- Anomaly Detection Methods:

Anomaly detection is the process of discover the event or the points which are unexpected at this position of the dataset or deviates from the normal pattern of the dataset.

So, the detection of those points very important; because it give us an early step to make the emergency movements to control that un usual change.

We used many techniques to reach best one to apply it on our way of the project:
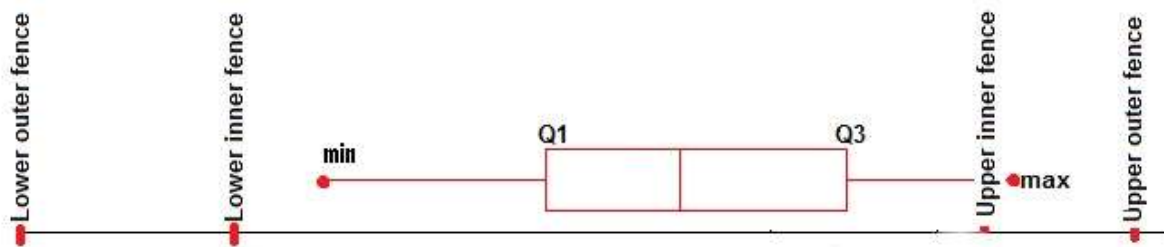
1- Tukey's box plot Method:

Figure 1 https://miro.medium.com/max/1342/1*vQyvZ7yZpLcFk7eDdoc5lg.png

in this method we depend on the pox plot to determine if the point is outlier or not and not only that it gives us the ability to decide if this outlier is possible or probable outlier point; by calculate the following parameters:

- 25th percentile (Q1)
- 75th percentile (Q3)
- interquartile range (IQR = Q3 – Q1)
- Lower inner fence: Q1 – (1.5 * IQR)
- Upper inner fence: Q3 + (1.5 * IQR)
- Lower outer fence: Q1 – (3 * IQR)
- Upper outer fence: Q3 + (3 * IQR)

Then decide if the point between the inner fence and outer fence it considered as a possible outlier point. And if the point lies outside the outer fence, it will be considered as probable outlier.

The following two graphs indicate what we said. on a random grid of the whole dataset. The red points are the anomalies points of this grid.

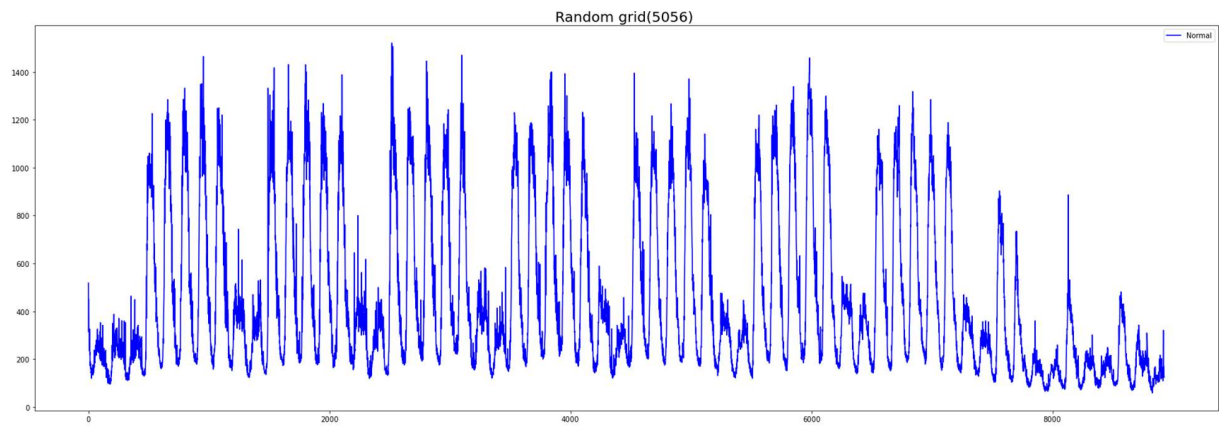The other two graphs indicate the anomalies points on the whole dataset.
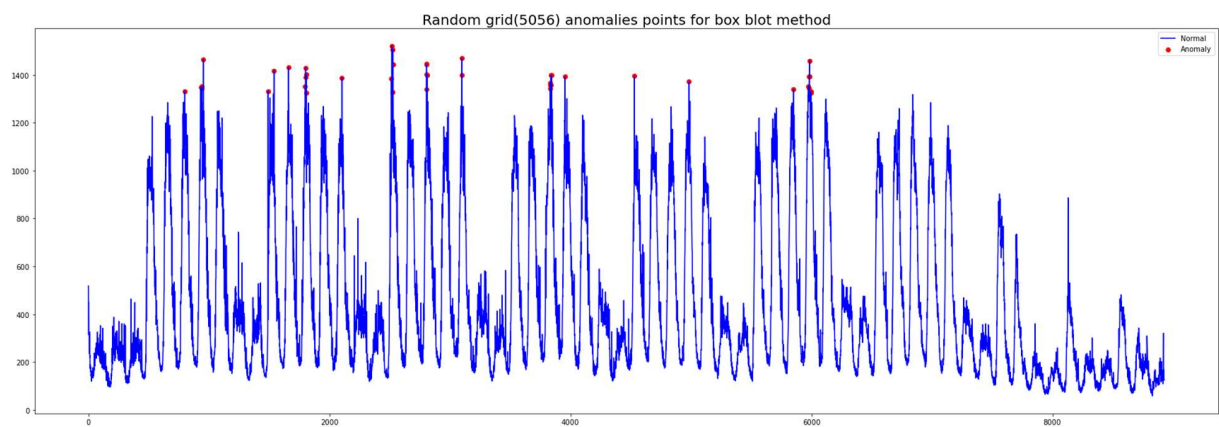
Figure 2 Random grid (5056)

