

# Investing Locally: 5 ZIP Codes in the Sacramento Region

Gina Durante

Flatiron School Data Science Boot Camp

Module 4 Project



# Agenda

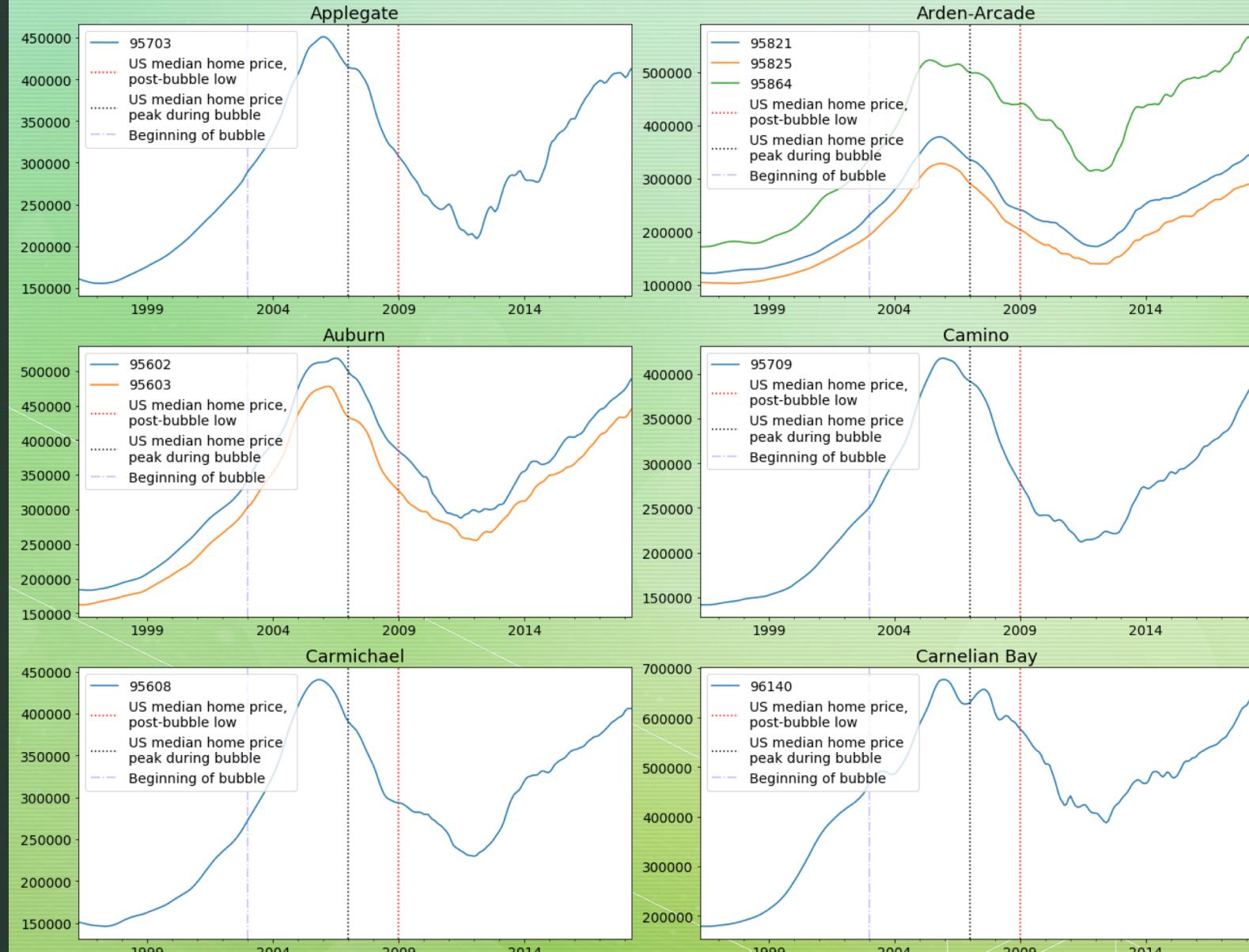
- Project Overview and Methodology
  - Evaluate ZIP Codes for potential investment opportunities
  - Methodology to analyze multiple ZIP codes and select 5
- Findings
- Recommendations
- Potential future work
- Q&A and acknowledgements

# Project Overview

- Evaluate 5 best ZIP codes for investment opportunity
  - Data source: Zillow monthly housing values by ZIP in US
- Methodology
  - What is meant by “best”
  - Data exploration and visualization
  - Selected 20 ZIP codes for further analysis (“semi-finalists”)
  - Time series modeling
  - Forecasting 24 months

# Creating multiple plots for certain geographies

(At right: example subset of Sacramento metro area home values by ZIP code and city)



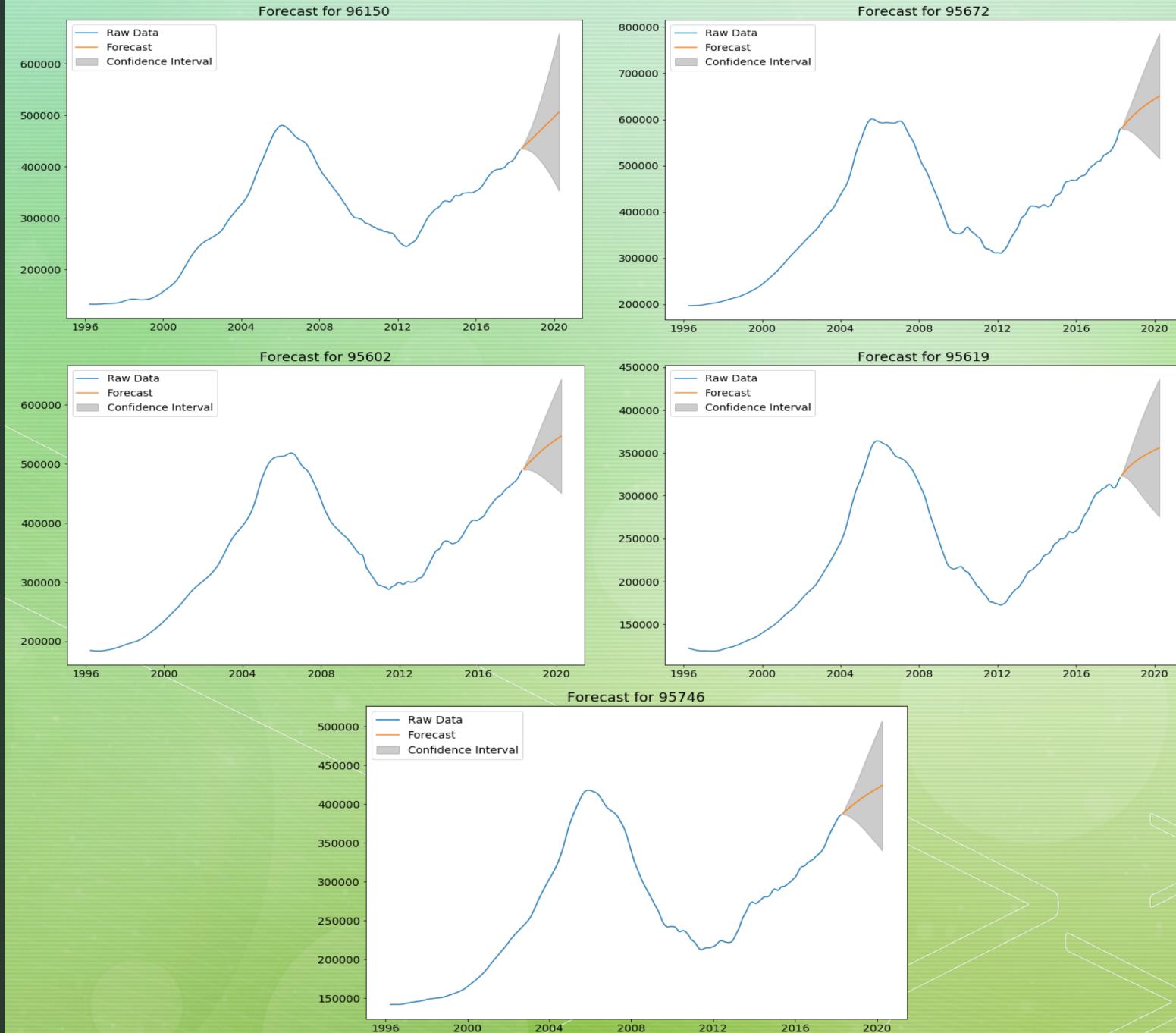
# Findings

- El Dorado and Placer county ZIPs seem to offer better investment opportunities
- Recommending a mix of populations and locations
  - Around Lake Tahoe
  - East of Sacramento in Placer County (foothills of the Sierra Nevada)

ZIP code	2018 value	City	Pop	County	Investment rating, based on predicted return	Predicted % Change	Worst Case % Change	Best Case % Change
96150	432500.0	South Lake Tahoe	30000	El Dorado	excellent	16.93	-18.47	52.33
95672	579300.0	Rescue	4592	El Dorado	excellent	12.30	-11.10	35.71
95602	488700.0	Auburn	18290	Placer	excellent	11.94	-7.86	31.73
95619	321100.0	Diamond Springs	4359	El Dorado	excellent	10.80	-14.27	35.86
95746	386000.0	Granite Bay	22482	Placer	good	9.80	-11.96	31.55
95623	455700.0	El Dorado	3986	El Dorado	good	9.53	-36.53	55.59
95650	631000.0	Loomis	12600	Placer	good	6.59	-14.19	27.37
96140	644600.0	Carnelian Bay	1170	Placer	good	6.55	-13.04	26.14
95709	386000.0	Camino	4354	El Dorado	good	6.37	-16.59	29.33
95616	692300.0	Davis	45500	Yolo	mediocre	3.69	-10.50	17.89
95663	600700.0	Penryn	2468	Placer	mediocre	2.03	-21.47	25.53
95765	509200.0	Rocklin	41810	Placer	mediocre	1.73	-19.44	22.91
95811	567500.0	Sacramento_DosRios	7630	Sacramento	poor	0.55	-19.01	20.10
95818	563900.0	Sacramento_LandPark	21825	Sacramento	poor	-0.01	-17.61	17.59
95747	483800.0	Roseville	72437	Placer	poor	-1.30	-20.48	17.89
95614	423300.0	Cool	3882	El Dorado	poor	-3.95	-28.06	20.15
95636	241200.0	Somerset	1000	El Dorado	poor	-4.60	-31.27	22.06
95630	542900.0	Folsom	74111	Sacramento	poor	-10.82	-29.63	7.98
95831	449300.0	Sacramento_Pocket	42952	Sacramento	poor	-13.87	-34.87	7.14
95864	552700.0	Arden-Arcade	92186	Sacramento	poor	-18.75	-40.73	3.23

## *Recommendations for investment:*

- 96150 (S. Lake Tahoe)
- 95672 (Rescue)
- 95602 (Auburn)
- 95619 (Diamond Springs)
- 95746 (Granite Bay)





# Possible Future Work

## What

- Evaluating other metro areas, e.g., Dallas, Pittsburgh
- Back testing
- Analyzing shorter time frames (e.g., 2013-2018)
- Scaling investment by population size
- Construct a hypothetical investment basket
- Look at size of standard deviations

## Why?

- Other metro areas might have greater upside
- Predict 2016-2018 values?
- More accurate if a shorter time frame were used?
- Larger populations in some ZIP codes
- Show potential returns from investment strategy
- Indicator of volatility/risk

