**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.

# 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 classify new visitors would generate revenue or not.

# What data are you using?

We would be using visitor’s website activities logged by the company. In this case we are using a Kaggle dataset.

1. Brief outline on solving this problem.

This is a classification problem. The column that we would be using to classify would be ‘Revenue’ where the values are either ‘True’ or ‘False’. 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. This model would require to have higher value of True Positive and lower value of False Negative.