Lead Scoring Case - Study

PANDURANG RAYFALWAR…

Problem Statement

X Education is an organization that provides online courses for industry professionals. The company marks its courses on several popular websites like Google.

X Education wants to select the 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 come through numerous modes like email, advertisements on websites, google searches etc.

The company has had a 30% conversion rate through the whole process of turning leads into customers by approaching those leads who are to be found having an interest in taking the course. The implementation process of lead generating attributes is not efficient in helping conversions.

Business Goal

The company requires a model to be built for selecting the most promising leads.

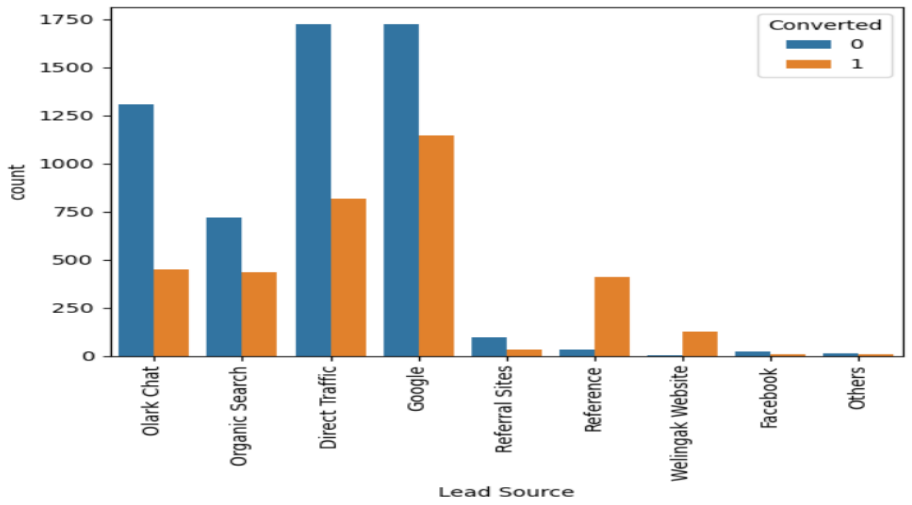
Lead score to be given to each lead such that it indicates how promising the lead could be. The higher the lead the more promising the lead to get converted, the lower it is the lesser the chances of conversion

The model is to be built with a lead conversion rate of around 80% or more.

Strategy

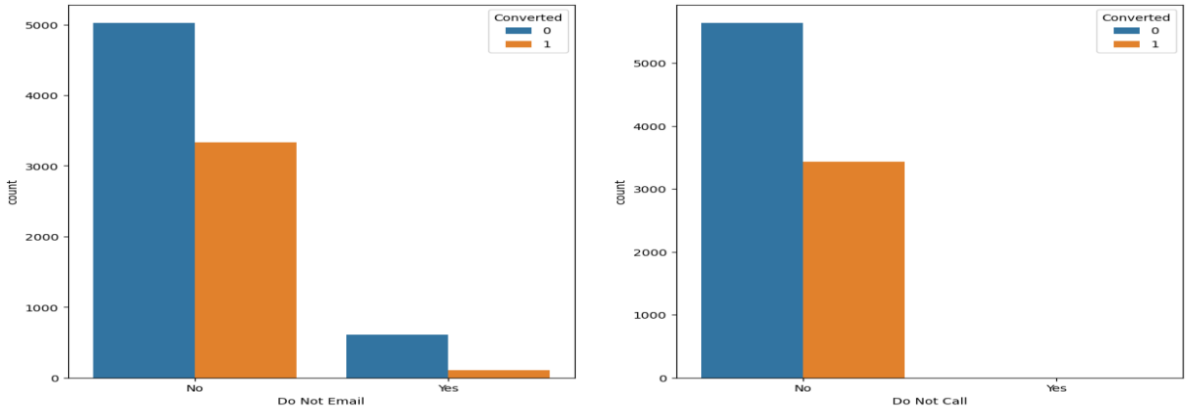
- Import data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis(EDA)
- Feature scaling
- Data preparation for the model building
- Build a logistic regression model
- Assign a lead score for each lead
- Test the model on the train data set
- Evaluate the model by different measures and metrics
- Test the Model
- Measure the accuracy of the model and other metrics for evaluation lead

