Zillow Home Price Analysis

The research question is can we predict which three zip codes provide the best investment opportunity for the Syracuse Real Estate Investment Trust (SREIT)?

- Using the base data available from Zillow (files.zillowstatic.com/research/public/Zip/Zip_Zhvi_SingleFamilyResidence.csv)
- Develop time series plots for the following Arkansas metro areas:

    o Hot Spring, Little Rock, Fayetteville, Searcy

    o Present all values from 1997 to present

    o Average at the metro area level

- Using data from Zillow: o Develop model(s) for forecasting average median housing value by zip code for 2018
    o Use the historical data from 1997 through 2017 as your training data
    o Integrate data from other sources
- Answer the following questions:
    o What technique/algorithm/decision process did you use to down sample?
    o What three zip codes provide the best investment opportunity for the SREIT?
    o Why?

The average Median Home prices for the Hot Springs, Little Rock, Fayetteville, and Searcy metro areas in Arkansas from 1997 to early 2019 are shown below in Figure 1.
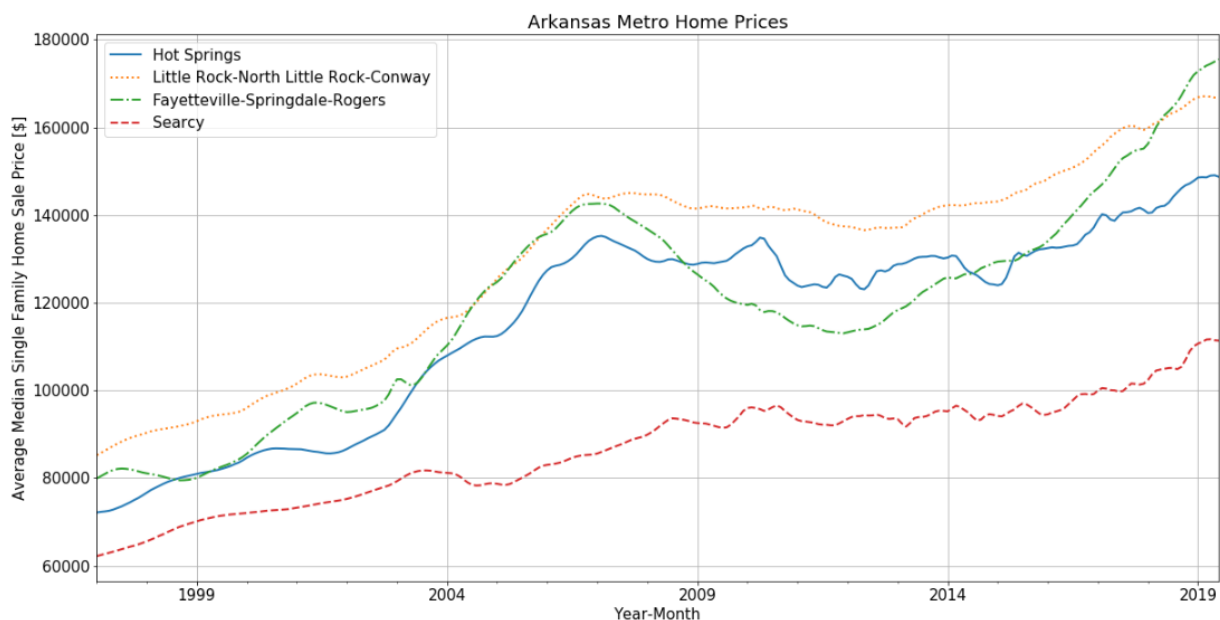

Figure 1: Arkansas Metro Average Median Home Prices

In order to determine the three best zip codes for an investment opportunity we need to build models for each zip code in the United States to forecast which ones will provide the best Return on Investment, and a metric to measure the quality of the forecast model. For a model metric we will use Mean Absolute Percent Error, or MAPE. MAPE is the average percent difference between the model prediction and the true value over the entire range of predicted values.

We used the [fbprophet](#) Python package, better known simply as Prophet, to build our prediction models. Prophet uses piecewise linear or logistic regression models for non-periodic trends in time-series data, and will also automatically detect and adjust for seasonality. It also has the capability to account for specific holidays, or non-regular special events that may have strongly influenced the time-series data. A single model can be built in a few seconds, though due to time constraints on this assignment we did not build a model for each of the more than 15,000 zip codes. Instead, we aggregated the average median home price for each of the 700+ metro areas in the provided dataset. Then, individual models for each zip code in the top performing metro areas would be used to select the best three zip codes to invest in. The top metro areas for initial consideration are shown in Table 1.

