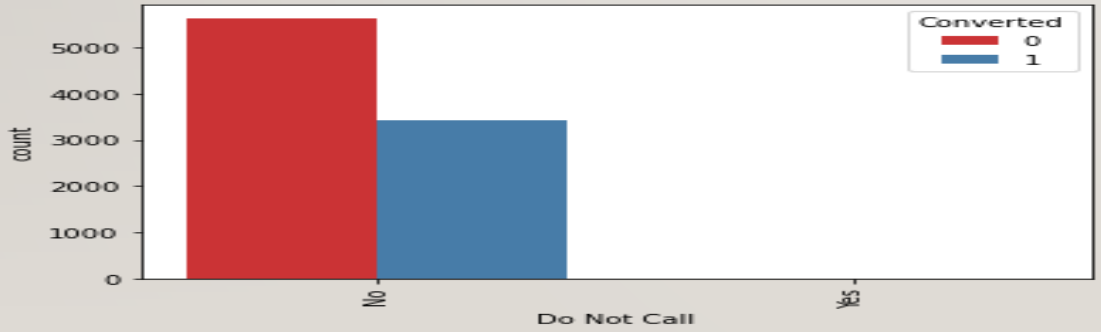
Google searches has had high conversions compared to other modes, whilst references has had high conversion rate.

- **Do Not Email Vs Converted**

Google searches has had high conversions compared to other modes, whilst references has had high conversion rate.

EXPLORATORY DATA ANALYSIS





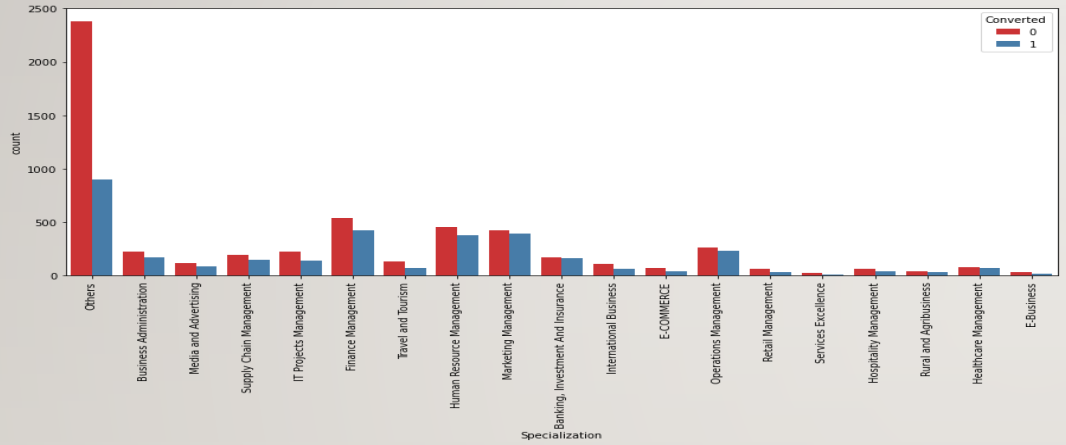
- **Last Activity Vs Converted**

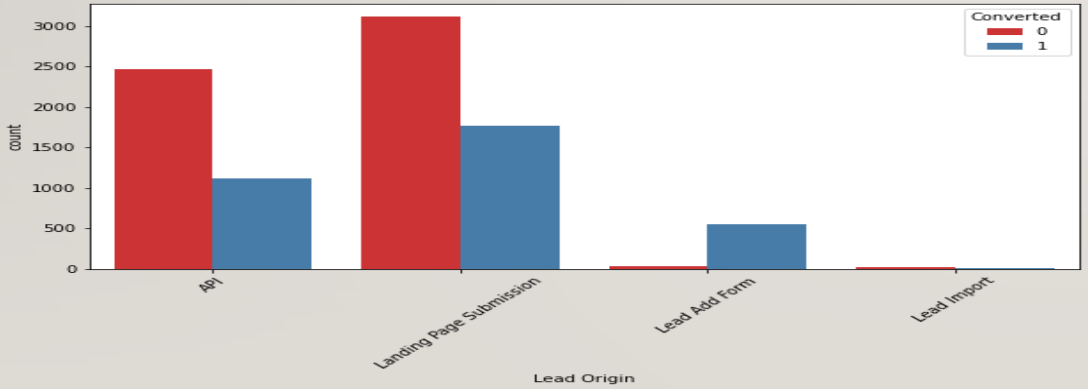
SMS has shown to be a promising method for getting higher confirmed leads, email also has high conversions.

