

# City of Trenton – 2019 Tax Sale

**Notes:** This python scripts scrapes a City of Trenton 2019 Tax Sale PDF and creates a csv output file. The following column headers are created and the associated data is collected: **Block & Lot, Tax, Munic., Interest, Cost, Total**. The PDF filename is not unique. The Block & Lot numbers differed and I used the column number to separate the columns.

**PDF**

5/24/19		CITY OF TRENTON	NOTICE OF 2019 TAX SALE	Date of Sale: 6/26/19			PG: 1
BLOCK & LOT	OWNER NAME	PROPERTY LOCATION	TAX	MUNIC.	INTEREST	COST	TOTAL
00101 0015	SMART APPLICATIONS INNOVATI	1 W STATE ST	124438.01	5249.85	11869.26	125.00	141,682.12
00101 0016	SMART APPLICATIONS INNOVATI	6 W FRONT ST	8634.04	364.34	823.54	125.00	9,946.92
00102 0004	FREEMAN, RONALD	30 S WARREN ST	5700.76	265.89	277.41	125.00	6,369.06
00102 0008	FOWLER, THOMAS S & RITA RIC	14 S WARREN ST	2156.69	90.80	38.89	70.73	2,357.11
00102 0012	SADIP CORP	1-7 E STATE ST	7856.82	331.11	202.26	125.00	8,515.19
00102 0013	OM NAMO BHAGAVATE V REALTY	9 E STATE ST	1588.72	0.00	20.77	57.19	1,666.68
00102 0015	RAMS REAL ESTATE, LLC	13 E STATE ST	0.00	105.07	1.28	40.00	146.35
00102 0019	LEE, SANG GON & JUNG SOOK H	25 E STATE ST	0.00	155.69	1.90	40.00	197.59
00102 0022	FARNA I, LLC	35 E STATE ST	3442.23	0.00	71.74	95.28	3,609.25
00102 0030	LE, TUAN M & PHUNG, YEN H H	7 S BROAD ST	3816.90	160.71	86.47	106.28	4,170.36
00104 0011	COPACETIC CULTURAL FACILITY	126 S WARREN ST	2512.82	0.00	46.19	76.18	2,635.19
00104 0013	FS PROPERTIES 171 LLC,C/O N	130-132 S WARREN ST	1155.79	48.66	33.12	49.75	1,287.32
00106 0002	EDISON HOLDINGS NJ, LLC	1 W LAFAYETTE ST	57723.67	0.00	1587.40	125.00	59,436.07
00107 0011.01	SPORTY JOE REALTY, LLC	113 S WARREN ST	2556.51	0.00	47.39	77.08	2,680.98
00107 0012	GIAQUINTO, ROBERT H SR & LIN	115 S WARREN ST	3877.29	245.26	123.51	109.92	4,355.98
00107 0014	MORGAN CLARKE ENT, INC, C/O M	119 S WARREN ST	10668.38	239.15	862.84	125.00	11,895.37
00108 0015	614 ROUTE 130, LLC, C/O WOODF	15 W FRONT ST	50139.48	2120.94	4638.90	125.00	57,024.32
00108 0018	MILLER, SCOTT	9 W FRONT ST	5120.03	237.38	213.94	125.00	5,696.35
00108 0019	MILLER, SCOTT	HONELLS	377.31	15.88	4.80	40.00	437.99
00203 0001	ARCHITECTS HOUSING CO I, TH	211-219 E FRONT ST	245.58	0.00	2.67	40.00	288.25
00203 0015	FLEURY, REINALDINE	209 E FRONT ST	9791.24	413.17	798.51	125.00	11,127.92
00303 0009	247 E FRONT ST ASSOC LLC C/	245-247 E FRONT ST	6180.13	260.20	229.39	125.00	6,794.72
00401 0002	SABABU SUMAI JAY	5-7 PERRY ST	4367.52	237.68	149.98	120.10	4,875.28
00401 0003	SABABU SUMAI JAY	9 PERRY ST	1258.33	50.88	33.35	51.85	1,394.41
00401 0004	SABABU SUMAI JAY	11 PERRY ST	289.30	0.00	8.72	40.00	338.02
00401 0010	PE ENERGY & ELECTRIC, LLC	155 N BROAD ST	2488.53	205.68	77.47	80.43	2,852.11
00401 0014	RUBY ASSOCIATES, LLC	147 N BROAD ST	4257.68	232.66	122.66	117.26	4,730.26
00401 0030	CRAWFORD CUSTOMS LLC	26-301/2 E HANOVER ST	5758.53	242.82	309.76	125.00	6,436.11
00502 0007	SABABU, SUMAI JAY	148 N BROAD ST	2139.48	152.00	62.65	72.08	2,426.21
00502 0008	SABABU, SUMAI JAY	150 N BROAD ST	4292.19	180.72	144.28	117.34	4,734.53
00502 0010	BOYD III LAWRENCE M & NATAS	213-215 N MONTGOMERY ST	891.98	71.24	9.99	44.46	1,017.67
00504 0014	SABABU SUMAI JAY	10-12 N BROAD ST	4861.49	204.69	116.40	125.00	5,307.58
00504 0020.01	SABABU JAY SUMAI	22-24 N BROAD ST	3748.06	303.64	110.55	108.25	4,270.50
00601 0014	HANOVER ACADEMY REDEVELOPME	108 N MONTGOMERY ST	1408.67	0.00	27.08	53.72	1,489.47
00601 0020	MAZARIEGOS, DAVID & ELDA	120 N MONTGOMERY ST	842.23	0.00	10.29	42.05	894.57
00602 0007	TORRES, CARLOS J & RODRIGUE	47 N STOCKTON ST	386.58	0.00	7.43	40.00	434.01
00602 0008	TORRES, CARLOS J & RODRIGUE	45 N STOCKTON ST	1816.44	0.00	39.75	62.12	1,918.31
00603 0005	MICHAELS, DAVID	36 COMMERCE ST	1680.30	0.00	153.78	61.68	1,895.76
00603 0006	HOLLYWOOD, DORA	34 COMMERCE ST	1918.08	0.00	130.80	65.98	2,114.86
00603 0007	THOMPSON, ETHEL	32 COMMERCE ST	1479.29	0.00	44.66	55.48	1,579.43
00603 0025	MAP TRENTON SF HOLDINGS, LL	307 ACADEMY ST	198.58	0.00	2.43	40.00	241.01

