

National Housing Prices

The background features a blue bar chart with vertical bars of varying heights. Overlaid on this is a red line graph that trends upwards, ending in a large red arrow pointing towards the top right. In the foreground, there are several 3D models of houses with white walls and red roofs, arranged in a cluster. The overall theme is housing and real estate.

By: Jazmaine Vitta

40% 50% 15%

Improving Hotels Operations and Revenues

The Problem

The ability to accurately draw insights on potential investments in the real estate market is essential for Brokerage Firms.

-

The Aim

How will the target audience benefit from my model?

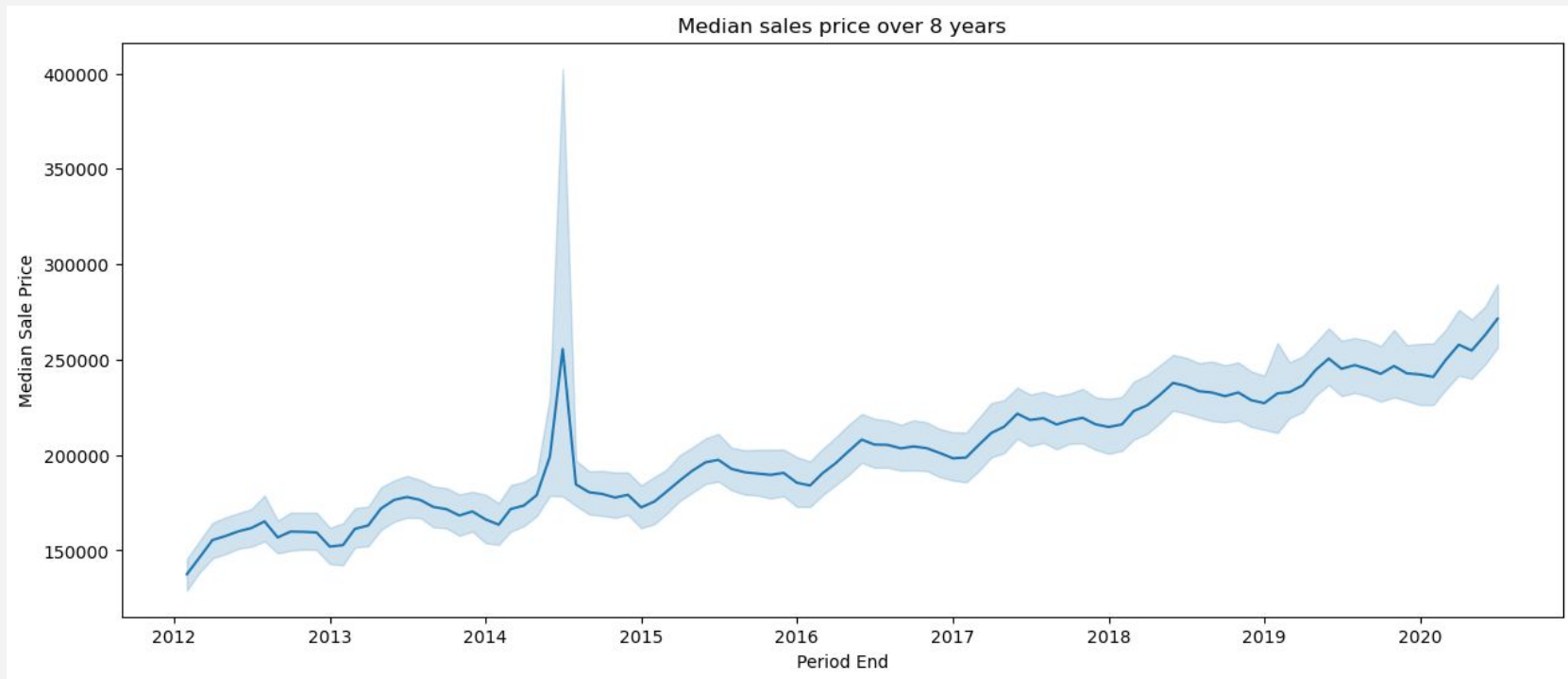
The Problem

The Data

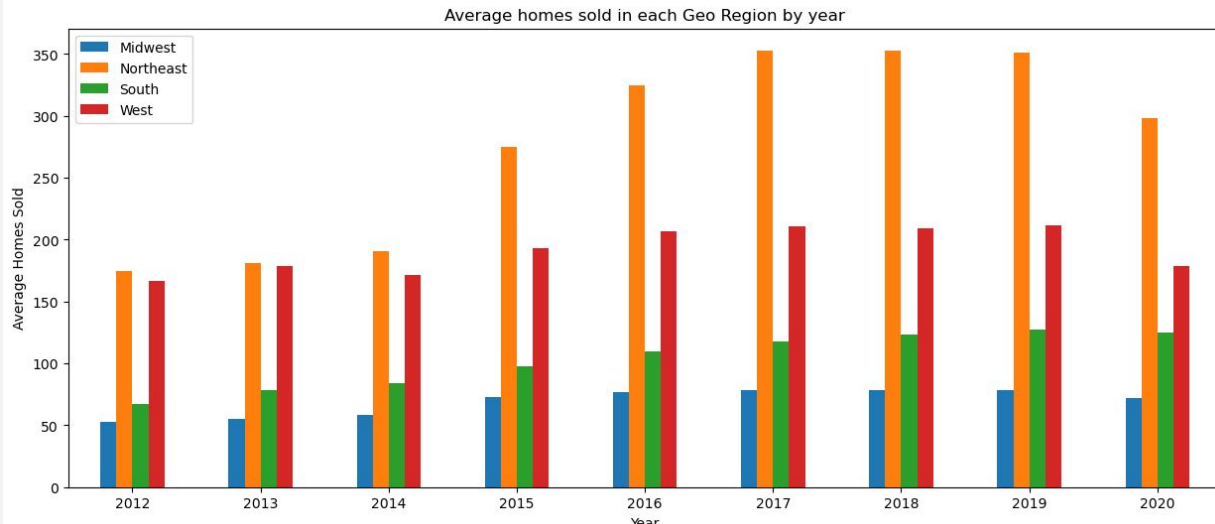
Data was collected from a brokerage database named Redfin. Redfin is a brokerage firm who connects buyers and sellers to complete a transaction for real estate amongst many other things.