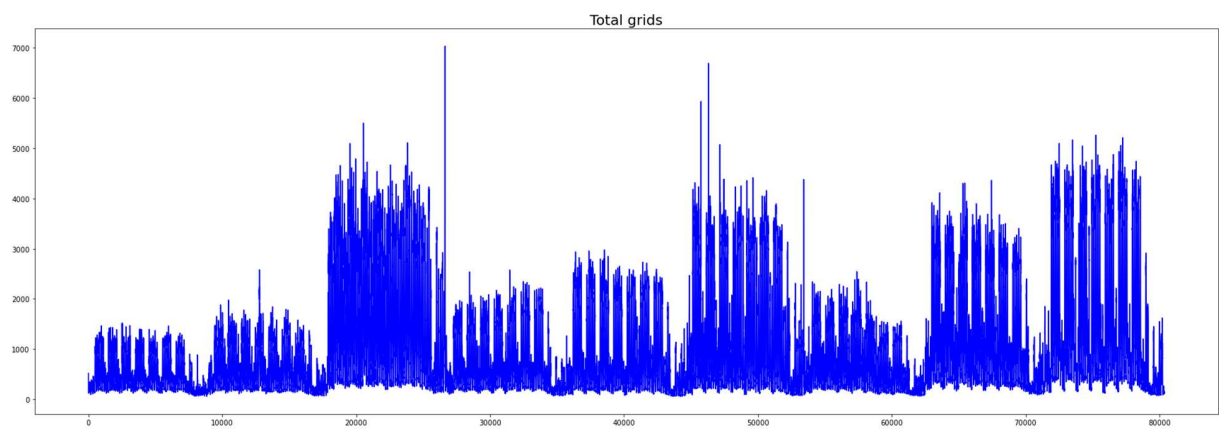*Figure 3 Random grid anomalies points for box plot method*

Figure 4 Total 9 grids

After using the box plot method as we shown before we will indicate the cons and pros of this method.

Pros:

- Quick representation and very fast method
- Easy to implement

Cons:

- The number of points is so small (215 anomalies points from total dataset points 80253).
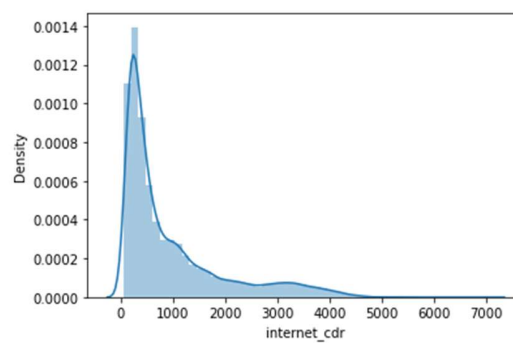- Not efficient

Figure 6 distribution plot

- If the data is skewed it will miss some anomalies our data.

2- Isolation Forest Method:

In this method we will depend on the detection using some Machine Learning algorithms. In the we will depend on Isolation Forest.

One of those algorithms is the Isolation Forest method. Isolation Forest build using the decision trees which depend on the points that go deeper into the tree are not anomalies and points which go short distance have big probability to be anomalies, and it is unsupervised learning model which used without labeled data.

The algorithm goes by selecting a sample of the dataset then branch it on the binary tress by setting a threshold if the sample we selected is less than this threshold it will be in the left branch and if it not it will be in the right branch. This process repeated until we every point in the dataset is isolated.

After building the algorithm we reached to the following output:

The following two graphs indicate what we said. on a random grid of the whole dataset. The red points are the anomalies points of this grid.

