Time Series Analysis on NYC Housing Data

By: Wesley Yu

House Price Forecasting

Business Problem

- NYC, the most populous city in the US

- The top 5 regions that provide the best opportunities for investing in real estate

Data

- Zillow Home Value Index for 106 zip codes in NYC, excluding Manhattan

- Reflects median house price of each region

- Contains monthly records from April 1996 to April 2018

Methodology

Criteria for best investment opportunity:

- Identify **Growth** and **Risk**

- Select regions with **high growth** and **low risk**

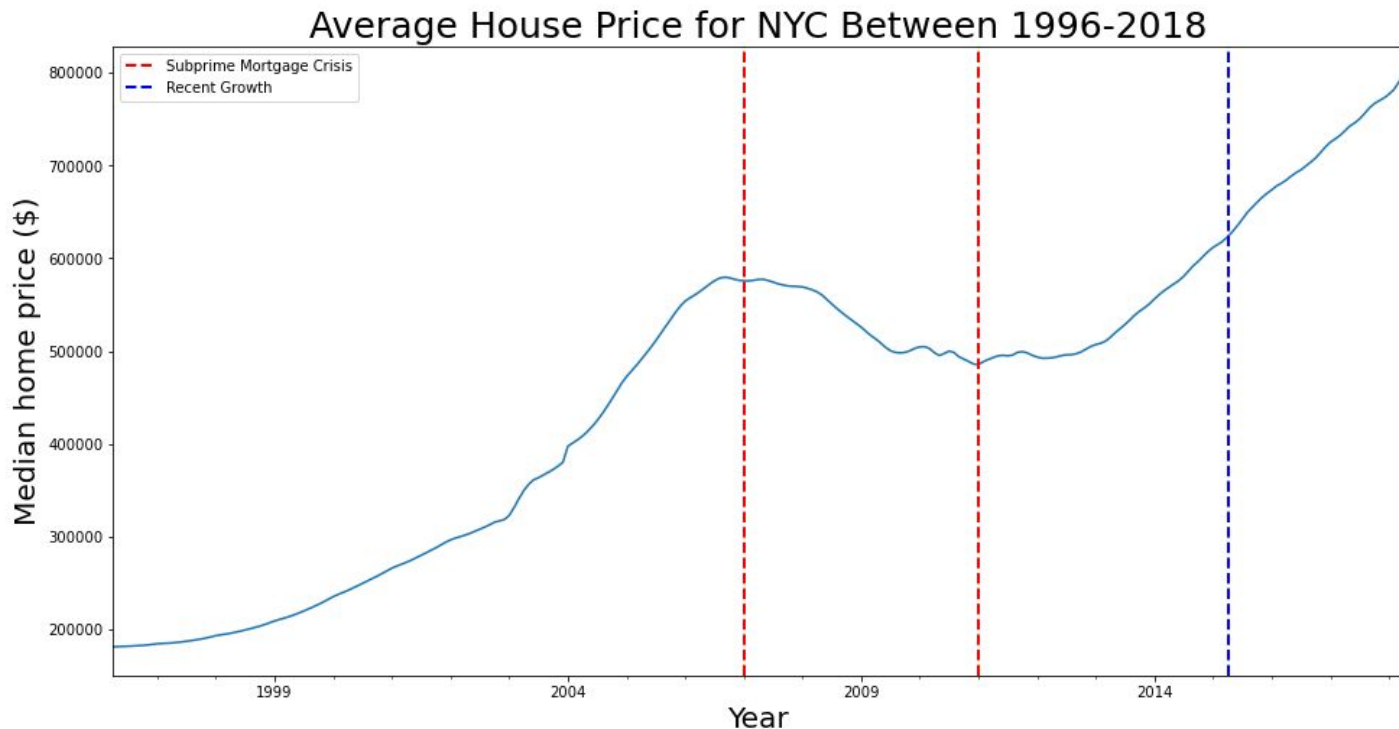- Assess forecast performance of ARIMA model for selected regions

