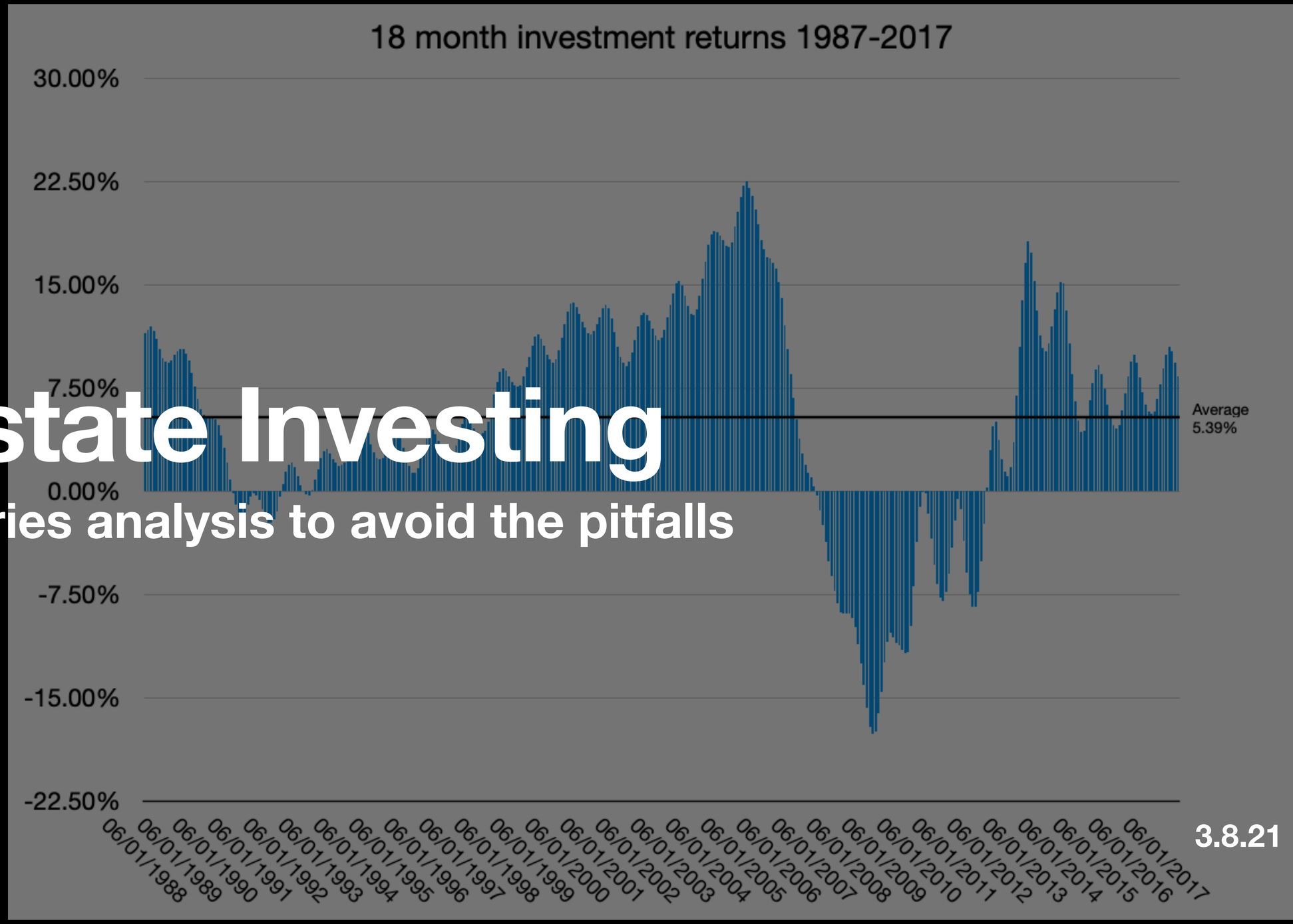
Real Estate Investing

Using time series analysis to avoid the pitfalls



Sawyer Shader-Seave

3.8.21

Case Study

As an analyst for a real estate investment firm:

Find areas with significant potential for increases in home prices over the next 18 months (1.5 years).

Focus on lower risk investments.

Data

	RegionID	RegionName	City	State	Metro	CountyName	SizeRank	1996-04	1996-05	1996-06	...
0	84654	60657	Chicago	IL	Chicago	Cook	1	334200.0	335400.0	336500.0	...
1	90668	75070	McKinney	TX	Dallas-Fort Worth	Collin	2	235700.0	236900.0	236700.0	...
2	91982	77494	Katy	TX	Houston	Harris	3	210400.0	212200.0	212200.0	...
3	84616	60614	Chicago	IL	Chicago	Cook	4	498100.0	500900.0	503100.0	...
4	93144	79936	El Paso	TX	El Paso	El Paso	5	77300.0	77300.0	77300.0	...

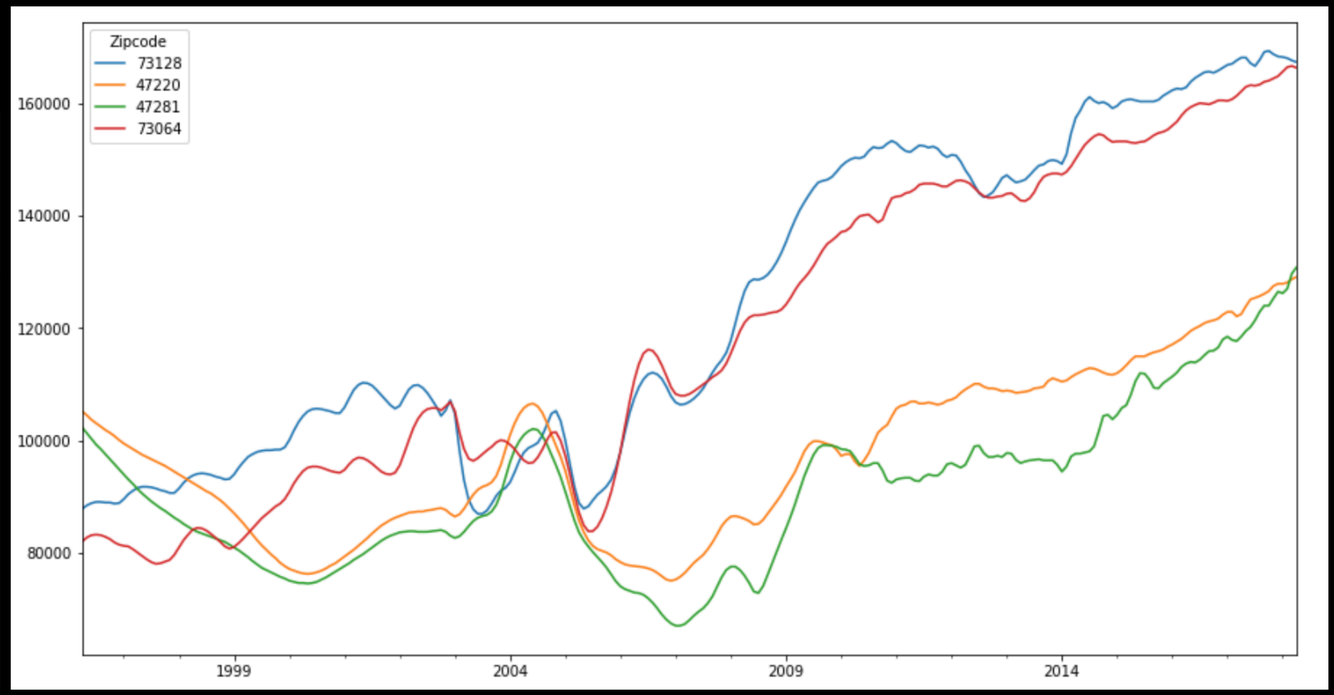
- Initial data:
  - US monthly median home prices from 1996-2018
  - 14,723 zipcode
  - City, State, Metro, County, Size rank
- Focused on “recessing proof” areas
  - top 5 zip codes by performance from before to after the great recession

