LEAD SCORE CASE STUDY

LOGISTIC REGRESSION

BY SWAYANSUDHA SAHOO

PROBLEM STATEMENT

X Education is an organization which provides online courses for industry professional. 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 come through numerous modes like email, advertisements on websites, 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. 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.

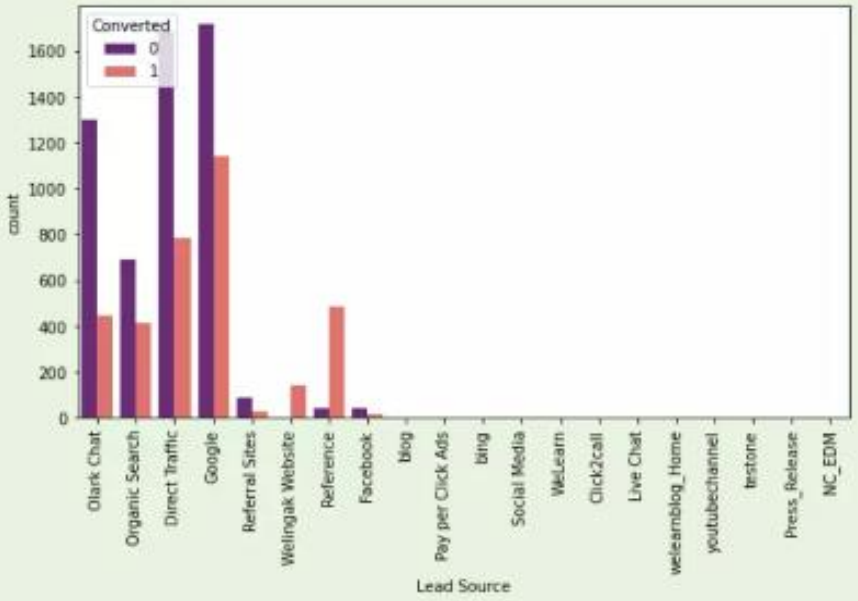
Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the more promising the lead to get converted, the lower it is the lesser the chances of conversion

The model to be built in lead conversion rate around 80% or more.

STRATEGY

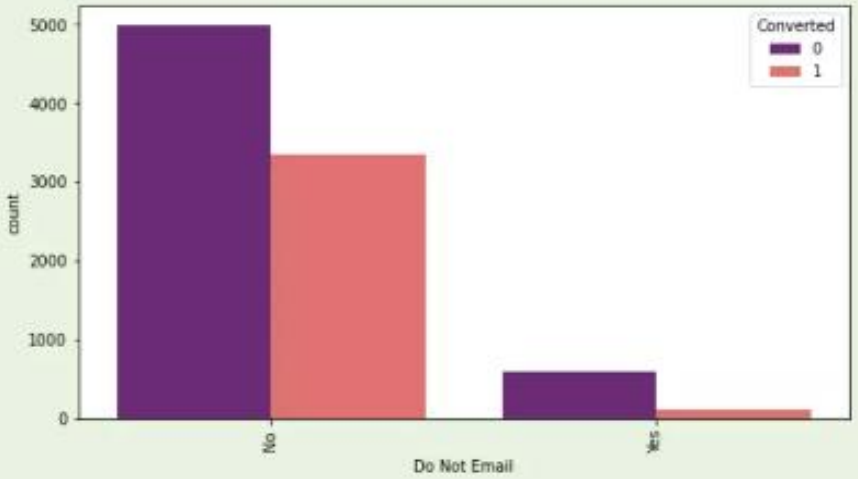
- Import data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model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



