LEAD SCORING CASE STUDY

MOSHEER KHAN

PROBLEM STATEMENT

Company named X Education gets a lot of leads. However, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

BUSINESS GOAL

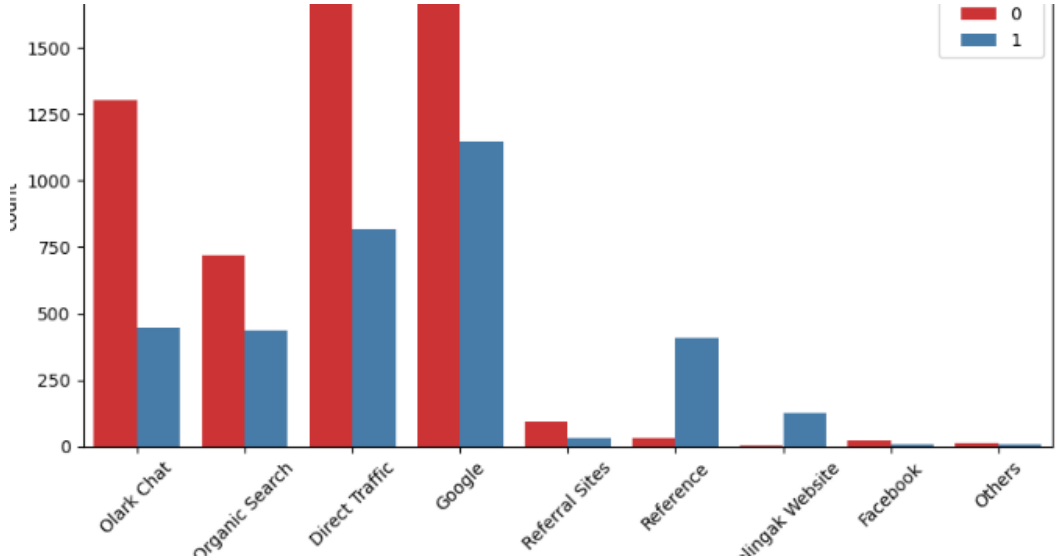
The Company requires a model to be built for selecting most promising leads.

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 lead to get converted, the lower it is the lesser the chances of conversion.

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

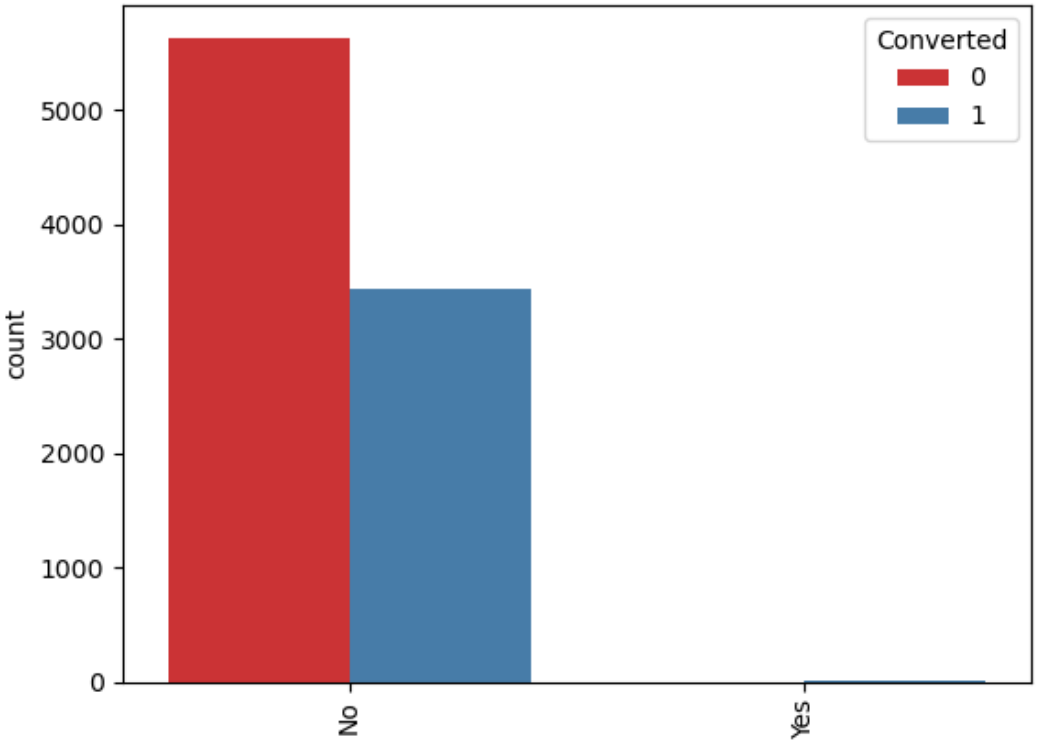
STRATEGY

- Import data
- Clear and prepare the acquired data for further analysis
- Exploratory data analysis for figuring most helpful attributes for conversion
- Scaling features
- Prepare the data for model building
- Build a logical regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Measure the accuracy of the model and other metrics for evaluations



