

Lead Score Case Study

Submitted By:- Varun Sharma



Problem Statement

- X Education sells online courses to industry professionals.
- In order to increase the lead conversion rate, the company first should identify the most potential leads
- The Company wants to increase their profit to 80%.
- The lead conversion rate should go up as the sales team, so that they focusing more on communicating with the potential leads rather than making calls to everyone.



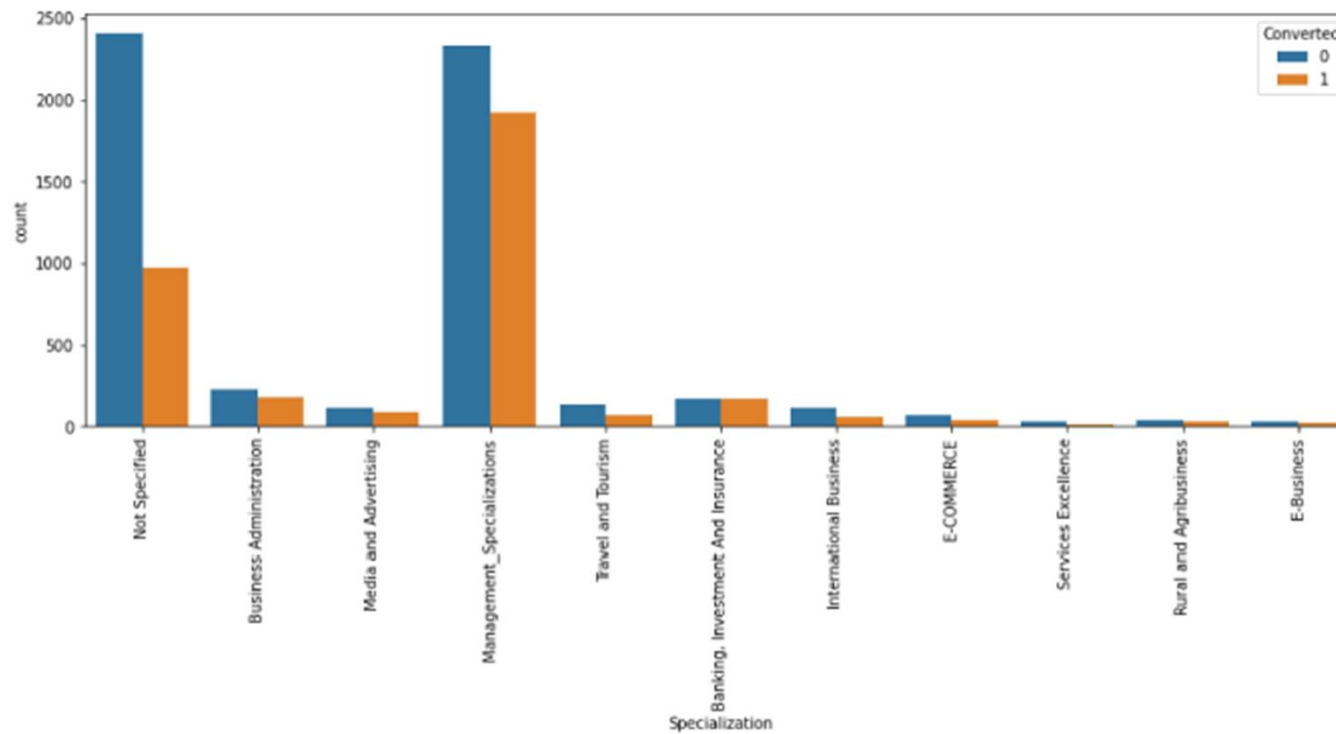
Goals and Solutions

- First of all we need to perform data cleaning and data manipulation on given DataSet:
- Handle the duplicates in data.
- Finding null and Missing values .
- Drop the unnecessary columns who contains large missing values then finding outliers for given variables.
- **Exploratory Data Analysis-** Univariate & bivariate
- Creating Dummy variables and Scaling
- Building logistics model and removing variables whose have high P-value.
- Making Predictions in the Model
- Finding best output values for Accuracy , sensitivity and specificity.

