

Upset Tax Sale List

Notes: This python scripts scrapes an Upset Tax Sale list in PDF form and creates a csv output file. The following column headers are created and the associated data is collected: **Sale No, PIN#, Assessed Value, Transfer Tax, Record Fee, Deed Fee, Upset Bid Price.** The PDF filename is not unique. Two formats of PIN#s where accounted for in the code.

PDF

09/20/2019 Upset Tax Sale List

SALE NO	DISTRICT	PROPERTY OWNER	PROPERTY LOCATION	PIN#	ASSESSED VALUE	TRANSFER TAX	RECORD FEE	DEED FEE	UPSET BID PRICE
1901001	CITY OF READING	CANAL STREET INC	515 CANAL ST	01-930648861606C02	\$ 335,900.00	\$ 27,031.90	\$ 82.00	\$ 25.00	\$ 47,000.00
1901003	CITY OF READING	TORRES MARISOL L & FIGUEROA ISAMACTORRES-	333 S 3RD ST	01-930694680025	\$ 29,900.00	\$ 2,398.90	\$ 82.00	\$ 25.00	\$ 2,000.00
1901005	CITY OF READING	LIBERTY FIRE CO	503 S 5TH ST	01-930643779324	\$ 28,800.00	\$ 2,318.40	\$ 82.00	\$ 25.00	\$ 3,000.00
1901006	CITY OF READING	CITIMORTGAGE INC AND/OR ENERBANK USA ITS SUCCESSORS	306 S 6TH ST	01-930635838403	\$ 25,300.00	\$ 2,086.65	\$ 82.00	\$ 25.00	\$ 5,000.00
1901007	CITY OF READING	ALCANTARA ERMES LONH	310 S 6TH ST	01-930635838309	\$ 22,000.00	\$ 1,771.00	\$ 82.00	\$ 25.00	\$ 2,000.00
1901009	CITY OF READING	HERNANDEZ MARTIN LEYVA	486 S 6TH ST	01-930635873609	\$ 6,600.00	\$ 531.30	\$ 82.00	\$ 25.00	\$ 2,000.00
1901011	CITY OF READING	LUKASEWYC BARBARA & NOWACZYNSKI DONNA	442 BINGAMAN CT	01-930611774427	\$ 56,100.00	\$ 4,516.05	\$ 82.00	\$ 25.00	\$ 7,000.00
1901012	CITY OF READING	MENSAH AKWASI J	327 BINGAMAN ST	01-930635881042	\$ 29,100.00	\$ 2,342.55	\$ 82.00	\$ 25.00	\$ 3,000.00
1901013	CITY OF READING	TINIO JUAN D	328 BINGAMAN ST	01-930635871971	\$ 27,500.00	\$ 2,213.75	\$ 82.00	\$ 25.00	\$ 5,000.00
1901014	CITY OF READING	FELICIANO LAUREANO/ASUNCION A	521 LAUREL ST	01-930643870447	\$ 29,300.00	\$ 2,358.65	\$ 82.00	\$ 25.00	\$ 4,000.00
1901015	CITY OF READING	CASTILLO LUIS E	408 MARKET AL	01-930635872819	\$ 10,000.00	\$ 805.00	\$ 82.00	\$ 25.00	\$ 2,000.00
1901017	CITY OF READING	DINH ERIC PHI/MARTINEZ DIANNE	424 PINE ST	01-930635775971	\$ 27,100.00	\$ 2,181.55	\$ 82.00	\$ 25.00	\$ 4,000.00
1901018	CITY OF READING	NO 1 SV CONSTRUCTION CORP & VICORIANO SANTIAGO	424A PINE ST	01-930635775981	\$ 30,000.00	\$ 2,415.00	\$ 82.00	\$ 25.00	\$ 4,000.00
1901019	CITY OF READING	FERNANDEZ CESAR & LUZ MARINA	420 SPRUCE ST	01-930635785424	\$ 32,400.00	\$ 2,608.20	\$ 82.00	\$ 25.00	\$ 3,000.00
1902020	CITY OF READING	MARTIN JERMAINE	622 PINE ST	01-930635876940	\$ 23,300.00	\$ 1,875.65	\$ 82.00	\$ 25.00	\$ 3,000.00
1902021	CITY OF READING	RODRIGUEZ JACK DANIEL	SPRINGS GARDEN ST	01-930636979547	\$ 100.00	\$ 8.05	\$ 82.00	\$ 25.00	\$ 1,000.00
1902022	CITY OF READING	BURFORD JERRY T & MABEL T	305 S 6TH ST	01-930635804404	\$ 30,600.00	\$ 2,469.30	\$ 82.00	\$ 25.00	\$ 3,000.00
1902024	CITY OF READING	HICKEY ROBERT F & GRAUER GEORGE F & REGINA A	417 S 6TH ST	01-930635874778	\$ 27,300.00	\$ 2,187.65	\$ 82.00	\$ 25.00	\$ 3,000.00
