**Student Names:**

Leonel Jerez

**Data Info:**

<https://www.kaggle.com/ntnu-testimon/paysim1>

Synthetic dataset of mobile money transactions. Each step represents an hour of simulation.

Step -Maps a unit of time in the real world. In this case 1 step is 1 hour of time.

Type -CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER

Amount -amount of the transaction in local currency

nameOrig -customer who started the transaction

oldbalanceOrg -initial balance before the transaction

newbalanceOrig -customer's balance after the transaction.

nameDest -recipient ID of the transaction.

oldbalanceDest -initial recipient balance before the transaction.

newbalanceDest -recipient's balance after the transaction.

isFraud- identifies a fraudulent transaction (1) and non fraudulent (0)

isFlaggedFraud- flags illegal attempts to transfer more than 200.000 in a single transaction.

**Down sampling method:**

1. *Half of the size*: I read the number of rows and put them into an array line by line. I then divided the array index by 2 and created a new file. The file was populated to half the size of the original array and stopped there.
2. *Under 100 MB*: I opened the file using a file reader and found its length (this returns the file size in bytes). I then made a variable that would signify 100 MB in bytes. Once I had the two, I started populating the array with the rows again. I told the program to stop once it reached 100 MB.

**3 vis tasks:**

1. Check the probability of frauds according to payment type using Type and isFraud.
2. Check the correlation between transaction amount and if it’s a fraud using Amount and isFraud.
3. Check how often a fraud is done using Step and isFraud to have a probability of their occurrence.