Growth and Risk

Growth

28% average growth from past 3 years

Risk

17% average decline during crisis

Average House Price for NYC Between 1996-2018

Median home price ($) vs Year

Legend:
- Subprime Mortgage Crisis (red dashed)
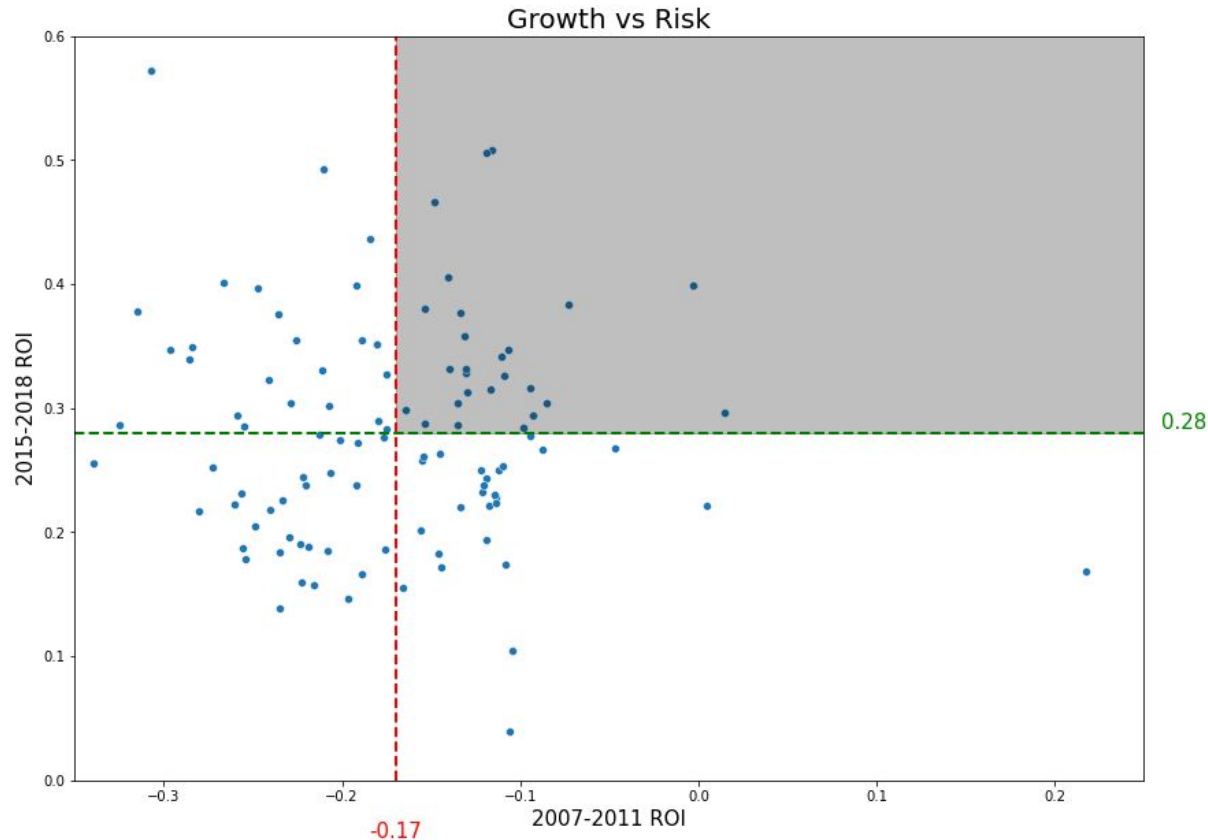- Recent Growth (blue dashed)

Selecting Regions

Regions selected based on combination of:

- High Growth
- Low Risk

High Growth:
Regions with ROI greater than 28% during 2015-2018

Low Risk:
Regions with ROI greater than -17% during 2007-2011

Growth vs Risk

Top Performing Regions

- 10 regions with highest growth while maintaining lowest risk

- Growth ROI in the top 35%

- Risk ROI in the top 25%

Zip Code	Growth ROI	Risk ROI
11211	50%	-11%
11102	50%	-11%
11238	39%	-0.2%
11222	38%	-7%
11694	34%	-10%
11229	34%	-11%
11105	32%	-10%
11224	31%	-9%
11354	31%	-11%
11355	30%	-8%

Forecasting Performance

Forecast Comparison of 11355 May 2017 - April 2018
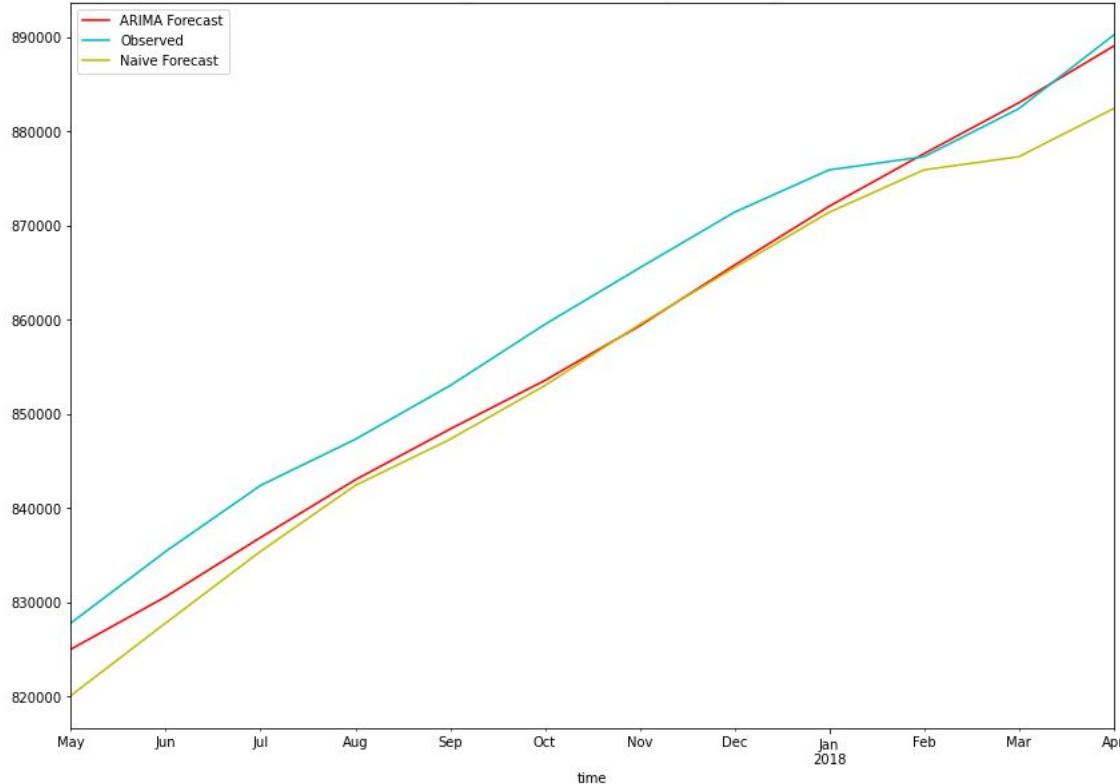
- Naive forecasts assumes previous actual price as next predicted value
- Forecast performance measured by comparing errors of both models
- Lower error indicates better performance