The data was loaded with 40,677 observations & 67 features

The Data

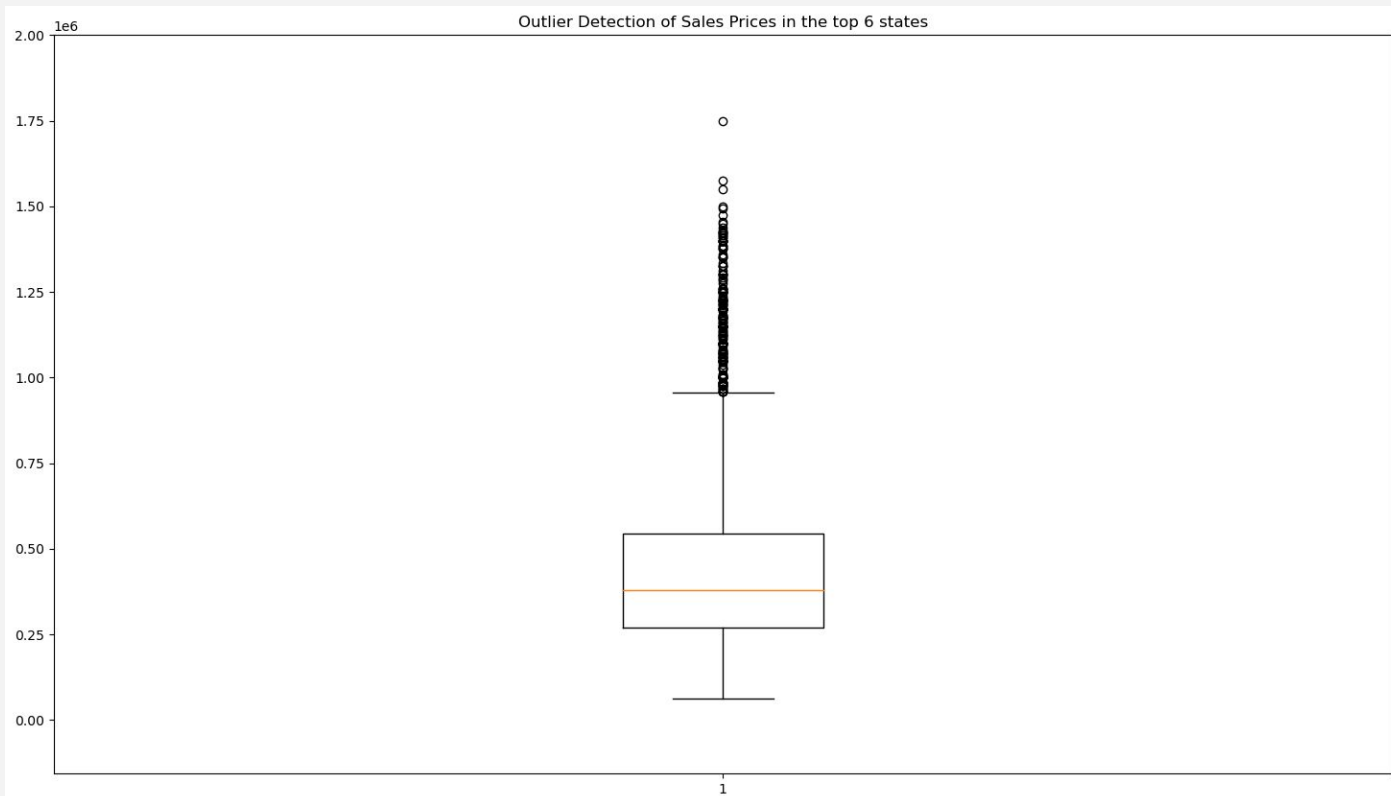


The Data



The Data

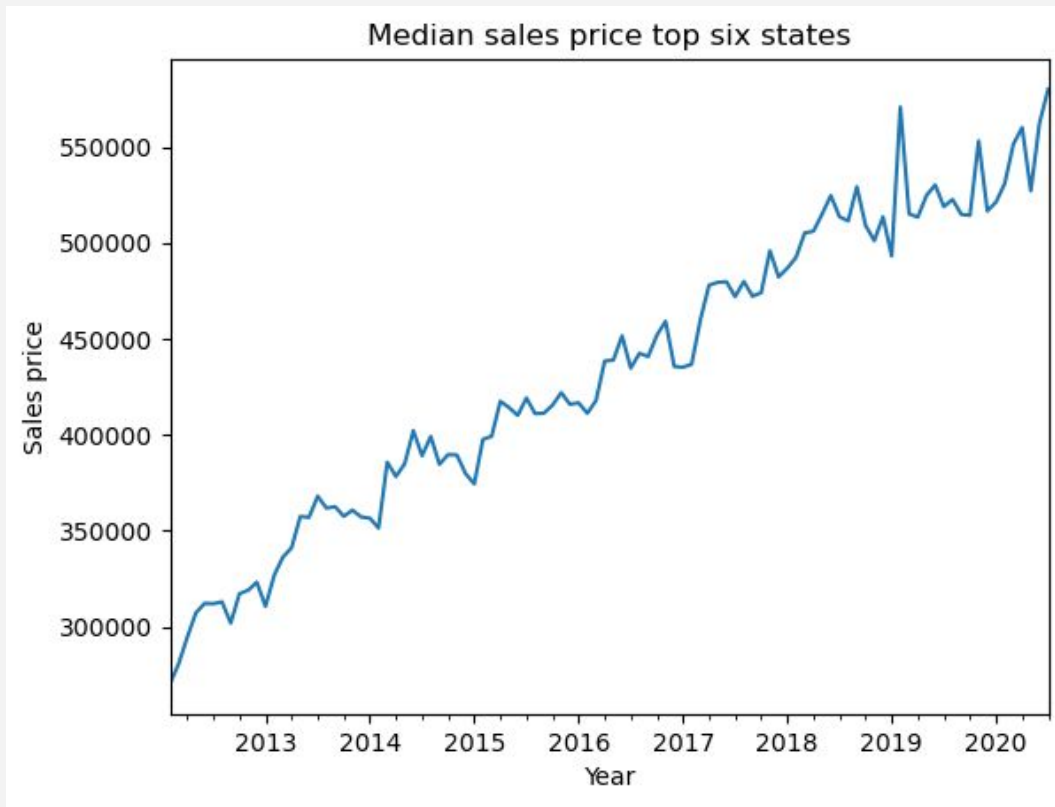
A Box Plot Detecting Outliers in the DataFrame



The Data

I have resampled the data, creating a mean sale price for each month from 2012 to 2020.

