

Orange County

Notes: This python scripts scrapes PDFs from an Orange County Mortgage Company Listing and creates a csv output file with the following items: **Account No., Customer Name, and Customer Address.** The PDF filename is not unique and is static.

PDF

DATE	09/14/2016	13:36	MORTGAGE COMPANY LISTING	ORANGE COUNTY TAX OF	PAGE	1			
MTG	ACCOUNT NO	NAME/ADDRESS	PROPERTY DESCRIPTION	LOAN #	DELETE	KEEP			
450	000001-001125	BURNS, BERMAN D. & PRISCELLA E. 4532 STALLION ORANGE	TX 77632-8779	1 ABST. 1 JOHN ALLEN, TRA CT 011, ACRES 6.934	---	---			
450	000001-001171	HICKEY, MICHAEL EDMUND 4401 STALLION RD ORANGE	TX 77632-4709	1 ABST. 1 JOHN ALLEN, TRA CT 011D, ACRES 6	---	---			
450	000001-001201	FOREMAN, JIMMIE R. 4616 STALLION RD ORANGE	TX 77632-8700	1 ABST. 1 JOHN ALLEN, TRA CT 011B, ACRES 2.1	---	---			
450	000001-001295	SIMON, SAMUEL WADE 4600 STALLION RD ORANGE	TX 77632-8700	1 ABST. 1 JOHN ALLEN, TRA CT 011H, ACRES 5	---	---			
450	000001-004101	STARK, CASEY DON & LEAH GUILLOTTE 3182 ALICE ST ORANGE	TX 77632-8700	1 ABST. 1 JOHN ALLEN, TRA CT 018A, ACRES 0.72	---	---			
450	000002-005701	FLOYD, BRYANT E. 3290 JERRY ST. ORANGE	TX 77632-6002	2 ABST. 2 RICHARD BALLEW , TRACT 062, ACRES 0.54	---	---			
450	000002-010401	COMPTON, RICK A., JR 3110 FM 3247 ORANGE	TX 77632-2066	ABST. 2 R. BALLEW @ EAST SIDE OF TR 126	---	---			
450	000002-011601	PEET, OSCAR J., JR 3264 W BLUFF RD ORANGE	TX 77632-3763	ABST. 2 RICHARD BALLEW P ART OF TR 4 JOHN PEET TRA CT TR 135	---	---			
450	000002-018911	PERRY, LEO & RHONDA, JR 1090 W BLUFF RD ORANGE	TX 77632-3798	2 ABST. 2 RICHARD BALLEW , TRACT 092A, ACRES 0.41	---	---			
450	000002-019730	LOVELACE, LESLIE M 1149 W BLUFF RD ORANGE	TX 77632-0728	2 ABST. 2 RICHARD BALLEW , TRACT 206, ACRES 1	---	---			
450	000002-024801	JENNINGS, JUSTIN MORGAN & PRISCILLA 192 GLIDDEN ST ORANGE	TX 77632-0780	2 ABST. 2 RICHARD BALLEW , TRACT 235, ACRES 0.492	---	---			
450	000002-026901	GOTTE, SHANE & ANGELA 8207 HIGHWAY 87 N ORANGE	TX 77632-1278	2 ABST. 2 RICHARD BALLEW , TRACT 274, ACRES 0.794	---	---			
450	000002-028901	BURNETT, DEON 4066 LITTLE CYPRESS DR ORANGE	TX 77632-0530	2 ABST. 2 RICHARD BALLEW , TRACT 271, ACRES 1	---	---			
450	000002-034001	BROOMES, ROSE MARY 3002 PARISH CEMETERY RD ORANGE	TX 77632-1720	2 ABST. 2 RICHARD BALLEW , TRACT 312, ACRES 3	---	---			
450	000002-039031	MCCORVY, MELISSA & BRIAN 2798 BEN MAC RD ORANGE	TX 77632-1782	2 ABST. 2 RICHARD BALLEW , TRACT 384E, ACRES 3.344	---	---			
450	000003-009501	HARRIS, SAMUEL EDWIN 285 BLUEBIRD ST VIDOR	TX 77632-7165	ABST. 3 T. H. BREECE TR 086	---	---			
450	000003-010801	FILLINGAME, JESSIE EARL, JR 2935 WIER DR BEAUMONT	TX 77662-4802	ABST. 3 T. H. BREECE TR 100	---	---			
450	000003-012101	PICKERING, JONATHAN N. P.O. BOX 1662 VIDOR	TX 77706-7832	ABST. 3 T. H. BREECE LOT 4 TR 114	---	---			
450	000003-015801	CLEVELAND, ELBERT, JR 660 PINE ST VIDOR	TX 77662-1662	ABST. 3 T. H. BREECE TR 147	---	---			
			TX 77662-4314						