Exploratory Data Analysis



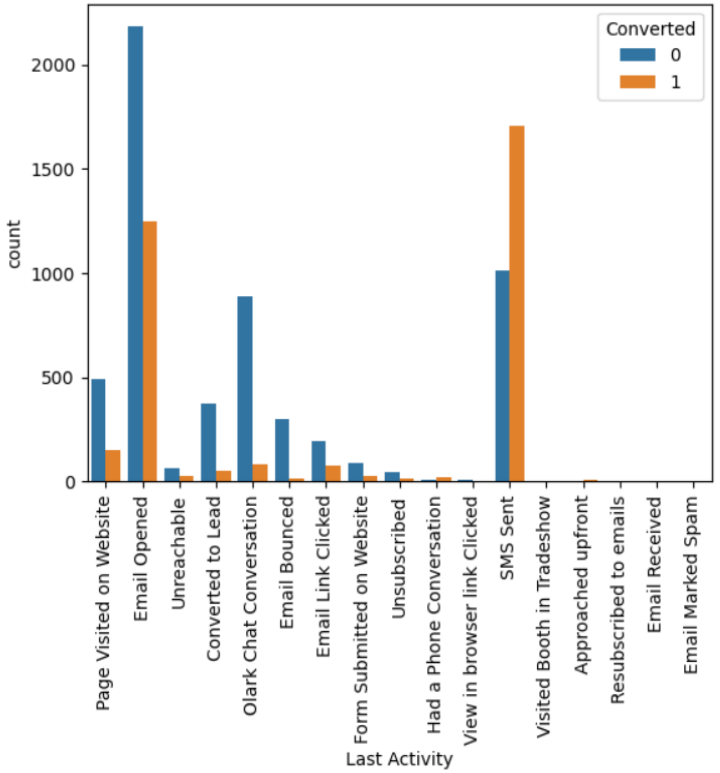
Lead source Vs Converted

Google search and direct traffic has a higher conversion rate.



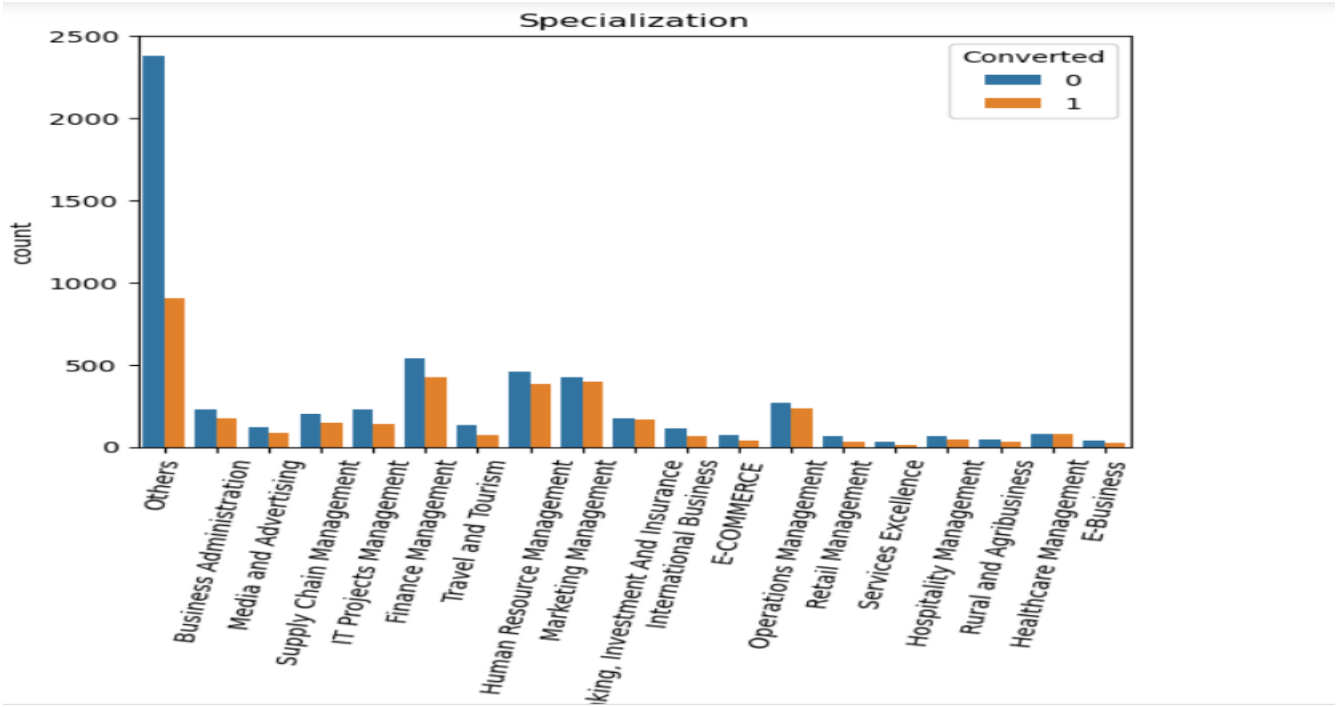
Do Not Call and do Not Email Vs Converted

