Name: MD Ekram Ullah

Batch: ML-DS 102

Submitted to: AILABS

1. Problem Statement:

Determination of production mode according to client requirement for a gas manufacturing plant

1. Description of dataset:

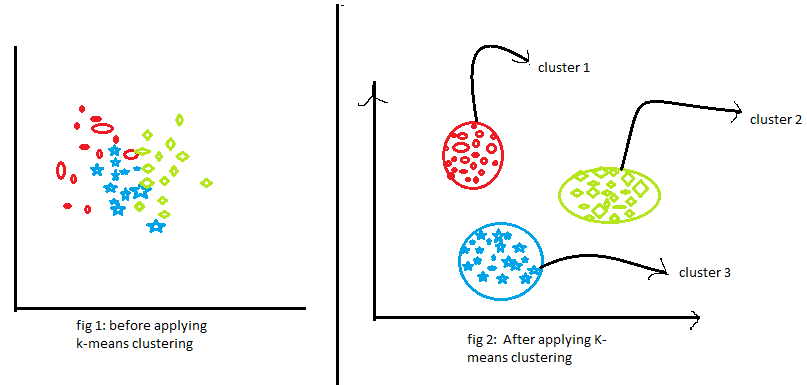
There were two datasets given for this problem. The first dataset was about production of three different gases according to dates. At the very beginning there was 384 different days and their production values.

The second dataset was about process variables according to 15 minutes time interval. There was in total 50 values and 35040 datapoints out there.

1. Theory
2. K-means clustering

*A K-means clustering algorithm tries to group similar items in the form of clusters. The number of groups is represented by K.*

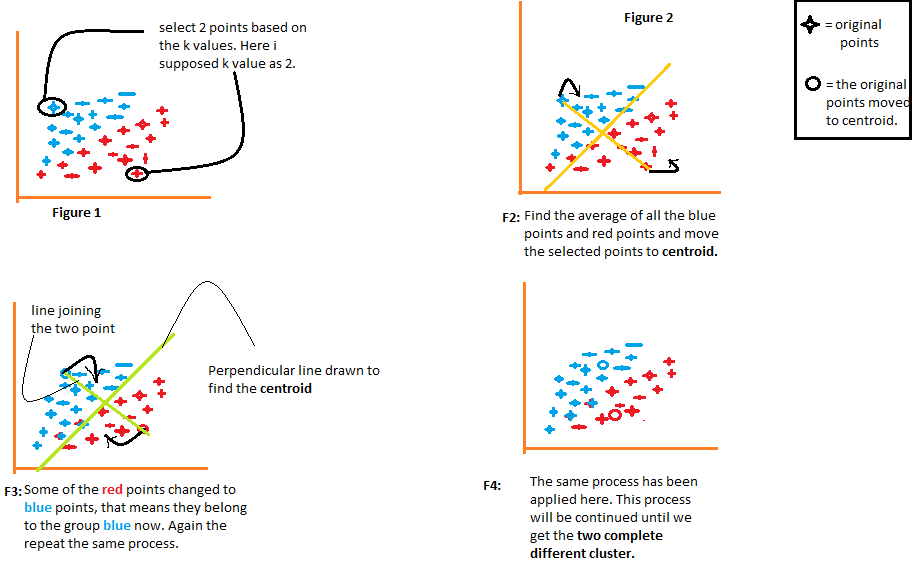
The algorithm will categorize the items into k groups of similarity. To calculate that similarity, we will use the Euclidean distance as measurement.



k-means clustering tries to group similar kinds of items in form of clusters. It finds the similarity between the items and groups them into the clusters. K-means clustering algorithm works in three steps. Let’s see what are these three steps.

1. Select the k values.
2. Initialize the centroids.
3. Select the group and find the average.

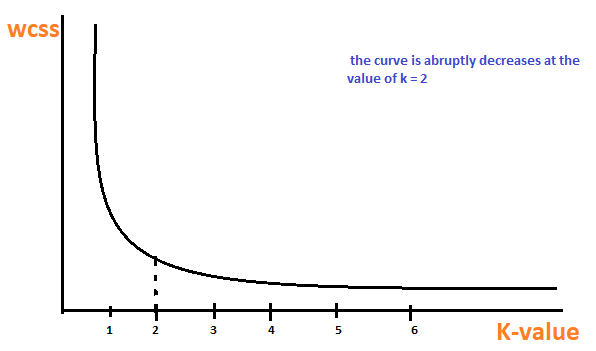
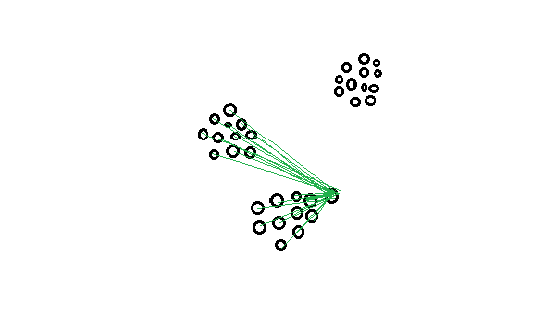
Let us understand the above steps with the help of the figure because a good picture is better than the thousands of words.



**How to choose the value of K?**

One of the most challenging tasks in this clustering algorithm is to choose the right values of k. What should be the right k-value? How to choose the k-value? Let us find the answer to these questions. If you are choosing the k values randomly, it might be correct or may be wrong. If you will choose the wrong value then it will directly affect your model performance. So there are two methods by which you can select the right value of k.

1. Elbow Method.
2. Silhouette Method.

Elbow method Silhouette Method.

1. Method:

Preprocessing:

I started my preprocessing step in the 1st dataset which contained the production values. Firstly, I dropped the unnecessary columns, re-indexed the dataset and dropped the rows and columns with all null values in it. Then I changed the dataframe to the 1st row which contained the feature names. Then I created another dataframe with same header name and concatenated with the previous dataframe which gave me a structured view of the whole dataset. Then I fixed the date-time format of the dataset and removed the datapoints that contained negative values. After that I added the percentage of each production value to the dataframe and the dataframe was ready for clustering part.

