

Phase 4 Project

Group 20

**Frank Kiptoo Ruto
Lilian Ngige
Abdihakim Mohamed**



FORECASTING REAL ESTATE PRICES FOR STRATEGIC INVESTMENT



Business Problem



The real estate investment firm aims to maximize ROI by strategically investing in areas poised for significant value appreciation. The challenge is analyzing extensive data to identify these profitable regions.



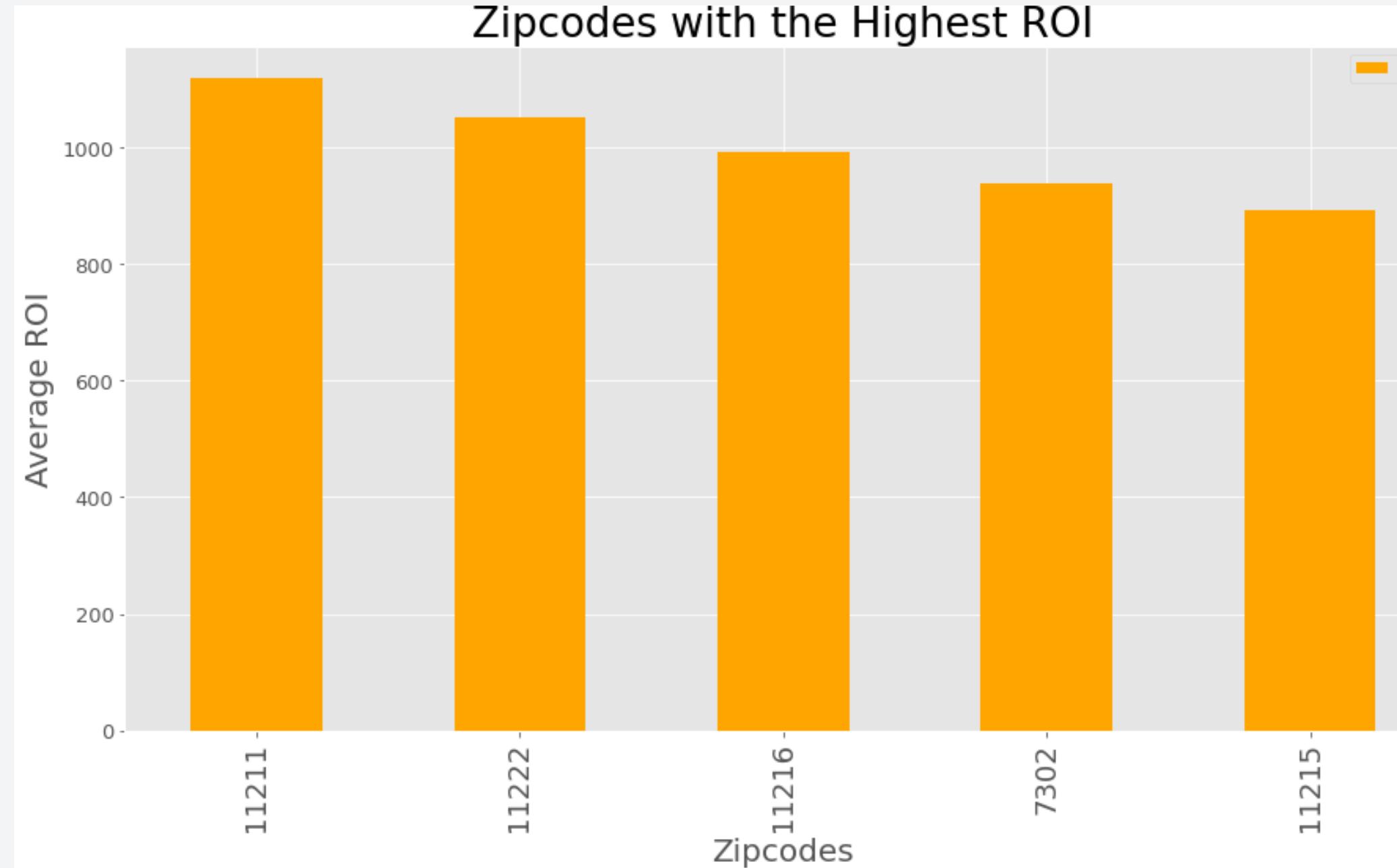
The focus is on gaining insights into future market trends rather than historical data.



