Real Estate Investment

Top Zip Codes in United States – 5 Year Return on Investment

A Time-Series Data Science Project

Jim Petoskey

Flatiron School

Phase 4

Outline

- Business Opportunity
- Summary
- Data
- Methods
- Results
- Conclusions
- Further Study

Business Opportunity

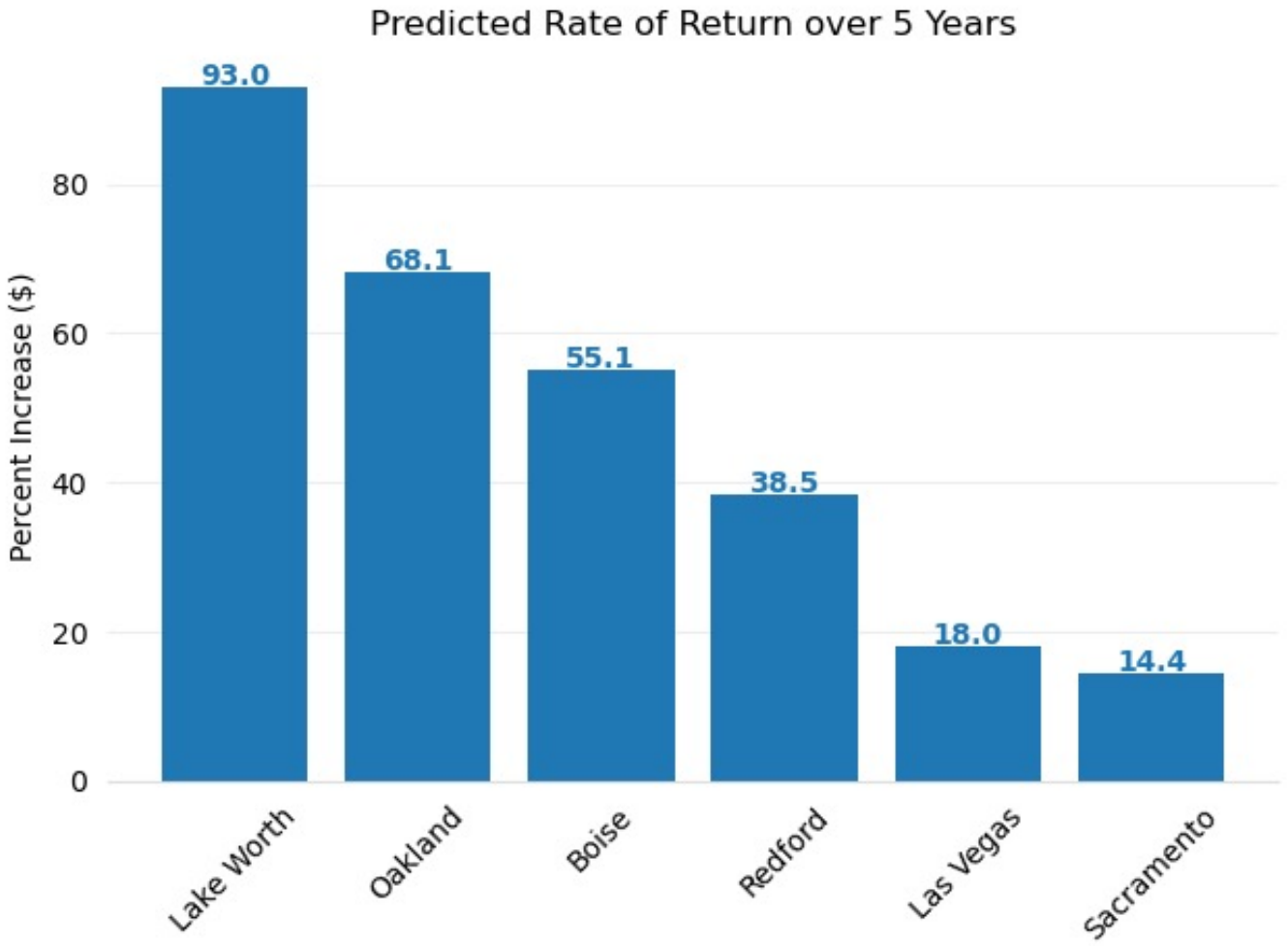
- High rate of return likely over the next 5 years.
  - Models for top zip codes predict price within 3% of market value
  - Market Value Increase of 55% - 93% for top zip codes
  - Options for higher or lower risk investments