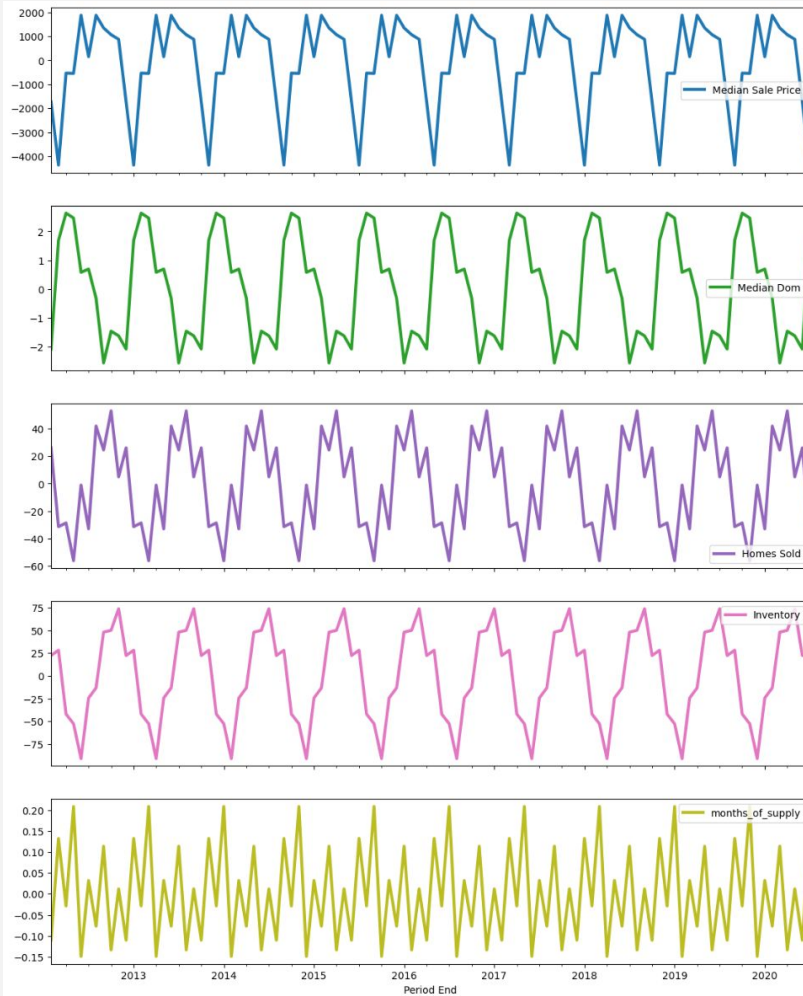
Now we have a Time Series!



