Nassau County Time Series Analysis

ZILLOW REAL ESTATE

AUTHOR: KYUNGHWAN WILLIAM KIM

Business Problem

o Our clients are a Real Estate Investment Firm looking to invest in Long Island New York, specifically Nassau County, with a budget of $1,000,000.00.

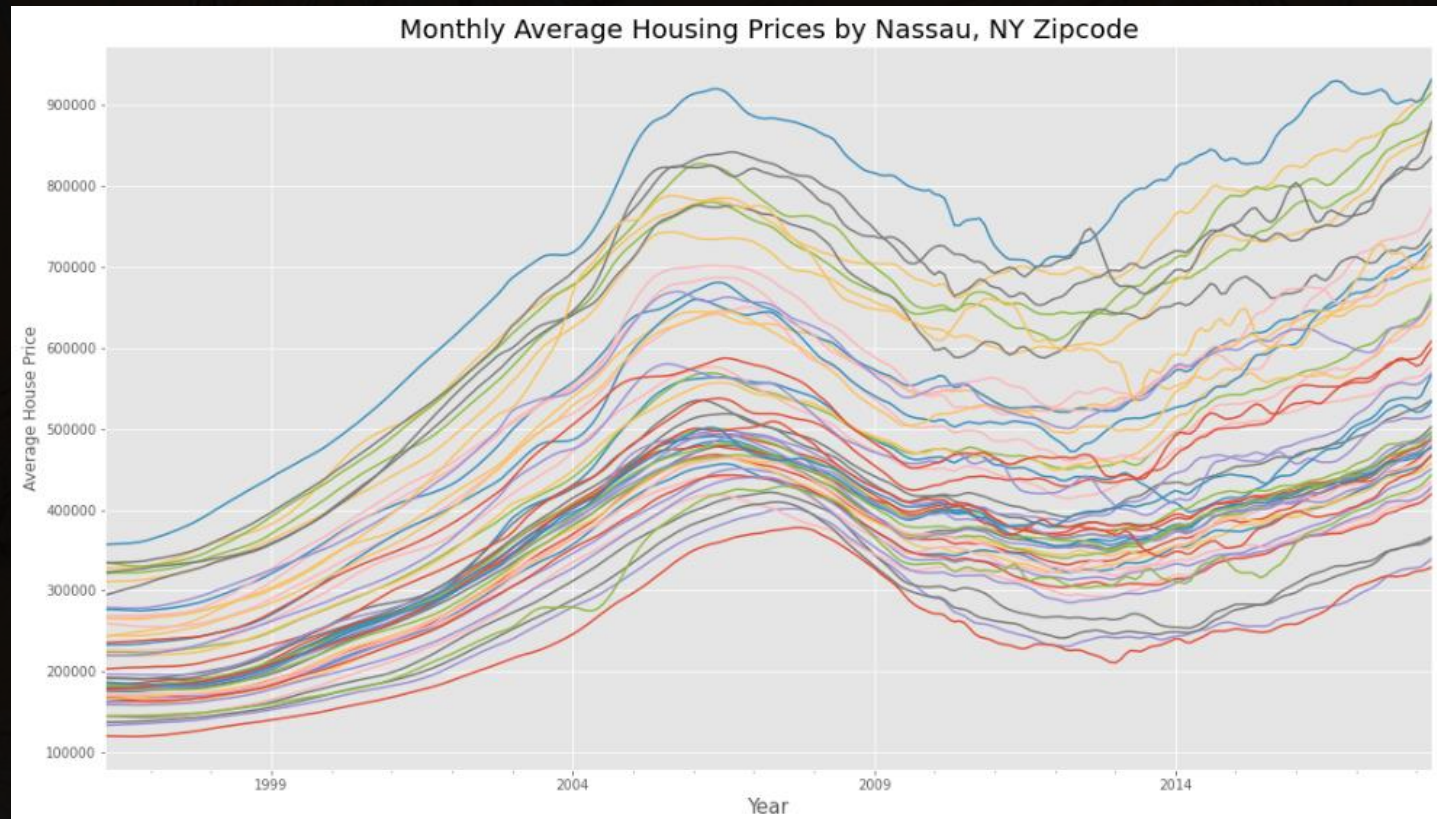o Our task is to conduct a time series analysis to predict the 5 best zip codes to invest in based on ROI.