EDA

SizeRank	1996-04	1996-05	1996-06	...	2017-08	2017-09	2017-10	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04	recession_return
10580	87900.0	88500.0	88900.0	...	167700	169100	169300	168700	168300	168200	168000	167600	167300	0.408126
10572	105200.0	104500.0	103700.0	...	125700	126100	126600	127500	127900	127900	128100	128700	129100	0.349367
14306	102300.0	101300.0	100300.0	...	122900	124000	124000	125300	126500	126200	127000	129700	130800	0.333333
4893	82100.0	82800.0	83200.0	...	163300	163800	164000	164400	164800	165600	166400	166600	166300	0.329982
22	3676700.0	3704200.0	3729600.0	...	7422400	7417600	7427300	7371400	7342700	7353300	7350300	7363000	7386600	0.326018

- 2 towns in Oklahoma
- 2 in Indiana
- I in Manhattan

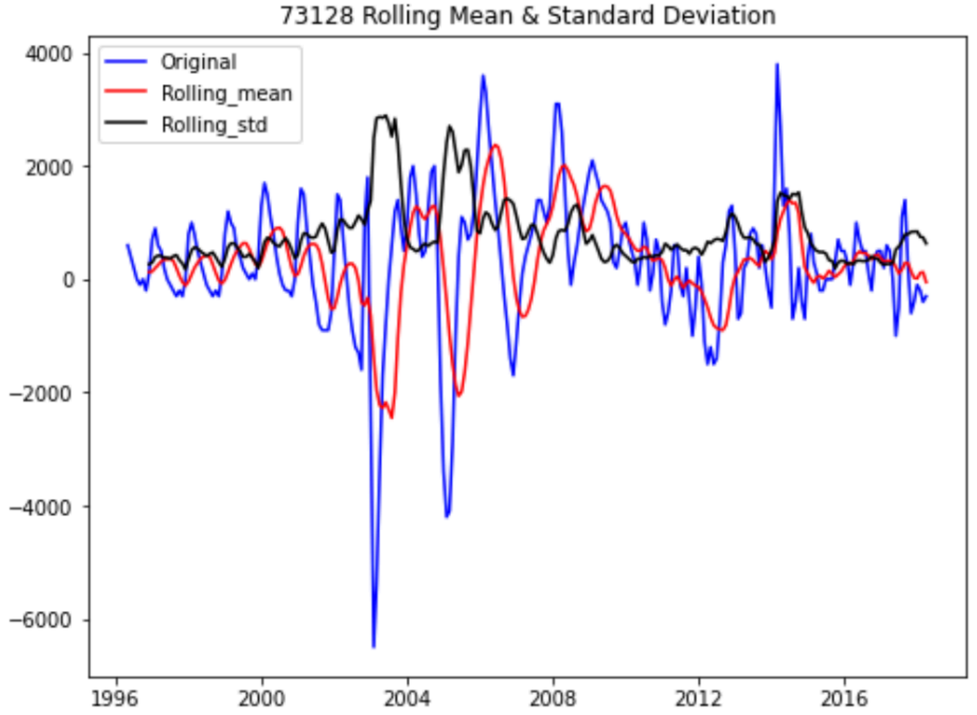
Zipcode	73128	47220	47281	73064	10128
Zipcode					
73128	1.000000	0.758325	0.698105	0.981098	0.880108
47220	0.758325	1.000000	0.961722	0.763331	0.736357
47281	0.698105	0.961722	1.000000	0.694567	0.688026
73064	0.981098	0.763331	0.694567	1.000000	0.889686
10128	0.880108	0.736357	0.688026	0.889686	1.000000



EDA

Stationarity

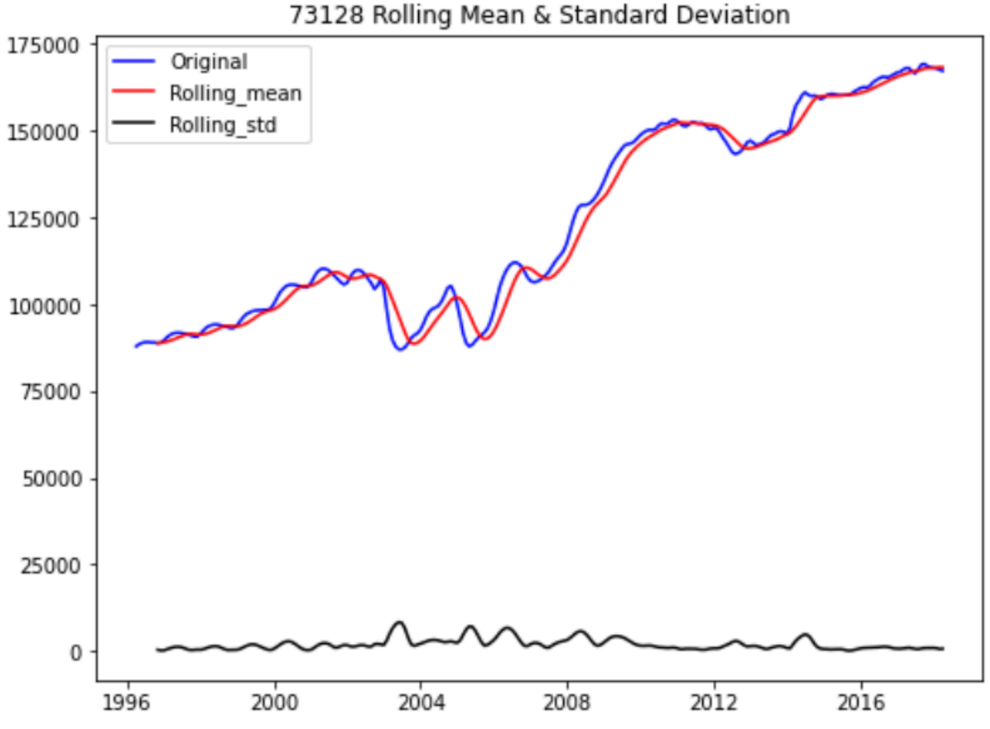
Differenced once




```
Results of df Test:

Test Statistic                  -5.198157
p-value                          0.000009
#Lags Used                       8.000000
Number of Observations Used    255.000000
Critical Value (1%)             -3.456257
Critical Value (5%)             -2.872942
Critical Value (10%)            -2.572846
dtype: float64
```

