Lead Scoring Case study

Logistic Regression

Problem Statement

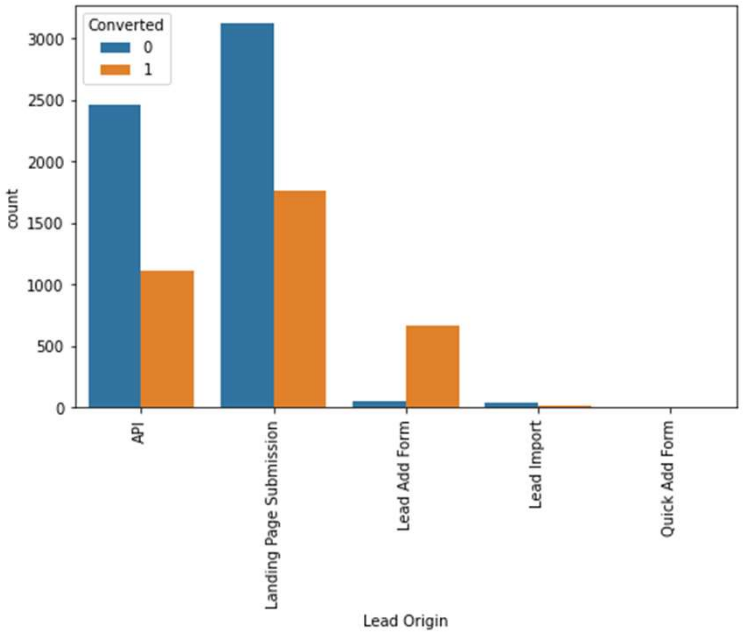
- An organization called X Education provides online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- Once the people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%. The company wishes to identify the most potential leads, also known as ‘Hot Leads’.

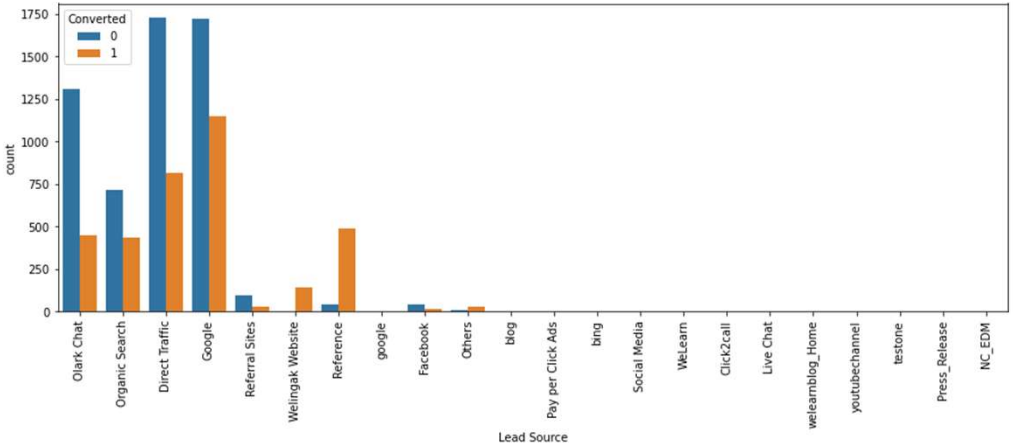
Business Goal

- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.
- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Strategy

- Source the data for analysis
- Clean and prepare the date
- Exploratory Data Analysis
- Feature Scaling
- Splitting the data into Test and Train Data Set
- Building the Logistic Regression Model and calculate Lead score
- Evaluating the Model by using different metrics – Specificity and Sensitivity or precision and recall
- Measure the accuracy of the model

