Projecting New York City Real Estate

Gabe Arnold

Background and Project Overview

Having lived in different cities in America, I am amazed at the cost of houses in New York. Homes offer a source of wealth, a place of security and a chance for families to set down their roots and help grow the surrounding areas. By looking at the asking price of homes, we can get a glance at one way an area is growing.

I wanted to take a look into the asking prices of homes across the five boroughs in order to see how much values change over time, and to build a model that could forecast these prices.

Process

Data Collection

- Gathered Data from StreetEasy.

- Created functions to clean up CSV files.

EDA & Statistical Analysis

- Visualized pricing data over time.

- Ran Dickey-Fuller test to check for stationarity.

Modeling

- AR (Baseline)

- ARIMA

- Sarimax

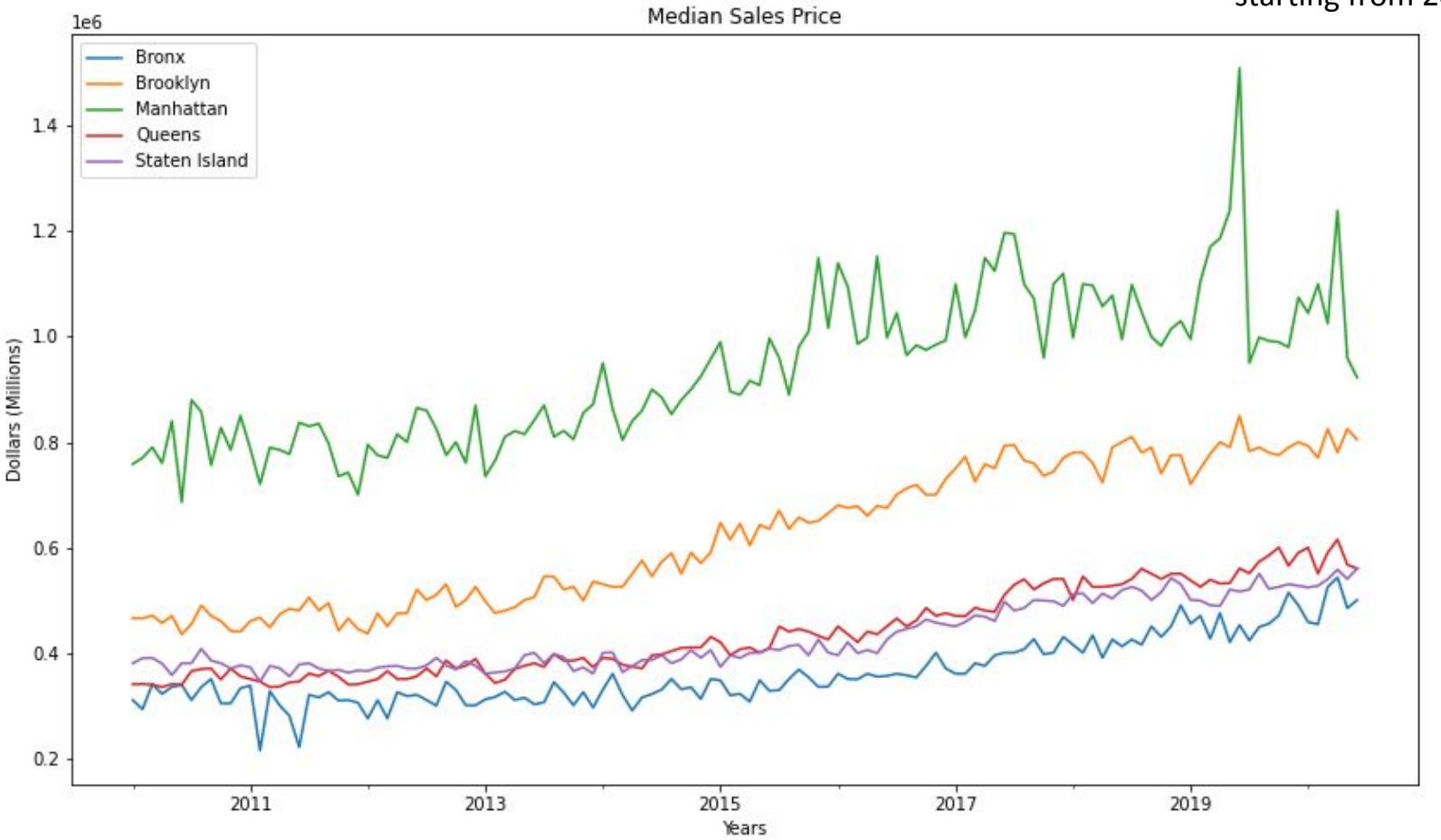
- LSTM

Conclusions

- Selected best model

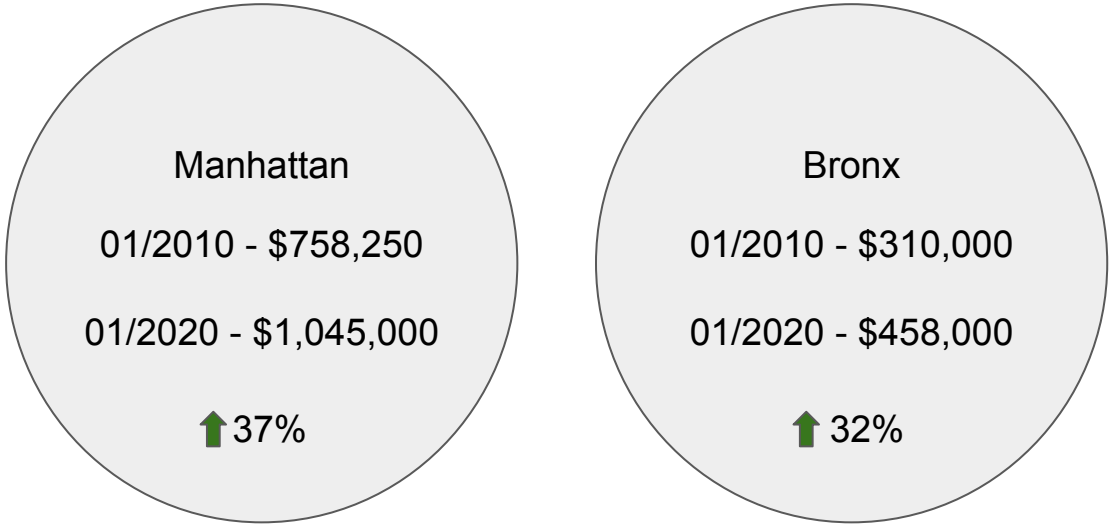
- Interpreted results of the model

Sales prices by borough starting from 2010.

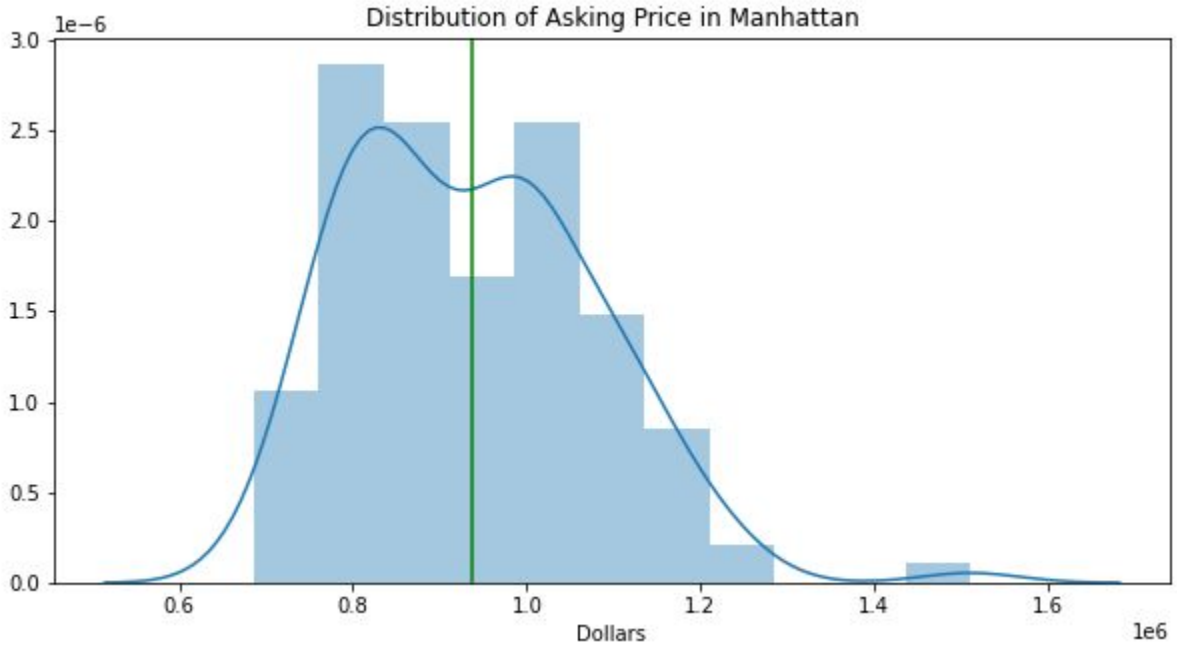


Borough Profiles

The data used for model is the median asking home prices for Manhattan and the Bronx. (All boroughs were tested but for brevity only Manhattan and the Bronx will be used.)