Naive Error: $6,087

ARIMA Error: $4,310

Results

- ARIMA model was able to outperform naive model in 3 regions

- Top 5 regions selected by lowest error margin between ARIMA and naive forecasts

Zip code	Naive Error	ARIMA Error	Error Margin
11354	$8,132	$5,819	-$2,313
11355	$6,087	$4,310	-$1,777
11694	$7,198	$7,075	-$123
11222	$15,125	$21,464	$6,339
11229	$6,870	$13,235	$6,365

Forecast Details
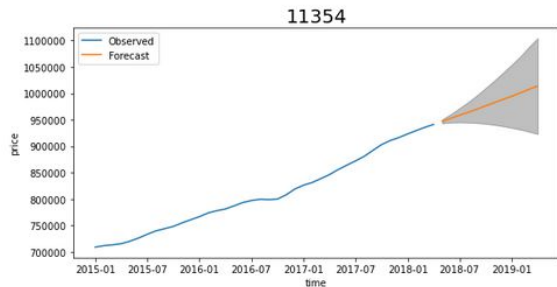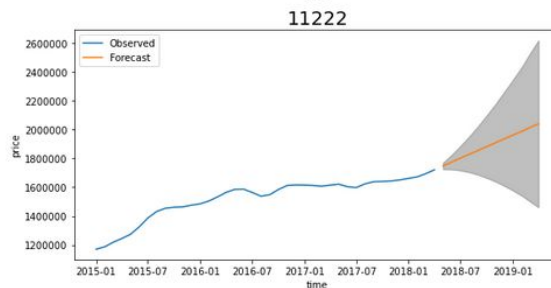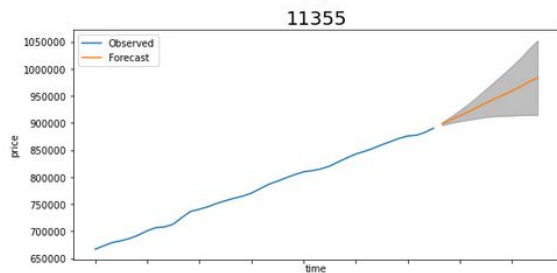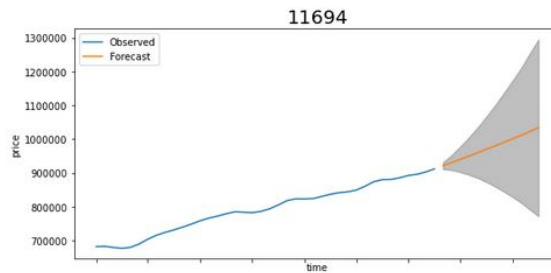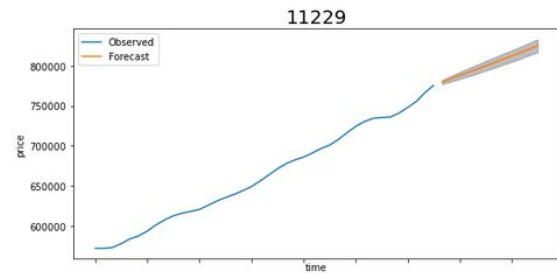
- Forecast horizon May 2018 to April 2019

- Upward trend in prices for all 5 regions

- Shaded area represents upper and lower bounds of forecasts, can be considered as risk assessment

Recommendations

Risk Level	Zip Code	Cost	1 Year Growth	Risk
Low	11229	$775,500	6.9%	±0.9%
Moderate	11355	$890,200	10.5%	±7.7%
	11354	$941,200	7.7%	±9.6%
High	11694	$912,100	13.3%	±28.7%
	11222	$1,721,100	18.6%	±33.7%

Conclusion

Limitations

- Forecasts based on past prices
- External factors can affect prices

Next Steps

- Test out other modeling methods
- Include external factors such as income, mortgage rates, inflation

THANKS!

For additional questions or comments please contact me below:

Email: to.wesleyyu@gmail.com

Github:
https://github.com/whyyoutoo/Forecasting-House-Price-Time-Series