Stafford County – Mortgage Listing

Notes: This python scripts scrapes PDFs from the Stafford County Mortgage Listing. The first page is ignored because it is descriptive only and does not contain any mortgage data. The CSV output file contains the following columns: **Map No., Bill No., Pin, Tax Half, Tax Annual, Delinquent, Tax Half + Delinquent**. The file output name is not unique.

PDF

STAFFORD COUNTY - MORTGAGE LISTING

958 ACCUMATCH   Tax Service: 000

	NAME and ADDRESS	LEGAL DESCRIPTION	
MAP NO: 58A 1 20A **Bill No:** 38116 **Pin:** 38949	ADAMS JAMES &MARY ALICE AVERY ADAMS 106 WINDSOR CIR FREDERICKSBURG VA 22405-3136	LOT 20A-1 14210 SQ FT ARGYLE HEIGHTS 106 WINDSOR CIR	**Tax Half:** $2,027.07 **Tax Annual:** $4,054.14 **Delinquent:** $0.00 **Tax Half + Delinquent:** $2,027.07
MAP NO: 37 44 **Bill No:** 23069 **Pin:** 23528	ALVAREZ NOAH MAURICIO & ALVAREZ SABI 1023 MOUNTAIN VIEW RD FREDERICKSBURG VA 22406-4429	POTOMAC RUN 6.0067 AC 1023 MOUNTAIN VIEW RD	**Tax Half:** $2,017.48 **Tax Annual:** $4,034.96 **Delinquent:** $0.00 **Tax Half + Delinquent:** $2,017.48
MAP NO: 30C 1 18 **Bill No:** 19342 **Pin:** 19719	ALVAREZ RAUL REYNOSO 8 WALKER WAY STAFFORD VA 22554-5335	LOT 18 SEC 1 STAFFORD HEIGHTS 8 WALKER WAY	**Tax Half:** $1,268.06 **Tax Annual:** $2,536.12 **Delinquent:** $0.00 **Tax Half + Delinquent:** $1,268.06
MAP NO: 44CC 7A 990 **Bill No:** 47163 **Pin:** 48392	APPLE GEORGIANA G 105 AMBROSE LN FREDERICKSBURG VA 22406-5454	LOT 990 SEC 7A 3412 SF CELEBRATE VIRGINIA NORTH 105 AMBROSE LN	**Tax Half:** $1,043.33 **Tax Annual:** $2,086.66 **Delinquent:** $0.00 **Tax Half + Delinquent:** $1,043.33
MAP NO: 53K 7 78 **Bill No:** 50266 **Pin:** 51750	BAUER MICHELLE MILLER 311 SPRING PARK LN FREDERICKSBURG VA 22405-0000	UNIT 78 PHASE 7 RAPPAHANNOCK LANDING CONDOMINIUM 311 SPRING PARK LN	**Tax Half:** $1,156.96 **Tax Annual:** $2,313.92 **Delinquent:** $0.00 **Tax Half + Delinquent:** $1,156.96
MAP NO: 20BB 2 34	BENNETT WILLIAM C 11 DARBYWOOD CT	LOT 34 10000 SQFT PHASE 2 MEADOWOOD PARK	**Tax Half:** $1,364.51