Q&A



*Thank you for your time!*



# Appendix

# Box plots and violin plots

## City of Sacramento

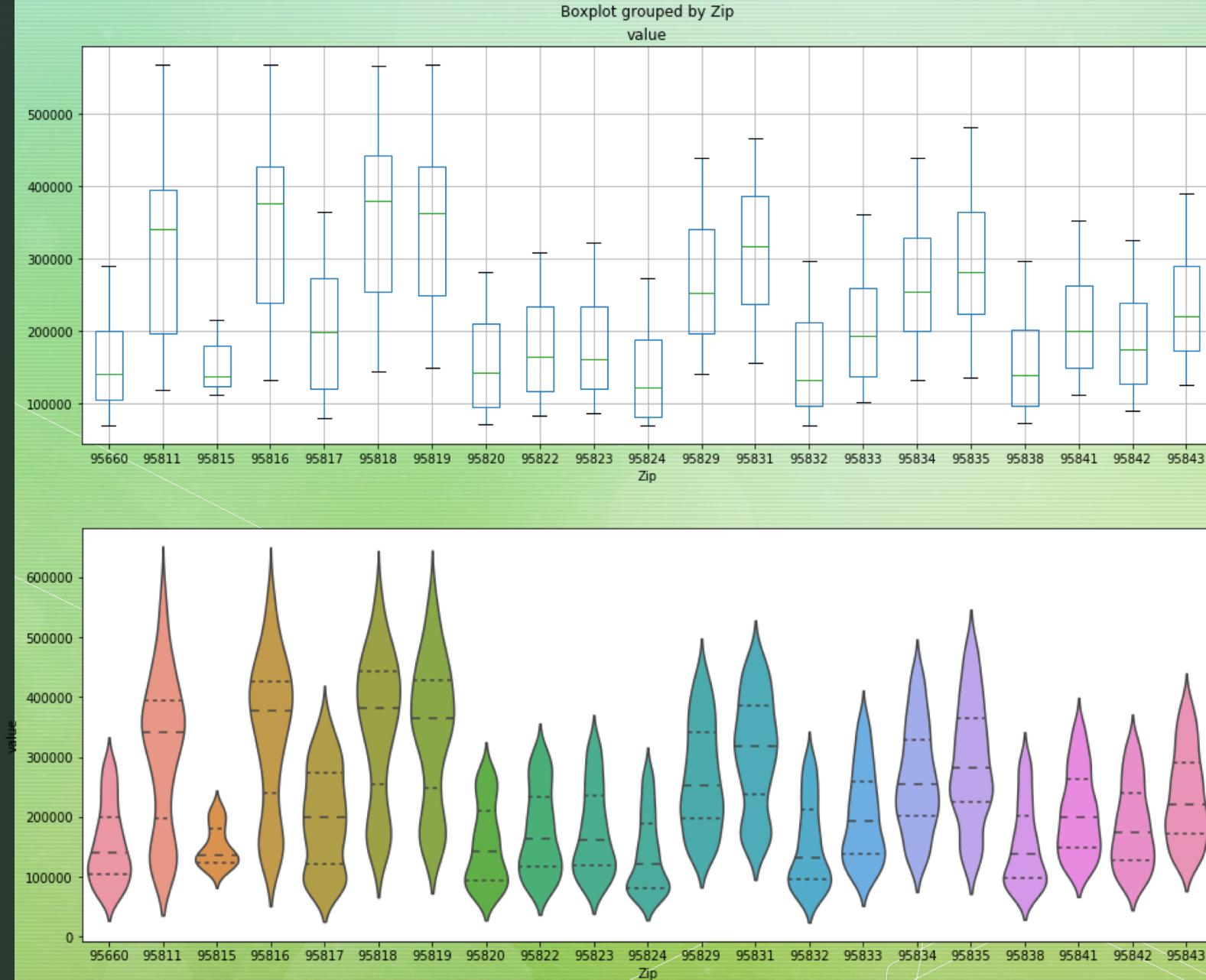


Plots show spread of values. Spread creates confidence intervals.

Plots also show the distributions of values around the median. One can see how prices move during various time frames of the housing market bubble build-up, collapse, and recovery.

# *Boxplots and violin plots show spread and distribution of home prices*

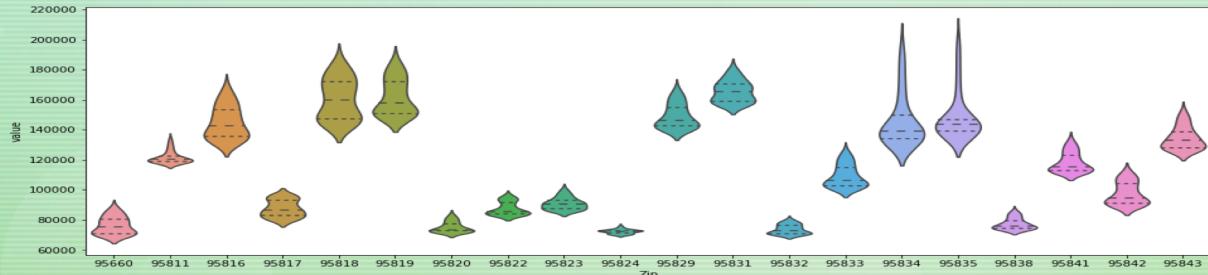
- ZIP codes in the City of Sacramento
  - Shown by boxplot and violin plot
  - Includes all values from 1996 to 2018
- Spread is correlated to forecast confidence intervals
  - Greater spread implies more uncertainty in forecasts



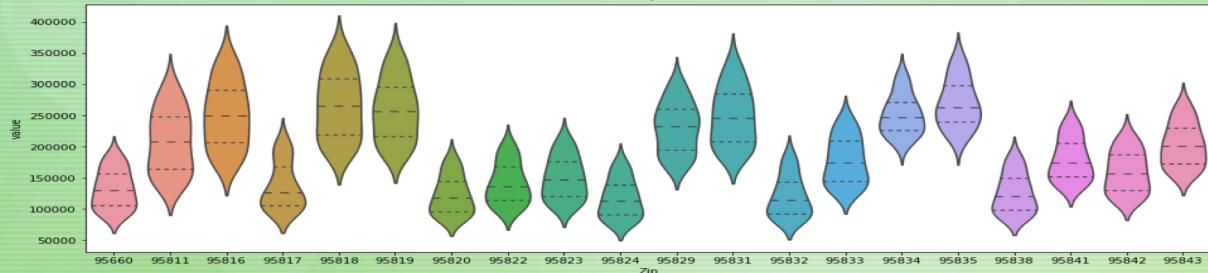
# *Violin plots of ZIP codes in the City of Sacramento during bubble stages*

- Plot shapes reveal:
  - *Spread of values in each ZIP code*
  - *Distribution* during each time frame
- Observe the progression of the housing market bubble:
  - Build-up (2003-2006)
  - Crash (2007-2012)
  - Gradual recovery (2013-2018)

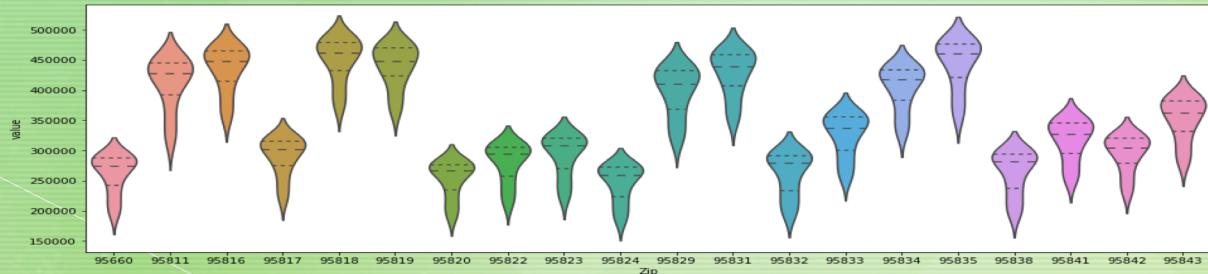
1996 – 1999  
(housing value distribution skewed low)



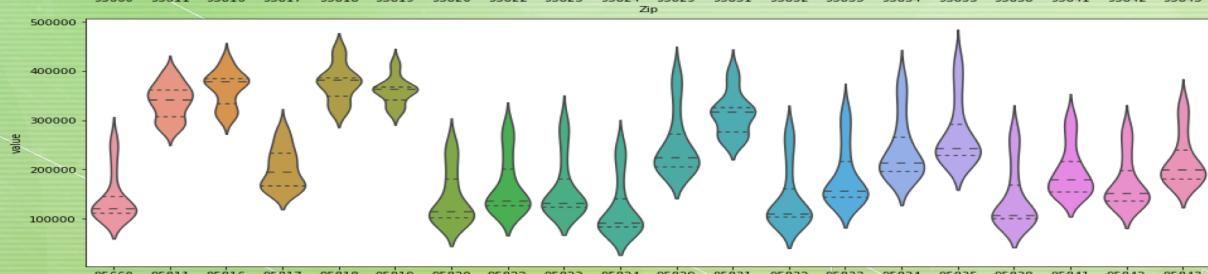
1999 – 2003  
(housing value distribution balanced)



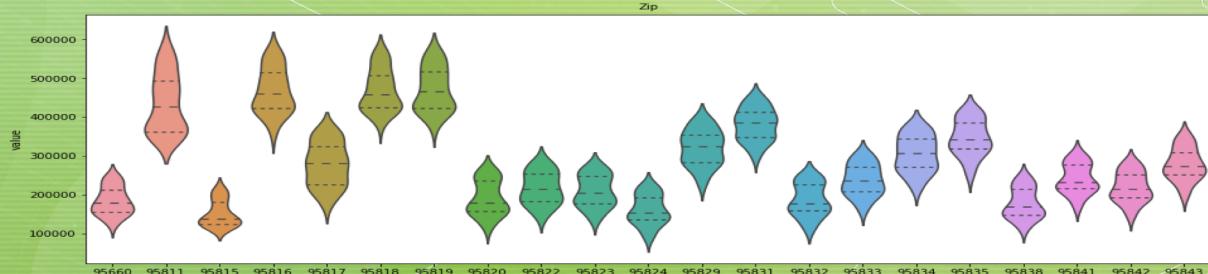
2003 – 2006  
(housing value distribution skewed high)



2007 – 2012  
(housing value distribution skewed low)



2013 – 2018  
(housing value distribution becoming more balanced)



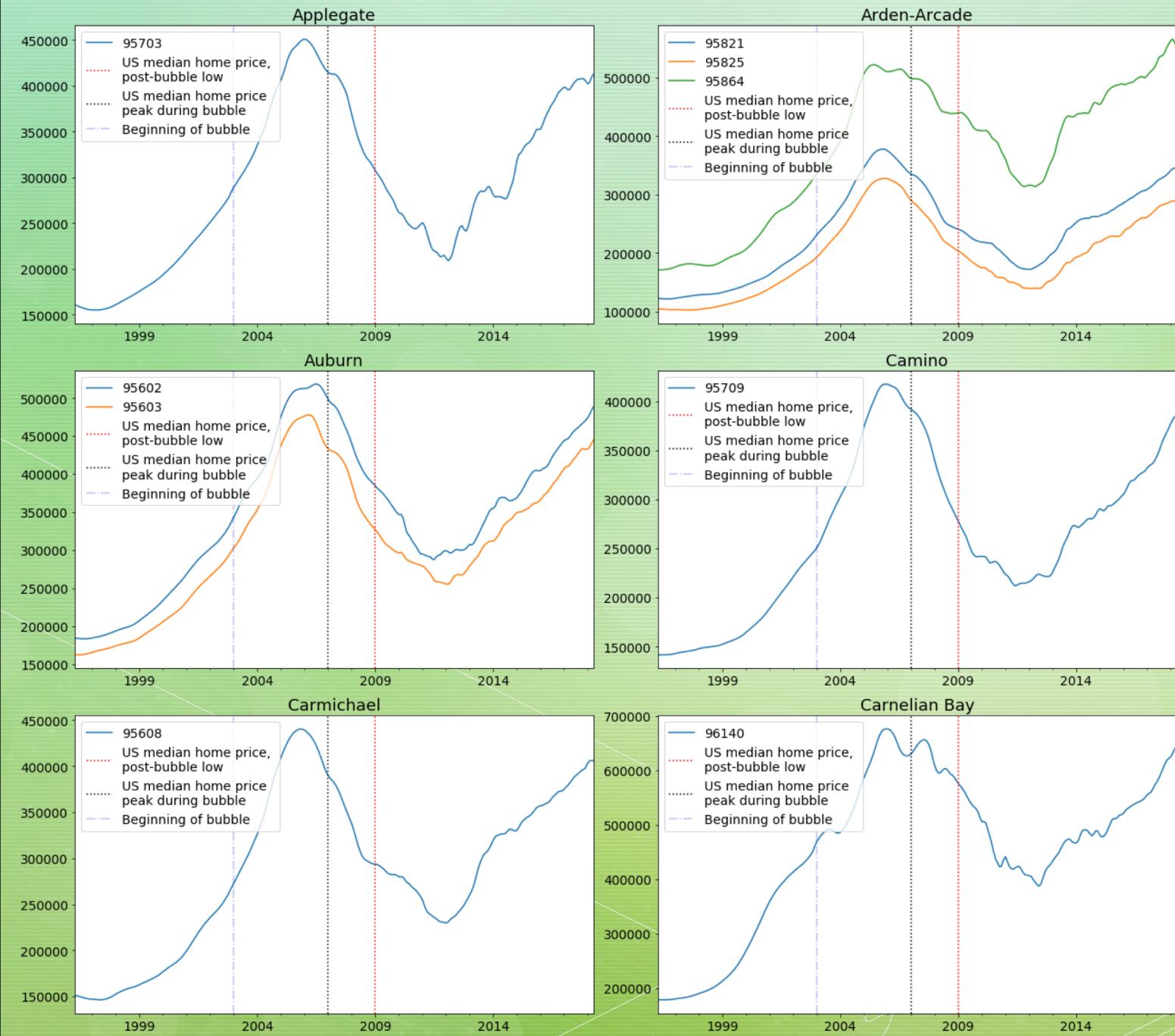
## Plots of historical values by city: Sacramento metro region

