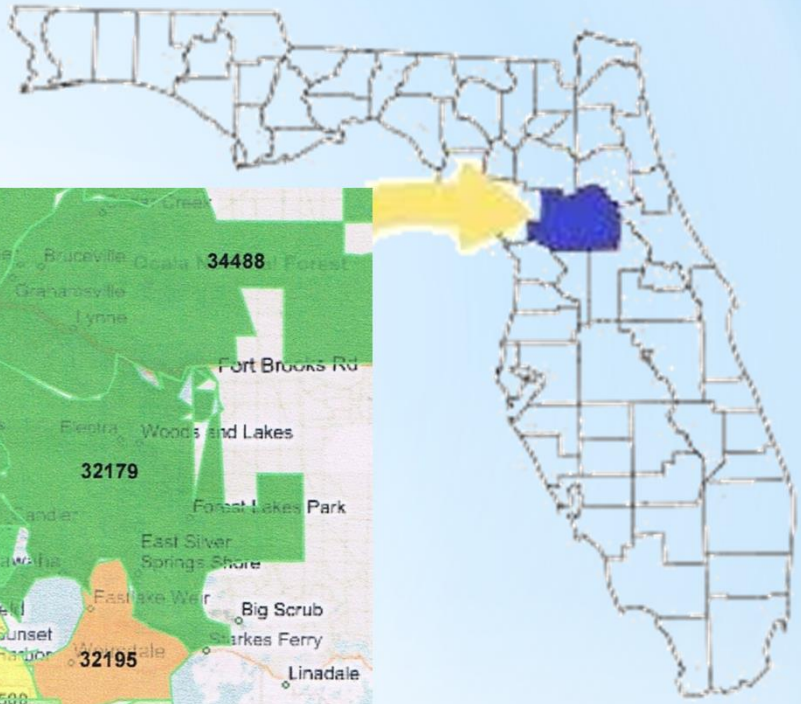
Zillow Ocala Best ZIPCODE Prediction

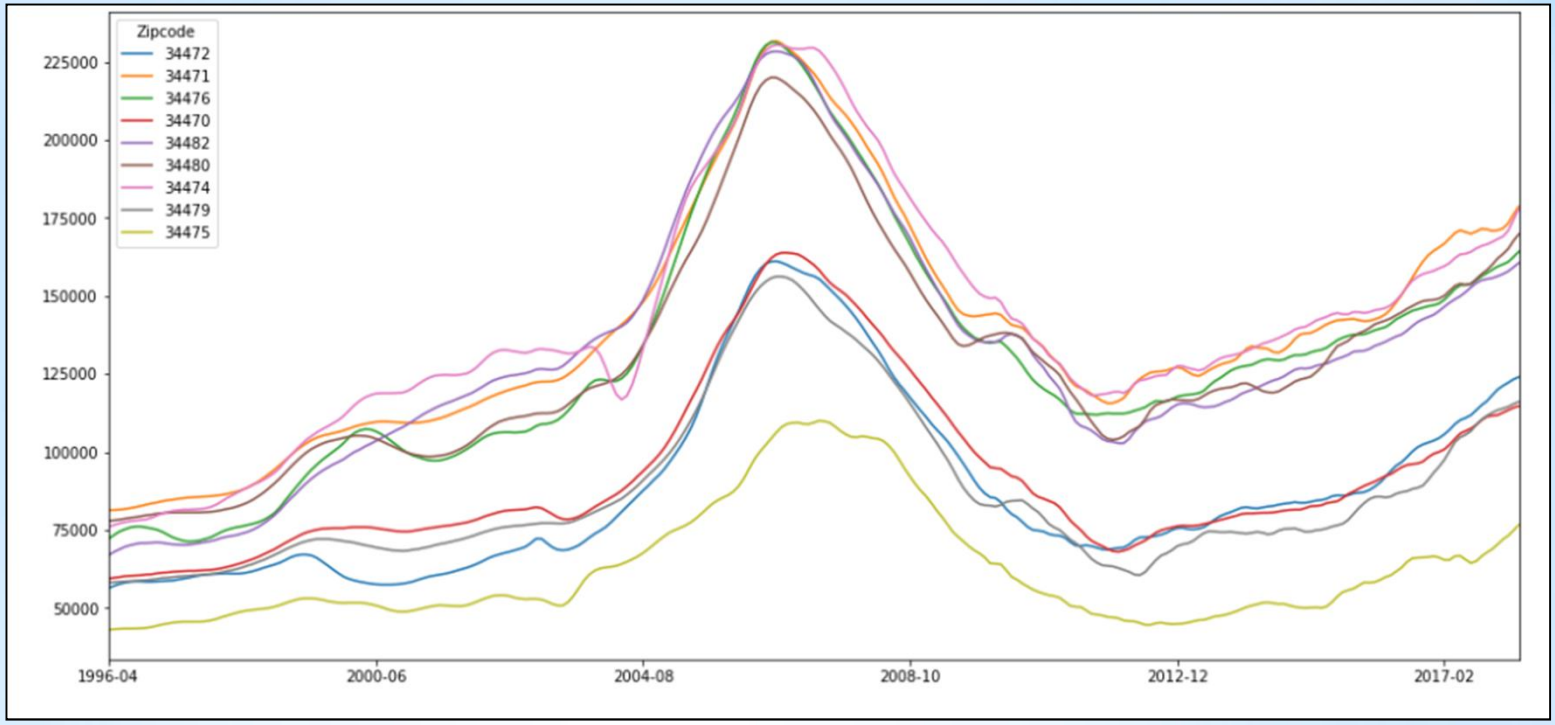
Goal :
To Predict the best ZIPCODES to invest to have assured returns



