Guzman Assignment 1

Gu Wu UNI: gw2406

February 10, 2021

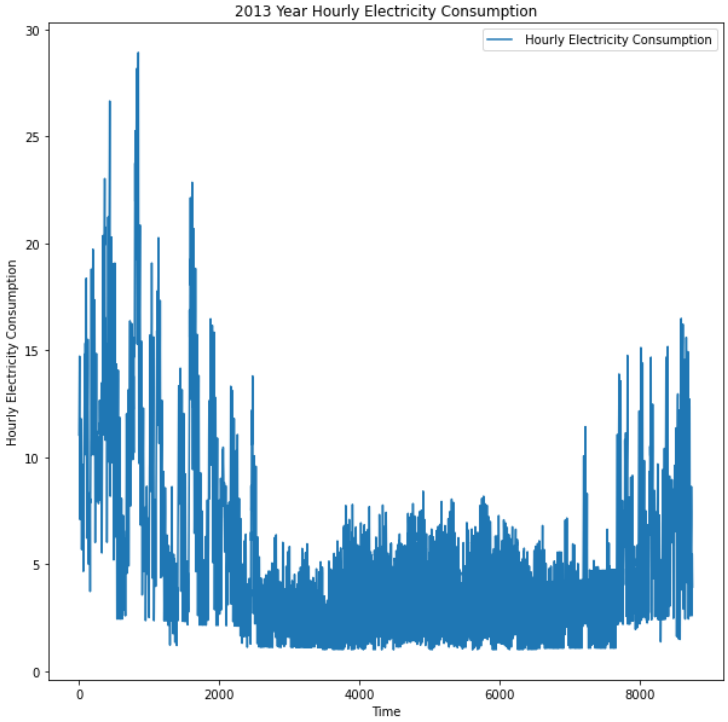
1. Meter Data formatting

```
@make_symbolic
def datetime_old(series):
    series=series.apply(lambda x:"2013/"+x[1:] if x.split()[-1].split(":")[0]!="24"
    else "2013/"+x.split()[0]+" 00:00:00")
    return series.apply(lambda x:datetime.strptime(x, '%Y/%m/%d  %H:%M:%S'))

@make_symbolic
def datetime_newapp(series):
    return series.apply(lambda x:datetime.strptime(x, '%m/%d/%Y %H:%M').replace(minute
    =0))

df_new = pd.read_csv('/Users/guwu/Desktop/nov2020_homework-main/data/Assignment 2 - new
    .app4.csv')
df_old = pd.read_csv('/Users/guwu/Desktop/nov2020_homework-main/data/Assignment 2 -
    USA_AL_Auburn-Opelika.AP.722284_TMY3_BASE.csv')

df_merged=(df_old>>rename(time="Date/Time")>> mutate(time=datetime_old(X.time))>>
    left_join((df_new>>transmute(Electricity_newapp=X.W_min/60000,
       time=datetime_newapp(X.time))>>
 group_by(X.time)>>
 summarize(Electricity_app=X.Electricity_newapp.sum())
),by='time')
          >>
          mutate(total_Electricity=X.sum(axis=1))
         )

df_merged
```

I achieved the data formatting using the code above. Essentially what I did is to add the year information 2013 to the table to make it consistent between two files, and I convert it into datetime format for future analysis. I used the package of dfply with the idea of sub-query to accomplish the data formatting.
This is part of resulting table below.

