Today, I’m going to talk about what I’ve already done for my project. To recap, for my project, I wanted to use regression and clustering to investigate in the trend and the efficiency of energy use in different timing based on 9 attributes given in the dataset.

The first thing that I want to see is the overall look of how the energy was used for the household according to month, year and type to get the basic idea of how the data looked like. I actually wanted to plot the data by minutes or hours rather than by month because normally for different types of appliances, the usage for different month. However, since the data set is quite big, rendering the graph with big data is quite a challenge. So, grouping by month, year and type stills gives us a general idea too.

Normally when we take a look at the energy bill, we only actually notice that if the total energy used is decreased or increased and assumed that decreased total energy means higher efficiency. However, it’s not usually that case. Total energy consists of the active energy, which is the energy that does the actual work, and the reactive energy, which is the loss energy that doesn’t do any actual work. Thus, in order to really specify the efficiency for the energy, we need to address the power factor that tells us the ratio of the actual energy that you use over the energy that you need to pay for. Higher power factor should be the main criteria when we want to talk about efficiency.

If you take a look more carefully at the graph, you can see that even though the active energy increases and deceases in different months, the reactive energy remains roughly the same and even increases when we use less active energy. It is not a good thing especially what we need to pay is both the active energy and the reactive energy. Those differences in pattern can be seen by the plotting the power factor that tells u the ratio of the actual energy that you use over the energy that you need to pay for. In general, good utility of energy should reflect higher power factor over time. Reducing active energy is good but if the power factor drops, it means there are some problems. If you take a look at the graph, as you can see that the active energy significantly drop in July and August. However, the power factor also significantly drops too. That leads me to further look in the data for those July and August month in the next session.