Assumption:

- $125,000 to deploy upfront.
- Minimum 3years and Max 10 years
- You seek to maximize growth potential by tapping into home value appreciation in“ Horse Capital of the World" OCALA

EDA



- 9 unique Zip Codes are there in Ocala
- The data time line is monthly from 1996-Apr to 2018 Apr

Top 5 ZIPCODES Based on Largest Mean

- 34474 : 141436.981132
- 34471 : 139155.094340
- 34482 : 132369.811321
- 34476 : 130906.415094
- 34480 : 129782.264151

	34471	34474	34476	34480	34482
Test Statistic	-2.195014	-2.588167	-2.448012	-2.137101	-2.756722
p-value	0.208060	0.095468	0.128641	0.229860	0.064727
#Lags Used	5.000000	15.000000	5.000000	9.000000	13.000000
Number of Observations Used	259.000000	249.000000	259.000000	255.000000	251.000000
Critical Value (1%)	-3.455853	-3.456888	-3.455853	-3.456257	-3.456674
Critical Value (5%)	-2.872765	-2.873219	-2.872765	-2.872942	-2.873125
Critical Value (10%)	-2.572752	-2.572994	-2.572752	-2.572846	-2.572944

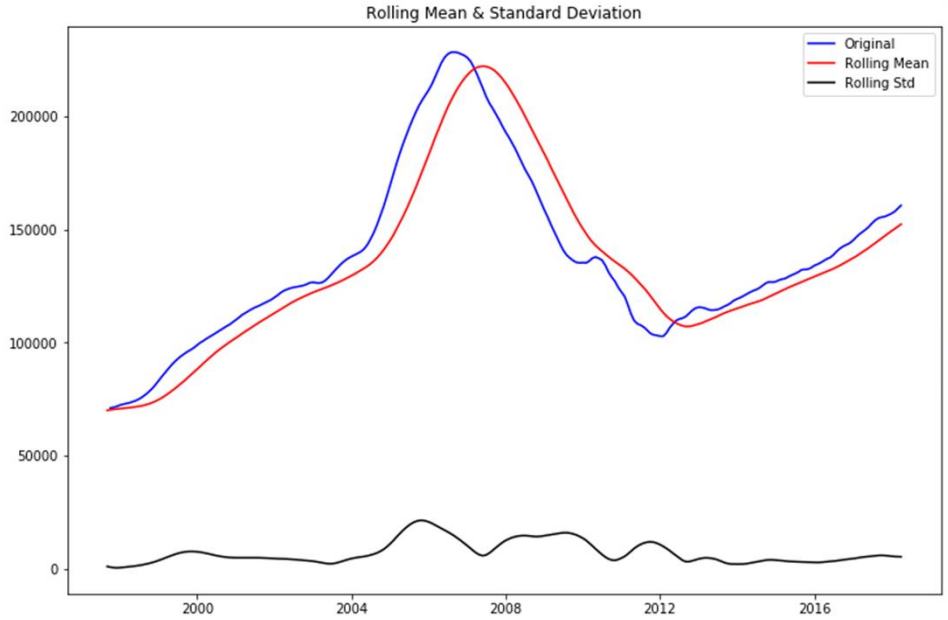
- 34474 & 34482 have the lowest p-value , higher lags
- Test statistics less than critical value (5%) Reject the null hypothesis
- All ZIPCODES Failed to Reject the null hypothesis & time series is Non Stationary

