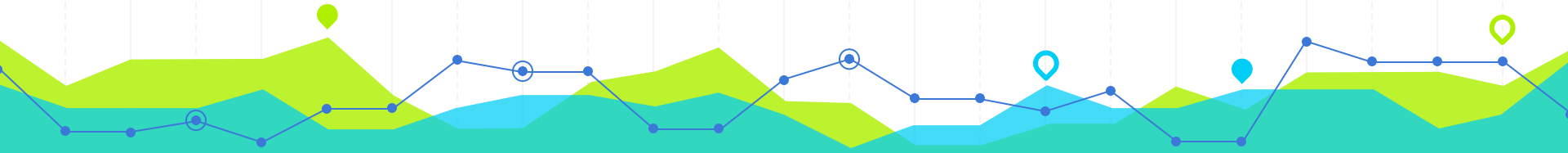


ZILLOW TIME SERIES ANALYSIS

By: Melody Peterson

Flatiron School Data Science Project - Phase 4

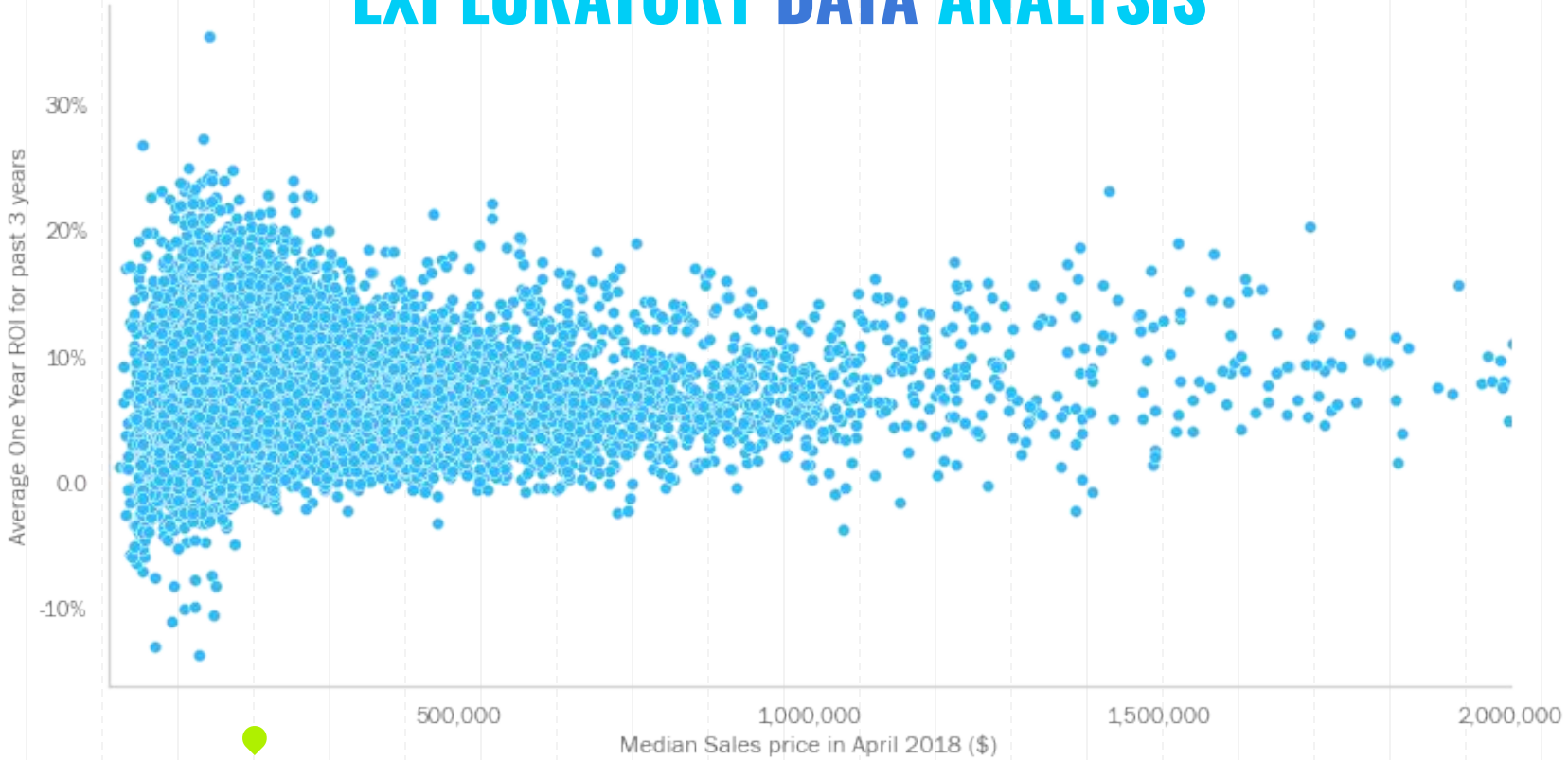
BUSINESS PROBLEM

- Client is a financial investment firm
- Looking for short-term real estate investments