1902025	CITY OF READING	LOPEZ LEAFAR J	509 S 6TH ST	01-930648874287	\$ 31,700.00	\$ 2,551.85	\$ 82.00	\$ 25.00	\$ 3,000.00
1902026	CITY OF READING	REYES MARIA & SEMEDO THELMA	545 S 6TH ST	01-930643864885	\$ 7,200.00	\$ 579.60	\$ 82.00	\$ 25.00	\$ 2,000.00
1902027	CITY OF READING	REYES MARIA & SEMEDO THELMA	551 S 6TH ST	01-930643864880	\$ 500.00	\$ 40.25	\$ 82.00	\$ 25.00	\$ 1,000.00
1902028	CITY OF READING	FENELON MARIE M	304 S 7TH ST	01-930636888459	\$ 20,500.00	\$ 1,650.25	\$ 82.00	\$ 25.00	\$ 2,000.00
1902029	CITY OF READING	NORMAN ALBERT	332 S 7TH ST	01-930636888069	\$ 1,200.00	\$ 96.60	\$ 82.00	\$ 25.00	\$ 1,000.00
1902030	CITY OF READING	DIAZ ANGEL E	404 S 7TH ST	01-930636878940	\$ 16,900.00	\$ 1,360.45	\$ 82.00	\$ 25.00	\$ 2,000.00
1902031	CITY OF READING	DIAZ EMMA	406 S 7TH ST	01-930636878840	\$ 24,000.00	\$ 1,932.00	\$ 82.00	\$ 25.00	\$ 2,000.00
1902032	CITY OF READING	ROMAN CARMEN & MELENDEZ JOSE JOEY	314 S 7TH ST	01-930644878044	\$ 22,000.00	\$ 1,771.00	\$ 82.00	\$ 25.00	\$ 2,000.00
1902033	CITY OF READING	RIVERA AIDA LUZ	322 S 8TH ST	01-930636983390	\$ 12,600.00	\$ 1,014.30	\$ 82.00	\$ 25.00	\$ 2,000.00
1902034	CITY OF READING	PROTRADE SERVICES LLC	328 S 8TH ST	01-930636983390	\$ 21,800.00	\$ 1,754.90	\$ 82.00	\$ 25.00	\$ 3,000.00
1902035	CITY OF READING	TAYLOR MYRON CARL	415 S 9TH ST	01-931629030085	\$ 27,400.00	\$ 2,205.70	\$ 82.00	\$ 25.00	\$ 3,000.00
1902036	CITY OF READING	BELTRE MARIA D	429 S 9TH ST	01-931629070973	\$ 12,800.00	\$ 1,030.40	\$ 82.00	\$ 25.00	\$ 1,000.00
1902037	CITY OF READING	MIRANDA FILOGONIO GOMEZ-	444 S 9TH ST	01-930636979835	\$ 6,600.00	\$ 531.30	\$ 82.00	\$ 25.00	\$ 1,000.00
1902038	CITY OF READING	SKY LIMIT ADVERTISING INC	528 S 9TH ST	01-930636979539	\$ 6,600.00	\$ 531.30	\$ 82.00	\$ 25.00	\$ 2,000.00
1902039	CITY OF READING	RIVERA VIRGINIA E	528S 9TH ST	01-930636979538	\$ 22,100.00	\$ 1,779.05	\$ 82.00	\$ 25.00	\$ 3,000.00
1902041	CITY OF READING	MOGEL RAYMOND F	240 S 10TH ST	01-931629084668	\$ 21,900.00	\$ 1,762.95	\$ 82.00	\$ 25.00	\$ 3,000.00
1902042	CITY OF READING	VALDEZ JUAN & CANDITA DELACRUZ	428 S 10TH ST	01-931629072982	\$ 29,300.00	\$ 2,406.95	\$ 82.00	\$ 25.00	\$ 3,000.00
1902043	CITY OF READING	NINHO SREGORY T & DIVINA M	602 S 10TH ST	01-931687074981	\$ 28,700.00	\$ 2,310.35	\$ 82.00	\$ 25.00	\$ 3,000.00
1902049	CITY OF READING	GONZALEZ HAYDEE	640 BINGAMAN ST	01-930636888241	\$ 39,800.00	\$ 3,203.90	\$ 82.00	\$ 25.00	\$ 4,000.00
1902047	CITY OF READING	XP PARTNERS LLC	640 BINGAMAN ST	01-930636888283	\$ 5,800.00	\$ 468.90	\$ 82.00	\$ 25.00	\$ 3,000.00
1902047	CITY OF READING	CANDLERIA BENJAMIN JR	647 BINGAMAN ST	01-930636888334	\$ 29,300.00	\$ 2,358.65	\$ 82.00	\$ 25.00	\$ 3,000.00
1902048	CITY OF READING	CANDLERIA BENJAMIN	649 BINGAMAN ST	01-930636888354	\$ 14,500.00	\$ 1,167.25	\$ 82.00	\$ 25.00	\$ 2,000.00
1902049	CITY OF READING	BABYKAKES PROPERTIES LLC	700 BINGAMAN ST	01-930636888425	\$ 18,500.00	\$ 1,489.25	\$ 82.00	\$ 25.00	\$ 3,000.00
