

Michigan Foreclosure Auction Sales

Notes: This python scripts scrapes a folder of PDFs from a Michigan Foreclosure Auction Sales and creates a csv output file for each PDF. The following column headers are created and the associated data is collected: **Parcel and Purchase Price**. The PDF filenames are not unique and each CSV file is named based on the current number in a series. Multiple counties were listed in the PDFs and each county had a separate regular expression to search for the parcel number.

PDF

2015 FORECLOSURE AUCTION SALES

COUNTY	MUNICIPALITY	PARCEL NUMBER	START OF LEGAL DESCRIPTION	SALE DATE	PURCHASE PRICE
BRANCH	Township of SHERWOOD	010-028-100-015-00	BEG AT C/L OF HWY IN NW 1/4 AT NW COR OF LDSOWNED BY V	8/13/2015	\$7,500.00
BRANCH	Township of SHERWOOD	010-028-100-060-03	COM AT INTER OF ARNEY RD & W LI VOGEL LD IN L211P387 TH 3	8/13/2015	\$450.00
BRANCH	Township of SHERWOOD	010-C30-000-009-00	LOT 9 CHEROKEE LAKE ESTATES SEC 12 T55 R8W L653PG698	8/13/2015	\$7,000.00
BRANCH	Township of SHERWOOD	010-C30-000-183-00	LOT 183 CHEROKEE LAKE ESTATES SEC 12 T55 R8W L555PG916	9/10/2015	\$200.00
BRANCH	Village of UNION CITY	021-000-017-022-00	UNION TOWNSHIP T55 R7W SECTION E 4/5 OF LOTS 19 & 20 EXC	8/13/2015	\$8,500.00
BRANCH	Village of UNION CITY	021-H05-013-021-00	LOT 8 BLK 13 HAMMOND ADD VILLAGE OF UNION CITY	9/10/2015	\$2,500.00
BRANCH	Village of UNION CITY	021-H05-014-006-00	LOT 8 BLK 14 HAMMOND ADD VILLAGE OF UNION CITY	9/10/2015	\$3,700.00
BRANCH	Township of MATTESON	050-A10-000-020-00	LOT 20 OF ASSESSOR'S PLAT OF SALEK FARM SEC 23T6S R8W	9/10/2015	\$14,000.00
BRANCH	Township of BATAVIA	060-030-400-010-01	PAR IN W 1/2 OF SE 1/4 COM 405 FT W FR SE COR THN 225 FT T	9/10/2015	\$2,500.00
BRANCH	Township of BATAVIA	060-030-400-010-02	COM 705 FT W OF THE SE COR OF THE W1/2 SE1/4 TH N225 FT T	8/13/2015	\$11,300.00
BRANCH	Township of BRONSON	090-016-400-045-01	BEG N 89DEG 49MIN W 990.02 FT FR E 1/4 POST TH N89DEG 49	8/13/2015	\$1,550.00
BRANCH	Township of BRONSON	090-028-200-005-03	BEG N 89DEG 21MIN 07SEC W 1961.79 FT FR E 1/4POST TH N 00	8/13/2015	\$500.00
BRANCH	Township of BRONSON	090-028-200-005-04	COM AT E 1/4 POST TH N 89DEG 21MIN 07SEC W1961.79 FT TH	8/13/2015	\$450.00
BRANCH	Township of BETHEL	100-561-000-012-00	LOT 12 SPRUNG LAKE PLAT NO 2 SEC 4 T7S R7W	9/10/2015	\$1,800.00
BRANCH	Township of OVID	110-F20-000-227-00	FIESTA SHORES LOT 227 SEC 19-30 T7S R6W	8/13/2015	\$700.00
BRANCH	Township of OVID	110-F20-000-228-00	FIESTA SHORES LOT 228 SEC 19-30 T7S R6W	8/13/2015	\$700.00
BRANCH	Township of ALGANSEE	120-K10-000-012-00	LOT 12 KOOLAMOOLAH SEC 7 T7S R5W	8/13/2015	\$33,000.00
BRANCH	Township of CALIFORNIA	160-R10-001-035-00	BEG AT A PT 29 RDS N OF SW COR OF NE 1/4 OF SEC22 TH E 11 F	8/13/2015	\$12,000.00
BRANCH	City of	200-003-000-172-00	BEG ON S LI OF CHICAGO RD 164.85 FT WLY FR N & S1/4 LI OF 58	8/13/2015	\$1,000.00
BRANCH	City of COLDWATER	301-000-000-395-00	COM 59 FT ELY FR NW COR LOT 13 BRADLEY CRIPPENSADDN TH	9/10/2015	\$25,000.00
BRANCH	City of COLDWATER	301-000-000-530-00	COM ON N LI SHERMAN ST 487 FT W OF PRESTON ST TH NLY AT	8/13/2015	\$20,500.00
BRANCH	City of COLDWATER	302-000-000-322-00	LOT 15 BLK 12 F V SMITH ADDN ALSO COM AT NE CORLOT 15 BL	9/10/2015	\$500.00