- Particularly to diversify the portfolio of smaller investors

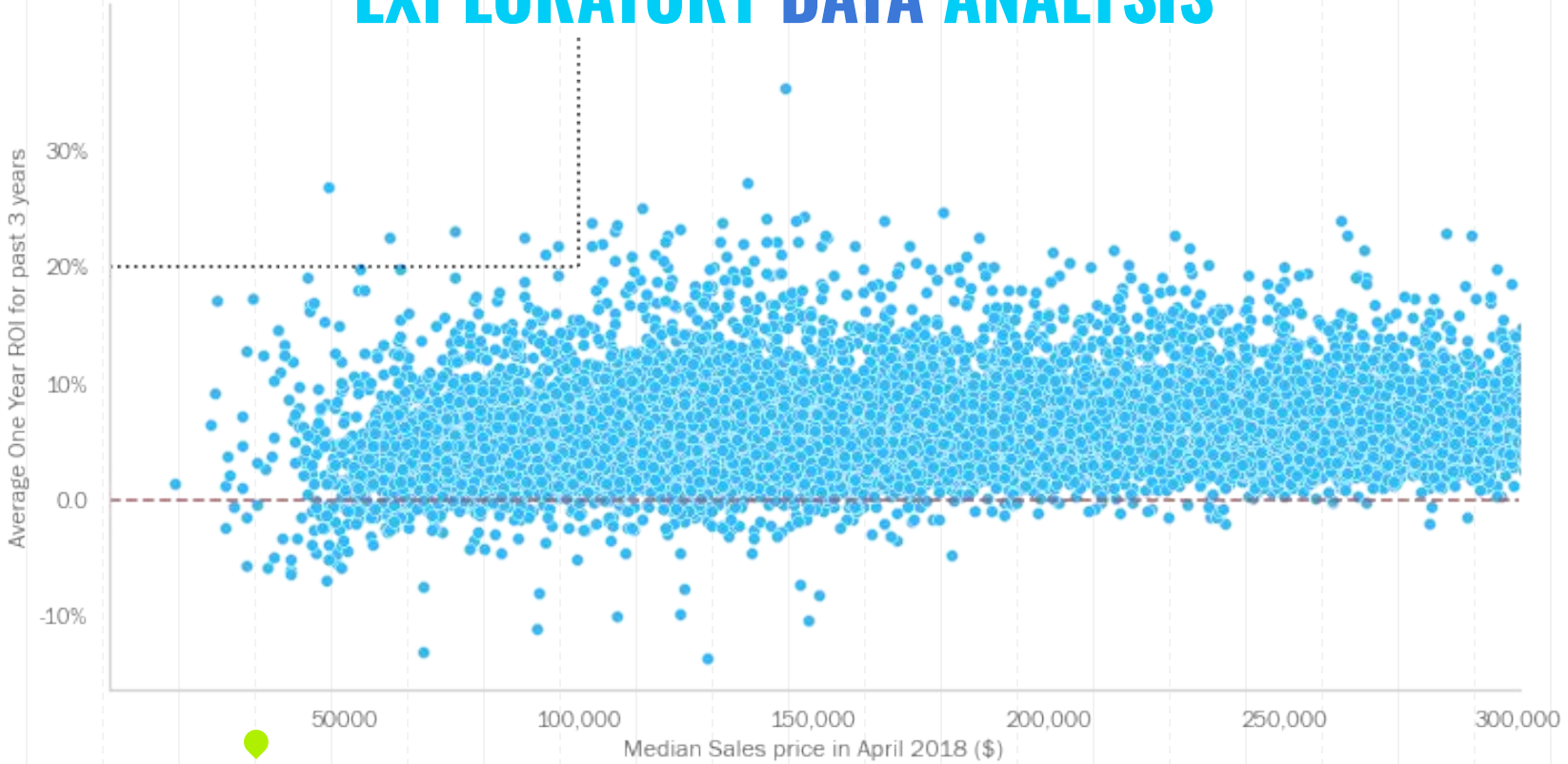
Analyze median monthly housing sales prices for over 14,000 United States zip codes and choose the best areas to further analyze for potential investment.