Model Selection

Seasonality

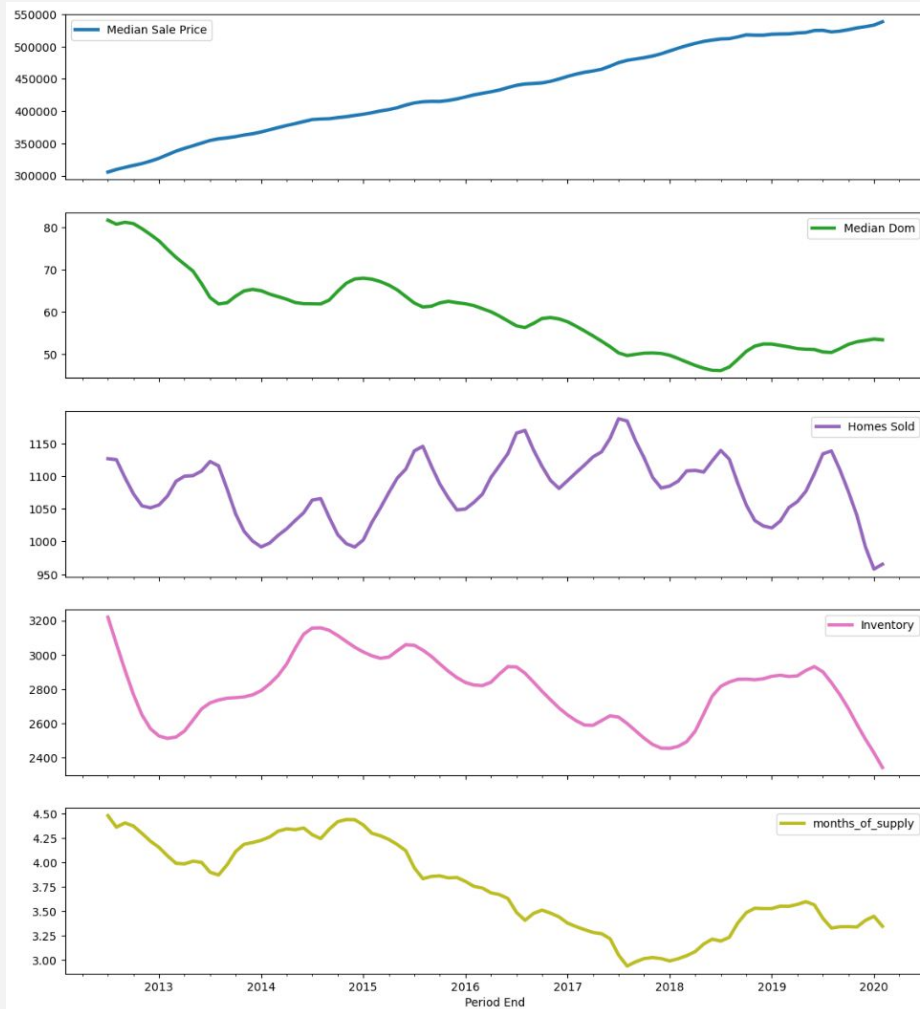
Some seasonality behaviors are apparent



Model Selection

TREND

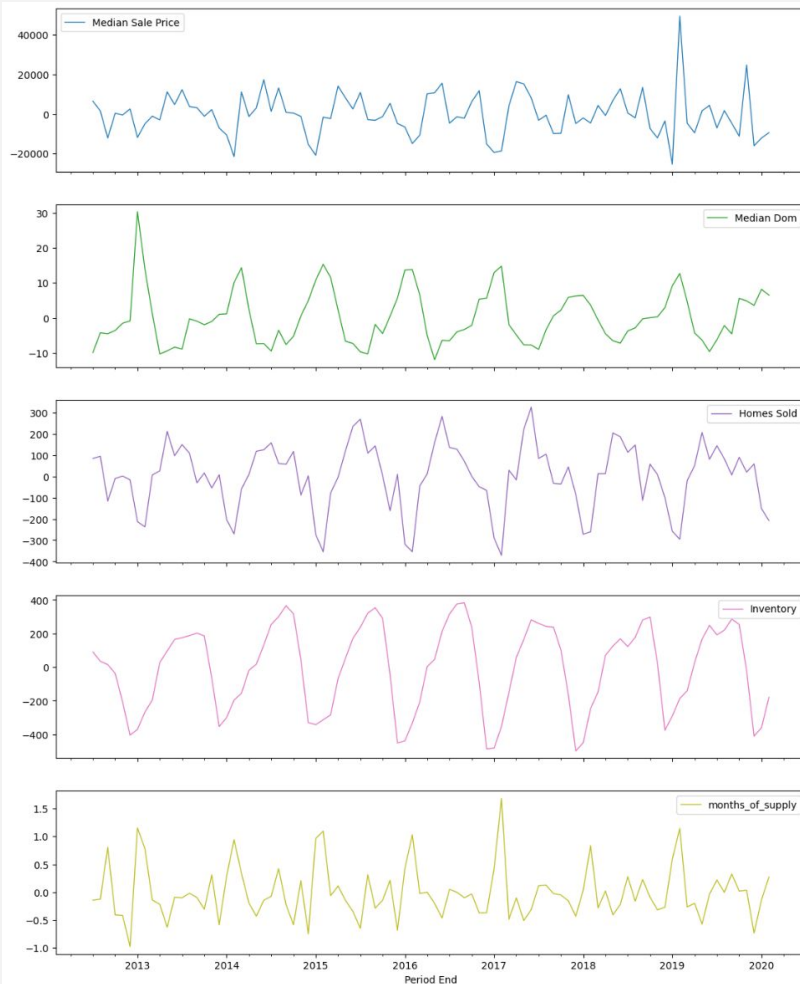
The trend shows that from 2013 house prices slowly and steadily start to increase.