1. **Determination of outlier data points:**

We have to show the max, min ,std and mean distance value for all clusters

A huge barrier for any dataset is its outliers. In the 1st dataset there were a few outliers in it which needed to be removed before moving further. So, after performing the clustering part, I isolated the outliers on the basis of its distance from the centroid. I set the threshold to mean+(2\*std). Firstly, I calculated mean, standard deviation, maximum value and minimum value for each cluster and then using the mean and standard deviation I managed to detect the outliers and added them in a separate column named ‘outliers’ where ‘1’ means the datapoint is an outlier and ‘0’ means it is not. After that I saved the whole dataset in a particular CSV file. Then, I isolated the outliers from my dataframe to work further.

1. **Determining optimum cluster for non-outlier datapoints**

After completing the preprocessing steps thoroughly, I divided the whole production dataset into suitable number of clusters. As the division was to made according to the percentage of the production value, so the K-means clustering was done with three features. To determine the cluster number, I used the Elbow method in which I found the suitable cluster number to be 5. Then I splitted the dataset into training and test where training set contained the 90% of the dataset and the test set contained only 10% . I again performed Elbow method in the training data and it also showed that suitable number of clusters will be 5. I plotted the clusters using 3D scatter plot. They were well separated from each other. Then to testify the test dataset, I measured the distance of each datapoints of the test data from the centroids of the training data and mapped the datapoint to the nearest cluster. After plotting it into 3D scatter plot, it showed quite accurate result. So my clustering was satisfactory.

1. **Description of process variable for all cluster in daily frequency and 15 min freq.**
2. **15 minute frequency :**

In the 15 mint frequency determination, I have generalized the whole timestamp column by taking only the day wise part at first.

Then marging the 2nd dataset with the 1st one according to ‘Date’ column.

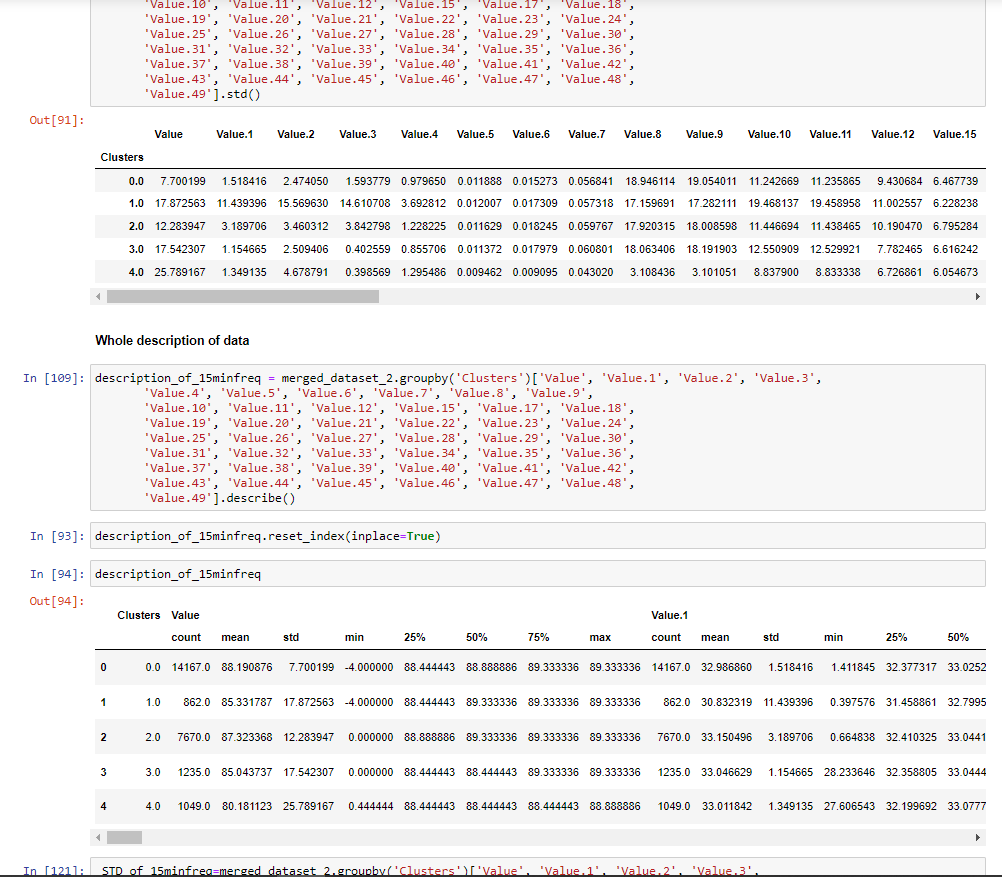
Then to determine the median and standard deviation of the merged dataset firstly I arranged the dataset according to cluster and then calculated the median and standard deviation. To view the whole description of the dataset, I used the describe method.