Original

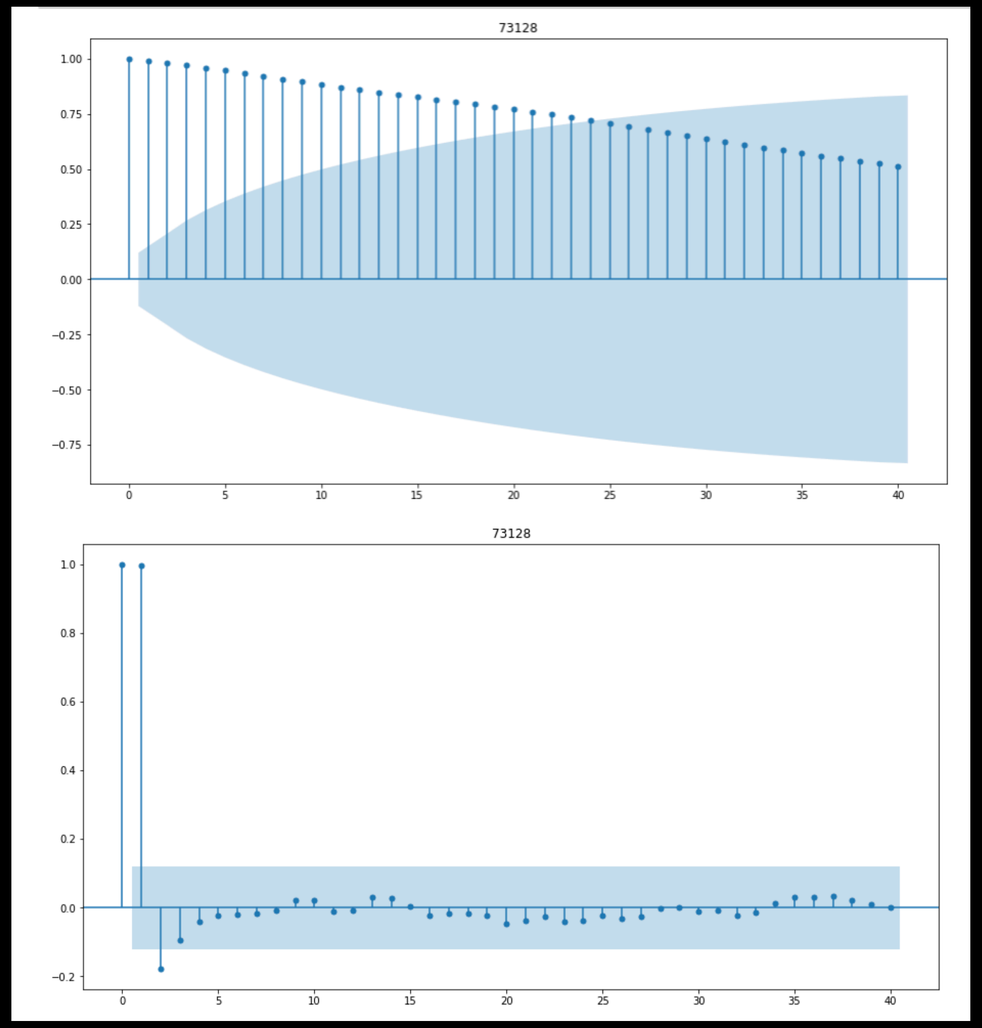


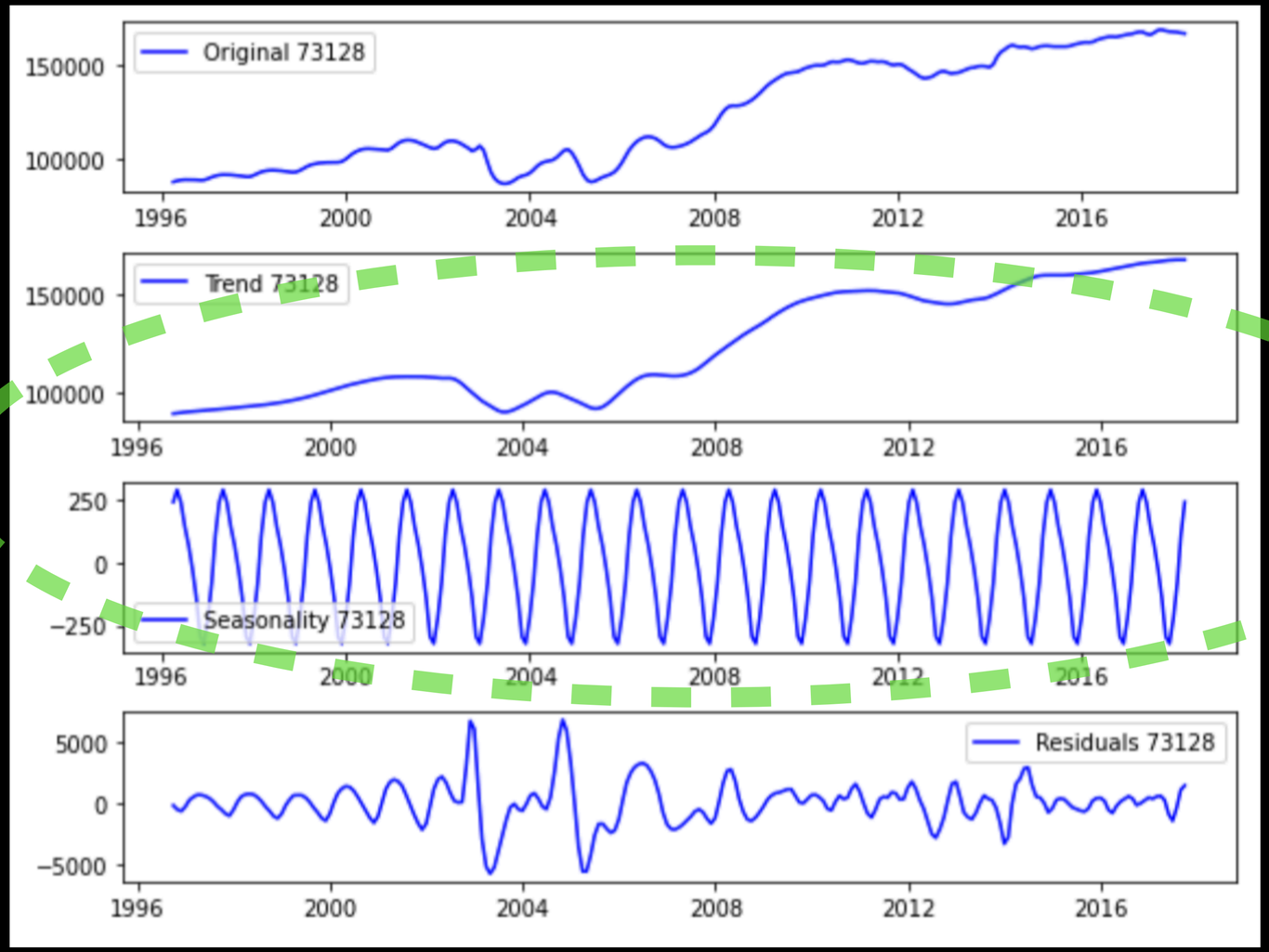

```
Results of df Test:

Test Statistic                  -0.369704
p-value                          0.915036
#Lags Used                       9.000000
Number of Observations Used    255.000000
Critical Value (1%)             -3.456257
Critical Value (5%)             -2.872942
Critical Value (10%)            -2.572846
dtype: float64
```

