

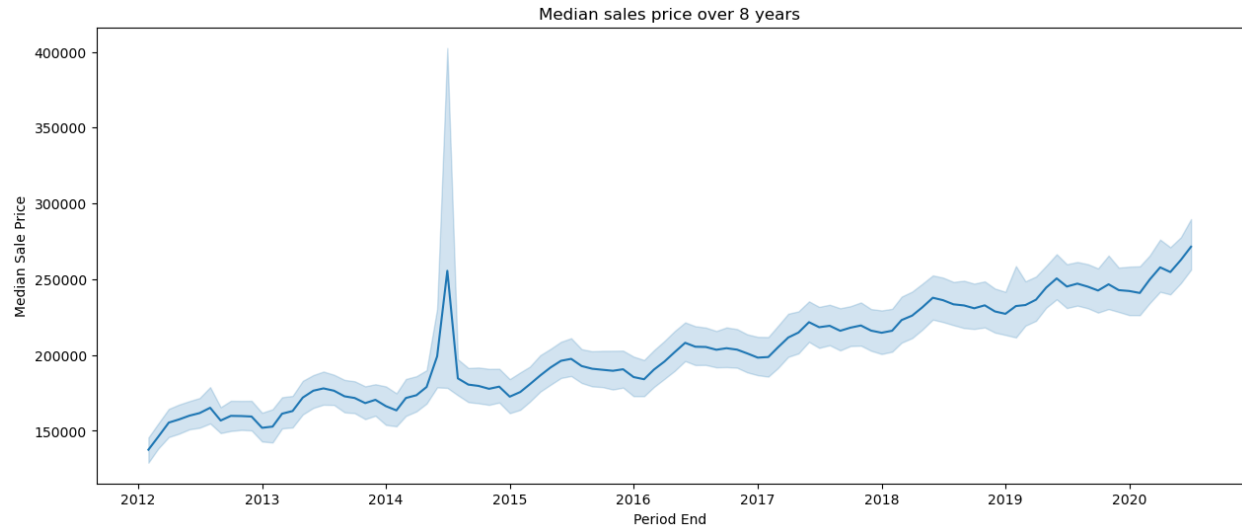
Jazmaine Vitta

December 14th, 2023

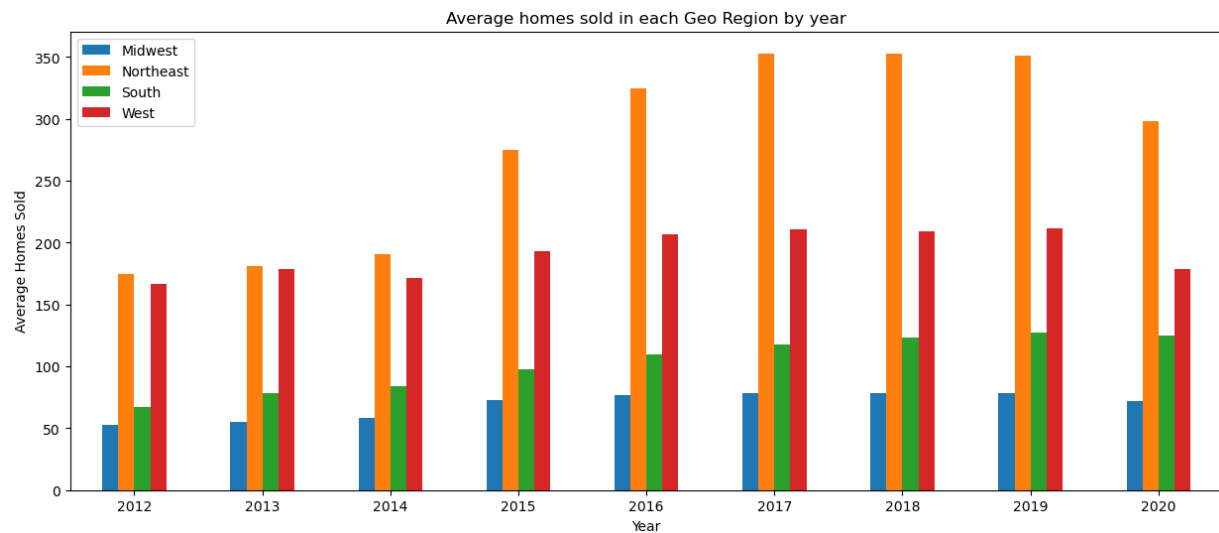
Predicting Housing Prices

The effects of making poor and uneducated financial decisions can be detrimental to an individual's existence. This can lead to things such as a low credit score, lack of savings, overreliance on debt and endless other consequences. As registered financial professionals, Brokerage Firm personnel are expected to be knowledgeable enough to prevent those possibilities. In order for brokers to facilitate transactions between traders, sellers, or buyers they will need to be able to make accurate and knowledgeable predictions based on past history. Using my ARIMA time series model to predict housing prices in the United States companies will be able to successfully draw insights and come to knowledgeable decisions. To assist brokerage firms in providing accurate insights and information on their potential investment decisions in regard to the real estate market.

