Assignment 1: Power Calendar function

In order to write the function get_hour(), first I get the start date and end data of every kind of time period.

Second, I divide the ISOs into two lists to differentiate their daylight-saving time.

Then, after creating the weekdays, hour, and NERC holiday columns, I create a peak column to judge whether each row is non-NERC holiday weekday, if it is the value will be 1 else 0.

Next, I get the sum of the peak column, which is the total number of hours of that peak type in that period.

Finally, put all the data into a dictionary and return it.

Assignment 2: Meter Data formatting

First, I load data and make the time column in new.app4 the same as the time in USA_AL_Auburn-Opelika.AP.722284_TMY3_BASE.

In this part, I use groupby() and sum()/1000 to get the electricity consumption evert hour in kw unit.

Then, I can merge 2 data frames on column Data/Time and create a new column to give total hourly consumption of electricity.

Finally, I create plots of the data and use groupby() again to get the average of the data by hour, weekday and month.

Assignment 3: EDA and forecast model

First, I load the data and delete the useless columns.

In order to do the EDA, I use .info() to get the information of this data frame and find that the number of NA values is so small that I directly use dropna() as a method of data cleaning.

After that, I use .describe() to see the statistical parameters of each columns.

Then, I draw a time series plot of RTLMP and create a corr matrix of each column and find that the corr values between RTLMP and other columns are small.

Since the hint say we should notice the timestamps, I use the ARIMA model to predict RTLMP.

I draw the acf and pacf plot to decide the range of p values and q values firstly and use the function evaluate_models() to get the best parameter and its RMSE in the range.