For my project, I want to use the energy dataset from UCI Irvine Machine Learning to detect correlation among time, voltage imbalance, waste power and the energy usage for different types of appliances used in a household. For a quick summary, this dataset contains the measurements of electric power consumption in one household with one-minute sampling rate over a period of almost 4 years. This dataset will provide significant insights into utility consumption trend and the correlation between minutes, hours, days or months and the energy usage of different types of appliances used at home. It, in turn, will help families or companies make better use of the power they are consuming.

For the first plot (<http://www.cis.jhu.edu/~trinh/graph1.png>), I want to make a general graph to get a basic idea of how much energy that the family uses for the past 4 years. In the plot of “Total Energy Consumed by Month, Year and Type”, we can see the distinct pattern of electricity usage in the family with the most electric consumption in December and January and the least electric consumption in June or August (It may be because during the summer, the family spends more time outdoor or on vacation). Throughout the years, the family seems to be able to reduce the energy consumed by other appliances even though it’s unfortunate that the total active energy consumed by the high electric water-heater and an air-conditioner is still kept at the same significantly high level. Moreover, if you take a look more carefully at the reactive energy column, actually the reactive energy (or unusable energy) doesn’t change. That leads me to plot another graph to compare the active energy and the reactive energy so see what’s going on.

If you click on the second graph (<http://www.cis.jhu.edu/~trinh/graph2.png>), you can see that even though the active energy, which is the energy that does the actual work, increases and deceases in different months, the reactive energy, which is the energy that doesn’t do any actual work, 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.

Secondly, it’s common sense that energy usage depends on the daily schedule of each family. If I am the one who usually go home around 11 pm everyday, you should probably see that my electricity usage only starts to increase significantly around 11pm to 2 am. Therefore, it is useful to plot the cycle of the average usage of active energy in each hour block in different months for different categories (<http://www.cis.jhu.edu/~trinh/graph3.png>). Based on this graph, we can see that this family has a pretty consistent routine of schedule: different types of appliances have different peaks throughout the day. For example, they use the kitchen more around 6pm to 10pm for every month even though it does drop in July and August. One thing to take notice is that even though all other types of energy have a significant drop in July and August, it seems to me the energy used for the laundry room doesn’t seem to decrease a lot. That leads me to plot the next graph to see more clearly the actual amount of energy used for each type of appliances in the house (<http://www.cis.jhu.edu/~trinh/graph4.png>).

In this graph, the area of each circle tells you how much energy was used. As you can see, in the laundry room, the energy used was pretty uniform through out the day and especially in July and August. That actually tells us there are some appliances that consistently run in the laundry room. It may be due to the ongoing usage of electricity of a washing machine, a tumble-dryer, a refrigerator and a light that the family unintentionally left them on.

Those plots are only done to investigate the power and the energy. We can also work with voltage and the intensity to see if we can detect any imbalance or harmonics. For example, as you may know, the electricity shouldn’t experience any voltage fluctuation over time, which is harmful to any household appliances. Voltage fluctuation indicates that there is a large utility of energy on your line that is being switched in and out.  One way to take a look at that is take a look at Box Plot (<http://www.cis.jhu.edu/~trinh/graph5.png>) to see the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum. I like box plot a lot because it’s non parametric and doesn’t rely on any assumption and it can basically give you some basic information about the spread of the data. Ideally, we wanted to see a very short box with min and max values are smaller. As you can see, the voltage fluctuates more during the winder and less during the summer. It can be explained by the fact that the family actually uses more energy in the winter for heaters that are well known to cause voltage imbalance and for some reasons, much less energy in the summer.

In conclusion, I believe this analysis will help electricity companies to understand the pattern of surges in power usage and the families to monitor their own utility use according to their own schedules. In terms of economics, it will provide an energy management system that brings a potential new source of savings and profitability in the form of the monthly energy bill without any substantial investment.