OUR DATA

o The Dataset consists of Housing Prices in the United States from the years 1996-2018, with our frequency being Monthly Mean Values.

o Dataset has 14,723 Zip codes

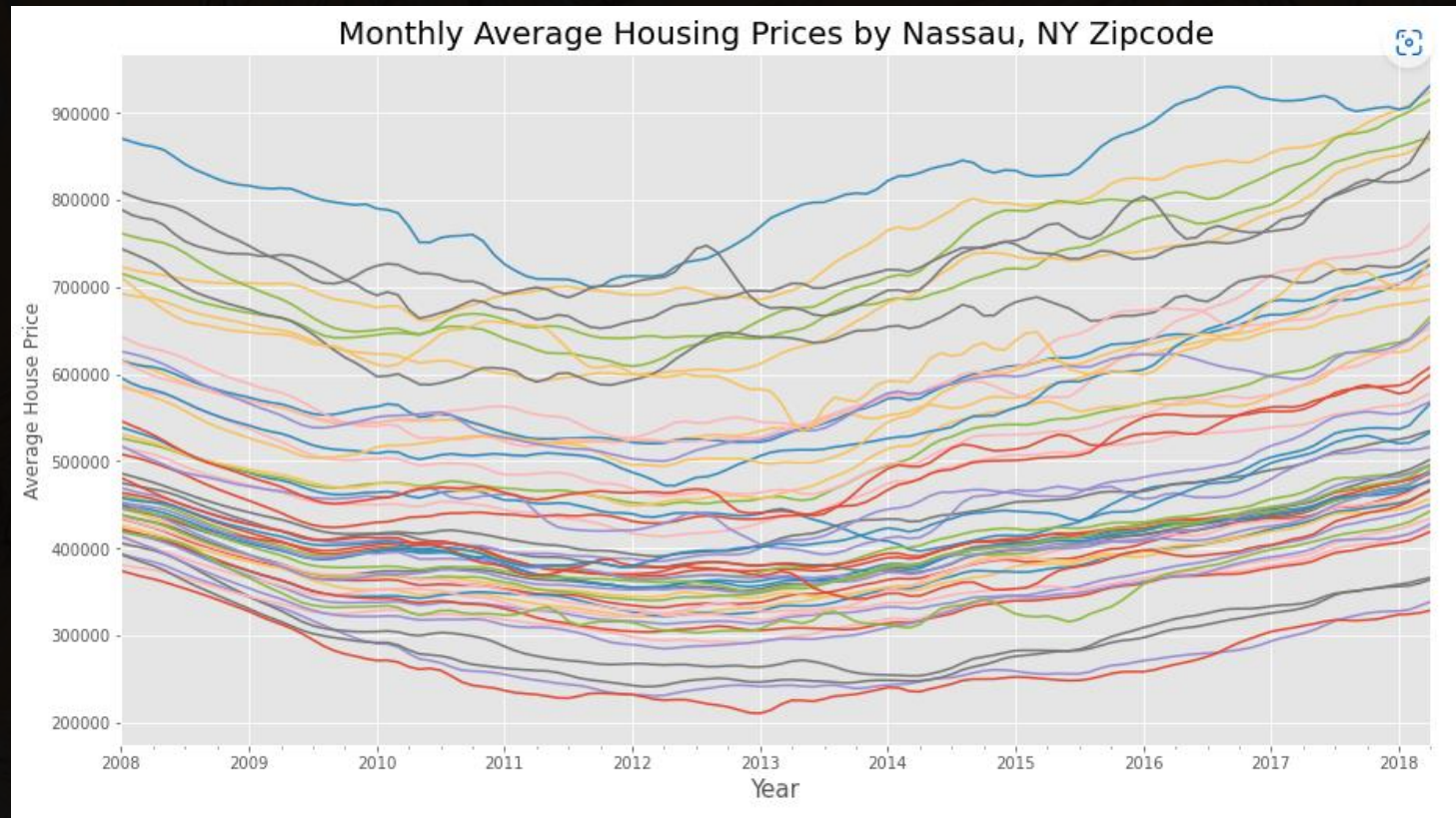o Due to the 2008 Housing Crisis, we decided to focus on 2012 and onwards.