1902050	CITY OF READING	AGUILERA SERGIO BAEZA & RODRIGUEZ NORMA A G	804 BINGAMAN ST	01-930636888587	\$ 28,200.00	\$ 2,270.10	\$ 82.00	\$ 25.00	\$ 4,000.00
1902051	CITY OF READING	DEMBSOWSKI DOROTHY	812 BINGAMAN ST	01-930636888649	\$ 19,800.00	\$ 1,593.90	\$ 82.00	\$ 25.00	\$ 2,000.00
1902052	CITY OF READING	BURGOS EDWIN	834 BINGAMAN ST	01-930636888648	\$ 22,800.00	\$ 1,835.40	\$ 82.00	\$ 25.00	\$ 4,000.00
1902053	CITY OF READING	RIVERA CLARIOLIA	839 BINGAMAN ST	01-930628987894	\$ 26,600.00	\$ 2,144.30	\$ 82.00	\$ 25.00	\$ 4,000.00
1902054	CITY OF READING	OYDLA JOHNNY	728 CHESTNUT ST	01-930628982948	\$ 25,200.00	\$ 2,028.60	\$ 82.00	\$ 25.00	\$ 3,000.00
1902055	CITY OF READING	CABA MANUEL	804 CHESTNUT ST	01-930628985948	\$ 40,100.00	\$ 3,228.05	\$ 82.00	\$ 25.00	\$ 4,000.00
1902056	CITY OF READING	SAMUEL ANAND & KHALMA NESHA	856 CHESTNUT ST	01-931621085929	\$ 46,300.00	\$ 3,727.15	\$ 82.00	\$ 25.00	\$ 5,000.00
1902057	CITY OF READING	FELICIANO ALEXIS SERRANO	919 COTTON ST	01-931629071960	\$ 23,500.00	\$ 1,891.75	\$ 82.00	\$ 25.00	\$ 2,000.00
1902058	CITY OF READING	SEMIDEV AGRIPINA	957 COTTON ST	01-931629073829	\$ 19,300.00	\$ 1,553.65	\$ 82.00	\$ 25.00	\$ 2,000.00
1902059	CITY OF READING	RASHID KARIM	959 COTTON ST	01-931629074889	\$ 16,000.00	\$ 1,288.00	\$ 82.00	\$ 25.00	\$ 1,000.00
1902060	CITY OF READING	PROTRADE SERVICES LLC	915 CULVERT ST	01-931629081028	\$ 14,100.00	\$ 1,135.05	\$ 82.00	\$ 25.00	\$ 2,000.00
1902061	CITY OF READING	LEBRON JOSUE D NEGRON	917 CULVERT ST	01-931629081048	\$ 22,400.00	\$ 1,803.20	\$ 82.00	\$ 25.00	\$ 3,000.00
1902062	CITY OF READING	PROTRADE SERVICES LLC	923 CULVERT ST	01-931629081088	\$ 19,100.00	\$ 1,537.55	\$ 82.00	\$ 25.00	\$ 3,000.00
1902063	CITY OF READING	PHILLIPS PETER J & HAAS SONNIE K	416 GILSON AL	01-930636971828	\$ 400.00	\$ 32.20	\$ 82.00	\$ 25.00	\$ 1,000.00
1902064	CITY OF READING	CRUZ PRISCILLA	802 LAUREL ST	01-930643874354	\$ 27,100.00	\$ 2,181.55	\$ 82.00	\$ 25.00	\$ 5,000.00
1902065	CITY OF READING	TORRES HECTOR L	612 LAUREL ST	01-930643875324	\$ 36,200.00	\$ 2,914.10	\$ 82.00	\$ 25.00	\$ 5,000.00
1902066	CITY OF READING	ADAM LEON JR	615 LAUREL ST	01-930643875478	\$ 27,200.00	\$ 2,189.60	\$ 82.00	\$ 25.00	\$ 3,000.00
1902070	CITY OF READING	AGUIRRE JESUS SALVADOR REYES	826 MASTER ST	01-930628986892	\$ 28,400.00	\$ 2,286.20	\$ 82.00	\$ 25.00	\$ 3,000.00
1902071	CITY OF READING	GERMOSEN ROBERT A HENRIQUEZ-	831 MASTER ST	01-930628987921	\$ 28,400.00	\$ 2,286.20	\$ 82.00	\$ 25.00	\$ 5,000.00
1902072	CITY OF READING	GONZALEZ JOSE O	726 MCILVAIN ST	01-930636973897	\$ 16,800.00	\$ 1,352.40	\$ 82.00	\$ 25.00	\$ 1,000.00
1902073	CITY OF READING	GLO MANAGEMENT INC	729 MCILVAIN ST	01-930636973952	\$ 16,600.00	\$ 1,336.30	\$ 82.00	\$ 25.00	\$ 2,000.00
1902074	CITY OF READING	NUNEZ INVESTMENT INC	733 MCILVAIN ST	01-930636973974	\$ 23,700.00	\$ 1,907.85	\$ 82.00	\$ 25.00	\$ 2,000.00

