# Summary

**Submitted By – Prathamesh Kulkarni** **April 2024**

## Assignment Objective:

Objective of this assignment was to use historical data from X Company, which offers educational courses online, to objectively predict which of the potential buyers (leads) of courses can be converted to actual paying customers.

## Approach:

1. This was a binary classification problem, which required use of Machine Learning technique to classify leads into good leads and bad leads.
2. Besides, we also needed to come up with a method to score each of the leads, so that the company can decide which leads can be targeted, depending on their chances of converting.
3. To tackle this classification problem, we used Logistic Regression Machine Learning model based on the historical data, where previous potential buyers were already classified as converted lead (1) or bad lead (0).
4. We neglected certain outlying or incomplete data and managed to use approximately 99% of historically available data to build the machine learning model.
5. After going through multiple iterations of building the model and evaluating its performance on the data that the model had not seen before, the final revision of the model predicted and used the most significant indicators of good or bad leads in the unseen data, with approximately 84% accuracy.

## Key findings:

1. Historically, the company had gathered data of 9240 potential course buyers.
2. A large number of attributes from this data had missing values, which we managed to either ignore or create, based on those attributes' significance in this analysis.
3. We also had to further treat the data so that some of the data attributes can be objectively used by the Machine Learning model. For example - some attributes such as the tags assigned to potential buyers had categorical information in text format, which we converted to machine readable numerical format.
4. We ignored data of a few potential buyers, because that data was statistically insignificant for this analysis due to certain unusual (outlying) information in that data. For example - certain customers visited X Company's website an unusually high number of times, which, if considered during this analysis, would have been detrimental to the final results that we got.

# Summary

**Submitted By – Prathamesh Kulkarni** **April 2024**

## Learnings:

1. The final ML model uses the following data attributes to assign Lead score to any customer:
    a. Tags assigned to the customers indicating the current status of the lead, such as whether the customer is going to revert after reading the email.
    b. The source of the lead, particularly the leads coming from ‘Welingak Website’,
    c. The last activity or last notable activity of the customer, such as visiting the company’s website pages,
    d. Lead Quality assigned to the lead,
    e. Asymmetrique Activity Index assigned to the customer based on their activity or profile.
2. Lead score of 34 is optimum for deciding class of a lead. Typically, beyond the score of 34 a lead can be considered good, with increasing likelihood of conversion.
3. Lead score threshold can be increased or decreased to gain more or less traction, as business demands seasonally.