**Process Flow**

**Problem Statement:**

Your goal is to build an outlier detection algorithm within 1 week of receiving this assignment

for the provided data set. The data consists of values of a variable observed every minute every

minute over a period of 1 week. Consider the training set to be the data for the first 6 days and

the test set as the last day data.

Expected output:

● Build the algorithm in python as a function where the input is the test data. The output is

a plot of the test data with potential outlier points highlighted.

● Share a public Github repository containing your code with us for review

● Your code should convey your overall approach clearly. It is a good idea to annotate your

code to make it easy to understand.

What are we looking for?

● We are more interested in your approach towards solving this problem. So no worries if

the output is not perfect...though you do get bonus points for a good output.

**Plan of Action**

As per the given instructions it has to be splited into train and test set and no target (label) is given So I am considering the problem as a unsupervised problem and also there is no any additional info about data like source of the data, domain of the data, percetange of outlier that can occur for this type of data etc.,

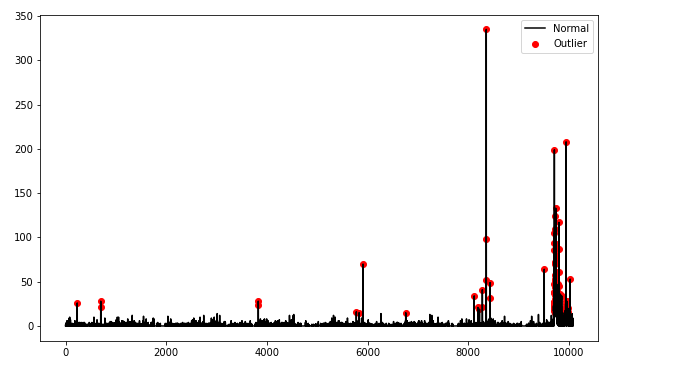
First Level Approach : I will apply IQR, z-score and other traditional outlier detection methods to detect the outliers

Second Level Approach : will apply some unsupervised or time series based outlier/anomaly detection methods

isolation Forest, Local outlier factor etc.,

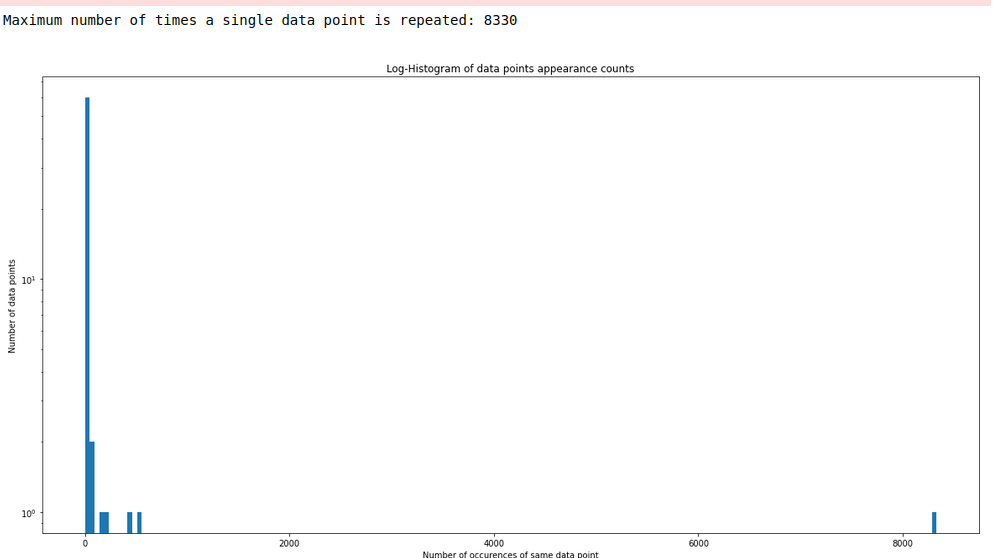
**First Level Approach results:**

Z-Score:

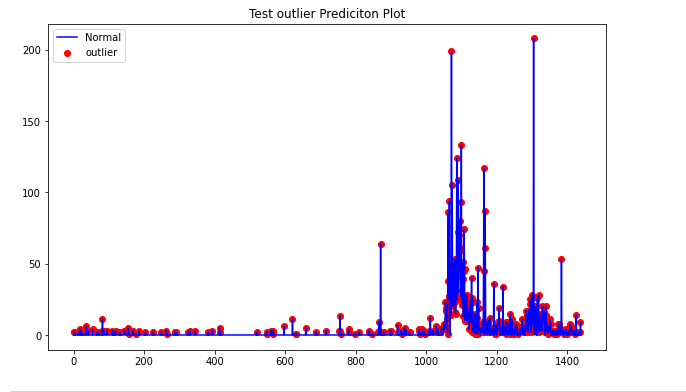
Z score has detected the outliers but it has detected very less outliers So it is better to use unsupervised methods to detect proper outliers I am trying to use IsolationForest and Local Outlier Factor methods.

The biggest challenge is the most of the values are nearer to zero 0 = 8330 out of 10080

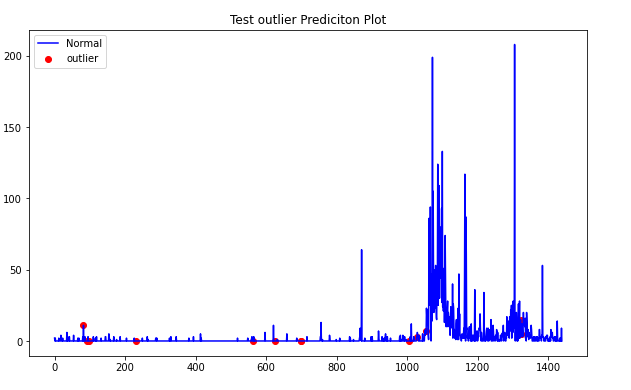
The confusion here is whether I should ignore these zero values or not you can observe in the bellow graph. I am considering zero values also for futher analysis



**Isolation Forest Results on Test Data :**

****

**Local Outlier Factor Results on TestData**

****

**If you can check among the all the results isolation forest has given good results.**

**So I have considered isolation forest as my final model or approach**

**Please check** finalOutlierDetection **Function in Outlierdetection.ipynb and FinalTestCode.ipynb files**

**Note: please try to open files using jupyter notebook**