- ▼ Looking at ZIP code plots shows patterns that can be used to inform ZIP code selection.

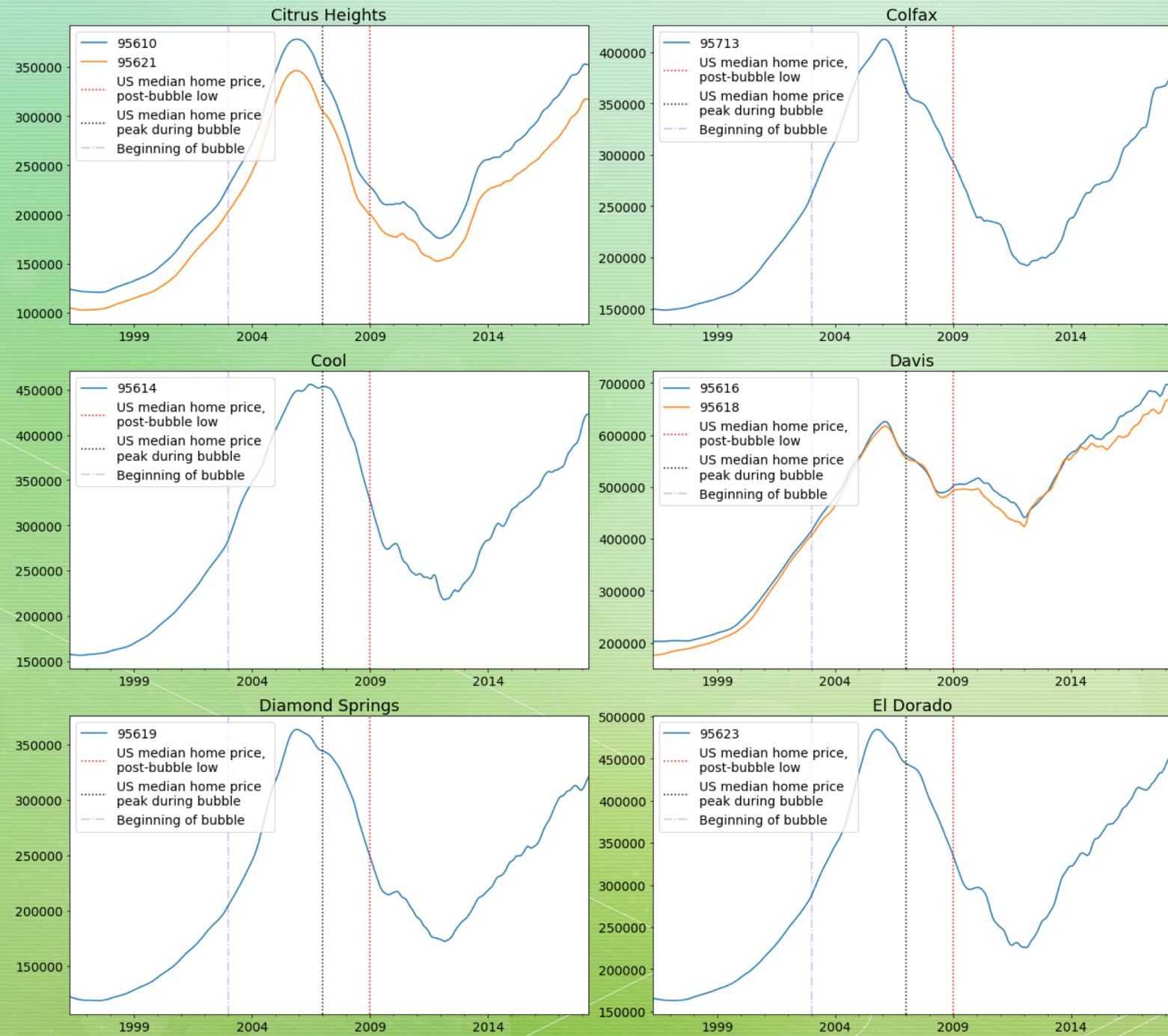
Several pattern elements can be identified, including:

- Volatility (how large was run-up and subsequent drop of housing values)
- Remaining upside potential after market recovery
- Momentum (likelihood of increased or decreased values based on shape of curve in the last time periods for which data is available)

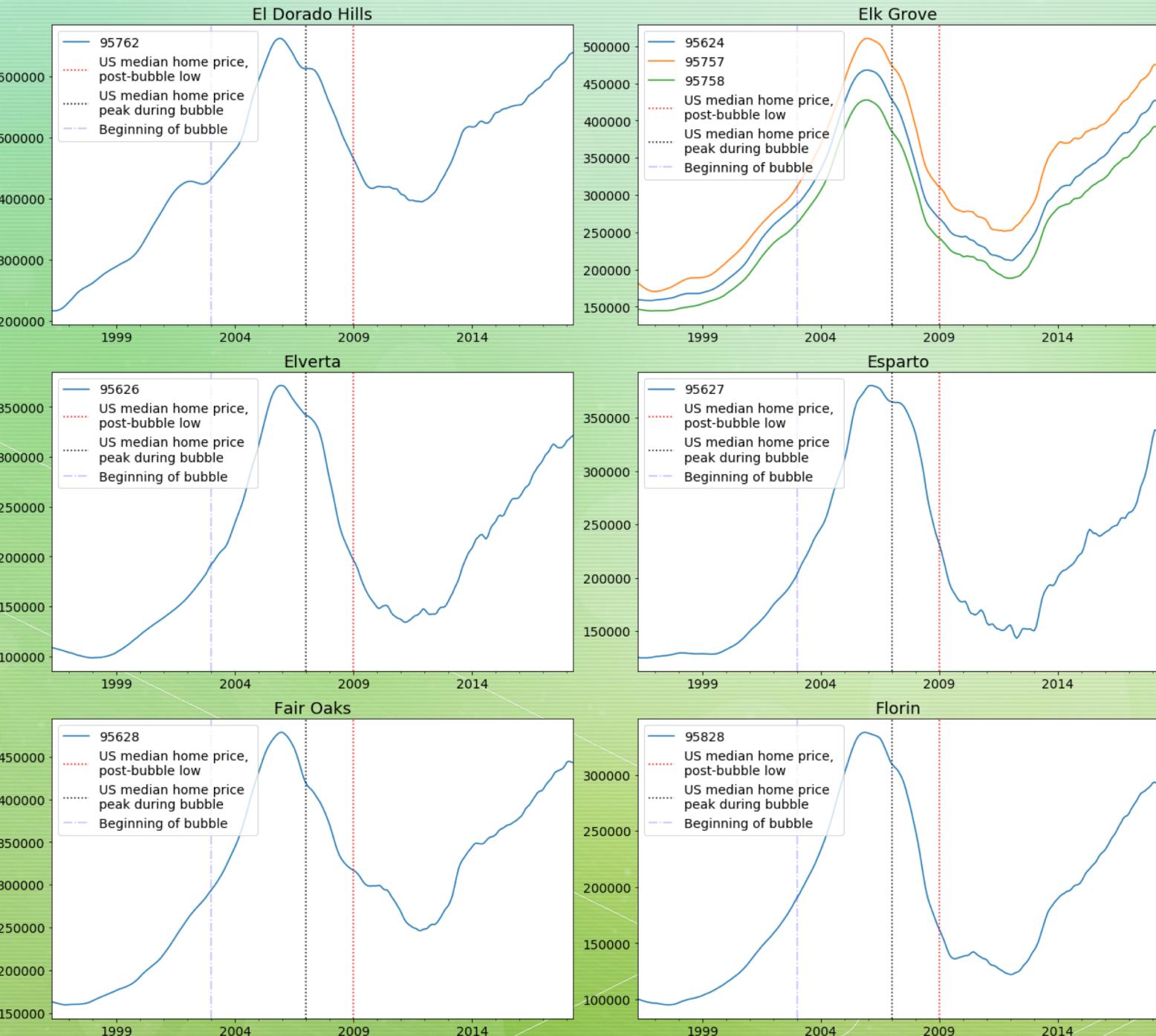
*Sacramento Metro Area:*  
*Plots by city*  
*(First group of six plots)*



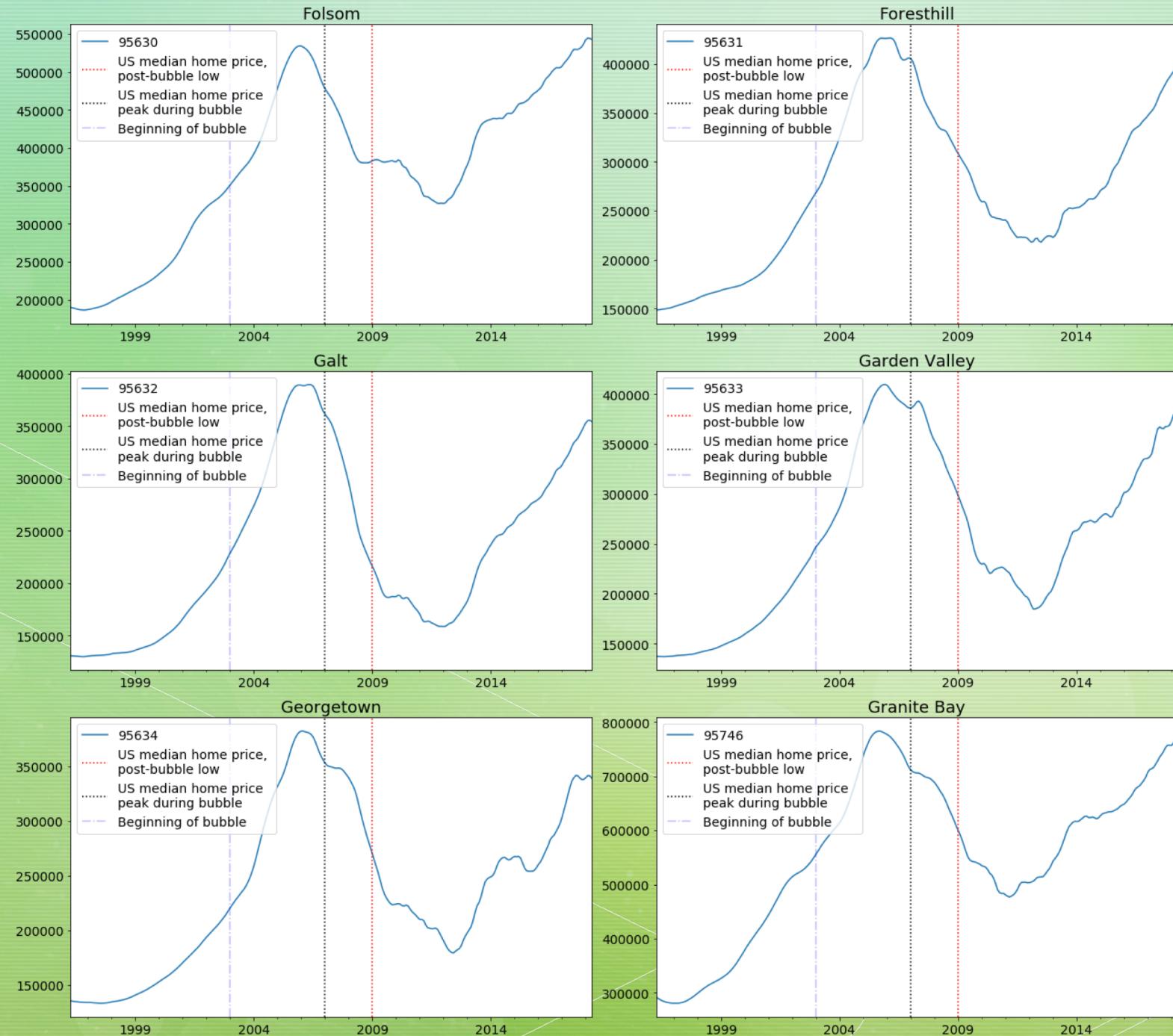
*Sacramento Metro Area:*  
*Plots by city*  
*(Second group of six plots)*



