Summary on Assignment 2

This part asks to combine 2 data frames with different formats. By using the package dfply. Since the goal is to find overlapping periods, the key is to understand the semantics of the package and to convert the dates into a common data type. Then functions that converts the excel format of dates into pandas datetime objects based were defined. The dates will be in the same data type for merging. Then since one of the files has time with precision to seconds while the other has time with precision to hours, the former time series was “truncated” so that data with the same dates and hours can be grouped together. The unit of consumption is W/min for the file new.app4 and these groups of observations were summed to achieve the unit of W/h. Then the consumption divided by 1000 derives the same consumption units as the other file. The group_by and summarize functions helped achieved that purpose. Then, the inner_join function was available to combine both files with overlapping time. On top of that, the assumption made with the electricity consumptions was observations that have the same hours on the same day were grouped together. For example, the minute consumption at 1:20pm belongs to the 1:00pm group on that date.

After merging the files, plots were drawn to find any existing patterns. While the appliance in the file new.app4 didn’t seem to demonstrate observable movements, some other equipment’s’ consumptions had some repetitive movements within the combined dates. The total consumption by hours and weekdays were also presented, but no observable patterns found either.