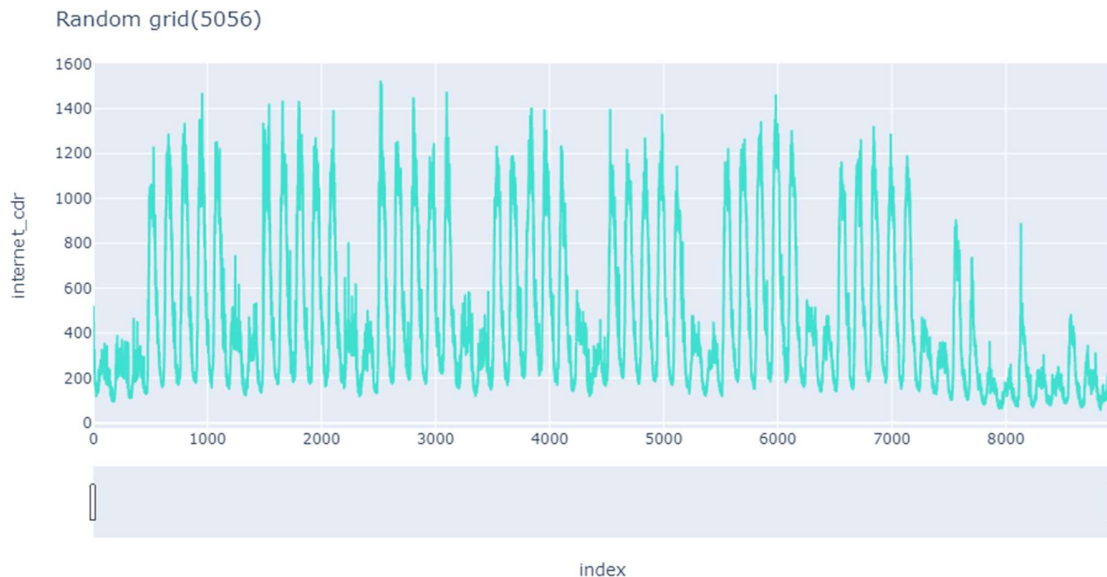The other two graphs indicate the anomalies points on the whole dataset.

Figure 7 random grid (5056)

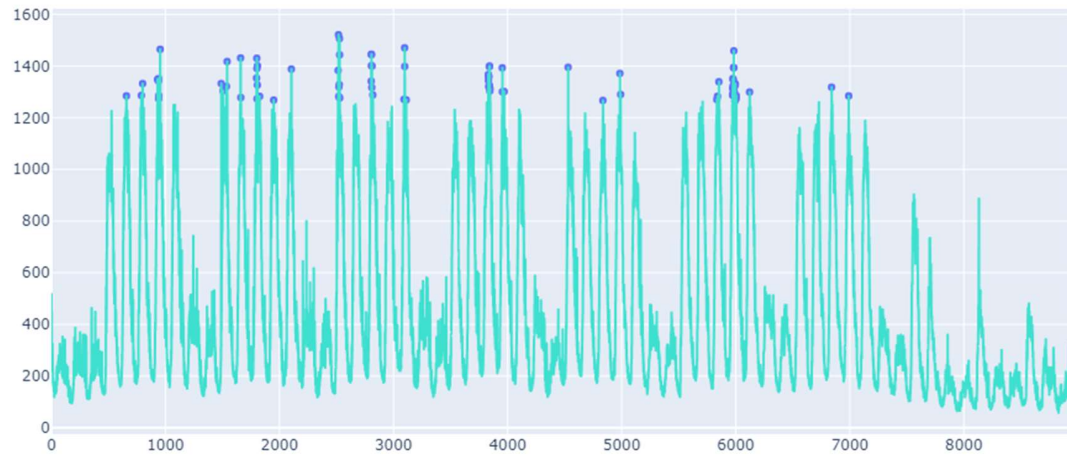Random grid(5056) anomalies points for isolation forest