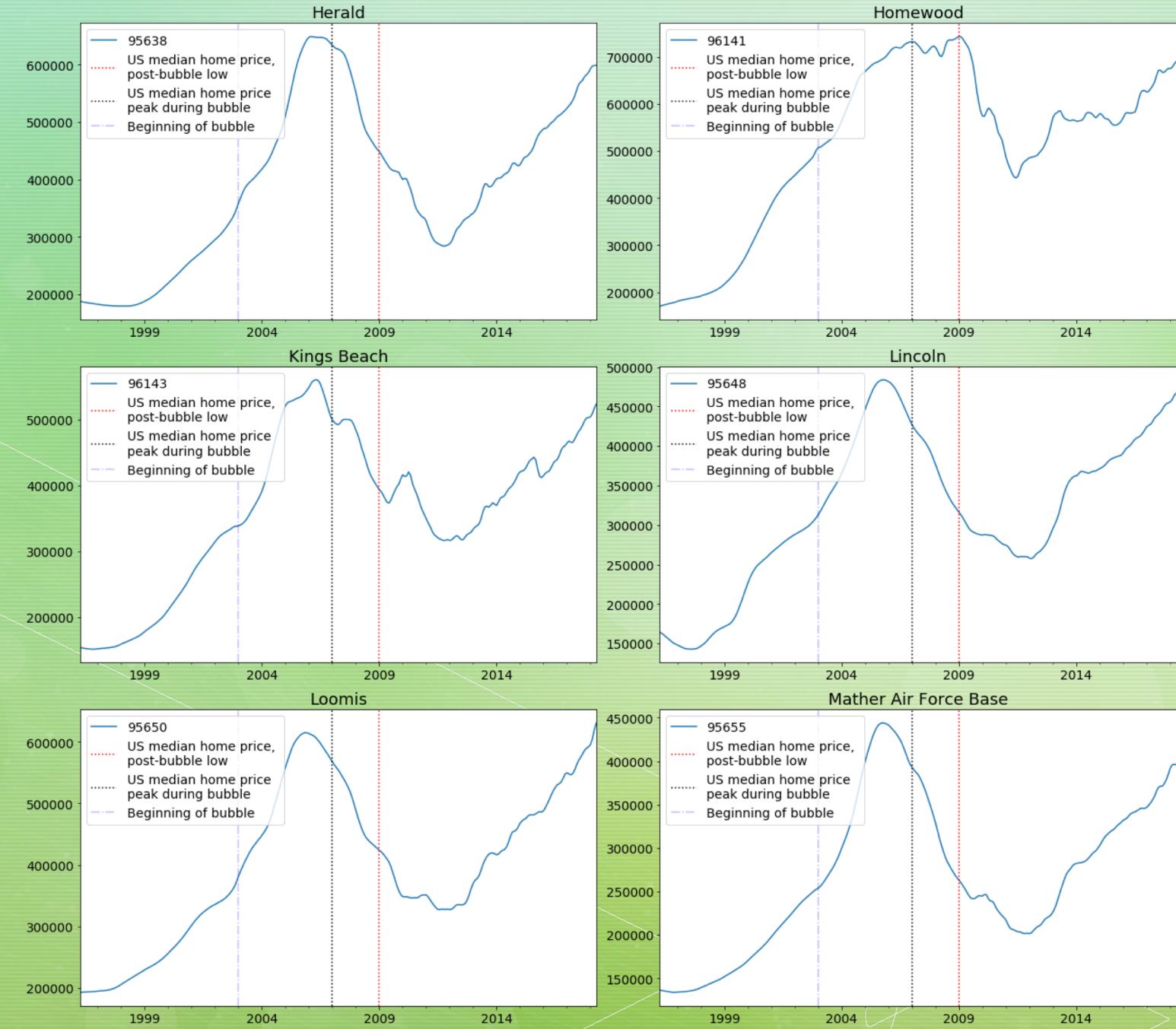
*Sacramento Metro Area:*  
*Plots by city*  
*(Third group of six plots)*



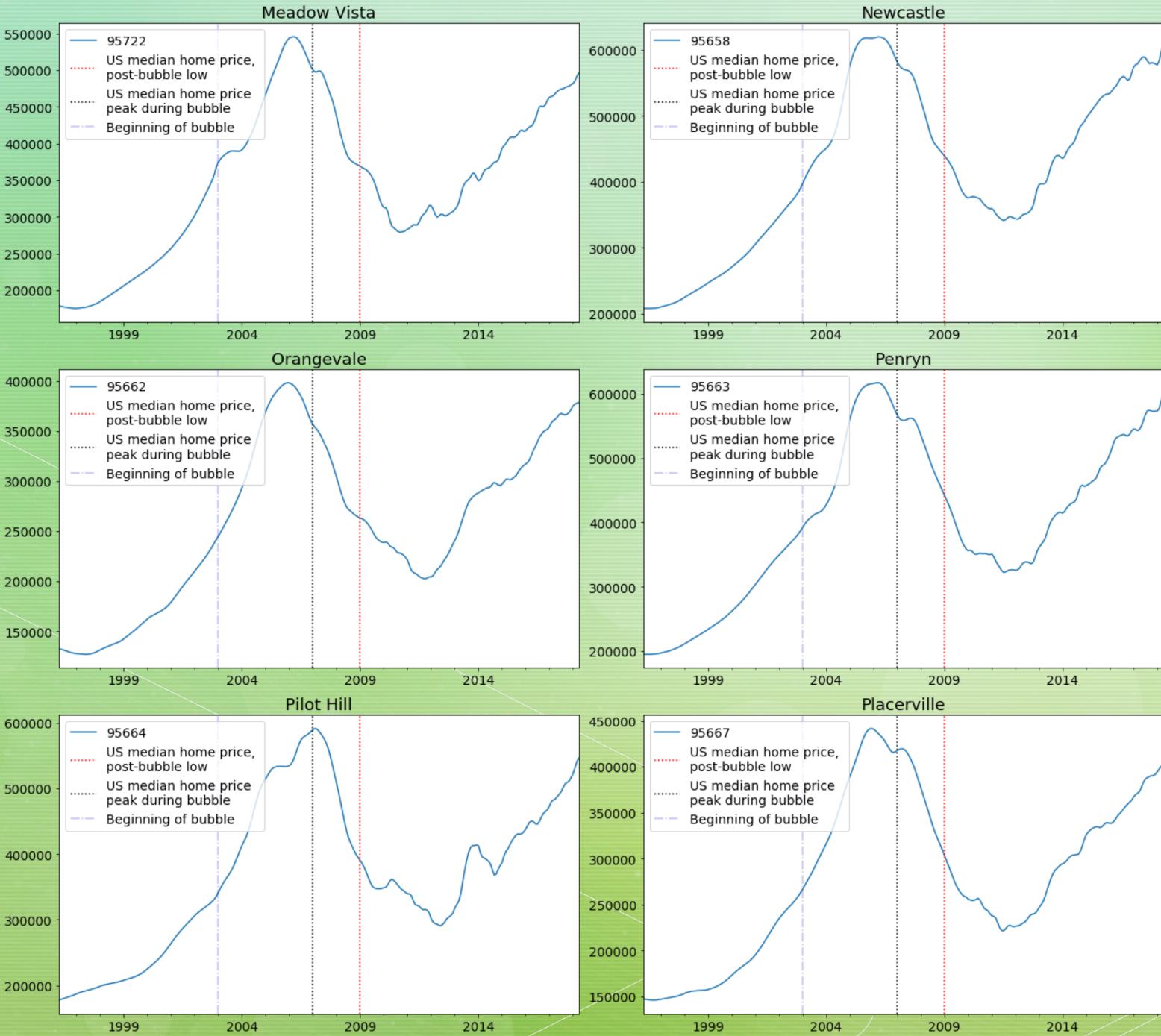
*Sacramento Metro Area:*  
*Plots by city*  
*(Fourth group of six plots)*



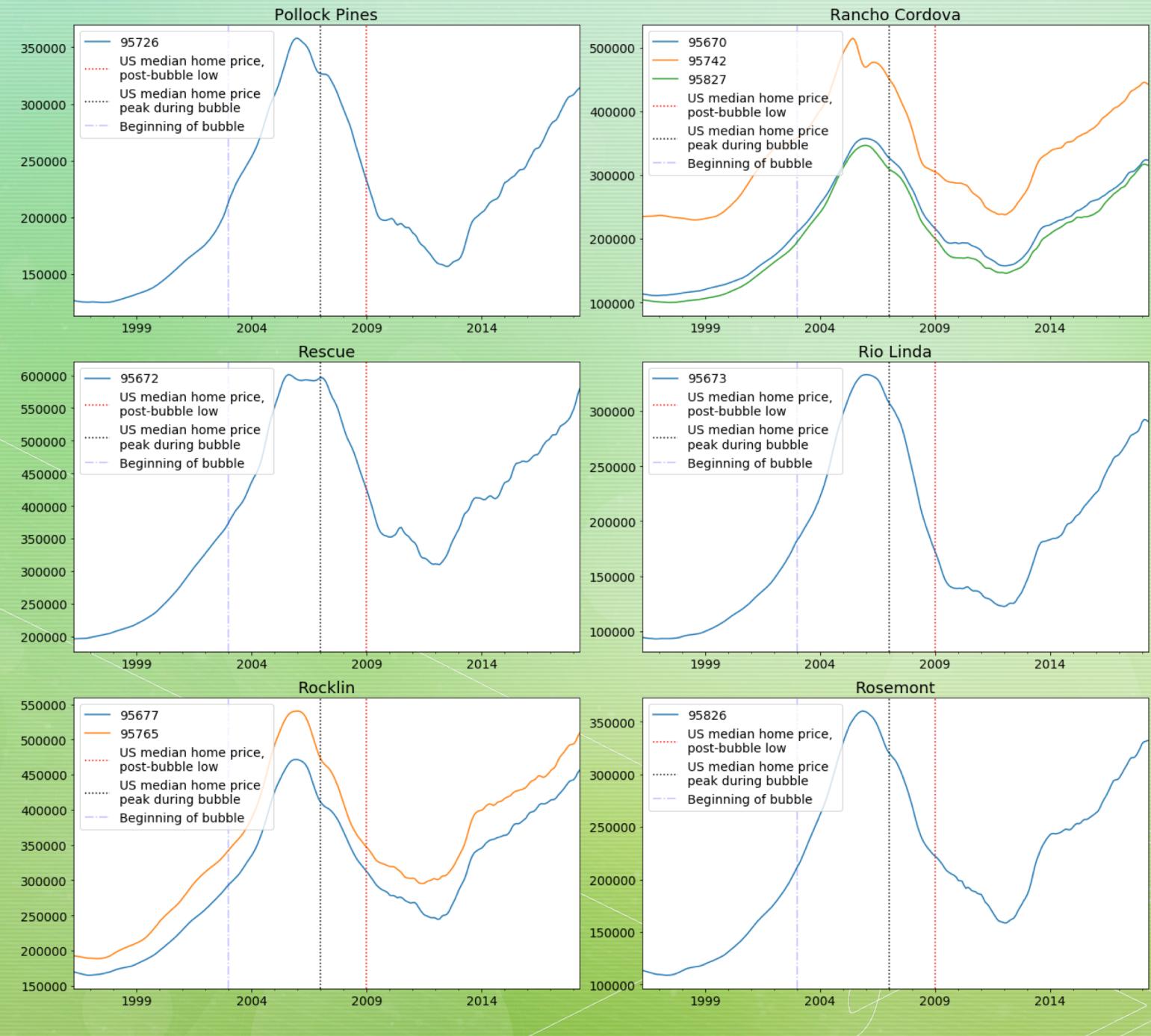
*Sacramento Metro Area:*  
*Plots by city*  
*(Fifth group of six plots)*