Begin by calculating 4 year ROI, Current year ROI, and Average ROI

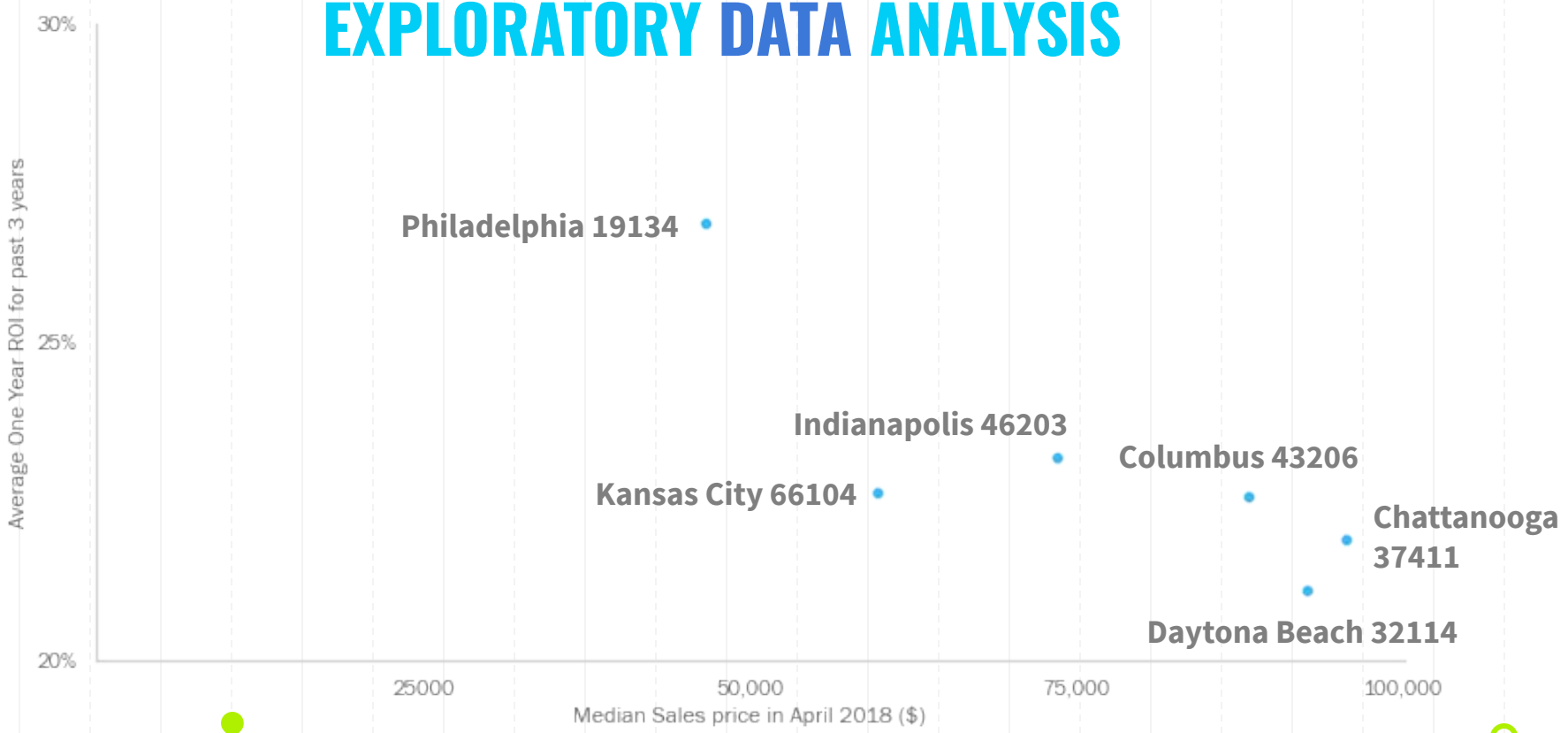
EXPLORATORY DATA ANALYSIS



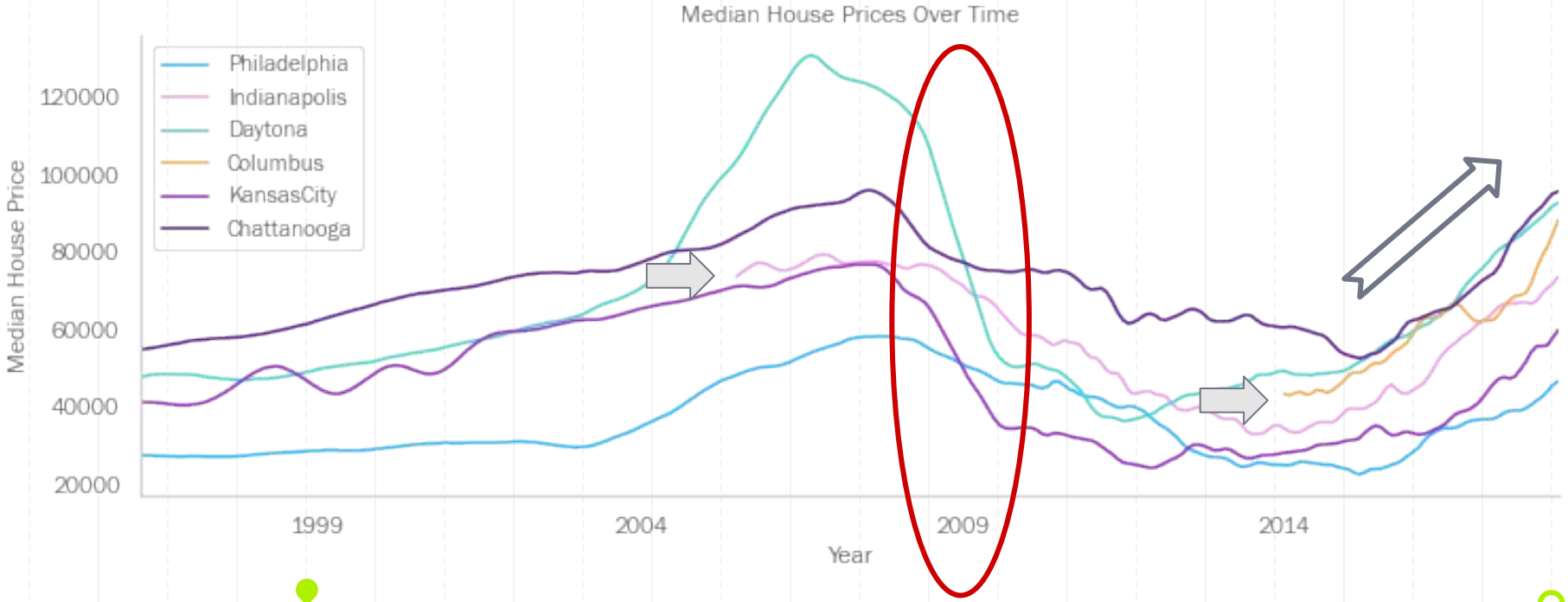
EXPLORATORY DATA ANALYSIS



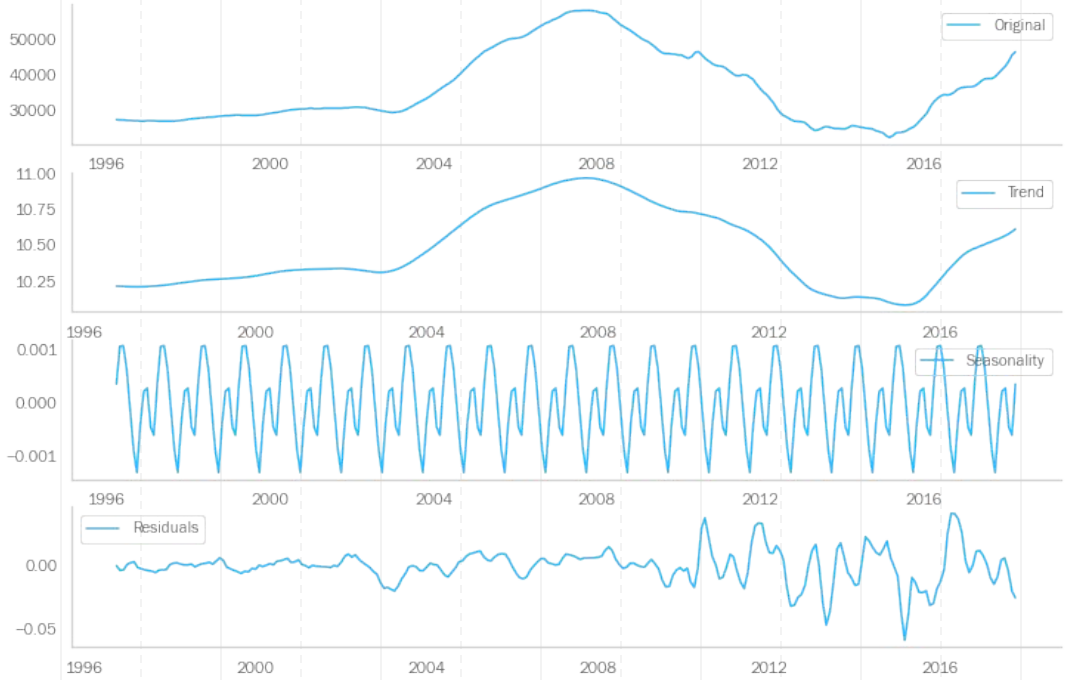
EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS



Seasonal Decomposition

Each time series can be broken down into its components of trend, seasonality, and residual or “noise” in the data.

MODELING - BASELINE

	Name	Order	Seasonal_Order	Fit_Time	Const	ar.L1	ma.L1	sigma2	AIC Score
0	Philadelphia	(1, 0, 1)	(0, 0, 0, 0)	0.1751	36279.03	0.9975	0.7643	105736.54	3660.12
1	Indianapolis	(1, 0, 1)	(0, 0, 0, 0)	0.1616	56143.06	0.9971	0.8044	365498.65	2237.24
2	Daytona	(1, 0, 1)	(0, 0, 0, 0)	0.1536	65794.06	0.9976	0.9413	541074.22	4074.94
3	Columbus	(1, 0, 1)	(0, 0, 0, 0)	0.0608	54074.93	0.9913	0.9994	362743.00	642.68
4	Kansas City	(1, 0, 1)	(0, 0, 0, 0)	0.1556	48263.62	0.9970	0.8875	209724.01	3834.07
5	Chattanooga	(1, 0, 1)	(0, 0, 0, 0)	0.2144	70698.40	0.9979	0.8266	167378.79	3776.88

