Using Python to Wrangle Public Datasets: Researching Property Ownership in Detroit

Eric Seymour, Ph.D.
Urban and Regional Planning Program
University of Michigan
eseymour@umich.edu

Network Detroit 2016

09/30/2016
Key Terms

- Real Estate Owned Properties (REOs)
- Government Sponsored Enterprises (GSEs)
- U.S. Department of Housing and Urban Development (HUD)
Detroit Foreclosures

Figure 1: Completed Mortgage Foreclosures, Detroit 2005–2013. Sources: CoreLogic, Wayne County Register of Deeds
Theory of Change (simplified)

Foreclosure → Vacancy → Blight → Lower home prices → Disinvestment

Eric Seymour  Using Python to Wrangle Public Datasets
The Role of Investors

- Foreclosure
  - Vacancy
  - Blight
- Investor purchase
- Lower home prices
- Disinvestment
Foreclosure Process

1. Delinquency
2. Shortsale
3. Foreclosure
4. Sold at auction?
   - yes → Foreclosure Sale
   - no → REO
5. FHA insured?
   - yes → Conveyed to HUD
   - no → Loan owned by GSE?
     - yes → Conveyed to GSE
     - no → Privately owned
   - no → Conveyed to GSE
Data Analysis

■ Tasks:
  ■ Identify REOs
  ■ Identify REO owners
  ■ Identify REO buyers
  ■ Link REOs to tax foreclosure and blight records

■ Data Sources:
  ■ Wayne County Register of Deeds
  ■ Data Driven Detroit
  ■ Detroit Open Data
Data Collection and Storage

- Python for cleaning and formatting
- SQLite for data storage

```python
import csv
import sqlite3 as sql

with open("myfile.csv", "r") as f: # open file in read-only mode
    reader = csv.reader(f) # create reader object
    for row in reader: # for every row in the datafile...
        clean row
        insert into database
```
import sqlite3 as sql

con = sql.connect("mydatabase.sqlite")  # connect to db
cur = con.cursor()  # create cursor object
query = """
    SELECT * FROM records
    WHERE deed = "Sheriff's Deed";
"""

cur.execute(query)  # run query
results = cur.fetchall()  # get results
for row in results:  # iterate over results
    print row

con.close()  # close connection to db
<table>
<thead>
<tr>
<th>Date</th>
<th>Amount ($)</th>
<th>Deed</th>
<th>Seller</th>
<th>Buyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-03-18</td>
<td>50,559</td>
<td>SHD</td>
<td>KOPANAKIS NICHOLAS S</td>
<td>MORTGAGE ELECTRONIC REGISTRATION SYSTEMS INC</td>
</tr>
<tr>
<td>2008-04-25</td>
<td>1</td>
<td>QCD</td>
<td>MORTGAGE ELECTRONIC REGISTRATION SYSTEMS INC NOM</td>
<td>FEDERAL NATIONAL MTG ASSN</td>
</tr>
<tr>
<td>2008-12-02</td>
<td>373</td>
<td>QCD</td>
<td>FEDERAL NATIONAL MTG ASSN, TROTT &amp; TROTT ATTY</td>
<td>HOMESOLUTIONS PROPERTIES</td>
</tr>
<tr>
<td>2010-08-23</td>
<td>2,165</td>
<td>QCD</td>
<td>HOMESOLUTIONS PROPERTIES LLC</td>
<td>HARRIS RICKY M</td>
</tr>
<tr>
<td>2011-08-15</td>
<td>3,045</td>
<td>QCD</td>
<td>TOUVAK ENTERPRISES LLC</td>
<td>KOREM FIVE LLC</td>
</tr>
<tr>
<td>2012-09-12</td>
<td></td>
<td>JOF</td>
<td>KOPANAKIS NICHOLAS S</td>
<td>WAYNE COUNTY TREASURER</td>
</tr>
<tr>
<td>2012-12-06</td>
<td>0</td>
<td>QCD</td>
<td>WAYNE COUNTY TREASURER</td>
<td>WS BUSINESS SOLUTIONS, INC</td>
</tr>
</tbody>
</table>
Finding REO Owners

Sheriff’s Auction

MERS buyer?

conveyed to Fed?