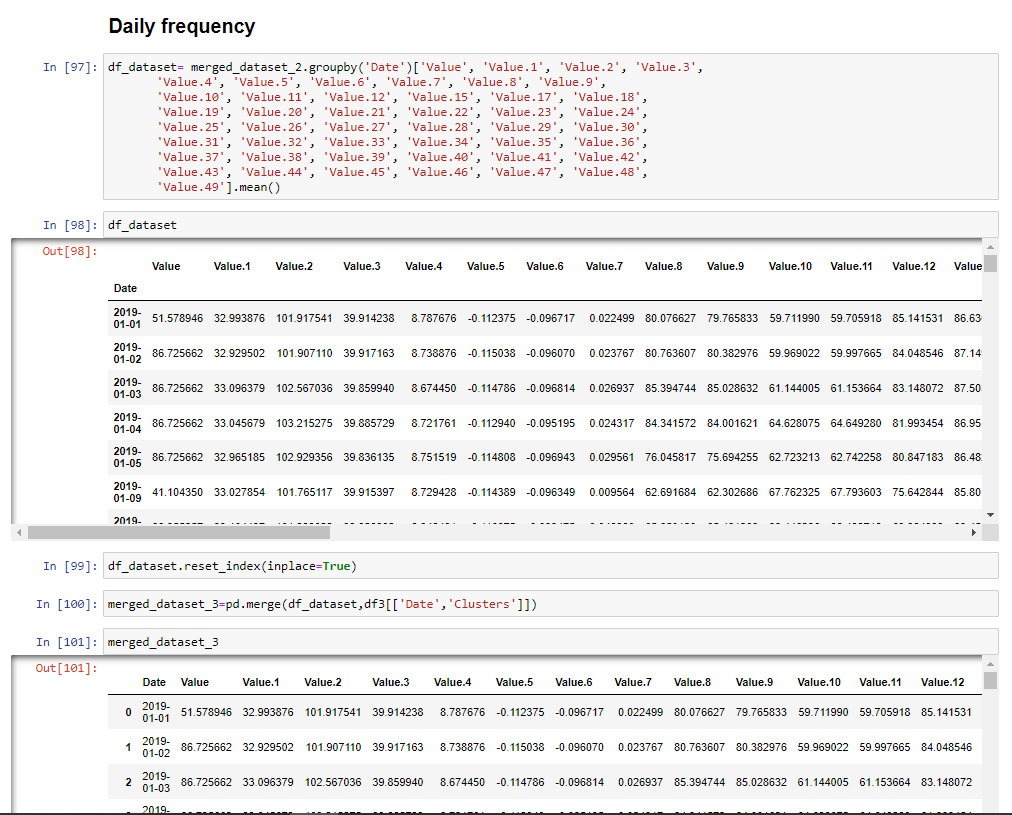
1. **Daily frequency:**

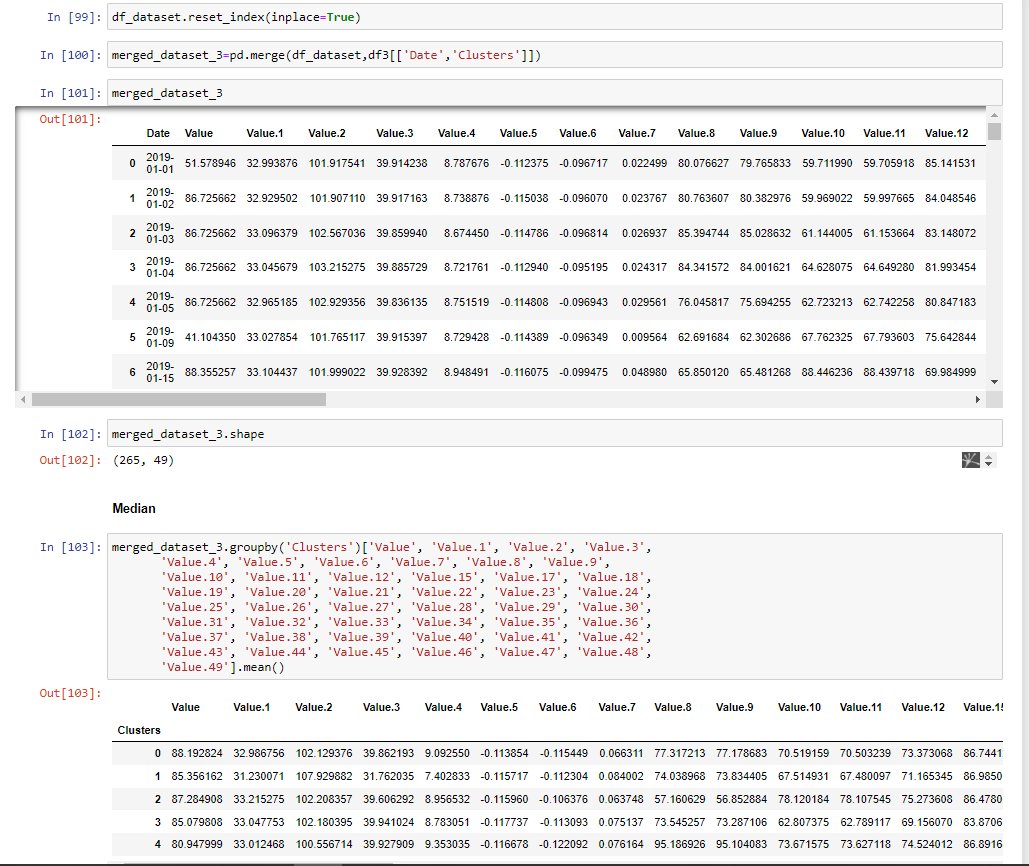
In case of daily frequency , firstly I arranged the dataset according to date and then took the mean of that. By which I got each process variable value of a certain date. Then I merged this dataset with the 1st dataset which contained the cluster values. Then grouped the dataset by cluster and calculated median and standard deviation. To view the whole description of the dataset, I used the describe method.