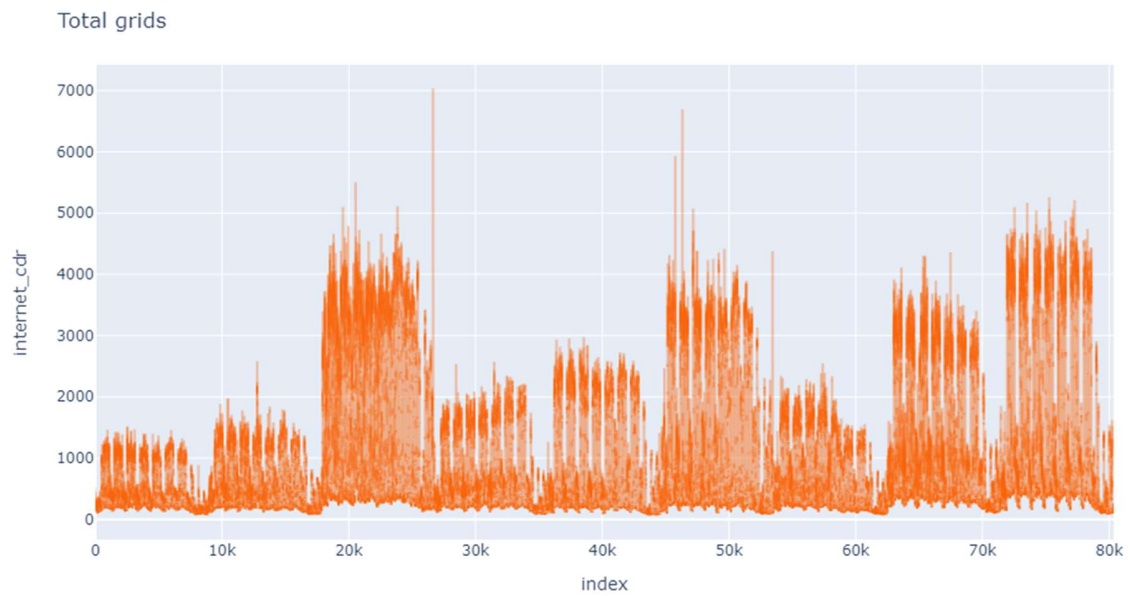*Figure 8 Random grid anomalies points for isolation forest method*

Total grids

Figure 9 Total 9 grids

Total grids anomalies points for isolation forest

Figure 10 Total grids anomalies points for isolation forest method

By looking on the previous graphs and results of the isolation forest method we will indicate the advantages and disadvantages of it.

Pros:

- We could use any number of the features which will make the model more accurate.
- It gives bigger number of the anomaly's points (1258 anomalies points from total dataset points 80253) than the box plot method.

Cons:

- There is biasing for the mode due to the brunching process.
- The greater number of features is good for the model but it will affect badly on the performance and the speed of the model so we should use our feature carefully.

3- LSTM Autoencoders Method:

In this method we will depend on the detection using the forecasting by Deep Learning algorithms. In the forecasting methods we depend on predict the next point with the addition of some noise and make comparison of this point and the true point at this timestamp by finding the difference between the two points then add threshold finally find the anomalies by

compare the difference of the two points with this threshold (we used the Mean absolute error MAE).

Autoencoders are type of self-supervised learning model which are a neural network that learn from the input data. We use autoencoder because the Principal Component Analysis (PCA), which we used in the previous method we depend on the linear algebra to do the models, but by using autoencoders we depended on the non-linear transformation like by use the activation functions; those non-linearity gives us the ability to go deep in the number of the neural network layers.

Long Short-Term Memory (LSTM) is a type of artificial recurrent neural network (RNN). which are designed to handle sequential data, with the previous step's output being fed as the current step's input.
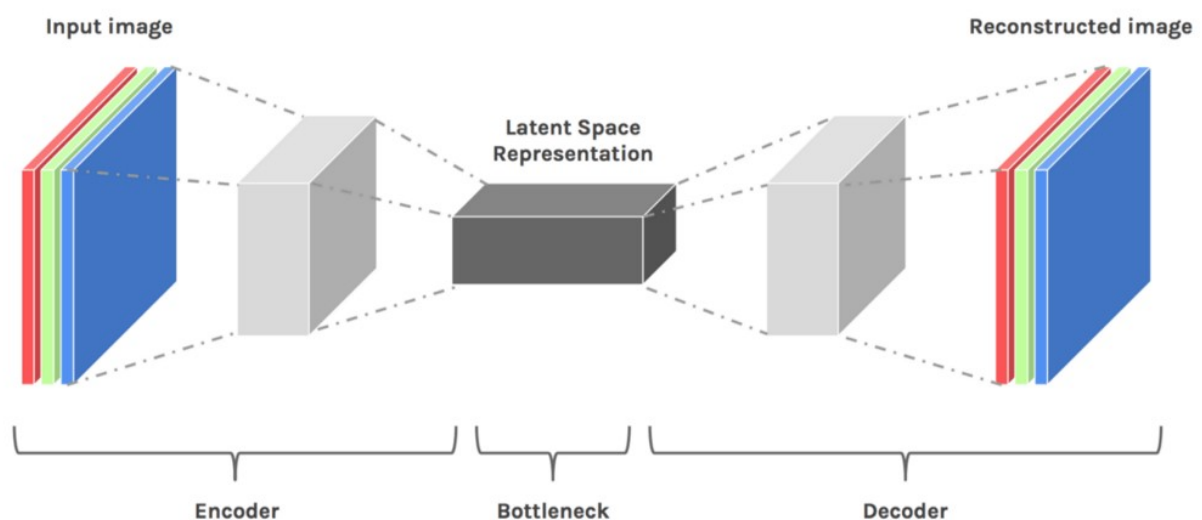

