

Investment opportunity for the Syracuse Real Estate Investment Trust

February 23rd, 2019

Introduction

Zillow home value estimates (Zestimate's) are designed to provide an unbiased opinion on the value of a real estate properties independent from sellers and buyers. Seller's motivations to list their properties at higher prices are often driven by a desire to leave room for negotiations—sellers can even be motivated to list their properties at lower prices compared to market value to attract more buyers specially in must-sell situations. Buyers, on the other hand, use tactics such as cherry-picking comparable sales in order to low offers. With an increasing number of automated home valuation models, real estate investors need to spend resources analyzing data, such as Zillow dataset, in order to make sound decisions when it comes to where and when to execute their investment strategies. The amount of information and data points that can be used to predict home values are way too many, but sometimes a simple approach can yield sound results.

About the data

The initial Zillow dataset contains data for every city in the U.S. in addition to average home values between April 1996 and December 2018.

Column Name	Column description	Comment
Zillow	RegionID	Zillow region ID. Not used in this model
RegionName	ZIP code. Primary identifier for data values	Needs to be renamed for more clarity
City	Primary city for the zip code	
State	State zip code resides in	
Metro	Major metropolitan area this zip code is classified under	
CountyName	Primary county for the zip code	
SizeRank	Numerical rank for the zip code by median housing value	
1996-04 ... 2018-06	Median housing value by year and month for the zip code	

Table 1: Zillow data set column definition

Objectives:

- 1- Develop time series plots for the following Arkansas metro areas:
Hot Springs, Little Rock, Fayetteville, Searcy
 - i. Present all values from 1997 to present
 - ii. Average at the metro area level Sample data:
- 2- Develop model(s) for forecasting average median housing value by zip code for 2018
- 3- Predict which three zip codes provide the best investment opportunity for the Syracuse Real Estate Investment Trust (SREIT)

Data Exploration:

The following is a sample of the Zillow dataset:

	RegionID	ZipCode	City	State	Metro	CountyName	SizeRank	1996-04	1996-05	1996-06	...	2017-09
0	84654	60657	Chicago	IL	Chicago	Cook	1	334200.0	335400.0	336500.0	...	1008100
1	90668	75070	McKinney	TX	Dallas-Fort Worth	Collin	2	235700.0	236900.0	236700.0	...	312300
2	91982	77494	Katy	TX	Houston	Harris	3	210400.0	212200.0	212200.0	...	320100
3	84616	60614	Chicago	IL	Chicago	Cook	4	498100.0	500900.0	503100.0	...	1288000
4	93144	79936	El Paso	TX	El Paso	El Paso	5	77300.0	77300.0	77300.0	...	120000

5 rows × 274 columns

Fig.1: Zillow dataset sample

Filtering, grouping, and calculating the average value for the metro areas of Hot Springs, Little Rock, Fayetteville, and Searcy:

Metro	Fayetteville	Hot Springs	Little Rock	Searcy
1996-04	78055.56	72400.00	84800.00	60640.00
1996-05	78183.33	72375.00	85017.24	60800.00
1996-06	78261.11	72375.00	85200.00	60940.00
1996-07	78350.00	72325.00	85365.52	61100.00
1996-08	78444.44	72250.00	85496.55	61240.00

Fig.2: Fayetteville sample data

Notes:

- The dataset has been transposed in order to be used for time series analysis; a couple of different methods are used to transpose the data.

- The time component of the dataset is converted to datetime object which is needed for modeling.

This allows for developing a graph for each of these metro areas in Arkansas:

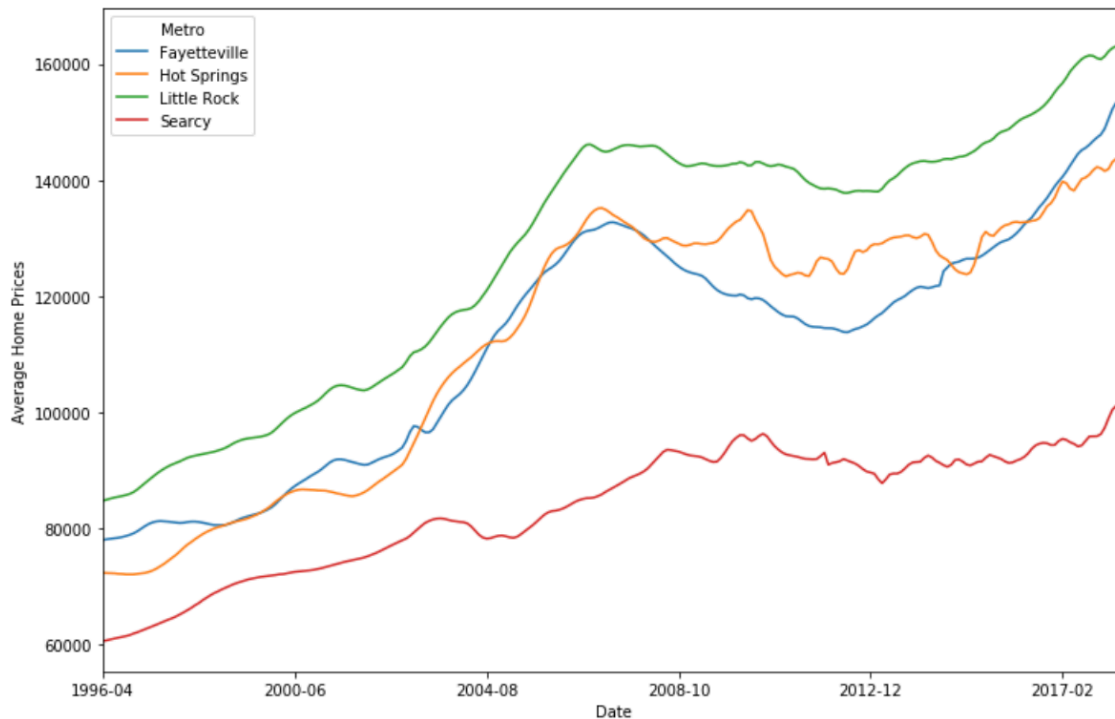


Fig. 4: Average home prices for selected metro areas in AR.

Observations:

- Average value reached a peak around 2007 for three of the metro areas except Fayetteville.
- Average value reached the lowest point in 2012 with significant growth afterwards.
- Investors in the area of Fayetteville had a great opportunity to invest since 2012 with a rapid growth.
- Though Searcy has experienced some growth, the average home value has remained significantly lower compared to other metro area.
- Hot Springs has experienced some value instability.

Data Analysis:

A model using Facebook Prophet python library is developed to predict the average home value for the area of Fayetteville in order to determine if this metro area presents a decent opportunity for investment.

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

The model was trained using timeseries data from January, 1997 until December of 2017. Following model training, a prediction is made for all months in 2018.

The following table shows the predicted values for Fayetteville:

ds	yhat	yhat_lower	yhat_upper
2018-01-01	145702.40	143798.06	147581.94
2018-02-01	146225.34	144285.01	148133.82
2018-03-01	146804.84	144957.81	148699.41
2018-04-01	147291.97	145399.23	149281.63
2018-05-01	147787.03	145929.10	149778.17
2018-06-01	148319.63	146270.11	150363.96
2018-07-01	148976.88	146732.73	151270.40
2018-08-01	149519.12	146917.48	152231.79
2018-09-01	150084.35	147451.63	152714.83
2018-10-01	150620.28	147733.25	153478.39
2018-11-01	151106.32	147831.33	154852.50
2018-12-01	151578.04	147953.12	155253.88

Fig. 5: 2018 Prophet predictions for Fayetteville

Plotting the Prophet model results for Fayetteville:

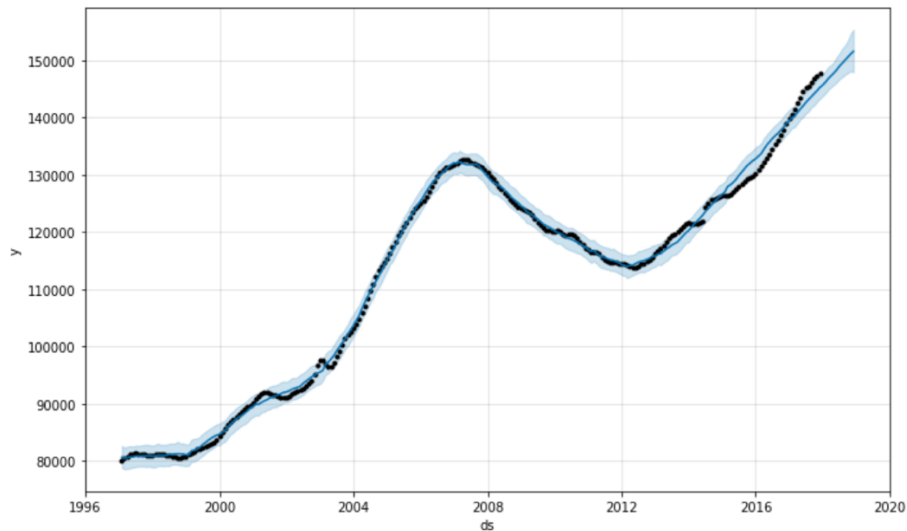


Fig.6: Plot of prophet model results for Fayetteville(1997-2018)

Observations:

- Model does a good job picking up the timeseries overtime.
- 2018 predictions indicate
- Is this timeseries pattern similar to other areas or even is it similar to the national average?

