Name:Atta Ullah Khan ASUID:1217207992

Requirements: 1) python3.6 2)pip3

Install the following libraries before running this project, open Terminal

```
pip3 install numpy
pip3 install pandas
pip3 install seaborn
pip3 install sklearn
```

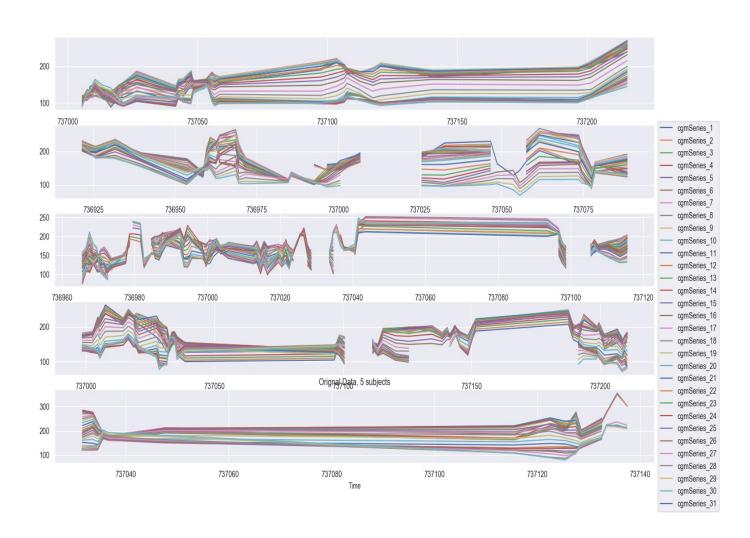
#### Tasks:

- a) Extract 4 different types of time series features from only the CGM data cell array and CGM timestamp cell array (10 points each) total 40
- b) For each time series explain why you chose such feature (5 points each) total 20 Answer to Both Part(a) and Part(b):

Let first Visualize the all the 5 sample with all the 31 cgm serieses ploted against its timestamp

```
python3 assignment1.py -m orignal
```

# Output will be:



Initially to extract 4 different type of features, we will use four methods 1:mean 2:Std(Standard deviation) 3: Max(Minimum) 4:Min(Minimum)

## Algorithm 1:

input = method, data ##possible values of method =[min, max, mean, std] 0

For each cgmSeries in {cgmseries1, cgmseries2.....cgmseriesN},

list = combine cgmseries of this type from all the 5 subjects excecute the method(Mean/std/min/max) on list

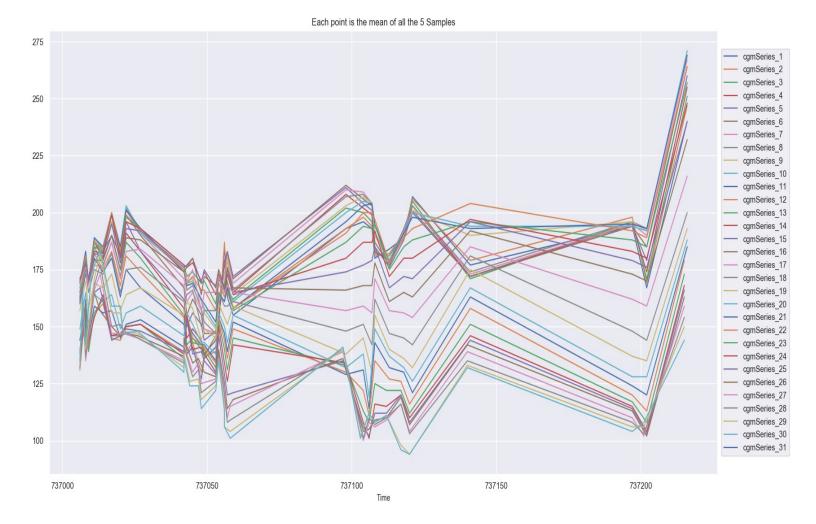
create a new data frame for this method return this new data frame;

this algorithm return that mean/std/min/max of all the 5 subjects

lets take the mean of all the 5 subject and plot it, run the following

python3 assignment1.py -m mean -p plot

output wil be Data plot for mean



This is the mean of all th 5 subjects

Now for mean lets extract the important features using corelation

Algorithm2, for extracting features:
method=input
data= call Algorthm1(method)
CorMatrix=find the correaltion matrix
selectedVariable =Select 1 variable from {cgmSeries}
independentVaribles={cgmSeries1,cgmseries2,.....N}-selectedVaribles

PotentialFeturesList=from the corMatrix Select all the variables such that for each variable V, CorMtrix[v,selectedVariable]>0.5

independent Varibles = Potential Fetures List-selected Varibles

CorMatrixIndependentV=Find correlation amngst independentVaribles

for each of the variables V1,V2 in independentVaribles: if CorMatrixIndependentV[V1,V2]>.9

Remove V1 From the PotentialFeturesList

return potentialFeatures;

In this algorithm we are slecting all the variable that are co related to the dependent variable, and droping all the independent variables which are co related

Lets extract the features from the Mean Data

python3 assignment1.py -m mean -f extract -e head

Cor relation matrix for mean:

#### Correlation Matrix

```
1.0
                       0.98 \ \ 0.96 \ \ 0.93 \ \ 0.91 \ \ 0.89 \ \ 0.85 \ \ 0.84 \ \ 0.84 \ \ 0.84 \ \ 0.8 \ \ \ 0.75 \ \ 0.65 \ \ 0.52 \ \ 0.39 \ \ 0.26 \ \ 0.120.0730.0530.017 - 0.08 - 0.14 - 0.18 - 0.120.0720.0540.0420.0520.048
camSeries 1
                   1 0.99 0.98 0.95 0.93 0.92 0.9 0.88 0.87 0.86 0.83 0.79 0.69 0.57 0.44 0.31 0.17 0.12 0.0950.059-0.040.0960.140.0860.0330.020.0110.0220.022
camSeries 2
cgmSeries 3
                             1 0.98 0.97 0.95 0.94 0.91 0.9 0.89 0.85 0.81 0.72 0.59 0.46 0.32 0.18 0.12 0.1 0.0640 0.280 0.71-0.11-0.050 0.0820 0.240 0.360 0.260 0.260
              0.96 0.98 1 1 0.99 0.98 0.97 0.95 0.93 0.91 0.91 0.86 0.83 0.74 0.62 0.49 0.35 0.21 0.15 0.13 0.093 00470.0340 0680.0130.0460.0630 0740.0680.067
              0.93 0.95 0.98 0.99 1 0.99 0.98 0.97 0.95 0.92 0.91 0.87 0.84 0.76 0.65 0.53 0.39 0.23 0.18 0.15 0.12 0.040 000 40.0360.0060.0580.0710.077 0.07 0.065
                                                                                                                                                                                   0.8
              0.91 0.93 0.97 0.98 0.99 1 1 1 0.99 0.96 0.94 0.93 0.89 0.87 0.8 0.69 0.56 0.41 0.24 0.19 0.16 0.13 0.03 0.011 0.03 10.011 0.0650.0740.0790.0740.069
cgmSeries_6
              0.89\ 0.92\ 0.95\ 0.97\ 0.98\ \ 1 \\ 1 \\ 1 \\ 0.98\ 0.96\ 0.95\ 0.91\ 0.89\ 0.82\ 0.71\ 0.57\ 0.41\ 0.24\ 0.18\ 0.15\ 0.12\ 0.032 0.094 0.039 0.064 0.058 0.065 0.071 0.066 0.06
cgmSeries 7
              0.88 0.9 0.94 0.95 0.97 0.99 1
                                                1 0.99 0.97 0.97 0.93 0.91 0.85 0.74 0.6 0.43 0.25 0.19 0.17 0.13 0.043 00090.0270.0240.0740.0770.0830.0760.069
camSeries 8
              camSeries 9
              0.84 0.87 0.9 0.91 0.92 0.94 0.96 0.97 0.99 1 1 0.99 0.97 0.92 0.82 0.68 0.52 0.35 0.29 0.25 0.22 0.12 0.0750.041 0.1 0.15 0.15 0.15 0.14 0.13
cgmSeries_10
                                                              1 0.99 0.98 0.93 0.83 0.7 0.54 0.36 0.29 0.26 0.22 0.13 0.0780.043 0.11 0.16 0.16 0.16 0.16 0.15
              0.84 0.86 0.89 0.91 0.91 0.93 0.95 0.97 0.99 1
                                                                                                                                                                                   0.6
cgmSeries 11
              0.8 0.83 0.85 0.86 0.87 0.89 0.91 0.93 0.97 0.99 0.99 1 0.99 0.96 0.88 0.75 0.6 0.42 0.36 0.32 0.29 0.19 0.130.085 0.14 0.19 0.18 0.17 0.16 0.15
cgmSeries_12
              0.75 0.79 0.81 0.83 0.84 0.87 0.89 0.91 0.95 0.97 0.98 0.99 1 0.98 0.92 0.8 0.66 0.48 0.41 0.38 0.35 0.26 0.2 0.16 0.21 0.25 0.23 0.22 0.21 0.19
ogmSeries_13
              0.65 0.69 0.72 0.74 0.76 0.8 0.82 0.85 0.9 0.92 0.93 0.96 0.98 1 0.97 0.89 0.77 0.61 0.54 0.51 0.48 0.39 0.33 0.28 0.31 0.35 0.31 0.29 0.28 0.25
cgmSeries 14
              0.52 0.57 0.59 0.62 0.65 0.69 0.71 0.74 0.8 0.82 0.83 0.88 0.92 0.97 1 0.97 0.88 0.75 0.69 0.65 0.63 0.55 0.49 0.45 0.46 0.48 0.44 0.41 0.39 0.36
cqmSeries 15
               0.44 0.46 0.49 0.53 0.56 0.57 0.6 0.67 0.68 0.7 0.75 0.8 0.89 0.97 1 0.96 0.86 0.81 0.77 0.76 0.7 0.64 0.59 0.58 0.59 0.53 0.49 0.47 0.43
camSeries 16
                                                                                                                                                                                  - 0.4
ogmSeries 17
              0.26 0.31 0.32 0.35 0.39 0.41 0.41 0.43 0.5 0.52 0.54 0.6 0.66 0.77 0.88 0.96 1 0.96 0.93 0.9 0.88 0.82 0.78 0.73 0.71 0.72 0.66 0.62 0.61 0.58
comSeries 18 0.12 0.17 0.18 0.21 0.23 0.24 0.24 0.25 0.32 0.35 0.36 0.42 0.48 0.61 0.75 0.86 0.96 1 0.99 0.96 0.94 0.9 0.87 0.83 0.81 0.82 0.77 0.73 0.72 0.69
ogmSeries_19 0.073 0.12 0.12 0.15 0.18 0.19 0.18 0.19 0.26 0.29 0.29 0.36 0.41 0.54 0.69 0.81 0.93 0.99 1 0.99 0.97 0.93 0.89 0.86 0.81 0.81 0.72 0.71 0.68
cgmSeries 20 0.0530.095 0.1 0.13 0.15 0.16 0.15 0.17 0.23 0.25 0.26 0.32 0.38 0.51 0.65 0.77 0.9 0.96 0.99 1 0.99 0.96 0.92 0.87 0.83 0.81 0.76 0.71 0.69 0.66
cgmSeries 21 0.0170.0590.0640.0930.12 0.13 0.12 0.13 0.19 0.22 0.22 0.29 0.35 0.48 0.63 0.76 0.88 0.94 0.97 0.99 1 0.98 0.94 0.9 0.84 0.82 0.76 0.71 0.67 0.63
                                                                                                                                                                                  - 02
comSeries 22 -0.08-0.04-0.028.004/0.0410.0390.0320.043.0.11 0.12 0.13 0.19 0.26 0.39 0.55 0.7 0.82 0.9 0.93 0.96 0.98 1 0.98 0.95 0.9 0.86 0.82 0.77 0.72 0.67
cgmSeries 23 -0.140.09@0.0740.0340001400140.009.00010050.0750.0780.13 0.2 0.33 0.49 0.64 0.78 0.87 0.89 0.92 0.94 0.98 1 0.99 0.95 0.92 0.89 0.84 0.8 0.75
comSeries 24 -0.18-0.14-0.110.0680.0360.0310.0390.0270.0210.0410.0430.0850.16 0.28 0.45 0.59 0.73 0.83 0.86 0.87 0.9 0.95 0.99 1 0.96 0.94 0.91 0.87 0.82 0.79
ogmSeries 25 -0.120.08@.05\0.0130.00@.010.006.0110.006\0.02\0.073 0.1 0.11 0.14 0.21 0.31 0.46 0.58 0.71 0.81 0.81 0.83 0.84 0.9 0.95 0.96 1 0.99 0.97 0.95 0.91 0.87
ogmSeries 26 -0.0720.033.0082.0460.0580.0650.0580.0740.12 0.15 0.16 0.19 0.25 0.35 0.48 0.59 0.72 0.82 0.81 0.81 0.82 0.86 0.92 0.94 0.99 1 0.98 0.96 0.93 0.9
                                                                                                                                                                                  -00
cgmSeries 27 -0.0540.020.0240.0630.0710.0740.0650.077 0.11 0.15 0.16 0.18 0.23 0.31 0.44 0.53 0.66 0.77 0.76 0.76 0.76 0.76 0.82 0.89 0.91 0.97 0.98 1 0.99 0.96 0.94
cgmSeries_28 -0.0420.0110.0360.0740.0770.0790.0710.0830.11 0.15 0.16 0.17 0.22 0.29 0.41 0.49 0.62 0.73 0.72 0.71 0.71 0.71 0.77 0.84 0.87 0.95 0.96 0.99 1 0.98 0.97
comSeries 29 -0.0520.0220.0260.068 0.07 0.0740.0660.076 0.1 0.14 0.16 0.16 0.21 0.28 0.39 0.47 0.61 0.72 0.71 0.69 0.67 0.72 0.8 0.82 0.91 0.93 0.96 0.98 1
camSeries 30 -0.0480.0220.0260.0670.0650.0690.060.0690.0930.13 0.15 0.15 0.15 0.19 0.25 0.36 0.43 0.58 0.69 0.68 0.66 0.63 0.67 0.75 0.79 0.87 0.9 0.94 0.97
cgmSeries 31
```

gmSeries\_12 gmSeries\_13

mSeries\_11

gmSeries\_15
gmSeries\_16
gmSeries\_17
gmSeries\_17

mSeries\_20

**☆ ← → ⊕** Q 至 🖺

gmSeries\_1

gmSeries 4

gmSeries

Output:

So our selected Features from Mean

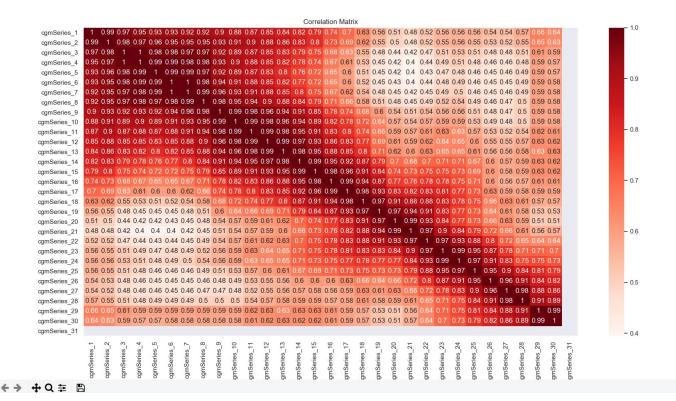
featuer1(mean)=['cgmSeries\_9', 'cgmSeries\_2', 'cgmSeries\_15', 'cgmSeries\_1'],

## for 2<sup>nd</sup> type of feature lets run(standard Deviation)

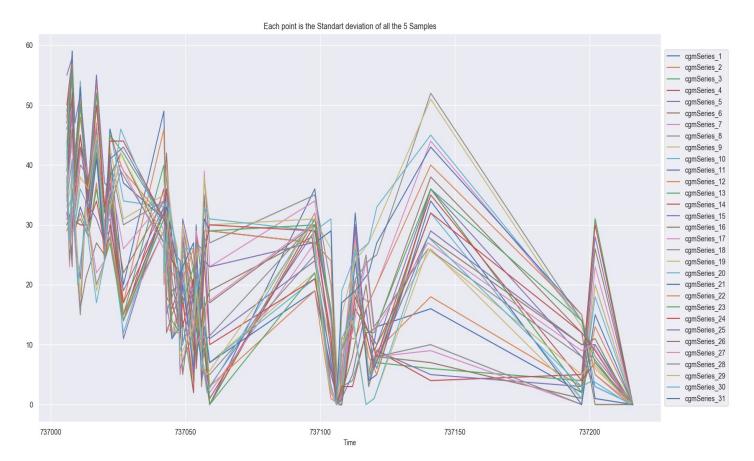
python3 assignment1.py -m std -f extract -p plot -e head

## Output

#### Cor relation matrix for std



## Data plot of std



dtype='object')

Independable variable thae are have strong correllation, Droping:{'cgmSeries\_9', 'cgmSeries\_22',
'cgmSeries\_13', 'cgmSeries\_25', 'cgmSeries\_24', 'cgmSeries\_7', 'cgmSeries\_27', 'cgmSeries\_3',
'cgmSeries\_20', 'cgmSeries\_28', 'cgmSeries\_6', 'cgmSeries\_17', 'cgmSeries\_18', 'cgmSeries\_4',
'cgmSeries\_30', 'cgmSeries\_12', 'cgmSeries\_10', 'cgmSeries\_14', 'cgmSeries\_5', 'cgmSeries\_15',
'cgmSeries\_8'}

Using Selected Features are :['cgmSeries\_2', 'cgmSeries\_1', 'cgmSeries\_19', 'cgmSeries\_29', 'cgmSeries\_16', 'cgmSeries\_26', 'cgmSeries\_23', 'cgmSeries\_11'],

features2(STD)=['cgmSeries\_2', 'cgmSeries\_1', 'cgmSeries\_19', 'cgmSeries\_29', 'cgmSeries\_16', 'cgmSeries\_26', 'cgmSeries\_23', 'cgmSeries\_11']

## for 3<sup>rd</sup> type of features lets run(Minimum)

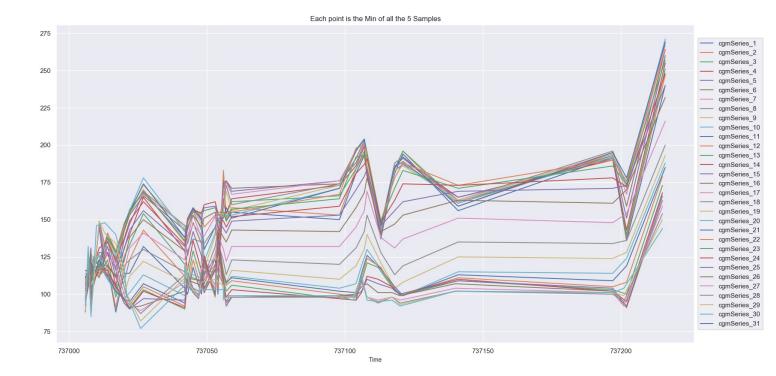
python3 assignment1.py -m min -f extract -p plot -e head

