Housing Investment Opportunities

Time Series Modeling and Analysis of
Housing Prices in Springfield, Missouri

Andrew Muller
Deric Williamson

Business Case

As Data Scientists working for a home renovation company, we were given the task to investigate and predict the average housing prices in the next two years using various zip codes in Springfield, MO.

We will use various time series models to determine which zip code would be the best investment to buy houses in to renovate.

Overview

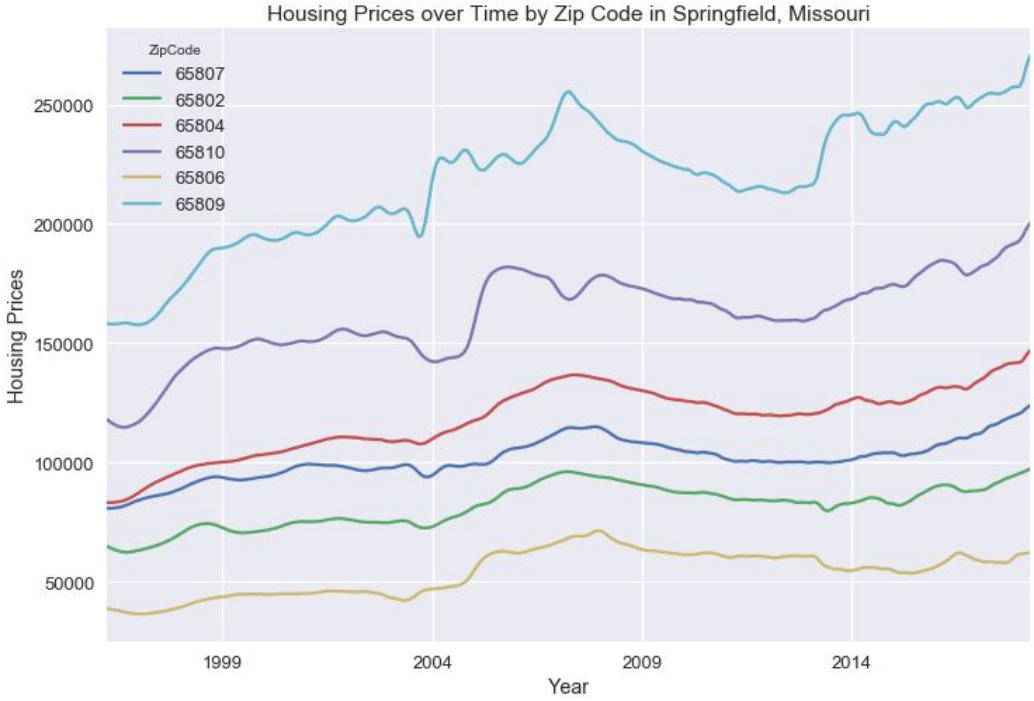
- Sourced from Zillow: https://www.zillow.com/research/data/
- Mean housing price by zip code from April 1996 to April 2018 (22 year)
- 15,000 different zip codes, narrowed to the 6 around Springfield, MO
- Forecasting prices over the next two years (May 2018 to April 2020)

- Libraries used:
  - Facebook Prophet
  - Matplotlib
  - NumPy
  - pandas
  - pmdarima
  - scikit-learn
  - SciPy
  - statsmodels

Exploratory Data Analysis

Dickey-Fuller test results:

- 65807: 0.788008
- 65802: 0.233519
- 65804: 0.391874
- 65810: 0.081698
- 65806: 0.211948
- 65809: 0.133164

