

 [jpetoskey](#) / [Zillow_Time_Series](#) Public

Identify which zip codes provide the highest rate of return with moderate risk for real estate investment.

☆ 0 stars  0 forks

☆ Star

👁 Unwatch ▾

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

 main ▾

...



jpetoskey ...

20 seconds ago 

[View code](#)

 README.md



Zillow_Time_Series

Overview

- Data from: [Time Series Zillow Starter Repository](#)

Identify which zip codes in the US provide the highest rate of return with low to moderate risk for real estate investment.

This is a time-series data set and SARIMA Models were implemented to predict and forecast median price of houses in selected zip codes.

Business Opportunity

- High rate of return likely over the next 5 years.
 - Models for top zip codes predict price within 3% of market value

- 14,723 zip codes, representing all 50 states and Washington DC have been analyzed for the best returns over the most recent 6 years in the data set (April 2012-2018)
- In addition, top mid-size cities and mountain towns, which have become more attractive since the covid pandemic, were modeled.

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

Investment recommended in the following zip codes:

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.

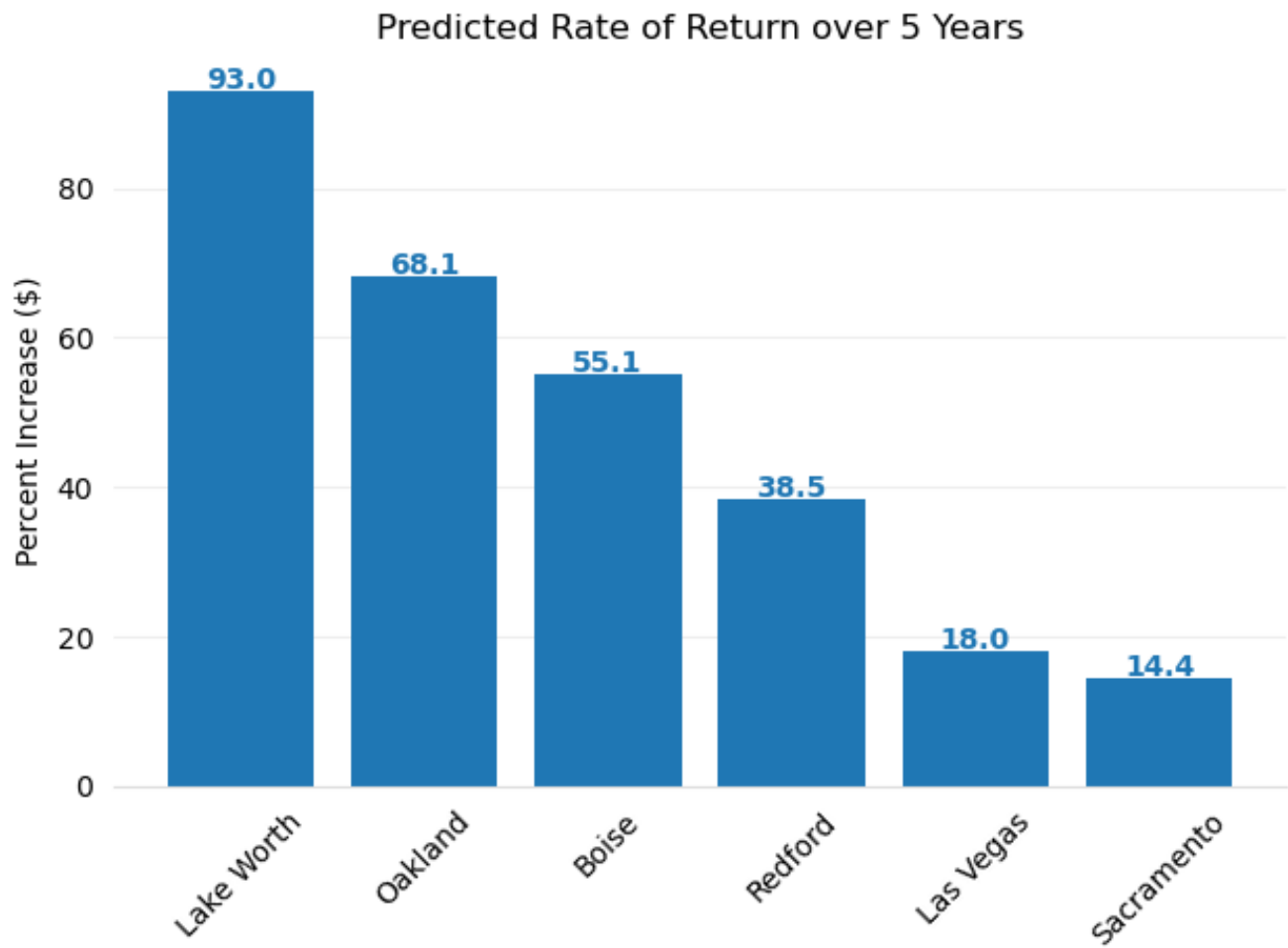
Lake Worth, FL 33460

- Highest predicted return of 93.0%
- Much higher 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.

Boise, ID 83703

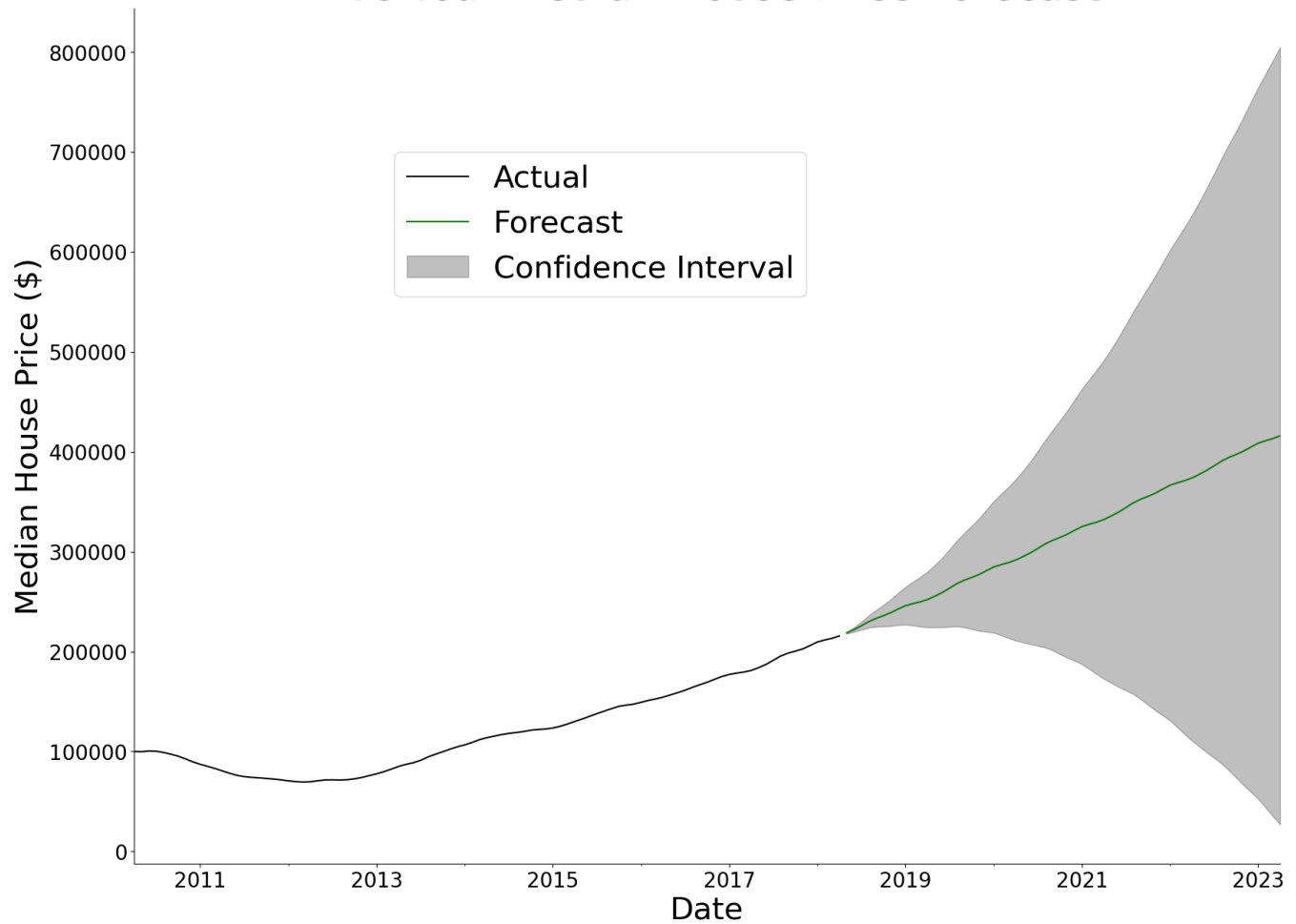
- Fairly high predicted return of 55.1%
- Low risk, with lowest confidence interval value being substantially higher than the models for other zip codes.
- Moderate level of model accuracy:
 - Predictions within 8% of median value on average.

