Summary of Findings:

- -The data has been explored with a goal to identity features from the weather data to model the Total Energy Consumption (simply referred as energy consumption herein).
- Total Energy Consumption shows strong periodical pattern
- -On an average, the energy consumption at diurnal (with in 1 day) times scale is higher during the daytime (between 5 AM-6 PM), and close to weekend days (between Friday-Monday) at weekly timescale
- -The cycle of maximum Energy consumption repeats at weekly timescale (seems to confirm the weekend energy loads)
- -The daily variations in temperatures and relative humidity inside the rooms are closely influenced by the outside temperature
- -Both inside and outside ambient conditions have an increasing trend as it progresses from winter to the summer
- -The lack of trend in the energy consumption suggests that there are other dominant factors, other than the weather conditions
- -The daily energy consumption is associated with the Outside air temperatures and higher windspeeds, with the later contains more small-scale fluctuations at hourly timescales
- The reasonable reason for the selection of the model and features are briefly discussed in the Python notebook
- Specific humidity and Temperature Differences (instead of absolute quantities) inside the rooms found to be no better compared to features such as outside Air temperature and wind speed
- -Outside air temperature has more potential to predict energy consumption at daily time scales, whereas windspeed has more potential at sub-daily (few hours) timescales
- -Extracting the predictability of these two features at different scales (daily and hourly) has more potential to model high-energy consumption episodes, thereby reducing the footprint

Future Scope:

Though I have only explored “Gaussian Process Regression” algorithm with “Outside Temperature” as the only feature to model the daily “energy consumption”, one needs to perform sensitivity studies and validate the results from other algorithms such as Random forests. The “Gaussian Process Regression” algorithm is sensitive to the initial conditions (such as theta and nugget). I haven’t fully explored the optimum values of solution convergence.