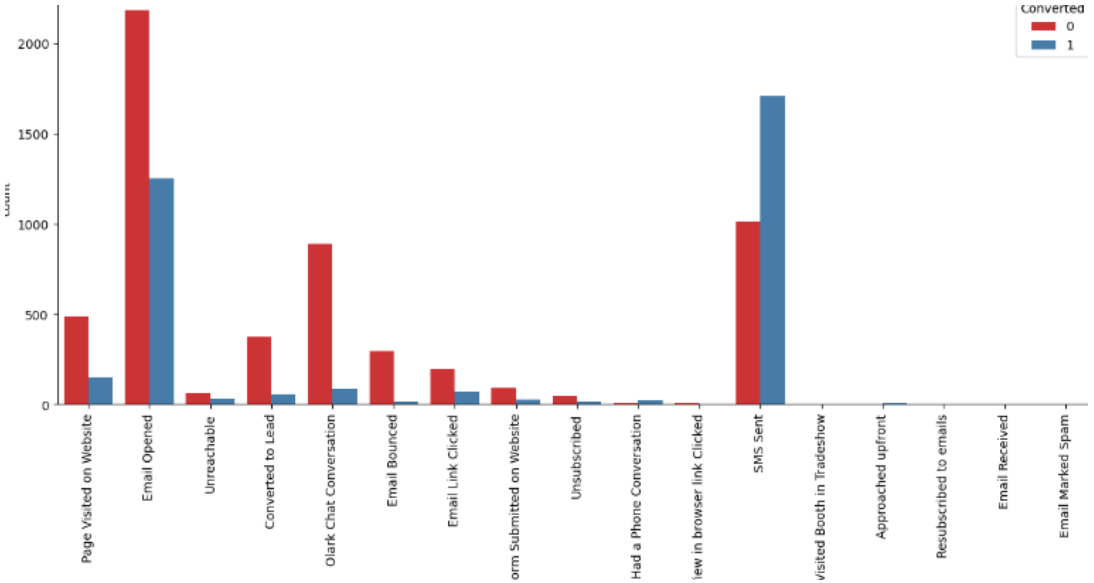
LEAD SOURCE VS CONVERTED

Google searches has had high conversion compared to other modes. Whilst references has had high conversion rate.