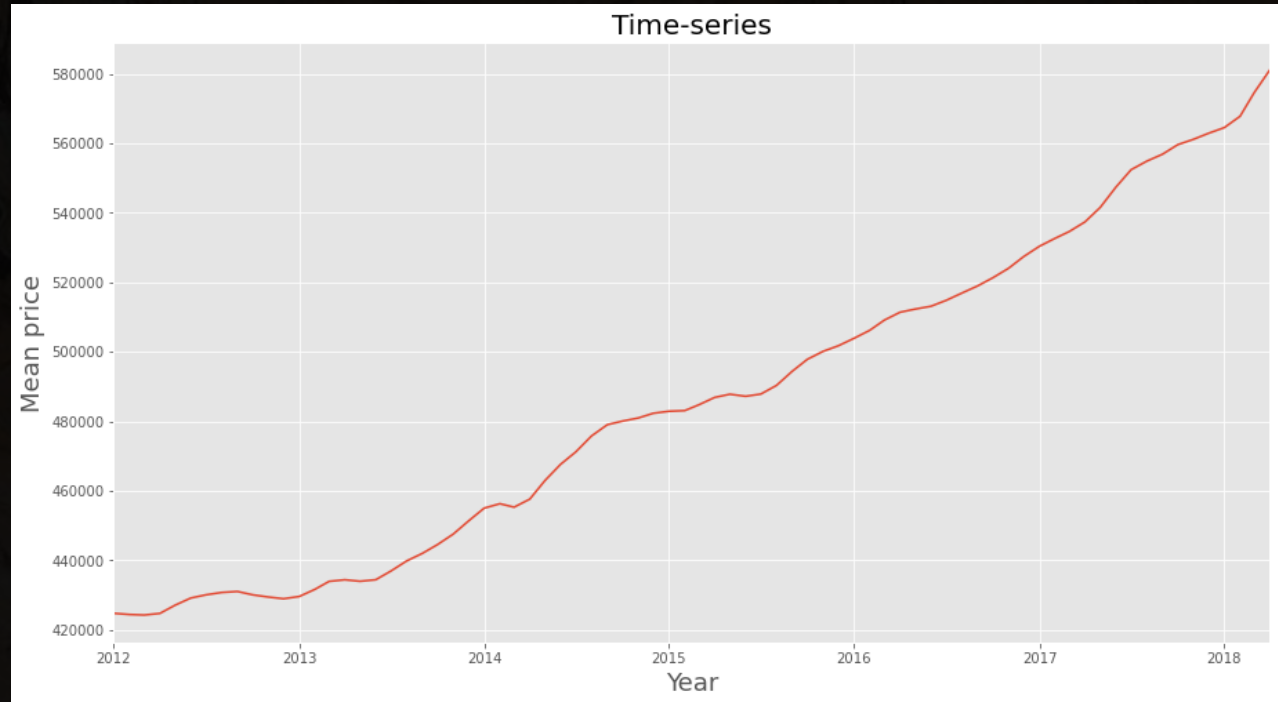# Price Distribution after Budget Filter
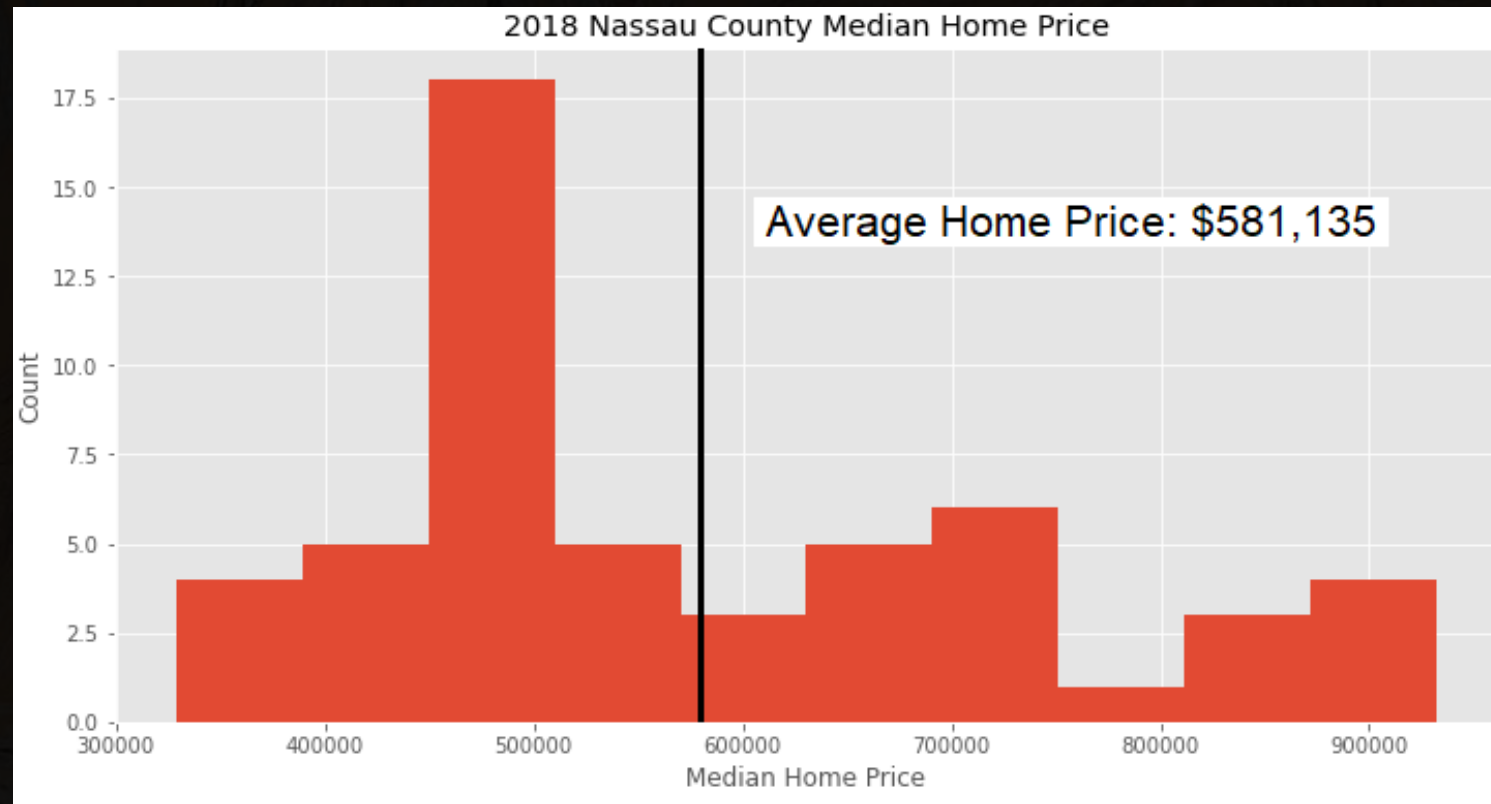