#### Output

#### Cor relation matrix for minimum

```
Correlation Matrix
                                 0.97, 0.95, 0.93, 0.91, 0.9, 0.88, 0.86, 0.85, 0.85, 0.86, 0.83, 0.79, 0.76, 0.73, 0.67, 0.55, 0.48, 0.39, 0.37, 0.32, 0.26, 0.22, 0.17, 0.15, 0.639, 0.0539, 0.460, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0.89, 0
 cgmSeries 1
 cgmSeries 2
                             1 0.99 0.97 0.95 0.94 0.92 0.91 0.89 0.88 0.88 0.87 0.85 0.8 0.77 0.74 0.68 0.55 0.48 0.39 0.36 0.32 0.25 0.21 0.15 0.13 0.0520.0170.057-0.1
                     0.97 0.99 1 0.99 0.98 0.97 0.96 0.94 0.93 0.92 0.91 0.9 0.87 0.82 0.79 0.76 0.7 0.56 0.48 0.38 0.36 0.31 0.24 0.2 0.14 0.120.0490.0180.0540.095
 cgmSeries 3
                    0.95 0.97 0.99 1 0.99 0.98 0.97 0.96 0.95 0.94 0.93 0.91 0.88 0.83 0.8 0.76 0.7 0.57 0.48 0.38 0.35 0.31 0.23 0.2 0.14 0.13 0.060.0007660340.076
 cqmSeries 4
                    0.93 0.95 0.98 0.99 1
                                                      1 0.99 0.97 0.96 0.94 0.93 0.91 0.88 0.83 0.8 0.77 0.71 0.57 0.48 0.39 0.36 0.31 0.23 0.2 0.12 0.12 0.060.00430.0290.071
 cgmSeries 5
                    0.91 0.94 0.97 0.98 1
                                                             1 0.99 0.98 0.96 0.95 0.93 0.9 0.85 0.83 0.8 0.73 0.6 0.49 0.4 0.37 0.32 0.24 0.21 0.12 0.12 0.0690.0180.0140.056
 camSeries 6
                                                                   1 0.99 0.97 0.96 0.94 0.92 0.87 0.84 0.82 0.75 0.61 0.5 0.4 0.37 0.32 0.24 0.21 0.12 0.11 0.06 0.0140.0240.063
                    0.9 0.92 0.96 0.97 0.99 1
 cgmSeries_7
                                                                          1 0.98 0.98 0.96 0.93 0.89 0.87 0.84 0.78 0.64 0.52 0.41 0.38 0.34 0.26 0.23 0.13 0.12 0.0650.0170.0180.06
                    0.88 0.91 0.94 0.96 0.97 0.99 1
 cgmSeries 8
                    0.86 0.89 0.93 0.95 0.96 0.98 0.99 1
                                                                          1 0.99 0.99 0.97 0.95 0.91 0.89 0.87 0.82 0.69 0.57 0.45 0.42 0.3
                                                                                                                                                                      0.3 0.27 0.16 0.150.0960.0470.00430.045
 cgmSeries 9
                     0.85 0.88 0.92 0.94 0.94 0.96 0.97 0.98 0.99 1
                                                                                       1 0.98 0.96 0.93 0.9 0.88 0.83 0.71 0.59 0.48 0.44 0.39 0.32 0.29 0.19 0.17 0.11 0.055.00840.043
cgmSeries_11
                    0.85 0.88 0.91 0.93 0.93 0.95 0.96 0.98 0.99 1 1 0.99 0.97 0.94 0.92 0.89 0.85 0.72 0.6 0.48 0.45 0.4 0.33 0.3 0.19 0.17 0.1 0.05.00050.051
                                                                                                                                                                                                                                                         0.6
                    0.86 0.87 0.9 0.91 0.91 0.93 0.94 0.96 0.97 0.98 0.99 1 0.99 0.97 0.95 0.93 0.89 0.77 0.63 0.5 0.47 0.42 0.3
camSeries 12
                    0.83 0.85 0.87 0.88 0.88 0.9 0.92 0.93 0.95 0.96 0.97 0.99 1 0.99 0.97 0.96 0.91 0.81 0.65 0.52 0.48 0.44 0.38 0.36 0.24 0.22 0.14 0.089 0.130 0.58
camSeries 13
                    0.79 0.8 0.82 0.83 0.83 0.85 0.87 0.89 0.91 0.93 0.94 0.97 0.99 1 0.99 0.98 0.95 0.85 <mark>0.69 0.56 0.52 0.48 0.43 0.41 0.3 0.27 0.19 0.140.0480.034</mark>
cgmSeries 14
                    0.76 0.77 0.79 0.8 0.8 0.8 0.83 0.84 0.87 0.89 0.9 0.92 0.95 0.97 0.99 1 0.99 0.97 0.88 <mark>0.72 0.59 0.55 0.51 0.45 0.43 0.31 0.28 0.2 0.150.0530.034</mark>
cgmSeries_15
                    0.73 0.74 0.76 0.76 0.77 0.8 0.82 0.84 0.87 0.88 0.89 0.93 0.96 0.98 0.99 1 0.99 0.91 0.76 0.62 0.58 0.53 0.47 0.45 0.32 0.28 0.2 0.150.0480.041
cgmSeries 16
                     0.67 0.68 0.7 0.7 0.71 0.73 0.75 0.78 0.82 0.83 0.85 0.89 0.91 0.95 0.97 0.99 1 0.96 0.83 0.69 0.64 0.6 0.54 0.51 0.39
cqmSeries 17
                                                                                                                                                                                                                                                        - 04
                     0.55 0.55 0.56 0.57 0.57 0.6 0.61 0.64 0.69 0.71 0.72 0.77 0.81 0.85 0.88 0.91 0.96 1 0.93 0.81 0.75 0.71 0.67 0.64 0.54 0.54
cqmSeries 18
                      48 0.48 0.48 0.48 0.48 0.49 0.5 0.52 0.57 0.59 0.6 0.63 0.65 0.69 0.72 0.76 0.83 0.93 1 0.96 0.91 0.87 0.83 0.79 0.72 0.65 0.56 0.43
camSeries 19
                      39 0.39 0.38 0.38 0.39 0.4 0.4 0.41 0.45 0.48 0.48 0.5 0.52 0.56 0.59 0.62 0.69 0.81 0.96 1 0.99 0.96 0.92 0.88 0.81 0.74 0.65 0.53 0.44 0
camSeries 20
                                                               37 0.38 0.42 0.44 0.45 0.47 0.48 0.52 0.55 0.58 0.64 0.75 0.91 0.99 1 0.99 0.96 0.92 0.84 0.77 0.68 0.56 0.47 0.
cgmSeries_21
ogmSeries_22 0.32 0.32 0.31 0.31 0.31 0.32 0.32 0.3
                                                                                0.39 0.4 0.42 0.44 0.48 0.51 0.53 0.6 0.71 0.87 0.96 0.99 1 0.99 0.96 0.87 0.8 0.7 0.58 0.49 0.4
- 0.2
                    0.22 0.21 0.2 0.2 0.2 0.2 0.21 0.21 0.23 0.27 0.29 0.3 0.34 0.36 0.41 0.43 0.45 0.51 0.64 0.79 0.88 0.92 0.96 0.99 1 0.92 0.86 0.77 0.65 0.55 0.4
cgmSeries_25 0.17 0.15 0.14 0.14 0.12 0.12 0.12 0.12 0.13 0.16 0.19 0.19 0.22 0.24 0.3 0.31 0.32 0.39 0.54 0.72 0.81 0.84 0.87 0.9 0.92 1 0.98 0.91 0.82 0.73 0.64
cgmSeries 26 0.15 0.13 0.12 0.13 0.12 0.12 0.11 0.12 0.15 0.17 0.17 0.19 0.22 0.27 0.28 0.28 0.33 0.48 0.65 0.74 0.77 0.8 0.83 0.86 0.98 1 0.97 0.91 0.84 0.76
cgmSeries_27 0.0630.0520.0490.0610.0610.0690.060.0650.0960.11 0.1 0.12 0.14 0.19 0.2 0.2 0.25 0.4 0.56 0.65 0.68 0.7 0.74 0.77 0.91 0.97 1 0.98 0.93 0.86
cgmSeries_28 -0.0058.0170.048.0000600430.0180.0110.0170.0470.0550.050.0690.0890.14 0.15 0.15 0.2
                                                                                                                                        0.34 0.47 0.53 0.56 0.58 0.62 0.65 0.82 0.91 0.98 1 0.98 0.92
                                                                                                                                                                                                                                                       - 00
cgmSeries_29 -0.0480.0570.0540.0340.0290.0140.0240.018.004@00®9000$300740.0130.0480.0530.0480.094 0.22 0.36 0.44 0.47 0.49 0.52 0.55 0.73 0.84 0.93 0.98 1 0.98
cgmSeries_30 -0.089-0.1-0.0990.0760.0740.0560.0630.060.0450.0450.0520.0580.0340.0340.040.00370.11 0.26 0.36 0.4 0.41 0.45 0.46 0.64 0.76 0.86 0.92 0.98 1
```