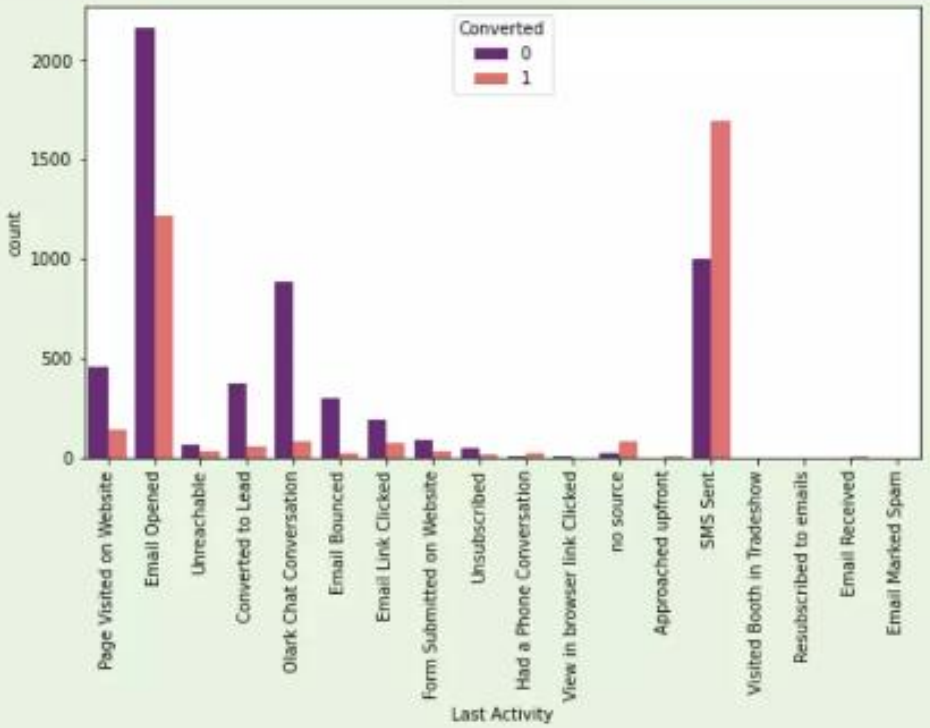
LEAD SOURCE VS CONVERTED

google searches has had high conversions compared t other modes, whilst references has had high conversion rate.