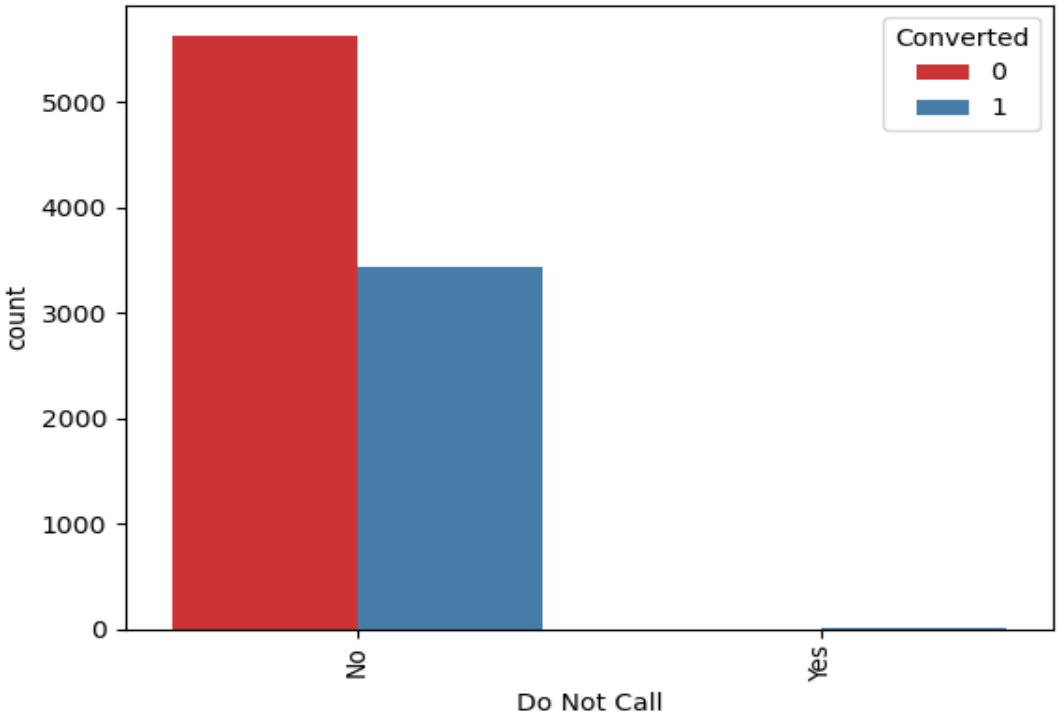
DO NOT EMAIL VS CONVERTED

Google search has had high conversion compared to other modes, Whilst reference has had high conversion rate.



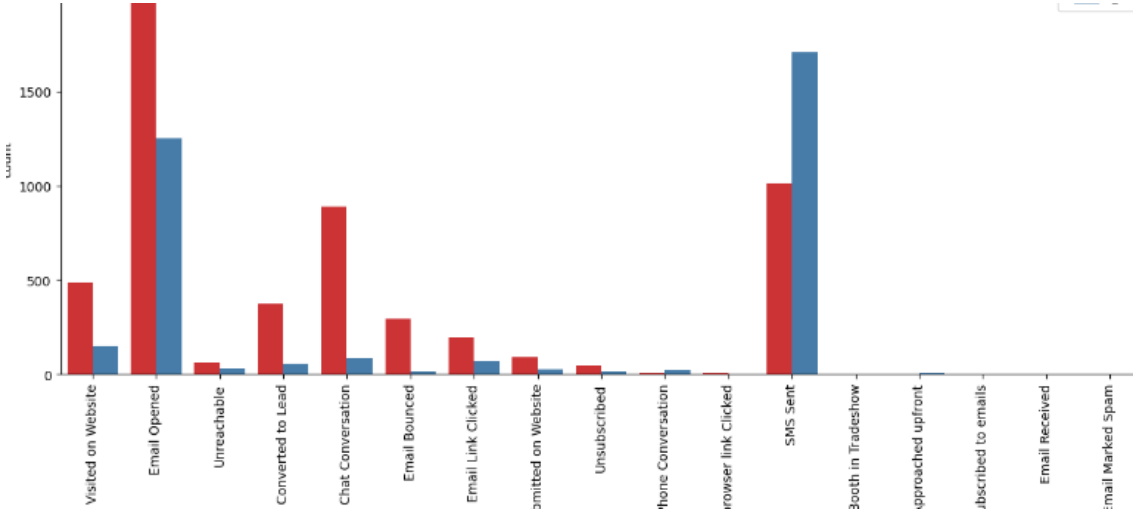
LAST ACTIVATE VS CONVERTED

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



DO NOT CALL VS CONVERTED

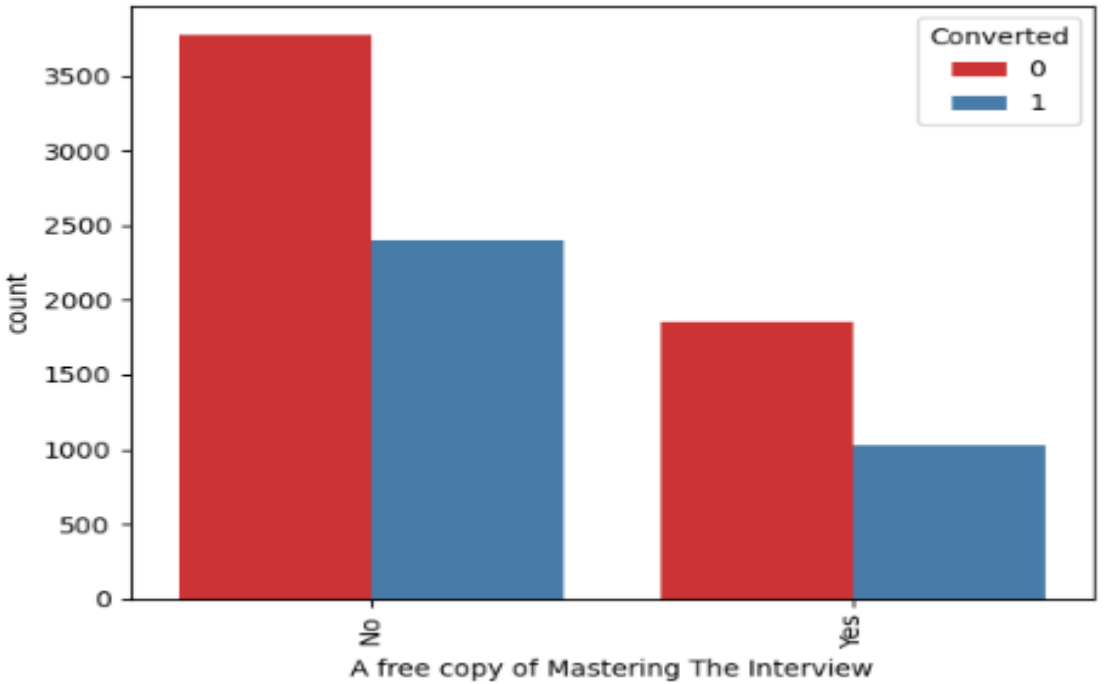
Most leads prefer not to informed through phone.