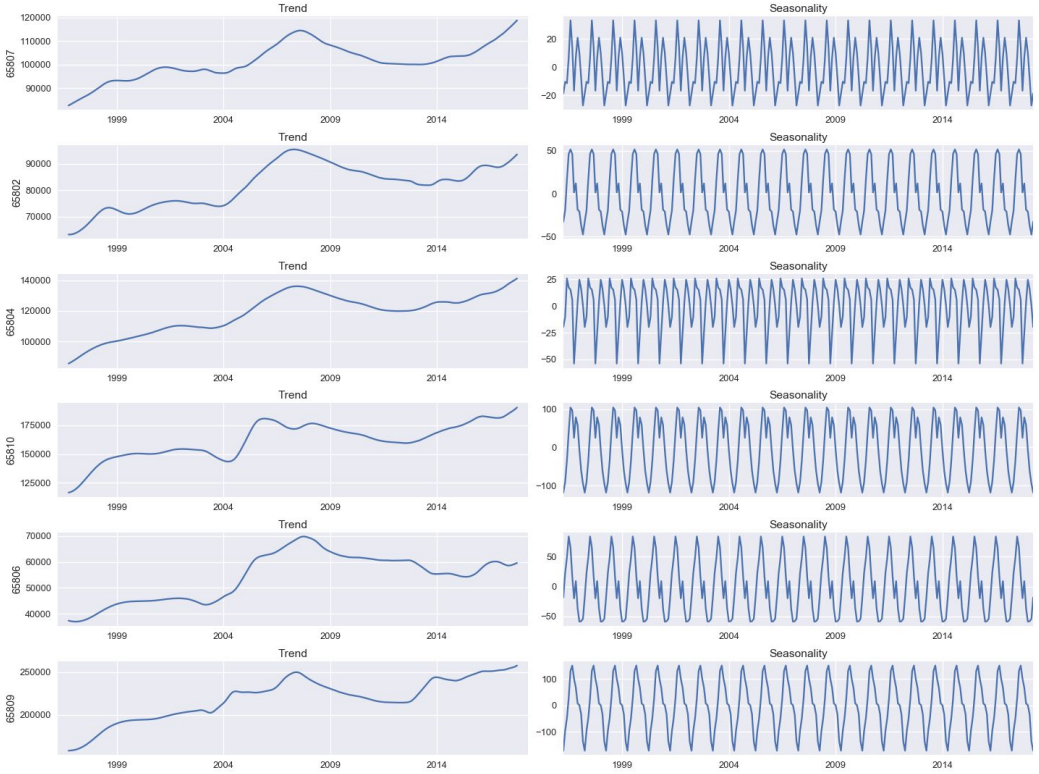


Exploratory Data Analysis

The scale of our seasonality is ±$100.

This accounts for ~0.1% of our values.

We can safely disregard all seasonality concerns.

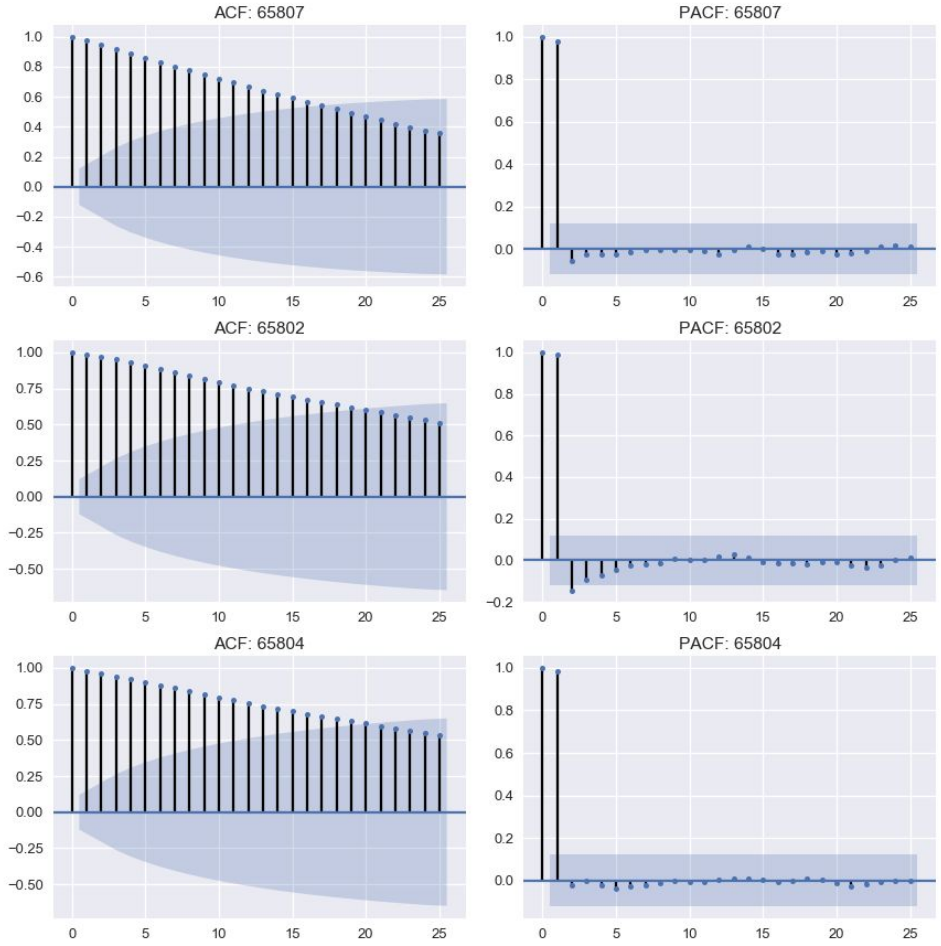


ARIMA

Autocorrelation details how each time series observations is related to the past.