BRANCH	City of COLDWATER	302-000-000-418-00	S 3 RDS IN WIDTH OF N 7 RDS IN WIDTH LOT 27 HULLSADDN EX	8/13/2015	\$2,000.00
BRANCH	City of COLDWATER	302-000-000-975-29	UNIT 29 EAST POINTE SUBD SEC 15 T6S R6W	9/10/2015	\$100.00
BRANCH	City of COLDWATER	302-000-000-975-30	UNIT 30 EAST POINTE SUBD SEC 15 T6S 6W	9/10/2015	\$100.00
BRANCH	City of COLDWATER	303-000-000-702-00	E 1/2 LOT 17 94/100 RD IN WI OFF W SIDE OF LOT 18MARTHA A	9/10/2015	\$100.00
BRANCH	City of COLDWATER	303-000-000-824-00	E 1/2 OF W 1/2 OF LOTS 289 & 290 O P	8/13/2015	\$16,500.00
CLINTON	Township of BATH	010-380-000-195-00	THE E 202.2 FT OF THE S 66 FT OF OUTLOT G, SUPERVISORS PLAT	8/13/2015	\$8,000.00
CLINTON	Township of BINGHAM	030-027-300-020-51	PART OF SW 1/4 OF SEC 27 T7N R2W; BEG AT THE W 1/4 COR O	9/10/2015	\$41,000.00
CLINTON	Township of BINGHAM	030-027-300-021-00	PART OF SW 1/4 OF SEC 27 T7N R2W; BEG AT THE W 1/4 COR, T	9/10/2015	\$100.00
CLINTON	Township of DEWITT	050-340-000-082-00	LOT 82 NORTHDALE FARMS SEC 33 T5NR2W	8/13/2015	\$6,500.00
CLINTON	Township of ESSEX	080-150-000-025-00	ORCHARD ESTATES SITE CONDOMINIUM, UNIT #25, SEC 8, T8N R	9/10/2015	\$300.00
CLINTON	Township of ESSEX	080-150-000-036-00	ORCHARD ESTATES SITE CONDOMINIUM, UNIT #36, SEC 8, T8N R	9/10/2015	\$400.00
CLINTON	Township of ESSEX	080-150-000-037-00	ORCHARD ESTATES SITE CONDOMINIUM, UNIT #37, SEC 8, T8N R	9/10/2015	\$300.00
CLINTON	Village of OVID	121-000-035-030-00	E 1/2 OF THE E 1/2 OF BLK F, EXC W 50 FT THEREOF, ALSO EXC T	8/13/2015	\$10,000.00
CLINTON	Township of VICTOR	140-001-400-045-00	COM 752.04 FT N OF THE SE COR SEC 1 T6N R1W, TH W 292.50	8/13/2015	\$3,100.00
CLINTON	Township of VICTOR	140-011-100-010-00	COM 345 FT S OF NE COR SEC 11 T6N R1W, TH W 1320 FT, S 345	8/13/2015	\$91,000.00
CLINTON	Township of VICTOR	140-017-200-015-00	BEG AT SW COR OF E 1/2 OF NW 1/4 OF SEC 17 T6N R1W, TH N.	8/13/2015	\$10,200.00
CLINTON	Township of VICTOR	140-024-400-110-00	BEG AT INTSN MERIDIAN & ROUND LK RDS SEC 24, T6N R1W, TH	9/10/2015	\$20,000.00
CLINTON	Township of WATERTOWN	150-018-100-015-00	COM 13 RODS W OF NE CORNER, SEC 18 T5N R3W, S 10 RODS, W	8/13/2015	\$17,900.00
CLINTON	City of ST. JOHNS	300-000-029-013-00	T7N R2W, CITY OF ST JOHNS, ORIGINAL PLAT BLK 29, LOT 14.	9/10/2015	\$300.00
CLINTON	City of ST. JOHNS	300-000-032-010-00	T7N R2W, CITY OF ST JOHNS, ORIGINAL PLAT BLK 32, LOT 7 EXC.	8/13/2015	\$26,000.00
CLINTON	City of ST. JOHNS	300-000-065-001-00	T7N R2W, CITY OF ST JOHNS, ORIGINAL PLAT BLK 65, E 56 FT OF	9/10/2015	\$100.00
CLINTON	City of ST. JOHNS	300-000-088-007-00	T7N R2W, CITY OF ST JOHNS, ORIGINAL PLAT, BLOCK 88, LOT 9 A	8/13/2015	\$14,500.00
CLINTON	City of ST. JOHNS	300-000-090-009-00	T7N R2W, CITY OF ST JOHNS, ORIGINAL PLAT BLK 90, LOT 10 & S	8/13/2015	\$9,200.00
CLINTON	City of ST. JOHNS	300-008-100-002-10	COM AT NE COR OF SEC 8, T7N R2W, TH S 01D 07' 17" W 1317 F	9/10/2015	\$300.00
DICKINSON	Township of BREITUNG	002-673-007-00	MAP #-2387C. LOTS 7-8 BLK 13. MULLEN'S 1ST ADD TO THE VILL	9/9/2015	\$16,100.00
DICKINSON	Township of FELCH	003-055-003-00	FEL-5 662A SEC 5 T43N R28W PRT OF THE SW 1/4 OF SW 1/4 BE	9/9/2015	\$100.00

