

White Pine County

Notes: This python scripts scrapes PDFs from a Delinquent tax payer report from different counties and creates a csv output file with the following items: **Parcel, Taxes, Penalty and Interest, Mailing Costs, and Total.** The PDF filename is not unique and is static.

PDF

PARCEL #	NAME	TAXES	INTEREST/P ENALTY FEES	PUBLISH/ MAILING COSTS	TOTAL
*** A ***					
004-053-07	A'BELL, WAYNE P & EDITA B	\$282.00	\$30.08	\$3.50	\$315.58
001-043-09	ADAMOLI, TONY ET AL	\$181.81	\$27.40	\$3.50	\$212.71
004-164-14	ADAMS, JASON S & AMY J	\$538.51	\$80.96	\$3.50	\$622.97
002-069-03	ALDERMAN, MINNIS A	\$57.31	\$9.29	\$3.50	\$70.10
004-051-11	ALGHAZAWI, NAHED Z	\$58.00	\$2.32	\$3.50	\$63.82
005-591-57	ANDERSON, BARBARA	\$312.00	\$21.84	\$3.50	\$337.34
005-018-04	ANDERSON, FRED	\$11.55	\$2.54	\$3.50	\$17.59
005-018-06	ANDERSON, FRED	\$112.30	\$16.87	\$3.50	\$132.67
005-018-03	ANDERSON, FREDDIE & ALLE	\$23.14	\$5.10	\$3.50	\$31.74
005-018-05	ANDERSON, FREDDIE & ALLE	\$11.55	\$2.54	\$3.50	\$17.59
005-018-07	ANDERSON, FREDERICK J	\$34.68	\$7.63	\$3.50	\$45.81
001-023-02	ANDERSON, FREDRICK J	\$333.85	\$50.20	\$3.50	\$387.55
004-166-07	ASHDOWN, DEBRA A	\$239.68	\$16.57	\$3.50	\$259.75
*** B ***					
099-060-30	BAMBERGER INVESTMENT CO	\$18.30	\$4.03	\$3.50	\$25.83
002-077-04	BARELA, GEORGE JR	\$817.82	\$122.80	\$3.50	\$944.12
002-033-07	BARRY, DEONNA A	\$965.97	\$145.04	\$3.50	\$1,114.51
003-074-16	BATES, EDWARD	\$413.75	\$61.22	\$3.50	\$478.47
005-581-61	BECKETT, JANET	\$123.00	\$4.92	\$3.50	\$131.42
008-610-09	BENCHLIH, OTHMANE	\$242.03	\$36.44	\$3.50	\$281.97
004-152-12	BLACKHAM, ELAINE & SAMI JO	\$463.04	\$69.66	\$3.50	\$536.20
002-136-09	BOUNDY, JEAN MARIE ET AL	\$397.32	\$59.69	\$3.50	\$460.51
002-026-10	BOWEN, ANGELA FULLER	\$79.89	\$10.31	\$3.50	\$93.70
003-072-20	BURRILL, MICHAEL ROY	\$349.81	\$52.60	\$3.50	\$405.91
004-156-11	BUTCHER, ROY J.	\$279.00	\$29.76	\$3.50	\$312.26
003-061-10	BYRNE, GEORGE & PATRICIA	\$410.56	\$61.76	\$3.50	\$475.82
*** C ***					
005-651-19	CAMBRON, KELLY	\$129.13	\$19.46	\$3.50	\$152.09
013-087-10	CARTER, DANNIEL G & JENNIFER P.	\$603.43	\$90.76	\$3.50	\$697.69

Python Code (20190726_WhitePine.py)

```
# Read and dump PDF file data to text file for Stafford County
# Accumatch
# by Michael Keller
# July 19, 2019
# requires pdfplumber, datetime

# Libraries
import pdfplumber
from datetime import datetime

# Start Timer
start = datetime.now()

# Booleans
First_Parcel_Found = False

# Variables
astericks = '***'
checkAstericks = ''
parcel = ''
total = ''
mailingCosts = ''
pAndI = ''
taxes = ''
restOfLine = ''
comma = ','
space = ' '
nospace = ''
dollarSign = '$'
count = 0
outputType = '.csv'
outputFile = 'WhitePine'
outString = []
_return = '\r'
colHeadings = 'Parcel,Taxes,P&I,Mailing Costs,Total'
```

```

# Messages
messageRun = 'Reading \'input.pdf\' file . . .'
messageSaveBeg = 'Saving \''
messageSaveEnd = '\' file . . .'
messageEOL = '. . . end of line'

# PDF File Open
pdf = pdfplumber.open("2018 First Year Delinquent Ad.pdf")

print(messageRun)

for page in pdf.pages:
    p1 = pdf.pages[count]
    count = count + 1
    pltext = p1.extract_text()
    text = pltext.splitlines()
    for items in text:
        carp = items.strip()
        checkAstericks = carp[0:3]
        if((First_Parcel_Found == True) and (checkAstericks != astericks)):
            carp1 = carp.split(space,1)
            parcel = carp1[0].strip()
            restOfLine = carp1[1].strip()
            restOfLine = restOfLine.replace(comma,nospace)
            restOfLine = restOfLine.replace(dollarSign,nospace)
            carp2 = restOfLine.rsplit(' ',4)
            total = carp2[4].strip()
            mailingCosts = carp2[3].strip()
            pAndI = carp2[2].strip()
            taxes = carp2[1].strip()
            outString.append(parcel + comma + taxes + comma + pAndI + comma + mailingCosts +
comma + total)

        if((checkAstericks == astericks) and (First_Parcel_Found == False)):
            First_Parcel_Found = True

```

```
# 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
```

CSV Output

	A	B	C	D	E
1	Parcel	Taxes	P&I	Mailing Costs	Total
2	004-053-07	282	30.08	3.5	315.58
3	001-043-09	181.81	27.4	3.5	212.71
4	004-164-14	538.51	80.96	3.5	622.97
5	002-069-03	57.31	9.29	3.5	70.1
6	004-051-11	58	2.32	3.5	63.82
7	005-591-57	312	21.84	3.5	337.34
8	005-018-04	11.55	2.54	3.5	17.59
9	005-018-06	112.3	16.87	3.5	132.67
10	005-018-03	23.14	5.1	3.5	31.74
11	005-018-05	11.55	2.54	3.5	17.59
12	005-018-07	34.68	7.63	3.5	45.81
13	001-023-02	333.85	50.2	3.5	387.55
14	004-166-07	239.68	16.57	3.5	259.75
15	099-060-30	18.3	4.03	3.5	25.83
16	002-077-04	817.82	122.8	3.5	944.12
17	002-033-07	965.97	145.04	3.5	1114.51
18	003-074-16	413.75	61.22	3.5	478.47
19	005-581-61	123	4.92	3.5	131.42
20	008-610-09	242.03	36.44	3.5	281.97
21	004-152-12	463.04	69.66	3.5	536.2
22	002-136-09	397.32	59.69	3.5	460.51
23	002-026-10	79.89	10.31	3.5	93.7
24	003-072-20	349.81	52.6	3.5	405.91
25	004-156-11	279	29.76	3.5	312.26
26	003-061-10	410.56	61.76	3.5	475.82
27	005-651-19	129.13	19.46	3.5	152.09
28	013-087-10	603.43	90.76	3.5	697.69

Michael Keller
Accumatch
20190726_WhitePine
July 26, 2019