Figure 11 https://cdn.hackernoon.com/hn-images/1*8ixTe1VHLsmKB3AquWdxpQ.png

We apply some dimensionality reduction on our dataset by use encoder to make the dimension small then use the decoder to get it back and that minimize the reconstruction loss. In fact, that will make us lose some information but it gives us the ability to know the main pattern of the information and thought that we could define any information out hits pattern under sone threshold will be outlier.

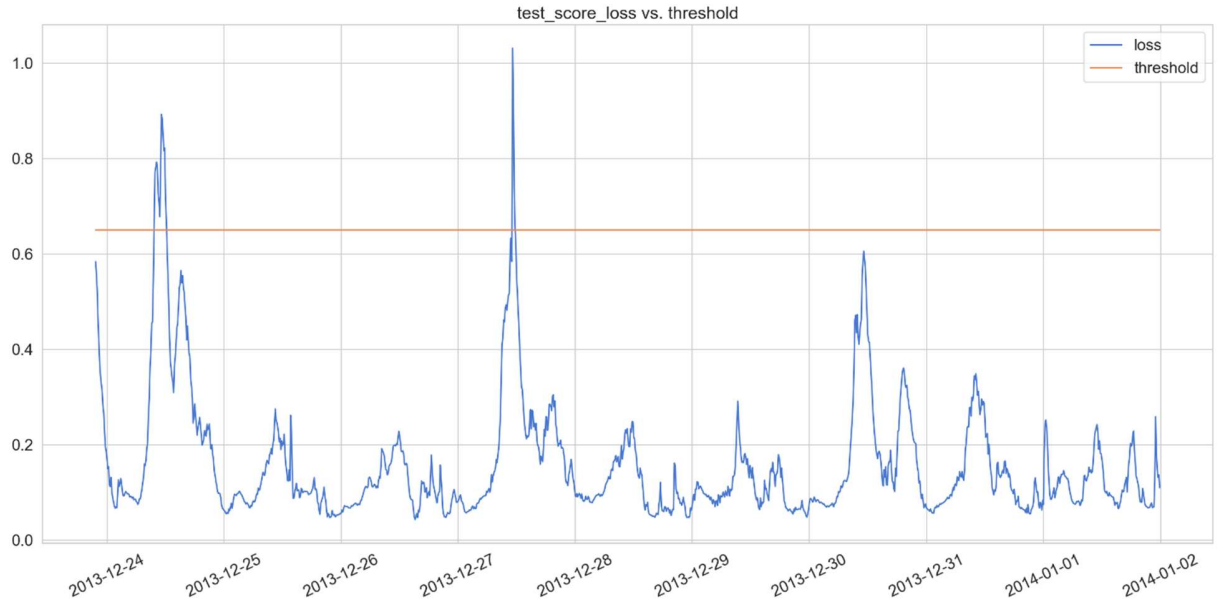After applying the model, we get the following results:

Figure 12 test_score_loss vs. threshold

To determine the cutoff point we use the Mean Absolute Error (MAE). We use the MAE because it so sensitive toward outliers. MAE find the mean absolute error between the actual value y and predicted value $\hat{y}$ of the dataset then find the threshold like the following equation:

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} |y - \hat{y}_i| \qquad (1)$$

