Data processing for PM data

GSA project

January 5, 2016

Contents

1 Introduction

The document records the process of calculating EUI for different meter type for Jihyun

2 Process

2.1 Convert natural gas unit to kbtu

There are two units used in PM data:

```
# Gas consumption unit types
gas_unit
cf (cubic feet)    83928
therms               275

# Electric - Grid consumption unit types
elec_unit
kWh (thousand Watt-hours)    115504
```

From the Energy Star PM website document, we get the energy conversion to be:

$$1kbtu = 1.026cf$$

$$1kbtu = 100therms$$

The converted result is as follows

Meter Type	Input Unit Options	U.S. Property Assumptions[1]	
		Multiplier to get kBtu	Heat Content
Electricity (Grid Purchase and Onsite Renewable)	kBtu	1	Not Applicable
	MBtu	1,000	
	kWh	3.412	
	MWh	3,412	
	GJ	947.817	
Natural Gas	kBtu	1	1,026 Btu/cf
	MBtu	1,000	
	cf	1.026	
	ccf	102.6	
	kcf	1,026	
	Mcf	1,026,000	
	Therms	100	
	cubic meters	36.303	
	GJ	947.817	

Figure 1: Conversion table from EnergyStar Website

```
     gas_amt gas_unit  gas_amt_kbtu
153      NaN                   NaN
154      NaN                   NaN
155      NaN                   NaN
156      NaN                   NaN
157      NaN                   NaN
158      NaN                   NaN
159      NaN                   NaN
160      NaN                   NaN
161      NaN                   NaN
162      NaN                   NaN

[10 rows x 3 columns]
cf (cubic feet)
   gas_amt         gas_unit  gas_amt_kbtu
0  1890000  cf (cubic feet)       1939140
1  2440000  cf (cubic feet)       2503440
2  3213000  cf (cubic feet)       3296538
3  4430000  cf (cubic feet)       4545180
4  2656000  cf (cubic feet)       2725056
5  3323000  cf (cubic feet)       3409398
6  2508000  cf (cubic feet)       2573208
7  1554000  cf (cubic feet)       1594404
8  1062000  cf (cubic feet)       1089612
9   348000  cf (cubic feet)        357048

[10 rows x 3 columns]
therms
       gas_amt gas_unit  gas_amt_kbtu
24208       39   therms          3900
24209       14   therms          1400
24210       20   therms          2000
24211      209   therms         20900
24212       98   therms          9800
24213      571   therms         57100
24214      658   therms         65800
24215      113   therms         11300
24216       91   therms          9100
24217      100   therms         10000
```

2.2 Adding Gas and Electricity

```
     elec_amt                  elec_unit  gas_amt gas_unit  gas_amt_kbtu  \
153    523200  kWh (thousand Watt-hours)      NaN                   NaN
154     13440  kWh (thousand Watt-hours)      NaN                   NaN
155     15880  kWh (thousand Watt-hours)      NaN                   NaN
156     14760  kWh (thousand Watt-hours)      NaN                   NaN
```

```
157     16800  kWh (thousand Watt-hours)      NaN                    NaN
158     15720  kWh (thousand Watt-hours)      NaN                    NaN
159     13600  kWh (thousand Watt-hours)      NaN                    NaN
160     14920  kWh (thousand Watt-hours)      NaN                    NaN
161     12080  kWh (thousand Watt-hours)      NaN                    NaN
162     12880  kWh (thousand Watt-hours)      NaN                    NaN

     total_amt
153        NaN
154        NaN
155        NaN
156        NaN
157        NaN
158        NaN
159        NaN
160        NaN
161        NaN
162        NaN

[10 rows x 6 columns]
cf (cubic feet)
   elec_amt                   elec_unit  gas_amt         gas_unit  \
0    739200  kWh (thousand Watt-hours)  1890000  cf (cubic feet)
1    616000  kWh (thousand Watt-hours)  2440000  cf (cubic feet)
2    684400  kWh (thousand Watt-hours)  3213000  cf (cubic feet)
3    714800  kWh (thousand Watt-hours)  4430000  cf (cubic feet)
4    620400  kWh (thousand Watt-hours)  2656000  cf (cubic feet)
5    638800  kWh (thousand Watt-hours)  3323000  cf (cubic feet)
6    682400  kWh (thousand Watt-hours)  2508000  cf (cubic feet)
7    719600  kWh (thousand Watt-hours)  1554000  cf (cubic feet)
8    764800  kWh (thousand Watt-hours)  1062000  cf (cubic feet)
9    770400  kWh (thousand Watt-hours)   348000  cf (cubic feet)

   gas_amt_kbtu  total_amt
0       1939140    2678340
1       2503440    3119440
2       3296538    3980938
3       4545180    5259980
4       2725056    3345456
5       3409398    4048198
6       2573208    3255608
7       1594404    2314004
8       1089612    1854412
9        357048    1127448

[10 rows x 6 columns]
therms
       elec_amt                   elec_unit  gas_amt gas_unit  gas_amt_kbtu  \
24208    105840  kWh (thousand Watt-hours)       39   therms          3900
```