Summary

- Top 15 zip codes by rate of return from April 2012 – April 2018 were modeled, in addition to 20 zip codes from 5 Mountain Towns and 5 Mid-Size Cities.
  - Boise is the only city added from Mountain Towns and Mid-Size cities due to high return and low risk.

- Highly accurate modeling.
  - Each model in this visualization predicted median price within 3% of market value.

  - Except for Boise, which predicted price within 8% of market value.



Data

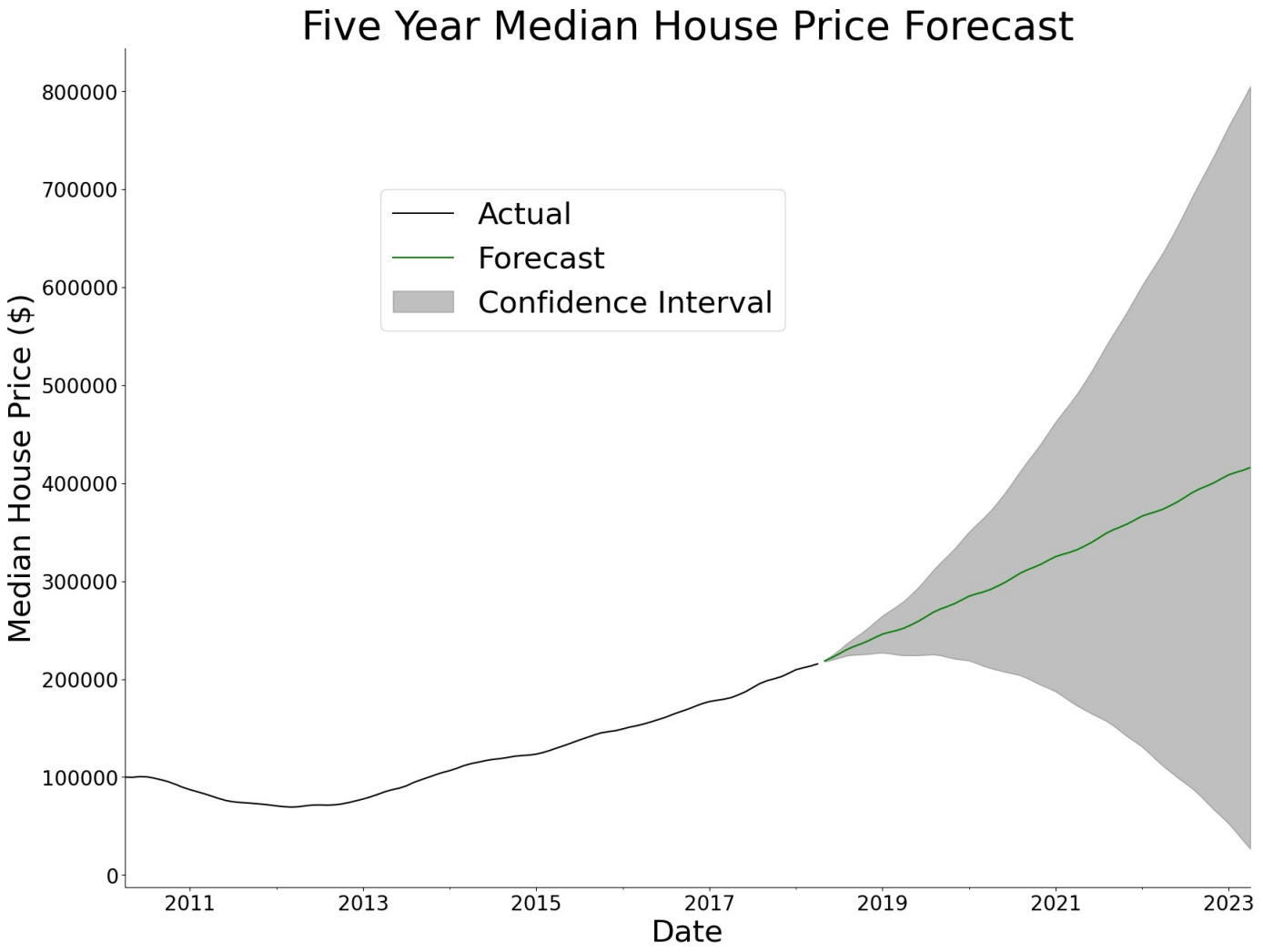
- Represents 14,723 zip codes, including all 50 states in the US
- Contains median price of homes in each zip code on a monthly basis from April 1996 to April 2018.
  - Some zip codes contain missing data – one of these zip codes was removed from the top 15 zip codes by rate of return from April 2012-2018.
- Initially, stored in wide format, converted to long format with time-series index for modeling purposes.