LAST NOTABLE ACTIVITY VS CONVERTED

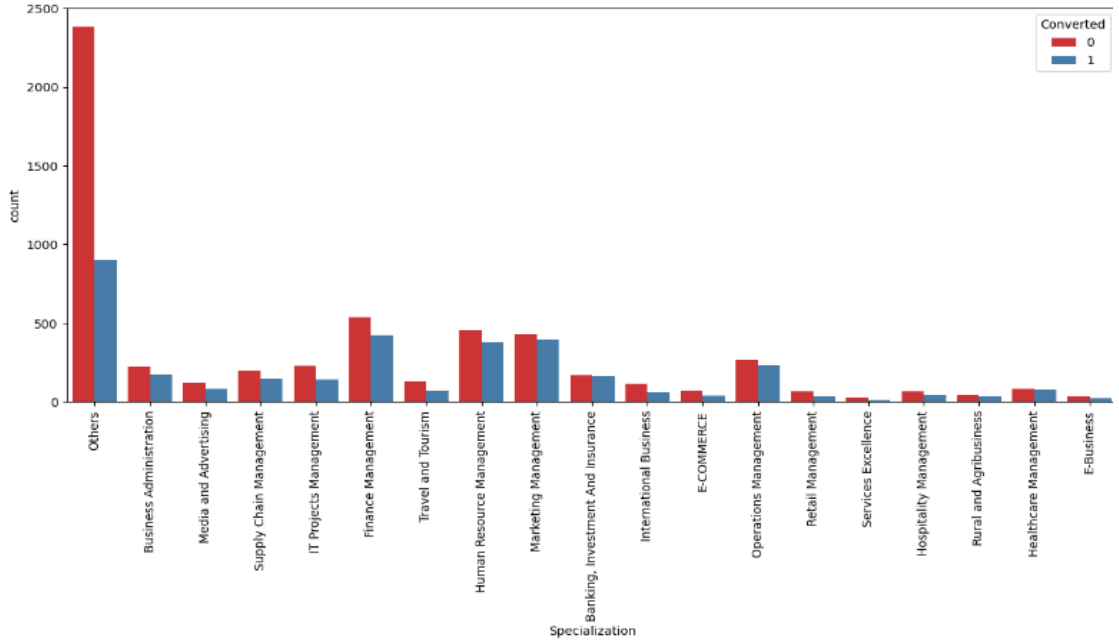
Most lead are converted with message.

Email also include leads.



A FREE COPY OF MASTERING THE INTERVEIW VS CONVERTED

Most leads prefer not to informed through phone.