## Python Code (20190821\_CityOfTrenton.py)

```
# Read a City of Trenton 2019 Tax Sale and extract
# Block & Lot, Tax, Munci., Interest, Cost, Total
# 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
Data_Found = False

# Variables
fn = 'TAX SALE LIST 052419 FOR CITY WEBSITE.pdf'
messageEOL = '. . . end of line'
dollarSign = '$'
count = 0
nospace = ''
space = ' '
comma = ','
outString = []
outFilename = 'CityOfTrenton'
outputType = '.csv'
dataLineStart = 'BLOCK & LOT'
dataLineStop = 'CITY OF TRENTON'
```

```

dataLineFinal = 'GRAND TOTAL'
colHeadings = 'Block & Lot,Tax,Munic.,Interest,Cost,Total'
_return = '\r'
BlockLot = ''
Tax = ''
Munic = ''
Interest = ''
Cost = ''
Total = ''

# Regular Expressions
regAmount = "[,0-9]+[']\d{2}"
spaces = '\s{2,}'

# 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 + outputFile
openFile(fileName)

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

## Break for Testing
## if(count == 4):
##     break

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

    if((blob.find(dataLineStop) != -1) or (blob.find(dataLineFinal) != -1)):
        lineStop = True
        lineStart = False

    if((lineStart == True) and (lineStop == False)):
        BlockLot = blob[0:34].strip()
        carp2 = blob[35:len(blob)].strip()
        Data_Found = re.search(regAmount, carp2)
        if not (Data_Found is None):
            restOfLine = carp2[Data_Found.start():len(carp2)]
            restOfLine = re.sub(spaces, space, restOfLine)
            carp = restOfLine.split(space, 4)

            Tax = stripAmount(carp[0], True)
            Munic = stripAmount(carp[1], True)
            Interest = stripAmount(carp[2], True)
            Cost = stripAmount(carp[3], True)
            Total = stripAmount(carp[4], True)

```

```
tempString = BlockLot + comma + Tax + comma + Munic + comma
tempString = tempString + Interest + comma + Cost + comma + Total
outString.append(tempString)

if(blob[0:12].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
1	Block & Lot	Tax	Munic.	Interest	Cost	Total
2	00101 0015	124438.01	5249.85	11869.26	125	141682.12
3	00101 0016	8634.04	364.34	823.54	125	9946.92
4	00102 0004	5700.76	265.89	277.41	125	6369.06
5	00102 0008	2156.69	90.8	38.89	70.73	2357.11
6	00102 0012	7856.82	331.11	202.26	125	8515.19
7	00102 0013	1588.72	0	20.77	57.19	1666.68
8	00102 0015	0	105.07	1.28	40	146.35
9	00102 0019	0	155.69	1.9	40	197.59
10	00102 0022	3442.23	0	71.74	95.28	3609.25
11	00102 0030	3816.9	160.71	86.47	106.28	4170.36
12	00104 0011	2512.82	0	46.19	76.18	2635.19
13	00104 0013	1155.79	48.66	33.12	49.75	1287.32
14	00106 0002	57723.67	0	1587.4	125	59436.07
15	00107 0011.01	2556.51	0	47.39	77.08	2680.98
16	00107 0012	3877.29	245.26	123.51	109.92	4355.98
17	00107 0014	10668.38	239.15	862.84	125	11895.37
18	00108 0015	50139.48	2120.94	4638.9	125	57024.32
19	00108 0018	5120.03	237.38	213.94	125	5696.35
20	00108 0019	377.31	15.88	4.8	40	437.99
21	00203 0001	245.58	0	2.67	40	288.25
22	00203 0015	9791.24	413.17	798.51	125	11127.92
23	00303 0009	6180.13	260.2	229.39	125	6794.72
24	00401 0002	4367.52	237.68	149.98	120.1	4875.28
25	00401 0003	1258.33	50.88	33.35	51.85	1394.41
26	00401 0004	289.3	0	8.72	40	338.02
27	00401 0010	2488.53	205.68	77.47	80.43	2852.11
28	00401 0014	4257.68	232.66	122.66	117.26	4730.26

Michael Keller  
Accumatch  
20190821\_CityOfTrenton  
Aug 21, 2019