Data Science Coding Challenge

Clustering

Version 1.0

**Problem Summary**

Attached you will find a data set sampled from three sensors (dataset.csv.zip). Using whatever techniques are appropriate construct a model that will detect clusters in the data stream.

**Hints**

* You should consider this a ‘real world’ data set, so it may have problems. Your solution should articulate those problems (if any) and have a strategy for dealing with them.
* Your model, data cleaning, and any data exploration should be coded in R or Python.
* Visualizations and explanations are worth as much as code. A working model is good, a defensible model is better.
* Making your code executable by the evaluator is extremely helpful. Your solution should include any setup instructions needed.
* It is important that the evaluator be able to reproduce the results of your model.

**Submitting Your Results**

In a zip file submit the following files:

* The original dataset.csv file
* The cleaned dataset in a file named clean\_dataset.csv
* The data cleaning code (if any) in a file named clean.(py | ipynb) or clean.R
* The data exploration code (if any) in a file named exp.(py | ipynb) or exp.R
* Any plots or charts produced in PDF, word, png or other format
* The model code in a file named model. (py | ipynb) or model.R
* The output of the model in a file named results.csv, each row of the file will contain one point and it’s associated cluster number. It should be formatted like this (cluster #, x,y,z) e.g. (1, 0.8234,0.9932,-1.001)