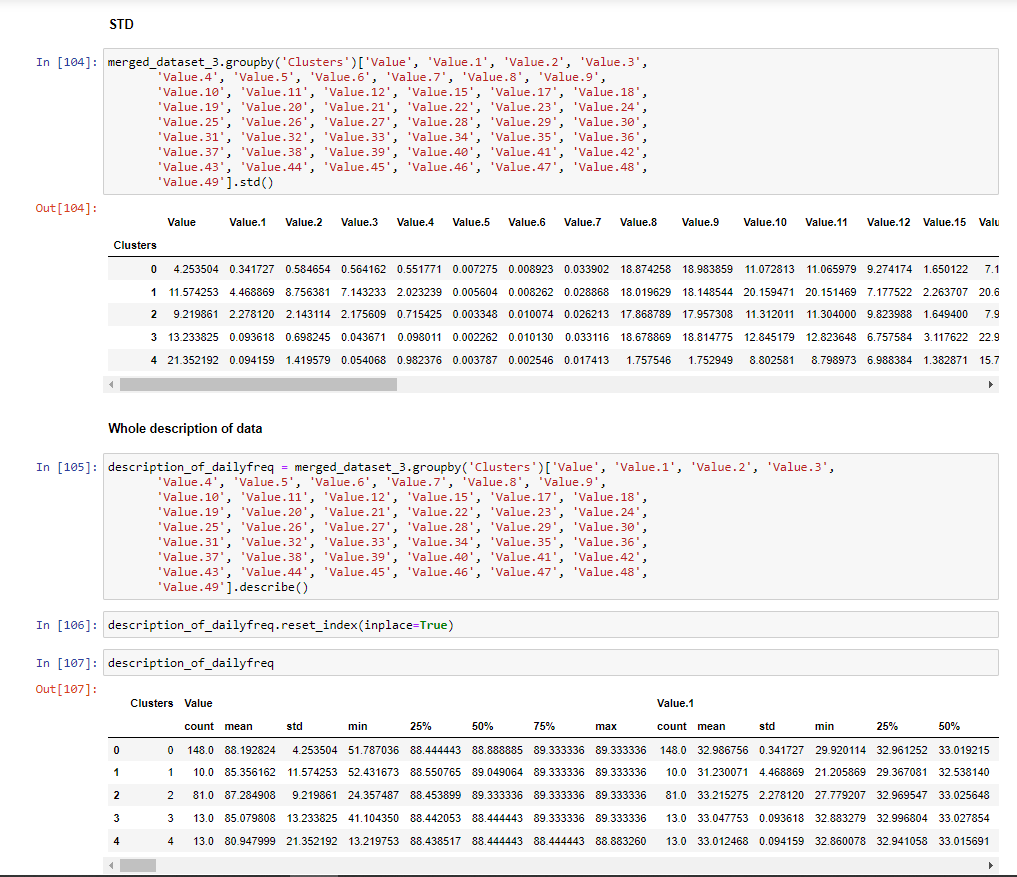






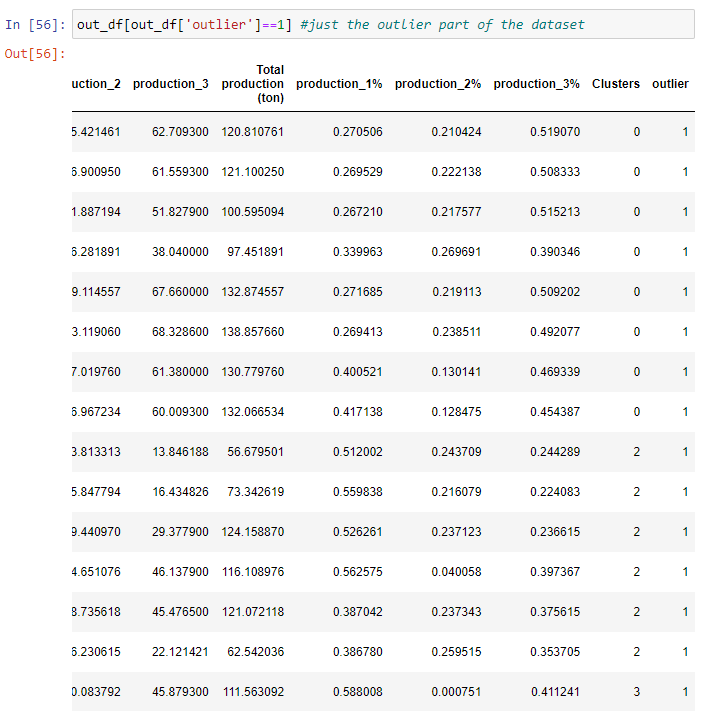




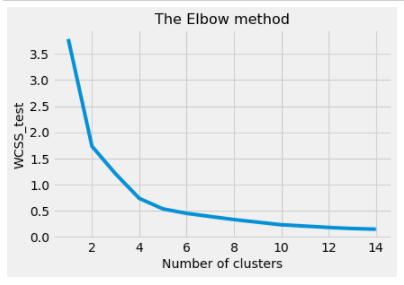


1. Results:

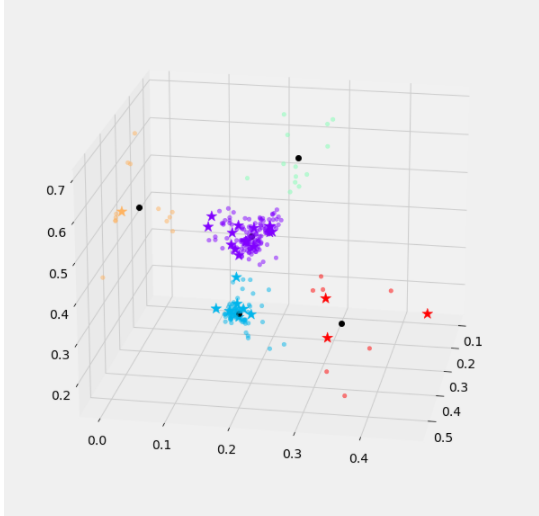
Outliers:



Elbow method:



Clusters:



The training and test cluster points performed well.

