Miguel Arabi

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.

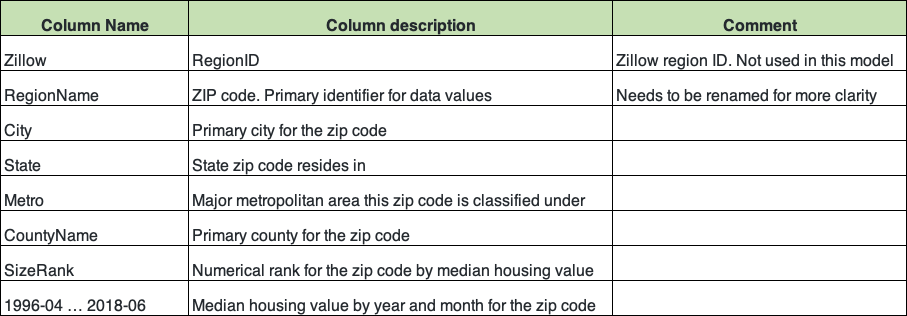


Table 1: Zillow data set column definition

**Objectives:**

1. Develop time series plots for the following Arkansas metro areas:

Hot Springs, Little Rock, Fayetteville, Searcy

* + 1. Present all values from 1997 to present
    2. Average at the metro area level Sample data:

1. Develop model(s) for forecasting average median housing value by zip code for 2018

1. 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:

A screenshot of a cell phone

Description automatically generated

Fig.1: Zillow dataset sample

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

A screenshot of a cell phone

Description automatically generated

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:

A close up of a map

Description automatically generated

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.
* ot spHHot 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:

**A screenshot of a cell phone

Description automatically generated**

Fig. 5: 2018 Prophet predictions for Fayetteville

Plotting the Prophet model results for Fayetteville:

**A close up of a map

Description automatically generated**

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

**A close up of a map

Description automatically generated**

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:

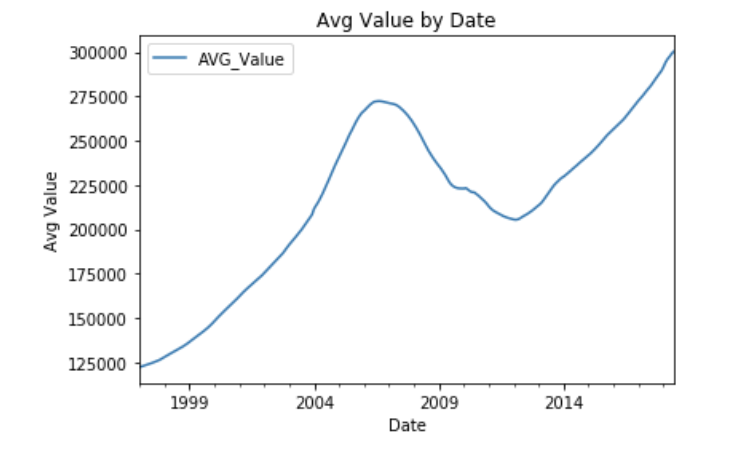


Fig. 8: Average home value nationwide (1997-2017)

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.