When we apply the threshold to the predicted values which will give us the anomalies at the points which corresponding to the locations of the signal which above the threshold line the previous graph, we get the following graph for the anomalies.
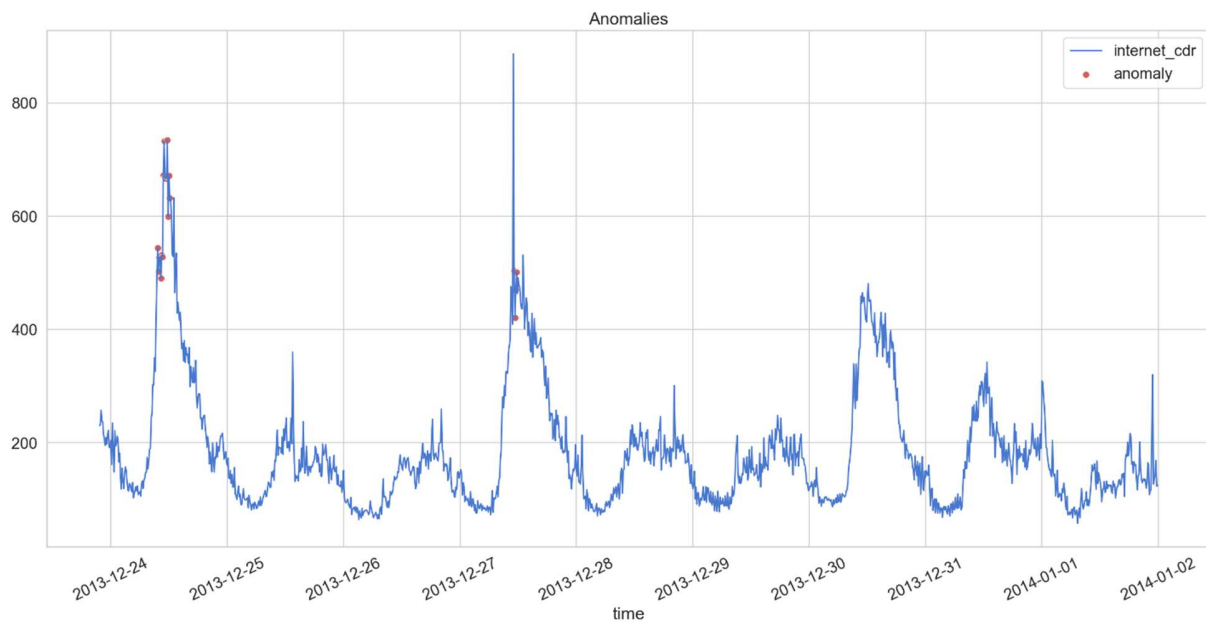
Figure 13 predicted values vs. the anomalies for the LSTM Autoencoders method

Pros:

-   Autoencoder could easily deals with data which have high dimensions.
-   As the good use of the activation function, it gives big ability to deal with complex datasets.

Cons:

-   Need big data to train the deep learning models.
-   High cost and low speed when deal with those big datasets.

4- Seasonal-Trend Decomposition Method:

Now we will go to the final method which is decomposition. Signal decomposition aims to analysis our signal to its main three components Seasonal, trend and the residual (S, T, R). Seasonal is the signal component which contain the most rapidly pattern which occurs regular every cerin time. Trend contain the general shape of the data over the whole dataset and finally the residual is the rest of the signal after extract the seasonal and trend of it, it is in somehow a random part over the signal which indicate it.
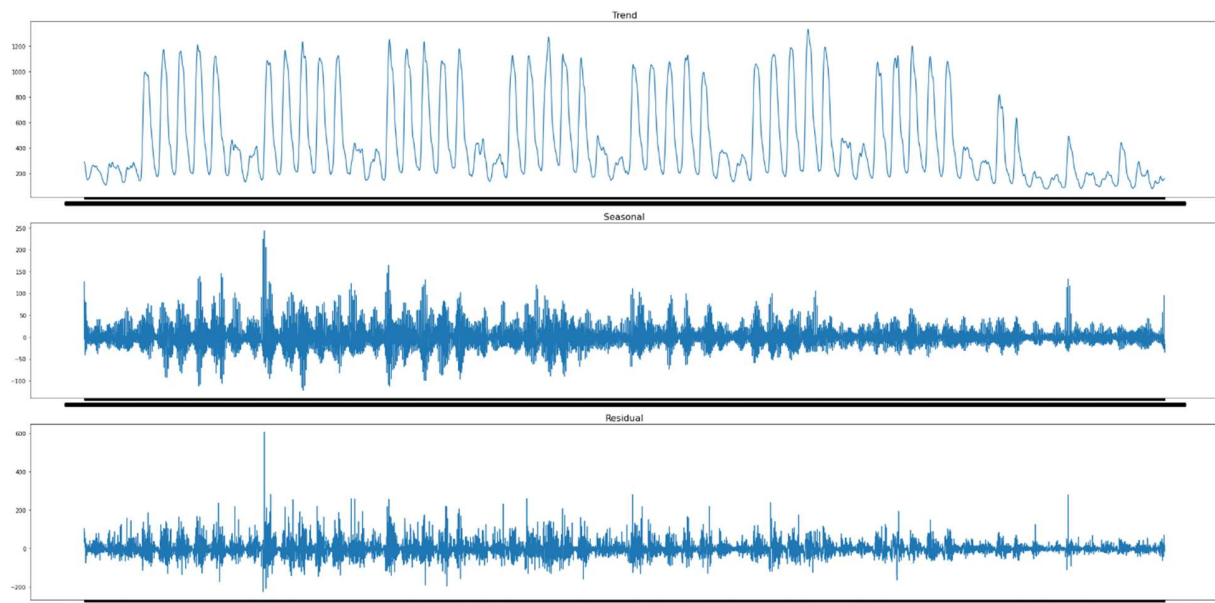
Figure 14 random grid 5056 decomposition three parts (S, T, R)