Python Code

```python
# Read and dump PDF file data to text file for Stafford County

# Accumatch

# by Michael Keller

# July 19, 2019

# requires pdfplumber, re, datetime


# Libraries

import pdfplumber

import re

from datetime import datetime


# Start Timer

start = datetime.now()


# Booleans

MAP_NO_Found = False

PO_BOX_Found = False

mapNoFound = False

Bill_No_Found = False

billNoFound = False

Pin_Found = False

pinFound = False

TaxHalf_Found = False

taxHalfFound = False


# Variables

flagMap = 'MAP NO:'

flagBill = 'Bill No:'

flagPin = 'Pin: Delinquent:'

flagTaxHalf = 'Tax Half + Delinquent:'

mapNo = ''

billNo = ''

Pin = ''

taxHalf = ''

taxAnnual = ''
```

```python
deliquent = ''

taxHalfDelinquent = ''

comma = ','

space = ' '

nospace = ''

dollarSign = '$'


count = 0

outputType = '.csv'

outputFile = 'StaffordCounty'

outString = []

_newline = '\n'

_return = '\r'

colHeadings = 'Map No.,Bill No.,Pin,Tax Half,Tax Annual,Delinquent,Tax Half + Delinquent'


regAddressStart = "[' ']\d+[' '][a-zA-Z]"

regTaxHalf = '[$].{1,}[.]\d{2}'

regPOBox = "[' ']PSC"


# Messages

messageRun = 'Reading \'input.pdf\' file . . .'

messageSaveBeg = 'Saving \''

messageSaveEnd = '\' file . . .'

messageEOL = '. . . end of line'


# PDF File Open

pdf = pdfplumber.open("Accumatch.pdf")


print(messageRun)


for page in pdf.pages:

    p1 = pdf.pages[count]

    count = count + 1

    pltext = p1.extract_text()


    if(count != 1):
```

```python
        text = pltext.splitlines()

    for items in text:


        if(mapNoFound == True):
            mapNoFound = False
            PO_BOX_Found = re.search(regPOBox,items)
            MAP_NO_Found = re.search(regAddressStart,items)
            if(PO_BOX_Found is not None):
                carp = items.lstrip()
                carp = carp[0:PO_BOX_Found.start()]
                mapNo = carp.strip()
            else:
                if(MAP_NO_Found is not None):
                    carp = items.lstrip()
                    carp = carp[0:MAP_NO_Found.start()]
                    mapNo = carp.strip()
            TaxHalf_Found = re.search(regTaxHalf,items)
            if(TaxHalf_Found is not None):
                carp = items[TaxHalf_Found.start()+1:TaxHalf_Found.end()]
                taxHalf = carp.replace(comma,nospace)
            mapNoFound = False
        if(billNoFound == True):
            carp = items.lstrip()
            carp = carp.split(space,1)
            billNo = carp[0]
            TaxHalf_Found = re.search(regTaxHalf,carp[1])
            if(TaxHalf_Found is not None):
                carp2 = carp[1][TaxHalf_Found.start()+1:TaxHalf_Found.end()]
                taxAnnual = carp2.replace(comma,nospace)
            billNoFound = False
        if(pinFound == True):
            carp = items.lstrip()
            carp = carp.split(space,1)
            Pin = carp[0]
            TaxHalf_Found = re.search(regTaxHalf,carp[1])
            if(TaxHalf_Found is not None):
```

```python
                carp2 = carp[1][TaxHalf_Found.start()+1:TaxHalf_Found.end()]

                deliquent = carp2.replace(comma,nospace)

                deliquent = deliquent.replace(dollarSign,nospace)

            pinFound = False

        if(taxHalfFound == True):

            taxHalfDelinquent = items.strip()

            taxHalfDelinquent = taxHalfDelinquent.replace(comma,nospace)

            taxHalfDelinquent = taxHalfDelinquent.replace(dollarSign,nospace)

            taxHalfFound = False

            outString.append(mapNo + comma + billNo + comma + Pin + comma + taxHalf + comma +
taxAnnual + comma + deliquent + comma + taxHalfDelinquent)


        if flagMap in items:

            mapNoFound = True

        if flagBill in items:

            billNoFound = True

        if flagPin in items:

            pinFound = True

        if flagTaxHalf in items:

            taxHalfFound = True


# PDF File Close

pdf.close()


filename = outputFile + outputType

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:
```

```python
    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	F	G
1	Map No.	Bill No.	Pin	Tax Half	Tax Annual	Delinquent	Tax Half + Delinquent
2	58A 1 20A	$38,116.00	$38,949.00	$2,027.07	$4,054.14	$0.00	$2,027.07
3	37 44	$23,069.00	$23,528.00	$2,017.48	$4,034.96	$0.00	$2,017.48
4	30C 1 18	$19,342.00	$19,719.00	$1,268.06	$2,536.12	$0.00	$1,268.06
5	44CC 7A 990	$47,163.00	$48,392.00	$1,043.33	$2,086.66	$0.00	$1,043.33
6	53K 7 78	$50,266.00	$51,750.00	$1,156.96	$2,313.92	$0.00	$1,156.96
7	20BB 2 34	$5,809.00	$5,958.00	$1,364.51	$2,729.02	$0.00	$1,364.51
8	20HH 10	$6,896.00	$7,048.00	$1,825.58	$3,651.16	$0.00	$1,825.58
9	21B 1032	$11,993.00	$12,195.00	$2,962.84	$5,925.68	$0.00	$2,962.84
10	53 81 2	$30,898.00	$31,612.00	$2,487.63	$4,975.26	$0.00	$2,487.63
11	44CC 6 814	$47,047.00	$48,276.00	$1,117.06	$2,234.12	$0.00	$1,117.06
12	46M 6C 95	$44,842.00	$45,949.00	$1,469.05	$2,938.10	$0.00	$1,469.05
13	34A 2 21	$22,113.00	$22,559.00	$1,480.66	$2,961.32	$0.00	$1,480.66
14	25C 3 39	$15,796.00	$16,021.00	$1,589.24	$3,178.48	$0.00	$1,589.24
15	44CC 6 692	$46,939.00	$48,162.00	$1,323.10	$2,646.20	$0.00	$1,323.10
16	44R 14B 1064	$49,713.00	$51,190.00	$1,642.77	$3,285.54	$0.00	$1,642.77
17	44CC 3A2 57	$51,469.00	$53,328.00	$1,681.65	$3,363.30	$0.00	$1,681.65
18	21B 1699	$12,619.00	$12,821.00	$1,644.79	$3,289.58	$0.00	$1,644.79
19	16P 3	$48,024.00	$49,292.00	$2,150.80	$4,301.60	$0.00	$2,150.80
20	20P 7 276	$8,263.00	$8,415.00	$2,097.77	$4,195.54	$0.00	$2,097.77
21	28H 2 29	$17,474.00	$17,762.00	$1,748.82	$3,497.64	$0.00	$1,748.82
22	49H 1 8	$30,691.00	$31,402.00	$5,085.35	$10,170.70	$0.00	$5,085.35
23	27F 1 18	$16,473.00	$16,735.00	$2,319.47	$4,638.94	$0.00	$2,319.47
24	46H 23	$29,413.00	$30,109.00	$790.33	$1,580.66	$0.00	$790.33
25	20JJ 2 47	$7,174.00	$7,326.00	$1,378.15	$2,756.30	$0.00	$1,378.15
26	29E 1 39	$43,396.00	$44,438.00	$2,514.40	$5,028.80	$0.00	$2,514.40
27	44R 13 985	$47,561.00	$48,805.00	$1,727.61	$3,455.22	$0.00	$1,727.61
28	19H 3 131	$4,672.00	$4,780.00	$1,593.78	$3,187.56	$0.00	$1,593.78

Michael Keller
Accumatch
20190722_StaffordCounty
July 22, 2019