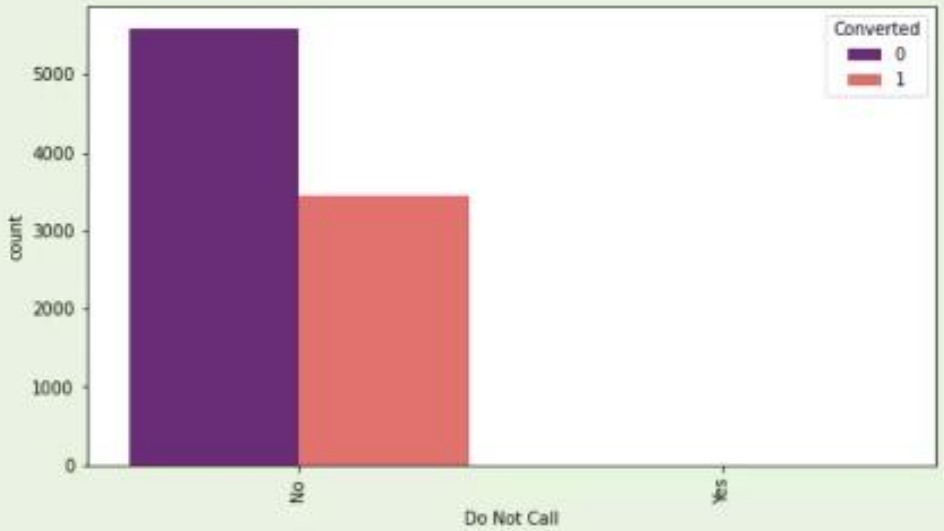
DO NOT EMAIL VS CONVERTED

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



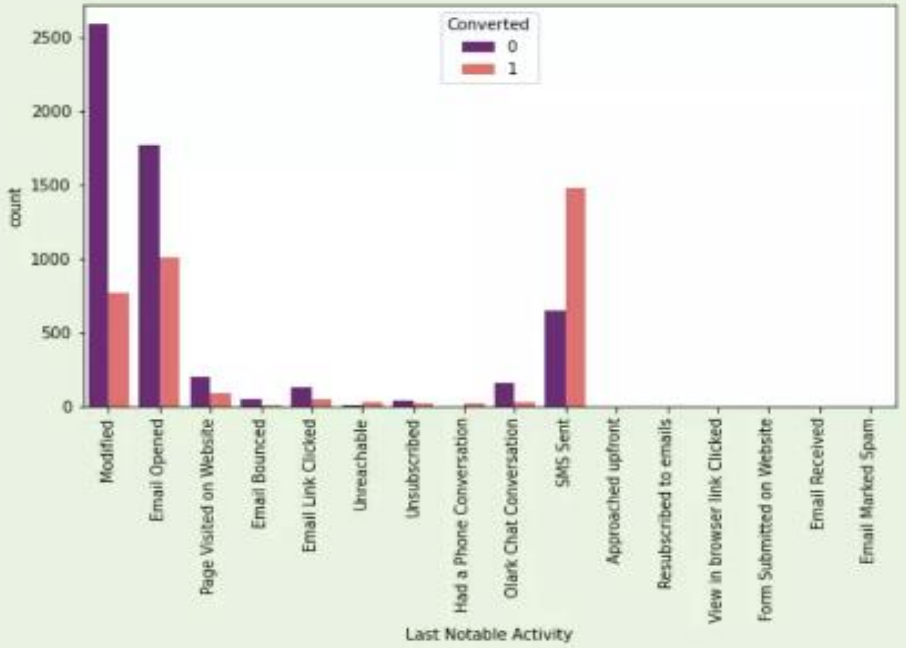
LAST ACTIVITY VS CONVERTED

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



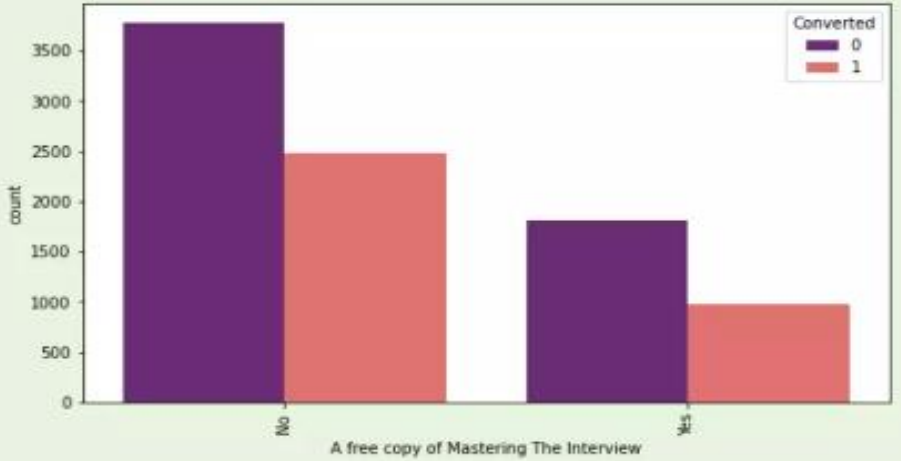
DO NOT CALL VS CONVERTED