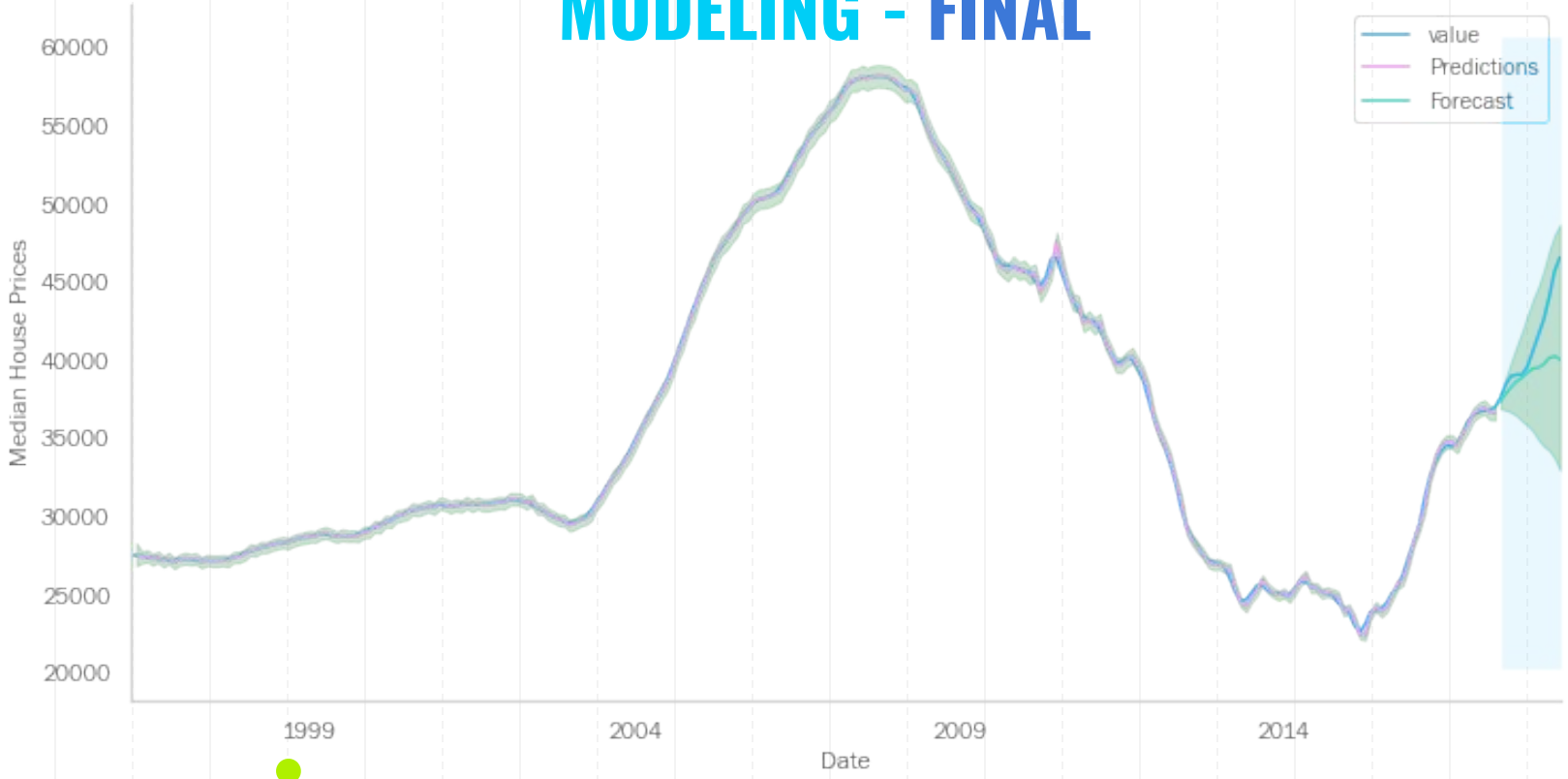
Auto Regressive(1) and Moving Average(1) model

Each time series is modeled as a function of the observation just before it as well as a function of the error value observed.

