CMG’25 Hackathon guidelines

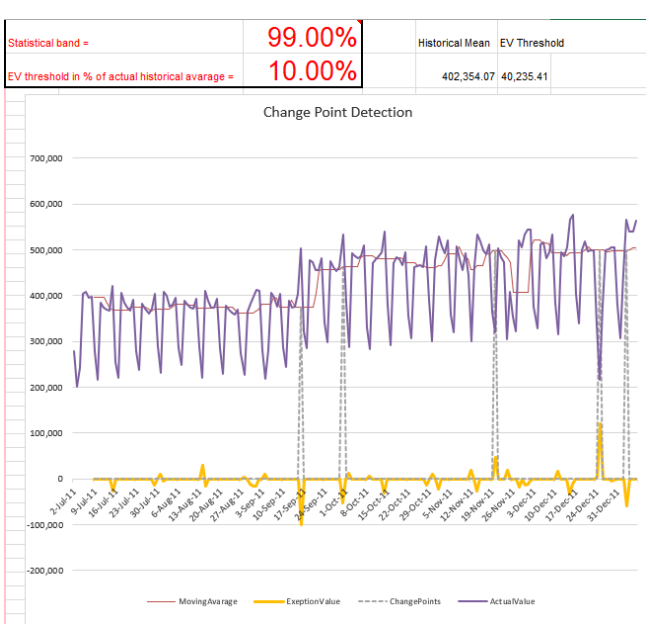
ANNONCE (to publish before conference)

The task is to find some change points and/or anomalies in the given time stamped data to see different phases/patterns .

Participants can use any tools or libraries/packages (in R, Python and so on) to detect change points and/or anomalies in the data.

Or they can use the free Change Points detection API (e.g. via Postman.com) described in Change Point Detection is implemented in the free web tool Perfomalist
Particularly the instruction on how to use that is described here.

To visualize the result any spreadsheet charting could be used or other means (e.g. python, R and so on). Examples are on the following picture:





The time to work on the task - 3 hours.
The result would be judged by CMG experts and the winner will get an award and time to make a short presentation.

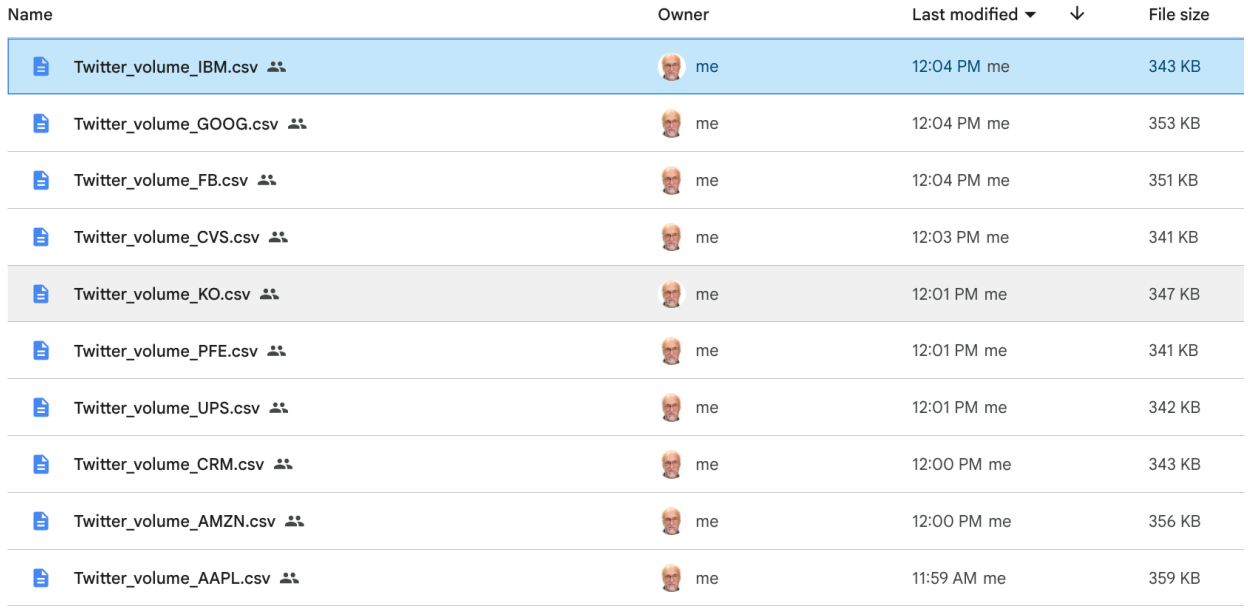
Vendors of similar tools are welcome to participate.