most leads prefer not to informed through phone



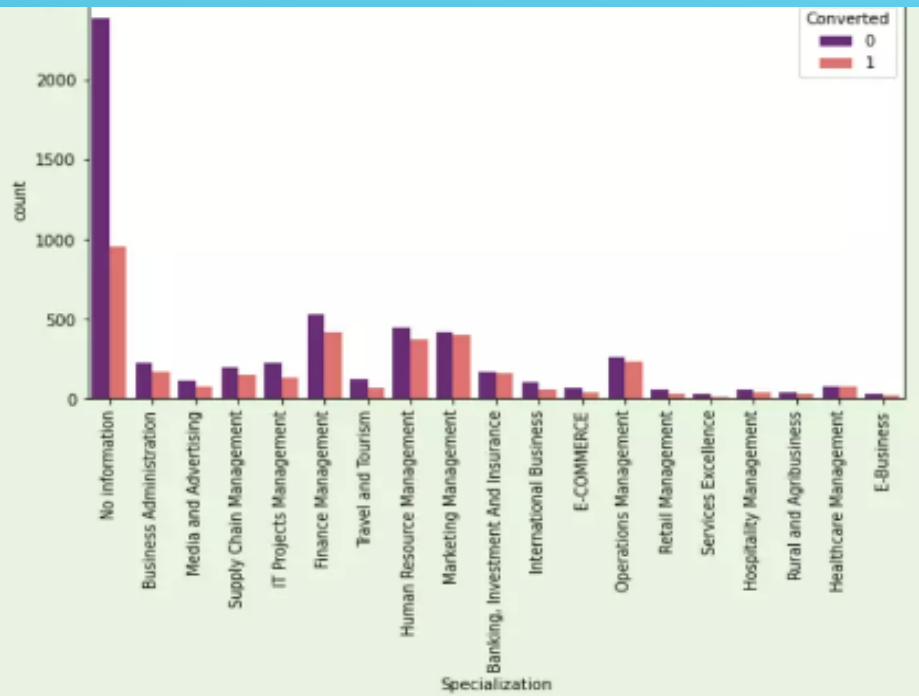
LAST NOTABLE ACTIVITY VS CONVERTED

most leads are converted with messages.
Emails also induce leads.



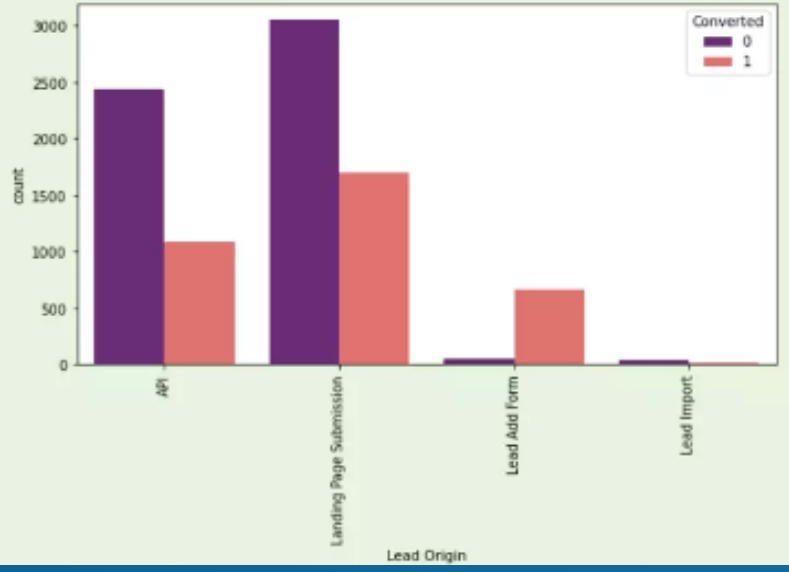
A FREE COPY OF MASTERING THE INTERVEIW VS CONVERTED

leads prefer less copies of interviews.