# Data plot of min



features3(Minimum)=['cgmSeries\_2', 'cgmSeries\_15', 'cgmSeries\_1', 'cgmSeries\_9', 'cgmSeries\_18']

for 4th type of features lets run(Mximum)

Output

## Cor relation matrix for max:

## Correlation Matrix

- 0.9

- 0.8

**-** 0.7

- 0.6

- 0.5

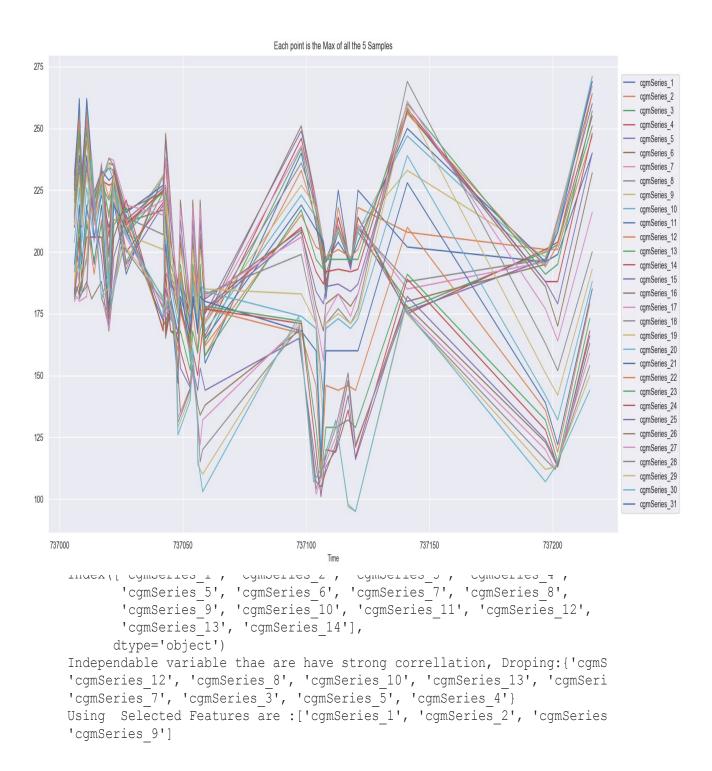
- 0.4

- 0.3

- 0.2

cgmSeries_1	1	0.99	0.95	0.93	0.89	0.89	0.89	0.88	0.84	0.82	0.79	0.76	0.7	0.54	0.42	0.33	0.3	0.3	0.26	0.29	0.23	0.21	0.2	0.17	0.22	0.26	0.24	0.27	0.28	0.31	
cgmSeries_2	0.99	1	0.97	0.95	0.91	0.91	0.92	0.91	0.89	0.86	0.84	0.81	0.75	0.58	0.45	0.35	0.32	0.33	0.3	0.33	0.28	0.26	0.25	0.23	0.27	0.31	0.29	0.31	0.33	0.36	
cgmSeries_3	0.95	0.97	1	1	0.97	0.97	0.97	0.95	0.85	0.81	0.79	0.73	0.65	0.47	0.35	0.25	0.23	0.23	0.21	0.26	0.23	0.23	0.26	0.24	0.29	0.31	0.3	0.32	0.35	0.37	
cgmSeries_4	0.93	0.95	1	1	0.98	0.98	0.97	0.96	0.86	0.81	0.79	0.73	0.65	0.46	0.34	0.25	0.23	0.23	0.21	0.28	0.24	0.25	0.29	0.27	0.31	0.33	0.32	0.34	0.37	0.4	
cgmSeries_5	0.89	0.91	0.97	0.98	1	0.99	0.98	0.95	0.85	8.0	0.77	0.71	0.64	0.47	0.36	0.28	0.26	0.26	0.24	0.29	0.26	0.28	0.32	0.3	0.33	0.35	0.32	0.35	0.39	0.42	
cgmSeries_6	0.89	0.91	0.97	0.98	0.99	1	0.99	0.96	0.87	0.82	0.79	0.73	0.65	0.49	0.37	0.28	0.26	0.26	0.23	0.29	0.26	0.28	0.33	0.31	0.34	0.37	0.34	0.36	0.4	0.42	
cgmSeries_7	0.89	0.92	0.97	0.97	0.98	0.99	1	0.99	0.9	0.86	0.83	0.77	0.69	0.53	0.4	0.29	0.27	0.26	0.23	0.29	0.26	0.28	0.32	0.3	0.33	0.36	0.33	0.35	0.39	0.42	
cgmSeries_8	0.88	0.91	0.95	0.96	0.95	0.96	0.99	1	0.94	0.9	0.88	0.83	0.76	0.61	0.48	0.36	0.32	0.31	0.29	0.33	0.3	0.32	0.35	0.33	0.37	0.39	0.37	0.39	0.43	0.45	
cgmSeries_9	0.84	0.89	0.85	0.86	0.85	0.87	0.9	0.94	1	0.99	0.98	0.95	0.91	0.79	0.66			0.49	0.46	0.47	0.44	0.43	0.42	0.41	0.44	0.46	0.43	0.45	0.48	0.5	
cgmSeries_10	0.82	0.86	0.81	0.81	0.8	0.82	0.86	0.9	0.99	1	0.99	0.98	0.94	0.83	0.71	0.58		0.54			0.48	0.46	0.44	0.43	0.47	0.48	0.45	0.47		0.52	
cgmSeries_11	0.79	0.84	0.79	0.79	0.77	0.79	0.83	0.88	0.98	0.99	1	0.99	0.95	0.86	0.74	0.62	0.59	0.59					0.49	0.48					0.54	0.56	
cgmSeries_12	0.76	0.81	0.73	0.73	0.71	0.73	0.77	0.83	0.95	0.98	0.99	1	0.99	0.91	0.81	0.7	0.67	0.66	0.62	0.61	0.58	0.57			0.55			0.54	0.56	0.57	
cgmSeries_13	0.7	0.75	0.65	0.65	0.64	0.65	0.69	0.76	0.91	0.94	0.95	0.99	1	0.96	0.88	0.78	0.74	0.73	0.69	0.67	0.64	0.62	0.57	0.56			0.56	0.57	0.59	0.6	
cgmSeries_14	0.54	0.58	0.47	0.46	0.47	0.49		0.61	0.79	0.83	0.86	0.91	0.96	1	0.97	0.91	0.87	0.85	0.81	0.76	0.75	0.74	0.68	0.67	0.68	0.66	0.61	0.6	0.61	0.61	
cgmSeries_15	0.42																0.95	0.91	0.87	0.81	0.81	0.81	0.75	0.73	0.73	0.71	0.63	0.62	0.64	0.64	
cgmSeries_16	0.33	0.35	0.25	0.25	0.28	0.28	0.29	0.36		0.58	0.62	0.7	0.78	0.91	0.98	1	0.98	0.94	0.91	0.84	0.85	0.86	8.0	0.77	0.76	0.73	0.65	0.63	0.66	0.65	
cgmSeries_17	0.3	0.32	0.23	0.23	0.26	0.26	0.27	0.32			0.59	0.67	0.74	0.87	0.95	0.98	1	0.98	0.96	0.91	0.91	0.91	0.86	0.83	0.81	0.78	0.7	0.68	0.7	0.69	
cgmSeries_18	0.3	0.33	0.23	0.23	0.26	0.26	0.26	0.31	0.49	0.54	0.59	0.66	0.73	0.85	0.91	0.94	0.98	1	0.99	0.95	0.95	0.94	0.88	0.85	0.83	8.0	0.72	0.71	0.71	0.71	
cgmSeries_19	0.26																													(300 Table )	