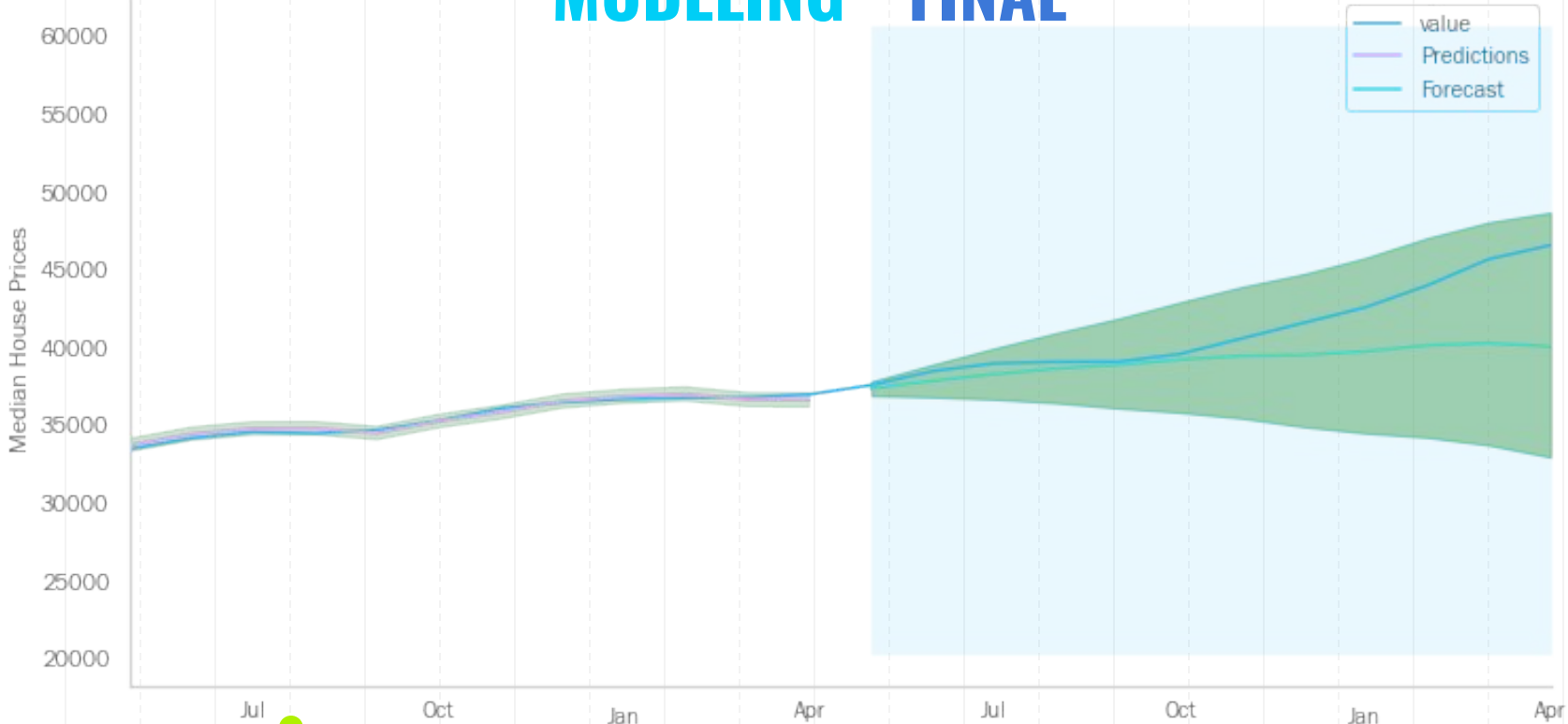
MODELING - FINAL

	name	order	seasonal order	ar.L1	ar.L2	ma.L1	ma.L2	ar.S.L12	ma.S.L12	sigma2	aic	train rmse	test rmse	ar.S.L24	ma.S.L24
0	Philadelphia	(2, 1, 2)	(1, 0, 1, 12)	1.1489	-0.1829	-0.1092	-0.3715	0.5673	-0.9261	0.0000	-1818.25	210.120410	2904.602710	NaN	NaN
1	Indianapolis	(1, 2, 2)	(2, 0, 2, 12)	0.4495	NaN	-0.2736	-0.5485	-0.2887	0.2439	0.0001	-867.45	473.983586	6108.436530	0.367	-0.6304
2	Daytona	(0, 5, 2)	(0, 0, 0, 12)	NaN	NaN	-1.9436	0.9541	NaN	NaN	0.0000	-1824.65	318.066481	23990.065072	NaN	NaN
3	Columbus	(2, 4, 1)	(1, 0, 0, 12)	-0.0722	-0.8108	-0.9405	NaN	-0.2056	NaN	0.0002	-194.60	844.184953	17797.268489	NaN	NaN
4	KansasCity	(1, 1, 2)	(0, 0, 2, 12)	0.7820	NaN	0.7766	0.2137	NaN	-0.1090	0.0000	-1801.54	243.298505	3745.083869	NaN	-0.1074
5	Chattanooga	(1, 1, 2)	(0, 0, 2, 12)	0.7161	NaN	0.7526	0.2969	NaN	-0.1871	0.0000	-2068.06	254.974561	13426.120075	NaN	-0.1307