EDA

Seasonality and autocorrelation





Models

Metrics

- SARIMA did best overall
- Autoregressive did best with training
- 10128 was outlier

AIC

	73128	47220	47281	73064	10128
basline(ar_aic)	4267.3587	4049.4541	4129.2836	4199.6011	6395.1331
ma_aic	5403.6332	5047.272	5003.7064	5400.6042	7358.0852
arma_aic	3998.1584	3778.5214	3868.752	3914.7435	6161.2386
arima_aic	4195.0679	4001.3068	4106.1145	4121.6675	6644.4452
autoARIMA_aic	3791.7243	3475.0392	3682.3084	3614.8004	6045.7595
autoSARIMA_aic	3521.2707	3225.2751	3419.6796	3333.6708	5567.3629

Train RMSE %

	73128	47220	47281	73064	10128
basline(ar_rmse)	0.0205	0.0092	0.0148	0.0209	0.0257
ma_rmse	0.1118	0.0683	0.0673	0.1169	0.1429
arma_rmse	0.0466	0.0703	0.0734	0.0456	0.049
arima_rmse	0.0472	0.0706	0.074	0.0462	0.0511
autoARIMA_rmse	0.0494	0.0764	0.0784	0.0491	0.0514
autoSARIMA_rmse	0.0596	0.09	0.0936	0.0583	0.0673

Test RMSE %

	73128	47220	47281	73064	10128
basline(ar_rmse)	0.0145	0.0395	0.0666	0.0248	0.0254
ma_rmse	0.2727	0.2349	0.271	0.2887	0.3219
arma_rmse	0.0159	0.0359	0.0662	0.0234	0.0294
arima_rmse	0.0145	0.0358	0.0674	0.0229	0.0203
autoARIMA_rmse	0.0186	0.0291	0.0689	0.0129	0.0451
autoSARIMA_rmse	0.0089	0.0157	0.057	0.0078	0.0727

Models

Final Models

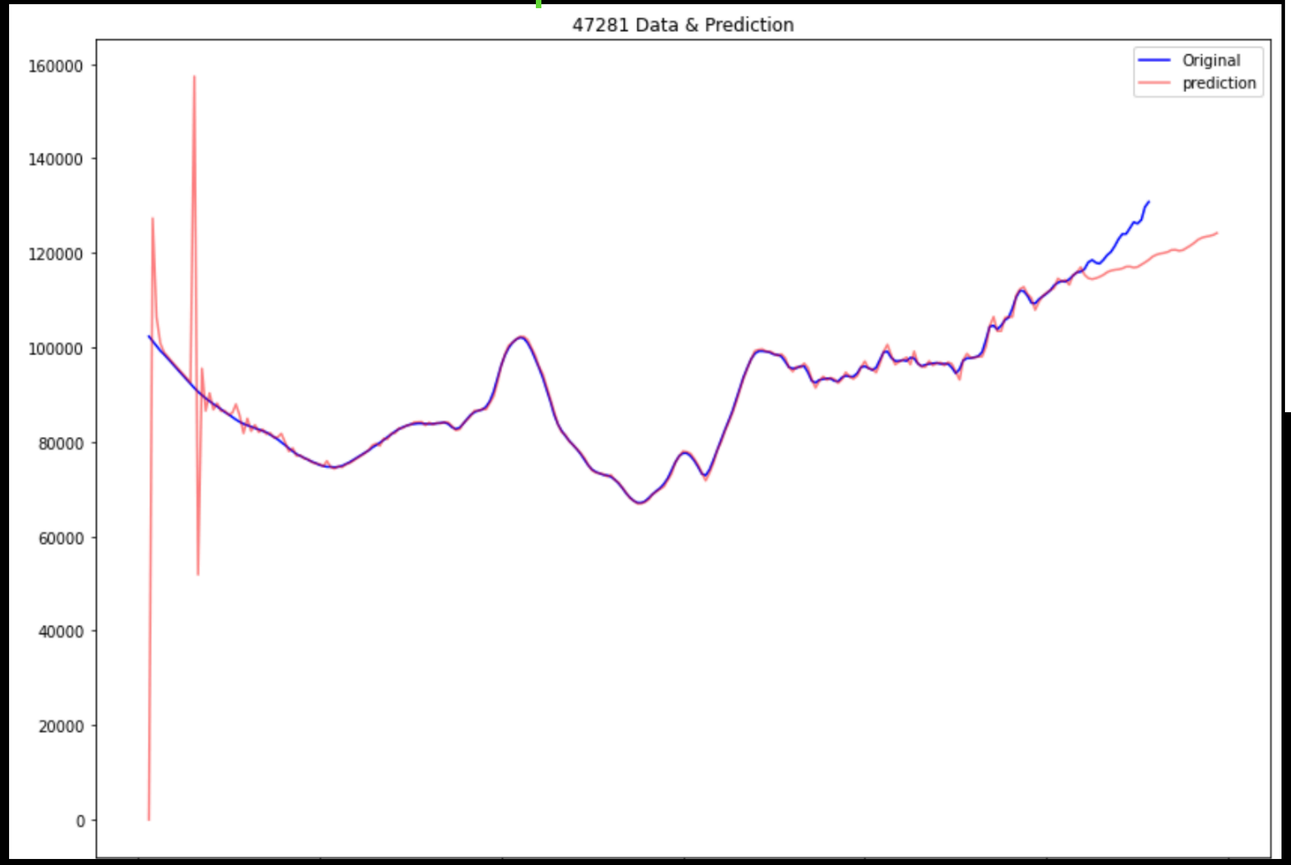
- Best model: SARIMAX(1, 1, 1)x(1, 1, 1, 12) | zipcode 10128
- Best model: SARIMAX(1, 1, 1)x(0, 1, 1, 12) | all other zip codes
- AIC lower by 17.75% (average)
- Test RMSE% lower by 76.85% (average)