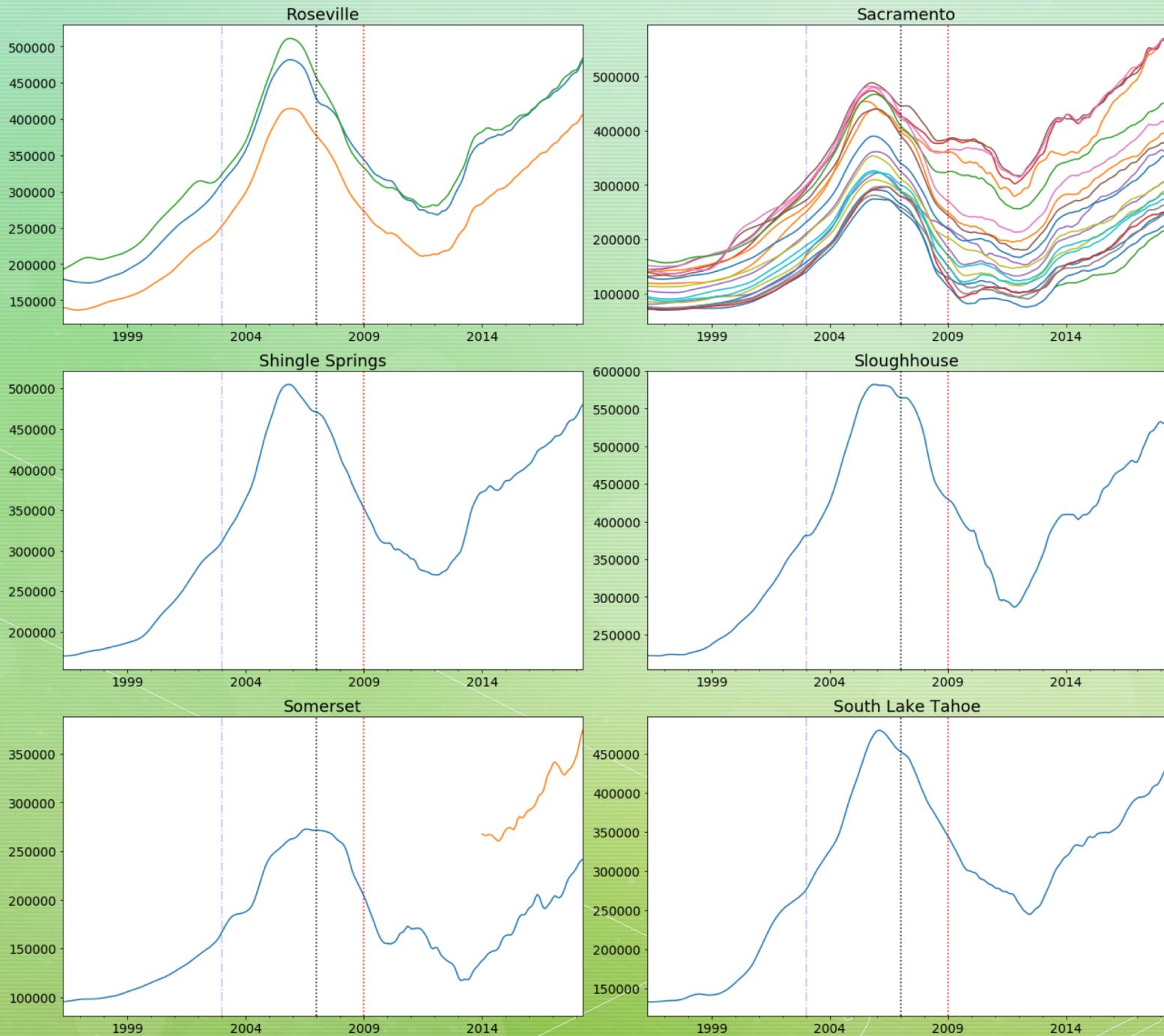
*Sacramento Metro Area:*  
*Plots by city*  
*(Sixth group of six plots)*



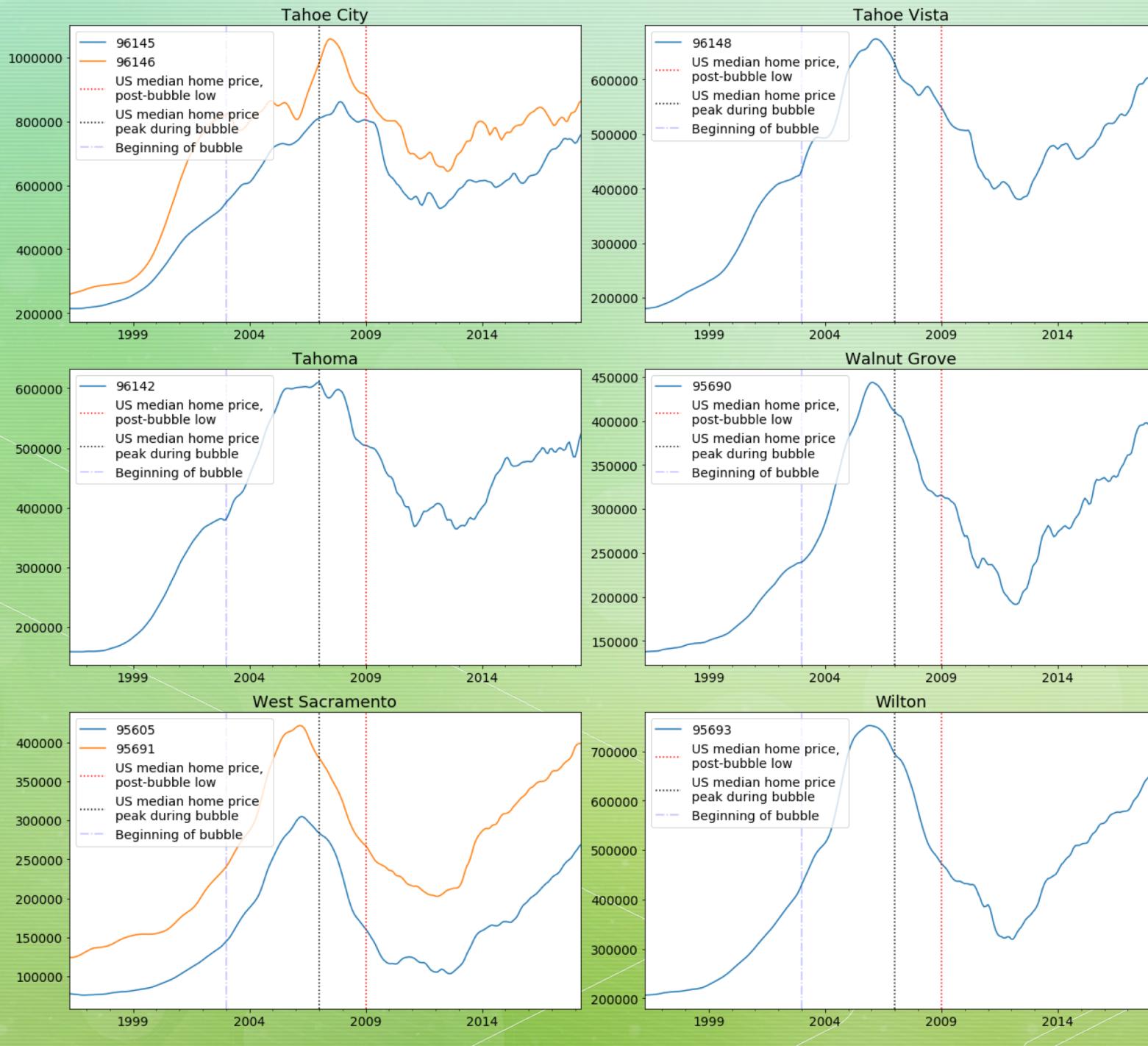
*Sacramento Metro Area:*  
*Plots by city*  
*(Seventh group of six plots)*



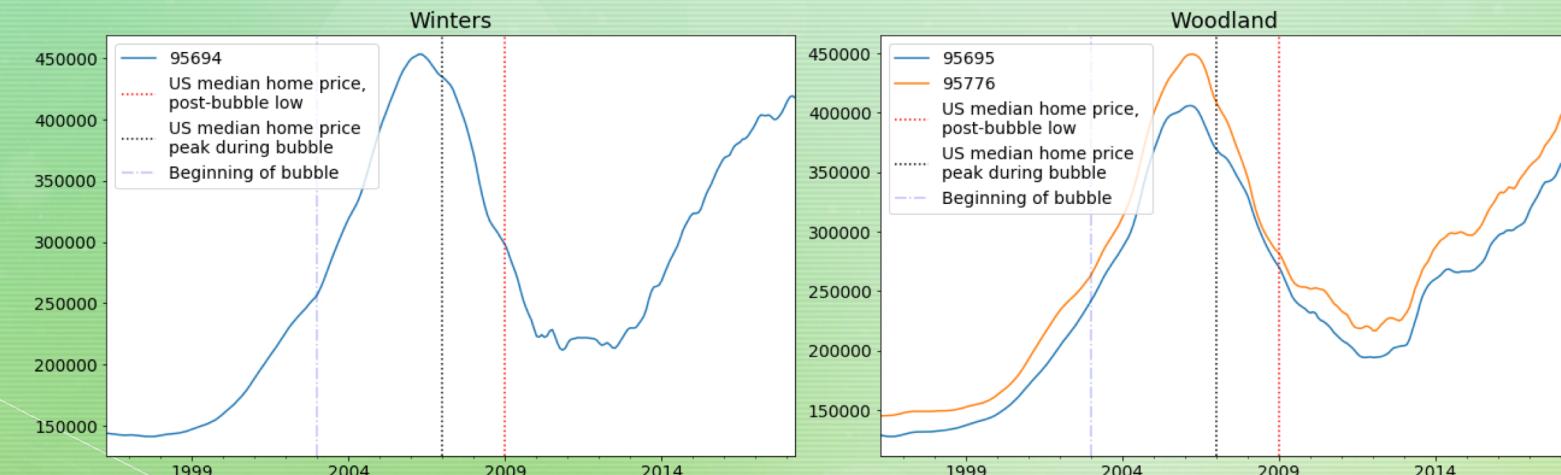
*Sacramento Metro Area:  
Plots by city  
(Eighth group of six plots)*