‘bought’ by bank?

search later records

fed owned (REO)

not REO

bank owned (REO)
Regular expressions (regex) in Python

```python
import re

fannie = re.compile('.*FNMA.*|^FAN.*MAE.*|^FED.*NAT.*')
names = ['"FEDERAL NATL MTG ASSN","FANNIE MAE","FNMA"]
for name in names:
    if fannie.match(name):
        print "This name matches: ", name
    else:
        print "This name does not match: ", name
```
regex = '.*\bMERS\b.*|.*\bMERS\d*\b.*|.*MORT.*ELEC.*'
mers_regex = re.compile(regex)

owner = None # empty variable to fill using script
for row in results: # from SQL query of property records
    buyer = row[0] # select buyer field from row of data
    if mers_regex.match(buyer): # does string match MERS?
        pass # if yes, skip and look at names in later records
    else: # if no, accept first buyer name as owner
        owner = buyer
        break # break from loop
Classifying Investors

- Large investors: $\geq 50$
- Medium investors: $\geq 10$ and $< 50$
- Small investors: $\geq 2$ and $< 10$

```python
# iterate over Python dictionary w/ buyer as k(e)y
for k, v in property_dict.iteritems():
    reo_buyer_category = None
    if v["count"] > 50:
        reo_buyer_category = "large_investor"
    elif v["count"] >= 10:
        reo_buyer_category = "medium_investor"
    elif v["count"] >= 2 or v["investor_flag"] == 1:
        reo_buyer_category = "small_investor"
```

Eric Seymour

Using Python to Wrangle Public Datasets
import pandas as pd
import sqlite3 as sql

con = sql.connect("mydatabase.sqlite") # connect to db
qry = "SELECT sale_date, sale_amount, owner FROM myTable;"
df = pd.read_sql(qry, con) # create dataframe from query
print df.head() # get first n rows
print df.describe() # get descriptive statistics
## Parties Taking Foreclosures, 2005–2013

<table>
<thead>
<tr>
<th>Owner</th>
<th>Foreclosures</th>
<th>Share of Total Foreclosures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Entities</td>
<td>41,460</td>
<td>59.70%</td>
</tr>
<tr>
<td>Fannie Mae</td>
<td>12,756</td>
<td>18.37%</td>
</tr>
<tr>
<td>HUD</td>
<td>9,981</td>
<td>14.37%</td>
</tr>
<tr>
<td>Freddie Mac</td>
<td>2,822</td>
<td>4.06%</td>
</tr>
<tr>
<td>Likely Investors</td>
<td>1,663</td>
<td>2.39%</td>
</tr>
<tr>
<td>VA</td>
<td>443</td>
<td>0.64%</td>
</tr>
<tr>
<td>Likely Individuals</td>
<td>248</td>
<td>0.36%</td>
</tr>
<tr>
<td>City and Nonprofit Entities</td>
<td>70</td>
<td>0.10%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>69,443</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>
REO Sales by Buyer Type

Strong Neighborhoods

Other Neighborhoods

Sales (left)
City Nonprofit
Individual
Small Investor
Medium Investor
Large Investor
<table>
<thead>
<tr>
<th>Category</th>
<th>Fannie Mae</th>
<th>Freddie Mac</th>
<th>HUD</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>36.6%</td>
<td>39.6%</td>
<td>51.3%</td>
<td>23.7%</td>
</tr>
<tr>
<td>Large Investors</td>
<td>15.7%</td>
<td>7.0%</td>
<td>5.8%</td>
<td>20.1%</td>
</tr>
<tr>
<td>Medium Investors</td>
<td>12.0%</td>
<td>12.7%</td>
<td>9.7%</td>
<td>16.9%</td>
</tr>
<tr>
<td>Small Investors</td>
<td>34.0%</td>
<td>40.0%</td>
<td>32.8%</td>
<td>38.3%</td>
</tr>
<tr>
<td>City and Nonprofits</td>
<td>1.7%</td>
<td>0.7%</td>
<td>0.4%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
## Tax Foreclosure by REO Buyer Type, 2008–2015

<table>
<thead>
<tr>
<th>Buyer Type</th>
<th>Properties</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prior to Second Sale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>4,179</td>
<td>22.9%</td>
</tr>
<tr>
<td>Large Investors</td>
<td>3,403</td>
<td>35.1%</td>
</tr>
<tr>
<td>Medium Investors</td>
<td>2,840</td>
<td>33.0%</td>
</tr>
<tr>
<td>Small Investors</td>
<td>6,944</td>
<td>32.2%</td>
</tr>
<tr>
<td>City and Nonprofits</td>
<td>100</td>
<td>17.0%</td>
</tr>
<tr>
<td><strong>Anytime after Purchase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>4,745</td>
<td>26.1%</td>
</tr>
<tr>
<td>Large Investors</td>
<td>6,558</td>
<td>67.7%</td>
</tr>
<tr>
<td>Medium Investors</td>
<td>4,490</td>
<td>52.1%</td>
</tr>
<tr>
<td>Small Investors</td>
<td>8,901</td>
<td>41.3%</td>
</tr>
<tr>
<td>City and Nonprofits</td>
<td>117</td>
<td>19.9%</td>
</tr>
</tbody>
</table>