Leads with not to mail and call have a higher conversion rate.



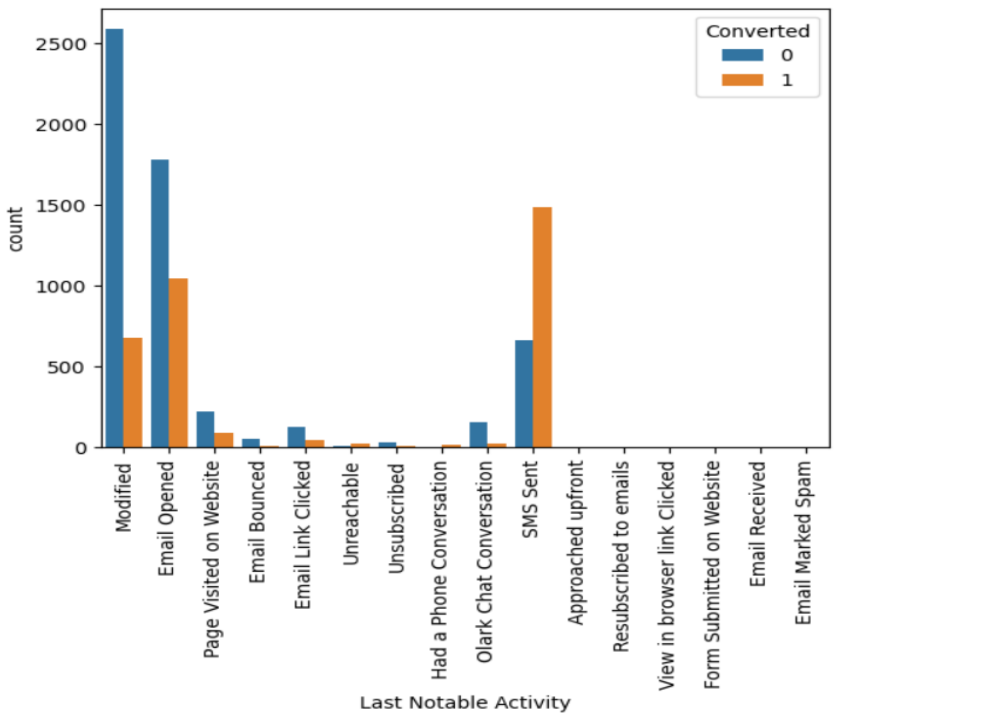
Last Activity Vs Converted

SMS has shown to be promising method for getting higher confirmed leads, emails also have high conversion