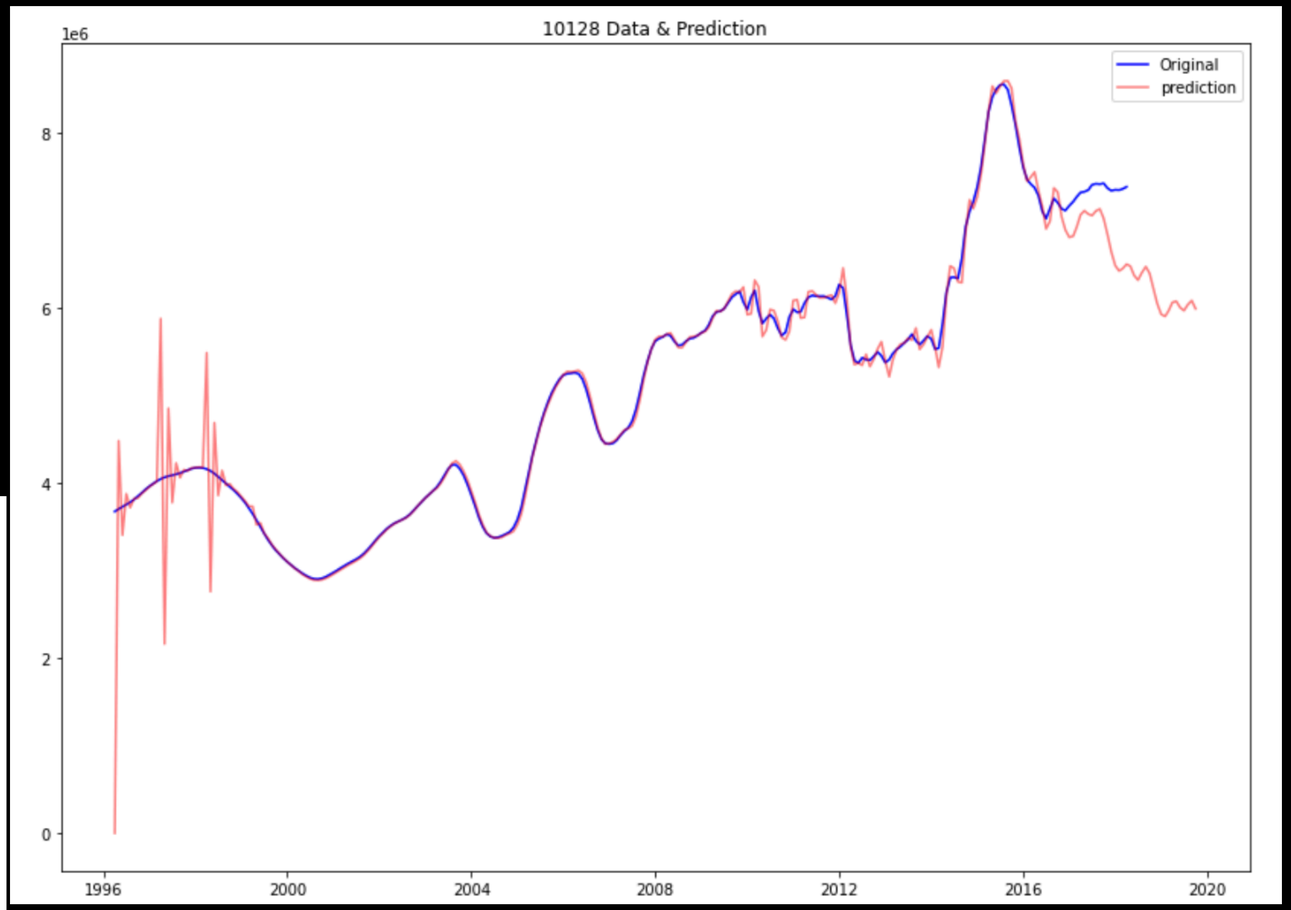
Results

- 73128 Oklahoma City, Oklahoma
  - total return% = 2.48 %
  - CAGR = 1.65 %
- 47220 Brownstown, Indiana
  - total return% = 3.04%
  - CAGR = 2.02 %
- 47281 Vallonia, Indiana
  - total return% = 4.19%
  - CAGR = 2.78%
- 73064 Mustang, Oklahoma
  - total return% = 3.97%
  - CAGR = 2.63%
- 10128 New York City, New York
  - total return% = -7.51%
  - CAGR = -5.07%
-

Results

47281 | Best





10128| Worst

Conclusions

- 47281 (Vallonia, Indiana) is potentially worth investing in:
  - Average national 18 month increase in home price is 5.39%
  - Best researched zip code 18 month increase in home price is 4.19%
- Avoid investing in 10128 (New York, New York)
- “Low Risk” areas seem to offer lower returns
- A broad portfolio of many zip codes/ regions is worth considering

