**Project Report**

# What is the problem you want to solve?

The boom of internet has given e-commerce websites accessibility to a large audience. But, this increase in e-commerce usage over the past few years has created potential in the market but the conversion rate is not as high.

Aim is to investigate the conversion rates, and understand what is happening.

# Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis?

The client is an e-commerce website looking to understand the behaviour of website visitors and proactively determine if the visitor is likely a potential customer and is going to generate revenue or not.

# What data are you using?

Visitor’s website activities logged by the company. In this case, dataset is from Kaggle.

1. Brief outline on solving this problem.

This is a classification problem.  'Revenue' will be used as the target/dependent variable, the values are either ‘True’ or ‘False’. ‘True’ value means revenue is generated by the visitor and ‘False’ value means revenue is not generated. Examples of other type of columns in this data set are type of visitor, page value, special day etc.

But the data is imbalanced and hence would be using confusion matrix and F1 score as performance matrix. We would also need to resample test and train data so that the models do not overfit or under fit.

# What are your deliverables?

Creating a model that would classify new visitors. With a higher F1 score, of around 70%.