Top Variables impacting the Conversion

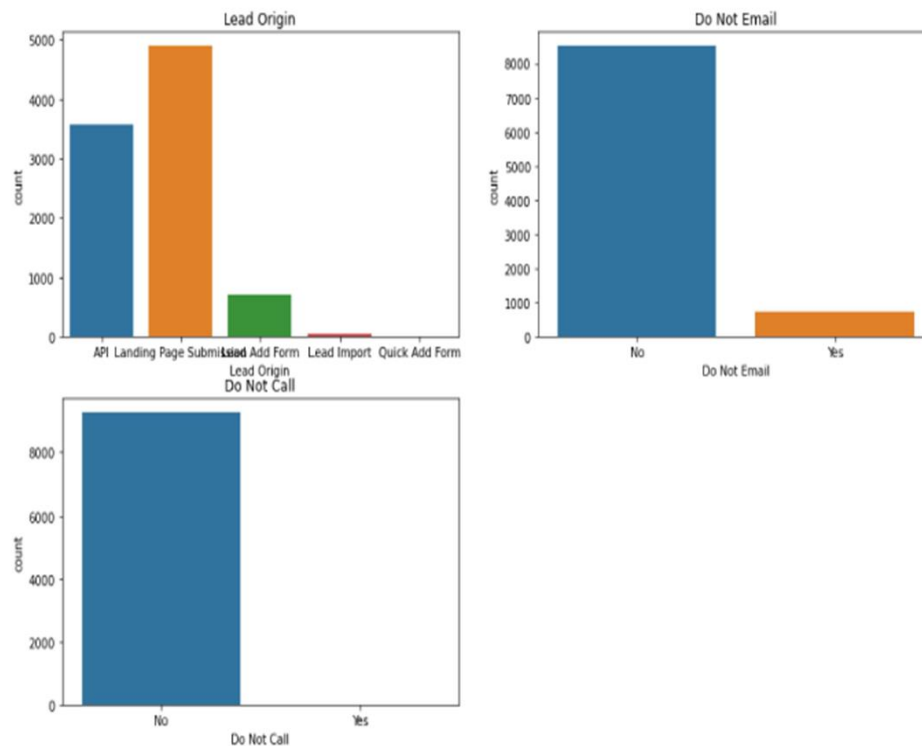
const
Lead Source_Welingak Website
Tags_Already a student
Tags_Closed by Horizzon
Tags_Diploma holder (Not Eligible) -
Tags_Graduation in progress
Tags_Interested in full time MBA
Tags_Interested in other courses
Tags_Lost to EINS
Tags_Not doing further education
Tags_Ringing
Tags_Will revert after reading the email
Tags_invalid number
Tags_number not provided -
Tags_switched off
Tags_wrong number given -

