

EG_FY2011_Taxroll

October 19, 2017

1 Read EG FY2011 (billed in 2010) assessed 12/31/2009

Real Estate Tax Roll and write .csv

E. Quinn 10/17/2017

1.1 convert tax roll pdf to txt

pdf2txt.py -o EG_RE_Tax_Roll_assessed_12_31_2009.txt EG_RE_Tax_Roll_assessed_12_31_2009.pdf

Unlike every other year, this command failed to reproduce the line breaks in the pdf, and just produced a list of fields

A web-based pdf converter was used to recover the text in the pdf.

```
In [1]: import re
import numpy as np
import scipy as sc
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

1.2 Read the FY2011 tax roll (billed 8/2010, assessed values on 12/31/2009)

```
In [2]: with open("../EG_RE_Tax_Roll_assessed_12_31_2009.txt", "rt") as in_file:
    text = in_file.readlines()
```

```
fcsv = open("../EG_RE_Tax_Roll_assessed_12_31_2009.csv", "w")
```

```
fcsv.write('"account","bill_no","parcel","cat","state code","type","address","valuation"
```

```
tcsv = open("../EG_RE_Tax_Roll_assessed_12_31_2009_Totals.csv", "w")
```

```
tcsv.write('"account","total_tax"\n')
```

```
parcel_records_out = 0
```

```
totals_records_out = 0
```

```
records_read = len(text)
```

```
for line in text[0:records_read]:
```

```
    if (len(line)==134):
```

#only full lines

```

if (line[0:9].isdigit()):
    account = line[0:9]
    parcel=''
if (line.find("TOTALS") == 82):
    total_tax = (line[120:133].strip()).replace(',', ' ')
    outtotal = account + ',' + total_tax + '\n'
    tcsv.write(outtotal)
    totals_records_out = totals_records_out + 1
bill_no = line[35:44].strip()
cat = line[44:46].strip()
state_code = line[47:49]
mid_string = line[50:97]
words = mid_string.split()
if (len(words) > 0):
    for word in words:
        if ((len(word)==16) & ("-" in word)):
            parcel = word
            atype = mid_string[0:mid_string.find(parcel)].strip()
            address = mid_string[mid_string.find(parcel)+len(parcel)+1:].strip()
            tax = (line[120:133].strip()).replace(',', ' ')
            if (tax == ' '):
                tax = '0'
            exemption = (line[109:119].strip()).replace(',', ' ')
            if (exemption == ' '):
                exemption='0'
            valuation = (line[98:108].strip()).replace(',', ' ')
            outline = account + ',' + bill_no + ',' + parcel + ',' + cat + ',' +
            + atype + ',' + address + ',' + valuation + ',' + exemption + ','
            fcsv.write(outline)
            parcel_records_out = parcel_records_out+1

fcsv.close()
tcsv.close()

print("parcel records written: " + str(parcel_records_out))
print("totals records written: " + str(totals_records_out))

parcel records written: 10400
totals records written: 5174

```