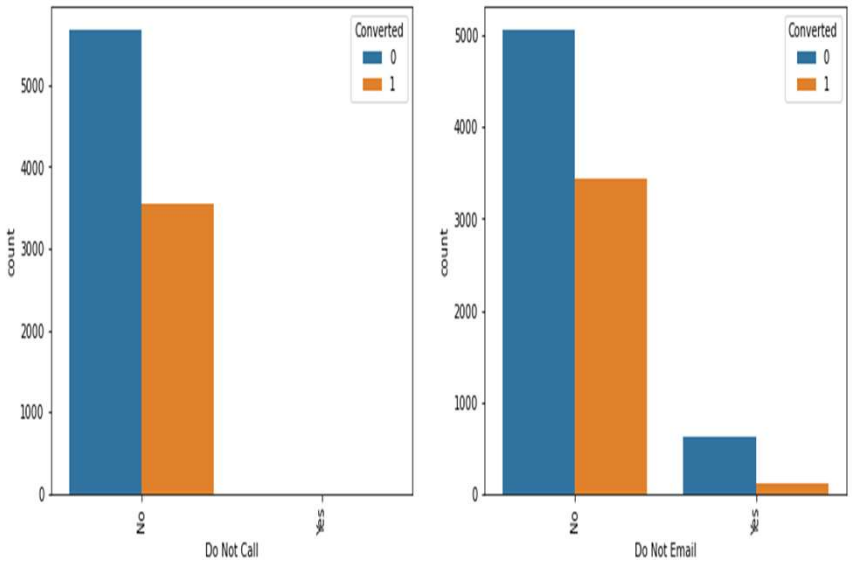




Exploratory Data Analysis

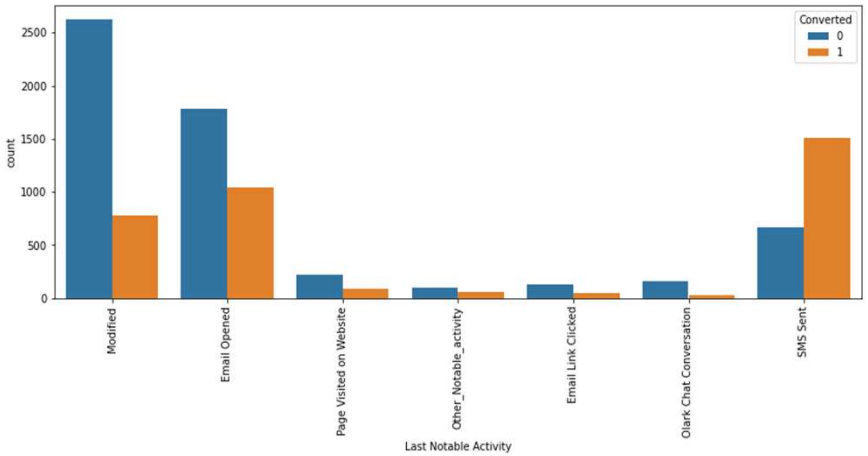
- Maximum number of leads are generated by Google and Direct traffic.
- Lead Add Form has a very high conversion rate but count of leads are not very high.

Do not Call and Do not Email



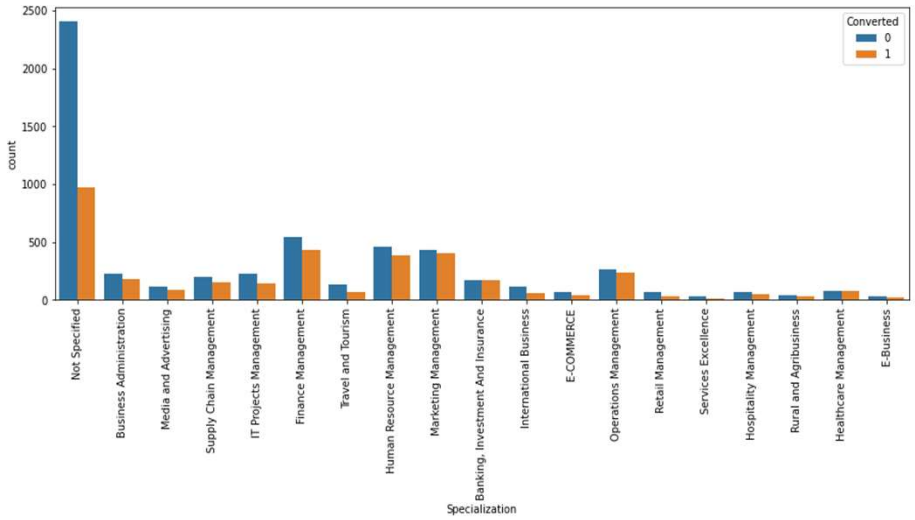
- There is no help from frequent calls and emails.
- We can use these options in times when there is no emergency situation for sales teams.
- Probably for providing the information about the product to the customers.