	34471	34474	34476	34480	34482
Test Statistic	-1.902749	-2.605347	-2.114425	-2.142880	-2.648984
p-value	0.330769	0.091897	0.238754	0.227625	0.083290
#Lags Used	7.000000	15.000000	5.000000	11.000000	13.000000
Number of Observations Used	257.000000	249.000000	259.000000	253.000000	251.000000
Critical Value (1%)	-3.456054	-3.456888	-3.455853	-3.456464	-3.456674
Critical Value (5%)	-2.872853	-2.873219	-2.872765	-2.873033	-2.873125
Critical Value (10%)	-2.572799	-2.572994	-2.572752	-2.572895	-2.572944

- 34474 & 34482 have the lowest p-value , higher lags
- Test statistics less than critical value (5%) Reject the null hypothesis
- All ZIPCODES Failed to Reject the null hypothesis & time series is Non Stationary

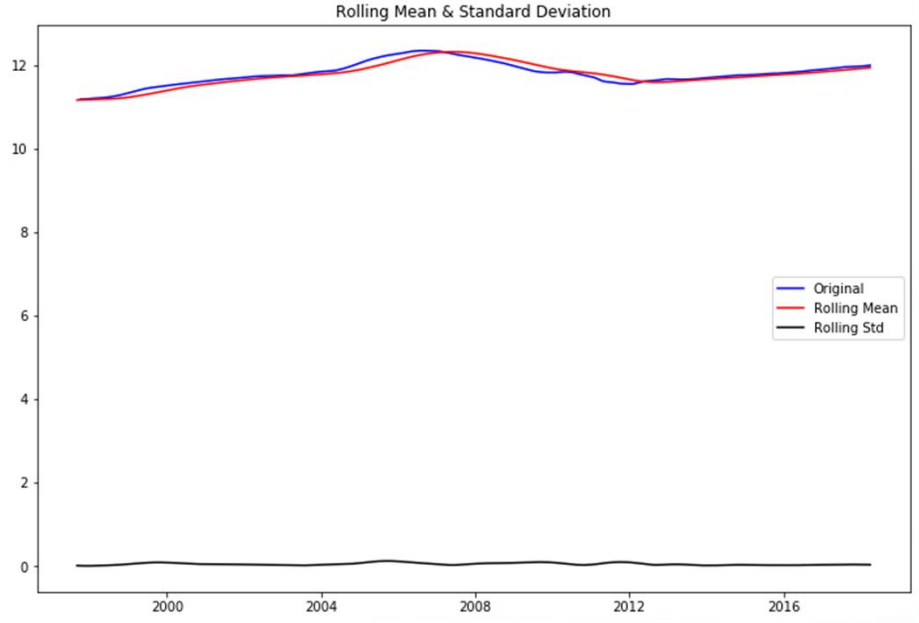
Testing Stationary

```
output34482 = test_stationarity(df_Ocala['34482'], 18)
```



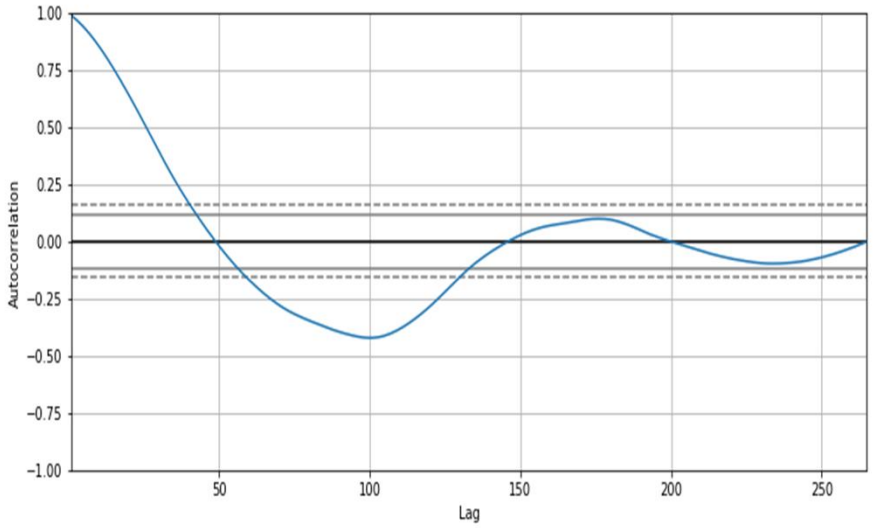
Testing Stationary post log_tran

```
log_output34482 = test_stationarity(data82_log, 18)
```