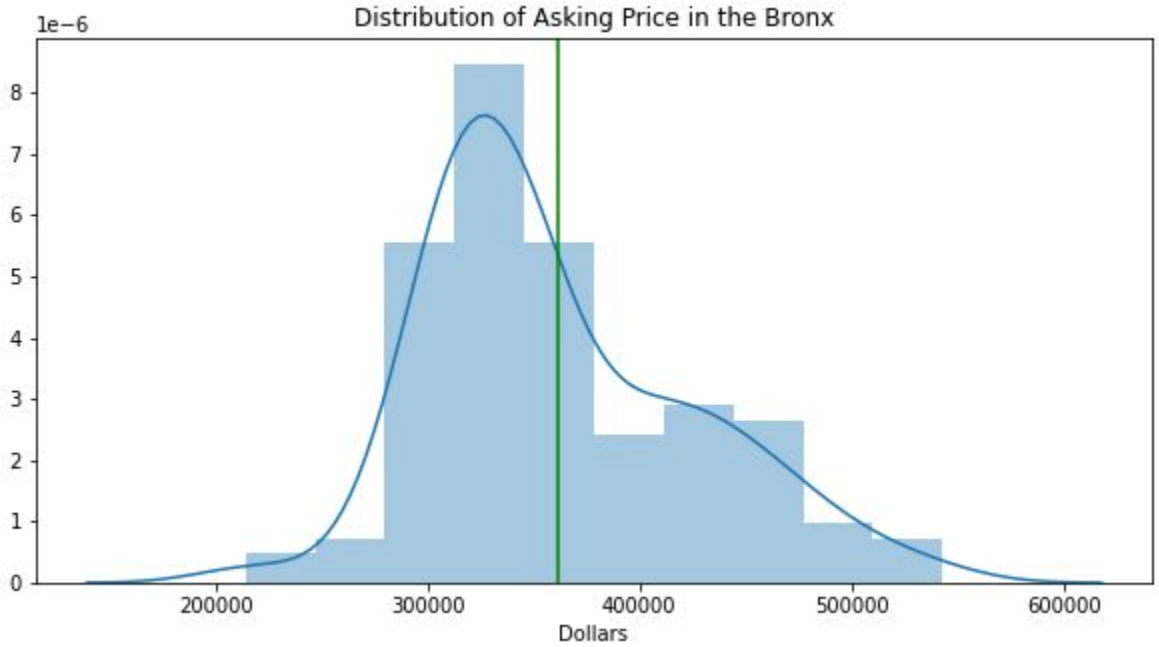


Borough Profiles



The distribution of prices in Manhattan. The majority of are values are located under a million dollars. The green line is the mean price.

Borough Profiles



The Distribution of prices in The Bronx. The majority of are values are less than $400,000. The green line is the mean asking price.