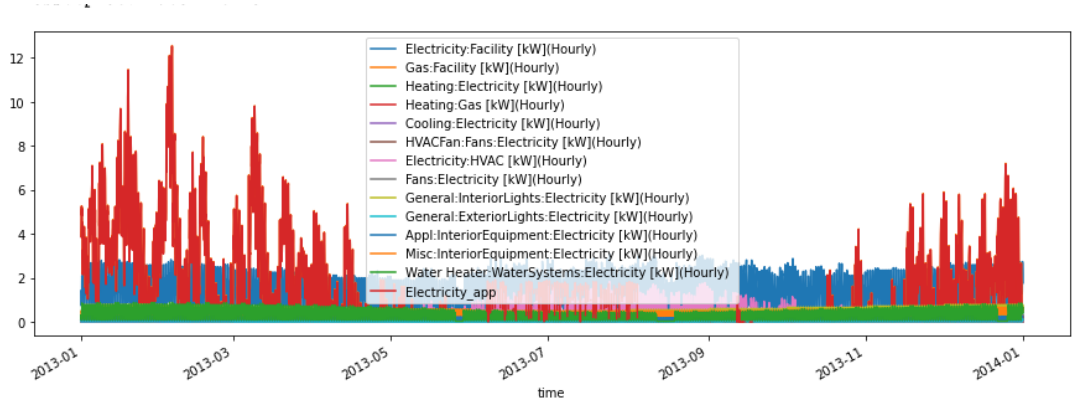
	time	Electricity:Facility [kW](Hourly)	Gas:Facility [kW] (Hourly)	Heating:Electricity [kW](Hourly)	Heating:Gas [kW](Hourly)	Cooling:Electricity [kW](Hourly)	HVACFan:Fans:Electricity [kW](Hourly)	Electricity:HVAC [kW](Hourly)	Fans:Electricity [kW](Hourly)	C
0	2013-01-01 01:00:00	0.974334	4.452977	0.0	4.425010	0.0	0.112709	0.112709	0.112709	
1	2013-01-01 02:00:00	0.796582	4.850317	0.0	4.824566	0.0	0.122617	0.122617	0.122617	
2	2013-01-01 03:00:00	0.735028	5.037645	0.0	5.012193	0.0	0.127099	0.127099	0.127099	
3	2013-01-01 04:00:00	0.727433	5.107562	0.0	5.082468	0.0	0.128391	0.128391	0.128391	
4	2013-01-01 05:00:00	0.778706	5.270878	0.0	5.246732	0.0	0.132549	0.132549	0.132549	
...	...	...	...	...	...	...	...	...	...	
8755	2013-12-31 20:00:00	2.601121	0.044507	0.0	0.000000	0.0	0.000000	0.000000	0.000000	
8756	2013-12-31 21:00:00	2.445630	0.046038	0.0	0.000000	0.0	0.000000	0.000000	0.000000	
8757	2013-12-31 22:00:00	2.206391	0.044963	0.0	0.000000	0.0	0.000000	0.000000	0.000000	
8758	2013-12-31 23:00:00	1.769166	0.295330	0.0	0.256420	0.0	0.006642	0.006642	0.006642	
8759	2013-12-31 00:00:00	1.335991	0.636988	0.0	0.603176	0.0	0.015653	0.015653	0.015653	

As I plot the hourly electricity consumption, it looks very volatile. There is a couple reasons for that, first of all, during the time in the middle of the year, people like to go out and have fun, therefore, the electricity consumption will be limited. When it is time to winter, people like to stay at home, turn their heat on and enjoy the TV, for this reason the electricity consumption will goes up.