34482

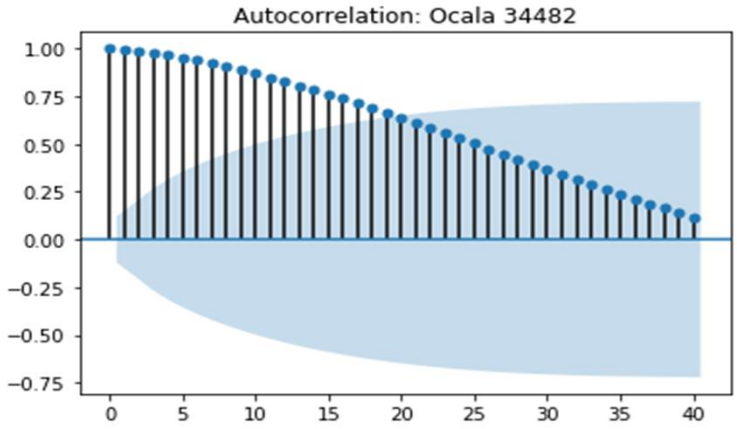
```
plt.figure(figsize=(12,5))
pd.plotting.autocorrelation_plot(log_82);
```



Auto Correlation at 40 lags

- The red line meets the dotted line at around 40 lags

```
from statsmodels.graphics.tsaplots import plot_acf,plot_pacf
title = 'Autocorrelation: Ocala 34482'
lags = 40
plot_acf(df_Ocala['34482'],title=title,lags=lags);
```

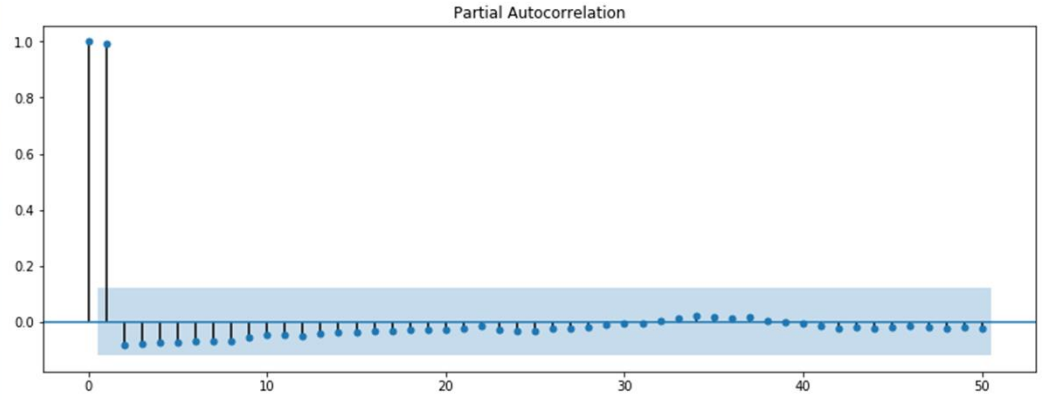


34482

```
from statsmodels.graphics.tsaplots import plot_pacf
from matplotlib.pylab import rcParams

rcParams['figure.figsize'] = 14, 5

plot_pacf(log_82, lags = 50);
```



STEPWISE: Best model: ARIMA(1,1,1)(0,0,0)[12] Total fit time: 27.304 seconds

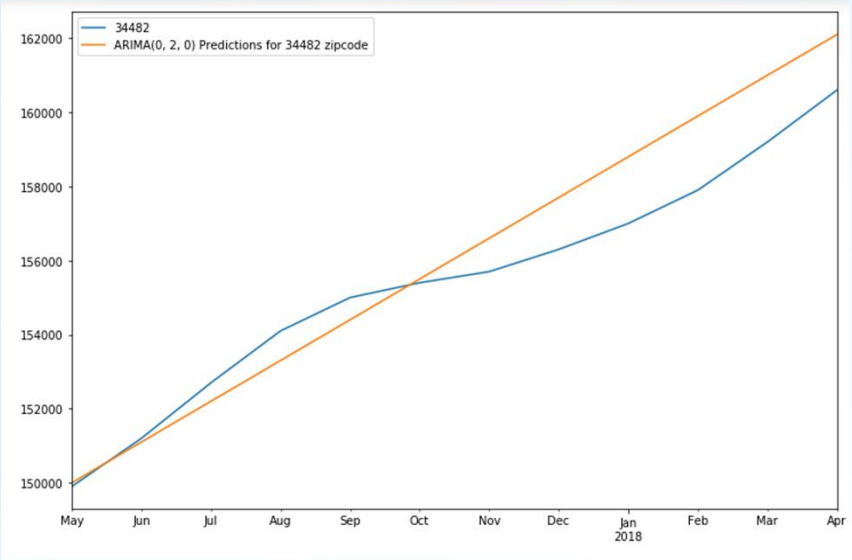
Train_test split
test data only for last 1 year