Table 1: Top Metro Areas

Metro	MAPE	Present Value / vs. 1997 Value (ROI)	Low ROI	High ROI	Present Avg Median Price
Corvallis	1.18135	4.91469	4.76774	5.03259	$350,166.67
Huntingdon	1.05684	4.48814	4.36494	4.6442	$120,850.00
Los Angeles-Long Beach-Anaheim	0.536661	4.15598	4.02278	4.268	$943,470.43
San Francisco-Oakland-Hayward	0.949296	4.55144	4.45091	4.67957	$1,342,274.02
San Jose-Sunnyvale-Santa Clara	1.27678	4.71931	4.59881	4.82194	$1,500,580.00
Selinsgrove	2.04086	4.39077	4.12105	4.70883	$158,866.67
Sioux Falls	0.728621	4.95986	4.89465	5.0539	$223,900.00
State College	1.03761	4.18036	4.09773	4.26764	$252,800.00
Vineyard Haven	0.601105	4.3887	4.30108	4.46276	$921,360.00

Each of these metro areas has an average ROI greater than 4.0 with MAPE less than 2%. Some of these metros however already have very high property values, close to or more than $1,000,000 dollars. These areas may not be the best candidates for investment, as the property values may require a significant capital investment that increases risk beyond what is tolerable for the SREIT, but putting "too many eggs in one basket" so to speak. However, some zip codes in those metro areas may be lower in current value, with the opportunity to dramatically increase in value due to being near other desirable neighborhoods.

After generating a model for each zip code in the top metro areas, we looked at the MAPE of each model vs. the ROI, to identify the zip codes with the highest returns that are also at a very high confidence level.
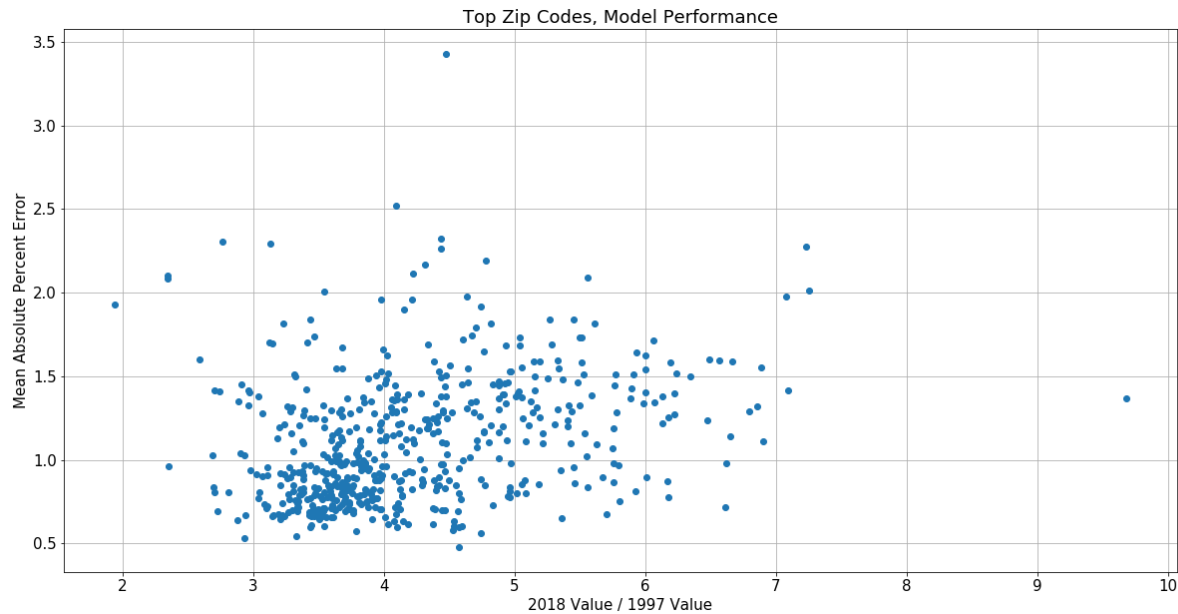

Figure 2: Top Metro Zip Code Model Results

In Figure 2, we see that all of the models have very good MAPE levels. Out of these we selected the top three zip codes with ROI greater that 6.5, and MAPE less than 1.5 %.