Average Housing Prices 1996 - 2018
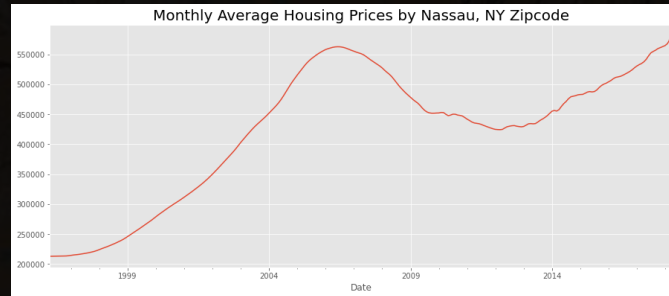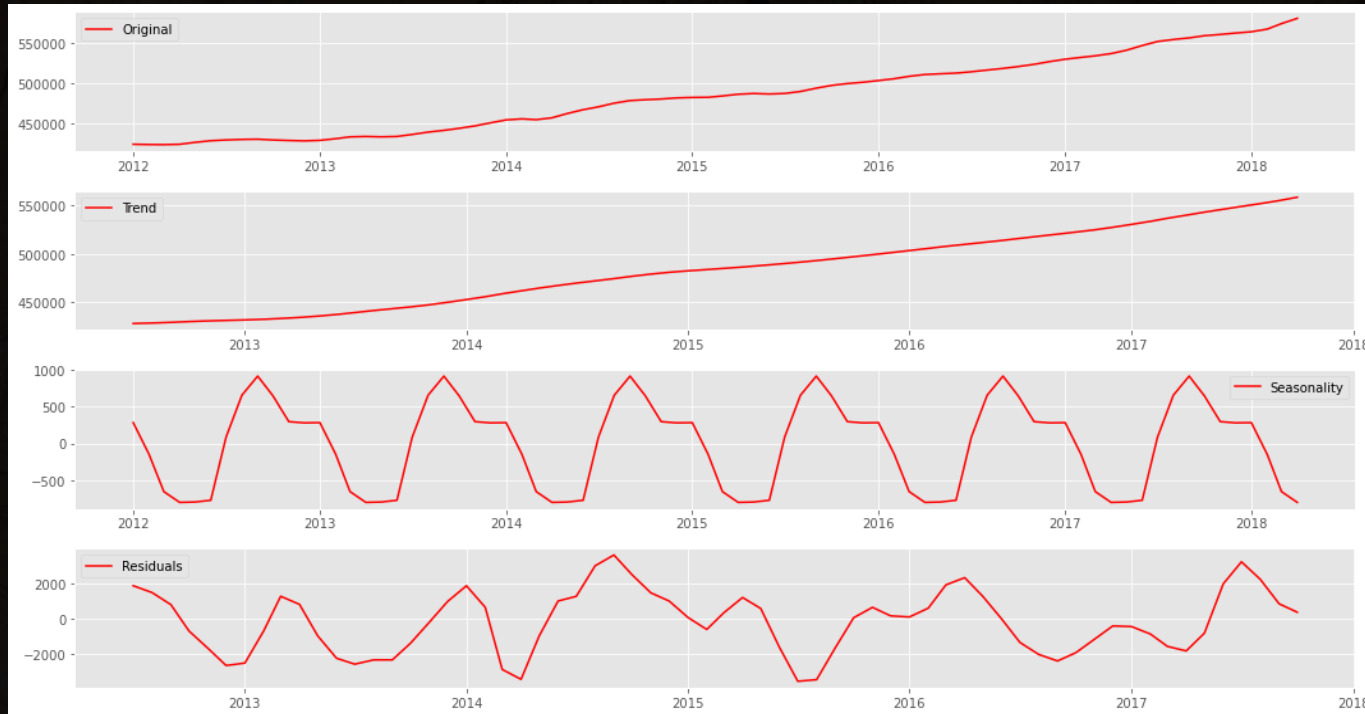