*Sacramento Metro Area:*  
*Plots by city*  
*(Ninth group of six plots)*



*Sacramento Metro Area:*  
*Plots by city*  
*(Tenth group—two plots)*



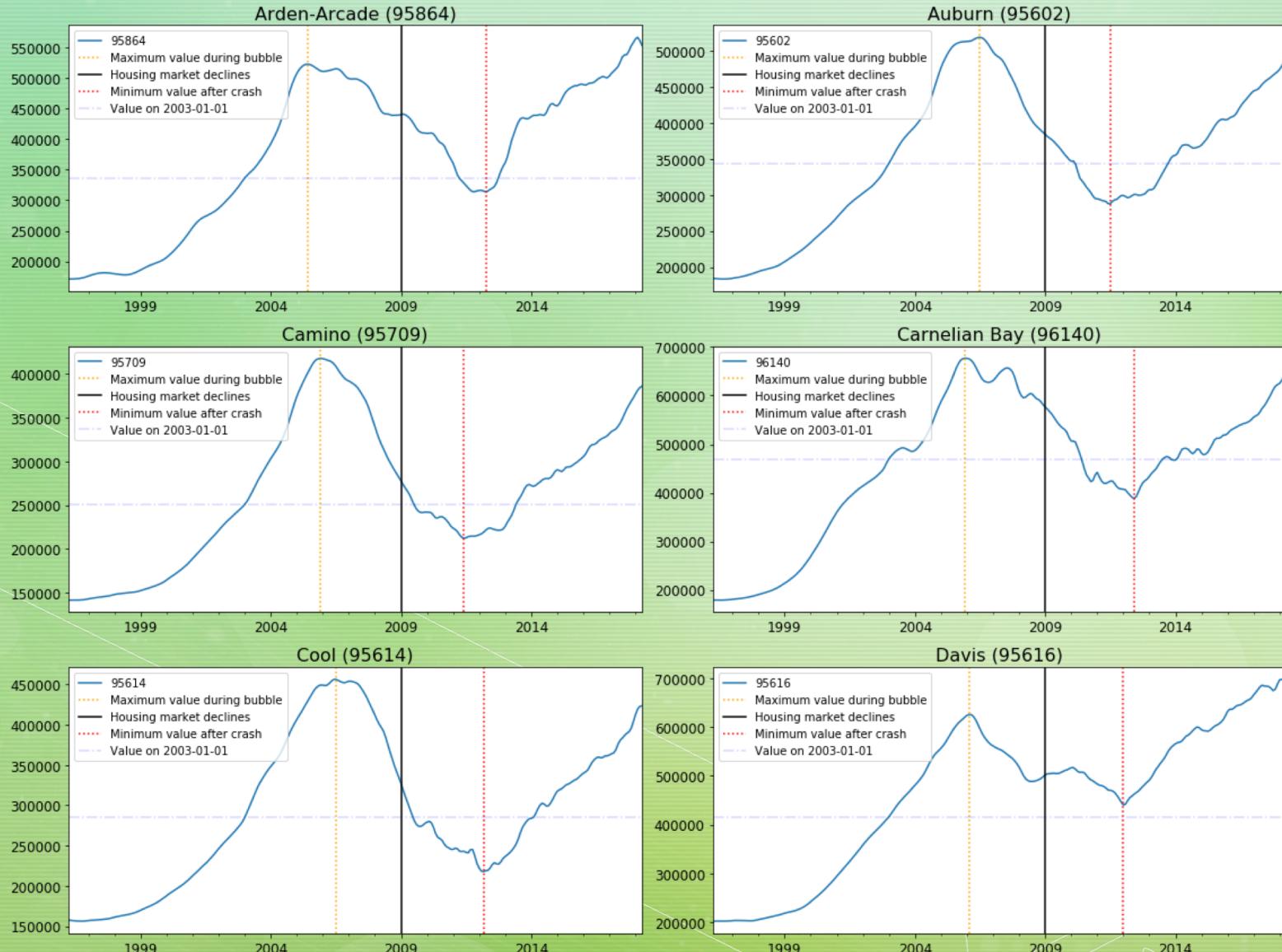
## Plots of historical values: Semi-finalist ZIP codes—Sacramento metro

- ▶ Looking at ZIP code plots shows patterns that can be used to inform ZIP code selection.

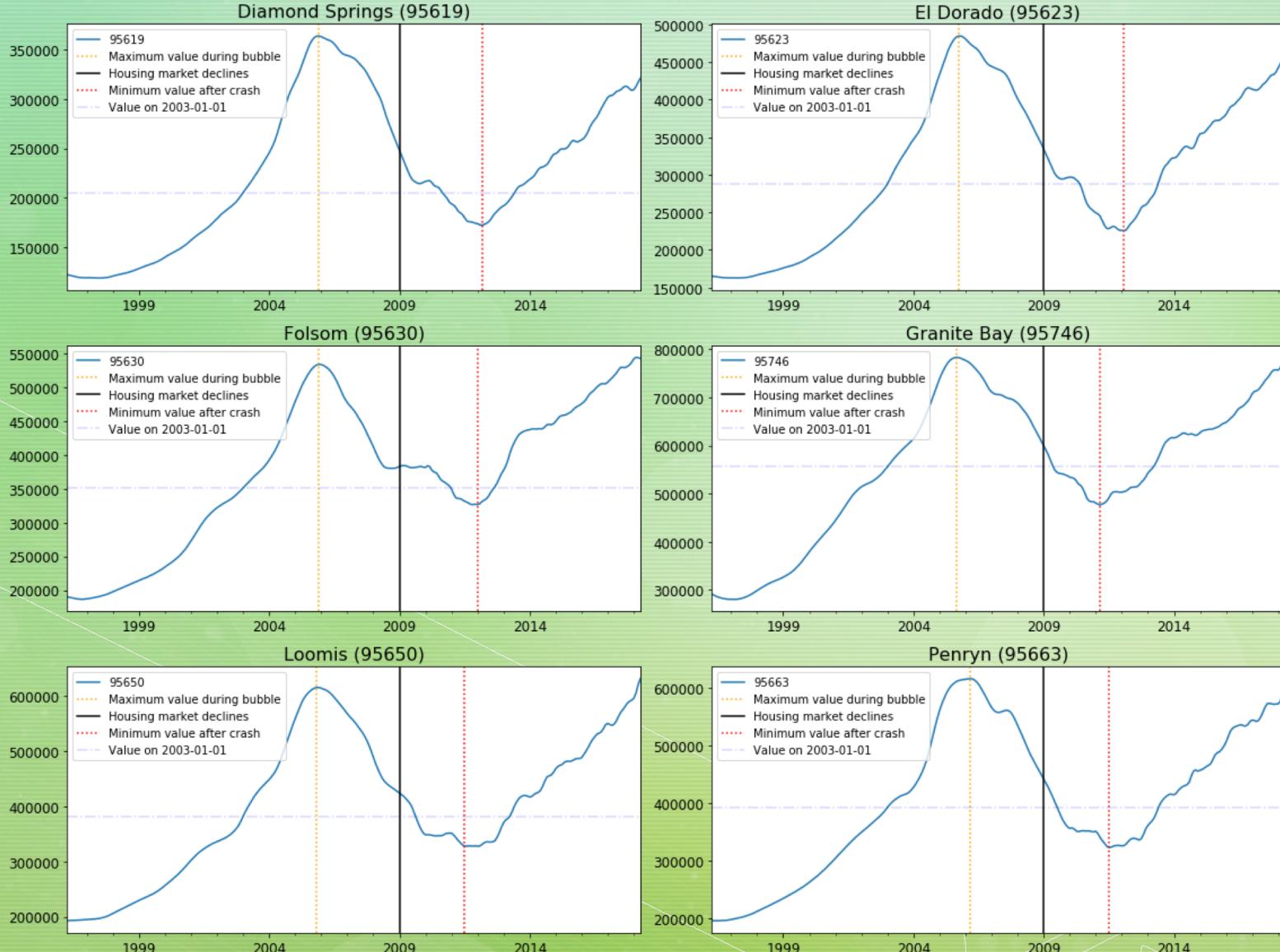
Several pattern elements can be identified, including:

- Volatility (how large was run-up and subsequent drop of housing values)
- Remaining upside potential after market recovery
- Momentum (likelihood of increased or decreased values based on shape of curve in the last time periods for which data is available)

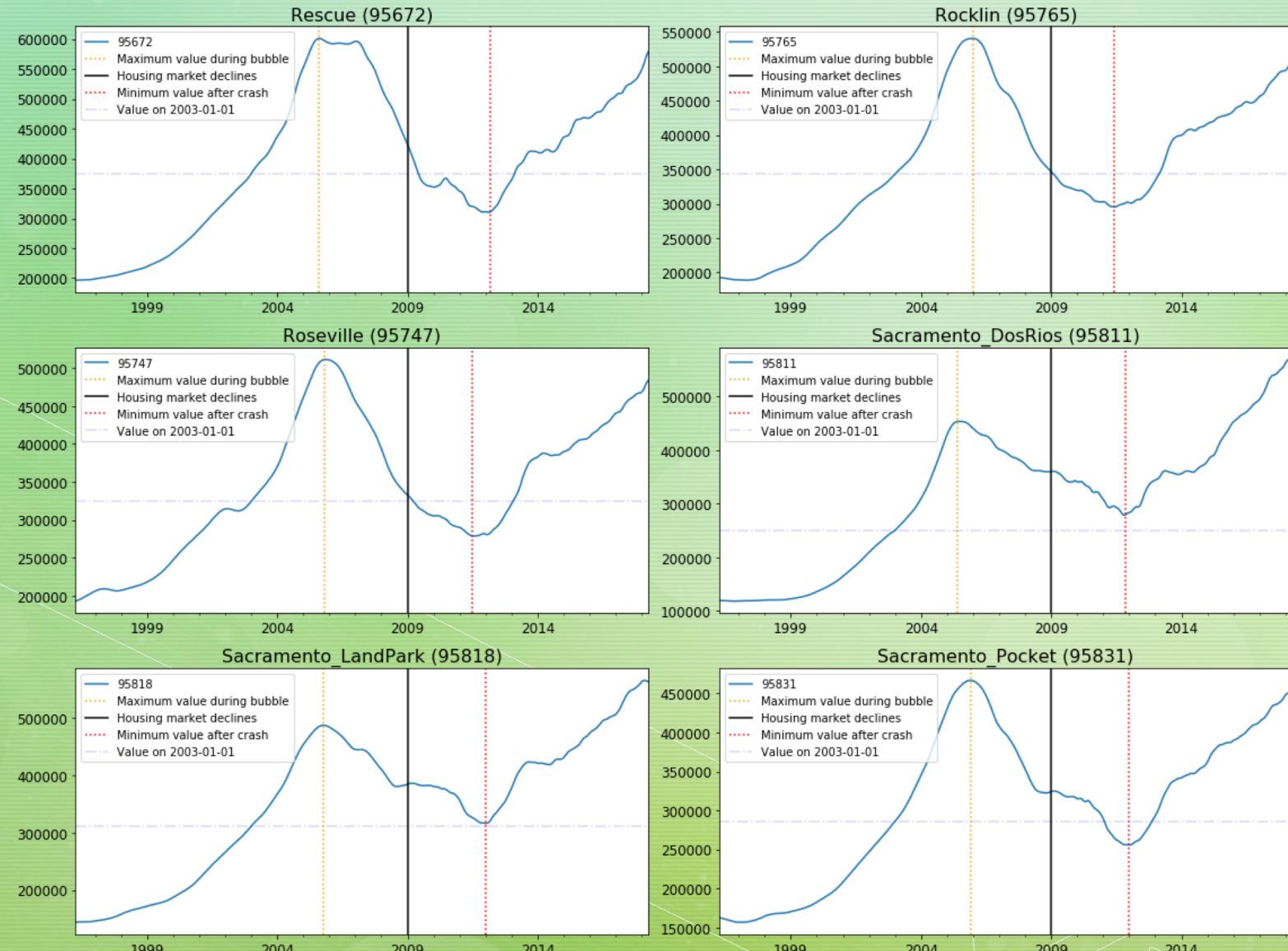
## *Semi-finalist ZIP codes: Sacramento Metro Area (First group of six)*