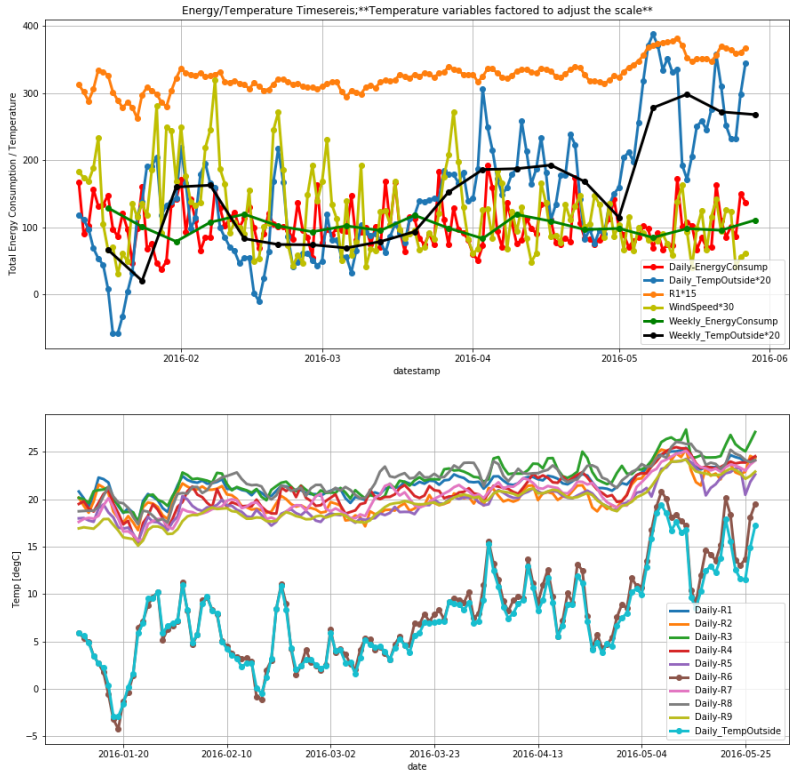


Figure 1: Timeseries of daily/weekly total energy consumption and weather data

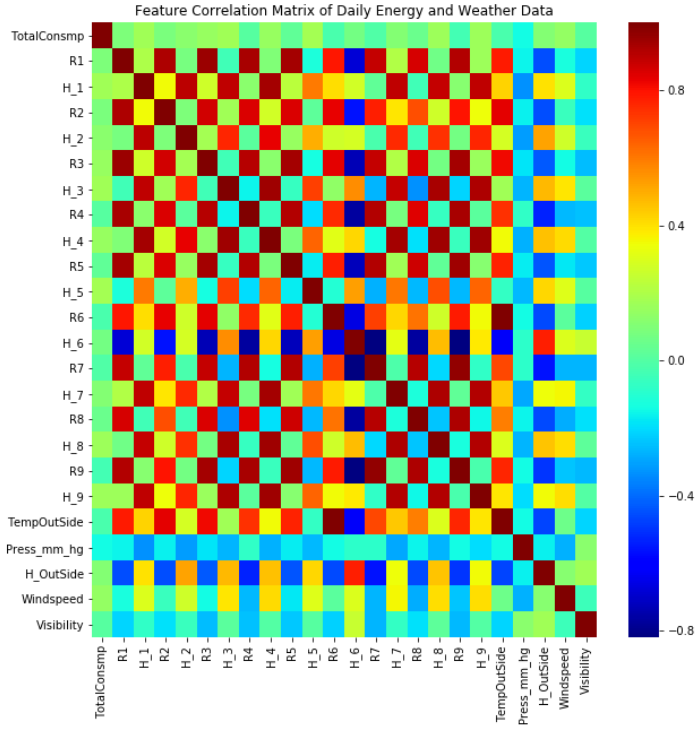
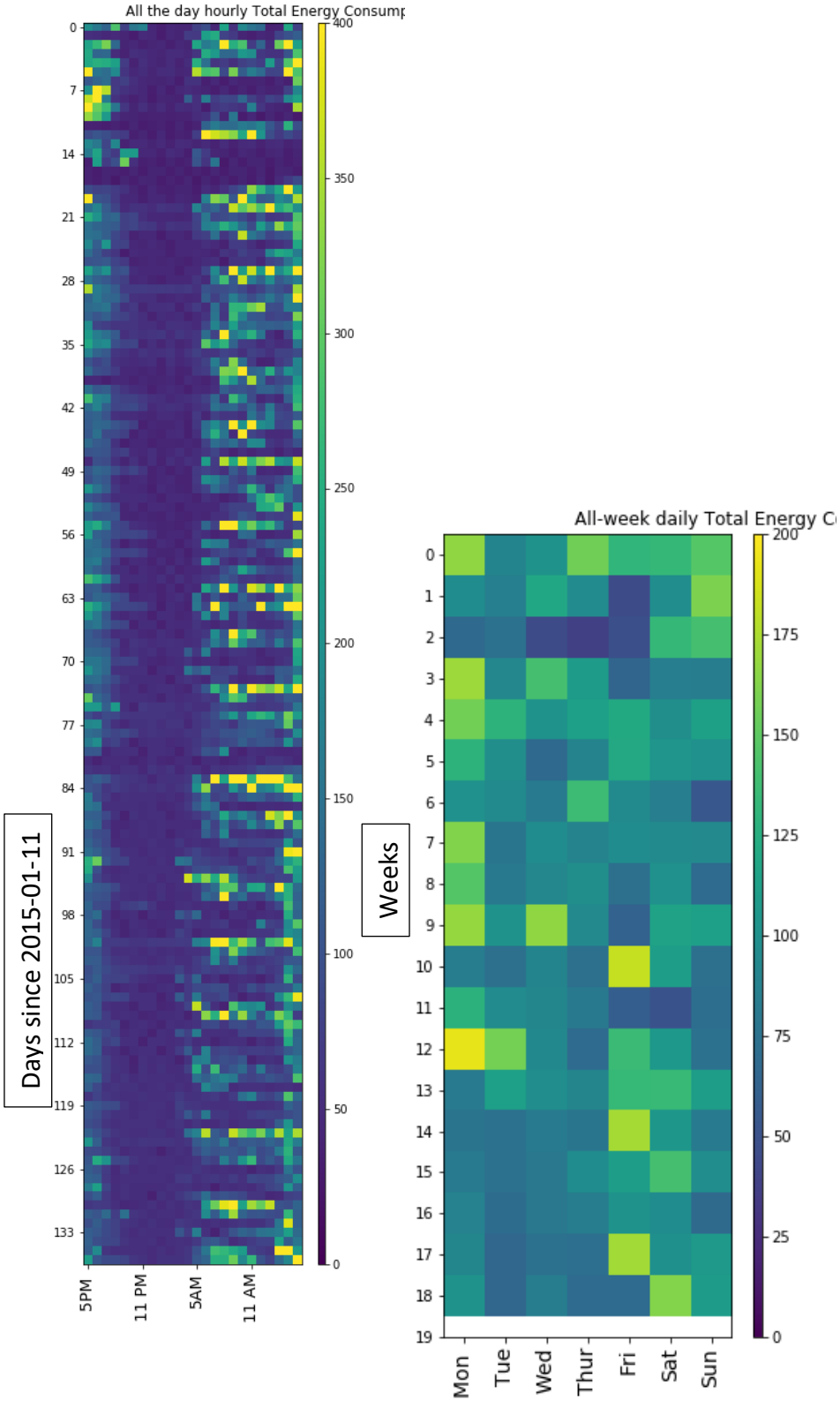


Figure 2: Feature correlation matrix at daily timescales.


All the day hourly Total Energy Consump
Days since 2015-01-11
0
7
14
21
28
35
42
49
56
63
70
77
84
91
98
105
112
119
126
133
5PM
11 PM
5AM
11 AM
400
350
300
250
200
150
100
50
0
All-week daily Total Energy C
Weeks
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Mon
Tue
Wed
Thur
Fri
Sat
Sun
200
175
150
125
100
75
50
25
0

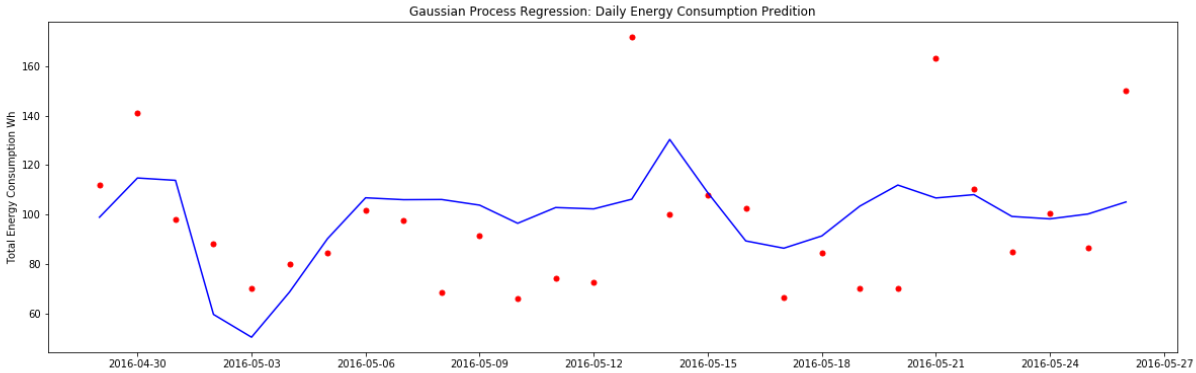


Figure 4: timeseries of predicted (red) and observed (blue) Total Energy Usage (Wh) using outside temperature as the sole feature.