Monthly Average Housing Prices by Nassau, NY Zipcode

Average Housing Prices 2008 - 2018

Monthly Average Housing Prices by Nassau, NY Zipcode

Time Series Model

Monthly Average Housing Prices by Nassau, NY Zipcode
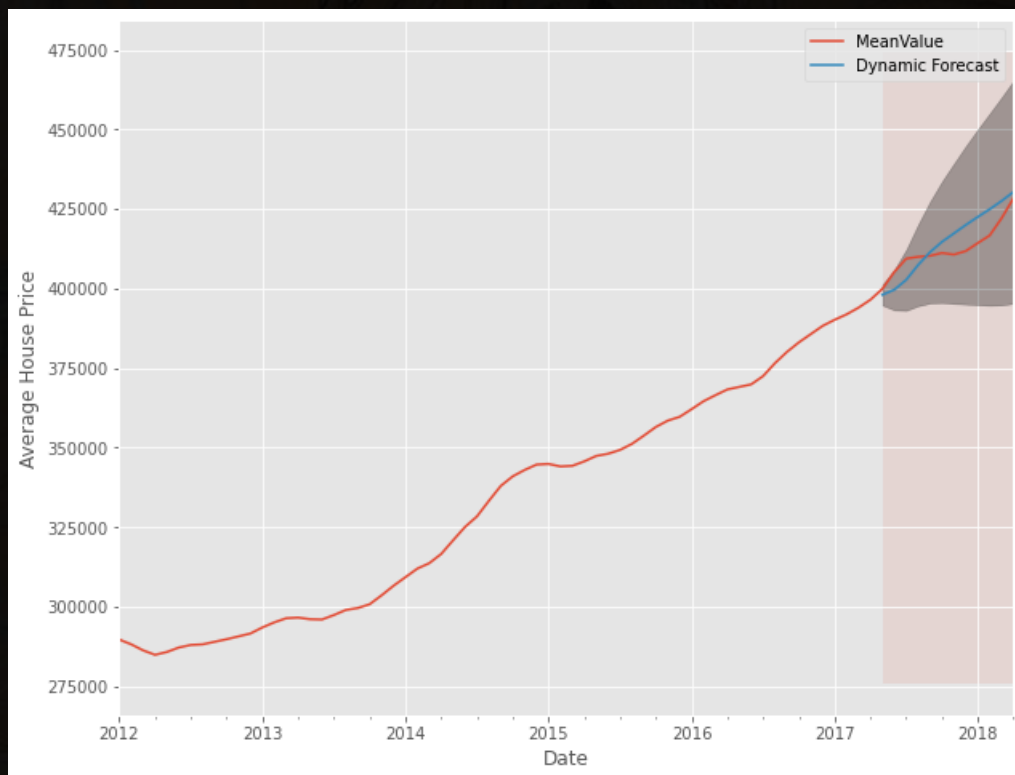