Variable based on converted values

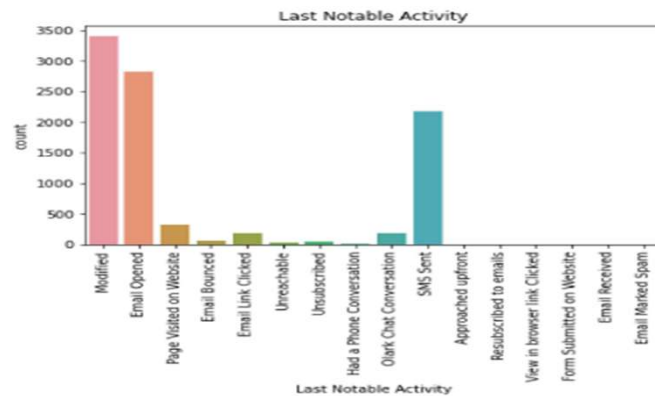




Variables shows Graphical Representation

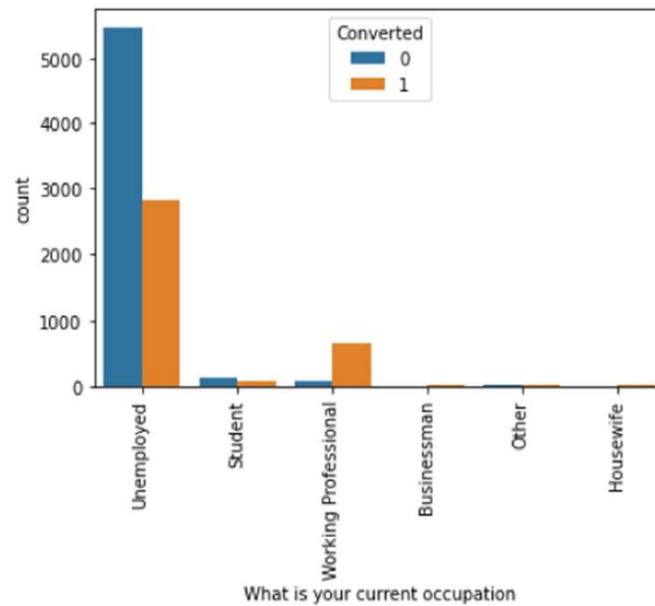


Leading Origin
Do not Email
Do not Call

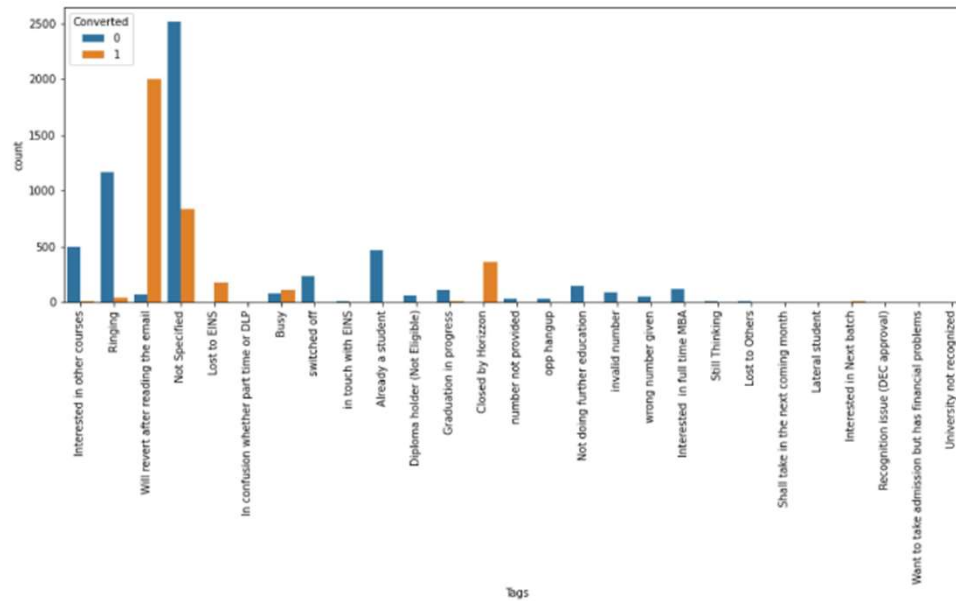


Variable Graphical analysis

- Last Notable Activity
- What is your current Occupation

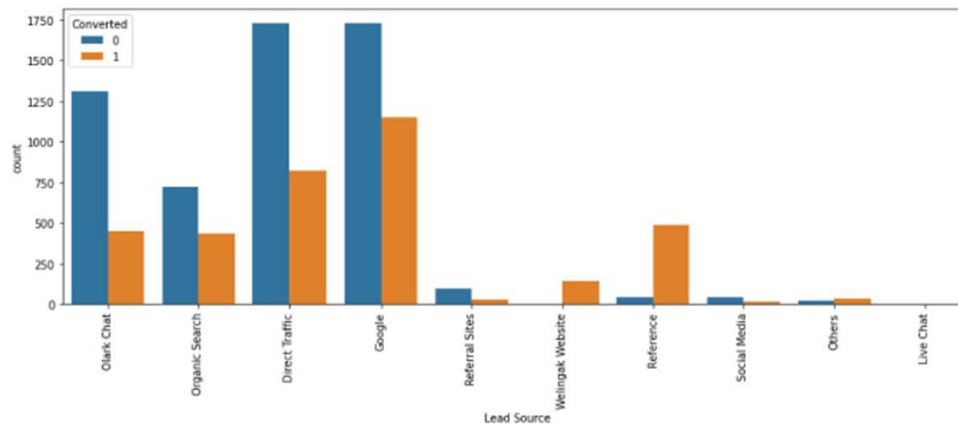


Variable Graphical analysis

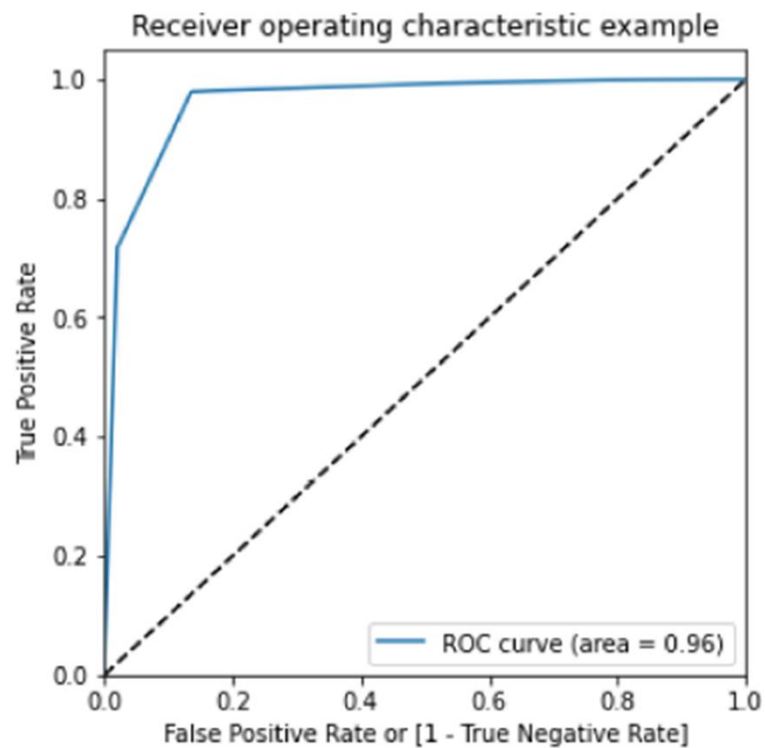


Tags – Highest values shows in Not Specified and Will revert after reading the email

Lead Source-



Lead Source- Max number
are for google and direct
traffic
Lead in welingak is high