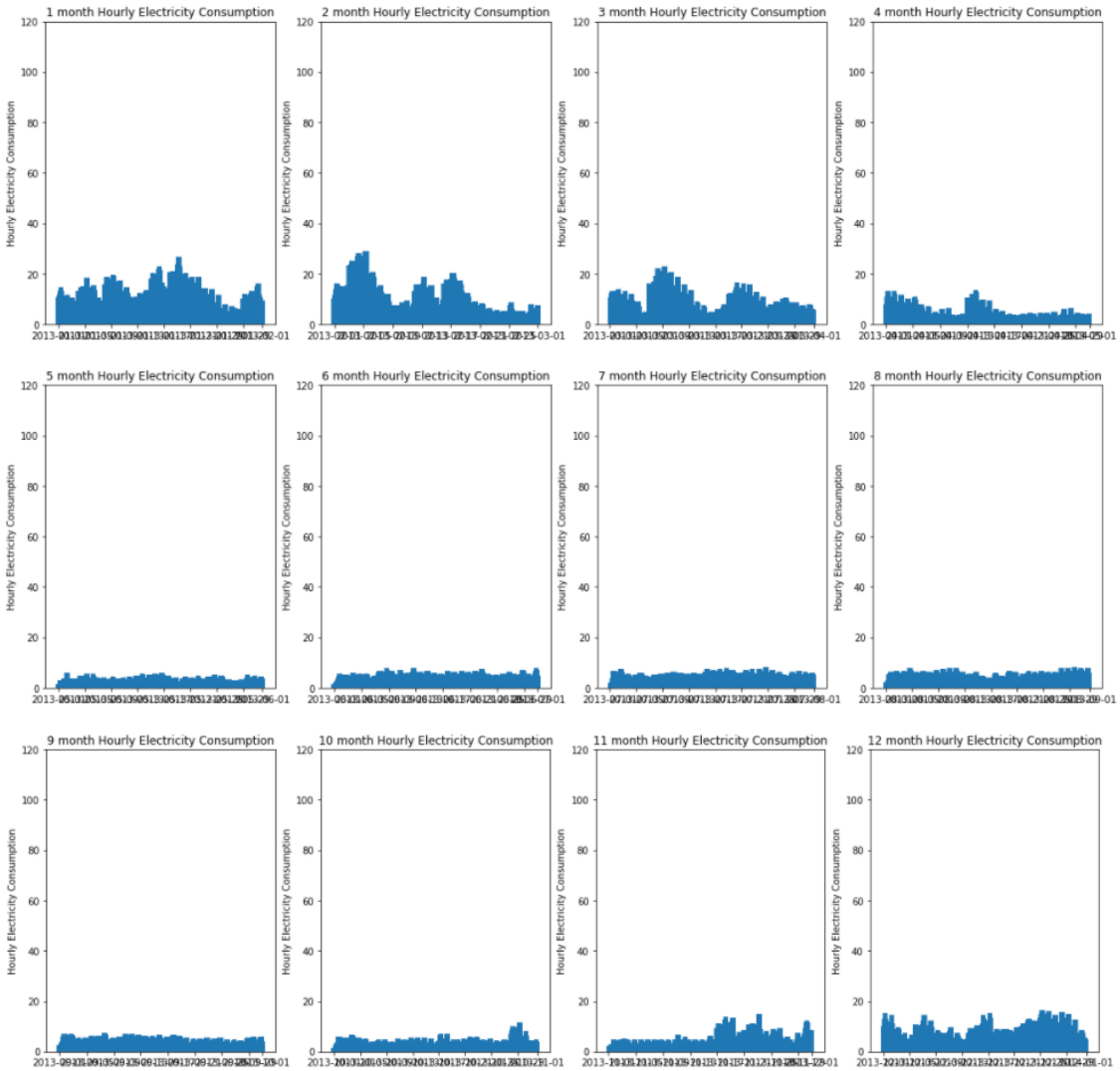


Then we have different time series for different things throughout the year, which is aligned with our expectation because we said that when it is time to winder, people like to stay at home and turn their

heat on, and from the graph below, we can clearly see that majority of the consumption actually came from the heat.