24209	107040	kWh (thousand Watt-hours)	14	therms	1400
24210	84240	kWh (thousand Watt-hours)	20	therms	2000
24211	60480	kWh (thousand Watt-hours)	209	therms	20900
24212	55200	kWh (thousand Watt-hours)	98	therms	9800
24213	58560	kWh (thousand Watt-hours)	571	therms	57100
24214	54960	kWh (thousand Watt-hours)	658	therms	65800
24215	64800	kWh (thousand Watt-hours)	113	therms	11300
24216	62400	kWh (thousand Watt-hours)	91	therms	9100
24217	121680	kWh (thousand Watt-hours)	100	therms	10000

	total_amt
24208	109740
24209	108440
24210	86240
24211	81380
24212	65000
24213	115660
24214	120760
24215	76100
24216	71500
24217	131680

3 Get monthly energy

3.1 Approach I

The first approach is to take the month portion of the "End Date" field and assign the reading to that month. From this approach, the result of some region turns out to be extremely large:

The box plot is as follows, we can see a lot of outliers with very large EUI

The problem is there are multiple records for a month. For example, for building 20600, there are two records for Aug-2012 (Aug-02-2012 and Aug-. The way the table with the desired form that there are individual columns for different meter type is generated as follows: the data of is grouped according to the value of the field "Meter Type", then the groups of energy consumption type is "joined" on "Portfolio Manager ID", "Year", and "Month". Thus if there are two Aug-2012 records

3.2 Approach II

I have considered another approach: for a date mm-dd-yyyy, if dd ¡ 15, month = mm, else month = mm + 1 This doesn't work, because there are also cases like building 20597 which has the following pattern of record time: 2010-Jul-31, 2010-Aug-31, 2010-Sep-06 In this

Table 1: Part of EUI table with Approach I in getting monthly data

Building ID	EUI	Region	Year
DC0001ZZ	12480.6184600518	11	2005
DC0001ZZ	12038.598586171	11	2010
DC0001ZZ	12037.9492157888	11	2003
DC0001ZZ	11914.3922932142	11	2007
DC0001ZZ	11889.482612042	11	2004
DC0001ZZ	11490.2126073281	11	2009
DC0001ZZ	11361.4174523634	11	2008
DC0001ZZ	11306.0507623555	11	2006
DC0001ZZ	10102.4922489585	11	2012
DC0001ZZ	10029.6403215708	11	2013
DC0001ZZ	9725.82060821062	11	2011
DC0001ZZ	9224.79333462762	11	2014
DC0001ZZ	9013.01753152796	11	2015
DC0001ZZ	3764.92990668652	11	2002
NY0234ZZ	1457.3329074403	2	2009
NY0234ZZ	1451.37965153371	2	2008

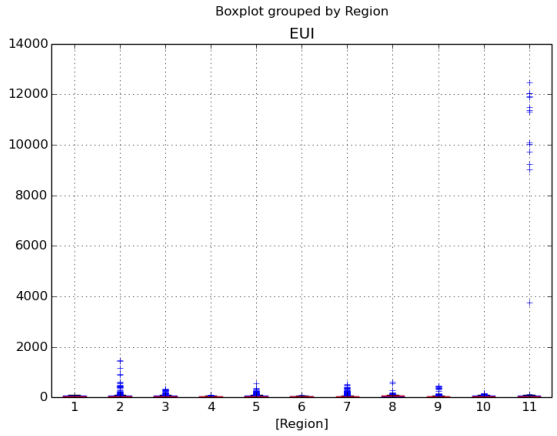

Figure 2: EUI by region, using Approach I in getting month data

case, there will still be two records of Aug. The next thought could be to calculate the time difference between adjacent rows and creating symbolic months. This will not give the right answer either considering the above example.