Plotting the forecast components. By default, you'll see the trend, yearly seasonality, and weekly seasonality of the time series

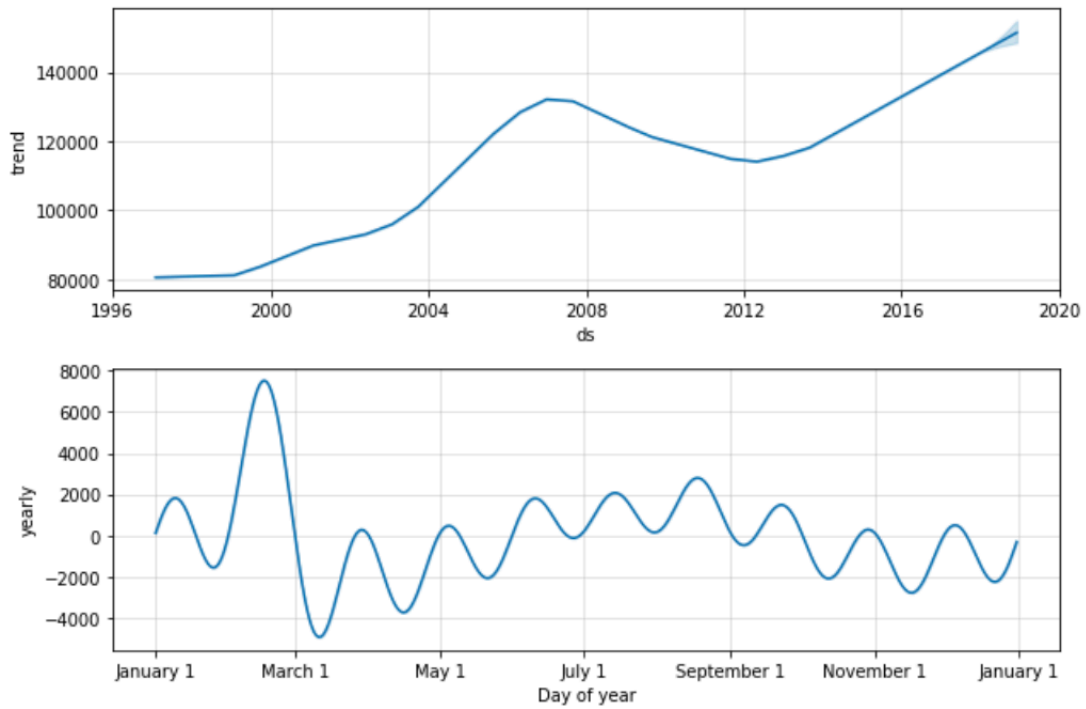


Fig. 7: Model components

Observations:

- Yearly trend indicates that February is perhaps the worst month to buy property in Fayetteville.
- Likewise, the month of March seems like the best month to buy as prices are lowest.

Using a similar approach, we can plot the average value nationwide to check if Fayetteville follows a similar trend:

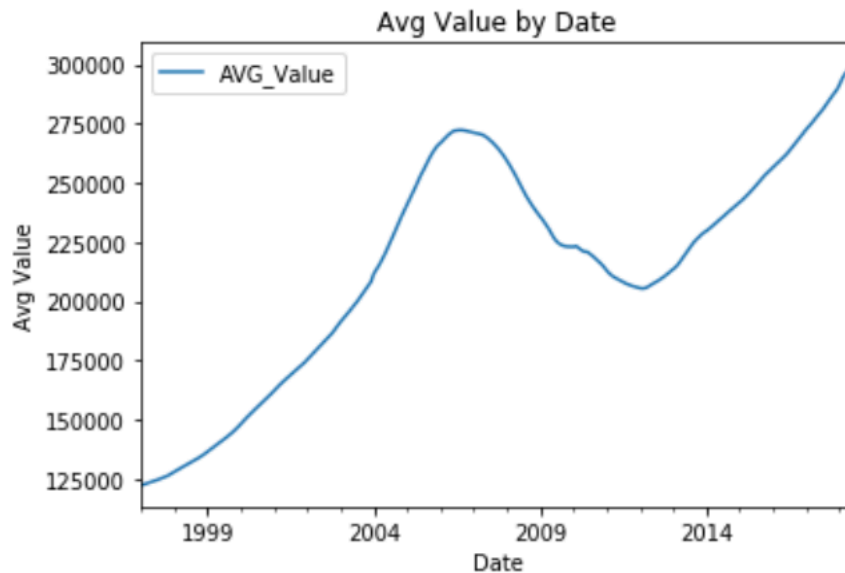


Fig. 8: Average home value nationwide (1997-2017)

Observation:

- At the national level, we can see the effect of the recession that hit the U.S. in 2008, when home prices plummeted in most US markets.

Plotting individual states should reveal if we see more states follow the same trend.

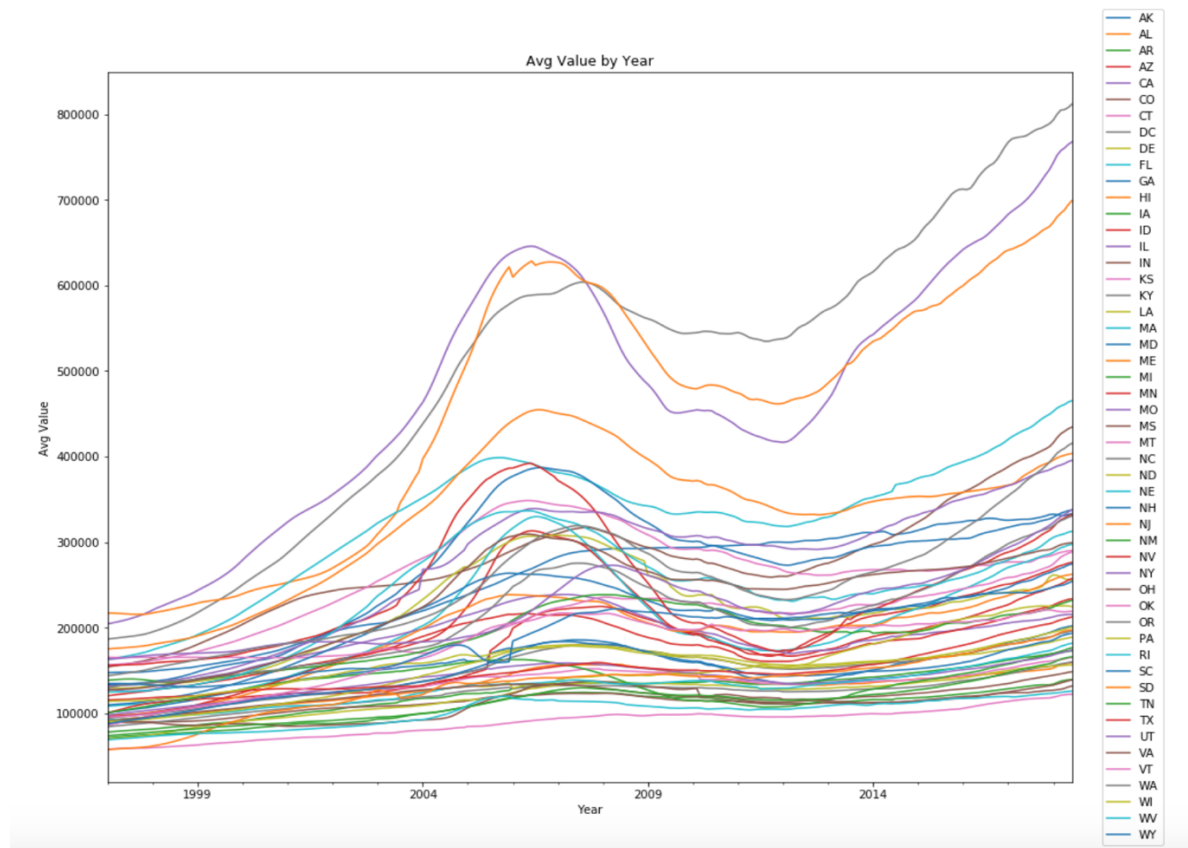


Fig.9: Average home value by state (1997-2017)

Observations:

- Some states follow the national trend.
- CA, HI, and DC exhibit the highest average values and fast growth following the 2012 fall.