If participants would like to use MATLAB for anomaly detection tools, we have a licensed version available here:

URL for event page: https://www.mathworks.com/licensecenter/classroom/4866200/
1. this will take you to a MathWorks account Sign-in page.
2. Create an account or use an existing account
3. Press "Access MATLAB Online" button followed by "Open MATLAB Online"

On site activities

Data to test: https://github.com/numenta/NAB/tree/master/data
Particularly the following csv files:



Name	Owner	Last modified	File size
Twitter_volume_IBM.csv	me	12:04 PM me	343 KB
Twitter_volume_GOOG.csv	me	12:04 PM me	353 KB
Twitter_volume_FB.csv	me	12:04 PM me	351 KB
Twitter_volume_CVS.csv	me	12:03 PM me	341 KB
Twitter_volume_KO.csv	me	12:01 PM me	347 KB
Twitter_volume_PFE.csv	me	12:01 PM me	341 KB
Twitter_volume_UPS.csv	me	12:01 PM me	342 KB
Twitter_volume_CRM.csv	me	12:00 PM me	343 KB
Twitter_volume_AMZN.csv	me	12:00 PM me	356 KB
Twitter_volume_AAPL.csv	me	11:59 AM me	359 KB

LINK TO THE FOLDER WITH DATA IS HERE

EXAMPLE 1

Simple case: art_daily_flatmiddle.csv

Tool is performalist.com Change Point detection API described HERE

1st step to change the format of the data (using EXCEL or Google sheet means):
From original:

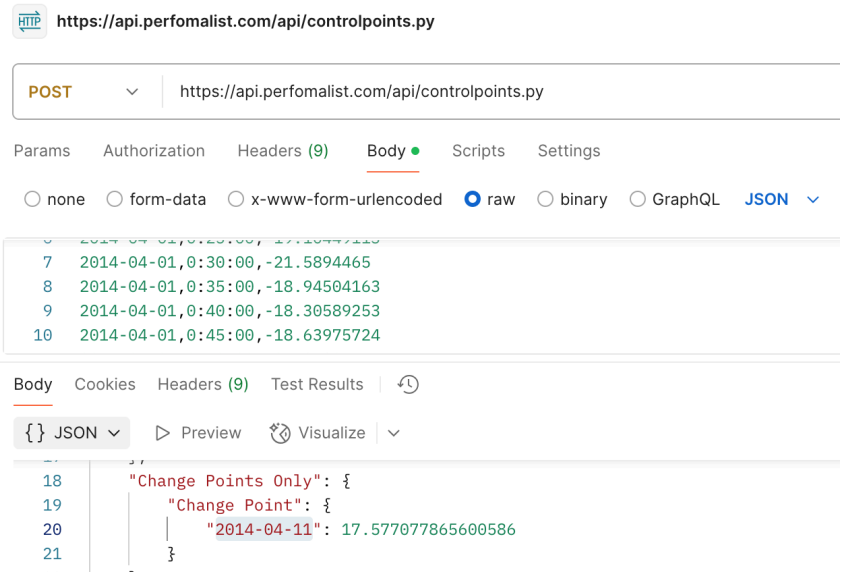
```
imestamp,value
2014-04-01 00:00:00,-21.0483826823
```

```
2014-04-01 00:05:00,-20.2954768676
2014-04-01 00:10:00,-18.127229468299998
2014-04-01 00:15:00,-20.1716653997
```