Median for process variables:

15 Minute freq:

| **Clusters** | **0.0** | **1.0** | **2.0** | **3.0** | **4.0** |
| --- | --- | --- | --- | --- | --- |
| **Value** | 8.888889e+01 | 8.933334e+01 | 89.333336 | 8.844444e+01 | 88.444443 |
| **Value.1** | 3.302528e+01 | 3.279955e+01 | 33.044193 | 3.304443e+01 | 33.077759 |
| **Value.2** | 1.021335e+02 | 1.025781e+02 | 101.995079 | 1.021416e+02 | 101.390053 |
| **Value.3** | 3.993277e+01 | 3.973395e+01 | 39.952063 | 3.995084e+01 | 39.925465 |
| **Value.4** | 9.058595e+00 | 8.716287e+00 | 9.002527 | 8.770806e+00 | 9.116302 |
| **Value.5** | -1.220703e-01 | -1.220703e-01 | -0.122070 | -1.220703e-01 | -0.122070 |
| **Value.6** | -1.220703e-01 | -1.220671e-01 | -0.101455 | -1.220703e-01 | -0.122070 |
| **Value.7** | 4.882812e-02 | 7.324219e-02 | 0.048828 | 5.954541e-02 | 0.070984 |
| **Value.8** | 8.611173e+01 | 6.482775e+01 | 54.640110 | 6.890981e+01 | 96.490746 |
| **Value.9** | 8.594911e+01 | 6.467065e+01 | 54.289152 | 6.862698e+01 | 96.437408 |
| **Value.10** | 6.953815e+01 | 6.365610e+01 | 78.064255 | 6.385006e+01 | 71.651939 |
| **Value.11** | 6.952341e+01 | 6.348219e+01 | 78.070122 | 6.389341e+01 | 71.617165 |
| **Value.12** | 7.347964e+01 | 6.907276e+01 | 78.585251 | 7.056946e+01 | 73.853065 |
| **Value.15** | 8.694755e+01 | 8.728450e+01 | 86.752617 | 8.352680e+01 | 87.047241 |
| **Value.17** | 1.708984e-01 | 1.708984e-01 | 0.176335 | 2.107469e-01 | 66.955940 |
| **Value.18** | 6.676537e+00 | 6.678433e+00 | 6.678928 | 6.683522e+00 | 6.681810 |
| **Value.19** | 1.707056e-01 | 1.971337e-01 | 0.202035 | 2.018382e-01 | 0.169874 |
| **Value.20** | -2.221680e-02 | -1.757812e-02 | -0.025314 | -2.579346e-02 | -0.017912 |
| **Value.21** | -9.765625e-02 | -9.399414e-02 | -0.105509 | -1.025398e-01 | -0.092773 |
| **Value.22** | -8.178580e-04 | -8.544920e-04 | -0.000444 | -7.324220e-04 | -0.001099 |
| **Value.23** | -2.441785e-03 | -4.882813e-03 | -0.004883 | -4.882813e-03 | -0.002441 |
| **Value.24** | -2.682734e-03 | -2.441406e-03 | -0.002441 | -2.441406e-03 | -0.008547 |
| **Value.25** | 4.635092e+00 | 4.670718e+00 | 4.737661 | 4.655334e+00 | 4.619235 |
| **Value.26** | 5.515203e+00 | 5.504785e+00 | 5.527457 | 5.528369e+00 | 5.517088 |
| **Value.27** | 9.094960e-12 | 6.462350e-26 | 0.000000 | 6.938890e-17 | 0.477100 |
| **Value.28** | 2.529539e+00 | 2.499080e+00 | 2.498735 | 2.497010e+00 | 2.551732 |
| **Value.29** | 1.790112e+00 | 1.714450e+00 | 1.733658 | 1.742810e+00 | 1.788846 |
| **Value.30** | 2.948017e+03 | 2.595559e+03 | 3034.948975 | 2.927256e+03 | 2962.288818 |
| **Value.31** | 4.263696e-02 | 4.213233e-02 | 0.038722 | 3.664399e-02 | 0.048429 |
| **Value.32** | 4.326805e+01 | 4.204710e+01 | 39.314028 | 3.947325e+01 | 43.487988 |
| **Value.33** | 5.556066e+01 | 5.257672e+01 | 50.647554 | 5.083588e+01 | 55.787552 |
| **Value.34** | 1.606999e+01 | 1.573085e+01 | 15.148449 | 1.592268e+01 | 16.179588 |
| **Value.35** | 1.641634e+00 | 1.648340e+00 | 1.631346 | 1.633428e+00 | 1.642348 |
| **Value.36** | 2.416387e+00 | 2.413350e+00 | 2.420140 | 2.419297e+00 | 2.418474 |
| **Value.37** | 2.413136e+01 | 2.407292e+01 | 24.104695 | 2.411551e+01 | 24.113434 |
| **Value.38** | 5.556101e+00 | 5.473299e+00 | 5.668023 | 5.575968e+00 | 5.543203 |
| **Value.39** | 5.404366e+00 | 1.951795e-01 | 5.443469 | 5.426128e+00 | 2.663213 |
| **Value.40** | 5.391072e+00 | 2.753756e+00 | 5.431991 | 2.334372e+00 | 5.394656 |
| **Value.41** | 2.750668e-01 | 2.750775e-01 | 0.264695 | 2.497224e-01 | 0.276719 |
| **Value.42** | 5.302186e+00 | 5.221852e+00 | 5.412594 | 5.319180e+00 | 5.282396 |
| **Value.43** | 5.197269e+00 | 5.131249e+00 | 5.306494 | 5.215588e+00 | 5.175781 |
| **Value.44** | 2.623604e-01 | 2.860917e-01 | 0.286520 | 2.681234e-01 | 0.256904 |
| **Value.45** | 6.753888e-01 | 6.797291e-01 | 0.712234 | 6.782092e-01 | 0.669836 |
| **Value.46** | -1.984404e-02 | -1.959362e-02 | -0.018988 | -1.955429e-02 | -0.020527 |
| **Value.47** | -1.464844e-02 | -1.464844e-02 | -0.015889 | -1.231584e-02 | -0.014648 |
| **Value.48** | 2.441406e-03 | 2.441406e-03 | 0.002627 | 2.442603e-03 | 0.002441 |
| **Value.49** | 3.297922e+01 | 3.187666e+01 | 32.442486 | 3.223877e+01 | 33.179192 |