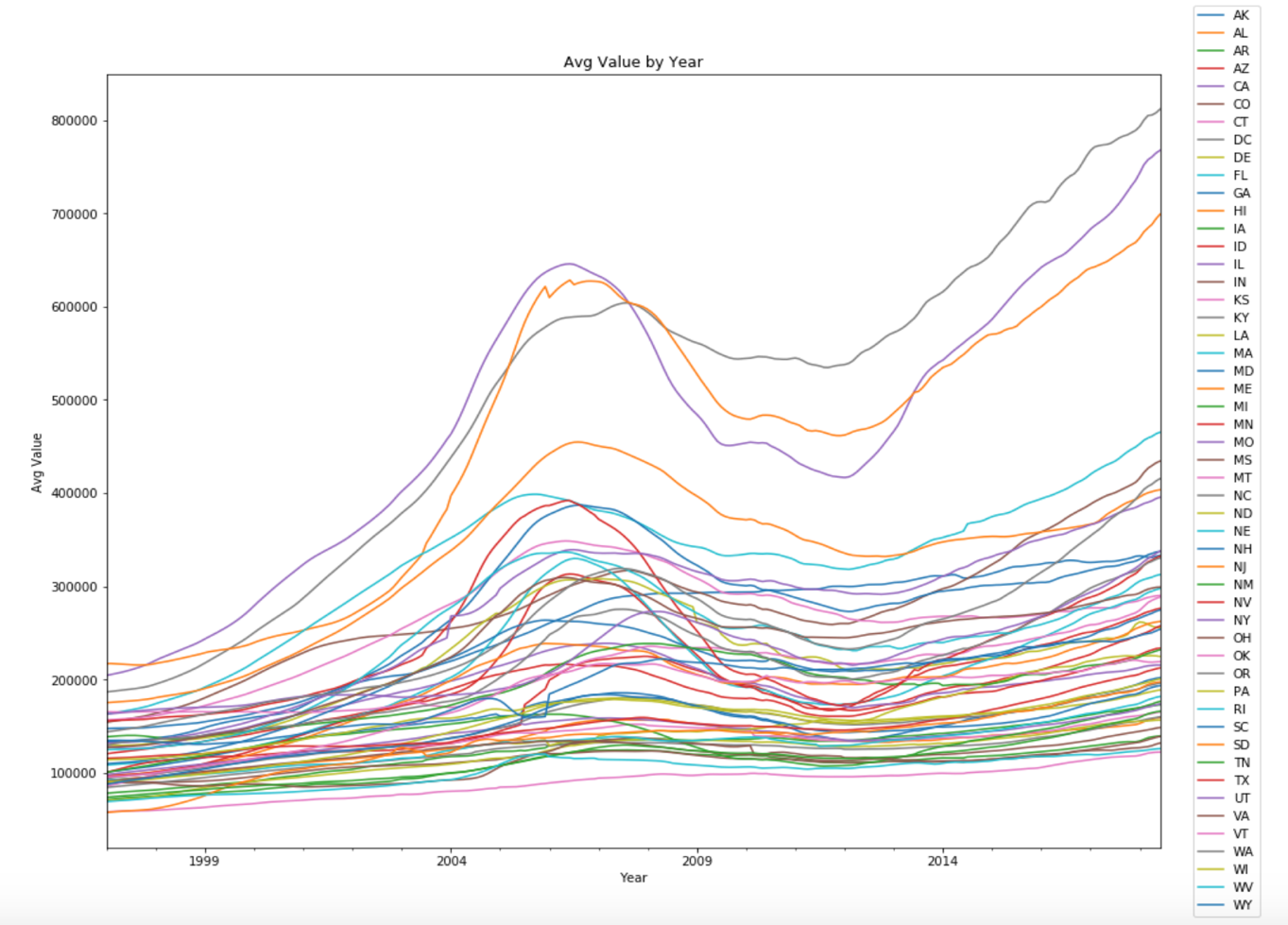


Fig.9: Average home value by state (1997-2017)

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 = (Dec2017/Dec2014 – 1)

3- Sort by % growth and get the top states (keep states with at least 25% growth)

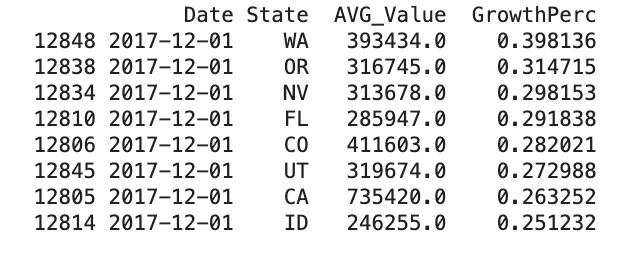


Fig.10: Top states by % Growth.

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.

A screenshot of a cell phone

Description automatically generated

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:

A picture containing indoor, text

Description automatically generated

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

A close up of a map

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a cell phone

Description automatically generated

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:

A close up of a map

Description automatically generated

Low-risk opportunity

Fig. 16: Projected value vs MAE for zip codes

A screenshot of text

Description automatically generated

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