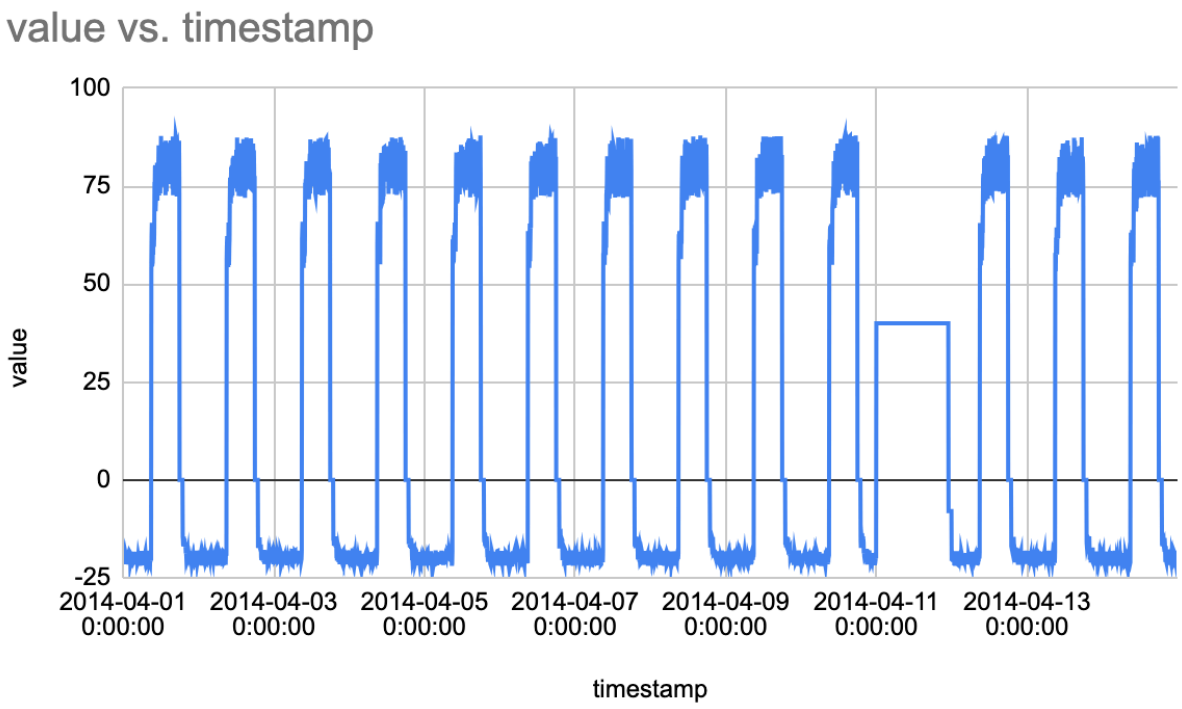
TO performalist form:

```
date,time,value
2014-04-01,0:00:00,-21.04838268
2014-04-01,0:05:00,-20.29547687
2014-04-01,0:10:00,-18.12722947
2014-04-01,0:15:00,-20.1716654
```

Then call perfomalist API (e.g. using postman.com)



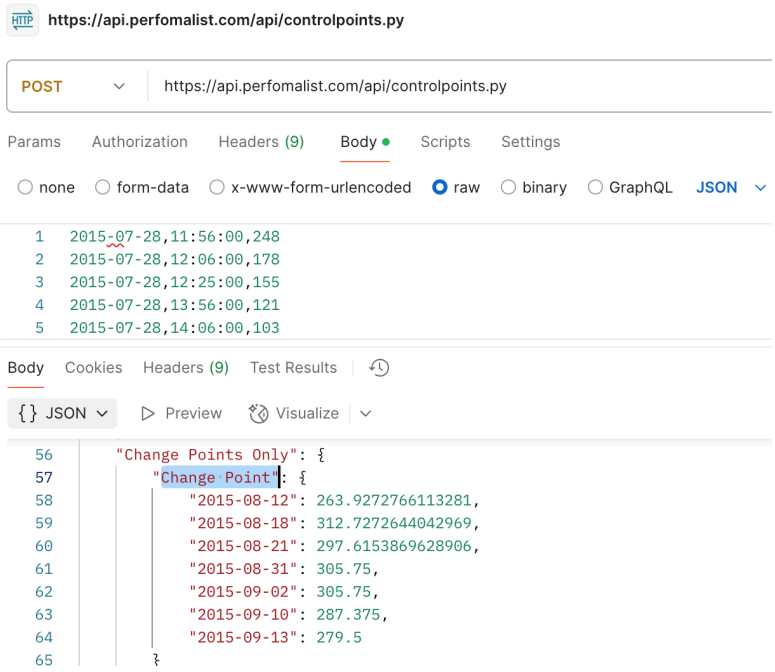
The result is one change point on 2014-04-11, which can be easily validated by building the spreadsheet chart (see below):



EXAMPLE 2

More difficult case:
https://github.com/numenta/NAB/blob/master/data/realTraffic/TravelTime_451.csv

After repeating the above steps (see Example 1) the result should show several change points:



To reduce the number of change points one can explicitly to provide as a 1st 3 lines in the data the following tuning parameters:

- **sValue** - Statistical band in %, where 100 is UCL=MAX, 0 is UCL=LCL=mean). (**normality)**
- **eValue** - Exception Value (EV) threshold in % of actual historical average. (**insensitivity)**
- **BaseLineLength** - The time period to compare current value against.