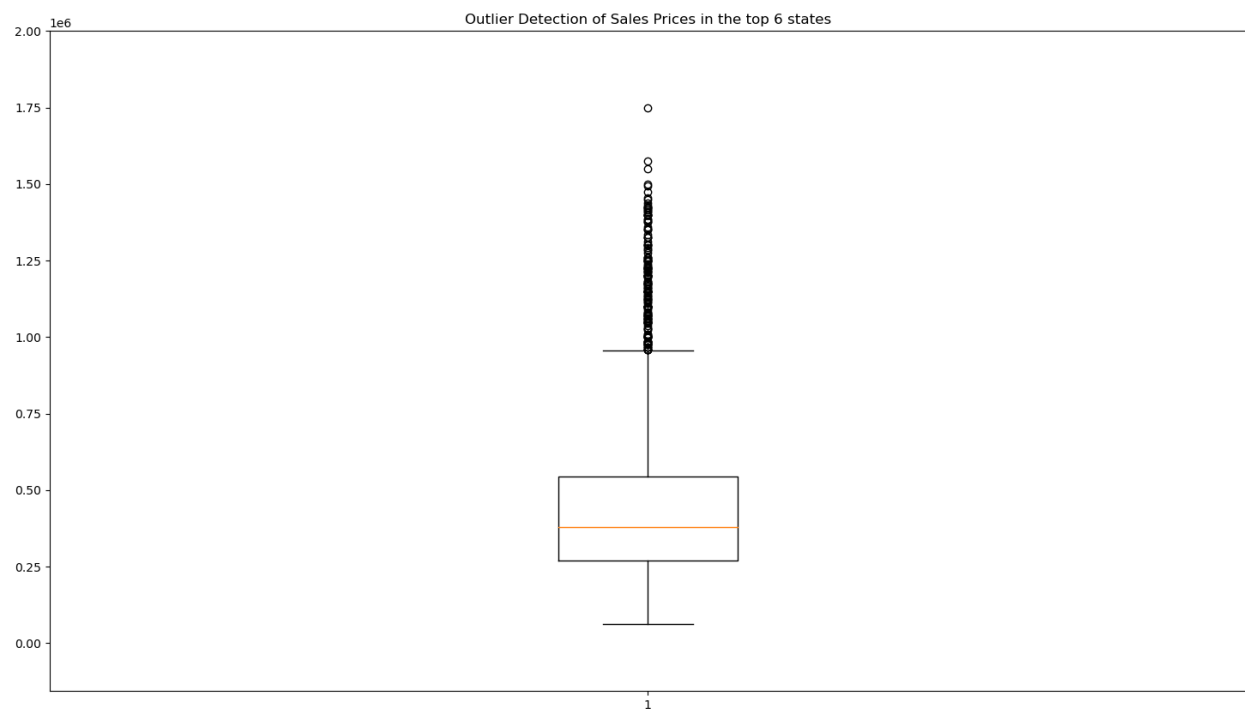
My time series model will be created with past housing sale data in order to predict what will happen in the future. Data collected from a brokerage database called Redfin. Redfin is a brokerage firm who connects buyers and sellers to complete a transaction for real estate amongst many other things. The data was made up of a total of 40,677 observations & 67 features throughout the years of 2012 to 2020. Throughout my process of preparing the data I was able to uncover the answers to many interesting questions. Based on the information the target feature is the 'Median Sale Price' column. Over the past eight years there has been a clear and steady increase in the sales pricing