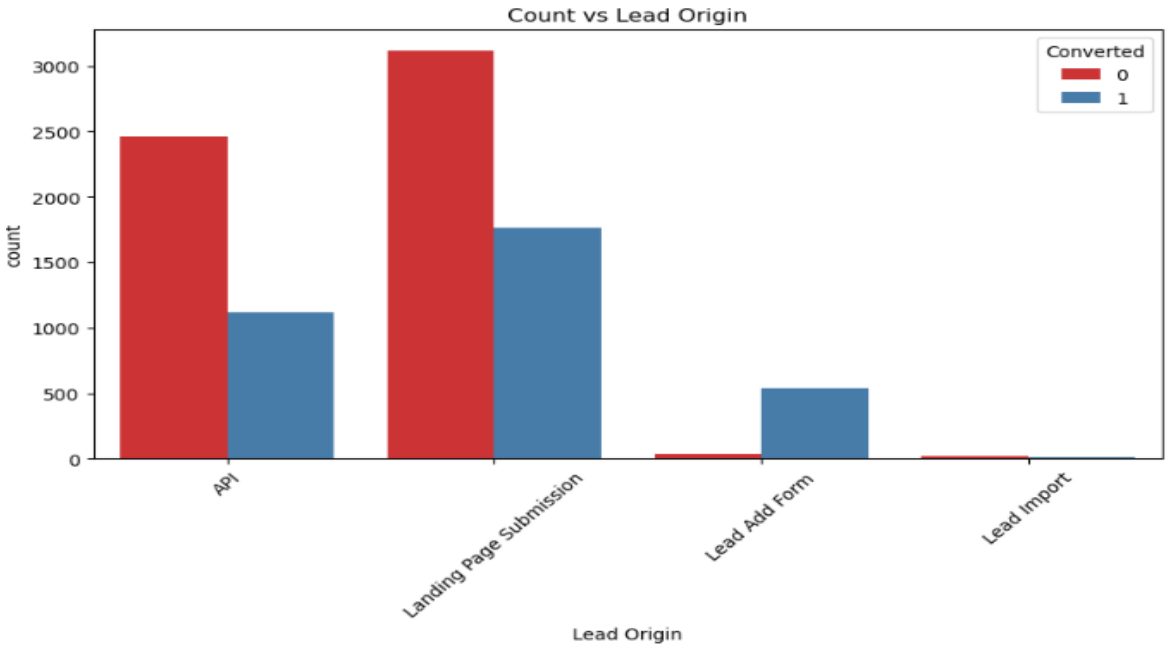


SPECILIZATION VS CONVERTED

Most of the lead have no information about specialization.

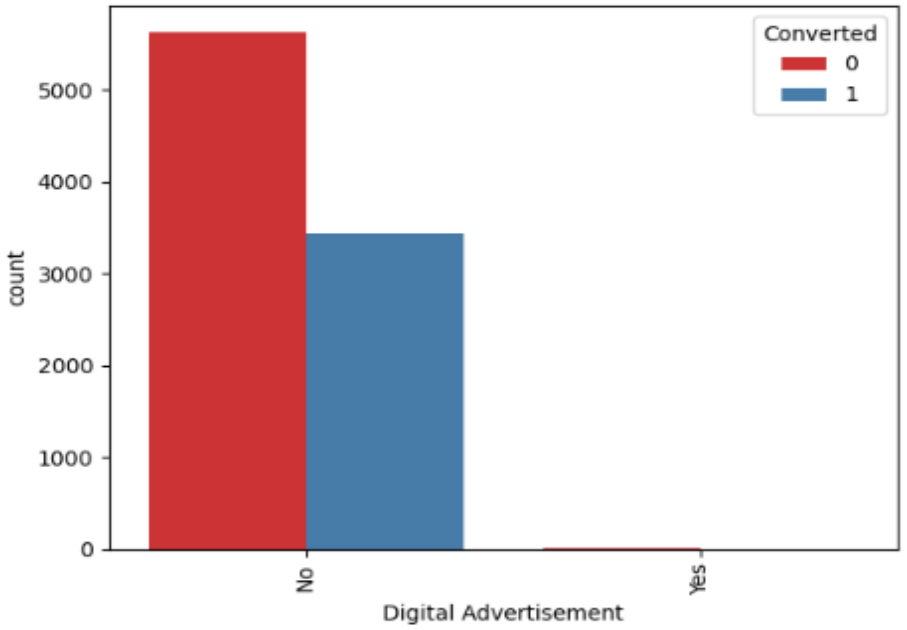
On the other hand, marketing management, human resource management has high conversion rates.

People from these specializations can be promising leads.



