# Lead Score Case Study

Using Logistic Regression

# About Business

- An education company named X Education sells online courses to industry professionals
- This company markets its courses on several websites and search engines.
- When people come to their website they fill up a form provides their email address or phone number and there conversion rate is 30%

# Problem with X Education

- Although X Education gets a lot of leads, its lead conversion rate is very poor
- To make this process more efficient, this 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.

# Expectation from Us

- Build a model wherein we 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.

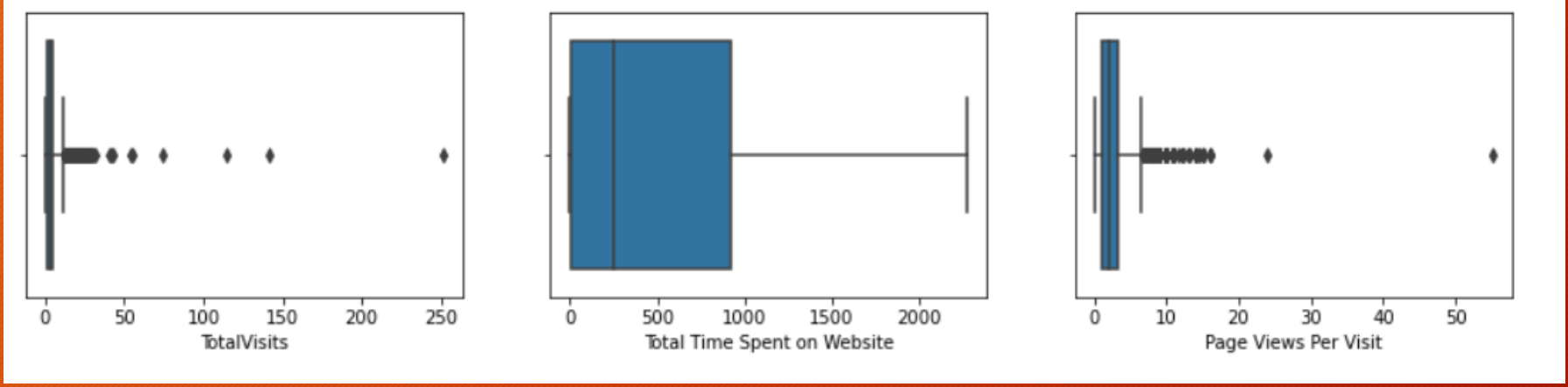
- The CEO, in particular, wants the target lead conversion rate to be around 80%.

# Our Analysis Approach

- We will build a logistic regression model to find whether a lead will convert or not based on the data filled by the allthe leads who fill the form
- Our major metrics will be sensitivity as we don’t want to miss any lead who has potential to convert and it should be about 80%
- In business terms 80% conversion rate

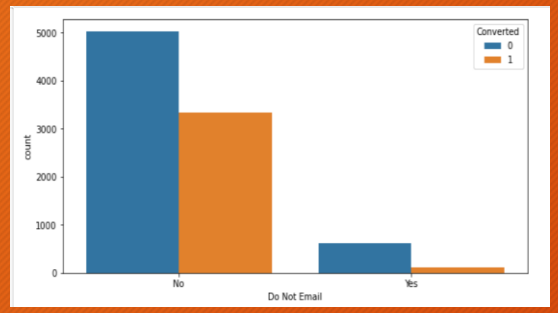
# Some inshights of data

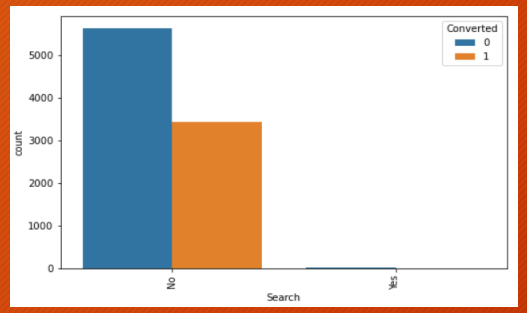
- There are few outliers in the data which can impact our analysis so it is required to remove those by capping it with 95 percentile value.

