

Michigan Auction List

Notes: This python scripts scrapes a folder of PDFs from a Michigan Auction Lists and creates a csv output file for each PDF. The following column headers are created and the associated data is collected: **Parcel, Minimum Bid, 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

Auction Results Report - UP Auction - September 17, 2002 - Marquette, MI

County	Jurisdiction	Local Par Num	Legal Desc	Sale Day	Sale Num	Min Bid	Purchase Price
KEWEENAW	TOWNSHIP OF GRANT	308-51-004-001	GJ-4-1 R LOT 1 BLK 4 PLAN OF THE VILLAGE OF WYOMING	09/17/02	1	\$550.00	\$1,150.00
KEWEENAW	TOWNSHIP OF GRANT	308-51-006-011	GJ-6-8 R N1/2 OF LOTS 11 & 12 BLK 6 PLAN OF THE VILLAGE OF WYOMING	09/17/02	2	\$700.00	\$3,000.00
KEWEENAW	TOWNSHIP OF HOUGHTON	403-51-008-005	HA-1-RB-5-6 R LOTS 5 & 6 LYING BETWEEN E & W MAIN ST. ALSO BETWEEN BLKS 8 & 9 RB TOWN OF EAGLE RIVER	09/17/02	3	\$650.00	\$2,100.00
BARAGA	TOWNSHIP OF BARAGA	002-156-003-000	TBG-6 0/4 102B SEC 6 T50N R34W COM AT FRL NW COR OF W 1/2 OF NE 1/4, THS 45 DEG E 46.6' TO POB, TH E 75', S 32 DEG 30M W 140', TH N 120' TO POB. .10 A.	09/17/02	4	\$500.00	unsold
BARAGA	TOWNSHIP OF BARAGA	002-229-006-000	TBG-29 1/4 302B SEC 29 T51N R34W PRT OF NW 1/4 OF SW 1/4 COM AT W 1/4 POST OF SEC 29, TH E 144', TH S 41 DEG W 229', TH N 176' TO POB. .29 A.	09/17/02	5	\$600.00	\$600.00
BARAGA	TOWNSHIP OF LANSE	004-106-012-000	TLA-6 9/2 303A SEC 6 T49N R32W FRL SW 1/4 OF SW 1/4 EXC COM AT COR POSTCOMMON TO SECS 1 & 12, IN 49-33 & SECS 6 & 7 IN 49-32, TH N 1,320', TH E 462' TO CEN OF HWY, TH SW ALG HWY 140' TO POB, TH SE AT RT ANGLE TO HWY 200', TH SW AT RT ANGLES 108', TH NW AT RT ANGLE 200', TH NE ALG HWY 108' TO POB. ALSO EXC 1.55 A FOR DSS & A RR R/W ALSO EXC COM AT SE COR OF SW 1/4 OF SW 1/4, TH N 1288', TH W 622', TH S 58 DEG W 50', TH S 32 DEG E 208', TH S 58 DEG W 104', TH N 32 DEG W 208', TH S 58 DEG W 246' TO A PT 50' FROM C/L OF FALLS RIVER, TH SELY ALG LN 50' FROM C/L OF RIVER TO A PT ON S LN OF SW 1/4 OF SW 1/4, TH E 250' TO POB. ALSO EXC THAT PRT OF THE SW 1/4 OF SW 1/4 LYING NWLY OF CO RD R/W ALSO EXC COM AT INT OF FALLS RIVER & CO RD, TH NELY ALG RD 50' TO POB, TH NELY ALG RD 200', TH SELY AT RT ANGLE TO RD 200', TH SWLY AT RT ANGLE 200', TH NWLY AT RT ANGLE 200' TO POB. 7.5 A.	09/17/02	6	\$1,050.00	unsold
BARAGA	TOWNSHIP OF LANSE	004-378-021-000	TLA-28 0/3 404J SEC 28 T50N R33W PRT OF SE 1/4 OF SE 1/4 LNG W OF US-41 COM 200' S OF NW COR, TH S 17.1', TH E TO SD HWY, TH N 17.1', TH W TO POB. .10 A.	09/17/02	7	\$450.00	unsold
BARAGA	TOWNSHIP OF LANSE	004-219-008-000	TLA-19 1/2 204G SEC 19 T51N R32W (TRACT NO 35, PRT OF SW 1/4 OF SE 1/4 OF NW 1/4) COM 990' W & 220' N OF CEN OF SEC 19, TH N 220', TH E 330', TH S 220', TH W 330' TO POB. 1.67 A.	09/17/02	8	\$600.00	\$600.00
BARAGA	TOWNSHIP OF SPURR	005-500-008-000	SP-P51 LOT 8 SUPERVISOR'S PLAT OF IMPERIAL HEIGHTS LOT 8.	09/17/02	9	\$750.00	unsold
BARAGA	VILLAGE OF L'ANSE	044-691-018-000	TLA-P33 18-19 1 VILLAGE OF L'ANSE KELSEY'S SUB NO 1 LOTS 18 & 19 BLK 1.	09/17/02	10	\$1,250.00	unsold
BARAGA	VILLAGE OF L'ANSE	044-549-017-75	TLA-9 0/3 202B6 VILLAGE OF L'ANSE SEC 9 T50N R33W PRT OF NW 1/4 OF NW 1/4 COM AT SE COR AS POB, TH N 0 DEG 18M W 30' ALG E LN, TH N 89 DEG 18M W 295.7', TH N 0 DEG 18M W 247.1' TO S LN OF KELSEY'S SUB #1, TH N 80 DEG W ALG WOODS AVE TO PT WHICH IS 300' W OF E LN OF NW 1/4 OF NW 1/4, TH S 0 DEG 18M E TO S LN OF NW 1/4 OF NW 1/4, TH E ALG S LN 300' TO POB. .23 A.	09/17/02	11	\$600.00	unsold
BARAGA	VILLAGE OF L'ANSE	044-549-017-000	TLA-9 0/3 202B2 VILLAGE OF L'ANSE SEC 9 T50N R33W PRT OF NW 1/4 OF NW 1/4 COM AT SE COR TH N 0 DEG 18M W 30' TO POB, TH N 0 DEG 18M W 195.7' TO SE COR OF KELSEY'S SUB #1, TH N 80 DEG W 120' ALG S LN OF WOODS AVE, TH S 0 DEG 18M E TO A PT 30' N OF S LN OF NW 1/4 OF NW 1/4, TH S 89 DEG 54M E PAR WITH S LN OF NW 1/4 OF NW 1/4 TO POB. .56 A.	09/17/02	12	\$3,300.00	unsold
LUCE	TOWNSHIP OF COLUMBUS	001-003-034-0250	SEC 34 T47N R11W SW 1/4 OF NE 1/4. 40 A.	09/17/02	13	\$4,300.00	unsold
LUCE	TOWNSHIP OF LAKEFIELD	002-001-010-1100	SEC 10 T45N R11W GOV'T LOT 2. 8 A.	09/17/02	14	\$500.00	\$2,000.00
LUCE	TOWNSHIP OF LAKEFIELD	002-002-026-2200	SEC 26 T45N R12W PT OF NW 1/4 OF NW 1/4 COM 209 FT S OF NW COR OF SAID NW 1/4 OF NW 1/4, TH S 616 FT, TH E 618.75 FT, TH S 495 FT, TH E 266.25 FT, TH N TO N LN OF NW 1/4 OF NW 1/4, TH W 27.24 FT, TH S 362.5 FT, TH W 402.5 FT, TH N 362.5 FT, TH W 246.26 FT, TH S 209 FT, TH W 209 FT TO POB. 15.65 A.	09/17/02	15	\$3,850.00	\$5,600.00