Down Sampling for General model:

Searching for the states with highest growth over a 3-year period (2014-2017), should help us reduce our target sample of zip codes to build and train our Prophet time series models. The objective is to forecast 2018 values using a reduced sample.

The following procedure is followed:

- 1- Obtain the values for December month.
- 2- Calculate % Growth for 3-Year period = $(\text{Dec2017}/\text{Dec2014} - 1)$
- 3- Sort by % growth and get the top states (keep states with at least 25% growth)

	Date	State	AVG_Value	GrowthPerc
12848	2017-12-01	WA	393434.0	0.398136
12838	2017-12-01	OR	316745.0	0.314715
12834	2017-12-01	NV	313678.0	0.298153
12810	2017-12-01	FL	285947.0	0.291838
12806	2017-12-01	CO	411603.0	0.282021
12845	2017-12-01	UT	319674.0	0.272988
12805	2017-12-01	CA	735420.0	0.263252
12814	2017-12-01	ID	246255.0	0.251232

Fig.10: Top states by % Growth.

Observations:

- Washington state shows the highest growth with nearly 40% growth
- California has the highest average prices of these states with over \$730k.
- We can use the zip codes from these states to train our forecast model.
- States with highest growth were Washington, Oregon, Nevada, Florida, Colorado, Utah, California, Indiana.

Building a model using an individual zip code (home zip: 33193) to test model accuracy. We use 1997-2017 as training data and 01/2018 though 06/2018 as test dataset.

	ds	yhat	yhat_lower	yhat_upper
	2018-07-01	350689.31	331739.23	366996.74
	2018-08-01	353365.64	330503.29	373807.81
	2018-09-01	355987.21	329342.38	381580.97
	2018-10-01	358478.55	328117.51	388480.11
	2018-11-01	360699.96	326267.15	393675.79
	2018-12-01	362970.29	323683.19	401993.50

Fig.11: Prophet forecast for FL zip code: 33193

Using sklearn metrics, the Mean Absolute Error (MAE) equals 4526.63

Plotting the results of the model for 33193 home zip code:

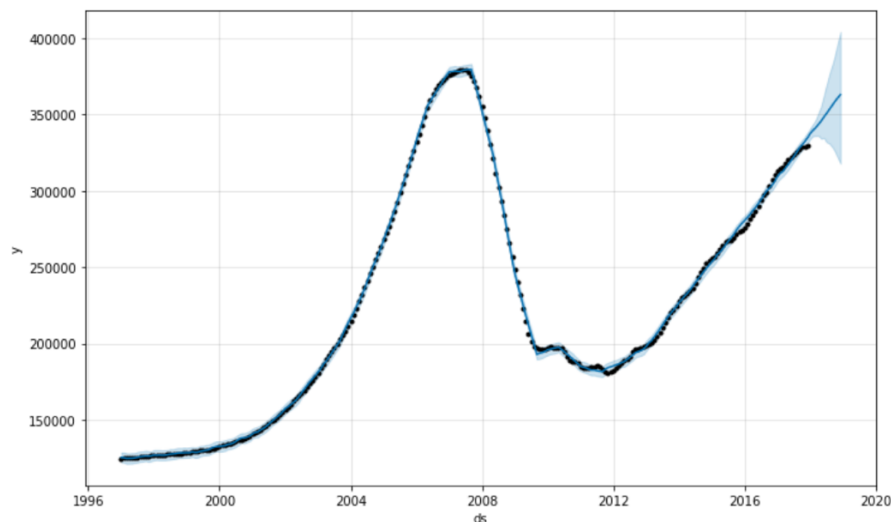


Fig. 12: Forecast plot for 33193, FL zip code

Observation

- Zip Code 33193 in FL shows a very similar trend compared to the nationwide trend.
- Prophet projects a 3.5% growth which is expected growth.
-

Exploring forecast components: general trend, yearly seasonality, and weekly seasonality of the time series