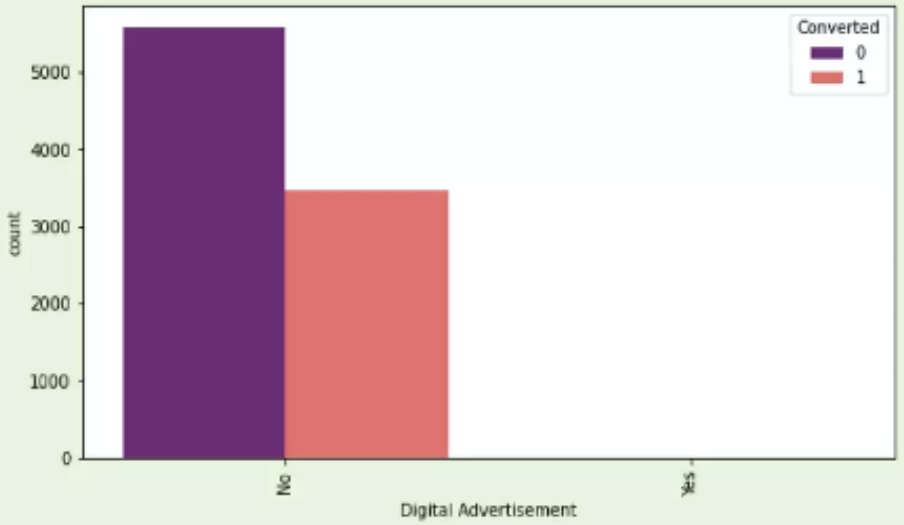
SPECIALIZATION VS CONVERTED

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



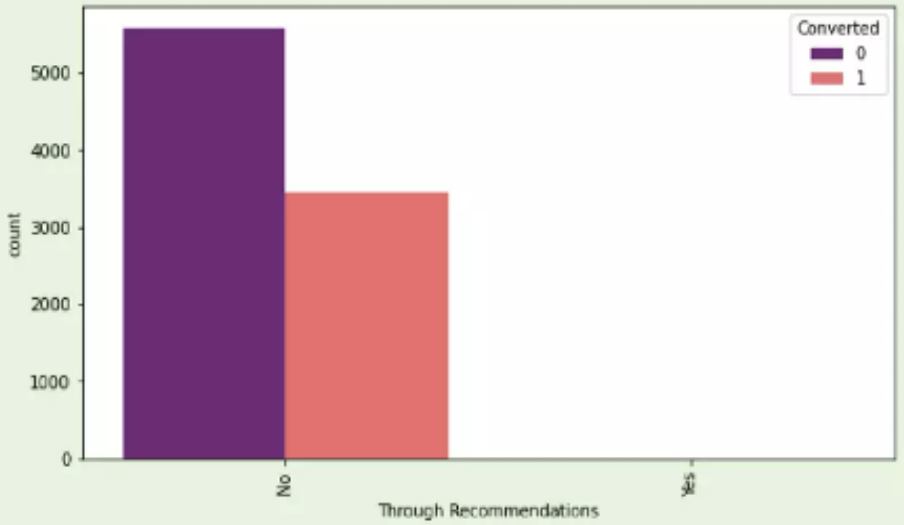
LEAD ORIGIN VS CONVERTED

landing page submissions has had high lead conversions



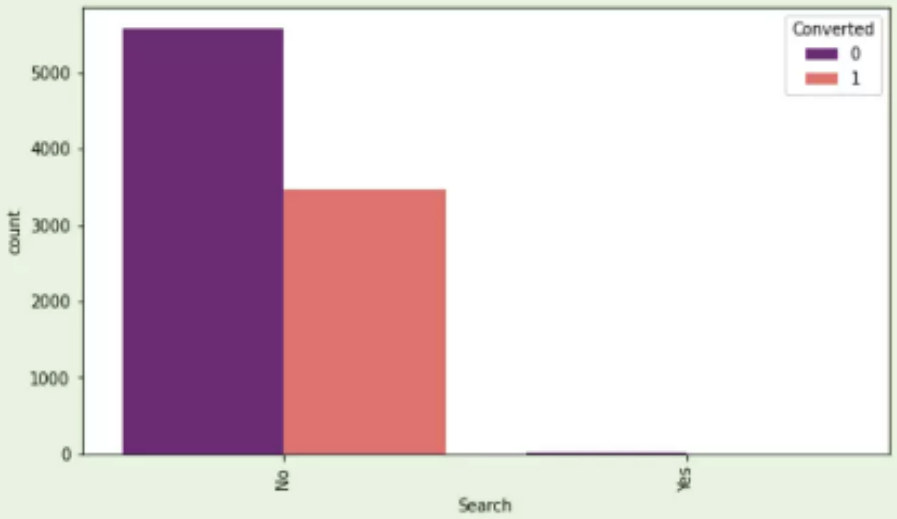
DIGITAL ADVERTISEMENTS VS CONVERTED

based on the above graph digital advertisements do not have promising leads



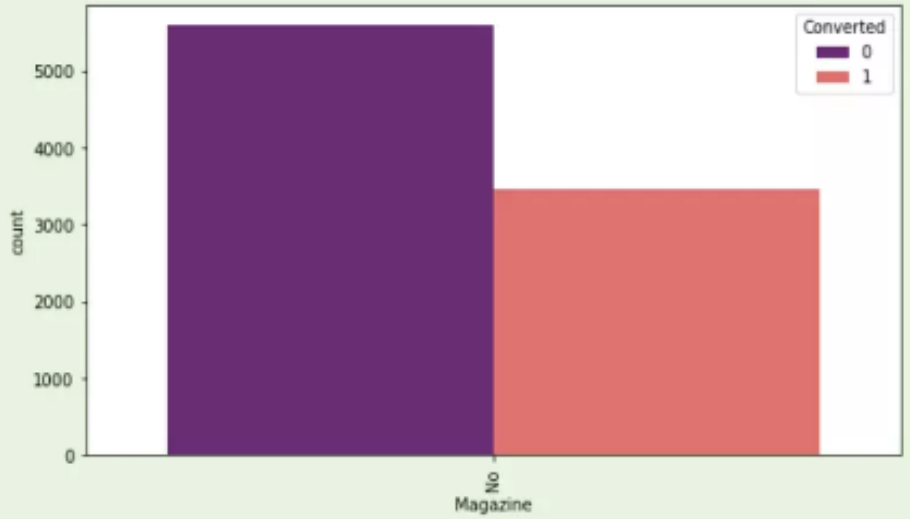
THROUGH RECOMMENDATIONS VS CONVERTED

from the above graph, recommendations are not a good source for promising leads



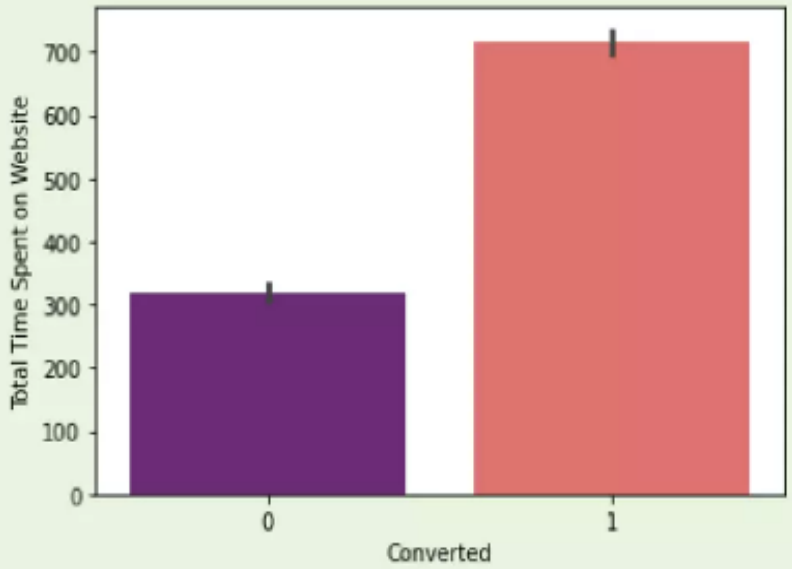
SEARCH VS CONVERTED