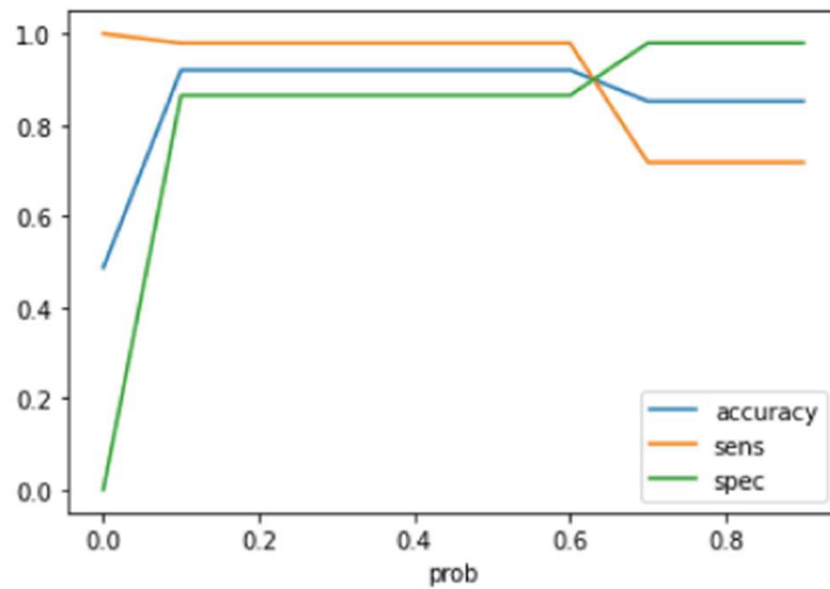
ROC Curve

To achieve a conversion rate of around 96%

The positive rate cutoff is around 0.3.

Hence using this model we achieve a right pint having around 0.3 probability prospect

Accuracy , sensitivity & Specificity



Final ,Graphical representation of Accuracy , sensitivity & Specificity of logistic model.



Conclusion

It was found that the variables that are the most of potential buyers are: The lead source was:

- Google
- Direct traffic
- Organic search
- Welingak website

The last activity from observation are:

- SMS
- Olark Chat conversation

The total time spend on the Website.

Total number of visits.