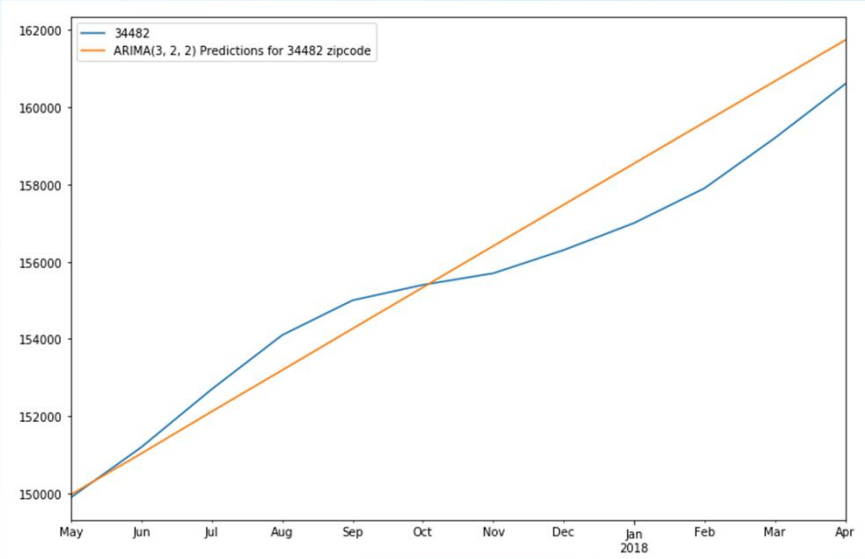
```
train_data = df_Ocala['1996-04-01':'2017-04-01']
test_data= df_Ocala['2017-05-01':'2018-04-01']
```

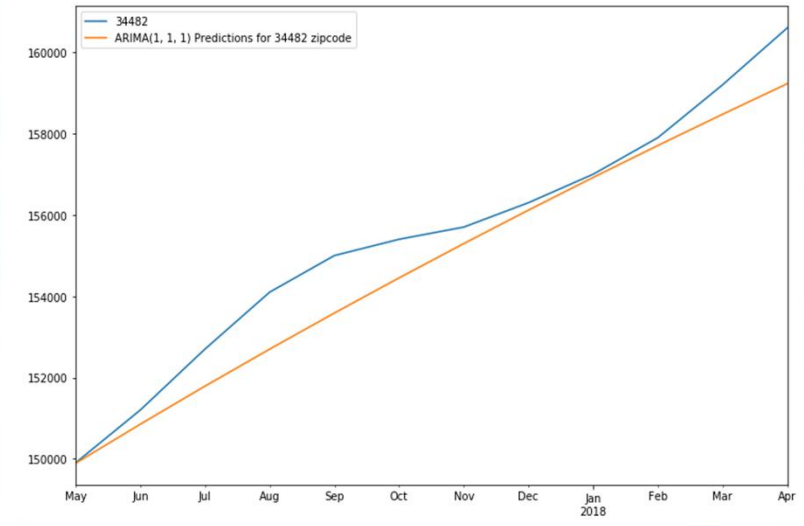
The best ARIMA model based on Auto Arima is (1,1,1)
RAN Three different model for each ZIPCODE for better understanding
1) (ARIMA (0,2,0) Named as model_ZIPCODE , result_ZIPCODE & Prediction_ZIPCODE
2) (ARIMA (3,2,2) Named as model_ZIPCODEa, result_ZIPCODEa & Prediction_ZIPCODEa
3) (ARIMA (1,1,1) Named as model_ZIPCODEb ,result_ZIPCODEb & Prediction_ZIPCODEb

Example ZIPCODE : 34474:
1) (ARIMA (0,2,0) Named as model_74 , result_74 & Prediction_74
2) (ARIMA (3,2,2) Named as model_74a , result_74a & Prediction_74a
3) (ARIMA (1,1,1) Named as model_74b , result_74b & Prediction_74b