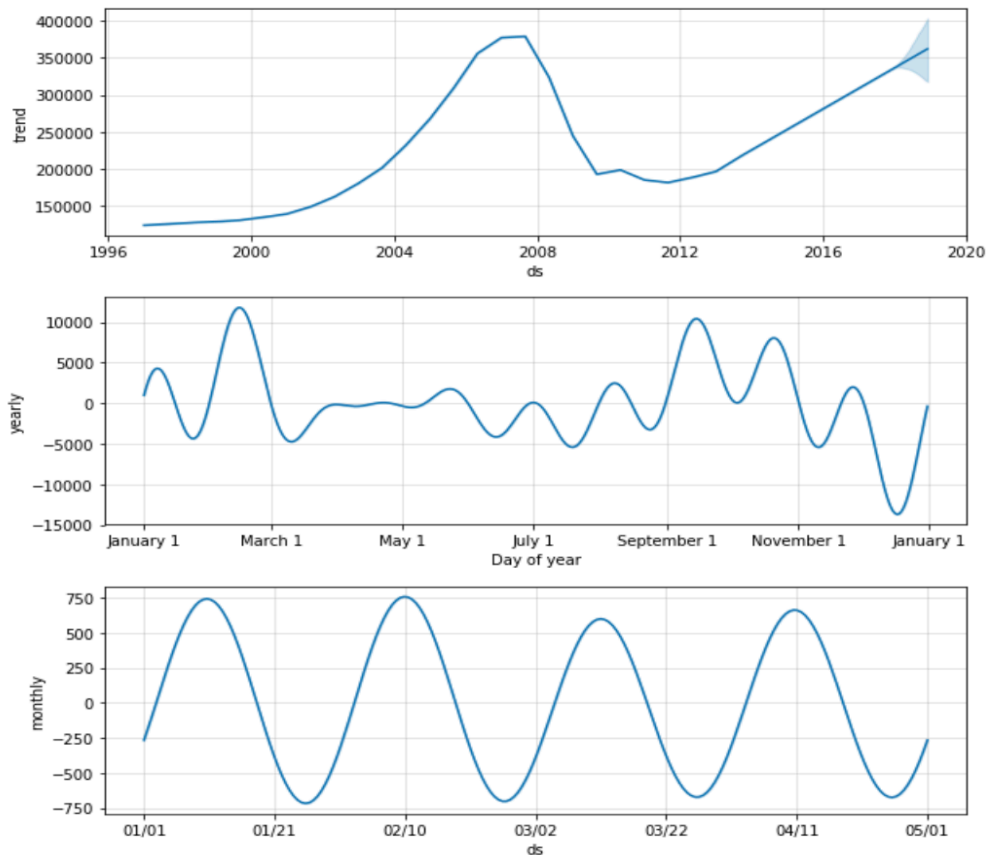


Fig. 13: Forecast components for 33193, FL

Observations:

- The months of February, September, and October are historically the worst months to purchase a property in this zip code. This might be explained by the beginning of school year.
- December is the best month to purchase a home as process tend to decrease.

Now applying the same model technique, we expand the general model to a reduced **sample of 3036 zip codes** from the states with highest average growth. This requires parallel processing using multiprocessing python library, to take advantage of multi-code CPU. The process takes roughly 1 hour and 4 minutes to complete.

Criteria for finding best opportunities:

- 2018 forecasted growth is calculated as (Dec-2018/Dec-2017)-1
- A dataframe is created with zip codes with low MAE and double-digit growth. This should help us find **low-risk, decent reward opportunities**. Results are shown in Fig.14
- Create another dataframe with high forecasted growth and MAE < 20k. This should reveal **high-reward, high-risk opportunities**. Results are shown in Fig.15

ZipCode	City	State	Metro	y	yhat	MSE	MAE	R2	2018 Growth
33919	Fort Myers	FL	Fort Myers	265500.00	292865.93	47076.37	168.00	0.98	0.10
34987	Port Saint Lucie	FL	Port St. Lucie	265100.00	293964.50	2512599.87	1216.05	0.70	0.11
33063	Margate	FL	Miami-Fort Lauderdale	263200.00	291079.55	6561301.55	2256.42	0.21	0.11
97114	Dayton	OR	Portland	259400.00	285057.96	7816272.38	2513.91	0.52	0.10
89433	Sun Valley	NV	Reno	258700.00	285392.59	11654038.43	3163.91	0.44	0.10
33170	Goulds	FL	Miami-Fort Lauderdale	256600.00	286671.69	15716654.25	3551.81	0.43	0.12
89701	Carson City	NV	Carson City	257200.00	283166.81	25669306.64	4585.14	0.26	0.10
95820	Sacramento	CA	Sacramento	257900.00	283330.87	23126166.77	4794.27	-5.95	0.10
89706	Carson City	NV	Carson City	255200.00	279899.46	34755534.25	4951.81	0.00	0.10
32764	Mims	FL	Melbourne	248400.00	285738.29	48545516.06	5044.27	-0.28	0.15