Methods

- SARIMA model used – currently best model for time series
  - Best parameters for model determined for each zip code individually.
- RMSE ratio to median market value calculated to determine accuracy of models.
  - Models for each zip code can reliably be compared to each other to determine which zip code is being modeled more accurately.
- Model Forecast predicts price of homes in 5 years.
  - Percent Market Value Increase and Confidence Interval calculated and plotted in visualizations for recommended zip codes.

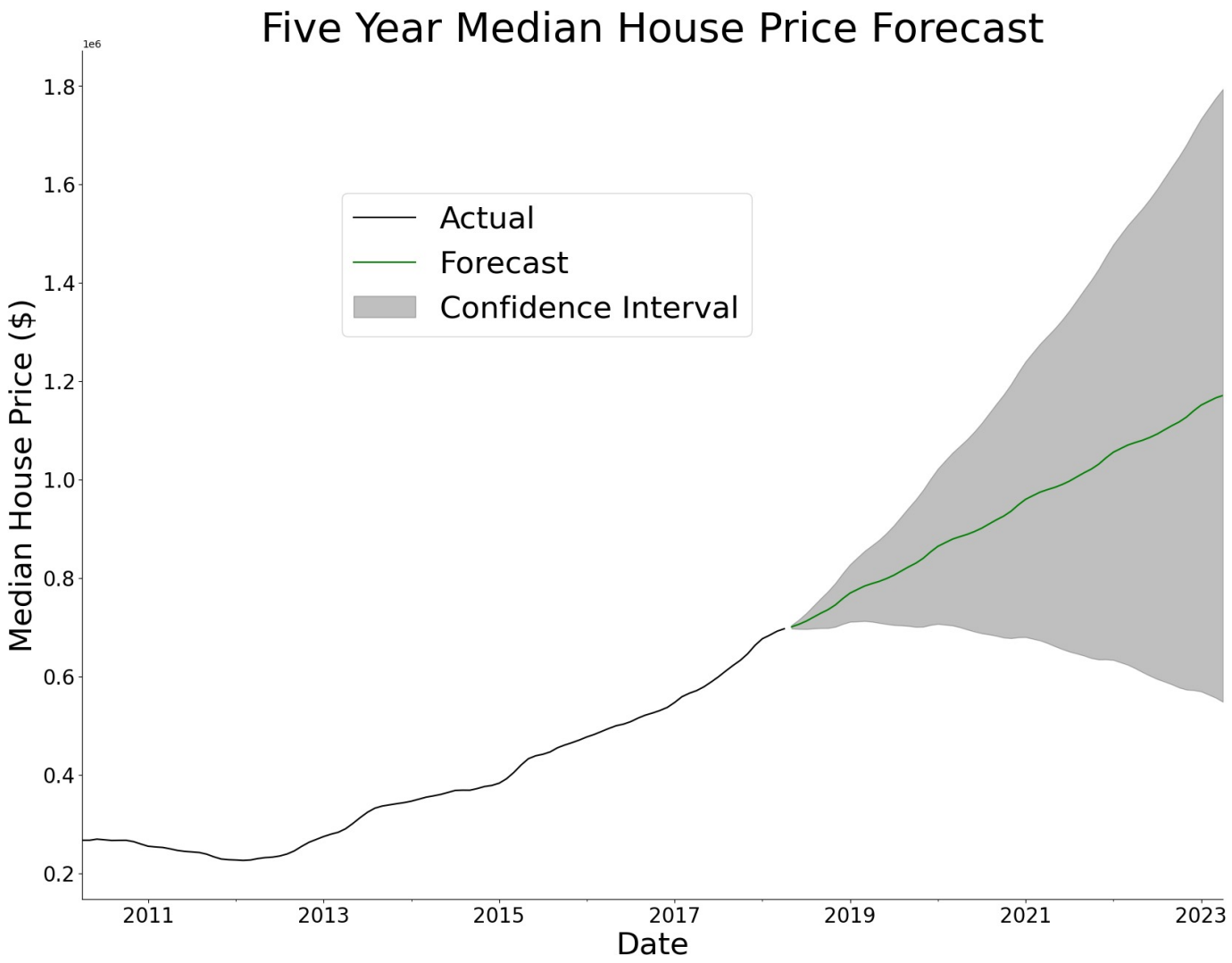
Results - Lake Worth, FL 33460

- Highest predicted return of 93.0%
- High risk with wide range and large negative potential for the confidence interval price at 5 years
- High level of model accuracy:
  - Predictions within 1.3% of median value on average.