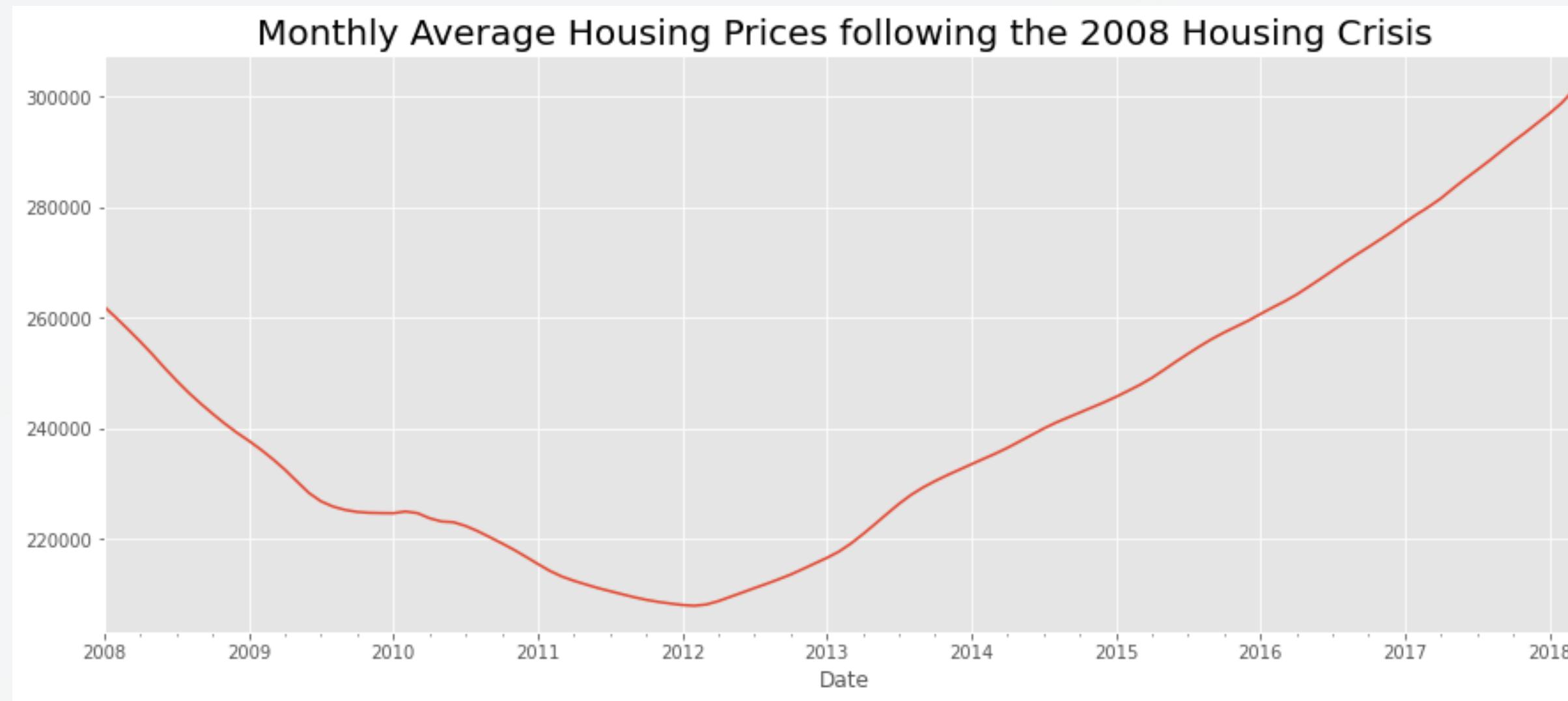


Top 5 Best Zip Codes to Invest in

Zipcodes with the Highest ROI



Time Series Plot showing the monthly Average Housing following the 2008 Housing crisis



Initial Decline

Starting in 2008, there's a clear downward trend in housing prices

Bottoming Out

Around late 2011, the trend in housing prices bottoms out

Recovery and Growth

From 2012 onwards, there is a noticeable upward trend in housing prices

Acceleration in Growth

Around 2016, the growth in housing prices accelerates, as indicated by the steeper slope of the line