Table 2: Example of two records for one month

Portfolio Manager ID	Portfolio Manager ID	Meter	Meter Type	End Date	Usage/Quantity
20600	4717275		Natural Gas	2012-Aug-02 0:00:00	13752
20600	4717275		Natural Gas	2012-Aug-30 0:00:00	80123

3.3 Approach III

If there are two or more records in a month, sum up the value of the records. In Pandas, it is achieved by "resample" to the month interval and aggregate the results with "sum" method.

4 EUI graphs

4.1 By region alone

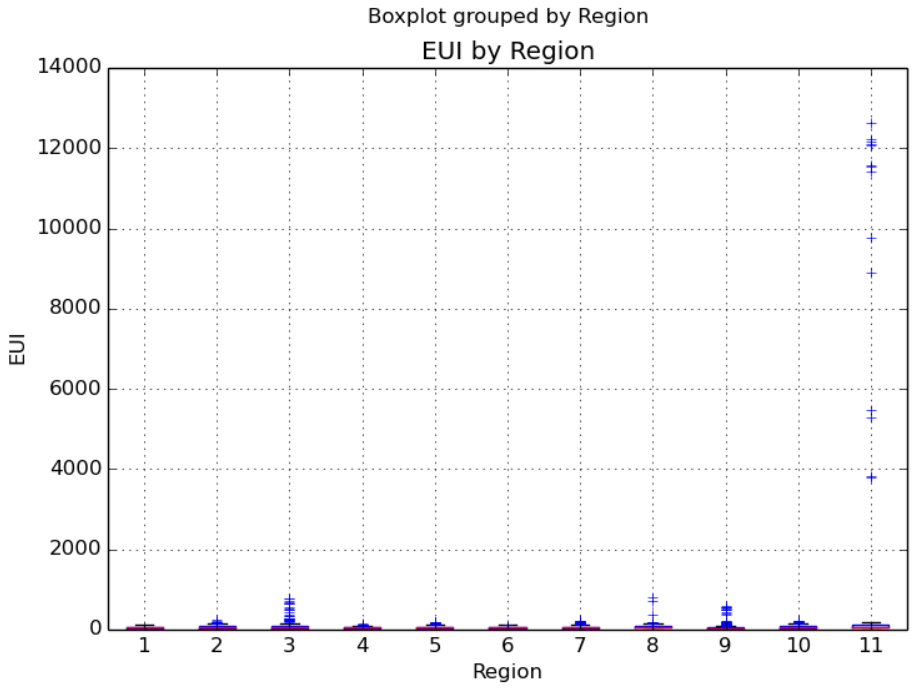

Figure 3: EUI by region

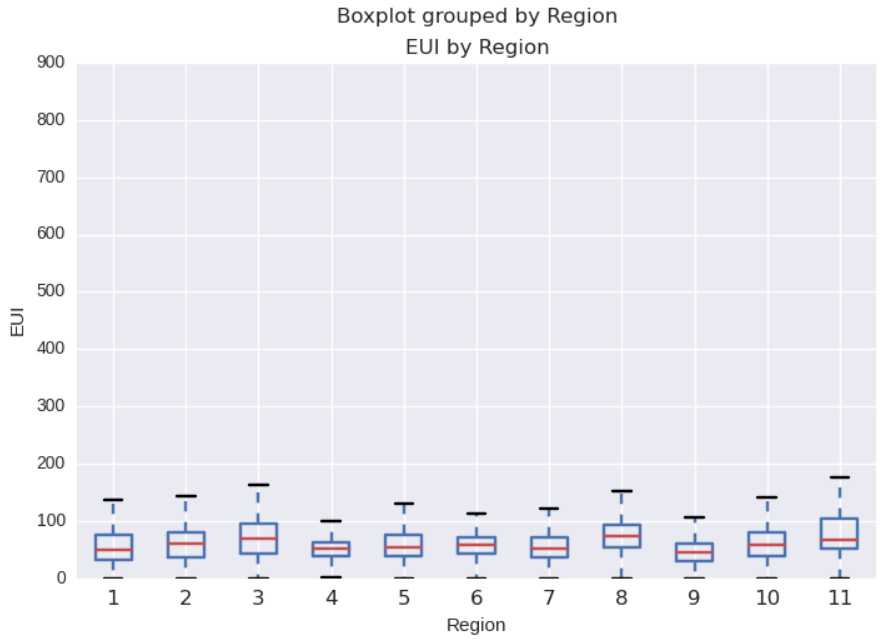

Figure 4: EUI by region, excluding a building with extremely high energy consumption, DC0001ZZ