Last Notable Activity



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

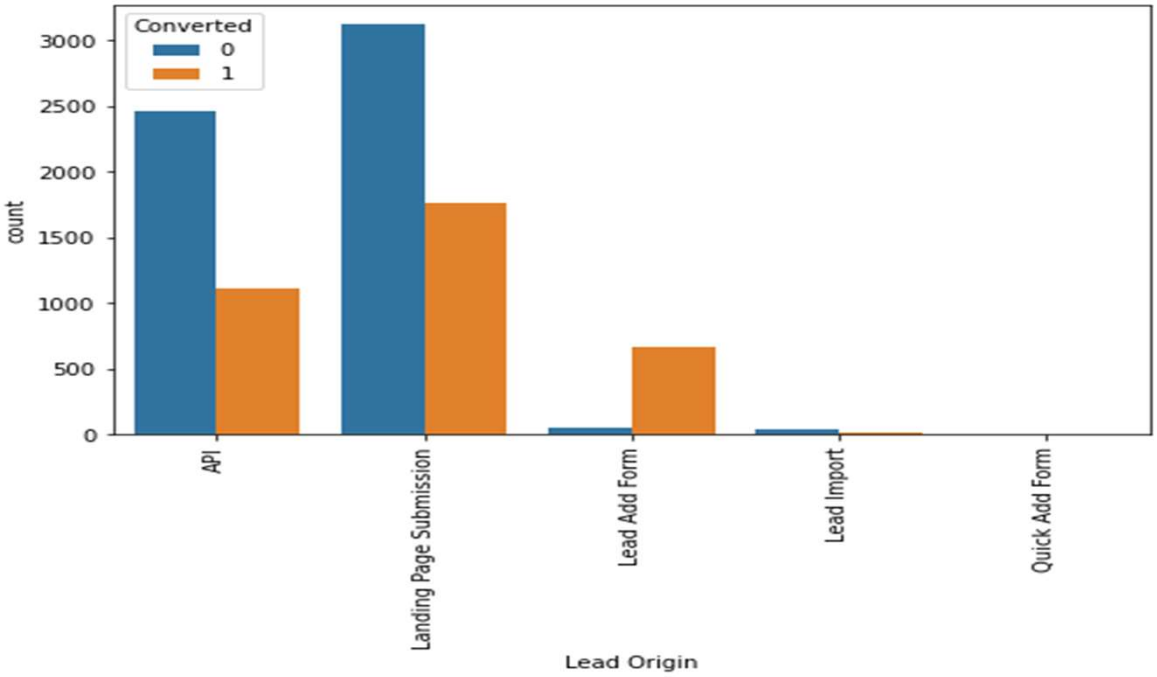
Specialization



- We see that specialization with Management in them have higher number of leads as well as leads converted.

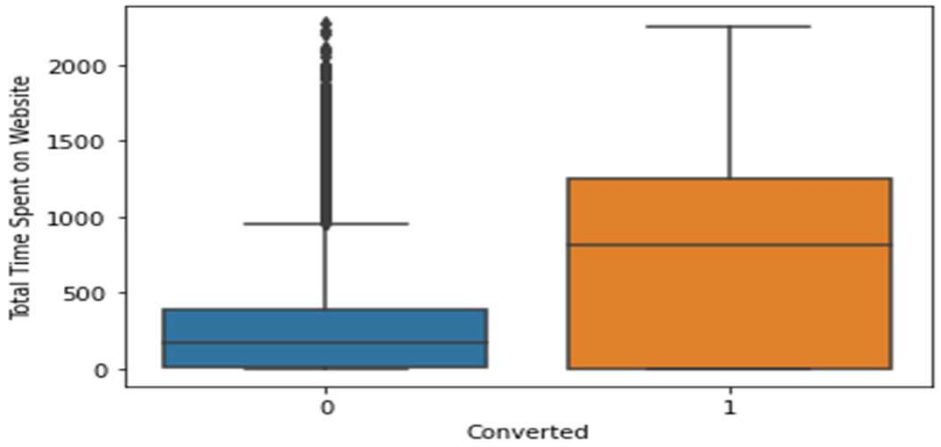
Lead origin

API and Landing page submission bring higher number of leads as well as conversion.



Total time spent on website

Leads spending more time on website are more likely to be converted.



Model Building

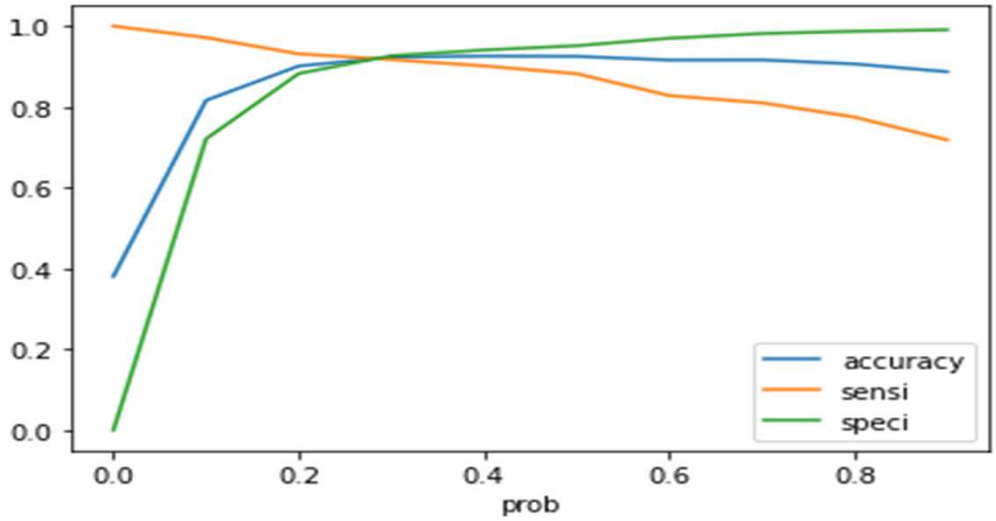
- Splitting into train and test set
- Scale variables in train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate variables based on high P – values
- Check the VIF values
- Predict using the train set
- Evaluate accuracy and other metrics
- Predict using Test set
- Precision and recall analysis on test predictions

Model Evaluation (Train)

Accuracy – 92.29%

Sensitivity – 91.70%

Specificity – 92.66%



Model Evaluation (Test)

- Accuracy – 92.78%
- Sensitivity – 91.98%
- Specificity – 93.26%

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.

Logistic Regression Model:

- The model shows high close to 92% accuracy
- The model shows 92% sensitivity and 93% specificity.
- The model finds correct promising leads and leads that have less chances of getting converted
- Overall this model proves to be accurate.