Future Work

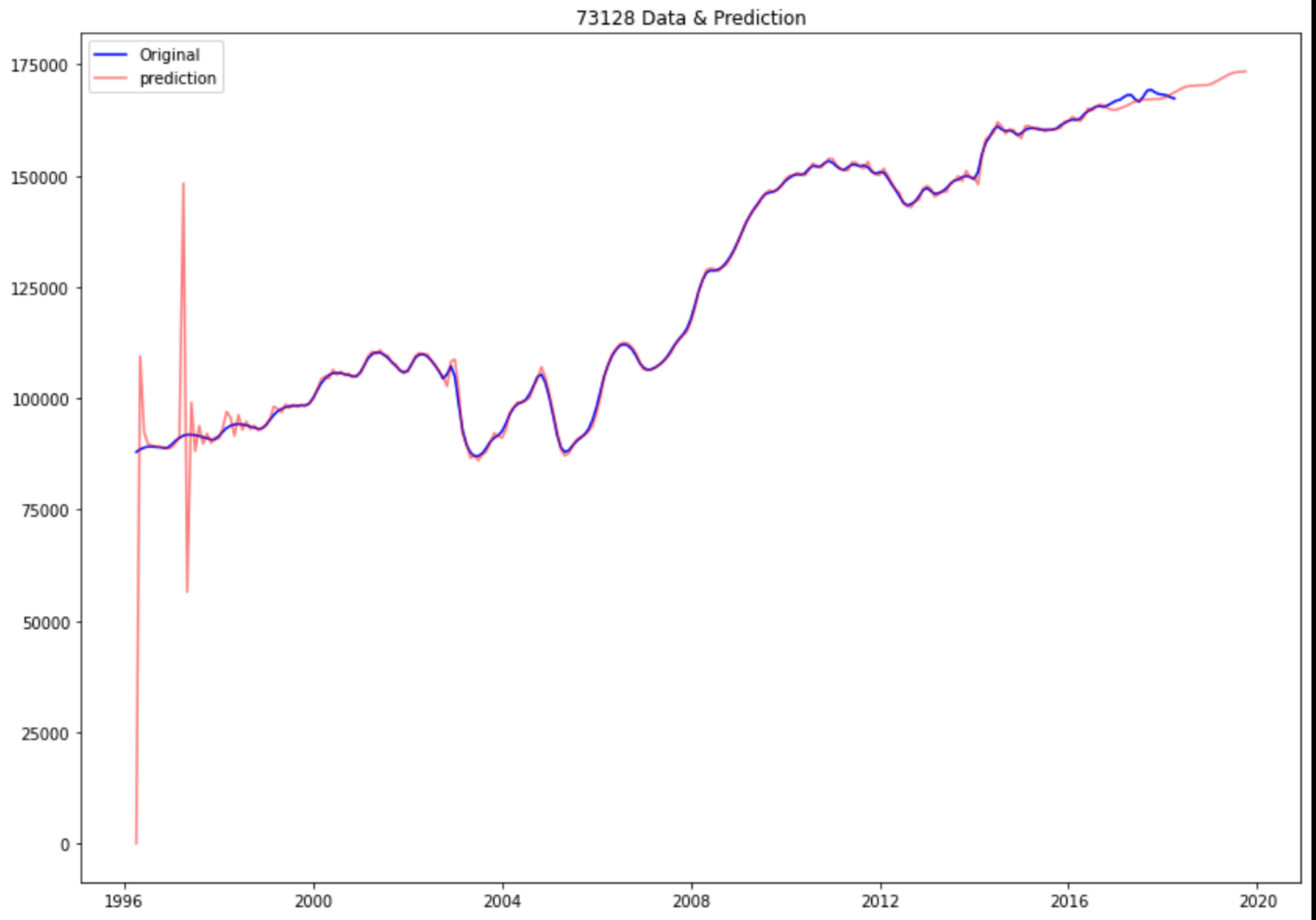
- Look at returns during other downturns
  - Is Vallonia, Indiana really a lower risk investment
- Leverage
  - Look into how using mortgages/loans could effect these returns
- Look at other zip codes that did well during the Great Recession
- Consider exogenous variables


Thank you
Questions?
Link: https://github.com/s-shader/real-estat-investment-analysis-and-projection

Appendix

73128: final model

```
pdq          (1, 1, 1)
pdqs     (0, 1, 1, 12)
aic       3800.430773
Name: 59, dtype: object
```

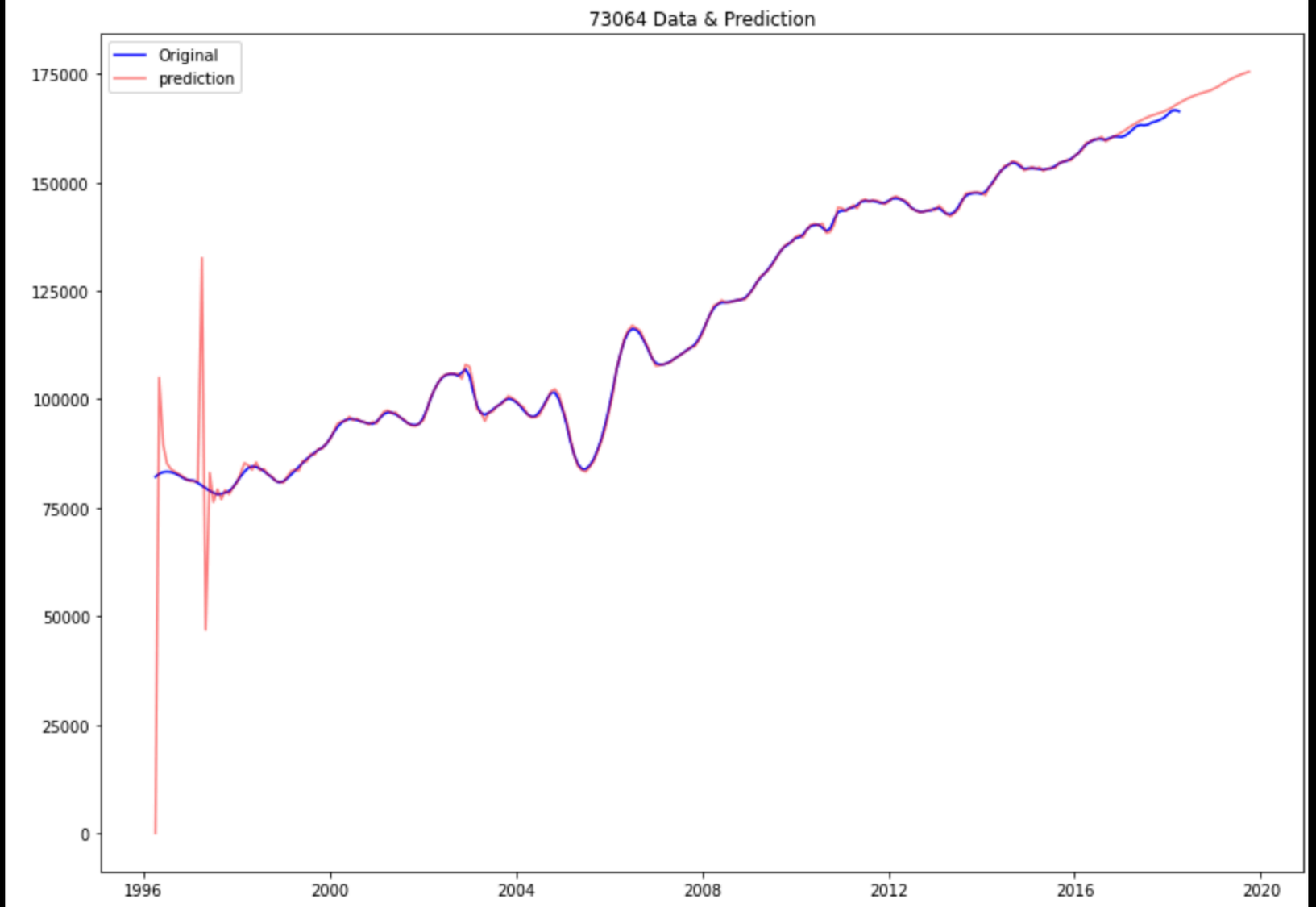



```
total return% =  2.481671530 %
CAGR =  1.647679095 %
```