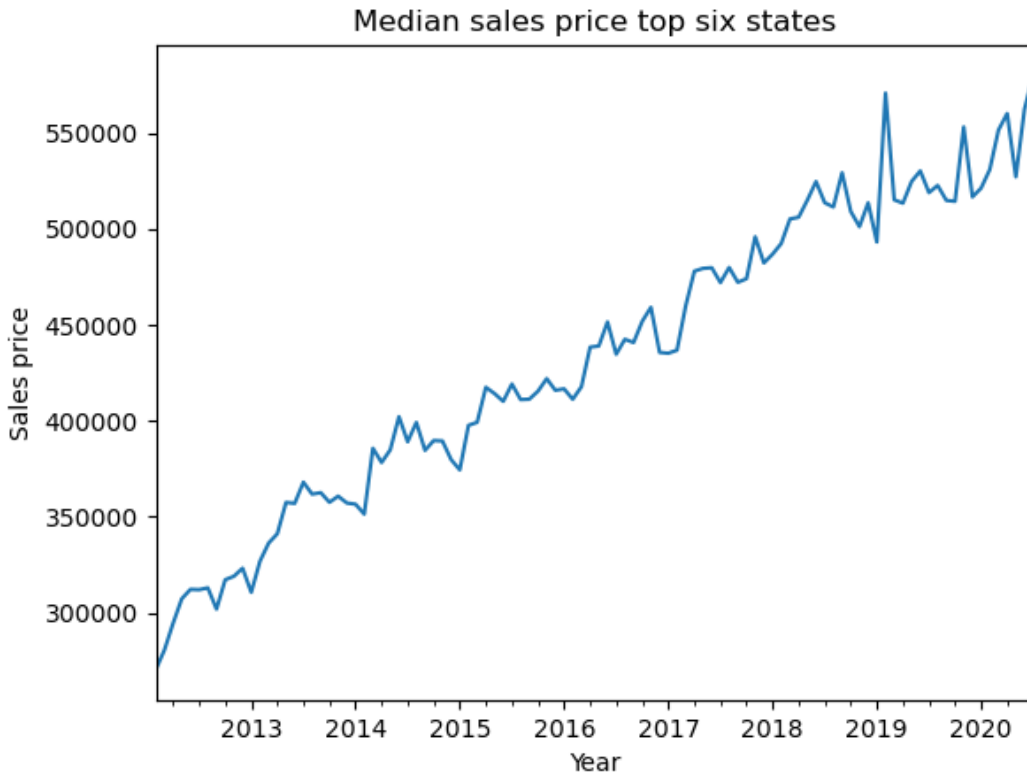
The reign with the overall highest average sale price appears to be the Northeast region with the next highest being the west. This will most likely be the areas where our focus will remain as they will have the highest profits.



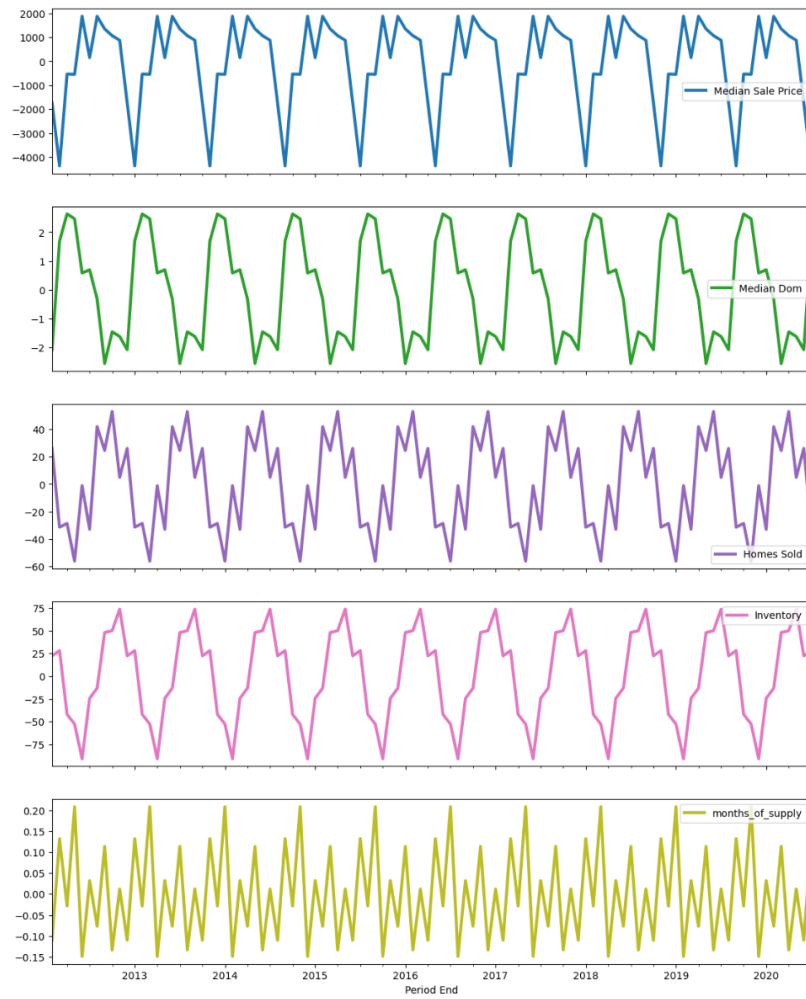
The six states with the highest median sale price are HI, CA, MA, DC, NM & CO. Using a box plot it is seen that there are quite a few outliers to deal with. This can be helped by resampling the data.