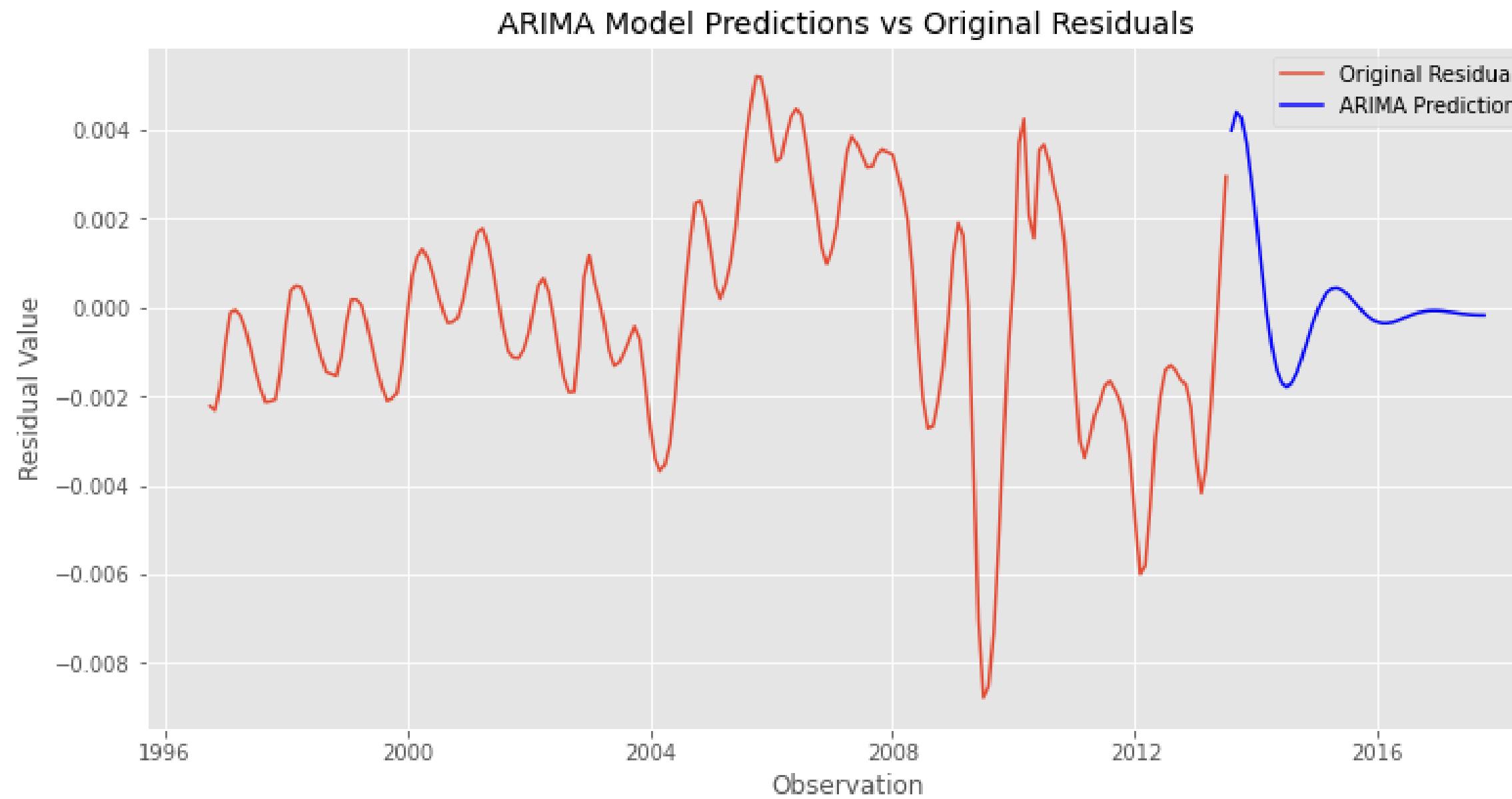
Current Prices Exceeding Pre-Crisis Levels

By the end of the plotted period (2018), housing prices not only recover from their post-crisis lows but also exceed the levels seen at the beginning of 2008.



