

Excel converter for Jihyun

SEED project

December 13, 2015

Contents

1	Introduction	2
2	Input Data Analysis	3
2.1	Number of missing value for selected field	3
2.2	Number of records for each energy type	4
2.3	ranges	5
2.4	Non-negativity	5
3	Cleaned up	6
3.1	Static info from EUAS template	6

1 Introduction

The document records the implementation of the data converter from excel to csv for Jihyn. The field of the target file (EUAS) include

Table 1: Excel merging script field mapping table

Source sheet name	field name	target	Condition
Properties	State/Province	State	
Properties	Gross Floor Area	GSF	
Properties	Year Built	Year Built	
Properties	Postal Code	Postal Code (first 5 digit)	
Property Use	Self-Selected Primary Function	Use Type	
Meter Consumption Data	Property Name	Building ID	
Meter Consumption Data	End Date (month)	Month	
Meter Consumption Data	End Date (year)	Year	
Meter Consumption Data	(Usage/Quantity, Meter Type)	Elec Amt	if Meter Type = Electric - Grid
		Gas Amt	if Meter Type = Natural Gas
		Oil Amt	if Meter Type = Fuel Oil (No. 2)
		Water Amt	if Meter Type = Potable: Mixed Indoor/Outdoor
Meter Consumption Data	(Usage/Quantity, Meter Type)	Elec Cost	if Meter Type = Electric - Grid
		Gas Cost	if Meter Type = Natural Gas
		Oil Cost	if Meter Type = Fuel Oil (No. 2)
		Water Cost	if Meter Type = Potable: Mixed Indoor/Outdoor
Meter Consumption Data	Portfolio Manager ID	Portfolio Manager ID	
Meter Consumption Data	Portfolio Manager Meter ID	Portfolio Manager Meter ID	

2 Input Data Analysis

2.1 Number of missing value for selected field

```
checking numer of missing values for columns
## -----##
Property Name
non_Null      850
dtype: int64
## -----##
Portfolio Manager ID
non_Null      850
dtype: int64
## -----##
State/Province
non_Null      850
dtype: int64
## -----##
Postal Code
non_Null      850
dtype: int64
## -----##
Year Built
non_Null      850
dtype: int64
## -----##
Gross Floor Area
non_Null      850
dtype: int64
checking numer of missing values for columns
## -----##
Portfolio Manager ID
non_Null      344509
dtype: int64
## -----##
Portfolio Manager Meter ID
non_Null      344509
dtype: int64
## -----##
Meter Type
non_Null      344509
dtype: int64
## -----##
End Date
non_Null      344470
Null          39
dtype: int64
## -----##
```

```

Usage/Quantity
non_Null      344509
dtype: int64
## -----##
Usage Units
non_Null      344509
dtype: int64
## -----##
Cost ($)
non_Null      331567
Null          12942
dtype: int64

```

From the output of the script, one can see there are missing data in the following fields:

- End Date: 39 missing data.
- “Cost\$” contains 12942 missing data.

For “End Date”, we will discard these records with, for “Cost\$”, we’ll first mark the missing data with “-1”, and discard it when doing cost related analysis

After processing, there are

2.2 Number of records for each energy type

Meter Type

District Chilled Water - Absorption	153
District Chilled Water - Electric	636
District Chilled Water - Engine	49
District Chilled Water - Other	5925
District Hot Water	371
District Steam	15784
Electric - Grid	111065
Electric - Solar	1900
Electric - Wind	12
Fuel Oil (No. 2)	20534
Natural Gas	79492
Other Indoor	14
Other:	580
Other: Mixed Indoor/Outdoor	467
Potable Indoor	1177
Potable: Mixed Indoor/Outdoor	98474
Power Distribution Unit (PDU) Input Meter	16
Power Distribution Unit (PDU) Output Meter	106
Uninterruptible Power Supply (UPS) Output Meter	7754
dtype: int64	

2.3 ranges

Portfolio Manager ID	600	4428021
Portfolio Manager Meter ID	519	15550834
End Date	inf	2015-09-01 00:00:00
Usage/Quantity	-1385200.0	513798464.0
Cost (\$)	0.0	7858632.0

Note: when pandas read in date time, it converts missing data to current date by default