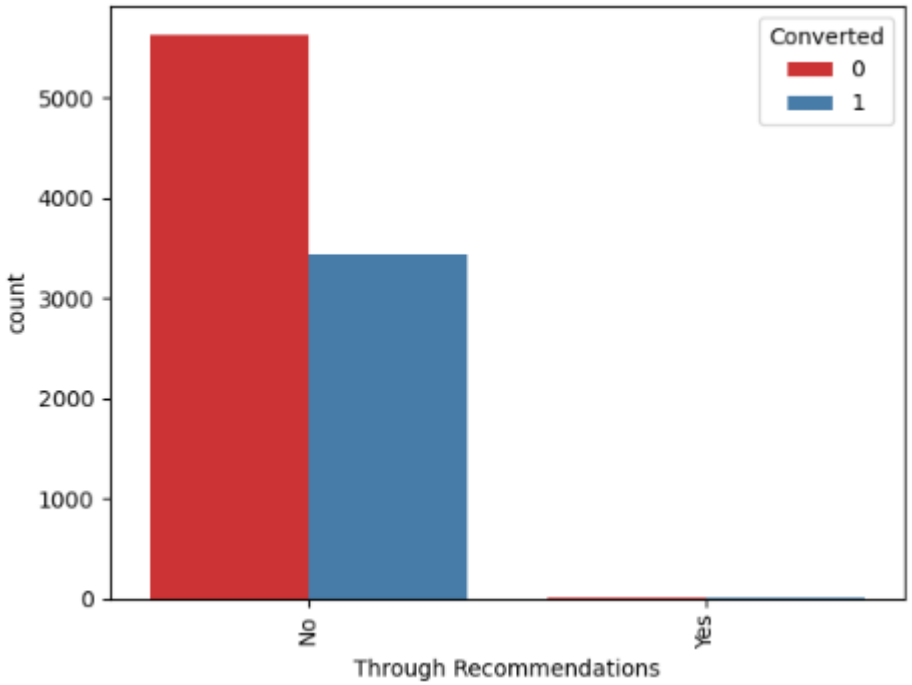
LEAD ORIGIN VS CONVERTED

Landing page submission has had high lead conversions.



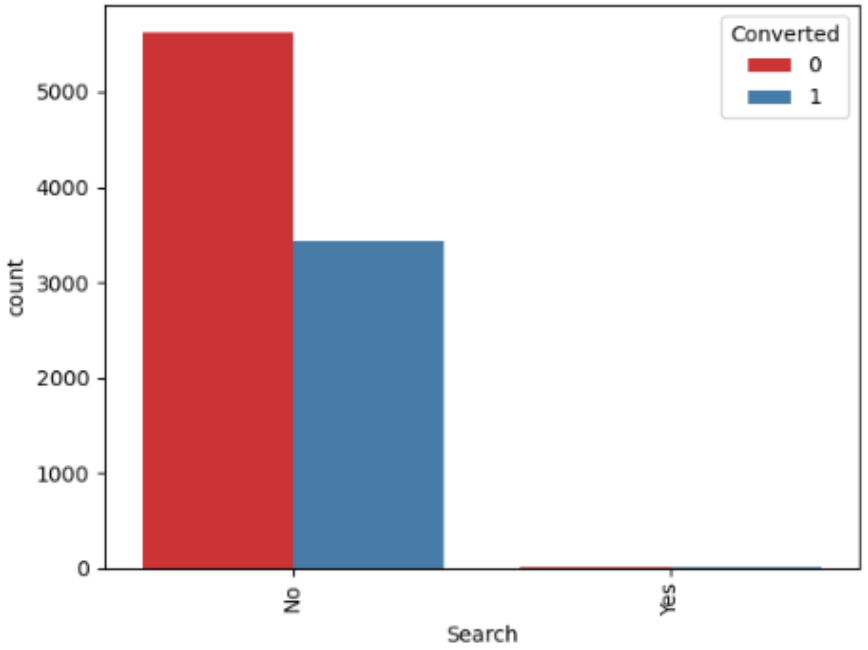
DIGITAL ADVERTISEMENTS VS CONVERTED

Based on the above graph digital advertisements do not have promising leads.



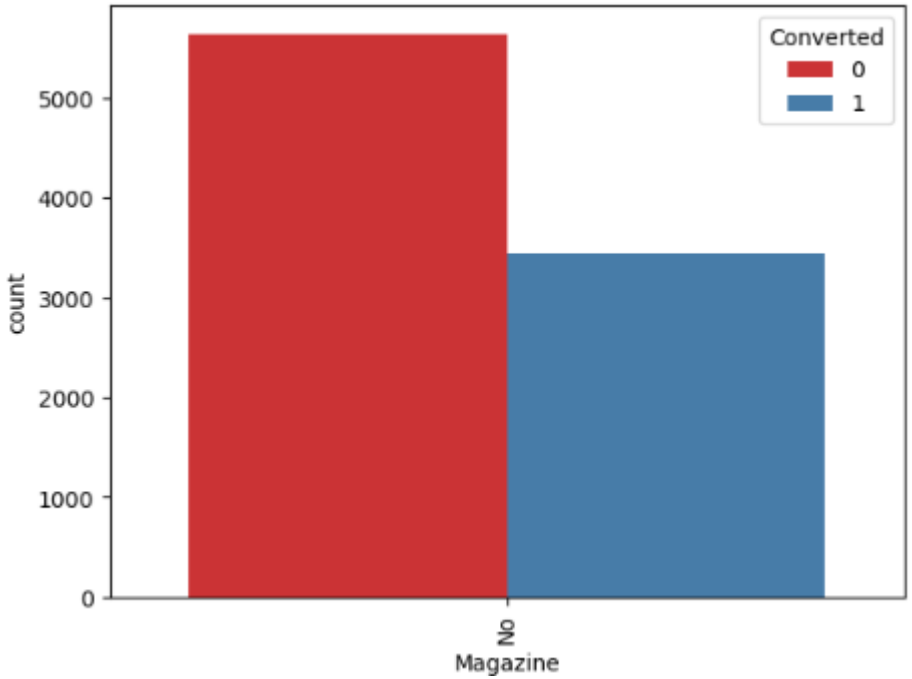
THROUGH RECOMMENDATIONS VS CONVERTED

From the above graph, recommendation are not a good source from promising leads.