Daily freq:

| **Clusters** | **0** | **1** | **2** | **3** | **4** |
| --- | --- | --- | --- | --- | --- |
| **Value** | 88.888885 | 89.049064 | 89.333336 | 88.444443 | 88.444443 |
| **Value.1** | 33.019215 | 32.538140 | 33.025648 | 33.027854 | 33.015691 |
| **Value.2** | 102.066515 | 106.420797 | 102.004723 | 102.113900 | 100.881354 |
| **Value.3** | 39.921209 | 32.346835 | 39.952282 | 39.949092 | 39.937559 |
| **Value.4** | 8.837922 | 8.315419 | 8.844706 | 8.788507 | 9.153443 |
| **Value.5** | -0.116215 | -0.117522 | -0.116226 | -0.117312 | -0.116649 |
| **Value.6** | -0.119978 | -0.110803 | -0.101999 | -0.112059 | -0.122071 |
| **Value.7** | 0.068674 | 0.092573 | 0.060815 | 0.083641 | 0.074095 |
| **Value.8** | 86.447001 | 72.012359 | 54.326875 | 68.585806 | 95.568855 |
| **Value.9** | 86.355310 | 71.907120 | 53.953167 | 68.248283 | 95.496650 |
| **Value.10** | 68.940940 | 64.781824 | 78.303920 | 62.723213 | 71.476193 |
| **Value.11** | 68.913042 | 64.696502 | 78.335646 | 62.742258 | 71.481327 |
| **Value.12** | 73.409279 | 69.346970 | 78.519480 | 70.254250 | 73.977808 |
| **Value.15** | 86.798541 | 87.302058 | 86.661916 | 84.700748 | 87.029495 |
| **Value.17** | 0.191598 | 0.499588 | 0.191919 | 0.194637 | 59.813984 |
| **Value.18** | 6.675508 | 6.677519 | 6.678310 | 6.678306 | 6.680644 |
| **Value.19** | 0.173209 | 0.196073 | 0.202923 | 0.202350 | 0.171312 |
| **Value.20** | -0.020254 | -0.009822 | -0.029148 | -0.014685 | -0.013112 |
| **Value.21** | -0.093945 | -0.087055 | -0.113180 | -0.094047 | -0.089813 |
| **Value.22** | -0.000783 | -0.000838 | -0.000347 | -0.000623 | -0.001089 |
| **Value.23** | -0.003556 | -0.004492 | -0.004809 | -0.004492 | -0.003030 |
| **Value.24** | 0.003539 | 0.003554 | 0.007356 | 0.004901 | -0.001831 |
| **Value.25** | 4.631555 | 4.381945 | 4.740848 | 4.656216 | 4.616496 |
| **Value.26** | 5.516138 | 5.357307 | 5.525942 | 5.531553 | 5.517676 |
| **Value.27** | 0.004328 | 0.070368 | -0.000058 | 0.000101 | 0.524546 |
| **Value.28** | 2.530757 | 2.405159 | 2.498292 | 2.495281 | 2.551110 |
| **Value.29** | 1.789720 | 1.542577 | 1.730894 | 1.745082 | 1.787859 |
| **Value.30** | 2993.300292 | 2229.976608 | 3068.365781 | 2901.671623 | 2988.174699 |
| **Value.31** | 0.041063 | 0.035699 | 0.026231 | 0.027250 | 0.045009 |
| **Value.32** | 665.835460 | 620.236961 | 833.724662 | 659.685182 | 684.915909 |
| **Value.33** | 55.651587 | 53.694417 | 50.670845 | 50.685335 | 56.831407 |
| **Value.34** | 1007.423711 | 916.621528 | 786.301759 | 1255.690592 | 1253.242553 |
| **Value.35** | 1.641570 | 1.647412 | 1.631241 | 1.638648 | 1.641890 |
| **Value.36** | 2.416058 | 2.415546 | 2.419910 | 2.419521 | 2.419157 |
| **Value.37** | 24.125162 | 24.115512 | 24.098618 | 24.113763 | 24.113658 |
| **Value.38** | 5.542729 | 5.191578 | 5.669555 | 5.582645 | 5.536668 |
| **Value.39** | 2.861873 | 2.417474 | 2.904849 | 3.303713 | 2.705874 |
| **Value.40** | 2.849946 | 2.769942 | 2.888402 | 2.408888 | 3.010775 |
| **Value.41** | 0.275186 | 0.314090 | 0.264024 | 0.248652 | 0.276662 |
| **Value.42** | 5.296945 | 4.922225 | 5.414001 | 5.322879 | 5.282844 |
| **Value.43** | 5.189770 | 4.849870 | 5.306970 | 5.219730 | 5.175668 |
| **Value.44** | 0.257248 | 0.368582 | 0.283561 | 0.271194 | 0.253914 |
| **Value.45** | 0.673725 | 0.690429 | 0.712259 | 0.679739 | 0.669001 |
| **Value.46** | -0.020242 | -0.019951 | -0.018492 | -0.020066 | -0.020400 |
| **Value.47** | -0.014153 | -0.013218 | -0.016125 | -0.011912 | -0.012453 |
| **Value.48** | 0.002468 | 0.002582 | 0.003008 | 0.002909 | 0.002442 |
| **Value.49** | 32.981024 | 28.861054 | 32.442882 | 32.243907 | 33.216310 |

Standard deviation for process variable:

15mint freq:

| **Clusters** | **0.0** | **1.0** | **2.0** | **3.0** | **4.0** |
| --- | --- | --- | --- | --- | --- |
| **Value** | 7.700199 | 17.872563 | 12.283947 | 17.542307 | 25.789167 |
| **Value.1** | 1.518416 | 11.439396 | 3.189706 | 1.154665 | 1.349135 |
| **Value.2** | 2.474050 | 15.569630 | 3.460312 | 2.509406 | 4.678791 |
| **Value.3** | 1.593779 | 14.610708 | 3.842798 | 0.402559 | 0.398569 |
| **Value.4** | 0.979650 | 3.692812 | 1.228225 | 0.855706 | 1.295486 |
| **Value.5** | 0.011888 | 0.012007 | 0.011629 | 0.011372 | 0.009462 |
| **Value.6** | 0.015273 | 0.017309 | 0.018245 | 0.017979 | 0.009095 |
| **Value.7** | 0.056841 | 0.057318 | 0.059767 | 0.060801 | 0.043020 |
| **Value.8** | 18.946114 | 17.159691 | 17.920315 | 18.063406 | 3.108436 |
| **Value.9** | 19.054011 | 17.282111 | 18.008598 | 18.191903 | 3.101051 |
| **Value.10** | 11.242669 | 19.468137 | 11.446694 | 12.550909 | 8.837900 |
| **Value.11** | 11.235865 | 19.458958 | 11.438465 | 12.529921 | 8.833338 |
| **Value.12** | 9.430684 | 11.002557 | 10.190470 | 7.782465 | 6.726861 |
| **Value.15** | 6.467739 | 6.228238 | 6.795284 | 6.616242 | 6.054673 |
| **Value.17** | 10.372024 | 20.048714 | 9.569854 | 25.403597 | 31.816090 |
| **Value.18** | 0.039071 | 0.689266 | 0.125281 | 0.039618 | 0.039838 |
| **Value.19** | 0.018427 | 0.128549 | 0.028248 | 0.020678 | 0.016108 |
| **Value.20** | 0.024708 | 0.023956 | 0.022277 | 0.025728 | 0.017533 |
| **Value.21** | 0.030978 | 0.025809 | 0.028951 | 0.030106 | 0.017947 |
| **Value.22** | 0.000438 | 0.000299 | 0.000474 | 0.000335 | 0.000319 |
| **Value.23** | 0.001943 | 0.001742 | 0.001973 | 0.001688 | 0.001361 |
| **Value.24** | 0.016061 | 0.016828 | 0.019858 | 0.018155 | 0.016484 |
| **Value.25** | 0.049451 | 1.616335 | 0.294766 | 0.057753 | 0.035459 |
| **Value.26** | 0.053717 | 1.369901 | 0.346037 | 0.070033 | 0.039808 |
| **Value.27** | 0.197203 | 0.327916 | 0.161331 | 0.293015 | 0.391561 |
| **Value.28** | 0.034583 | 0.647410 | 0.121705 | 0.029032 | 0.027243 |
| **Value.29** | 0.046920 | 0.617818 | 0.119348 | 0.047019 | 0.021032 |
| **Value.30** | 250.853387 | 1194.741470 | 425.560064 | 209.835496 | 217.834940 |
| **Value.31** | 0.134985 | 0.027669 | 0.187092 | 0.028199 | 0.020282 |
| **Value.32** | 2672.793084 | 2663.219526 | 2840.774439 | 2929.336438 | 2644.022637 |
| **Value.33** | 352.929051 | 2.582622 | 349.565728 | 2.964895 | 721.848413 |
| **Value.34** | 3037.631967 | 2788.685490 | 2809.251649 | 3232.888697 | 3134.004425 |
| **Value.35** | 0.020517 | 0.432492 | 0.081611 | 0.018149 | 0.012853 |
| **Value.36** | 47168.929217 | 0.616865 | 0.137957 | 0.013566 | 0.011405 |
| **Value.37** | 0.108764 | 6.319231 | 1.319588 | 0.072781 | 0.082143 |
| **Value.38** | 0.067275 | 1.671035 | 0.315971 | 0.045473 | 0.047929 |
| **Value.39** | 2.710766 | 2.590749 | 2.753607 | 2.704863 | 2.705992 |
| **Value.40** | 2.699663 | 2.584324 | 2.747045 | 2.706417 | 2.683280 |
| **Value.41** | 0.008486 | 0.211781 | 0.043241 | 0.015369 | 0.006737 |
| **Value.42** | 0.061646 | 1.748977 | 0.316452 | 0.043001 | 0.047014 |
| **Value.43** | 0.058717 | 1.718749 | 0.308506 | 0.046484 | 0.044037 |
| **Value.44** | 0.032896 | 0.209499 | 0.053358 | 0.022779 | 0.028870 |
| **Value.45** | 0.018644 | 0.143524 | 0.024813 | 0.014541 | 0.016515 |
| **Value.46** | 0.001460 | 0.001280 | 0.001535 | 0.001794 | 0.001073 |
| **Value.47** | 0.006735 | 0.007359 | 0.007269 | 0.006811 | 0.007605 |
| **Value.48** | 0.000344 | 0.000310 | 0.000501 | 0.000467 | 0.000144 |
| **Value.49** | 0.633746 | 12.455947 | 2.384754 | 0.576777 | 0.303047 |

Daily freq:

| **Clusters** | **0** | **1** | **2** | **3** | **4** |
| --- | --- | --- | --- | --- | --- |
| **Value** | 4.253504 | 11.574253 | 9.219861 | 13.233825 | 21.352192 |
| **Value.1** | 0.341727 | 4.468869 | 2.278120 | 0.093618 | 0.094159 |
| **Value.2** | 0.584654 | 8.756381 | 2.143114 | 0.698245 | 1.419579 |
| **Value.3** | 0.564162 | 7.143233 | 2.175609 | 0.043671 | 0.054068 |
| **Value.4** | 0.551771 | 2.023239 | 0.715425 | 0.098011 | 0.982376 |
| **Value.5** | 0.007275 | 0.005604 | 0.003348 | 0.002262 | 0.003787 |
| **Value.6** | 0.008923 | 0.008262 | 0.010074 | 0.010130 | 0.002546 |
| **Value.7** | 0.033902 | 0.028868 | 0.026213 | 0.033116 | 0.017413 |
| **Value.8** | 18.874258 | 18.019629 | 17.868789 | 18.678869 | 1.757546 |
| **Value.9** | 18.983859 | 18.148544 | 17.957308 | 18.814775 | 1.752949 |
| **Value.10** | 11.072813 | 20.159471 | 11.312011 | 12.845179 | 8.802581 |
| **Value.11** | 11.065979 | 20.151469 | 11.304000 | 12.823648 | 8.798973 |
| **Value.12** | 9.274174 | 7.177522 | 9.823988 | 6.757584 | 6.988384 |
| **Value.15** | 1.650122 | 2.263707 | 1.649400 | 3.117622 | 1.382871 |
| **Value.17** | 7.156894 | 20.635807 | 7.953718 | 22.928100 | 15.744889 |
| **Value.18** | 0.007778 | 0.412737 | 0.048630 | 0.009662 | 0.010737 |
| **Value.19** | 0.015073 | 0.047081 | 0.017538 | 0.018270 | 0.008935 |
| **Value.20** | 0.018253 | 0.014259 | 0.013543 | 0.015937 | 0.008530 |
| **Value.21** | 0.024322 | 0.016947 | 0.020069 | 0.021429 | 0.008564 |
| **Value.22** | 0.000412 | 0.000272 | 0.000440 | 0.000300 | 0.000310 |
| **Value.23** | 0.001335 | 0.001256 | 0.001138 | 0.000878 | 0.000732 |
| **Value.24** | 0.006382 | 0.005751 | 0.009498 | 0.003887 | 0.010134 |
| **Value.25** | 0.041058 | 1.077860 | 0.239781 | 0.056162 | 0.027231 |
| **Value.26** | 0.022744 | 0.983714 | 0.277947 | 0.062755 | 0.016304 |
| **Value.27** | 0.186823 | 0.243982 | 0.143599 | 0.278573 | 0.260664 |
| **Value.28** | 0.029097 | 0.392103 | 0.100364 | 0.030095 | 0.027775 |
| **Value.29** | 0.033850 | 0.370508 | 0.096458 | 0.047499 | 0.017852 |
| **Value.30** | 129.360548 | 579.020536 | 302.696328 | 93.808377 | 70.478907 |
| **Value.31** | 0.036868 | 0.014171 | 0.053597 | 0.013848 | 0.007999 |
| **Value.32** | 323.144672 | 431.135645 | 391.135689 | 420.678133 | 404.148217 |
| **Value.33** | 146.694010 | 2.670559 | 134.276458 | 3.072372 | 287.037561 |
| **Value.34** | 483.181468 | 502.253148 | 509.645274 | 356.055281 | 800.468153 |
| **Value.35** | 0.019144 | 0.248044 | 0.049467 | 0.013743 | 0.011326 |
| **Value.36** | 4857.810102 | 0.340173 | 0.093631 | 0.006029 | 0.006713 |
| **Value.37** | 0.098913 | 3.392459 | 0.846765 | 0.057569 | 0.080222 |
| **Value.38** | 0.048856 | 0.936434 | 0.239577 | 0.041413 | 0.043152 |
| **Value.39** | 0.427718 | 0.632979 | 0.428389 | 0.484648 | 0.462105 |
| **Value.40** | 0.421509 | 0.573426 | 0.429988 | 0.454415 | 0.480307 |
| **Value.41** | 0.007264 | 0.119765 | 0.035250 | 0.014891 | 0.005173 |
| **Value.42** | 0.046624 | 1.168877 | 0.263151 | 0.039551 | 0.042266 |
| **Value.43** | 0.045421 | 1.156296 | 0.256301 | 0.043542 | 0.038727 |
| **Value.44** | 0.026371 | 0.108758 | 0.039185 | 0.015163 | 0.020696 |
| **Value.45** | 0.017003 | 0.074846 | 0.017502 | 0.013414 | 0.015938 |
| **Value.46** | 0.001075 | 0.000737 | 0.001162 | 0.001380 | 0.000755 |
| **Value.47** | 0.004713 | 0.005846 | 0.004601 | 0.004552 | 0.006798 |
| **Value.48** | 0.000222 | 0.000124 | 0.000292 | 0.000296 | 0.000069 |
| **Value.49** | 0.401865 | 7.658277 | 1.979728 | 0.588633 | 0.278532 |

CSV for 1st dataset production

The CSV is attached naming **“1st dataset production.csv”**

In this csv file I have saved the production dataset with a separate column of outliers which indicates if a datapoint belongs to outlier or not.

CSV for 2nd dataset

1. **for all cluster in daily frequency**

The CSV is attached naming **“description\_of\_dailyfreq\_process\_variable.csv”**

This csv contains the whole description of daily frequency process variable.

1. **for all cluster in 15 min freq.**

The CSV is attached naming **“description\_of\_15minfreq\_process\_variable.csv”**

This csv contains the whole description of 15minute frequency process variable.

1. Discussion

If we look through the result of each dataset we can observe that, outcome from the 1st dataset was good enough. The clustering was perfect. The test dataset performed well. The outlier was isolated smoothly as it was a small number.

But if we look into the 2nd dataset, the scenario is a bit complex there. I managed to calculate the suitable process variables according to the date and cluster. Then I calculated the mean, median, standard deviation of each process variable where the standard deviation for some particular variables, such as: **Value.30, Value.32, Value.33, Value.34, Value.36** were much higher than the other values. Which was clearly caused by the outliers lying in the data. I could have removed the outliers from the data but as it was given by the client, I didn’t do the cleaning part with the 2nd dataset. As the ‘mean’ was fluctuating much I preferred working with the ‘median’ value as it was quite nearer to each other. So, taking the ‘median’ and ‘standard deviation’ of each value, the production can be made for any cluster condition.