Python Code (20190726_OrangeCounty.py)

```
# Read and dump PDF file data to text file for Orange County
# Accumatch
# by Michael Keller
# July 29, 2019
# requires pdfplumber, re, datetime

# Libraries
import pdfplumber
import re
from datetime import datetime

# Start Timer
start = datetime.now()

# Reg Expressions
regDateField = "^\\d{1,2}[\\./]\\d{1,2}[\\./]\\d{4}"
spaces = '\\s{2,}'

# Booleans
dateFound = False
recordFound = False

# Variables
dateString = 'DATE'
recordString = 'MTG'
accumatchCode = '450'
addressCount = 0
custAccountNumber = ''
custName = ''
custAddress = ''
custCity = ''
custState = ''
custZip = ''
inputFile = 'Orange.pdf'
outputType = '.csv'
```

```
outputFile = 'OrangeCounty'
filename = ''
count = 0
outString = []
space = ' '
comma = ','
replacementChar = '~'
_return = '\r'
colHeadings = 'Account No,Name,Address,City,State,Zip'

# PDF File Open
pdf = pdfplumber.open(inputFile)

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

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:
        blob = item[0:76].strip()

        if(blob[0:4] == dateString):
            recordFound = False
```

```
    dateFound = True

    if((dateFound == True) and (recordFound == True)):
        blob = re.sub(comma, replacementChar, blob)

        if(blob[0:3] == accumatchCode):
            addressCount = 0
            custAccountNumber = blob[3:24].strip()
            custName = blob[29:len(blob)].strip()

            if(addressCount == 1):
                custAddress = blob[0:len(blob)].strip()

            if(addressCount == 2):
                custCity = blob[0:25].strip()
                custState = blob[26:28].strip()
                custZip = blob[29:39].strip()

            addressCount = addressCount + 1

            if(addressCount == 3):
                outString.append(custAccountNumber + comma + custName + comma + custAddress +
                comma + custCity + comma + custState + comma + custZip)

            if(blob[0:3] == recordString):
                recordFound = 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	F
1	Account No	Name	Address	City	State	Zip
2	000001-001125	BURNS~ BERMAN D. & PRISCELLA E.	4532 STALLION	ORANGE	TX	77632-8779
3	000001-001171	HICKEY~ MICHAEL EDMUND	4401 STALLION RD	ORANGE	TX	77632-4709
4	000001-001201	FOREMAN~ JIMMIE R.	4616 STALLION RD	ORANGE	TX	77632-8700
5	000001-001295	SIMON~ SAMUEL WADE	4600 STALLION RD	ORANGE	TX	77632-8700
6	000001-004101	STARK~ CASEY DON & LEAH GUILLOTTE	3182 ALICE ST	ORANGE	TX	77632-6002
7	000002-005701	FLOYD~ BRYANT E.	3290 JERRY ST.	ORANGE	TX	77632-2066
8	000002-010401	COMPTON~ RICK A.~ JR	3110 FM 3247	ORANGE	TX	77632-3763
9	000002-011601	PEET~ OSCAR J.~ JR	3264 W BLUFF RD	ORANGE	TX	77632-3798
10	000002-018911	PERRY~ LEO & RHONDA~ JR	1090 W BLUFF RD	ORANGE	TX	77632-0728
11	000002-019730	LOVELACE~ LESLIE M	1149 W BLUFF RD	ORANGE	TX	77632-0780
12	000002-024801	JENNINGS~ JUSTIN MORGAN & PRISCILLA	192 GLIDDEN ST	ORANGE	TX	77632-1278
13	000002-026901	GOTTE~ SHANE & ANGELA	8207 HIGHWAY 87 N	ORANGE	TX	77632-0530
14	000002-028901	BURNETT~ DEON	4066 LITTLE CYPRESS DR	ORANGE	TX	77632-1720
15	000002-034001	BROOMES~ ROSE MARY	3002 PARISH CEMETERY RD	ORANGE	TX	77632-1782
16	000002-039031	MCCORVY~ MELISSA & BRIAN	2798 BEN MAC RD	ORANGE	TX	77632-7165
17	000003-009501	HARRIS~ SAMUEL EDWIN	285 BLUEBIRD ST	VIDOR	TX	77662-4802
18	000003-010801	FILLINGAME~ JESSIE EARL~ JR	2935 WIER DR	BEAUMONT	TX	77706-7832
19	000003-012101	PICKERING~ JONATHAN N.	P.O. BOX 1662	VIDOR	TX	77662-1662
20	000003-015801	CLEVELAND~ ELBERT~ JR	660 PINE ST	VIDOR	TX	77662-4314
21	000004-002117	HUGHES~ WILLIAM ROBERT~ JR	5032 OLD PEVETO RD	ORANGE	TX	77632-7041
22	000004-002153	HENDERSON~ DALE A. & VICKY A.	5150 OLD PEVETO RD	ORANGE	TX	77632-7044
23	000004-002501	TARTER~ GERALD~ JR	4410 OLD PEVETO RD	ORANGE	TX	77632-5935
24	000004-003901	HARRIS~ FRANKLIN A.~ JR	3865 OLD PEVETO RD	ORANGE	TX	77632-5924
25	000004-003925	HARRIS~ FRANKLIN A.~ JR	3865 OLD PEVETO RD	ORANGE	TX	77632-5924
26	000004-003951	HARRIS~ FRANKLIN A.~ JR	3865 OLD PEVETO RD	ORANGE	TX	77632-5924
27	000004-004301	ZAVADA~ HELEN D.	3803 OLD PEVETO RD	ORANGE	TX	77632-5924
28	000004-004311	ZAVADA~ HELEN DARLENE	3803 OLD PEVETO RD	ORANGE	TX	77632-5924
29	000004-008927	MATTHEWS~ CECIL J.	3641 FM 1078	ORANGE	TX	77632-8371

Michael Keller
 Accumatch
 20190726_OrangeCounty
 July 29, 2019