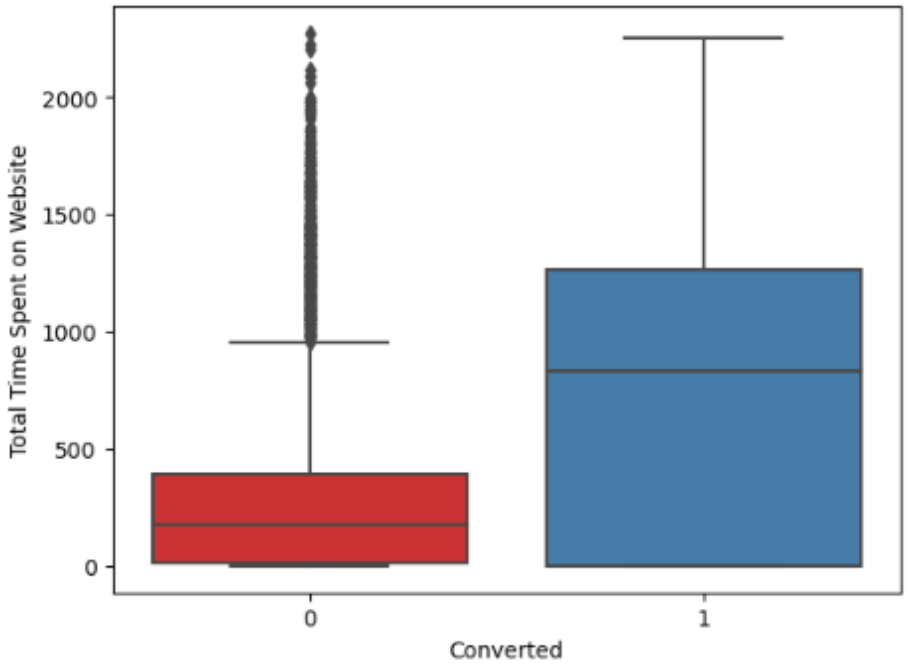
SEARCH VS CONVERTED

The above graph show search are not good source of leads.



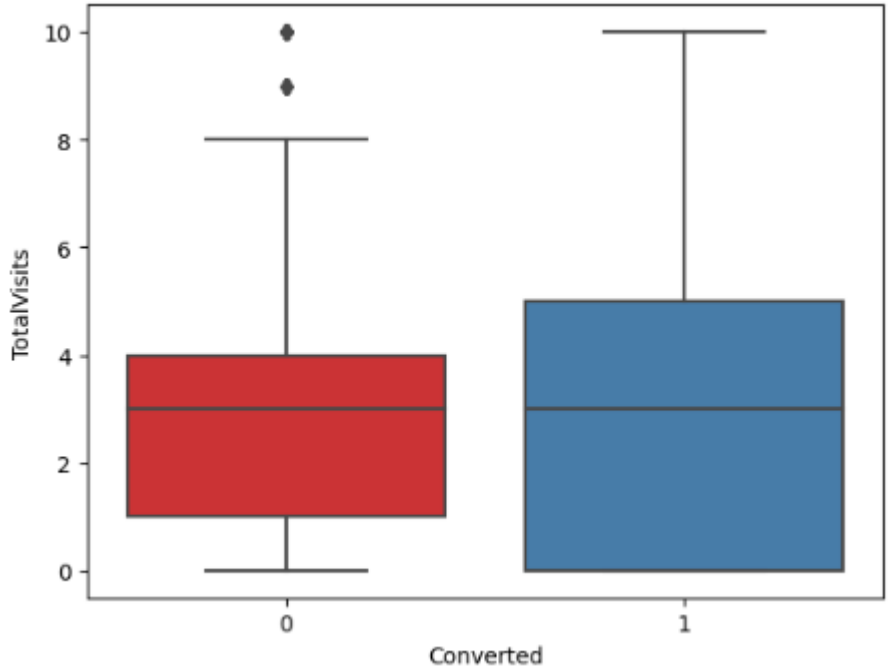
MAGAZINE VS CONVERTED

Magazine do not have higher conversion rate.



TOTAL TIME SPEND ON WEBSITE VS CONVERTED

People spending higher than average time are promising leads.

