

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

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

Detroit Foreclosures

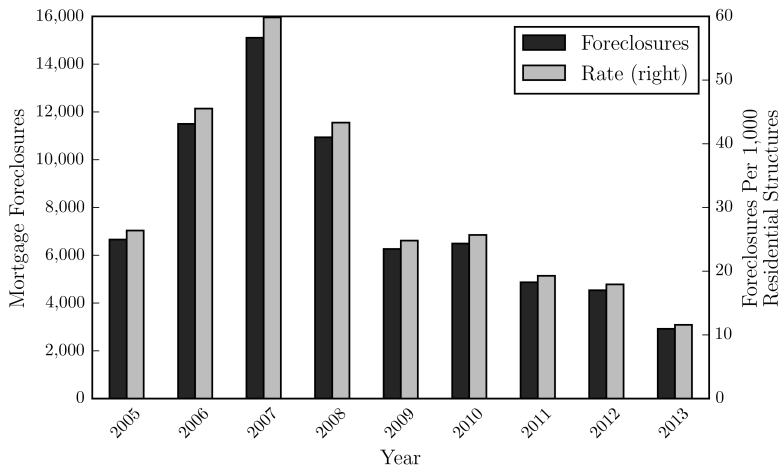
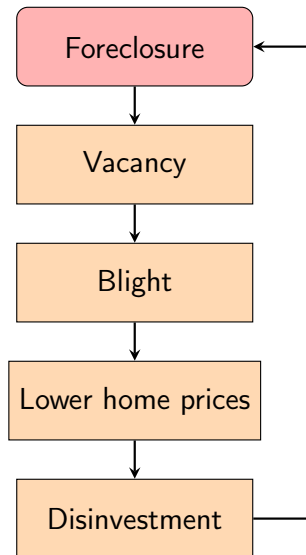
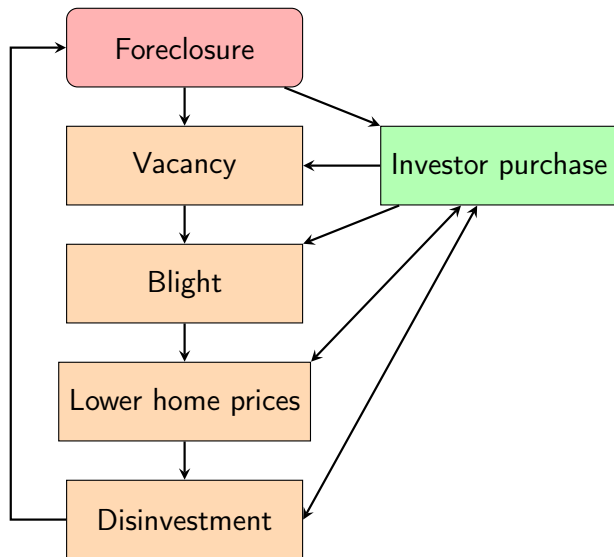


Figure 1: Completed Mortgage Foreclosures, Detroit 2005–2013. Sources: CoreLogic, Wayne County Register of Deeds

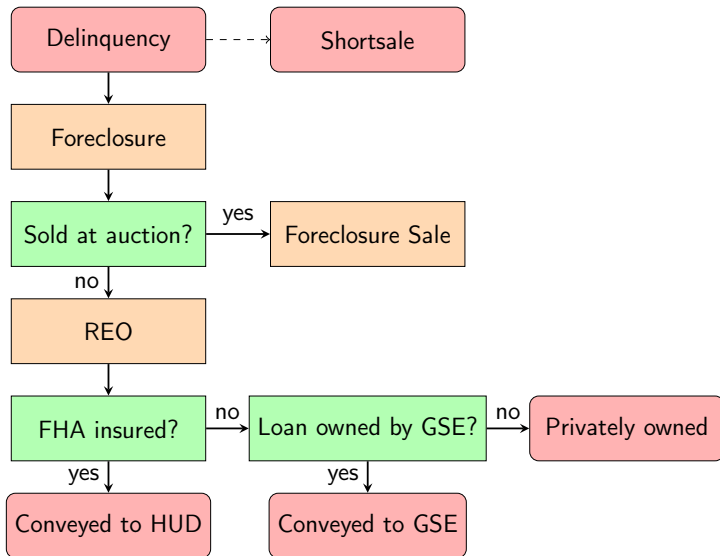
Theory of Change (simplified)



The Role of Investors



Foreclosure Process



- 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

```
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
```