Model Selection

Residual

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



Model Selection

The data is now stationary as the Test Statistic appears to be greater than the Critical Values. Below is AFC & PACF Plot of Median Sale Price column

Perform Dickey fuller test

ADF Statistic: -7.326265

p-value: 0.000000

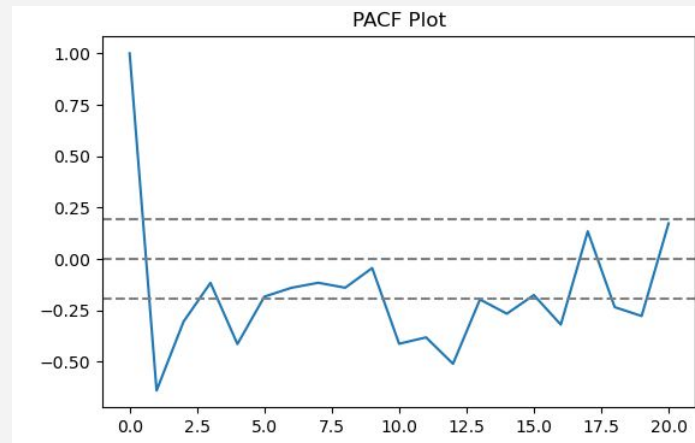
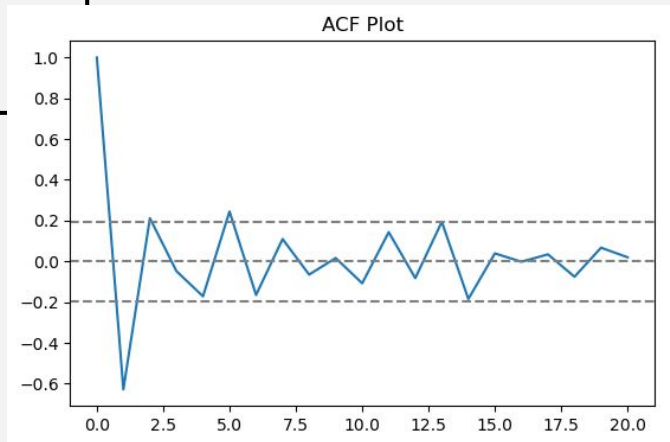
Critical Values:

1%: -3.508

5%: -2.895

10%: -2.585

The series is stationary.



Model Selection

SARIMAX Results

```
=====
Dep. Variable:   Median Sale Price  No. Observations:      70
Model:          ARIMA(2, 1, 1)  Log Likelihood      -751.331
Date:           Sat, 16 Dec 2023  AIC              1510.661
Time:           14:35:20  BIC              1519.597
Sample:         04-01-2012  HQIC              1514.206
                  - 01-01-2018
```