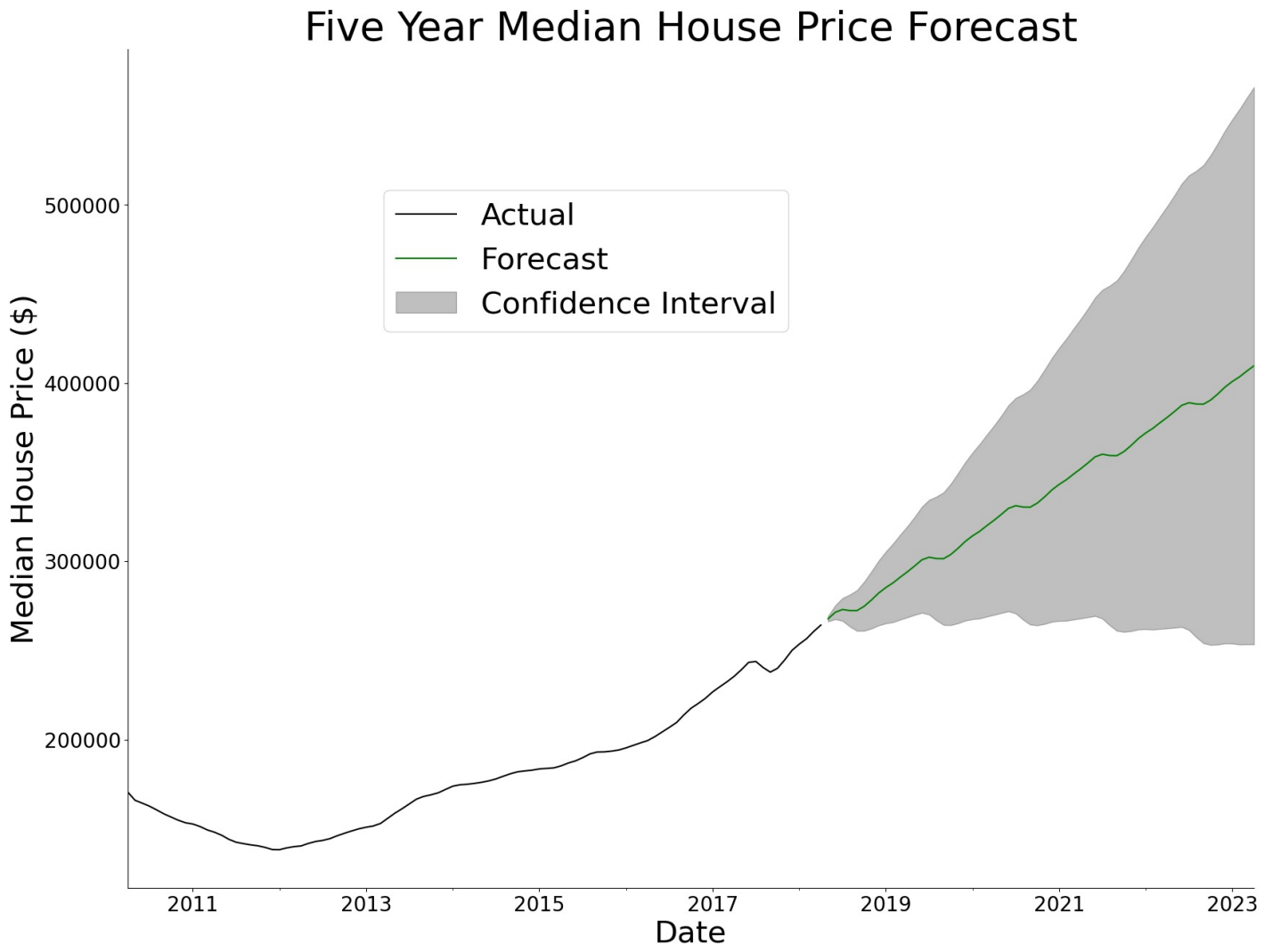
Results - Oakland, CA 94606

- Has a predicted rate of return of 68.1% over the next 5 years (assuming we are in April of 2018 when this dataset was current)
- Has the lowest margin of risk - with the lowest confidence interval value being substantially higher than the models for other zip codes.
- High level of model accuracy:
  - Predictions within 3% of median value on average.