Python Code (20190808_MichiganExtra.py)

```
# Read a list of PDF files in folder, open, scan, and dump data individual text files
# Accumatch
# by Michael Keller
# Aug 7, 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
Parcel_Found = False

# Variables
TownshipFlag = 'Parcel Count'
FileFlag = 'file:'
space = ' '
nospace = ''
comma = ','
_return = '\r'
count = 0
fileCount = 0
outString = []
outFilename = 'MichiganAuction_'
outputType = '.csv'
fileName = ''
parcel = ''
amount = ''
current_county = 'NotFound'
```

```

expression = ''
dollarSign = '$'
restOfLine = ''
PurchasePrice = ''
colHeadings = 'Parcel,Purchase Price'

# Regular Expression County Dictionary
counties = {
  'ALGER': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'ARENAC': '\d{3}\D{1}\d{1}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'BARAGA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'BARRY': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'BAY': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'BENZIE': '\d{12}',
  'BRANCH': '\d{3}\D{1}\S{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CALHOUN': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CASS': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CHARLEVOIX': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CHIPPEWA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CLARE': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CLINTON': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'CRAWFORD': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'DELTA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}',
  'DICKINSON': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{0,4}',
  'EATON': '\d{2}.{1}\d{2,3}.{1}\d{2,3}.{1}\d{2,3}.{1}\d{3}.{1}\d{2,3}',
  'EMMET': '\d{2}\D{1}\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
  'GOGEBIC': '\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
  'GRAND': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'HOUGHTON': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'IOSCO': '\d{3}\D{1}.{3}\D\d{3}\D\d{3}\D\d{2}',
  'INGHAM': '\d{2}\D{1}\d{2}\D{1}\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
  'IRON': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{0,4}',
  'LEELANAU': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
  'KALAMAZOO': '\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
  'KALKASKA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{0,4}',
  'KEWEENAW': '\d{3}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',