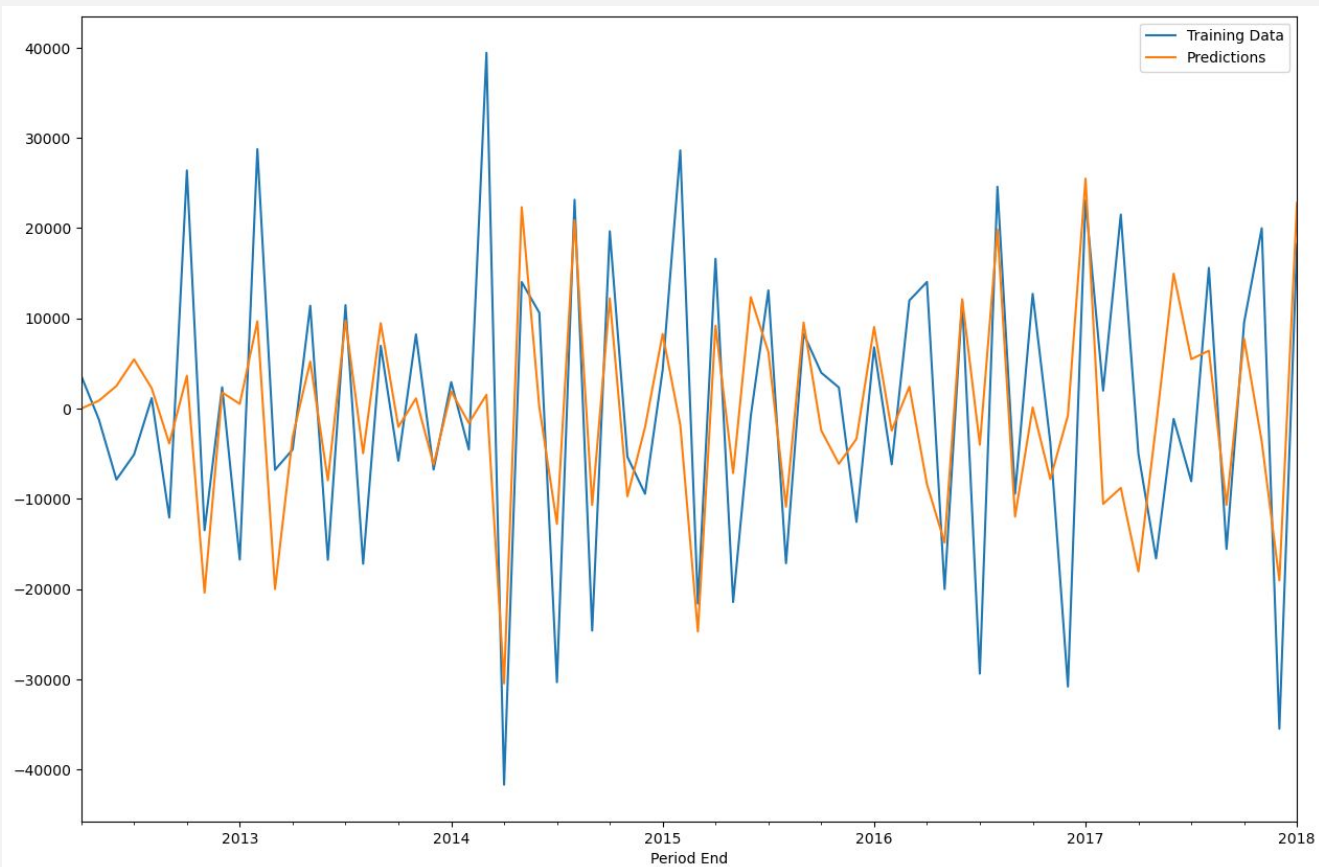
Covariance Type: opg

```
=====
              coef  std err          z      P>|z|    [0.025    0.975]
-----
ar.L1        -0.8132    0.140     -5.794    0.000    -1.088    -0.538
ar.L2        -0.2895    0.116     -2.492    0.013    -0.517    -0.062
ma.L1        -1.0000    0.148     -6.736    0.000    -1.291    -0.709
sigma2       1.571e+08  9.45e-10  1.66e+17    0.000  1.57e+08  1.57e+08
=====
```

```
=====
Ljung-Box (L1) (Q):          0.12  Jarque-Bera (JB):          3.70
Prob(Q):                   0.73  Prob(JB):                   0.16
Heteroskedasticity (H):      1.63  Skew:                   0.48
Prob(H) (two-sided):        0.25  Kurtosis:                3.61
=====
```

Model Selection

{'mape': 2.1044562217891163,
'me': 25.496449897446535,
'mae': 23844.76875217613,
'mpe': -1.2947719787574354,
'rmse': 30254.099538023274,
'corr': 0.6832096461602714}



Summary & Conclusions

Using my ARIMA model brokerage firms
will be able to accurately predict real estate prices

References

[Redfin Data Sorce](#)