Results – Boise, ID 83703

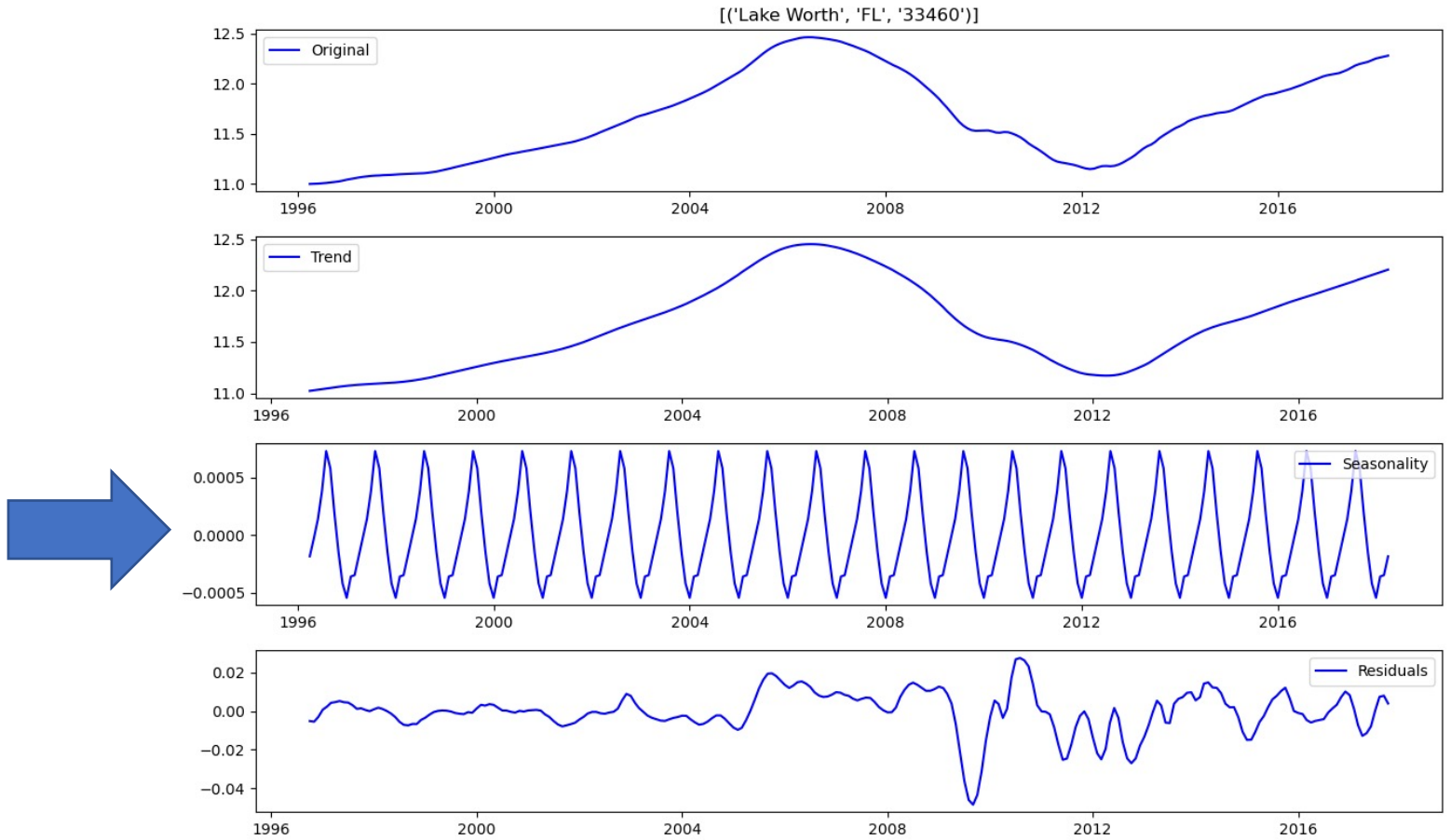
- High predicted return of 55.1%
- Low risk, with lowest confidence interval value being substantially higher than the models for other zip codes.
- Moderate level of model accuracy:
  - Predictions within 8% of median value on average.



Results – Seasonal Pattern

Seasonal Pattern

- Not visible in raw data, so important to note that seasonality is present in every zip code and prices are higher in the summer and lower in the winter.



Conclusions

- Allocate partial investment to all three recommended zip codes.
  - They offer a greater likelihood of strong returns when all are part of the investment portfolio.
- Optimize for highest returns, higher risk:
  - Allocate greater percentage of investment to Lake Worth
- Optimize for lower risk, still high returns:
  - Allocate greater percentage of investment to Boise and Oakland
- Note that if units of investment are of concern:
  - Lake Worth and Boise have lower overall cost per house, so more houses will need to be purchased to obtain the same return

Further Study

- **Adding to the model**
  - Need more recent data
  - Add external influences with more complex models, such as:
    - Average interest rates for mortgages
    - Overall employment, nationally or locally
    - Percent Increase or decrease in jobs in the county of that zip code
- **Ongoing testing**
  - Need monitoring and testing of price changes in real-time
  - Increase frequency of median house sale measures, if possible. (i.e. weekly)
- **Additional zip codes**
  - Ensure big cities, such as New York, Los Angeles, and Chicago are represented.

Q & A Time

- Jim Petoskey – Data Scientist
  - Jim.Petoskey.146@gmail.com
  - Phone: (734) 649-3182
  - linkedin.com/in/jimpetoskey/