Appendix

73064: final model

```
pdq          (1, 1, 1)
pdqs     (0, 1, 1, 12)
aic        3591.812249
Name: 59, dtype: object
```

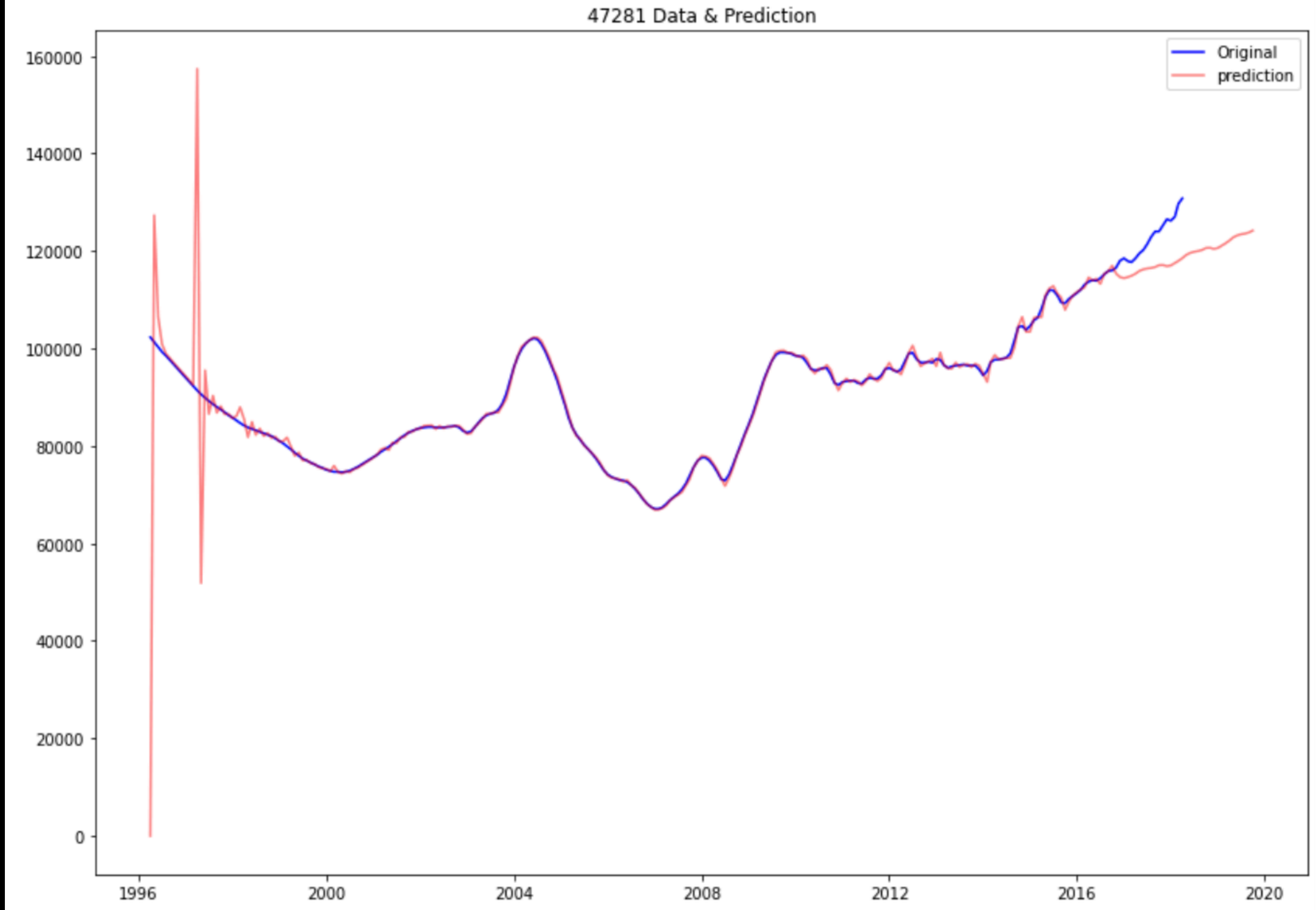



```
total return% =  3.967578398 %
CAGR =  2.627862983 %
```

Appendix

47281: final model

```
pdq          (1, 1, 1)
pdqs     (0, 1, 1, 12)
aic        3738.931681
Name: 59, dtype: object
```

