EG_FY2016_Taxroll

October 19, 2017

1 Read EG FY2016 (billed in 2015) assessed 12/31/2014

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_2014.txt EG_RE_Tax_Roll_assessed_12_31_2014.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 FY2016 tax roll (billed 8/2015, assessed values on 12/31/2014)

```
In [2]: with open("../EG_RE_Tax_Roll_assessed_12_31_2014.txt", "rt") as in_file:
            text = in_file.readlines()

        fcsv = open("../EG_RE_Tax_Roll_assessed_12_31_2014.csv","w")
        fcsv.write('"account","parcel","cat","state code","type","address","valuation","exemptio

        tcsv = open("../EG_RE_Tax_Roll_assessed_12_31_2014_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)==133):                                    #only full lines
                if (line[0:9].isdigit()):                           #first 9 all digits - assume thi
                    account = line[0:9]                             #save account number
                    parcel=''                                       #clear old parcel ID
```

```
                if (line.find("TOTALS") == 82):
                    total_tax = (line[120:133].strip()).replace(',', '')  #read the tax amount f
                    outtotal = account + ',' + total_tax + '\n'
                    tcsv.write(outtotal)
                    totals_records_out = totals_records_out + 1
                cat = line[35:37].strip()                          #category in columns 44-46
                state_code = line[38:40]                           #state code in columns 47-49
                mid_string = line[41:97]                           #string containing parcel ID and
                words = mid_string.split()                         #split the string into words
                if (len(words) > 0):                               #try to find parcel ID if it is
                    for word in words:                             #loop through words
                        if ((len(word)==16) & (word.count("-") == 3)):   #looking for parcel: a
                            parcel = word                          #save parcel ID
                            atype = mid_string[0:mid_string.find(parcel)].strip()   #everything
                            address = mid_string[mid_string.find(parcel)+len(parcel)+1:].strip()
                            tax = (line[120:133].strip()).replace(',', '')  #read the tax amount
                            if (tax == ''):                                 #spaces indicate zer
                                tax = '0'
                            exemption = (line[109:119].strip()).replace(',', '')   #read exempti
                            if (exemption == ''):                           #spaces indic
                                exemption='0'
                            valuation = (line[98:108].strip()).replace(',', '')    #read valuati
                            outline = account + ',' + parcel + ',' + cat + ',' + state_code + ',
                            + 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: 5569
totals records written: 5375
```