Python Code (20190821_UpsetTaxSale.py)

```
# Read an Upset Tax Sale List and extract Sale No, PIN, Assessed Value,  
# Transfer Tax, Record Fee, Deed Fee, and Upset Bid Price  
# Accumatch  
# by Michael Keller  
# Aug 21, 2019  
# requires pdfplumber, re, datetime, os  
  
# Libraries  
import pdfplumber  
import re  
from datetime import datetime  
import os  
import sys  
  
# Start Timer  
start = datetime.now()  
  
# Booleans  
lineStop = False  
lineStart = False  
PIN_Found = False  
  
# Variables  
fn = 'Upset Tax Sale List.pdf'  
messageEOL = '. . . end of line'  
dollarSign = '$'  
count = 0  
nospace = ''  
space = ' '  
comma = ','  
outString = []  
outFilename = 'UpsetTaxSaleList'  
outputType = '.csv'  
dataLineStart = 'SALE NO'  
dataLineStop = 'Updated'
```

```

colHeadings = 'Parcel,Year,Fees,Principal,Penalty,Interest,Total Due'

_return = '\r'

SaleNo = ''

PIN = ''

AssessedValue = ''

TransferTax = ''

RecordFee = ''

DeedFee = ''

UpsetBidPrice = ''

# Regular Expressions

regPIN = '\d{2}\S{1}\d{12}[A-Z0-9]*'

# Functions

def printPDFfilename(fn):

    return 'Reading ' + fn + ' file . . .'

def printNumberOfPages(np):

    return 'Number of Pages: ' + str(np)

def openFile(fn):

    # Create output file

    try:

        file = open(fn,'w+')

        # Write out lines from array

        file.write(colHeadings + _return)

        file.close()

        print('Opening ' + fn + ' file . . .')

    except FileNotFoundError:

        sys.exit('File: ' + fn + ' does not exist')

    except PermissionError:

        sys.exit('Unable to open output file for writing. Is the file currently open?')

def writeFile(fn,os):

    # Create and append to output file

    try:

```

```

    file = open(fn,'a+')
    for taxLine in os:
        file.write(taxLine + _return)
    file.close()
    print('Saving ' + fn + ' file . . .')
except FileNotFoundError:
    sys.exit('File: ' + fn + ' does not exist')
except PermissionError:
    sys.exit('Unable to open output file for writing. Is the file currently open?')

def stripAmount(amt,zero):
    ## Strip dollar sign, commas, and blanks from value
    ## if nothing left after stripping, change value to 0.00
    ## if zero is True, all amounts that are either blank, have
    ## dashes, do not have values, etc. will be converted to '0.00'
    regAmount = "[,0-9]+[']\d{2}"
    comma = ','
    dollarSign = '$'
    space = ' '
    Zero_Value = False

    amt = amt.strip()
    amt = amt.replace(comma,nospace)
    amt = amt.replace(dollarSign,nospace)
    amt = amt.replace(space,nospace)
    if(zero == True):
        Zero_Value = re.search(regAmount,amt)
        if ((Zero_Value is None) or (len(amt) == 0)):
            amt = '0.00'
    return amt

# PDF File Open
print(printPDFfilename(fn))
pdf = pdfplumber.open(fn)

lastPage = len(pdf.pages)

```

```

print(printNumberOfPages(lastPage))

fileName = outFilename + outputType
openFile(fileName)

for page in pdf.pages:
    p1 = pdf.pages[count]
    pltext = p1.extract_text()
    text = pltext.splitlines()

## Break for Testing
## if(count == 2):
##     break

for item in text:
    blob = item.strip()

    if(blob[0:7].strip() == dataLineStop):
        lineStop = True
        lineStart = False

    if((lineStart == True) and (lineStop == False)):
        blob2 = blob.split(space,1)
        SaleNo = blob2[0].strip()
        PIN_Found = re.search(regPIN,blob2[1])
        if not (PIN_Found is None):
            PIN = blob2[1][PIN_Found.start():PIN_Found.end()]
            carp = blob2[1].split(dollarSign,5)
            AssessedValue = stripAmount(carp[1],True)
            TransferTax = stripAmount(carp[2],True)
            RecordFee = stripAmount(carp[3],True)
            DeedFee = stripAmount(carp[4],True)
            UpsetBidPrice = stripAmount(carp[5],True)

            tempString = SaleNo + comma + PIN + comma + AssessedValue + comma + TransferTax +
comma
            tempString = tempString + RecordFee + comma + DeedFee + comma + UpsetBidPrice

```

```
        outString.append(tempString)

    if(blob[0:7].strip() == dataLineStart):
        lineStop = False
        lineStart = True

    count = count + 1

## Close PDF
pdf.close()

## Write CSV file
writeFile(fileName,outString)

#end timer
print('Runtime: ' + str(datetime.now() - start))

print(messageEOL)

# EOF
```

CSV Output

	A	B	C	D	E	F	G
1	Parcel	Year	Fees	Principal	Penalty	Interest	Total Due
2	1901001	01-530643863606C02	335800	27031.9	82	25	47000
3	1901003	01-530634688025	29800	2398.9	82	25	2000
4	1901005	01-530643779324	28800	2318.4	82	25	3000
5	1901006	01-530635883403	25300	2036.65	82	25	5000
6	1901007	01-530635883309	22000	1771	82	25	2000
7	1901009	01-530635873609	6600	531.3	82	25	2000
8	1901011	01-530611774427	56100	4516.05	82	25	7000
9	1901012	01-530635881042	29100	2342.55	82	25	3000
10	1901013	01-530635871971	27500	2213.75	82	25	5000
11	1901014	01-530643870447	29300	2358.65	82	25	4000
12	1901015	01-530635872819	10000	805	82	25	2000
13	1901017	01-530635775971	27100	2181.55	82	25	4000
14	1901018	01-530635775981	30000	2415	82	25	4000
15	1901019	01-530635785424	32400	2608.2	82	25	3000
16	1902020	02-530635876940	23300	1875.65	82	25	3000
17	1902021	02-530636978547	100	8.05	82	25	1000
18	1902022	02-530635884484	30600	2463.3	82	25	3000
19	1902024	02-530635874778	27300	2197.65	82	25	3000
20	1902025	02-530643874287	31700	2551.85	82	25	3000
21	1902026	02-530643864985	7200	579.6	82	25	2000
22	1902027	02-530643864980	500	40.25	82	25	1000
23	1902028	02-530636888453	20500	1650.25	82	25	2000
24	1902029	02-530636888069	1200	96.6	82	25	1000
25	1902030	02-530636878940	16900	1360.45	82	25	2000
26	1902031	02-530636878848	24000	1932	82	25	2000
27	1902032	02-530644878044	22000	1771	82	25	2000
28	1902033	02-530636983384	12600	1014.3	82	25	2000

Michael Keller
Accumatch
20190821_UpsetTaxSale
Aug 21, 2019