

# State of Alabama Delinquent Land Available for Sale

**Notes:** This python scripts scrapes PDFs from different Alabama counties and creates a csv output file with the **PARCEL ID** and the **Amt Bid at Tax Sale**. The filename is unique based on the county name listed at the top of the PDF. A sample of the directory is listed below.

## PDF

| 2753 Jefferson-Bham                              |                     | STATE OF ALABAMA-DEPARTMENT OF REVENUE-PROPERTY TAX DIVISION |     |      |       |      |                        | PAGE NO: 1   |
|--|---------------------|--|-----|------|-------|------|------------------------|--|
|  |                     | TRANSCRIPT OF TAX DELINQUENT LAND AVAILABLE FOR SALE         |     |      |       |      |                        | DATE: 7/12/2019  |
| Name   | Amt Bid at Tax Sale | CO.  | YR. | C/S# | CLASS | CODE | PARCEL ID              | DESCRIPTION  |
| ALLEN NETTIE HAYGOOD<br>520                      | 80.82               | 01   | 00  | 0029 | 2     | 1    | 0123000220140020000000 | LOT 11 BLK 4 HOMER HIGHLANDS   |
| ASSOCIATES HOME EQUITY SER INC<br>4400           | 365.94              | 01   | 00  | 0051 | 2     | 32   | 0129000420260130000000 | LOT 4 BLK 17 COMPTON RISING  |
| BAILEY ALLEN<br>5120                             | 423.00              | 01   | 00  | 0059 | 2     | 33   | 0123002020030140000000 | LOT 12 BLK D RESUR OF BLKS C D EBORN'S ADD TO EAST BHAM  |
| BATES SARAH A<br>1120                            | 125.53              | 01   | 00  | 0082 | 2     | 1    | 0123001430180040000000 | LOT 16 BLK 4 EAST LAKE HGLDS   |
| BATTS WILLIAM<br>1080                            | 230.39              | 01   | 00  | 0086 | 2     | 32   | 0123002320180080000000 | LOTS 3 & 4 BLK 3 CENTRAL ADD TO GATE CITY  |
| BEARDEN INEZ<br>500                              | 319.36              | 01   | 00  | 0089 | 2     | 1    | 0122001330010710000000 | LOTS 36 & 37 BLK 3 MITCHELLS ADD TO N BHAM   |
| BERRY EDNA L<br>780                              | 481.61              | 01   | 00  | 0111 | 2     | 1    | 0122001410300030000000 | P O B INTER NW COR LOT 8 BLK 43 PARK PL & 38TH AVE TH NE 100FT ALG 38TH AVE TH SE 130 FT S TO HWY 3  |
| BERRY EDNA L<br>260                              | 2,368.60            | 01   | 00  | 0112 | 2     | 1    | 0122001540170010000000 | LOTS 1 THRU 6 BLK 73 NORTH BHAM LAND COS ADD #5  |
| BLOCK BERNARD<br>620                             | 71.39               | 01   | 00  | 0125 | 2     | 1    | 0116002000000320000000 | LOT 6 BLK 6 TOWN OF FLAT CREEK   |
| BLOCK BERNARD<br>2460                            | 163.75              | 01   | 00  | 0127 | 2     | 1    | 0116002700000410000000 | LOT 96 PORTER  |
| JARVIS JAMES E & VONIE B<br>920                  | 143.48              | 01   | 00  | 0131 | 2     | 1    | 0116001100000290000000 | COM AT INT OF W BD LINE OF SE 1/4 AND N R O W OF COUNTY ROAD NO 12 TH E ALONG N R O W OF SD RD 400 F |
| STILLWELL SHARON L B & FRANK<br>460              | 71.58               | 01   | 00  | 0132 | 2     | 1    | 0116002000000290000100 | LOT 1 BLK 4 TOWN OF FLAT CREEK   |
| TUGGLE JAMES WESLEY JR & TUGGLE RUBY NELL<br>420 | 66.76               | 01   | 00  | 0133 | 2     | 1    | 0116002700000330000000 | LOT 130 PORTER   |
| BROWN LOUISE<br>440                              | 75.89               | 01   | 00  | 0180 | 2     | 1    | 0122001340420050000000 | LOT 7 BLK 1 J H POOLS SUR  |