cgmSeries_20	0.29																													CONTRACTOR OF	
cgmSeries_21	0.23	0.28	0.23	0.24	0.26	0.26	0.26	0.3	0.44	0.48		0.58	0.64	0.75	0.81	0.85	0.91	0.95	0.97	0.99	1	0.98	0.93	0.9	0.86	0.82	0.74	0.72	0.7	0.7	
cgmSeries_22	0.21																													100	
cgmSeries_23			0.26																					0.99	0.95	0.91	0.84	8.0	0.81	0.81	
cgmSeries_24	0.17																								0.97	0.93	0.89	0.85	0.84	0.84	
cgmSeries_25	0.22																										0.96				
cgmSeries_26	0.26																										0.95	0.93	0.89	0.89	
cgmSeries_27	0.24														0.63														0.94		
cgmSeries_28	0.27																												0.95	0.94	
cgmSeries_29	0.28																												1	1	
cgmSeries_30	0.31	0.36	0.37	0.4	0.42	0.42	0.42	0.45	0.5	0.52	0.56	0.57	0.6	0.61	0.64	0.65	0.69	0.71	0.7	0.69	0.7	0.75	0.81	0.84	0.9	0.89	0.93	0.94	1	1	
cgmSeries_31																														-13	
	-	Ν,	Э	4	-2	9	7	ω,	6	0	_	2	8	4	15	9	7	8	6	50	7	2	33	24	25	26	75	28	29	00	7
	es	es	es	es	es		es	es	es	S	S.	, s	L S	S 1	S <sub>I</sub>	L S	S	, s	S.	S	S	S	S		1000		S			S	S 3
	Ser	Ser	Ser	Ser	Ser	Seri	Seri	Ser	Ser	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie	erie
	ogmSeries_	gmSeries	ogmSeries	gmSeries_10	gmSeries_11	gmSeries_12	gmSeries_13	gmSeries	gmSeries_	gmSeries_16	gmSeries_17	gmSeries_18	gmSeries_19	gmSeries_20	gmSeries_21	gmSeries_22	gmSeries_23	gmSeries	gmSeries	gmSeries	gmSeries_27	gmSeries	gmSeries	gmSeries_30	gmSeries_31						
	J	J	0	J	J	0	0	J	J	Δ)	Ď	Δ)	Ď	Ď	Δĵ	ΔJ	Δĵ	Ď	Ď)	ζņ	ζŋ	Ď)	Ž)	Ď	Δĵ	Ö)	Ď)	Ď	Δj	Ö)	Ŏ)

## Data plot of max



```
Selected features using mean/std/min/max are following
featuer1(mean)=['cgmSeries_9', 'cgmSeries_2', 'cgmSeries_15', 'cgmSeries_1'],
features2(STD)=['cgmSeries_2', 'cgmSeries_1', 'cgmSeries_19', 'cgmSeries_29', 'cgmSeries_16',
'cgmSeries 26', 'cgmSeries 23', 'cgmSeries 11']
features3(Minimum)=['cgmSeries_2', 'cgmSeries_15', 'cgmSeries_1', 'cgmSeries_9', 'cgmSeries_18']
features4(Maximum)=['cgmSeries 1', 'cgmSeries 2', 'cgmSeries 14', 'cgmSeries 9']
   c) Show values of each of the features and argue that your intuition in step b is validated or
      disproved? (5 points each ) total 20
      Output for Feature1(Mean)
cgmSeries 15,cgmSeries 2,cgmSeries 9,cgmSeries 1
240.0,247.0,271.0,240.0
176.0,169.0,192.0,167.0
179.0,198.0,196.0,194.0
196.0,179.0,190.0,177.0
171.0,206.0,201.0,207.0
172.0,191.0,190.0,189.0
167.0,182.0,178.0,184.0
185.0,181.0,191.0,180.0
179.0,194.0,204.0,193.0
177.0,198.0,207.0,196.0
174.0,193.0,203.0,193.0
165.0,158.0,164.0,155.0
175.0,165.0,180.0,163.0
166.0,187.0,165.0,185.0
175.0,151.0,164.0,151.0
167.0,139.0,149.0,141.0
175.0,137.0,156.0,136.0
169.0,130.0,161.0,127.0
180.0,143.0,171.0,141.0
177.0,139.0,170.0,138.0
176.0,154.0,175.0,151.0
192.0,174.0,192.0,167.0
193.0,181.0,203.0,175.0
178.0,168.0,185.0,163.0
195.0,183.0,196.0,180.0
181.0,175.0,183.0,174.0
182.0,182.0,188.0,180.0
171.0,169.0,169.0,170.0
178.0,176.0,182.0,173.0
171.0,161.0,164.0,160.0
Output for Feature2(Std)
cgmSeries 2,cgmSeries 1,cgmSeries 19,cgmSeries 29,cgmSeries 16,cgmSeries 26,cgmSeries 23,cgm
Series 11
```