Modeling & Performance

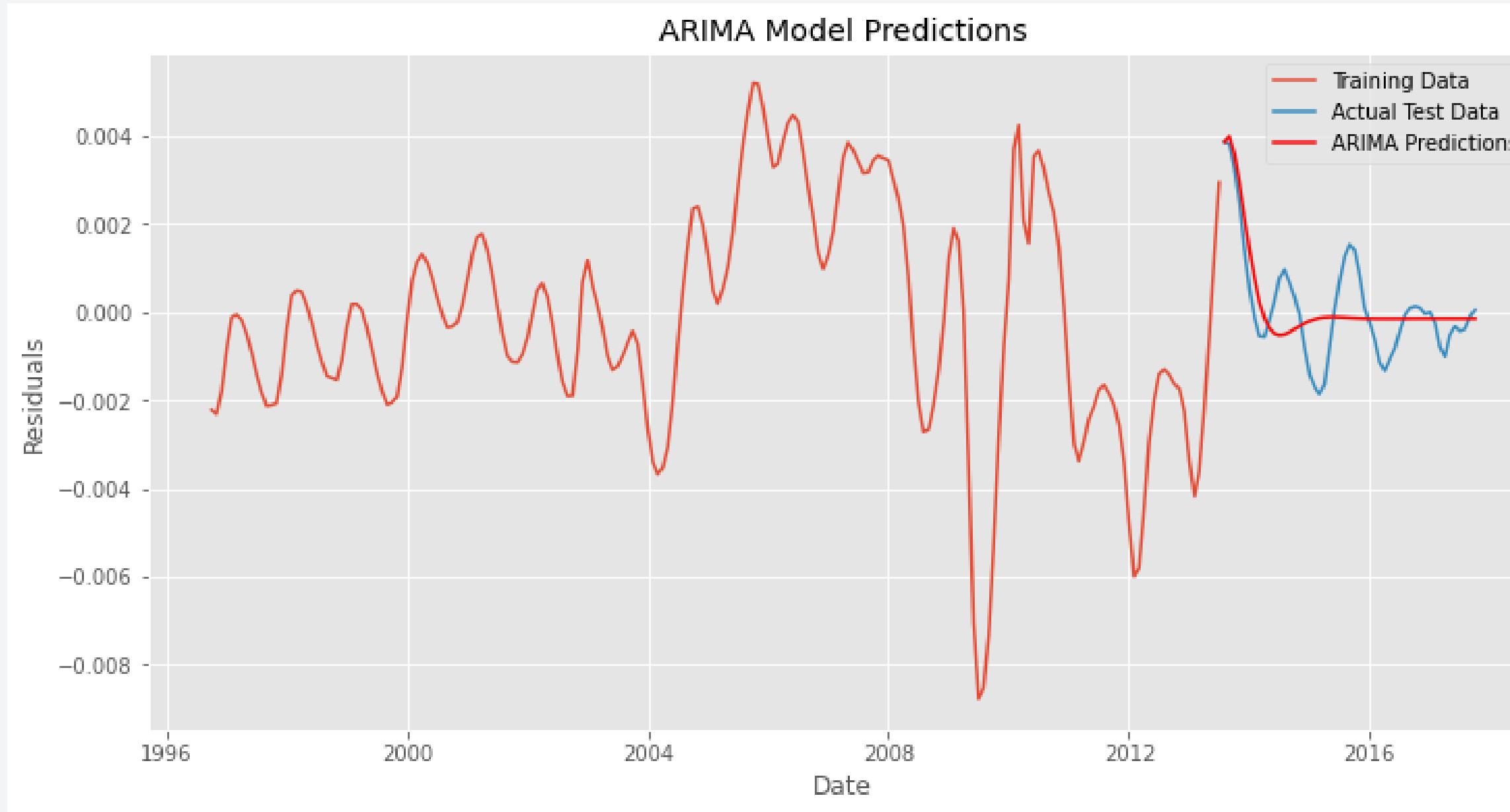
Model 1: ARIMA(2,0) Model



The model seems to fit the data well, as indicated by high log likelihood and low AIC/BIC