2.4 Non-negativity

Checking the number of negative records for each group. From the result, we can see there are 108 records in District Chilled Water and 151 records in District Hot Water with negative energy consumption records.

```
is_nn Meter Type
<0    District Chilled Water - Electric      108
      District Hot Water                    151
>=0   District Chilled Water - Absorption    153
      District Chilled Water - Electric      528
      District Chilled Water - Engine        49
      District Chilled Water - Other        5925
      District Hot Water                    220
      District Steam                       15784
      Electric - Grid                     111065
      Electric - Solar                     1900
      Electric - Wind                      12
      Fuel Oil (No. 2)                    20534
      Natural Gas                         79492
      Other Indoor                        14
      Other:                             580
      Other: Mixed Indoor/Outdoor          467
      Potable Indoor                      1177
      Potable: Mixed Indoor/Outdoor       98474
      Power Distribution Unit (PDU) Input Meter 16
      Power Distribution Unit (PDU) Output Meter 106
      Uninterruptible Power Supply (UPS) Output Meter 7754
dtype: int64
```

After removing 39 nan “End Date” and 259 negative “Usage/Quantity”, there are 344211 legal records to be further processed.

3 Cleaned up

After the initial clean up:

Checking non-negativity after initial clean

```
>=0    344211
```

```
dtype: int64
```

```
is_nn Meter Type
```

```
>=0    District Chilled Water - Absorption      153
      District Chilled Water - Electric        528
      District Chilled Water - Engine          49
      District Chilled Water - Other           5925
      District Hot Water                       220
      District Steam                           15784
      Electric - Grid                          111065
      Electric - Solar                         1900
      Electric - Wind                          12
      Fuel Oil (No. 2)                         20495
      Natural Gas                             79492
      Other Indoor                             14
      Other:                                   580
      Other: Mixed Indoor/Outdoor              467
      Potable Indoor                           1177
      Potable: Mixed Indoor/Outdoor            98474
      Power Distribution Unit (PDU) Input Meter 16
      Power Distribution Unit (PDU) Output Meter 106
      Uninterruptible Power Supply (UPS) Output Meter 7754
```

```
# Ranges of columns
```

```
      Portfolio Manager ID      600      4428021
Portfolio Manager Meter ID      519      15550834
      End Date      1998-06-30 00:00:00      2015-09-01 00:00:00
      Usage/Quantity      0.0      513798464.0
      Cost ($)      -1.0      7858632.0
```

3.1 Static info from EUAS template

```
# number of records
```

```
Building ID      11713 non-null object
```

```
Region          11713 non-null int64
```

```
# number of unique values
```

```
Building ID: 1065
```

```
Region: 11
```

Per Jiyhun's advice, I should look up the 'Region' field with building id in PM table from the EUAS template, since there are more buildings in EUAS than in PM. However, after

reading both tables, I found the EUAS template contains not enough matching building record information in the portfolio manager:

```
850 buildings in PM
1065 buildings in EUAS
120 common building records between PM and EUAS
```

Then Jiyhun pointed out the link to the definition of the GSA lookup, and GSA region map. I digitized it into a table and combined it with the data frame with the static information. State abbreviation table is retrieved from <http://www.stateabbreviations.us/>

There is a Canada state from the PM file, the following output shows the number of buildings in each state/Country in the PM file:

Canada	British Columbia	1
United States	Alabama	12
	Alaska	10
	Arizona	16
	Arkansas	9
	California	44
	Colorado	43
	Connecticut	6
	Delaware	1
	District of Columbia (D.C.)	44
	Florida	25
	Georgia	27
	Hawaii	3
	Idaho	4
	Illinois	23
	Indiana	10
	Iowa	6
	Kansas	4
	Kentucky	11
	Louisiana	14
	Maine	24
	Maryland	35
	Massachusetts	16
	Michigan	19
	Minnesota	14
	Mississippi	9
	Missouri	11
	Montana	10
	Nebraska	5
	Nevada	6
	New Hampshire	5
	New Jersey	11
	New Mexico	14
	New York	48
	North Carolina	14

North Dakota	15
Ohio	21
Oklahoma	7
Oregon	11
Pennsylvania	15
Puerto Rico	4
Rhode Island	2
South Carolina	14
South Dakota	5
Tennessee	12
Texas	80
Utah	9
Vermont	24
Virgin Islands of the U.S.	2
Virginia	17
Washington	34
West Virginia	12
Wisconsin	6
Wyoming	6

Need to drop the non-U.S. records