Fig.14: Zip codes with lowest MAE and greater than 10% growth

ZipCode	City	State	Metro	y	yhat	MSE	MAE	R2	2018 Growth
93927	Greenfield	CA	Salinas	284400.00	337748.21	695993855.07	26172.40	-28.96	0.19
95485	Upper Lake	CA	Clearlake	254000.00	295833.17	50465313.05	5744.27	-2.52	0.16
33955	Punta Gorda	FL	Punta Gorda	232000.00	269729.00	796460966.90	28094.27	-44.71	0.16
89408	Fernley	NV	Fernley	224300.00	257808.47	860935303.34	29127.60	-33.70	0.15
34231	Sarasota	FL	North Port-Sarasota-Bradenton	250600.00	288194.17	292012217.24	17044.27	-320.28	0.15
80621	Fort Lupton	CO	Greeley	266300.00	307055.60	150672805.93	11422.40	-3.17	0.15
33125	Miami	FL	Miami-Fort Lauderdale	271500.00	312392.78	273401866.55	15989.07	-7.28	0.15
95453	Lakeport	CA	Clearlake	262700.00	301951.02	75296072.24	7273.08	-0.54	0.15
34223	Englewood	FL	North Port-Sarasota-Bradenton	245400.00	281177.60	272824269.89	16444.27	-30.04	0.15
95467	Hidden Valley Lake	CA	Clearlake	285100.00	328961.03	627440130.63	24722.40	-19.49	0.15

Fig. 15: High growth zip codes with MAE<20k

Using Plotly library, a plot with the low-risk zip codes is generated for Predicted Values vs MAE:

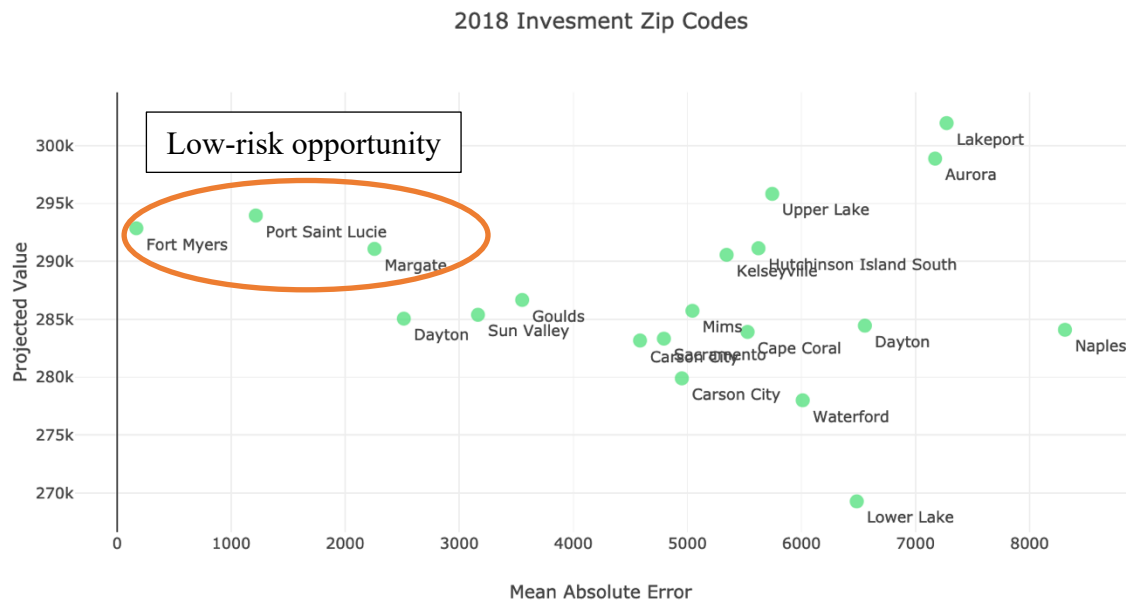


Fig. 16: Projected value vs MAE for zip codes

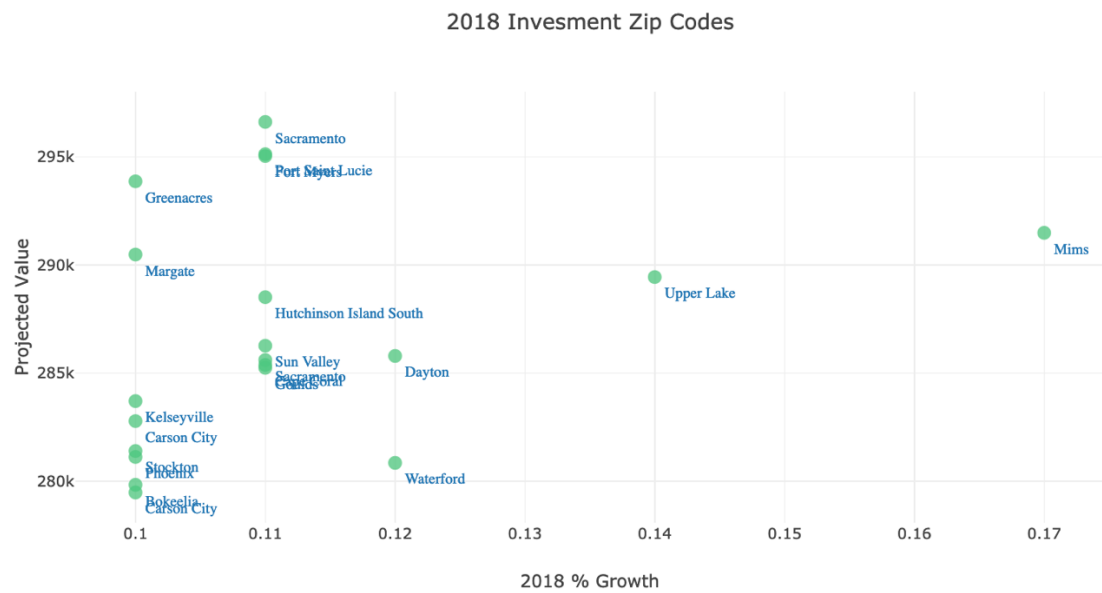


Fig. 17: Projected value vs forecasted growth in 2018

Conclusions:

The Zillow dataset with average home prices by zip code in the US, turns out to be a very useful dataset to predict future home prices. By finding the states with the highest growth in a 3-year period—from 2014 through December of 2017—a reduced dataset was used to forecast home prices in 2018 for each individual zip code in Washington, Oregon, Nevada, Florida, Colorado, Utah, California, and Indiana.

By running Prophet forecast models in parallel—minimizing runtime for all 3k zip codes—a dataset containing zip codes with low-risk (low MAE) and at least 10% return on investment was generated. From this dataset, we can recommend the following zip codes which happen to be in the state of Florida:

- 33919 in Fort-Myers, FL
- 34987 in Port Saint Lucie, FL
- 33063 in Margate (Miami-Fort Lauderdale market), FL

Margate and Port Saint Lucie are within one hour of each other in the south east of Florida, and Fort-Myers is in the south west corner of the state. The average home value in these zip codes is between \$263k and \$265k which offers an attractive entry point to market in 2018. A projected 10% increase in price represents a good opportunity for investment. Should the value growth remain steady for these 3 zip codes, a potential 30% return in investment over a 3-year period is fantastic. The fact that these zip codes are in relatively close proximity to each other, might help investors keep the cost of maintenance under control. On the negative side, Florida is always at risk of hurricanes, therefore high insurance premiums might work against the bottom line.

Higher reward zip codes are found in California, however, the risk of investing in such zip codes is higher given the accuracy of our models. Perhaps further analysis can help refine the forecast accuracy for CA and help expand the investment landscape.

Bibliography

- Facebook. (n.d.). *Prophet Github*. Retrieved from Forecasting at Scale: <https://facebook.github.io/prophet/>
- Gamboa, M. A. (2018, December 6). *Medium*. Retrieved from Forecasting multiple time-series using Prophet in parallel: <https://medium.com/devschile/forecasting-multiples-time-series-using-prophet-in-parallel-2515abd1a245>
- Plotly. (n.d.). *Plotly*. Retrieved from Jupyter Notebook Tutorial in Python: <https://plot.ly/python/ipython-notebook-tutorial/#getting-started>
- Thind, J. (2017, May 24). *Zillow AI blog*. Retrieved from Home Value Estimates: Understanding Their Purposes And Evaluating Their Results: <https://www.zillow.com/data-science/home-value-estimates/>
- Vincent, T. (2017, April 4). *Digital Ocean*. Retrieved from A Guide to Time Series Forecasting with Prophet in Python 3: https://www.digitalocean.com/community/tutorial_series/time-series-visualization-and-forecasting