features4(Maximum)=['cgmSeries\_1', 'cgmSeries\_2', 'cgmSeries\_14', 'cgmSeries\_9']

```
,,,,,,,
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
31.0,31.0,6.0,7.0,0.0,11.0,7.0,15.0
3.0,2.0,13.0,5.0,12.0,10.0,14.0,0.0
18.0, 16.0, 51.0, 26.0, 38.0, 28.0, 36.0, 34.0
9.0,13.0,30.0,1.0,13.0,9.0,12.0,5.0
5.0,5.0,27.0,0.0,12.0,20.0,12.0,4.0
29.0,32.0,24.0,14.0,16.0,8.0,4.0,24.0
2.0,0.0,15.0,10.0,0.0,2.0,3.0,0.0
0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0
14.0,14.0,26.0,4.0,9.0,3.0,16.0,6.0
32.0,36.0,31.0,30.0,30.0,28.0,30.0,19.0
0.0,0.0,30.0,5.0,17.0,19.0,29.0,7.0
11.0,12.0,38.0,7.0,35.0,16.0,27.0,13.0
3.0,3.0,26.0,10.0,21.0,14.0,21.0,7.0
28.0,30.0,25.0,19.0,29.0,8.0,9.0,13.0
19.0, 17.0, 22.0, 13.0, 9.0, 19.0, 23.0, 6.0
25.0,26.0,30.0,8.0,28.0,5.0,17.0,20.0
20.0,19.0,20.0,26.0,11.0,25.0,20.0,13.0
22.0,25.0,31.0,32.0,26.0,27.0,25.0,11.0
24.0,28.0,39.0,21.0,42.0,22.0,30.0,26.0
46.0,49.0,35.0,28.0,34.0,27.0,31.0,32.0
17.0,20.0,31.0,42.0,22.0,42.0,42.0,15.0
40.0,42.0,43.0,30.0,35.0,38.0,45.0,28.0
33.0,33.0,29.0,24.0,27.0,25.0,27.0,26.0
54.0,55.0,45.0,19.0,47.0,27.0,36.0,42.0
36.0,35.0,36.0,33.0,37.0,21.0,28.0,30.0
48.0,45.0,38.0,18.0,43.0,16.0,31.0,53.0
38.0,37.0,30.0,39.0,33.0,33.0,30.0,39.0
54.0,51.0,38.0,26.0,46.0,23.0,32.0,59.0
47.0,46.0,33.0,42.0,35.0,35.0,30.0,41.0
Output for Feature3(min)
cgmSeries 2,cgmSeries 15,cgmSeries 1,cgmSeries 9,cgmSeries 18
247.0,240.0,240.0,271.0,200.0
138.0,173.0,136.0,171.0,136.0
194.0,171.0,192.0,194.0,134.0
159.0,169.0,156.0,163.0,135.0
196.0,162.0,194.0,186.0,119.0
```

184.0,156.0,181.0,183.0,113.0 145.0,146.0,145.0,143.0,128.0 179.0,184.0,180.0,187.0,153.0 194.0,179.0,193.0,204.0,142.0 184.0,170.0,181.0,196.0,131.0 153.0, 153.0, 150.0, 173.0, 120.0 158.0, 149.0, 155.0, 156.0, 123.0 151.0,141.0,149.0,169.0,113.0 183.0,144.0,181.0,156.0,122.0 104.0, 150.0, 100.0, 147.0, 135.0 120.0, 159.0, 124.0, 139.0, 145.0 110.0, 158.0, 114.0, 117.0, 134.0 104.0, 150.0, 100.0, 147.0, 135.0 123.0,157.0,118.0,157.0,137.0 104.0,149.0,100.0,147.0,129.0 112.0,132.0,104.0,144.0,115.0 143.0,156.0,132.0,173.0,130.0 123.0,143.0,114.0,156.0,123.0 119.0,130.0,114.0,146.0,108.0

```
112.0,128.0,113.0,113.0,120.0
116.0,117.0,115.0,116.0,117.0
112.0,107.0,113.0,113.0,118.0
98.0,106.0,98.0,109.0,106.0
Output for Feature4(min)
cgmSeries 1,cgmSeries 2,cgmSeries 14,cgmSeries 9
,,,
240.0,247.0,248.0,271.0
199.0,201.0,188.0,212.0
196.0,201.0,188.0,198.0
202.0,208.0,258.0,233.0
225.0,218.0,193.0,209.0
193.0, 195.0, 192.0, 196.0
225.0,218.0,193.0,209.0
181.0,184.0,192.0,196.0
193.0,194.0,187.0,204.0
211.0,213.0,192.0,218.0
240.0,233.0,210.0,227.0
155.0,158.0,177.0,171.0
180.0,178.0,204.0,191.0
189.0,191.0,182.0,174.0
180.0,178.0,204.0,191.0
158.0,158.0,168.0,156.0
180.0,178.0,204.0,191.0
147.0,153.0,187.0,189.0
180.0,178.0,204.0,191.0
167.0,166.0,238.0,205.0
231.0,231.0,225.0,231.0
191.0,193.0,211.0,209.0
220.0,223.0,233.0,231.0
210.0,216.0,206.0,236.0
223.0,225.0,233.0,229.0
206.0,211.0,216.0,220.0
239.0,245.0,236.0,261.0
215.0,220.0,210.0,223.0
239.0,245.0,236.0,261.0
```

91.0,117.0,88.0,123.0,106.0 119.0,128.0,119.0,139.0,110.0

215.0,220.0,210.0,223.0

I have already explained in the previous part how I got these values and these all are valid important features in the data, as we can see from the data there are a lot of spikes and slop so these are important features

d) Create a feature matrix where each row is a collection of features from each time series. SO if there are 75 time series and your feature length after concatenation of the 4 types of features is 17 then the feature matrix size will be 75 X 17 (10 points) to create matrix for pca run

python3 assignment1.py -m matrix

it will write the matrix data to 'derivedMatrix.csv'

here is the output

MATRIX	X FOR PCA	> cgmS	Series_9_mean	cgmSeries_2_m	ean cgmSeries	_15_mean
cgmSeri	es_1_mean	cgmSeries_1	l1_std	cgmSeries_18_	min cgmSer	ies_2_min
cgmSeri			in cgmSeries_1			
2	271.0	247.0	240.0	240.0	0.0	200.0
247.0	271.0	240.0	240.0			
3	192.0	169.0	176.0	167.0	15.0	136.0
138.0	171.0	173.0	136.0			
4	196.0	198.0	179.0	194.0	0.0	134.0
194.0	194.0	171.0	192.0			
5	190.0	179.0	196.0	177.0	34.0	135.0
159.0	163.0	169.0	156.0			
6	201.0	206.0	171.0	207.0	5.0	119.0
196.0	186.0	162.0	194.0			
7	190.0	191.0	172.0	189.0	4.0	113.0
184.0	183.0	156.0	181.0			
8	178.0	182.0	167.0	184.0	24.0	128.0
145.0	143.0	146.0	145.0			
9	191.0	181.0	185.0	180.0	0.0	153.0
179.0	187.0	184.0	180.0			
10	204.0	194.0	179.0	193.0	0.0	142.0
194.0	204.0	179.0	193.0			
11	207.0	198.0	177.0	196.0	6.0	131.0
184.0	196.0	170.0	181.0			
12	203.0	193.0	174.0	193.0	19.0	120.0
153.0	173.0	153.0	150.0			
13	164.0	158.0	165.0	155.0	7.0	123.0
158.0	156.0	149.0	155.0			
14	180.0	165.0	175.0	163.0	13.0	113.0
151.0	169.0	141.0	149.0			
15	165.0	187.0	166.0	185.0	7.0	122.0
183.0	156.0	144.0	181.0			
16	164.0	151.0	175.0	151.0	13.0	135.0
104.0	147.0	150.0	100.0			
17	149.0	139.0	167.0	141.0	6.0	145.0
120.0	139.0	159.0	124.0			