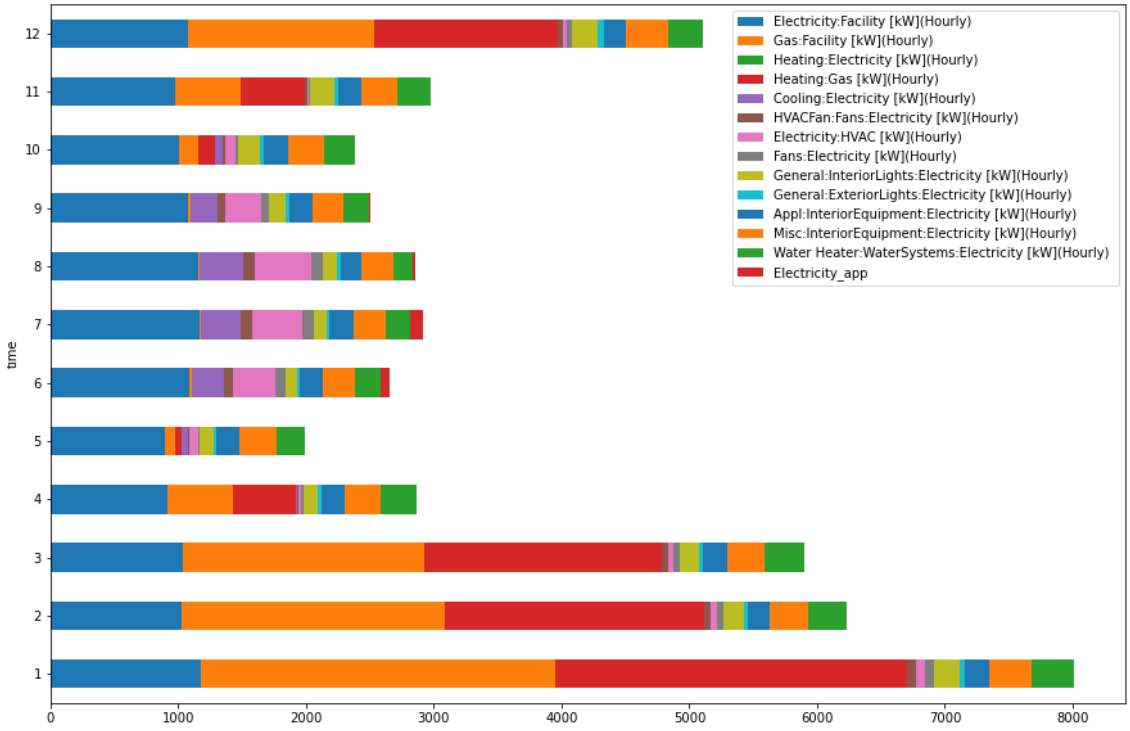
The figures below is for clarity purpose and we can see even more clear that the electricity consumption is high in the first three months and last two months



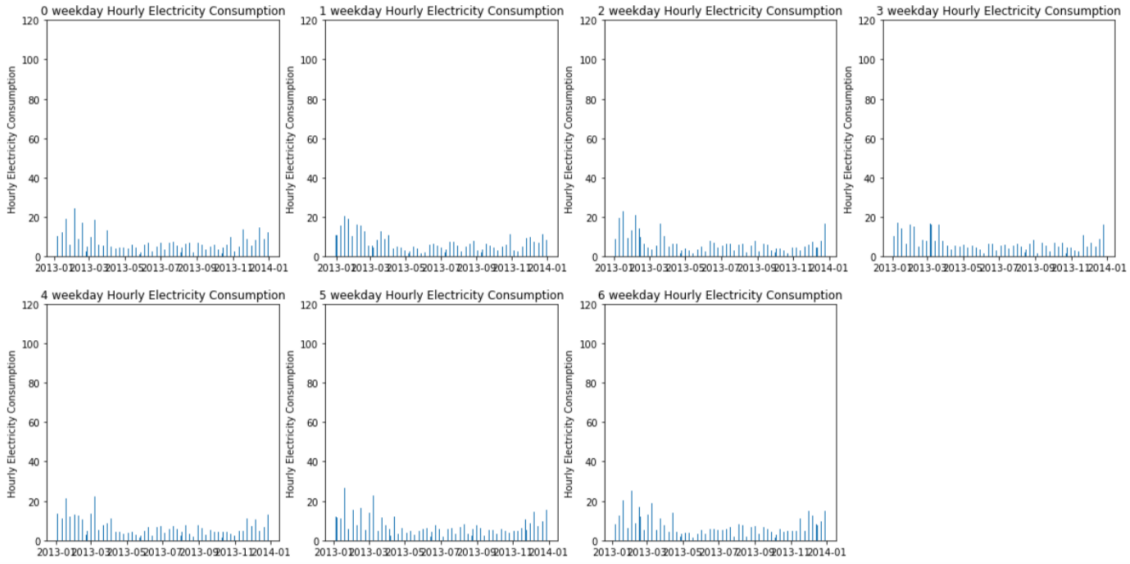
and I achieved the above graph by the code below

```
fig = plt.figure(figsize=(20,20))
for month in range(1,13):
    ax = fig.add_subplot(3,4,month)
    mask=merged.index.month==month
    ax.set(title="{} month Hourly Electricity Consumption".format(month),
        ylabel="Hourly Electricity Consumption")
    ax.set_ylim(0,120)
    ax.bar(merged.loc[mask,"total_Electricity"].index,height=merged.loc[mask,"
    total_Electricity"],label="{}".format(month))
```