Test Data MSE: 1.2113099061625732e-06

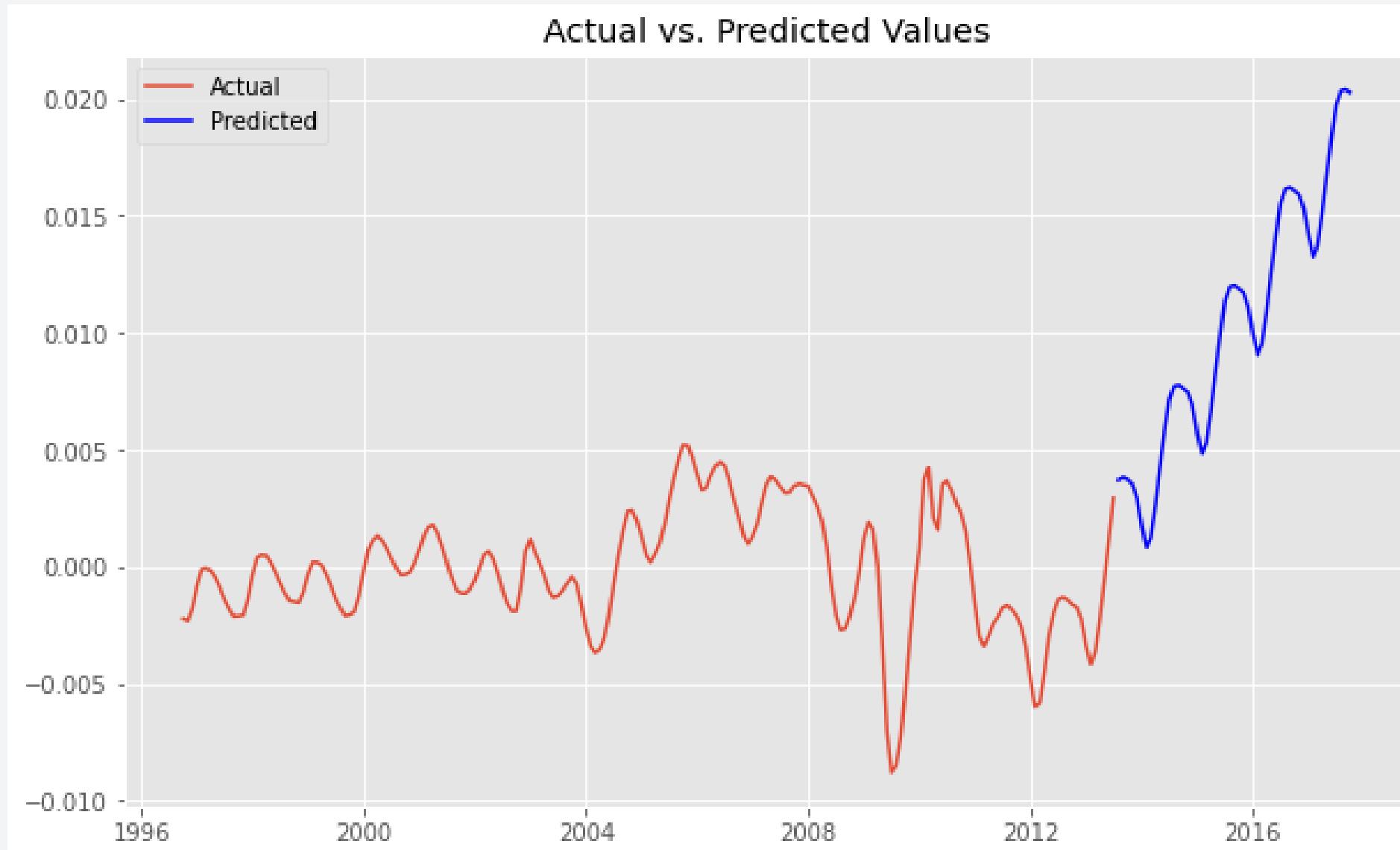
Model 2: ARIMA(2,1) Model



The model captures certain patterns, and similar to the previous model 1, further investigation might be needed to address potential issues such as serial correlation and non-normality.