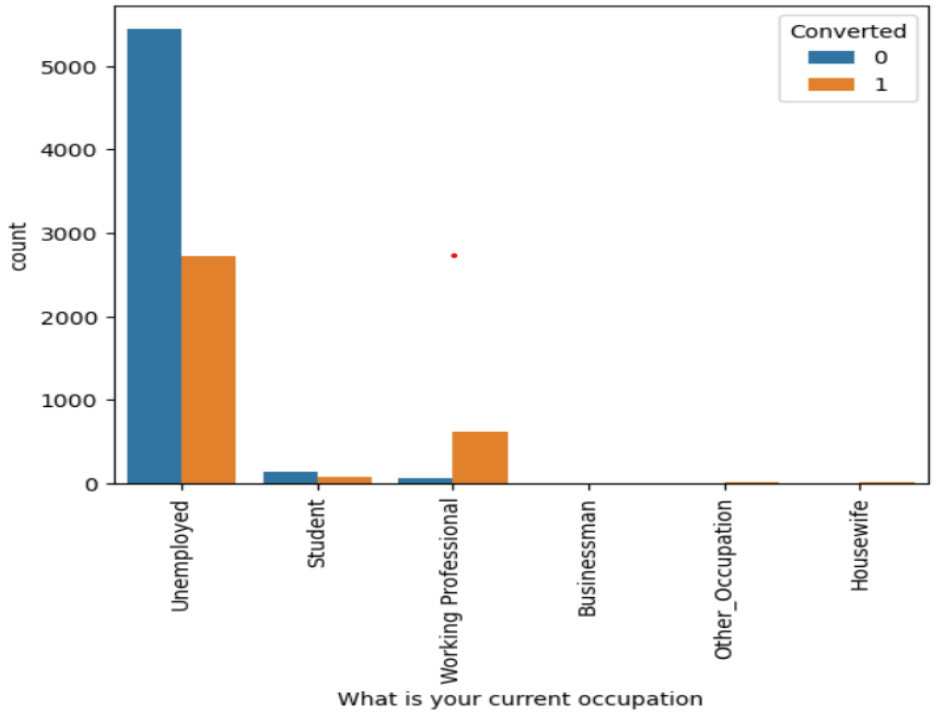
Specialization Vs Converted

most of the leads have no information about specialization.
on the other hand, marketing management, human resources management have high conversion rates.
people from these specializations can be promising leads.