Partial Autocorrelation (PACF) gives us a correlation of a time series with its own lagged values.

These charts will help us determine the starting orders in a ARMA model.



ARIMA

Baseline order (1, 2, 0) determined by ACF and PACF analysis.

cross-validated RMSE values:

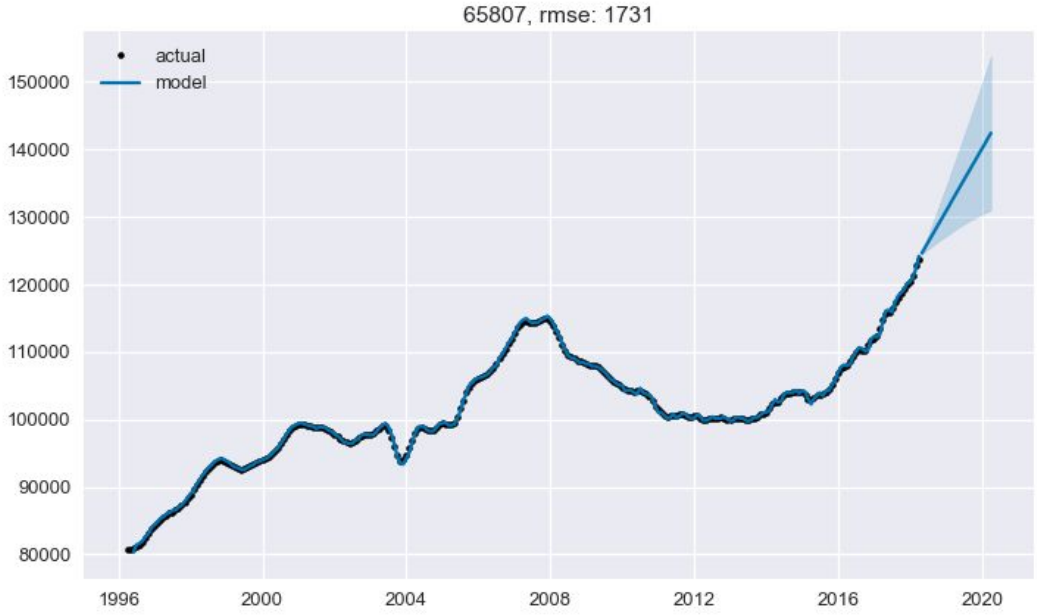
- 65807: 2042
- 65802: 3416
- 65804: 2661
- 65810: 4826
- 65806: 3368
- 65809: 11553

ARIMA

Optimal order determined by auto-ARIMA per model

order and cross-validated RMSE values:

- 65807: (2, 2, 1), 1731
- 65802: (0, 2, 1), 3342
- 65804: (1, 2, 2), 2229
- 65810: (2, 2, 2), 4222
- 65806: (0, 2, 1), 3073
- 65809: (1, 2, 0), 11553