```

```
'LIVINGSTON': '\d{4}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
'LUCE': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{0,4}',
'MANISTEE': '\d{12}',
'MARQUETTE': '\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'MASON': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'MECOSTA': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}',
'MISSAUKEE': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'MONROE': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'MONTMORENCY': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'MUSKEGON': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{4}\D{1}\d{2}',
'OCEANA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'OGEMAW': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'ONTONAGON': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'OSCEOLA': '\d{12}',
'OSCODA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'OTSEGO': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'OTTAWA': '\d{2}\D{1}\d{2}\D{1}\d{2}\D{1}\d{3}\D{1}\d{3}',
'SAINT CLAIR': '\d{14}',
'SAINT JOSEPH': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'SAGINAW': '\d{1}\D{1}\d{2}\D{1}\d{1}\D{1}\d{5}\D{1}\d{3}\D{1}\d{2}',
'SCHOOLCRAFT': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
'SHIAWASSEE': '\d{3}\D{1}\d{2,3}\D{1}\d{3}\D{1}.{3,6}',
'TUSCOLA': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{4}\D{1}\d{2}'
}
```

```
## 'EATON': '\d{2}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{2}',
## 'IOSCO': '\d{3}\w\d{10,11}',
## 'SHIAWASSEE': '\d{3}\D{1}\d{3}\D{1}\d{3}\D{1}\d{3}',
```

```
# Messages
```

```
messageSaveBeg = 'Saving \'
```

```
messageSaveEnd = '\ file . . .'
```

```
messageEOL = '. . . end of line'
```

```
# Functions
```

```
def getCurrentDirectory():
```

```

os.chdir(os.path.dirname(__file__))

return os.getcwd()

def getFiles(ext):
    for(dirpath,dirnames,filenames) in os.walk(getCurrentDirectory()):
        return (f for f in filenames if f.endswith(ext))

def printPDFfilename(fn):
    return 'Reading ' + fn + ' file . . .'

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 able 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 able to open output file for writing. Is the file currently open?')

```

```
PDFs = getFiles('.pdf')

for x in PDFs:
    ## Message Run
    print(printPDFfilename(x))

    ## Open PDF file for reading
    pdf = pdfplumber.open(x)

    lastPage = len(pdf.pages)

    ## Reset counters
    count = 0
    fileCount = fileCount + 1

    print('Number of Pages: ' + str(lastPage))

    fileName = outFilename + str(fileCount) + outputType
    openFile(fileName)

    for page in pdf.pages:
        p1 = pdf.pages[count]
        count = count + 1
        pltext = p1.extract_text()
        text = pltext.splitlines()

    ## Break for Testing
    ## if(count == 6):
    ##     break

    for item in text:
        Item_Found = False
        if (current_county == item[0:len(current_county)]):
            Item_Found = True
        else:
            for x in counties:
```

```

        if(x == item[0:len(x)]):
            current_county = x
            Item_Found = True
            break

    if(Item_Found == True):
        expression = counties.get(current_county)
        Parcel_Found = re.search(expression,item)
        if not (Parcel_Found is None):
            Parcel = item[Parcel_Found.start():Parcel_Found.end()]
            Parcel = Parcel.replace('-', '-')
            if(space in Parcel):
                blob = Parcel.split()
                Parcel = blob[0]
            restOfLine = item[Parcel_Found.end():len(item)]
            restOfLine = restOfLine.replace(comma,nospace)
            restOfLine = restOfLine.replace(dollarSign,nospace)
            carp2 = restOfLine.rsplit(space,1)
            PurchasePrice = carp2[1].strip()
            outString.append(Parcel + comma + PurchasePrice)

## Close PDF
pdf.close()

## Write CSV file
writeFile(fileName,outString)

## Clear variables
pl = ''
pltext = ''
text = ''
_TownshipFound = False
## Make sure outString is clear for next file's data
outString.clear()

#end timer

```

```
print('Runtime: ' + str(datetime.now() - start))
```

```
print(messageEOL)
```

```
# EOF
```


CSV Output

	A	B
1	Parcel	Purchase Price
2	010-028-100-015-00	7500
3	010-028-100-060-03	450
4	010-C30-000-009-00	7000
5	010-C30-000-183-00	200
6	021-000-017-022-00	8500
7	021-H05-013-021-00	2500
8	021-H05-014-006-00	3700
9	050-A10-000-020-00	14000
10	060-030-400-010-01	2500
11	060-030-400-010-02	11300
12	090-016-400-045-01	1550
13	090-028-200-005-03	500
14	090-028-200-005-04	450
15	100-S61-000-012-00	1800
16	110-F20-000-227-00	700
17	110-F20-000-228-00	700
18	120-K10-000-012-00	33000
19	160-R10-001-035-00	12000
20	200-003-000-172-00	1000
21	301-000-000-395-00	25000
22	301-000-000-530-00	20500
23	302-000-000-322-00	500
24	302-000-000-418-00	2000
25	302-000-000-975-29	100
26	302-000-000-975-30	100
27	303-000-000-702-00	100
28	303-000-000-824-00	16500

Michael Keller
Accumatch
20190808_MichiganExtra
Aug 9, 2019