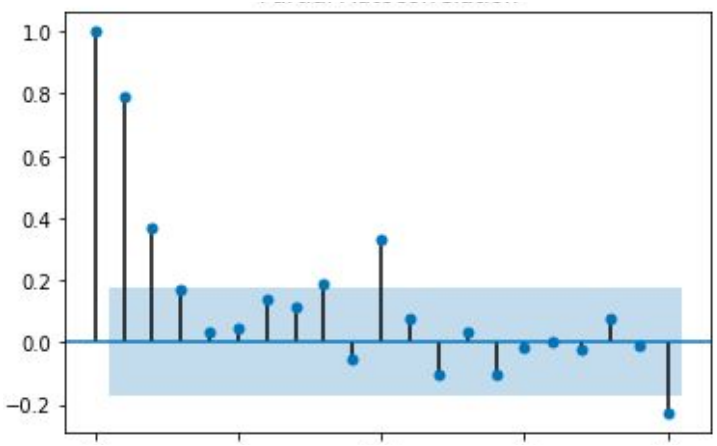
Evaluation Metrics

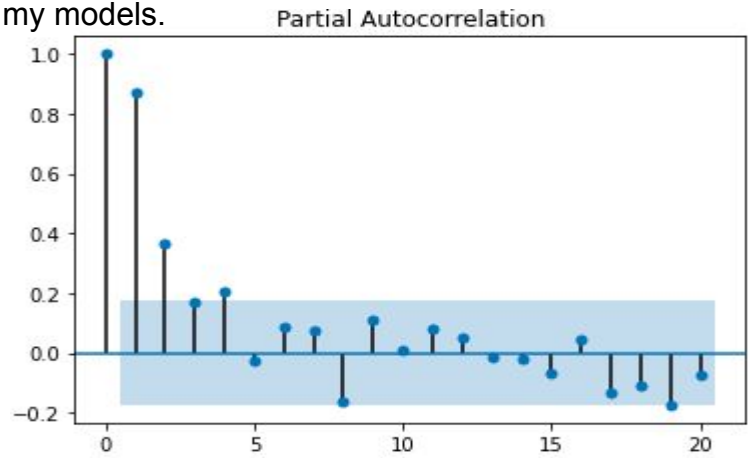
SMAPE - Accuracy measure based on percentage (or relative) errors. A perfect SMAPE score is 0.0, and a higher score indicates a higher error rate.

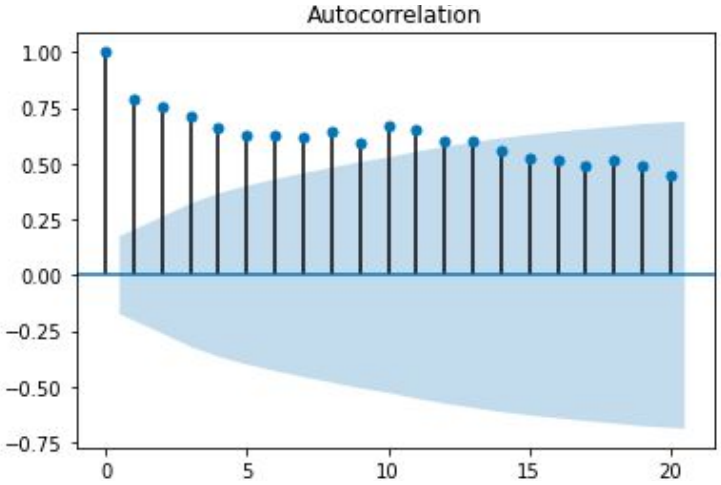
RMSE - Used in order tell the square root of the all squared errors. The answer is delivered in units that were passed through. In other words, my error will be in dollar amounts. One drawback to RMSE is sensitive to outliers and Manhattan has a large spike in 2019.

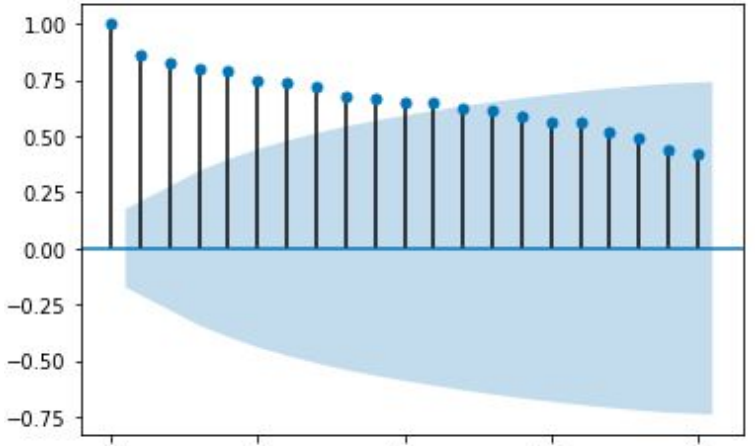
PACF and ACF

The PACF and ACF for Manhattan (Right) and the Bronx (Left). Used to set the order of my models.