Querying the Database

```
import sqlite3 as sql

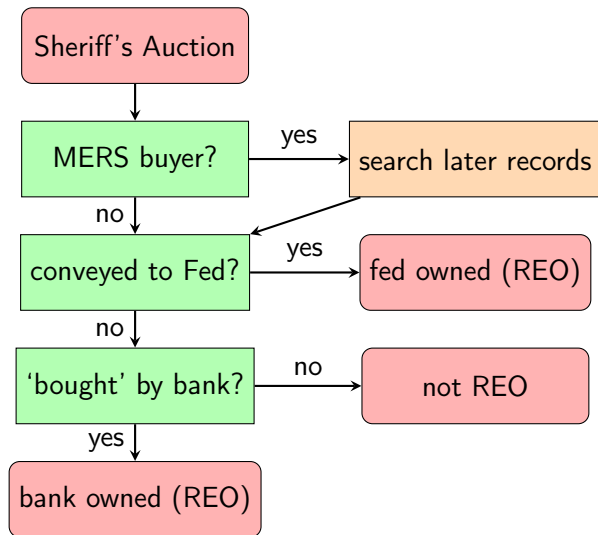
con = sql.connect("mydatabase.sqlite") # connect to db
cur = con.cursor() # create cursor object
query = '''
    SELECT * FROM records
    WHERE deed = "Sheriff's Deed";
    '''

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
```

8268 Asbury Park

Date	Amount (\$)	Deed	Seller	Buyer
2008-03-18	50,559	SHD	KOPANAKIS NICHOLAS S	MORTGAGE ELECTRONIC REGISTRATION SYSTEMS INC
2008-04-25	1	QCD	MORTGAGE ELECTRONIC REGISTRATION SYSTEMS INC NOM	FEDERAL NATIONAL MTG ASSN
2008-12-02	373	QCD	FEDERAL NATIONAL MTG ASSN, TROTT & TROTT ATTY	HOMESOLUTIONS PROPERTIES
2010-08-23	2,165	QCD	HOMESOLUTIONS PROPERTIES LLC	HARRIS RICKY M
2011-08-15	3,045	QCD	TOUVAK ENTERPRISES LLC	KOREM FIVE LLC
2012-09-12		JOF	KOPANAKIS NICHOLAS S	WAYNE COUNTY TREASURER
2012-12-06	0	QCD	WAYNE COUNTY TREASURER	WS BUSINESS SOLUTIONS, INC

Finding REO Owners



Finding Federal Owners

Regular expressions (regex) in 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
```

Decision Tree (simplified)

```
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: ≥ 50
- Medium investors: ≥ 10 and < 50
- Small investors: ≥ 2 and < 10

```
# iterate over Python dictionary w/ buyer as k(ey)
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"
```

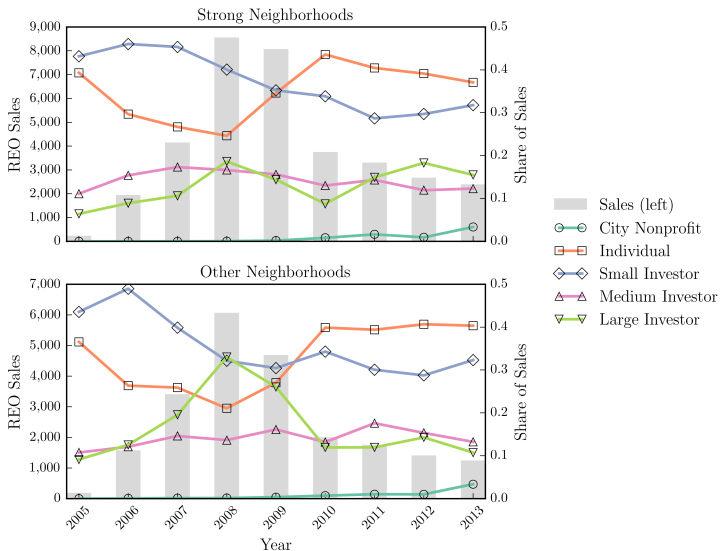
```
import pandas as pd
import sqlite3 as sql

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

Owner	Foreclosures	Share of Total Foreclosures
Private Entities	41,460	59.70%
Fannie Mae	12,756	18.37%
HUD	9,981	14.37%
Freddie Mac	2,822	4.06%
Likely Investors	1,663	2.39%
VA	443	0.64%
Likely Individuals	248	0.36%
City and Nonprofit Entities	70	0.10%
Total	69,443	100.0%

REO Sales by Buyer Type



REO Sales by Owner

	Fannie Mae	Freddie Mac	HUD	Private
Individuals	36.6%	39.6%	51.3%	23.7%
Large Investors	15.7%	7.0%	5.8%	20.1%
Medium Investors	12.0%	12.7%	9.7%	16.9%
Small Investors	34.0%	40.0%	32.8%	38.3%
City and Nonprofits	1.7%	0.7%	0.4%	0.9%
Total	100.0%	100.0%	100.0%	100.0%

Tax Foreclosure by REO Buyer Type, 2008–2015

	Properties	Share
Prior to Second Sale		
Individuals	4,179	22.9%
Large Investors	3,403	35.1%
Medium Investors	2,840	33.0%
Small Investors	6,944	32.2%
City and Nonprofits	100	17.0%
Anytime after Purchase		
Individuals	4,745	26.1%
Large Investors	6,558	67.7%
Medium Investors	4,490	52.1%
Small Investors	8,901	41.3%
City and Nonprofits	117	19.9%