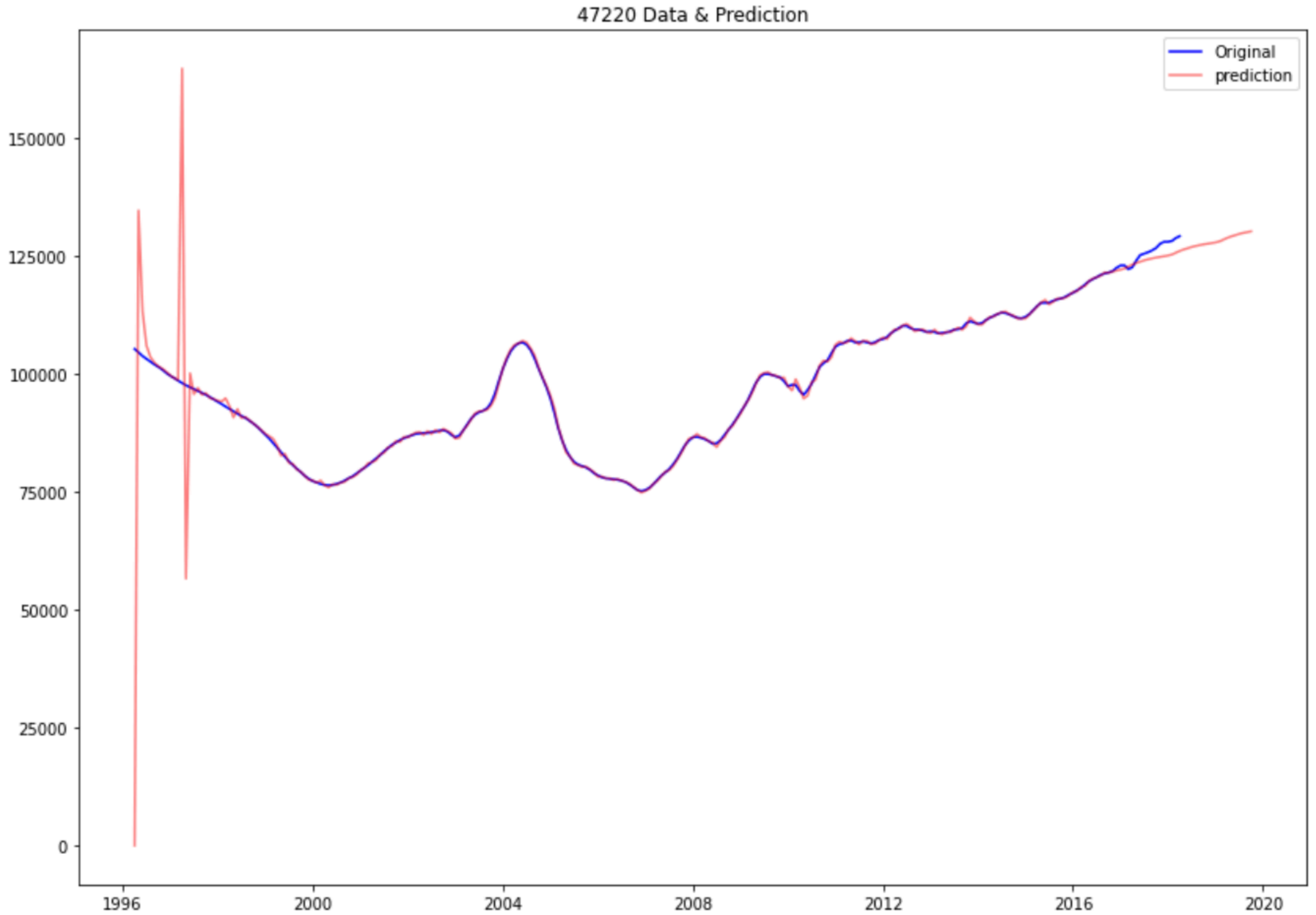



```
total return% =  4.191992211 %
CAGR =  2.775491266 %
```

Appendix

47220: final model

```
pdq          (1, 1, 1)
pdqs     (0, 1, 1, 12)
aic         3513.63664
Name: 59, dtype: object
```

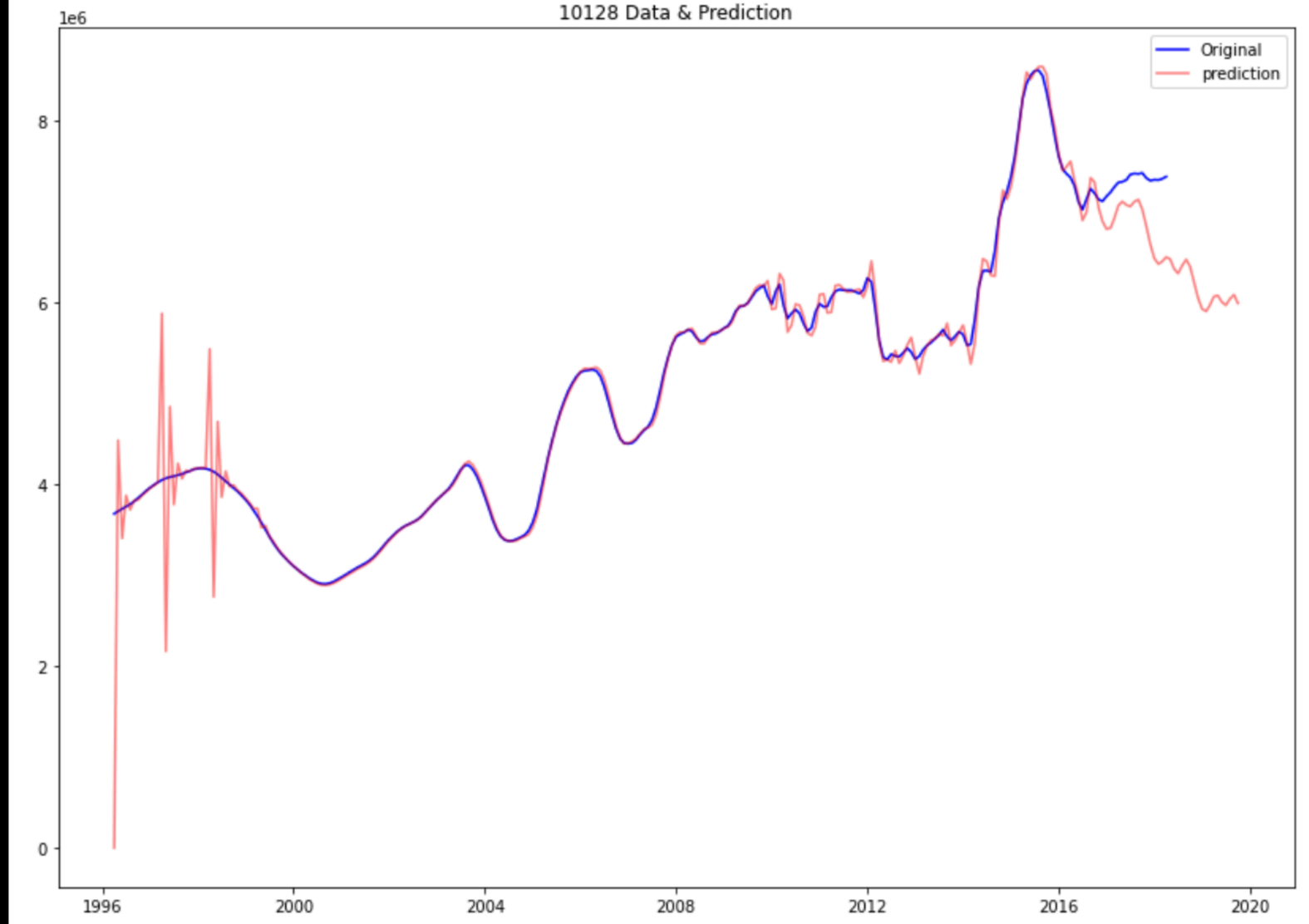



```
total return% =  3.041004162 %
CAGR =  2.017197340 %
```

Apendix

10128: final model

```
pdq          (1, 1, 1)
pdqs     (1, 1, 1, 12)
aic         6015.795433
Name: 63, dtype: object
```




```
total return% =  -7.505679441 %
CAGR =  -5.068565317 %
```