Test Data MSE: 6.675563129800525e-07

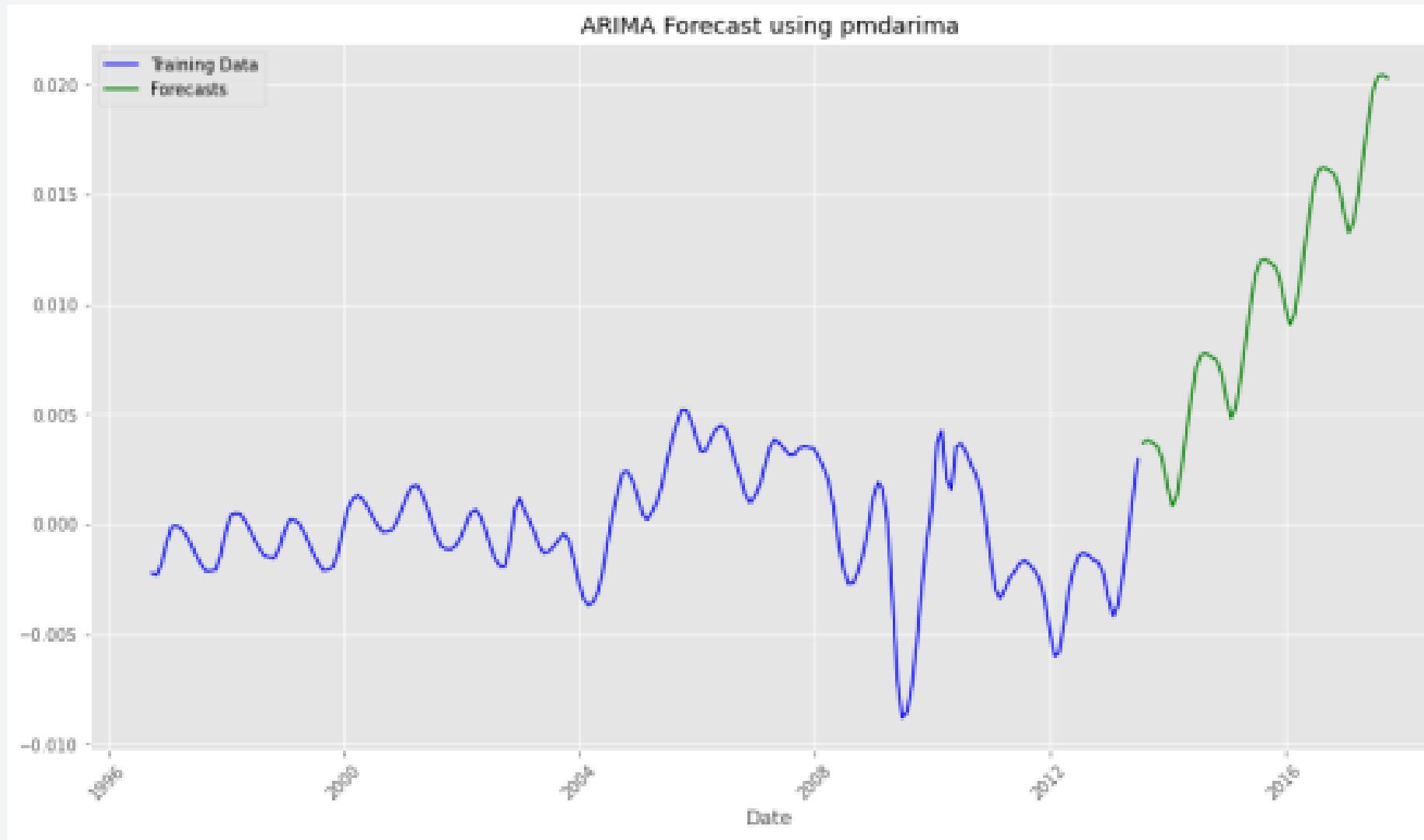
Model 3: SARIMA Model



This model is a SARIMA model, which extends ARIMA by adding seasonal components.

Test Data MSE: 0.00013748286477854483

Model 4: PMDARIMA(auto_arima)



PMDARIMA also performs very well, almost as good as Model 2

Test Data MSE: 7.125702038342685e-07t

Evaluation

1. ARIMA(2,0) Model:

- MSE: 1.21e-06
- Solid performance in forecasting.

2. ARIMA(2,1) Model:

- MSE: 6.68e-07
- Improved accuracy with first-order differencing.

3. SARIMA Model:

- MSE: 0.00014
- Higher MSE suggests challenges in capturing patterns or potential overfitting.

4. PMDARIMA (auto_arima) Model:

- MSE: 7.13e-07
- Competitive performance, aligning with Model 2.



Conclusion

Based on the MSE values:

The best performing model is Model 2: ARIMA(2,1), with the lowest MSE.

This suggests that the model is best at capturing the underlying data patterns among the models tested.

Model 2 is expected to have more accurate predictions on unseen test data compared to the other models.

In summary, for this specific dataset and based on the MSE evaluation metric, the ARIMA(2,1) model is the most suitable choice.

Recommendations

The top 5 zip codes share certain characteristics that yield attractive risk adjusted return.

- Healthy employment and strong economics anchored by the higher than usual opportunity in these high tier cities.
- Proximity to amenities such as parks, entertainment, or transit.
- Population growth fueling housing demand.

We would recommend that future investment consideration adheres to these characteristics. Together, they create strong demand that results in efficient ROI. Further analysis that uses employment, income growth, and population growth may reveal other lucrative zip codes.^{Inc.}



Limitations

Property values are influenced by more than just time; they are also affected by the balance of supply and demand.

Next steps in this project include added exogenous factors to our model:

- Interest rate
- Crime data
- Median income over time
- Population growth
- Experience
- Mortgage rates

The Zillow dataset used for this analysis had its most recent data points in April 2018.

Finding more recent data will allow us to more accurately forecast future home value into 2025 and beyond.



THANK YOU