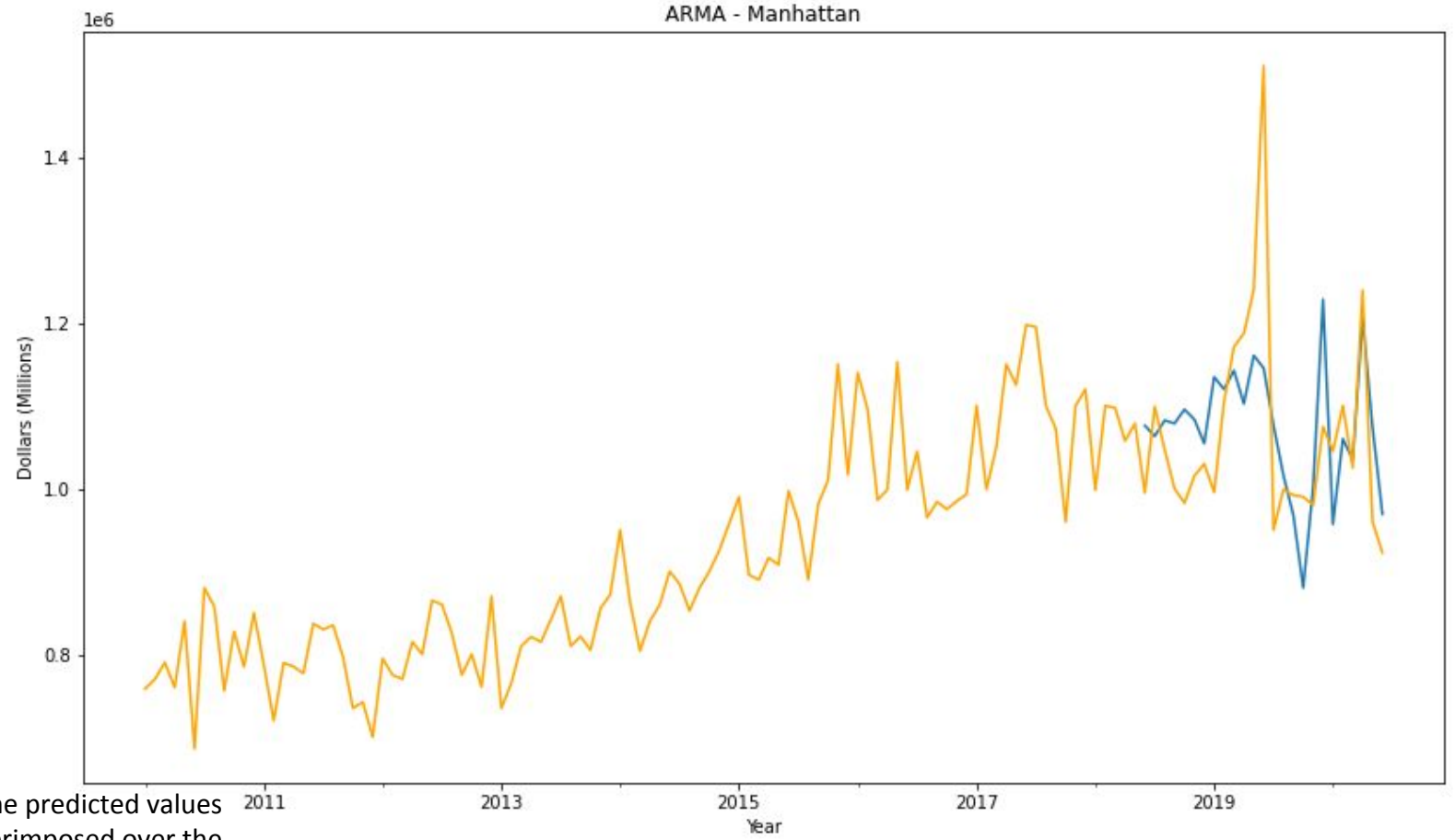




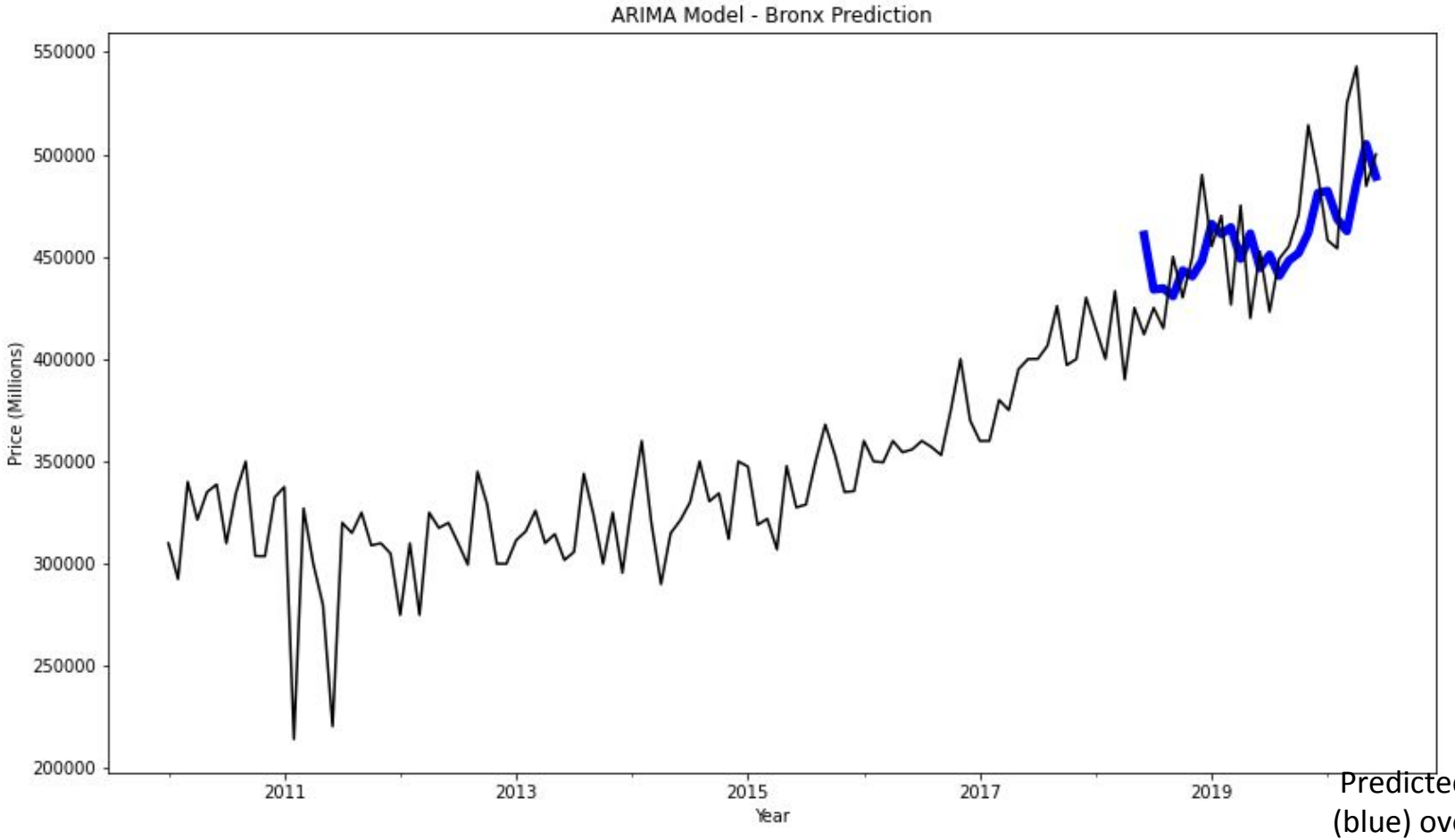




	Manhattan Models		Bronx Models	
Model	RMSE	SMAPE	RMSE	SMAPE
AR Model (First Order)	120,553	8.08	30,306	5.32
ARMA (Mhtn Order = 2,5) (BX Order = 2,2)	105,597	7.05	29,594	5.23
ARIMA	106,578	8.07	26,285	4.55
AUTO ARIMA	120,355	7.73	29,444	5.00
SARIMAX Model	123,594	7.97	29,468	4.79
LSTM Model	136,129	8.87	32,796	5.85



Here are the predicted values (blue) superimposed over the actual values.



Predicted values (blue) over actual values.

Takeaways from Modeling

- Adjusting for both p (Lag Order) and q (Moving Average) yielded the most significant changes among my models.
- None of my models accurately predicted the large spike in median asking price in Manhattan except LSTM, however it wasn't the best model.

Conclusion

When look to settle in New York, buyers will be faced with some of the highest prices in the U.S. Median asking price in Manhattan is $1 million dollars and about 500K for a home in Bronx.

Current prices are high which is good for sellers, tougher for buyers.

Lower cost opportunities Bronx, Staten Island.

If you are able to purchase NYC Real Estate, the nominal returns will be fairly high, and you may be able to double your asking price in 10 years.

Additionally, if the next 10 years are like the previous 10, the median asking price of home values across all boroughs will raise by at least 83%.

Stay in Touch

LinkedIn: Gabe Arnold

GitHub: GabeSArn