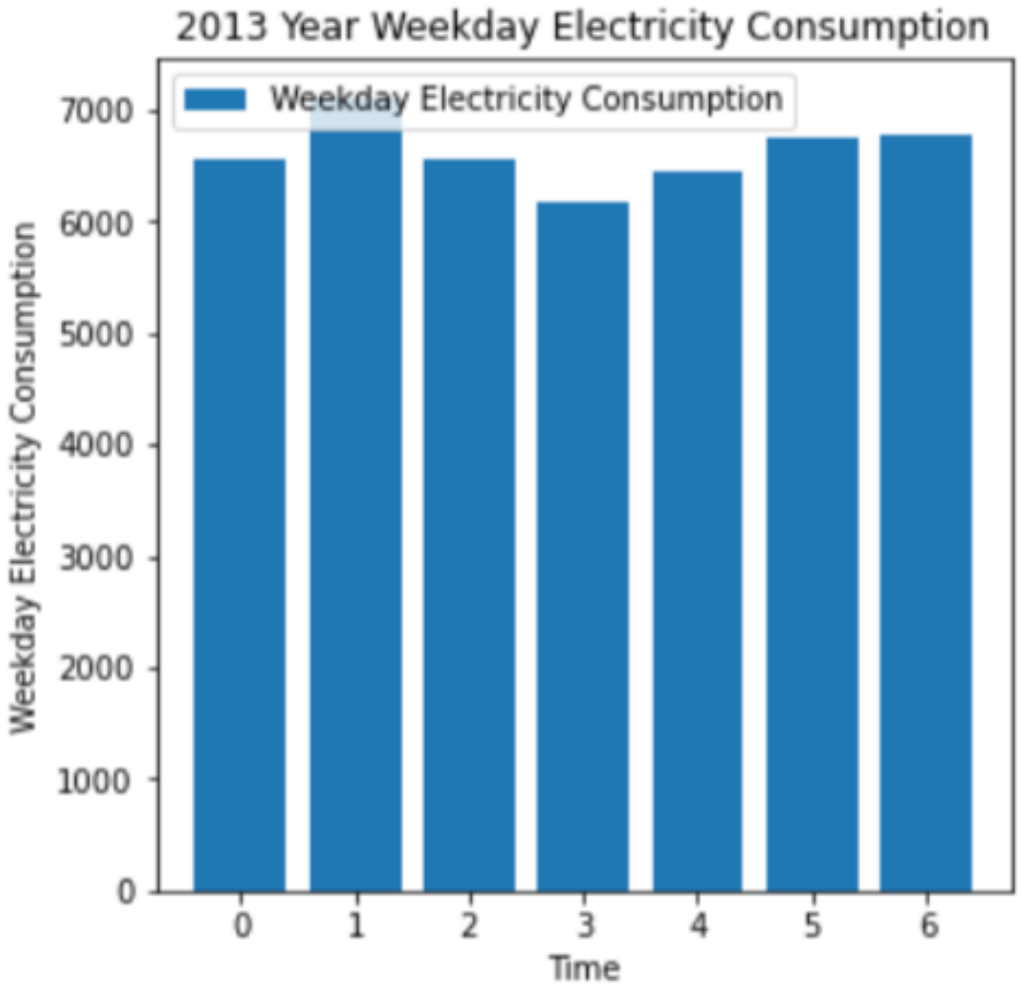
The graph below make is even more clear about our prior assumption. When it is on December or the first three months of the year. It is cold, and a lot of electricity consumption came from heat and gas. During other time, it is relatively hotter, therefore, we need AC, cooling and fans. Consequently, during this time, a lot of the electricity consumption came from heat and AC or fans.



As we drew the weekday electricity consumption, we did not see a major differences in the electricity consumption.



The graph below confirm our thought that there is no major different throughout a week



If we take at the hourly consumption distribution, we found something interesting. During morning as well as during the evening, our electricity consumption is higher than other time in the rest of the day, but this is also intuitive because, during that time, people are at home, spend time with their family or other activities. While in the rest of the day, people go our to work, or children go to school. There is no one at home, therefore, the consumption is lower. But during the pandemic last year, if we take a look at the consumption, it may turn out that the distribution will be relatively flatter than this.