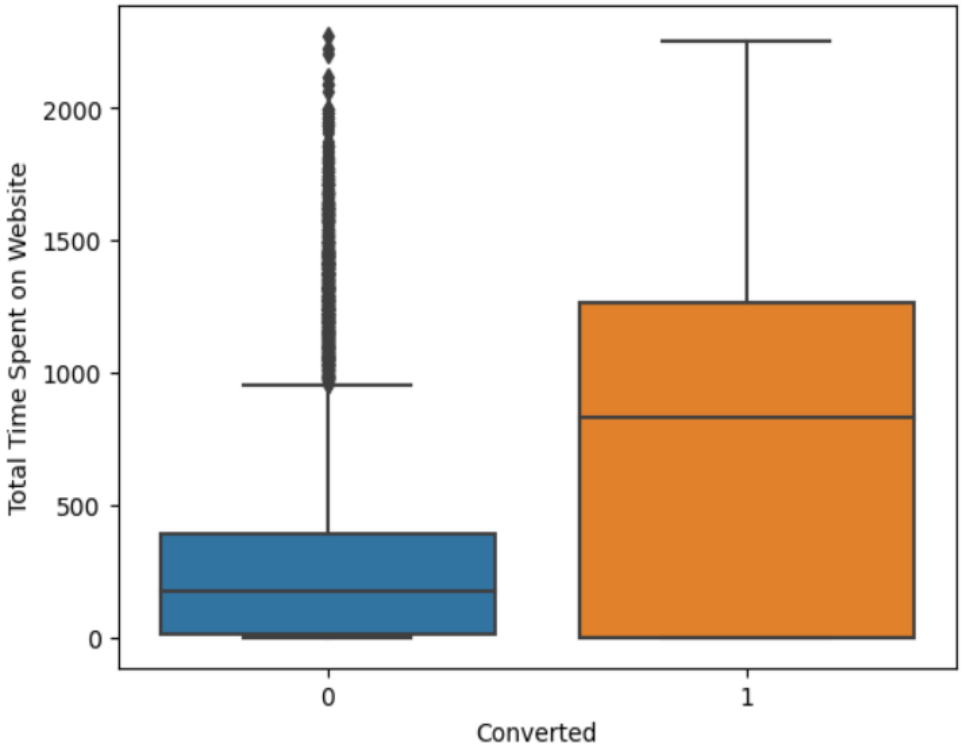
Last Notable Activity Vs Converted

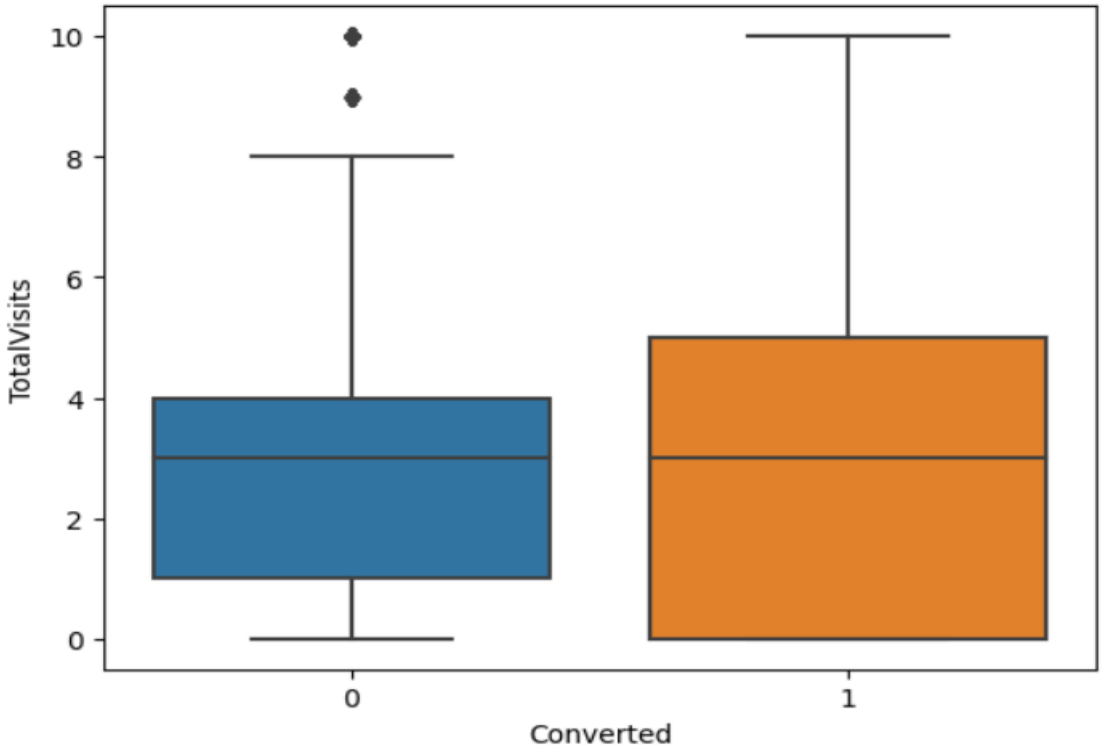
Most leads are converted with messages Email also includes leads.



Current Occupation Vs Converted

Most leads are converted to those who are unemployed.





Time spent on website vs. converted

People spending higher than average time are promising leads.