Time-series

Seasonal Decomposition

- Trend is Increasing
- Seasonality exists
- Residuals are not stationary
- Variations of these components must be removed for a proper time series model

Dynamic Forecasting

ZIPCODE	Sqrt-MSE
11565	3,446
11509	4,077
11518	4,102
11575	4,665
11756	4,890

Avg RMSE = 13,855

Top 5 Zip Codes with Highest 5yr ROI

Zip code	City	Current Value	5 Years Value	5 Years Upper	5Yr-ROI	5Yr-ROI-Upper
11735	Farmingdale	$ 467,900.00	$ 799,889.10	$ 1,459,824.00	70.1%	211%
11803	Plainview	$ 880,000.00	$ 1,491,852.00	$ 3,412,622.00	69.5%	287%
11598	Woodmere	$ 488,500.00	$ 786,229.50	$ 1,140,029.00	60.9%	133%
11003	Elmont	$ 566,900.00	$ 910,321.40	$ 1,444,101.00	60.5%	154%
11552	W. Hempstead	$ 502,500.00	$ 798,562.20	$ 1,076,203.00	58.9%	114%

Recommendations

Recommend the below 5 Zip Codes

- 11735: Farmingdale
- 11803: Plainview
- 11598: Woodmere
- 11003: Elmont
- 11552: W. Hempstead
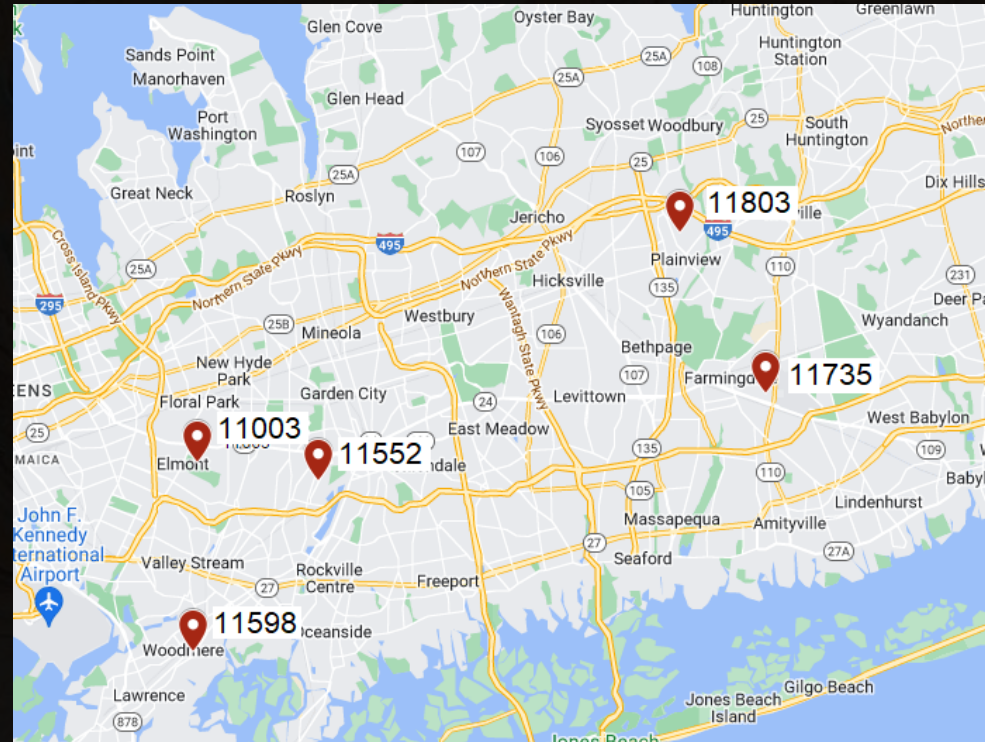
Future Considerations

o   The methods we will employ in this project example will only take in data from a uni-variate time series. That means we really are only considering the relationship between the y-axis value the x-axis time points. We're not considering outside factors that may be effecting the time series.

o   Build a linear model with additional features such as federal interest rate, housing grade score, etc.

Q&A