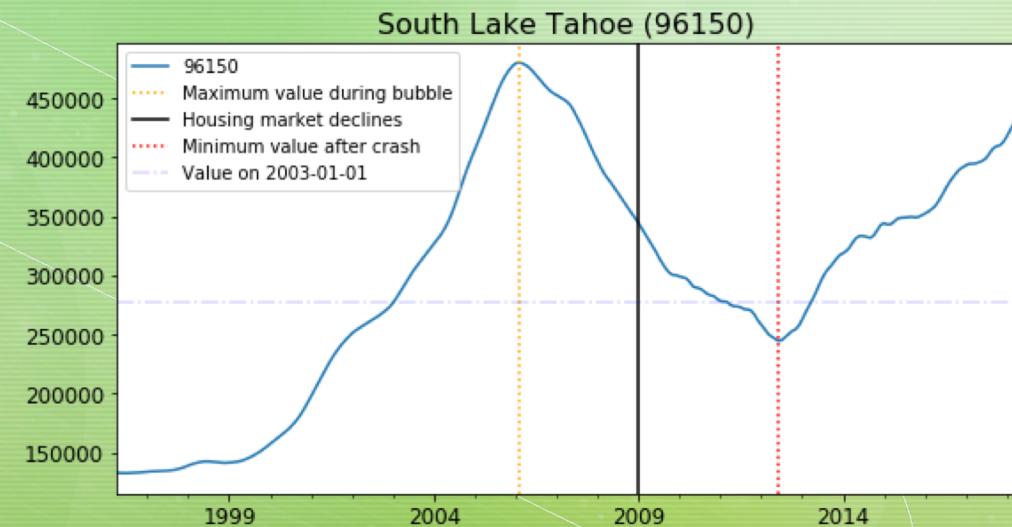
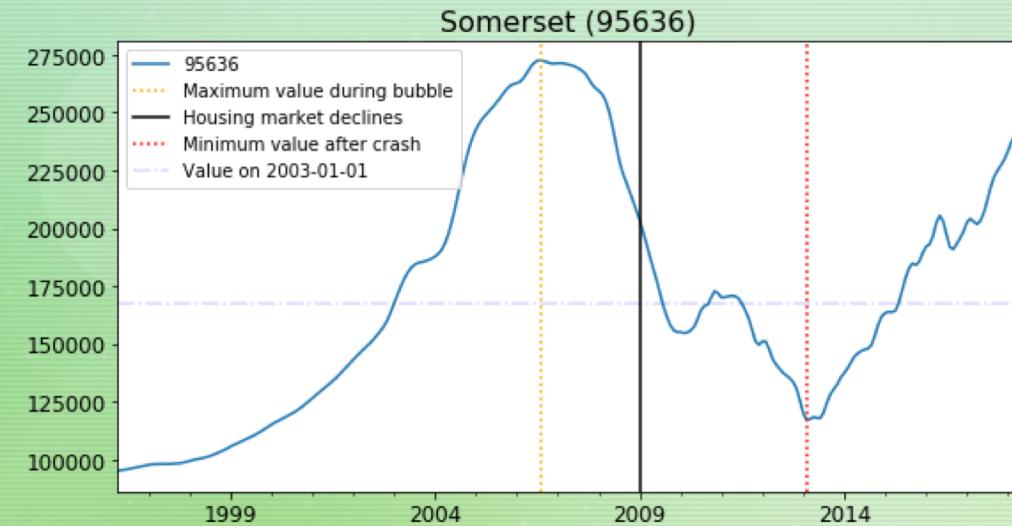
## *Semi-finalist ZIP codes: Sacramento Metro Area (Second group of six)*



*Semi-finalist ZIP codes:  
Sacramento Metro Area  
(Third group of six)*



*Semi-finalist ZIP codes:  
Sacramento Metro Area  
(Fourth group—last  
two ZIP codes)*



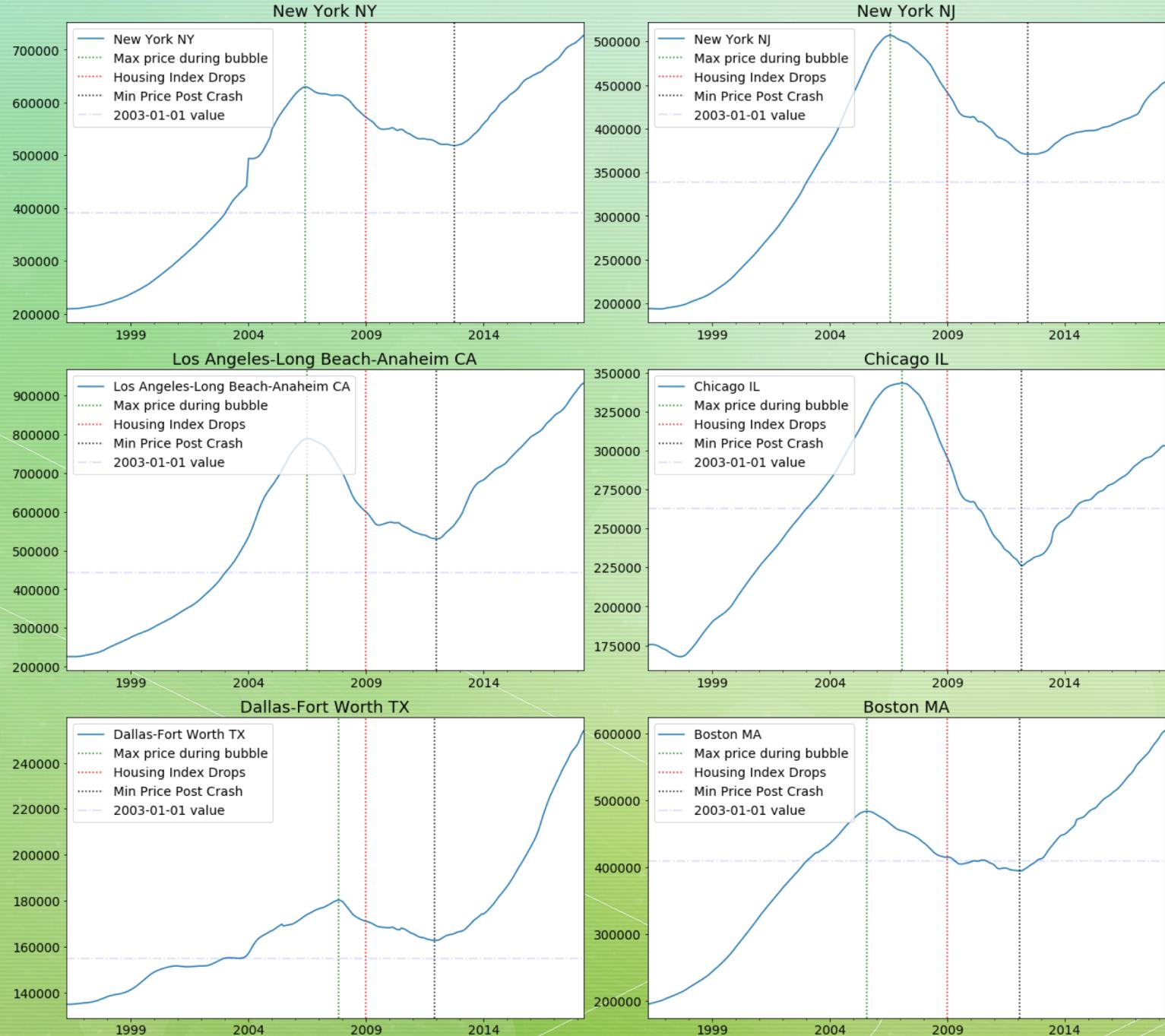
# Historical housing market values: Top 30 U.S. metro areas



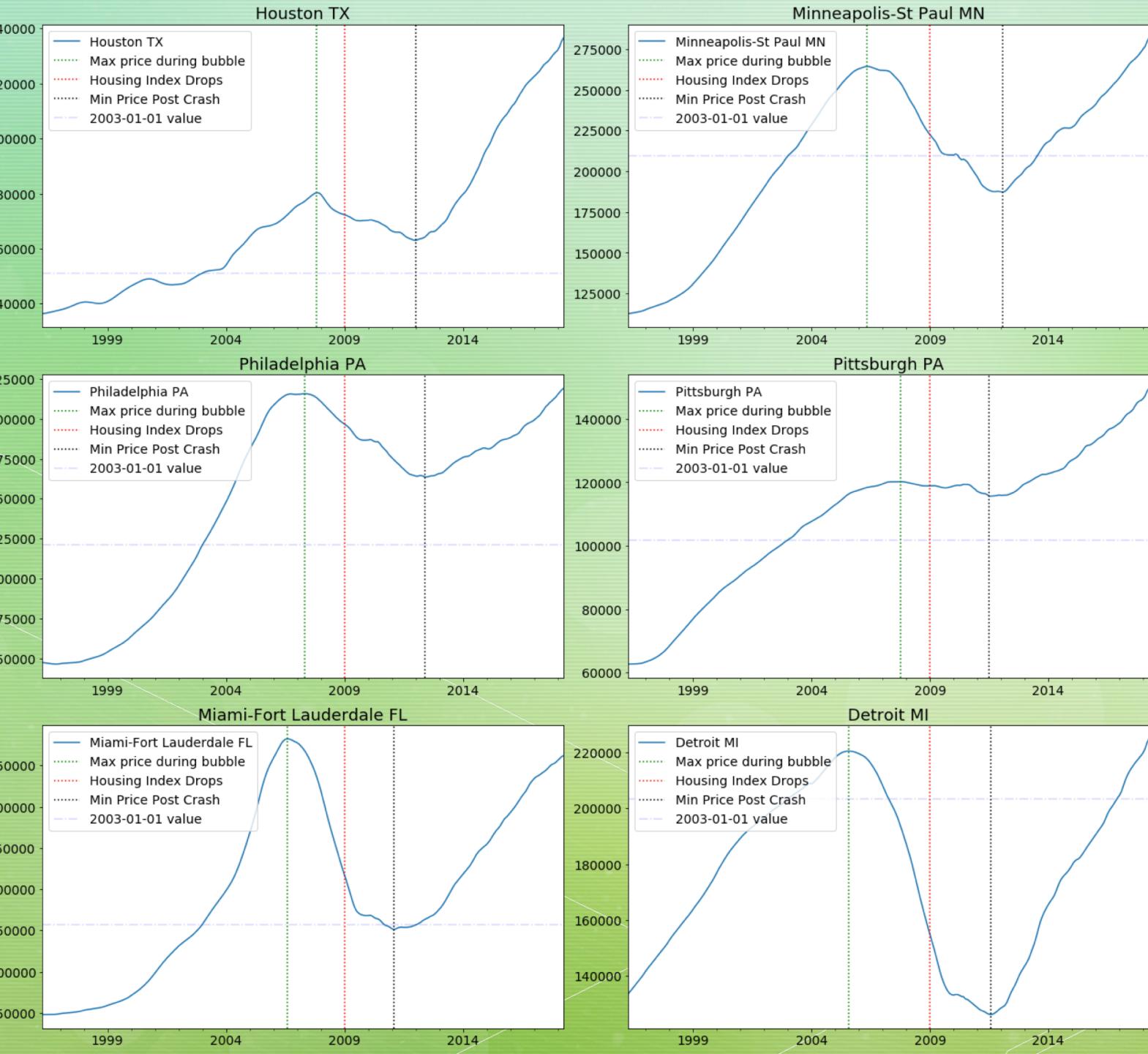
Plots aggregated at the metro area can reveal which metro areas might hold the greatest promise for investment.

Using the same dataframe construction, dictionary creation, and plotting functions that were demonstrated in the Sacramento metro area analysis, we can see which metro areas were most volatile during the crisis, and which ones might contain ZIP codes that still have upside potential after gains during the economic recovery.