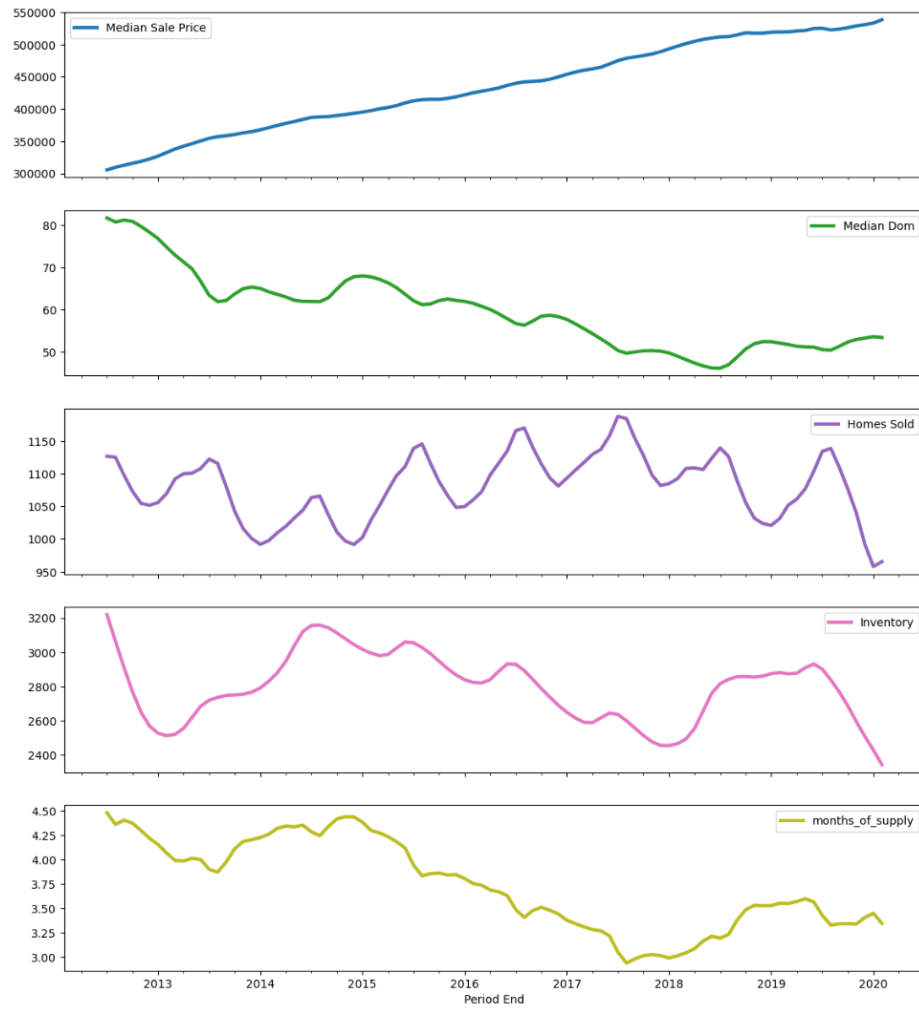


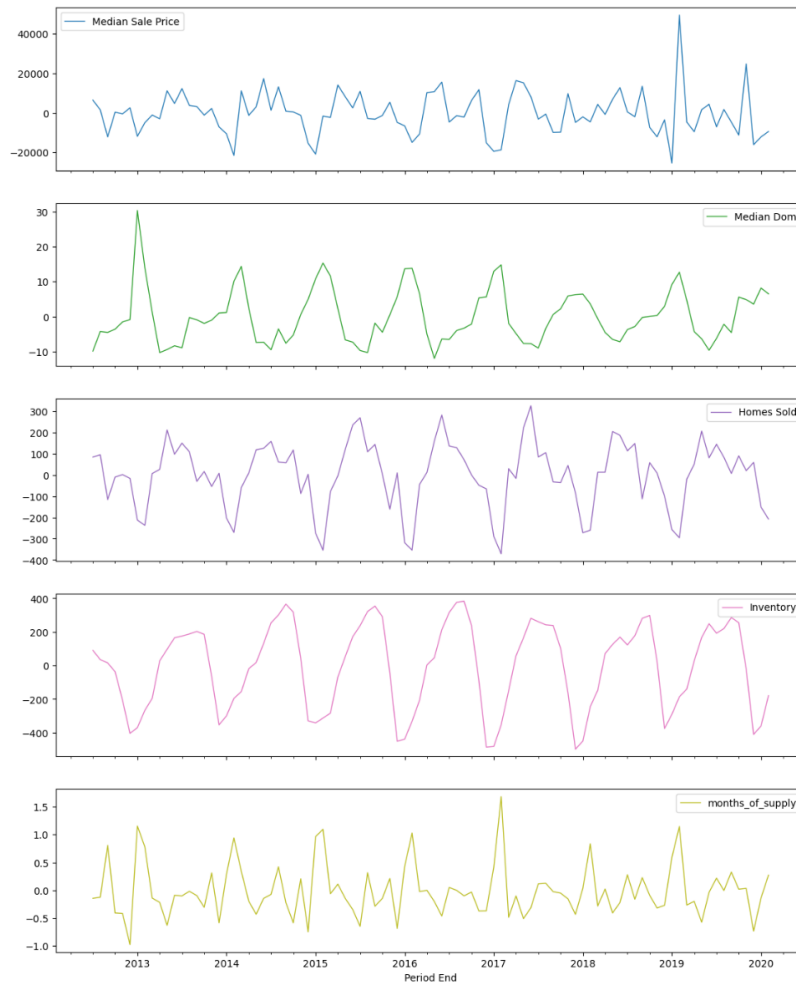
By resampling the data I now have a time series with a mean sale price for each month from 2012 to 2020. The line plot below shows the almost steady growth of the sales price over time.



The trend shows that from 2013 house prices slowly started to increase. The residual shows the time-series with the trend and seasonality removed. Looking at the residual plot, it is clear that house prices fluctuate more heavily and sporadically from 2019 onwards.