the above graph shows searches are not good source of leads



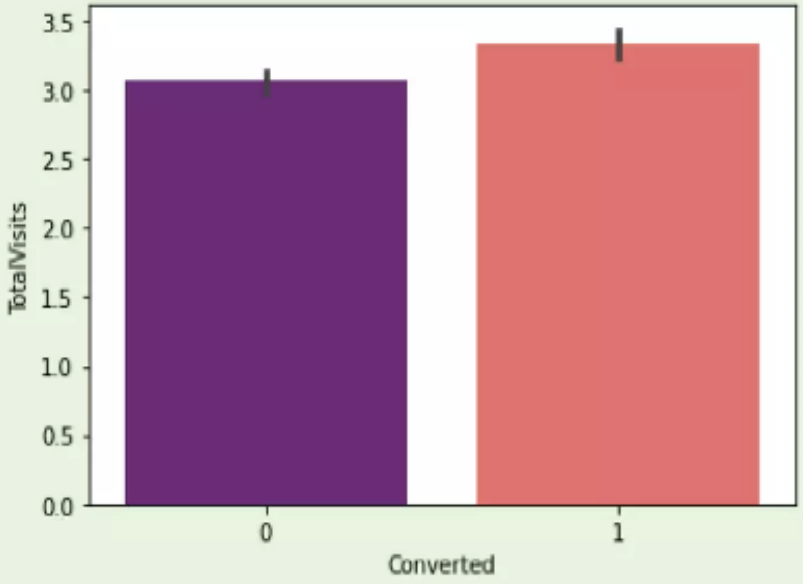
MAGAZINE VS CONVERTED

magazines do not have higher conversion rate



TOTAL TIME SPENT ON WEBSITES VS CONVERTED

people spending higher than average time are promising leads



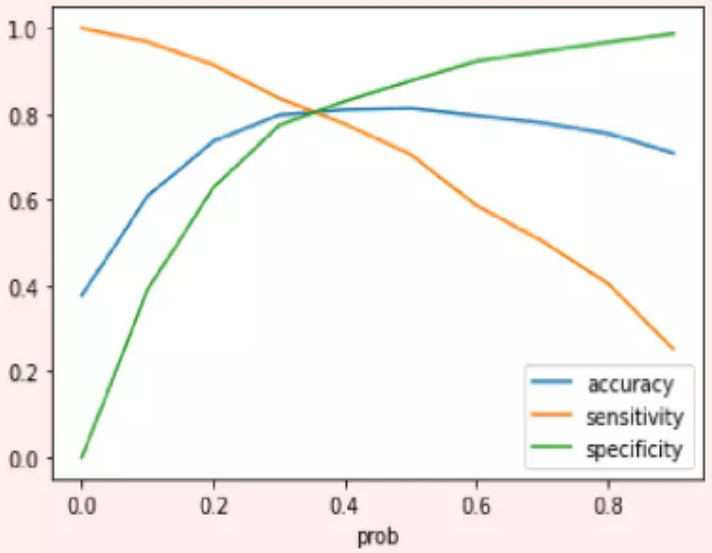
TOTAL VISITS VS CONVERTED

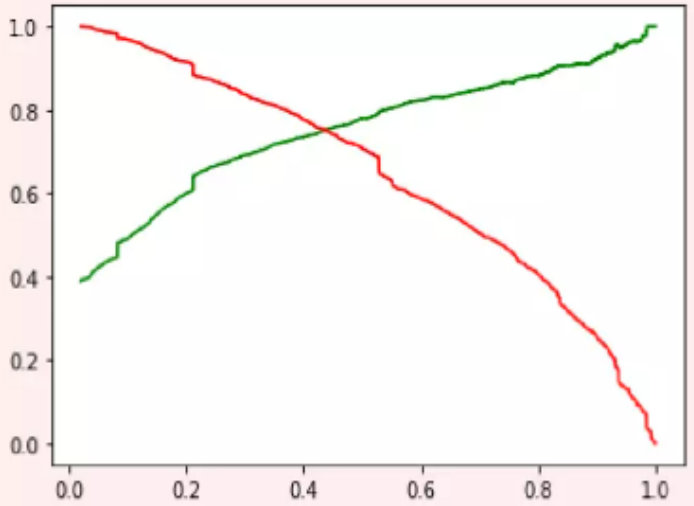
higher total visits have a slight higher chances of being a promising lead

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 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

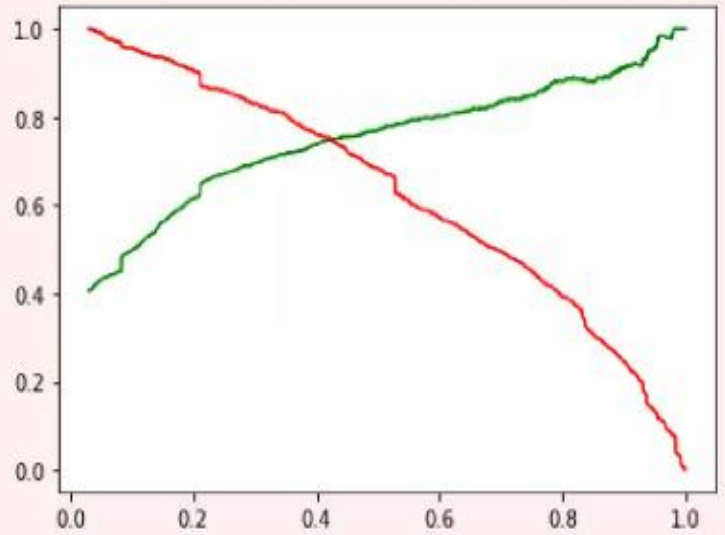
3270	671
534	1859

- 80.9% Accuracy
- 77.6% Sensitivity
- 82.9% Specificity

PRECISION AND RECALL

- 73.4% Precision
- 77.6% Recall

MODEL EVALUATION (TEST)



PRECSION AND RECALL

- 74.4% Precision
- 75.5% Recall

Test set threshold has been set as 0.41

ACCURACY SENSITIVITY AND SPECIFICITY

1370	277
261	807

- 80.1% Accuracy
- 75.5% Sensitivity
- 83.1% Specificity

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, 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 higher conversions
- An alert messages or information has seen to have high lead conversion rate

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 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