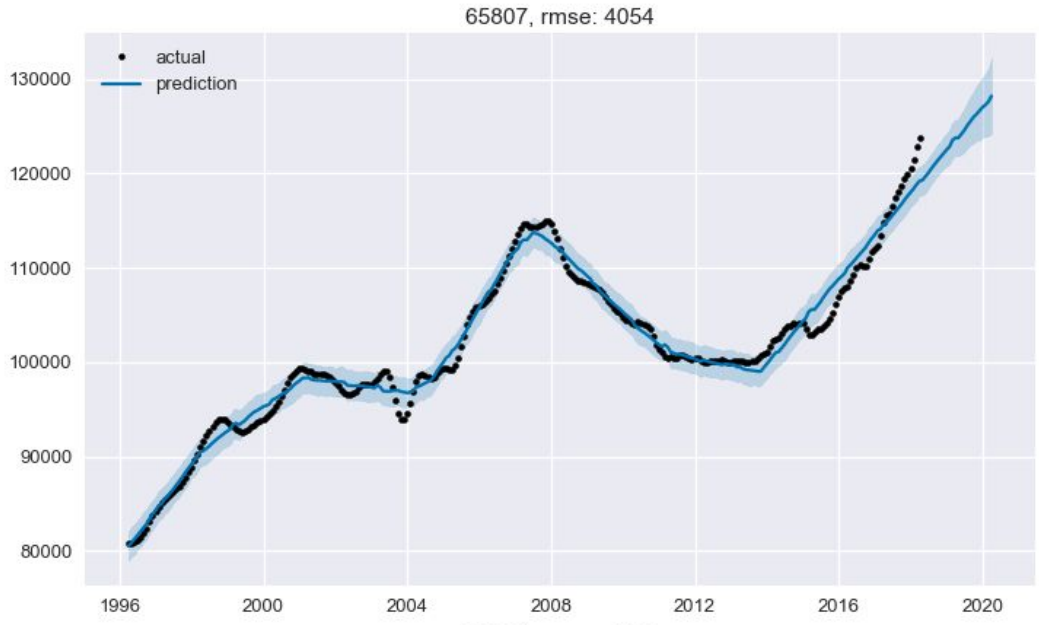


Facebook Prophet

Default parameters used for baseline

cross-validated RMSE values:

- 65807: 4054
- 65802: 3295
- 65804: 4492
- 65810: 7879
- 65806: 3781
- 65809: 14392

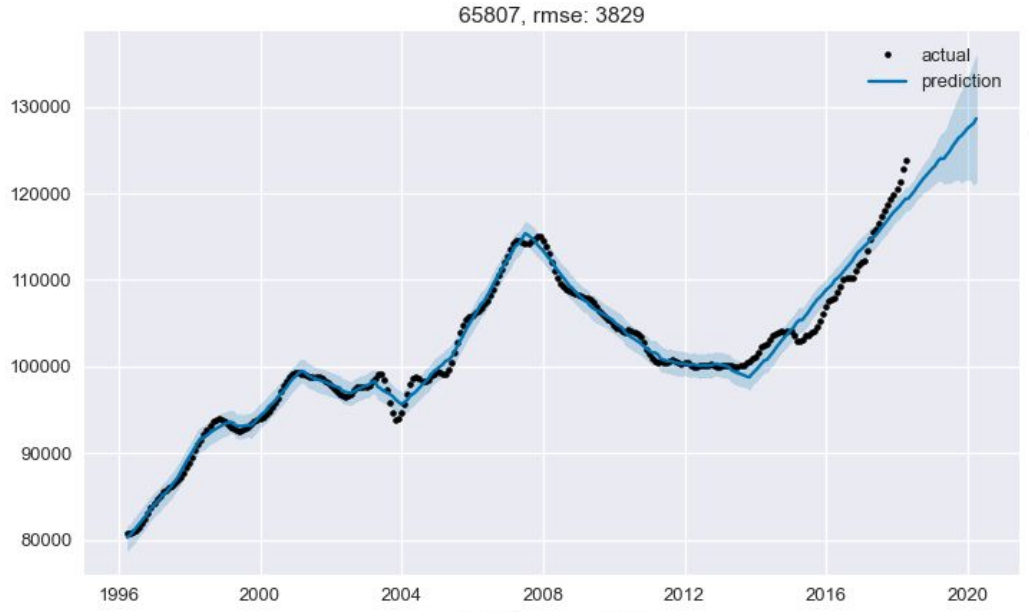


Facebook Prophet

Optimal parameters chosen by gridsearch

cross-validated RMSE values:

- 65807: 3829
- 65802: 3074
- 65804: 4339
- 65810: 6539
- 65806: 3740
- 65809: 13054



Forecast Analysis

The best model for each zip code was chosen based on RMSE.

Zip Code	ARIMA RMSE	Prophet RMSE	Best Model
65807	1731	3829	ARIMA
65802	3342	3074	Prophet
65804	2229	4339	ARIMA
65810	4222	6539	ARIMA
65806	3073	3740	ARIMA
65809	11553	13054	ARIMA

We calculated the slope of our models to help determine which zip code would be the best investment.

- 65809: 69.41
- 65804: 34.19
- 65807: 22.76
- 65810: 11.8
- 65802: 8.21
- 65806: 4.32

All areas of Springfield have potential housing market growth.

65809 has the steepest slope, but also had a significantly higher RMSE.

65804, the second steepest slope with a relatively lower RMSE seems to be promising.

Conclusions and Future Work

65804 is the best zip code in the Springfield, Missouri area to invest in.

The average housing prices in the area will continue to increase over the next two years.

- Our data is outdated by 2.5 years, so we'd like to use more recent data on top of our current data used,
- The ARIMA models appear to overfit and have too-large confidence intervals.
- More recent data could also be used as a holdout dataset, verifying our current conclusions.
- The 2-year forecast range was chosen arbitrarily. We could instead perform an analysis on what range these models are most accurate over, and what time frame is useful in the field of housing investments.

Thank You

Any Questions?

https://github.com/MullerAC/springfield-housing-analysis