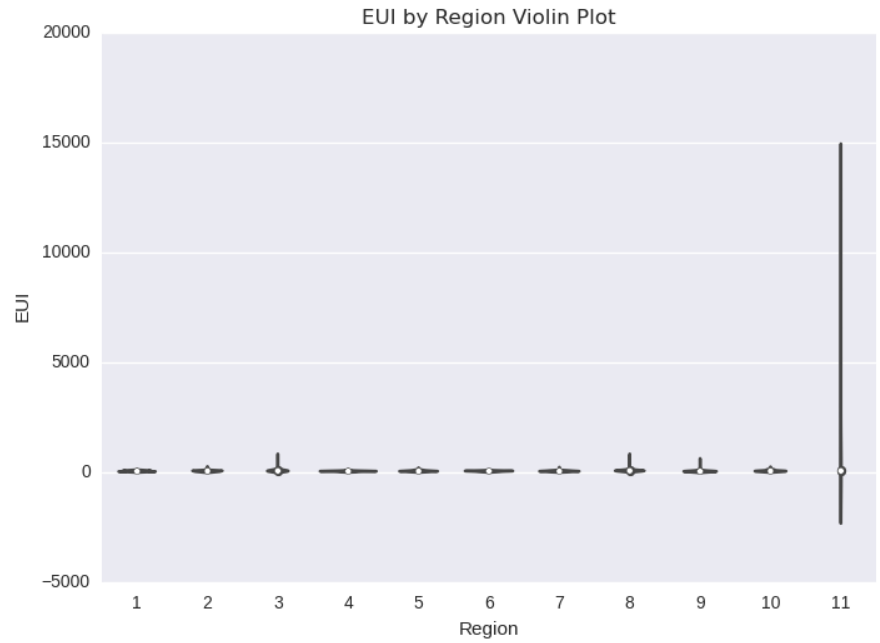

Figure 5: EUI by region violin plot

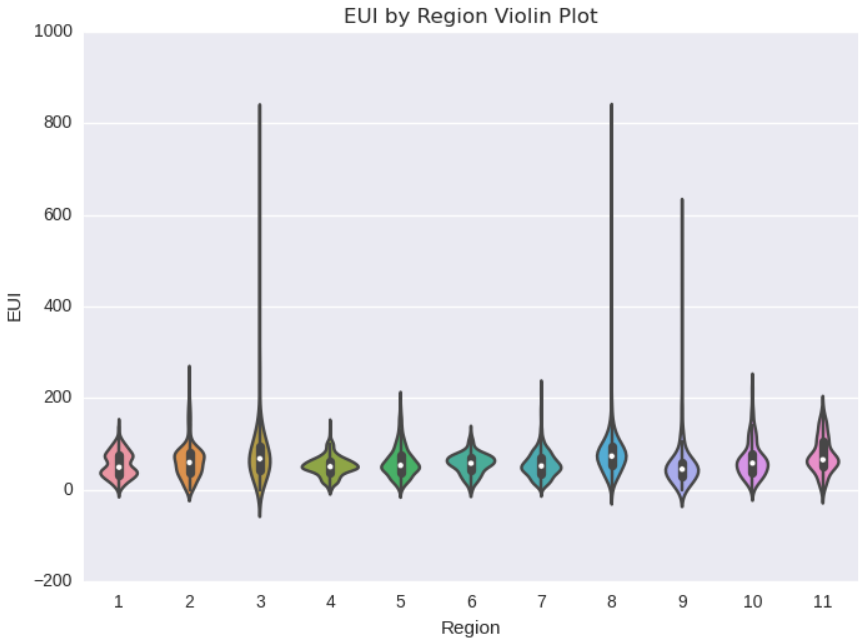

Figure 6: EUI by region violin plot, excluding a building with extremely high energy consumption, DC0001ZZ