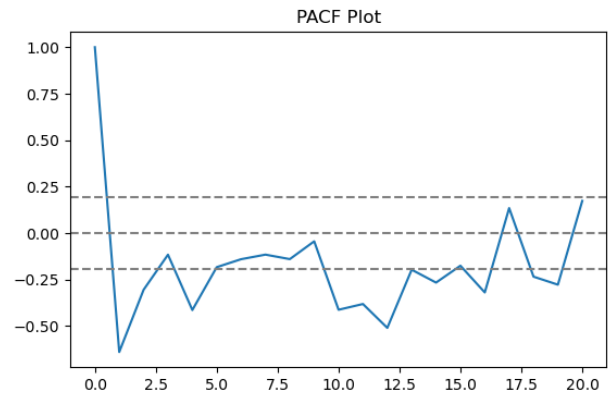
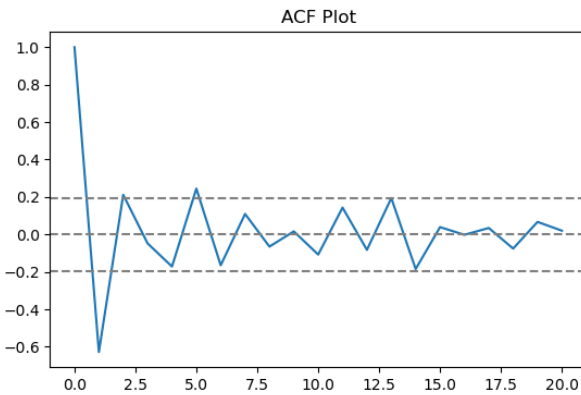






The dotted lines in confidence interval, this can be used to determine p and q.

p: The lag value where the PACF chart crosses upper chart for first time. q: The lag value where ACF chart crosses upper chart for first time. Here $p = 1$, $q = 1$, order = (1,1,1)



Perform Dickey fuller test

ADF Statistic: -7.326265

p-value: 0.000000

Critical Values:

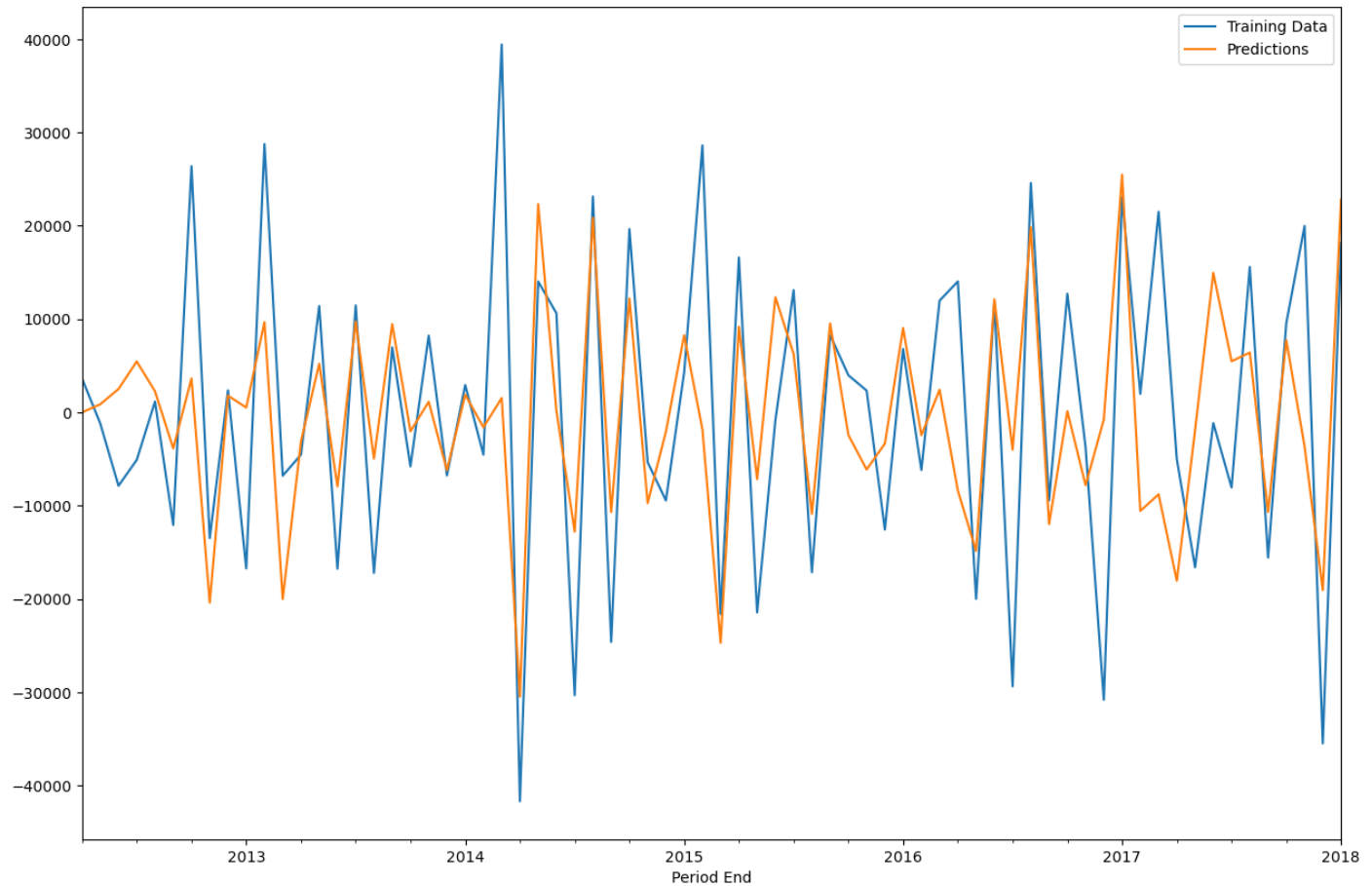
1%: -3.508

5%: -2.895

10%: -2.585

The series is stationary.

The Test Statistic is -7.326265, which is less than any of the critical values. The p-value obtained is less than the significance level of 0.05, so we can reject the null hypothesis. The time series is, in fact, stationary. Stationarity is an important property of time series data that indicates that the statistical properties of the data do not change over time. It is essential for various time series analysis techniques, including forecasting and modeling.



Predicting housing prices is not an easy task for an individual not equipped with the right tools and knowledge. With the past history of housing sale prices and the help of my ARIMA Time Series model it soon will be an easy task. It will no doubt assist brokerage firms in coming to more data driven insights and logical conclusions. People need experienced Brokers to make educated decisions, Brokers need accurate predictions of future potential investments to supply to their customers. By gathering even more observations my model will no doubt assist in strengthening the predictability of the model. This recommendation would help the model become even more of an asset to your business.