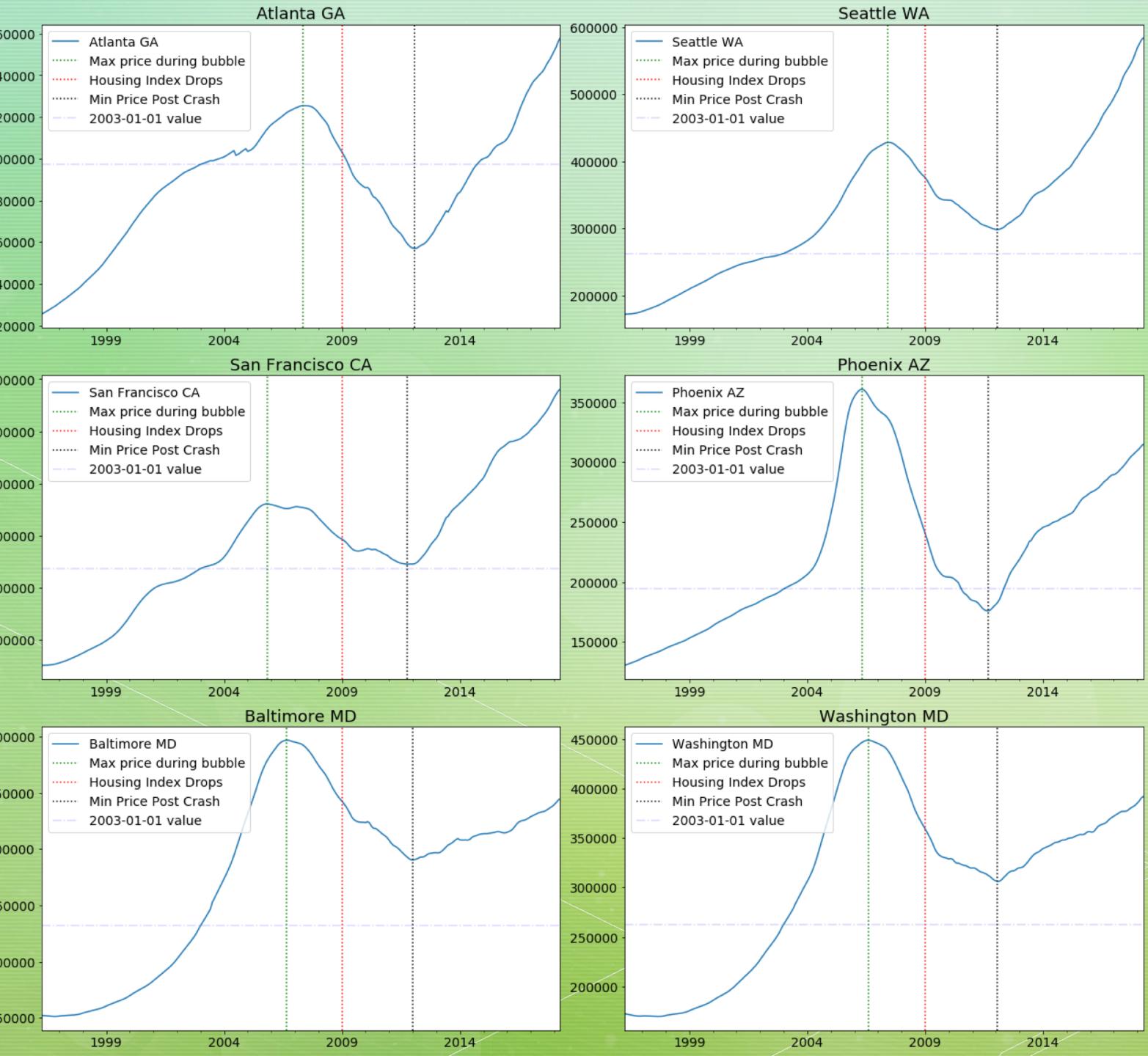
*Top 30 US metro areas  
(First group of six)*



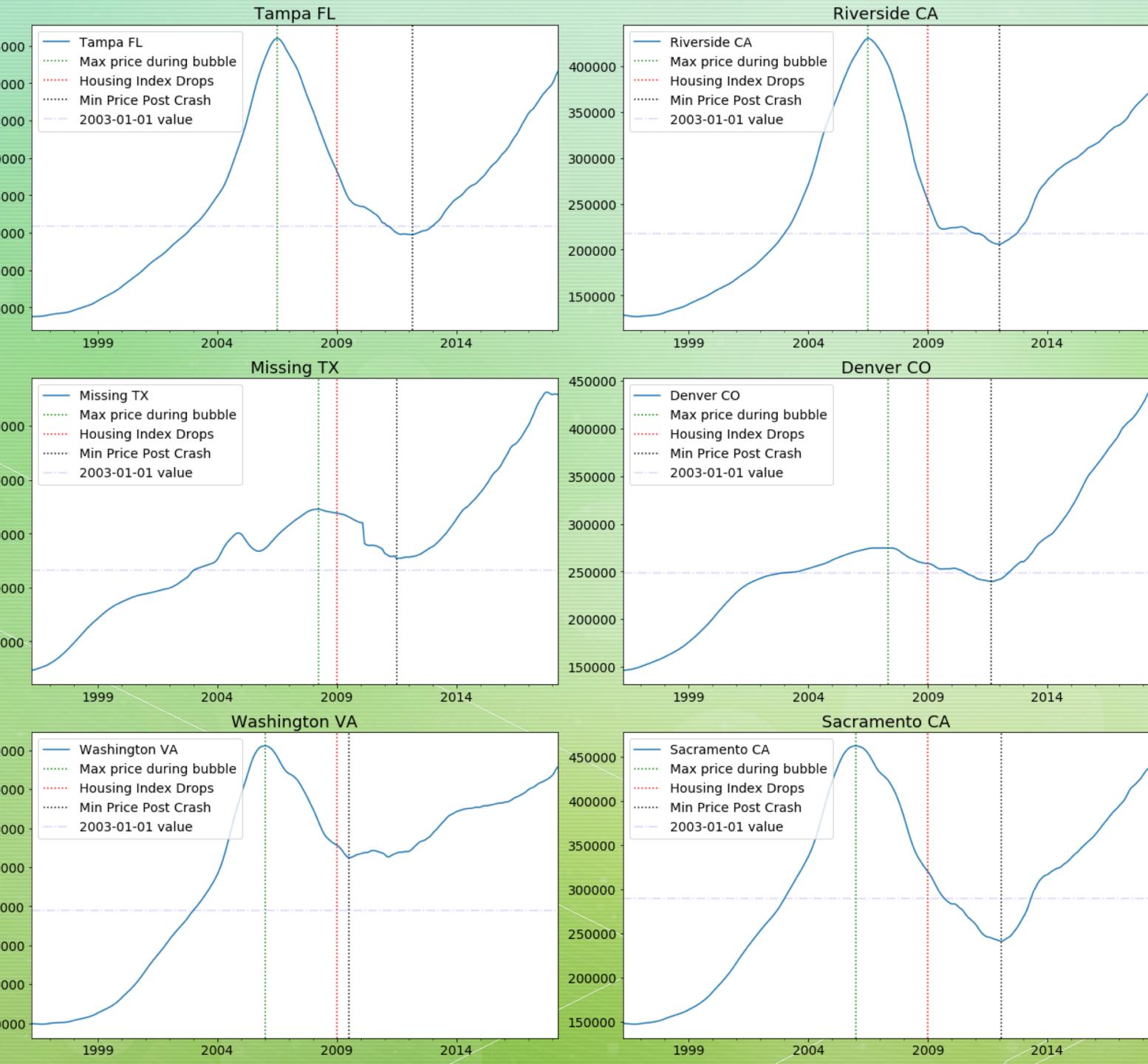
## *Top 30 US metro areas (Second group of six)*



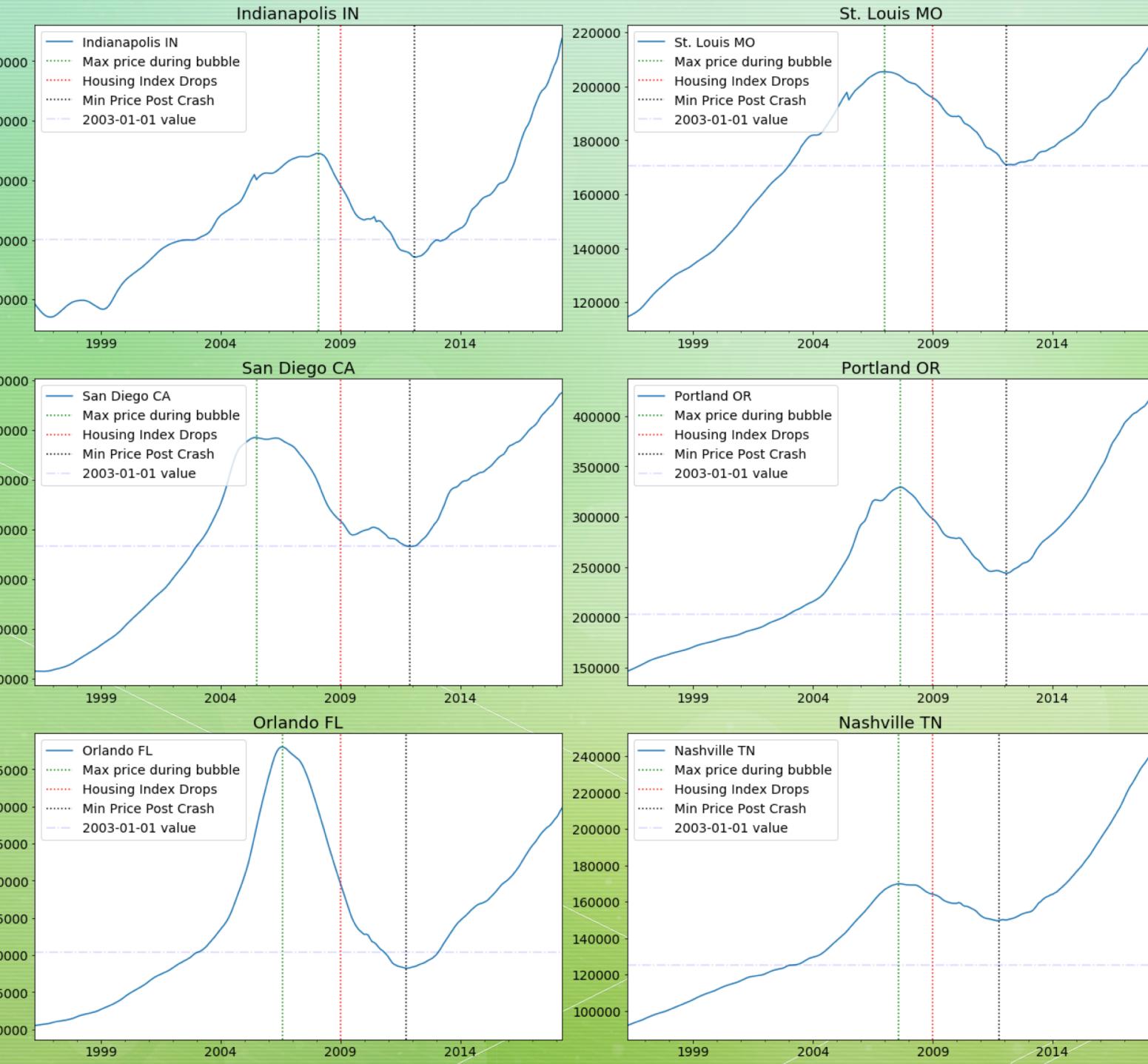
## *Top 30 US metro areas (Third group of six)*



*Top 30 US metro areas  
(Fourth group of six)*



*Top 30 US metro areas  
(Fifth group of six)*



# Historical and forecast plots: Dallas-Ft Worth TX

ARIMA modeling suggests that this area has excellent investment potential.

## *Dallas-Ft Worth TX metro area graphs*