- **Do Not Call Vs Converted**

Most leads prefer not to be informed through phone

EXPLORATORY DATA ANALYSIS





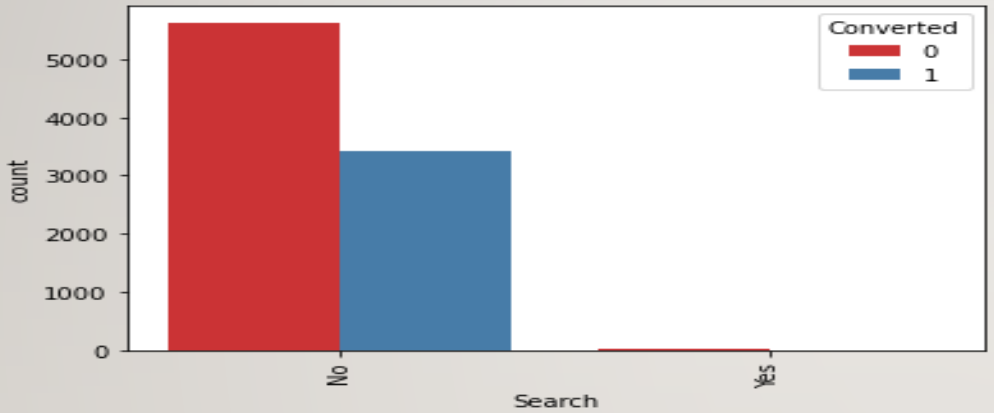
- **Specialization Vs Converted**

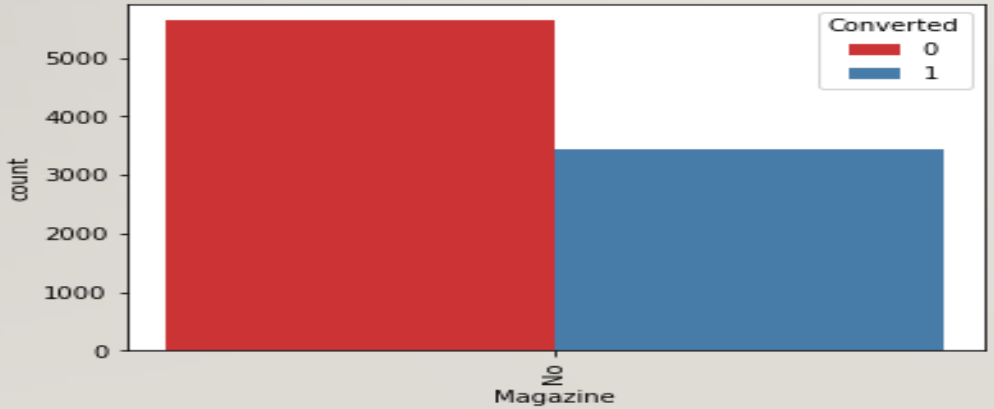
Most of the leads have no information on specialization. Marketing Management, HR management has high conversion rates. People from these specializations can be promising leads

- **Lead Origin Vs Converted**

Landing Page submissions has had high lead conversions.

EXPLORATORY DATA ANALYSIS





- **Search Vs Converted**

Graph says Searches are not good source of lead.