To make the residual more obvious to us we will plot the signal with and without the residual component as the following:
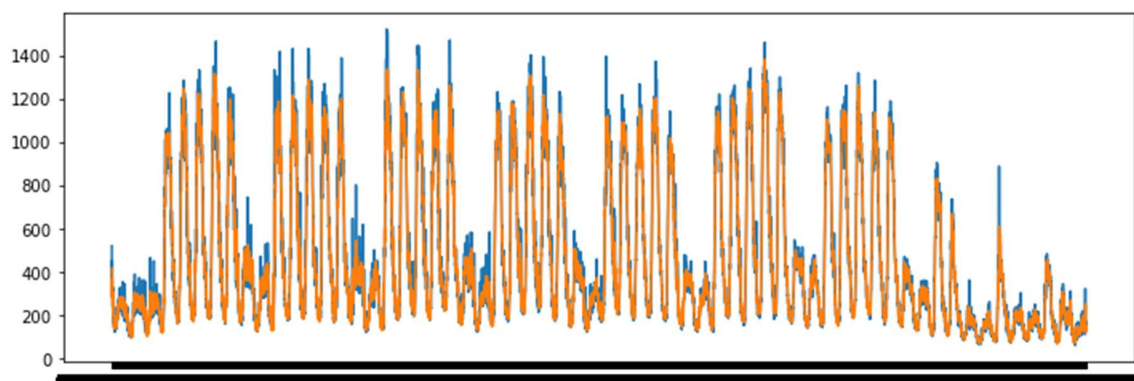

Figure 15 signal with and without the residual component

The residual will be our focus here, we will first analysis the signal to its main three component and take the residual to work on it.

We will apply the model by define the threshold which depend on the he confidence interval, then apply it for the residual then decide if this point is an anomaly or not.

We will define certain two limits lower limit and the upper limit on the dataset points. First is the lower limit which define through this equation:

$$Lower = \mu - 3 * S \qquad (2)$$

Second is the upper limit which define through this equation:

$$Upper = \mu + 3 * S \qquad (3)$$

After applying those constrain on the residual component, we will have the ability to define if this point is anomaly or not.

By input every grid of the nine grids on this condition we will the following outputs one for rand grid (5056) and after combine the all nine grids we will have the other graph.
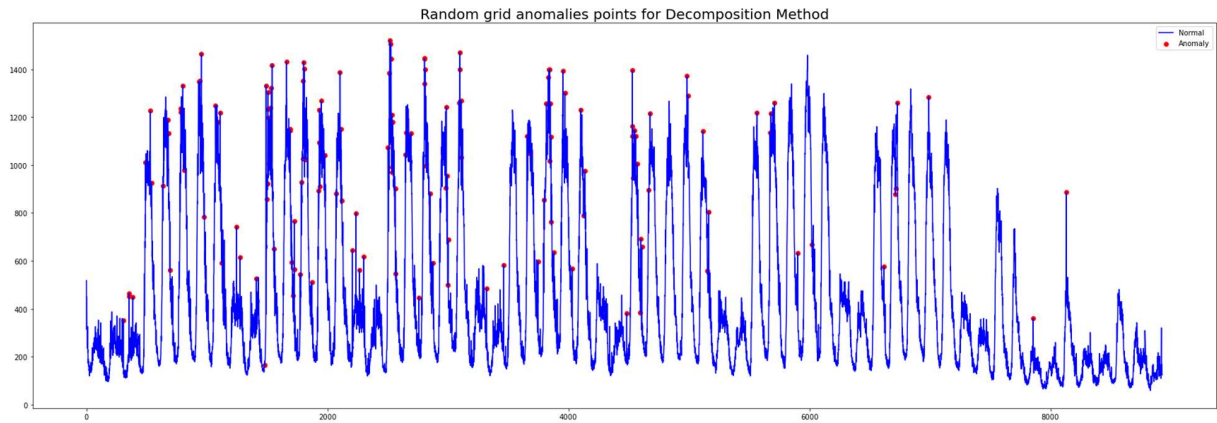