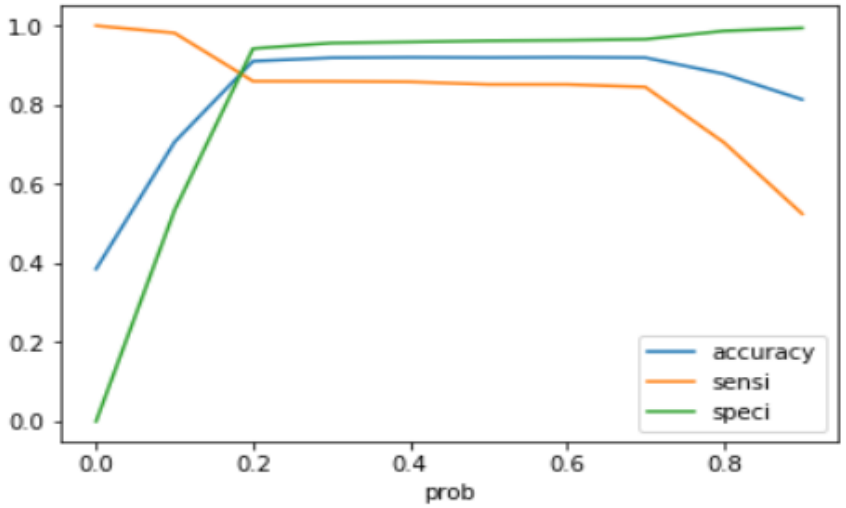
Total Visits Vs. Converted

Higher total visits have a slightly higher chance of being a promising lead.

Model Building

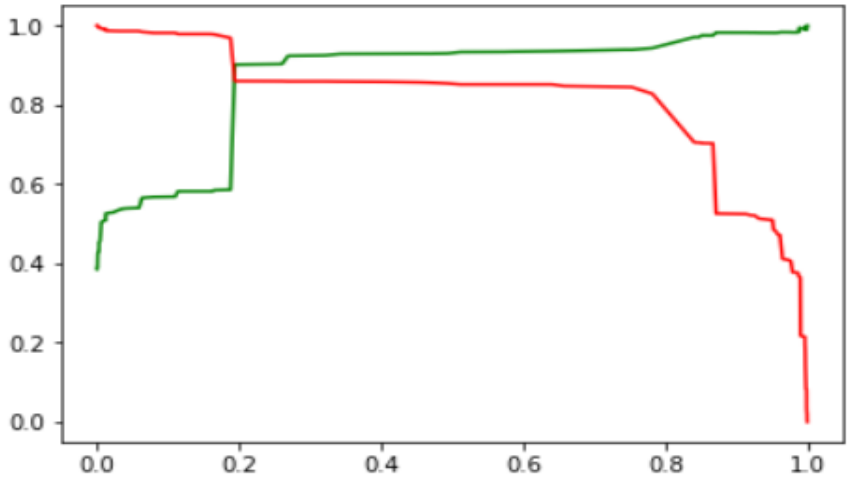
- Splitting into train and test set
- Scale variables in the train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate variables based on high p-values
- Check the VIF value for all the existing columns
- Predict using train set
- Predict using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

Model Evaluation



Accuracy Sensitivity and Specificity

- 91% Accuracy
- 85% Sensitivity
- 94% Specificity



Precision and Recall

- 90.8% Precision
- 86.3% 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.
- Landing page submissions can help find out more leads.

- Marketing management and human resource management have high conversion rates. People from these specializations can be promising leads.
- References and offers for referring a lead can be good sources for higher conversions.
- An alert message or information has been seen to have a high lead conversion rate.

Logistic Regression Model

- The model shows close to 91% accuracy.

- The threshold has been selected from Accuracy, Sensitivity, Specificity measures precisions and recall curves.

- The model shows 85% sensitivity and 94% specificity.

- The model finds correct promising leads and leads that have less chance of getting converted.

- Overall this model proves to be accurate