

## Summary

- Lead score assignment was done using logistic regression as a choice of model building method
- The problem statement was of X education company which runs online education platform, and want the customers to be onboarded as soon as possible and increase the chances of converting the cold lead to hot lead
- Most number of leads are from India and in terms of city highest number was from Mumbai
- Few columns had 'Select' as a value which represents the customer while filling the form hasn't selected any value, e.g., occupation, background, city, etc. In order to get most information about the data there should be compulsion of selection from the drop downs
- A lot of columns show biased data, as only one value from given has been selected such as Magazine, Newspaper, Digital advertisement, etc, and these columns doesn't support or contribute towards the model building
- There are some outliers' data points in total visits as some customers spent a lot of time on website which skewed the data
- Most leads' current occupation was mentioned as unemployed; hence it states that the unemployed category are the most likely ones to join the program
- SMS Sent value in last notable activity supports major lead conversion than Email opened and modified. Hence, company should focus more on sending sms as many people might not have access to emails
- The reference as a lead source as provided better conversion rates compared to others as lead source
- A lot of tags have shown no major contribution towards conversion except will revert after reading email
- A lot of customers opting for the course had selected better career opportunities as a joining reason