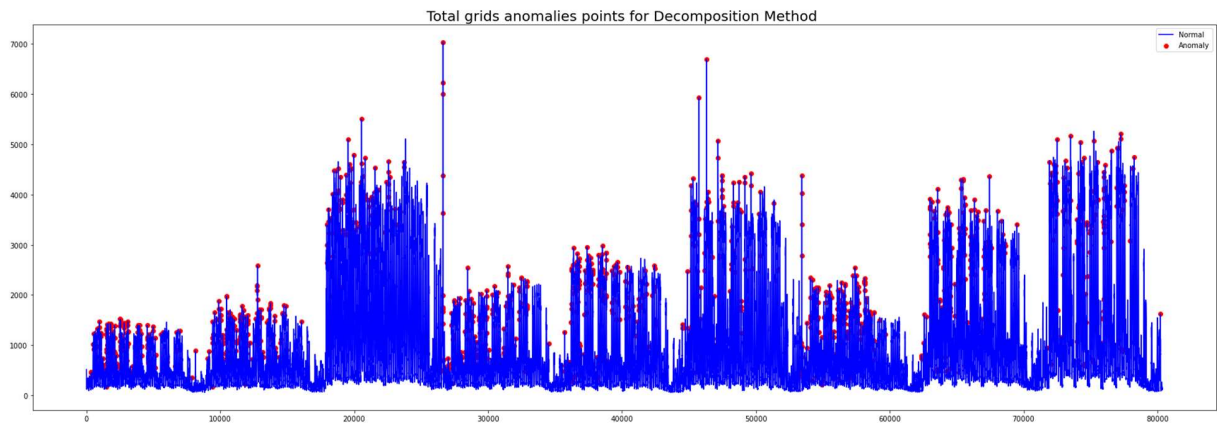*Figure 16 Figure 8 Random grid (5056) anomalies points for Decomposition method*

Figure 17 Total grids anomalies points for Decomposition method

Let us give our intuition of the results of this method by showing its cons and pros:

Pros:

- So simple method didn't need to the forecasting to define the anomalies.
- Robust which could work under any type of the datasets.

Cons:

- If there is a suddenly change made by us on the process, we work we should study the process individually after and before this change happen; as we using the confidence interval to determine the threshold.

- **What is next?**

After determining the anomalies points what we should do about them? In the most common application like Microsoft anomaly detection, they have some web application to send some emails to the concerned persons about the sudden and didn't expected change, then they take the right decision to solve this issue.

But in our scope here we concerned with if this point is outlier or point of interest.
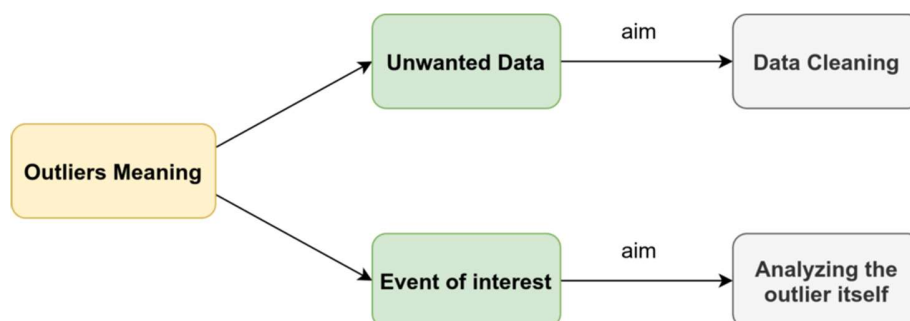

Figure 18 https://i0.wp.com/neptune.ai/wp-content/uploads/Time-series-outliers.png?resize=1000%2C345&ssl=1

So, if the point we detected as outlier is a data that we did not need we will do on it some cleaning data processing. But in the other hand some time did not mean if there is some unusual event that we didn't need it. Some event most be important to us because might this event will happen in the future so by studying it will make us have the ability to avoid this sudden change in the future by handle it by some control flow process.

Here we after studying those points we will make all anomalies point as a nan value so will handle it in the coming part, **Missing value imputation**.

References:

1- Time-Series Anomaly Detection Service at Microsoft Hansheng Ren, Bixiong Xu, Yujing Wang, Chao Yi, Congrui Huang, Xiaoyu Kou∗ Tony Xing, Mao Yang, Jie Tong, Qi Zhang Microsoft Beijing, China.
2- Iglewicz B, Hoaglin DC (1993) How to detect and handle outliers. ASQC Quality Press.
3- https://towardsdatascience.com/detecting-and-treating-outliers-in-python-part-1-4ece5098b755
4- https://blogs.sap.com/2020/12/21/anomaly-detection-in-time-series-using-seasonal-decomposition-in-python-machine-learning-client-for-sap-hana/
5- https://neptune.ai/blog/anomaly-detection-in-time-series
6- Raghavendra Chalapathy and Sanjay Chawla. 2019. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407 (2019).