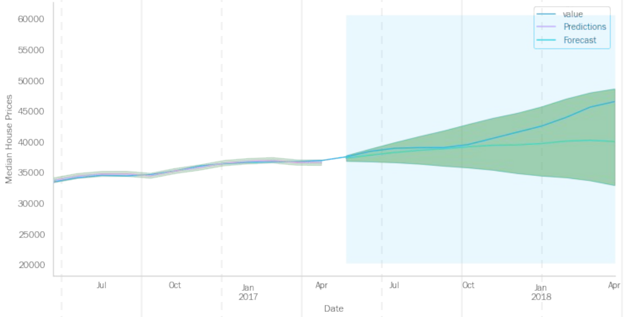
MODELING - FINAL

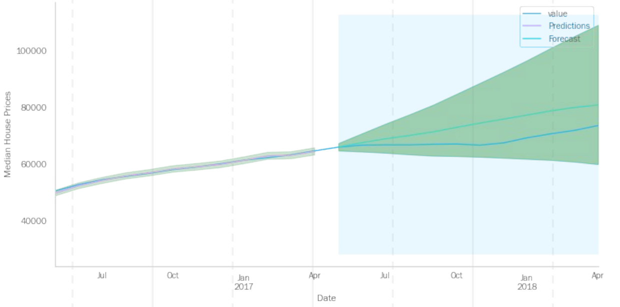


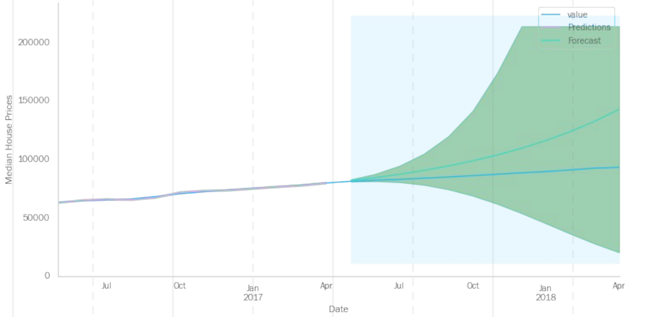
MODELING - FINAL

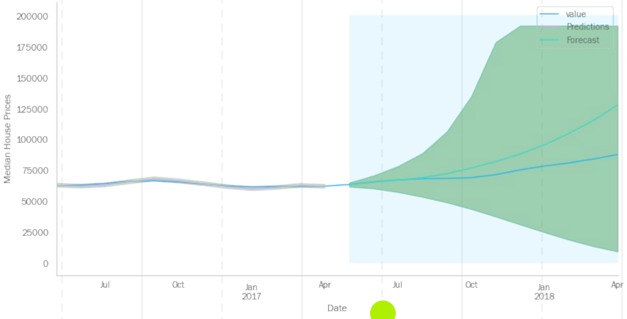


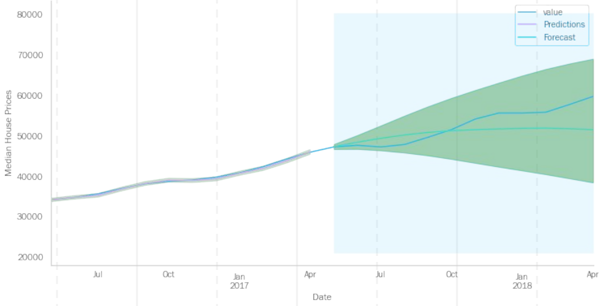
MODELING - FINAL

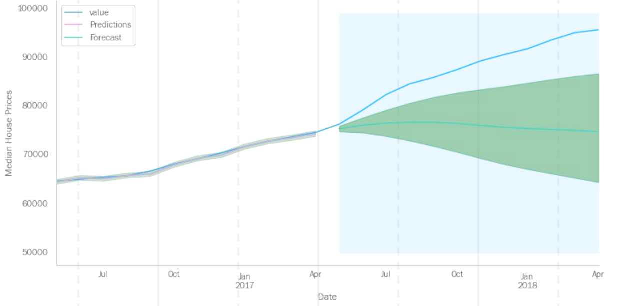












MODELING - FINAL

	Philadelphia	Indianapolis	Daytona Beach	Columbus	Kansas City	Chattanooga
Zip Code	19134	46203	32114	43206	66104	37411
median housing price	46600	73500	92600	88100	59800	95600
actual 2018 ROI	25.95	13.78	16.92	40.96	30	28.32
forecast 2018 ROI	7.21	22.42	75.04	101.29	8.87	-0.85

SUMMARY AND CONCLUSIONS

Results:

- All training data outperformed the test data, indicating overfitting.
- The models are all very skewed because of the market crash in 2009.
- Columbus and Daytona had very large confidence intervals and overly high forecasts.
- Chattanooga has outperformed even the confidence intervals of the model.
- Philadelphia is potentially a good 50K investment , Indianapolis at 75K and Chattanooga at 100K investment

Caveats:

- Logged and differenced the data but some still did not test as stationary according to the Dickey Fuller test.
- Real estate predictions can vary due to unseen fluctuations in the market

NEXT STEPS / FUTURE WORK

- Obtain current data after 2018 for current predictions. Found zip code data on Redfin but it is rolling average by zip code.
- Investigate why some of the models seem so far off in their forecasts.
- Try Facebook Prophet with each of the chosen zip codes
- Try other methods of choosing zip codes, including clustering to find trends
- Look for exogenous data to add to the models

THANKS!

Any questions?

You can find me at

GitHub / LinkedIn / melodygr@aol.com

Slide template by Slides Carnival