After adding there

```
sValue, 99

eValue, 20

BaseLineLength , 7
```

The API returns only 3 change points:

https://api.perfomalist.com/api/controlpoints.py

POST https://api.perfomalist.com/api/controlpoints.py

Params Authorization Headers (9) Body Scripts Settings

none form-data x-www-form-urlencoded raw binary GraphQL JSON

```
1  sValue, 99
2  eValue, 20
3  BaseLineLength , 7
4  2015-07-28,11:56:00,248
5  2015-07-28,12:06:00,178
```

Body Cookies Headers (9) Test Results

{} JSON Preview Visualize


```
56    "Change Points Only": {
57        "Change Point": {
58            "2015-08-18": 312.7272644042969,
59            "2015-09-02": 305.75,
60            "2015-09-13": 279.5
61        }
```


Puting the API output to EXCEL or Google sheet one can visualize the result by showing phases in the data between change points (see below):

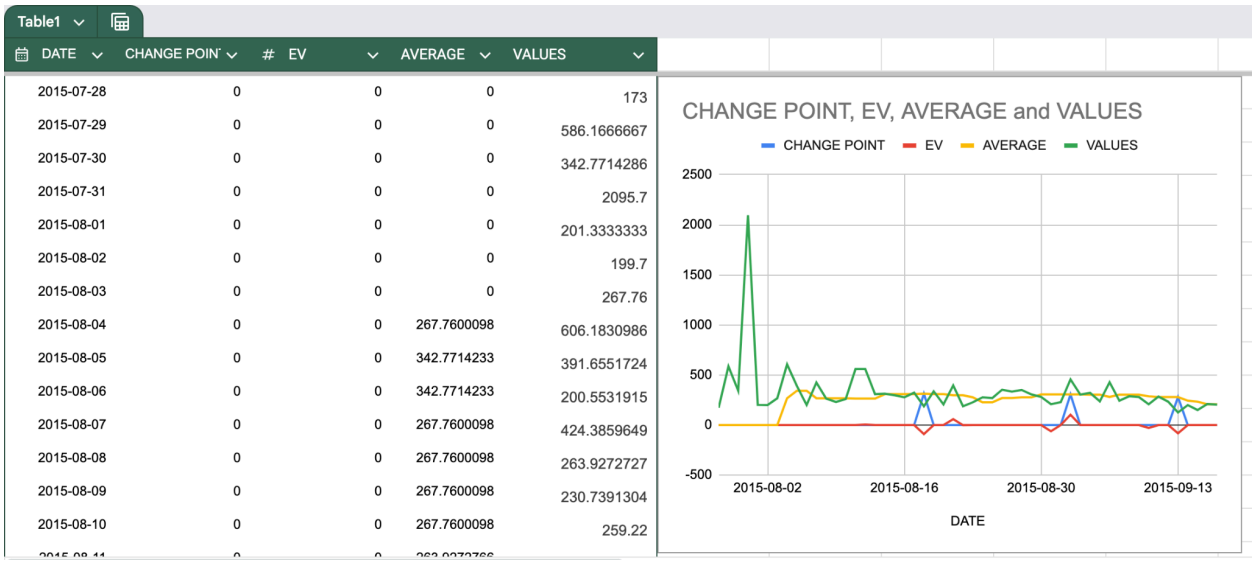


Table1

DATE	CHANGE POIN	# EV	AVERAGE	VALUES
2015-07-28	0	0	0	173
2015-07-29	0	0	0	586.1666667
2015-07-30	0	0	0	342.7714286
2015-07-31	0	0	0	2095.7
2015-08-01	0	0	0	201.3333333
2015-08-02	0	0	0	199.7
2015-08-03	0	0	0	267.76
2015-08-04	0	0	267.7600098	606.1830986
2015-08-05	0	0	342.7714233	391.6551724
2015-08-06	0	0	342.7714233	200.5531915
2015-08-07	0	0	267.7600098	424.3859649
2015-08-08	0	0	267.7600098	263.9272727
2015-08-09	0	0	267.7600098	230.7391304
2015-08-10	0	0	267.7600098	259.22