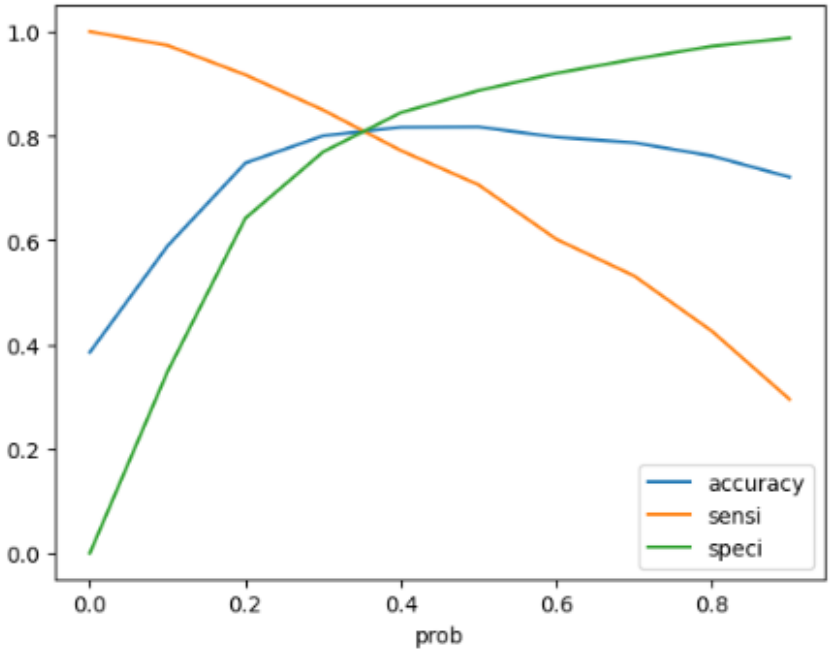


TOTAL VISITES VS CONVERTED

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

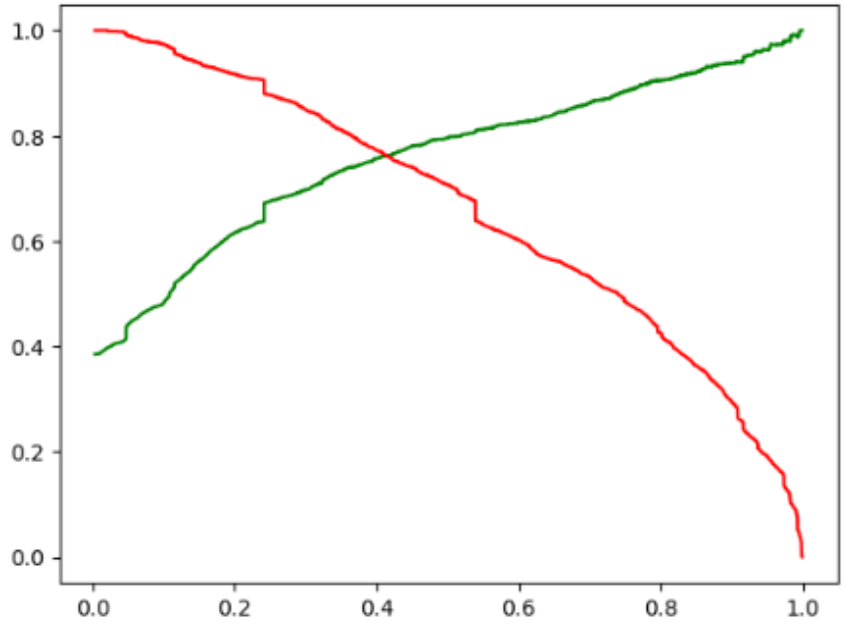
MODEL BUILDING

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



ACCURACY SENSITIVITY AND SPECIFICITY

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



PRECISION AND RECALL

- 73.4% Precision
- 77.6% Recall

CONCLUSION

EDA:

- People spending higher than average time are promising leads, so targeting then and approaching them can be helpful in conversion.
- SMS messages can have a high impact on lead conversion.
- Landing page submission can help find out more leads.
- Marketing management , human resource management has conversion rates. People from these specialization can be promising leads.
- Reference and offers for referring a leads can be good source for higher conversions.
- An alert message or information has seen to have high lead conversion rate.

Logistic Regression Model:

- The model show high close to 81% accuracy
- The threshold has been selected from accuracy, sensitivity, specificity measures and precision. Recall curves.
- The model show 76% sensitivity and 83% specificity
- The model finds correct promising lead and leads that chance of getting converted
- Overall this mode proves to be accurate

THANKS.