Table 2: Final Selected Zip Codes

Zip Code	MAPE	ROI	Low ROI	High ROI	City, State	Present Avg Median Price
90291	1.36747	9.68036	9.44992	9.95638	Los Angeles, California	$2,087,400.00
90048	1.41811	7.0938	6.93593	7.2514	Los Angeles, California	$2,185,500.00
90026	1.10998	6.90274	6.69389	7.08274	Los Angeles, California	$980,000.00

Table 2 shows the top 3 zip codes in terms of ROI over the previous 20 years, and the model for each has very good MAPE less than 1.5%. Each of these zip codes is in Los Angeles, California and are currently at very high Average Median Prices from $980,000 to $2,000,000 dollars. Investing in these areas is going to be expensive, but the projected returns are the best of those zip codes currently analyzed, and the model plots in Figures 3-5 show that home values are currently on significant upward trends following local minimums in the 3-4 years after the financial crisis of 2007-2008. Over an additional 20 year holding of investments, these zip codes offer great promise of returning investor value.
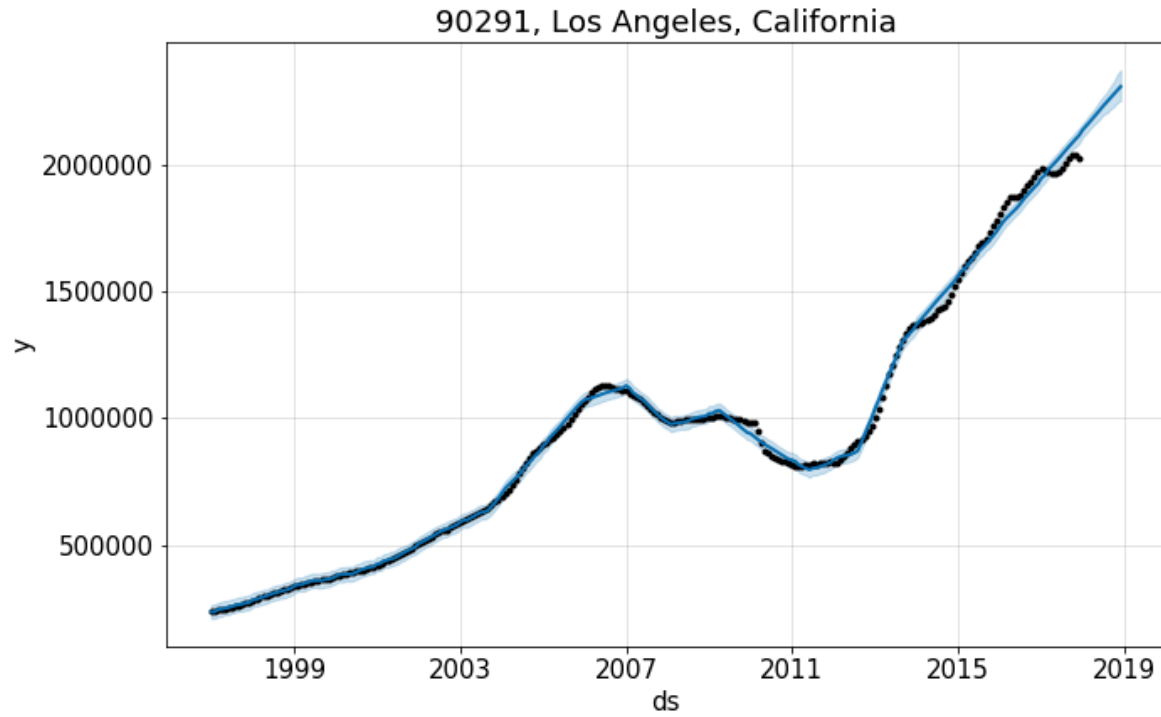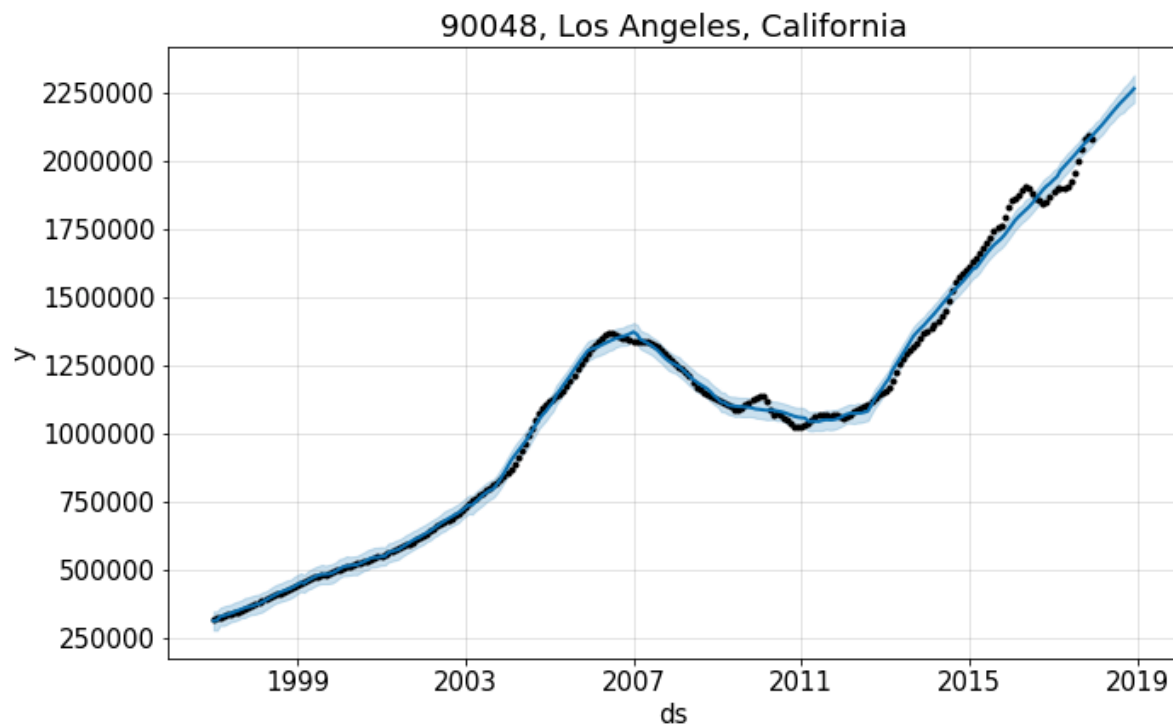
Figure 3: 90291 Data and Model
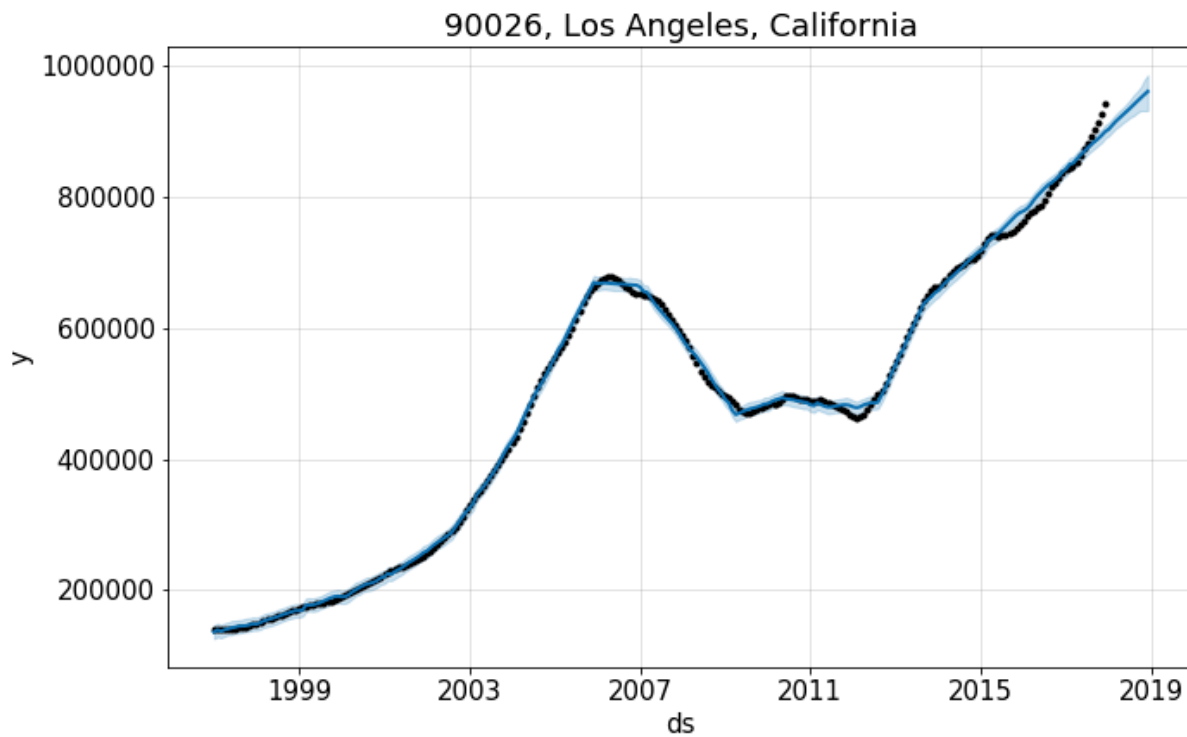

Figure 4: 90048 Data and Model

Figure 5: 90026 Data and Model

We did attempt to incorporate Population and Income data, but did not see drastic improvements in model performance versus the models using solely the time-series data, and it was not deemed worth the investment to include for this analysis at this time. Further analysis should be conducted by building a model for every zip code in the available dataset, and to search for features in could indicate larger future growth.