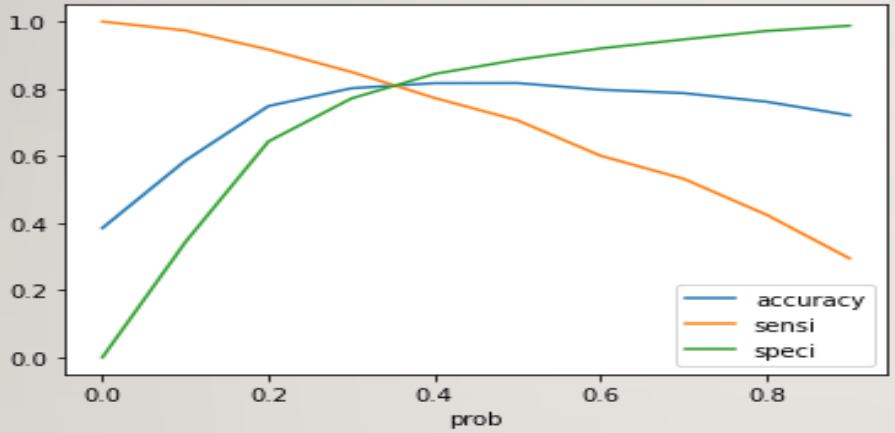
- **Magazines Vs Converted**

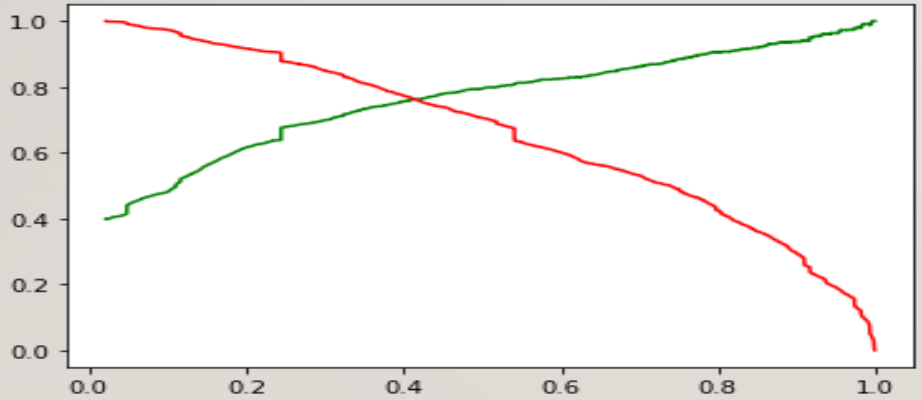
Magazines do not have higher conversion rate.

MODEL BUILDING

- Split into train and test set
- Scale variables in train set
- Built the first model
- Use RFE to eliminate less relevant variables
- Built the next model
- Eliminate variables based on high p-values
- Check VIF value for all the existing columns
- Predict using train set
- Evaluate accuracy and other metric
- Predict using test set
- Precision and recall analysis on test predictions

MODEL EVALUATION (TRAIN)





- **Accuracy, Sensitivity and Specificity**

81.0% Accuracy

81.7% Sensitivity

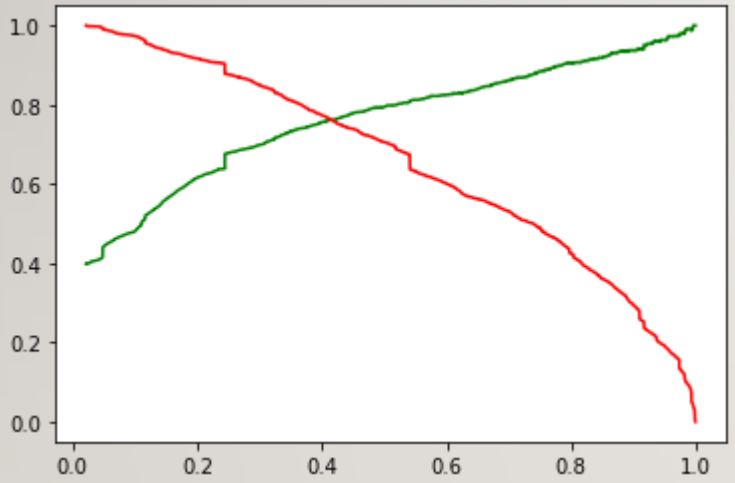
80.6% Specificity

- **Precision and Recall**

79.6% Precision

70.6% Recall

MODEL EVALUATION (TEST)



- **Accuracy, Sensitivity and Specificity**

80.4% Accuracy

80.4% Sensitivity

80.5% Specificity

- **Precision and Recall**

79.6% Precision

70.6% Recall

CONCLUSION

EDA:

- People spending higher than average time are promising leads, so targeting them and approaching them can be helpful in conversions.
- SMS messages can have a high impact on lead conversion.
- Land page submissions can help find out more leads.
- Marketing management, human resources management has high conversion rates. People from these specializations can be promising leads
- References and offers for referring a lead can be good source for conversions.

Logistic Regression Model:

- The model shows high close to 81%accuracy
- The threshold has been selected from Accuracy, sensitivity, specificity measures and precision ,recall curves.
- The model shows 80% Sensitivity and 80% specificity.
- The model finds correct promising leads and leads that have less chances of getting converted.
- Overall this model proves to be accurate.