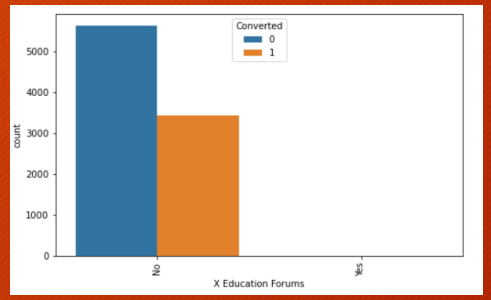


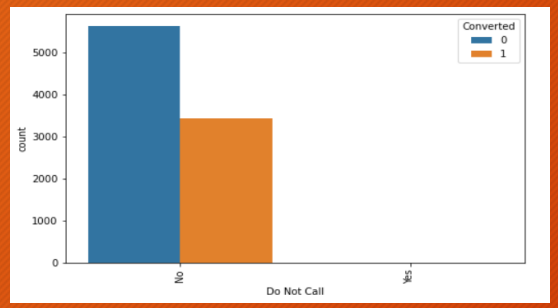
# Skewed value for some feature

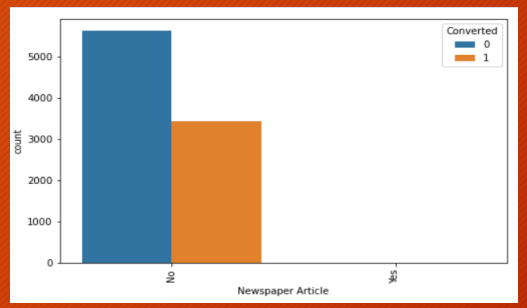
- There are many fratures which has skewed value so its good to drop all such fratures as they will not help in our analysis and model building.

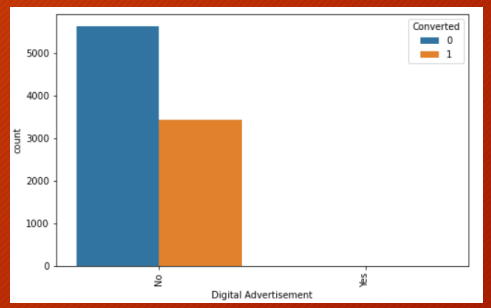


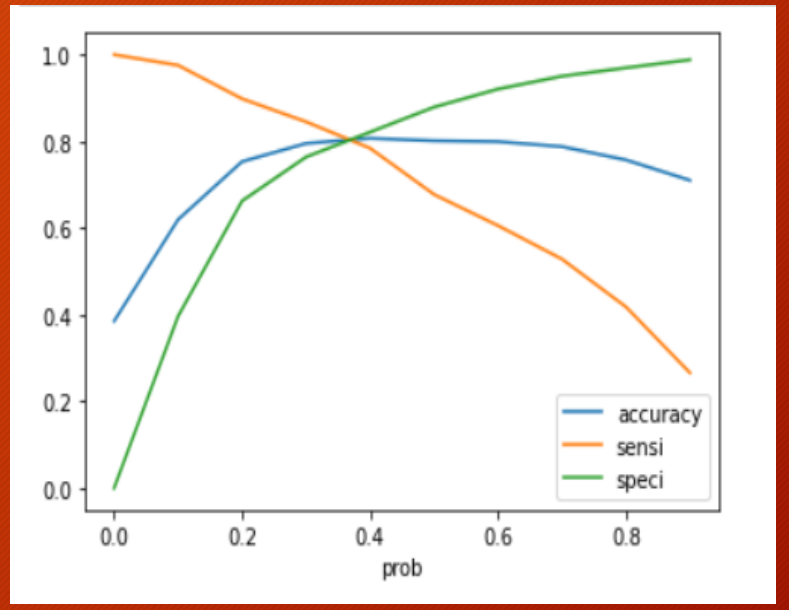



# Cut-off probability

- After performing the model and analysing the few metrics we came at a point that a cut-off probability should be 37.5

- According to the model if the predicted probability is greater than 37.5 the lead will convert



# Top 3 variables

- We have prepared multiple models and the final model is one with high sensitivity above 80%.

- The top 3 variable which the X Education should focus on are
- Lead Origin_Lead, Occupation, Last Activity

- Top three categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion are:
- Lead Origin_Lead Add Form, Lead Origin_Lead Import, Occupation_Other_occupation

# Result

- As a result of the modelling we 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.

# Recommendation

- At the time when the number of sales person is more try calling to all the leads whose score is greater than 40

- At the time when the company don’t want to spend much time and effort on calls go with higher lead score only