## Python Code

```
# Read and dump PDF file data to text file for Alabama Delinquent
# Accumatch
# by Michael Keller
# July 17, 2019
# requires pdfplumber, re, datetime

# Libraries
import pdfplumber
import re
from datetime import datetime

# Start Timer
start = datetime.now()

# Booleans
Parcel_Found = False
County_Found = False
State_Found = False
parcelFound = False

# Variables
parcel = ''
outputType = '.csv'
outputFile = ''
count = 0
outString = []
spaces = '\s{2,}'
space = ' '
nospace = ''
comma = ','
_return = '\r'
regParcel = '\d{13}'
stateLine = 'STATE OF ALABAMA'
colHeadings = 'Parcel ID,Amt Bid'

# Messages
messageRun = 'Reading \'input.pdf\' file . . .'
```

```

messageSaveBeg = 'Saving \''
messageSaveEnd = '\ ' file . . .'
messageEOL = '. . . end of line'

# PDF File Open
pdf = pdfplumber.open("WWW_TRANS_01.pdf")

print(messageRun)

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

## Break for Testing
## if(count == 3):
##     break

    for item in text:
        if((County_Found == False) and (State_Found == True)):
            # Do not run this if statement again i.e. change boolean to True
            State_Found = False
            County_Found = True
            # Read second line in PDF and create filename based on the text found
            carp = re.sub(spaces, space, item)
            carp = carp.replace(comma,nospace)
            carp = item.split(space,2)
            county = carp[1]
            outputFile = stateLine + ' - ' + county

        if((stateLine in item) and (County_Found == False)):
            State_Found = True

            # Search for Parcel number in line read from PDF
            Parcel_Found = re.search(regParcel, item)

            if(Parcel_Found is not None):

```

```
    parcel = item[Parcel_Found.start():Parcel_Found.end()]
    parcelFound = True
    elif(parcelFound == True):
        carp = re.sub(spaces, space, item)
        carp = carp.replace(comma,nospace)
        carp = carp.split(space,2)
        parcelFound = False
        outString.append(parcel + comma + str(carp[1]))

# PDF File Close
pdf.close()

filename = outputFile + outputFile
print(messageSaveBeg + filename + messageSaveEnd)

# Create and open output file
try:
    file = open(filename,'w')
    # Write out lines from array
    file.write(colHeadings + _return)
    # Now dump string array into file
    for taxLine in outString:
        file.write(taxLine + _return)
    file.close()
except PermissionError:
    print('ERROR: Unable able to open output file for writing. Is the file currently open?')

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

print(messageEOL)

# EOF
```




## Python Output (with run times)

```
RESTART: G:\Documents\Code\Python\code\20190717_AlabamaDelinquent\20190717_Delinquent.py
Reading 'input.pdf' file . . .
Saving 'STATE OF ALABAMA - Jefferson-Bham.csv' file . . .
Runtime: 0:01:56.479667
. . . end of line
>>>
RESTART: G:\Documents\Code\Python\code\20190717_AlabamaDelinquent\20190717_Delinquent.py
Reading 'input.pdf' file . . .
Saving 'STATE OF ALABAMA - Mobile.csv' file . . .
Runtime: 0:00:49.417039
. . . end of line
>>>
RESTART: G:\Documents\Code\Python\code\20190717_AlabamaDelinquent\20190717_Delinquent.py
Reading 'input.pdf' file . . .
Saving 'STATE OF ALABAMA - Montgomery.csv' file . . .
Runtime: 0:00:12.266812
. . . end of line
```

## CSV Output

|    | A            | B       |
|----|--------------|---------|
| 1  | Parcel ID    | Amt Bid |
| 2  | 123000220140 | 80.82   |
| 3  | 129000420260 | 365.94  |
| 4  | 123002020030 | 423     |
| 5  | 123001430180 | 125.53  |
| 6  | 123002320180 | 230.39  |
| 7  | 122001330010 | 319.36  |
| 8  | 122001410300 | 481.61  |
| 9  | 122001540170 | 2368.6  |
| 10 | 116002000000 | 71.39   |
| 11 | 116002700000 | 163.75  |
| 12 | 116001100000 | 143.48  |
| 13 | 116002000000 | 71.58   |
| 14 | 116002700000 | 66.76   |
| 15 | 122001340420 | 75.89   |
| 16 | 113003440010 | 140.77  |
| 17 | 123000310240 | 188.51  |
| 18 | 129000230070 | 246.76  |
| 19 | 129002120010 | 210.25  |
| 20 | 123001510010 | 443.01  |
| 21 | 113003330010 | 46.98   |
| 22 | 122001210010 | 100.01  |
| 23 | 129002120150 | 294.04  |
| 24 | 123002010020 | 429.97  |
| 25 | 129000840470 | 51.06   |
| 26 | 129000840470 | 64.21   |
| 27 | 123000310220 | 508.26  |

## Directory (Filenames of output files)

-  STATE OF ALABAMA - Jefferson-Bham.csv
-  STATE OF ALABAMA - Mobile.csv
-  STATE OF ALABAMA - Montgomery.csv

Michael Keller  
Accumatch  
20190717\_ALDelinquentLand  
Aug 17, 2019