Visual 1



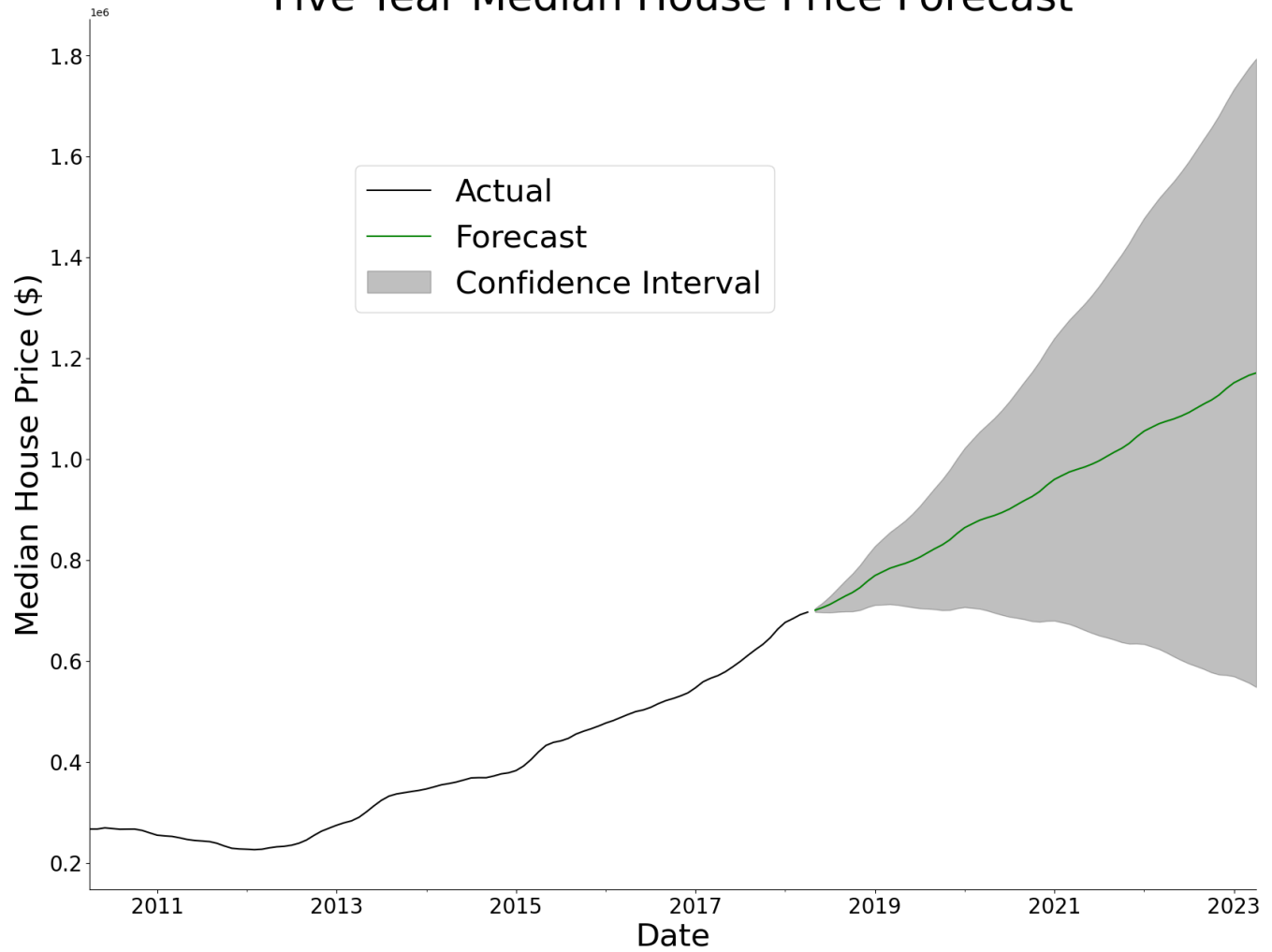
Visual 2

Five Year Median House Price Forecast



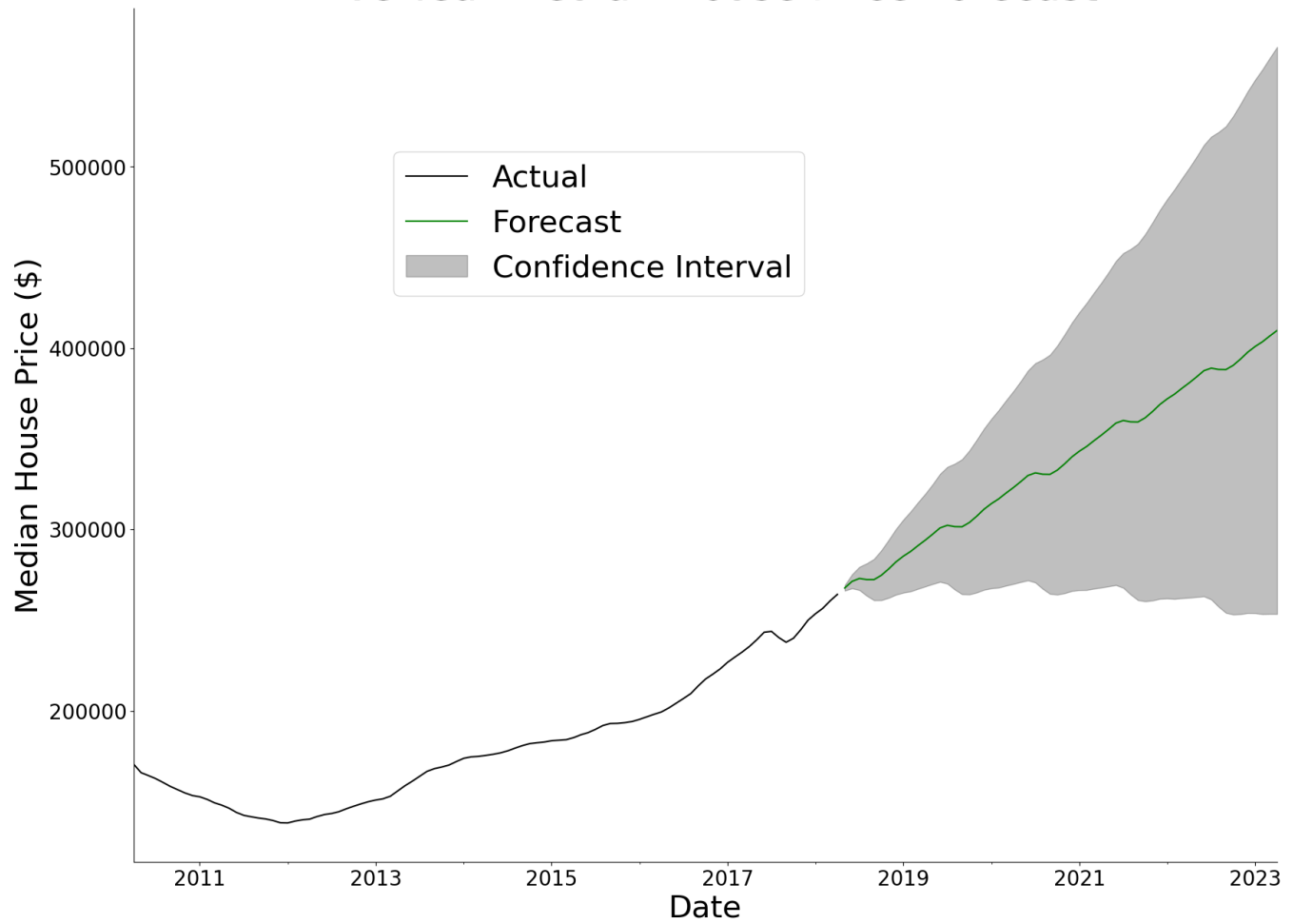
Visual 3

Five Year Median House Price Forecast

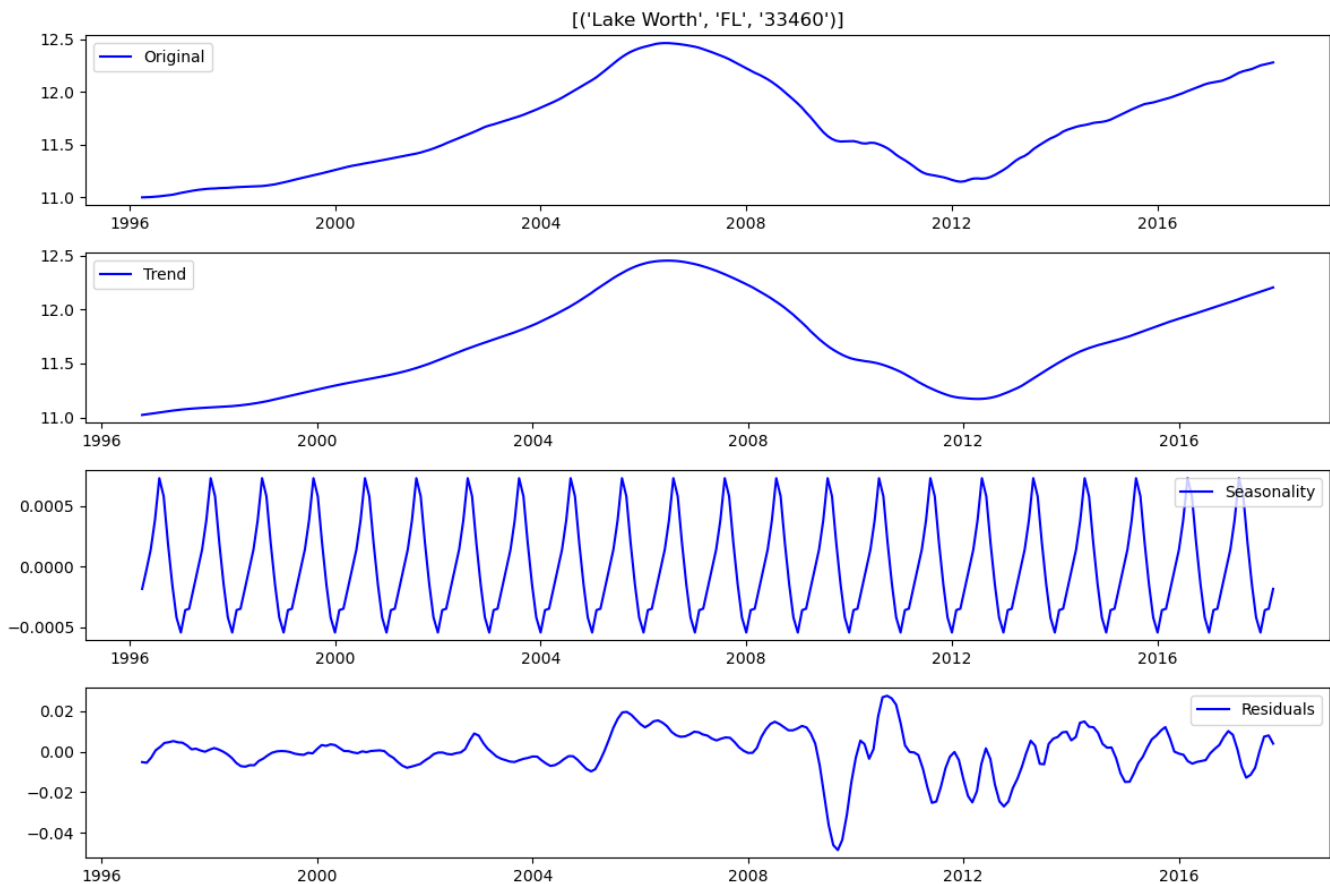


Visual 4

Five Year Median House Price Forecast



Visual 5



Conclusion

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 Studies Recommended

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.

For More Information

Please review my full analysis in my [Jupyter Notebook](#) or my [presentation](#).

For any additional questions, please contact **Jim Petoskey** - Jim.Petoskey.146@gmail.com

Repository Structure

— README.md	<— The top-level README for reviewers of this project
— Real Estate Investment - Top Zip Codes.ipynb	<— Narrative documentation of analysis in Jupyter notebook
— Preprocessing and Practice - Zillow Time Series with SARIMA.ipynb	<— Extended First Draft with focus on data exploration
— Real Estate Investment(...)pdf	<— PDF version of project presentation
— zillow_data.csv	<— Median House Price Data - Monthly - Zillow
— images	<— Generated from code with matplotlib

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● **Jupyter Notebook** 100.0%