Python Code (20190806_MichiganAuction.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 = 'Michigan_'
outputType = '.csv'
fileName = ''
parcel = ''
amount = ''
current_county = 'NotFound'
```

```

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

# Regular Expression County Dictionary
counties = {
  'ALGER': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'ARENAC': '\d{3}[-]\d{1}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'BARAGA': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'BARRY': '\d{2}[-]\d{3}[-]\d{3}[-]\d{2}',
  'BAY': '\d{2}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'BENZIE': '\d{12}',
  'BRANCH': '\d{3}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CALHOUN': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CASS': '\d{2}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CHARLEVOIX': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CHIPPEWA': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CLARE': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CLINTON': '\d{3}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'CRAWFORD': '\d{3}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'DELTA': '\d{3}[-]\d{3}[-]\d{3}[-]\d{3}',
  'DICKINSON': '\d{3}[-]\d{3}[-]\d{3}[-]\d{0,4}',
  'EATON': '\d{2}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'EMMET': '\d{2}[-]\d{2}[-]\d{2}[-]\d{3}[-]\d{3}',
  'GOGEBIC': '\d{2}[-]\d{2}[-]\d{3}[-]\d{3}',
  'GRAND': '\d{2}[-]\d{3}[-]\d{3}[-]\d{2}',
  'HOUGHTON': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'IOSCO': '\d{3}\w\d{10,11}',
  'INGHAM': '\d{2}[-]\d{2}[-]\d{2}[-]\d{2}[-]\d{3}[-]\d{3}',
  'IRON': '\d{3}[-]\d{3}[-]\d{3}[-]\d{0,4}',
  'LEELANAU': '\d{3}[-]\d{3}[-]\d{3}[-]\d{2}',
  'KALAMAZOO': '\d{2}[-]\d{2}[-]\d{3}[-]\d{3}',
  'KALKASKA': '\d{3}[-]\d{3}[-]\d{3}[-]\d{0,4}',

```

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

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()]
        restOfLine = item[Parcel_Found.end():len(item)]
        restOfLine = restOfLine.replace(comma,nospace)
        restOfLine = restOfLine.replace(dollarSign,nospace)
        carp2 = restOfLine.rsplit(space,2)
        PurchasePrice = carp2[2].strip()
        MinimumBid = carp2[1].strip()
        outString.append(Parcel + comma + MinimumBid + comma + PurchasePrice)
## print('LINE: '+current_county+', '+Parcel+', '+MinimumBid+',
'+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	C
1	Parcel	Minimum Bid	Purchase Price
2	308-51-004-001	550	1150
3	308-51-006-011	700	3000
4	403-51-008-005	650	2100
5	002-156-003-00	500	unsold
6	002-229-006-00	600	600
7	004-106-012-00	1050	unsold
8	004-378-021-00	450	unsold
9	004-219-008-00	600	600
10	005-500-008-00	750	unsold
11	044-691-018-00	1250	unsold
12	044-549-017-75	600	unsold
13	044-549-017-00	3300	unsold
14	001-003-034-0250	4300	unsold
15	002-001-010-1100	500	2000
16	002-002-026-2200	3850	5600
17	003-002-031-1500	450	900
18	003-003-023-1500	850	unsold
19	003-003-024-2700	900	unsold
20	003-008-029-0900	500	unsold
21	041-100-240-1300	750	unsold
22	041-206-070-6100	750	unsold
23	051-003-016-00	1150	unsold
24	001-008-001-12	1550	1850
25	001-008-001-60	1550	unsold
26	001-018-009-60	1350	1350
27	001-018-009-10	1100	unsold
28	002-118-014-00	1200	2200

Michael Keller
Accumatch
20190806_MichiganAuction
Aug 8, 2019