18	156.0	137.0	175.0	136.0	20.0	134.0
110.0	117.0	158.0	114.0			
19	161.0	130.0	169.0	127.0	13.0	135.0
104.0	147.0	150.0	100.0			
20	171.0	143.0	180.0	141.0	11.0	137.0
123.0	157.0	157.0	118.0			
21	170.0	139.0	177.0	138.0	26.0	129.0
104.0	147.0	149.0	100.0			
22	175.0	154.0	176.0	151.0	32.0	115.0
112.0	144.0	132.0	104.0			
23	192.0	174.0	192.0	167.0	15.0	130.0
143.0	173.0	156.0	132.0			
24	203.0	181.0	193.0	175.0	28.0	123.0
123.0	156.0	143.0	114.0			
25	185.0	168.0	178.0	163.0	26.0	108.0
119.0	146.0	130.0	114.0			
26	196.0	183.0	195.0	180.0	42.0	106.0
91.0	123.0	117.0	88.0			
27	183.0	175.0	181.0	174.0	30.0	110.0
119.0	139.0	128.0	119.0			
28	188.0	182.0	182.0	180.0	53.0	120.0
112.0	113.0	128.0	113.0			
29	169.0	169.0	171.0	170.0	39.0	117.0
116.0	116.0	117.0	115.0			
30	182.0	176.0	178.0	173.0	59.0	118.0
112.0	113.0	107.0	113.0			
31	164.0	161.0	171.0	160.0	41.0	106.0
98.0	109.0	106.0	98.0			

[30 rows x 21 columns]

Note 0.0=NAN values

e) Provide this feature matrix to PCA and derive the new feature matrix. Chose the top 5 features and plot them for each time series. (5 points)

Lets run the PCA:

## Output

Fisrt Pricncpal Component[0.21733285 0.08614395 0.2451086 0.2386494 0.09612052 0.14096799

0.09644583 0.08640853 0.13058543 0.16055111 0.10857047 0.16072668 0.02347255 0.03594191 0.04659873 0.0469037 0.14950063 0.28111919 0.46201078 0.40163057 0.46864806]

[20 18 19 17 2]

Top 5 Featues from PCA=cgmSeries\_2\_min

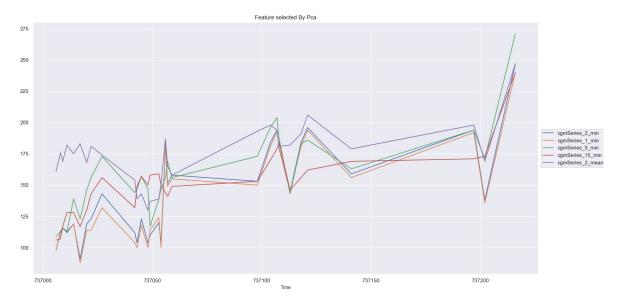
Top 5 Featues from PCA=cgmSeries\_1\_min

Top 5 Featues from PCA=cgmSeries\_9\_min

Top 5 Featues from PCA=cgmSeries\_15\_min

Top 5 Featues from PCA=cgmSeries\_2\_mean

Plot for top 5 features2

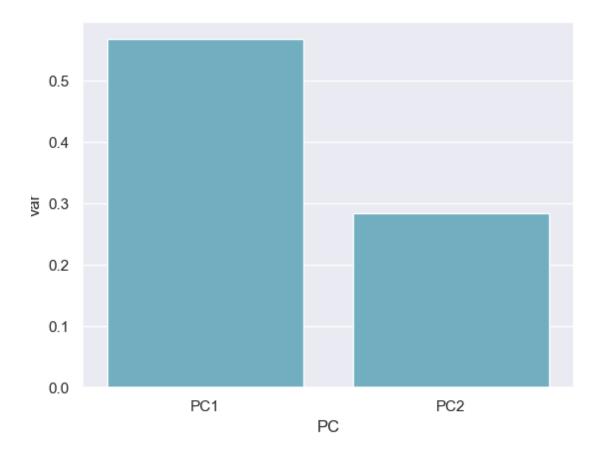


f) For each feature in the top 5 argue why it is chosen as a top five feature in PCA? (3 points each) total 15



python3 assignment1.py -m pca

output



Fisrt Pricncpal Component[0.21733285 0.08614395 0.2451086 0.2386494 0.09612052 0.14096799

0.09644583 0.08640853 0.13058543 0.16055111 0.10857047 0.16072668

0.02347255 0.03594191 0.04659873 0.0469037 0.14950063 0.28111919

0.46201078 0.40163057 0.46864806]

[20 18 19 17 2]

Top 5 Featues from PCA=cgmSeries\_2\_min

Top 5 Featues from PCA=cgmSeries\_1\_min

Top 5 Featues from PCA=cgmSeries\_9\_min

Top 5 Featues from PCA=cgmSeries\_15\_min

Top 5 Featues from PCA=cgmSeries\_2\_mean

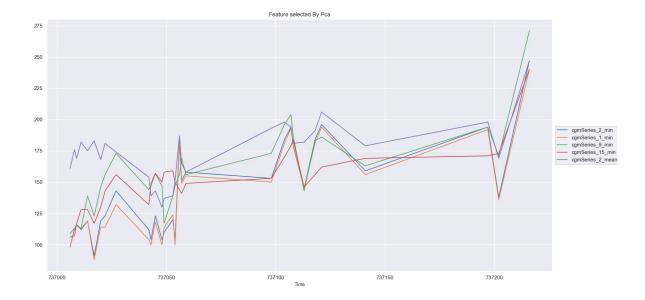
As we can see, 73% of varience belongs to pc1, and 27% to PC2

We will select pca1, we selected the 5 features with maximum pc1 values and 5 selected features based on maximum PC1 values are

1)cgmSeries\_2\_min

2)cgmSeries\_1\_min 3)cgmSeries\_9\_min 4)cgmSeries\_15\_min 5)cgmSeries\_2\_mean1

## PLot



if we compare this graph to the original data, the original looks smooth while the 5 features selected are the features with high variation.