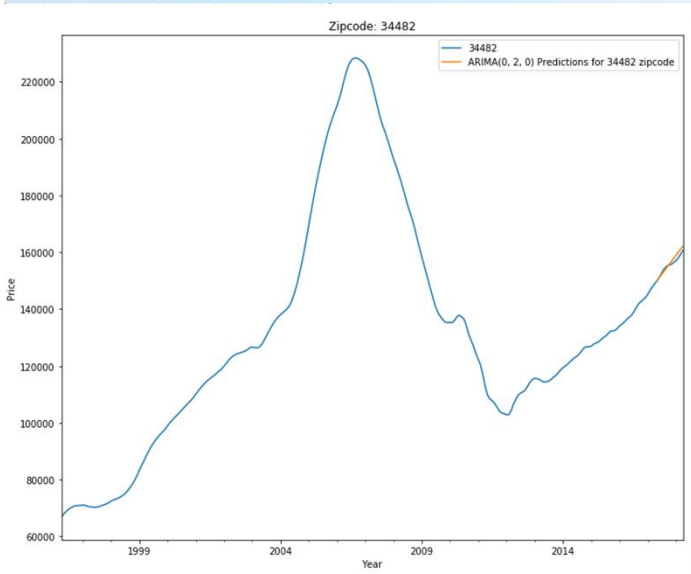
AIC	34471	34474	34476	34480	38882
ARIMA (0,2,0)	3742.858788	3922.84396	3692.912718	3690.450184	3723.647066
ARIMA (3,2,2)	3633.48803	3768.987127	3555.694891	3562.947234	3629.923014
ARIMA (1,1,1)	3650.647672	3811.322639	3574.664799	3612.233775	3665.85778

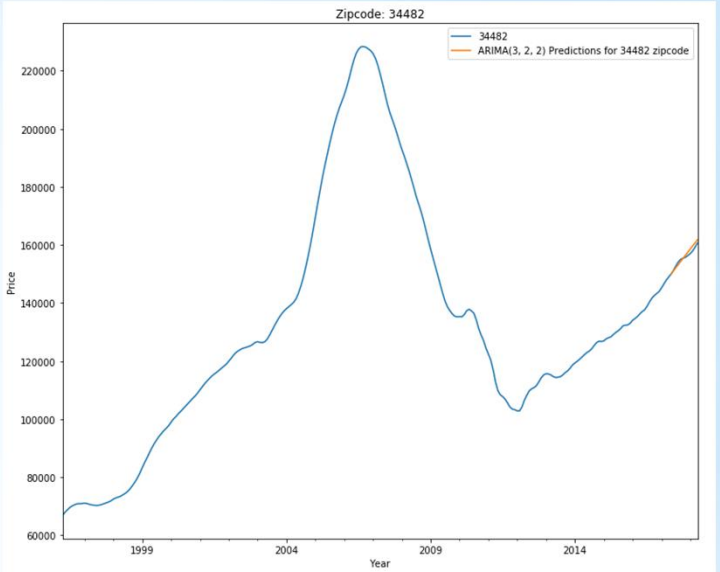


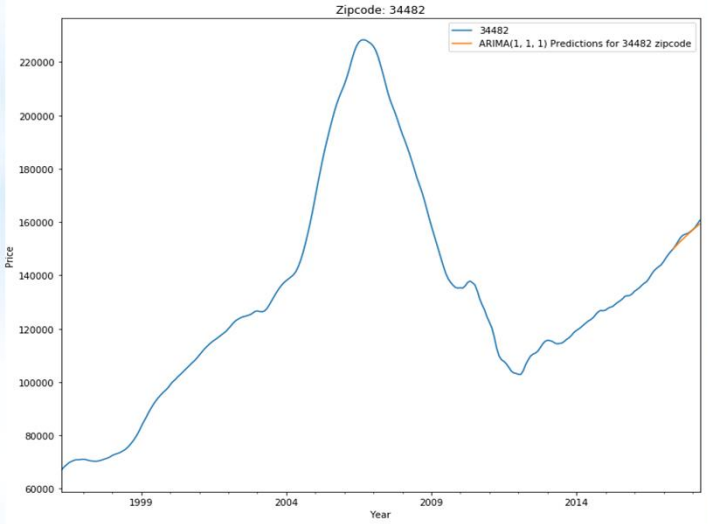




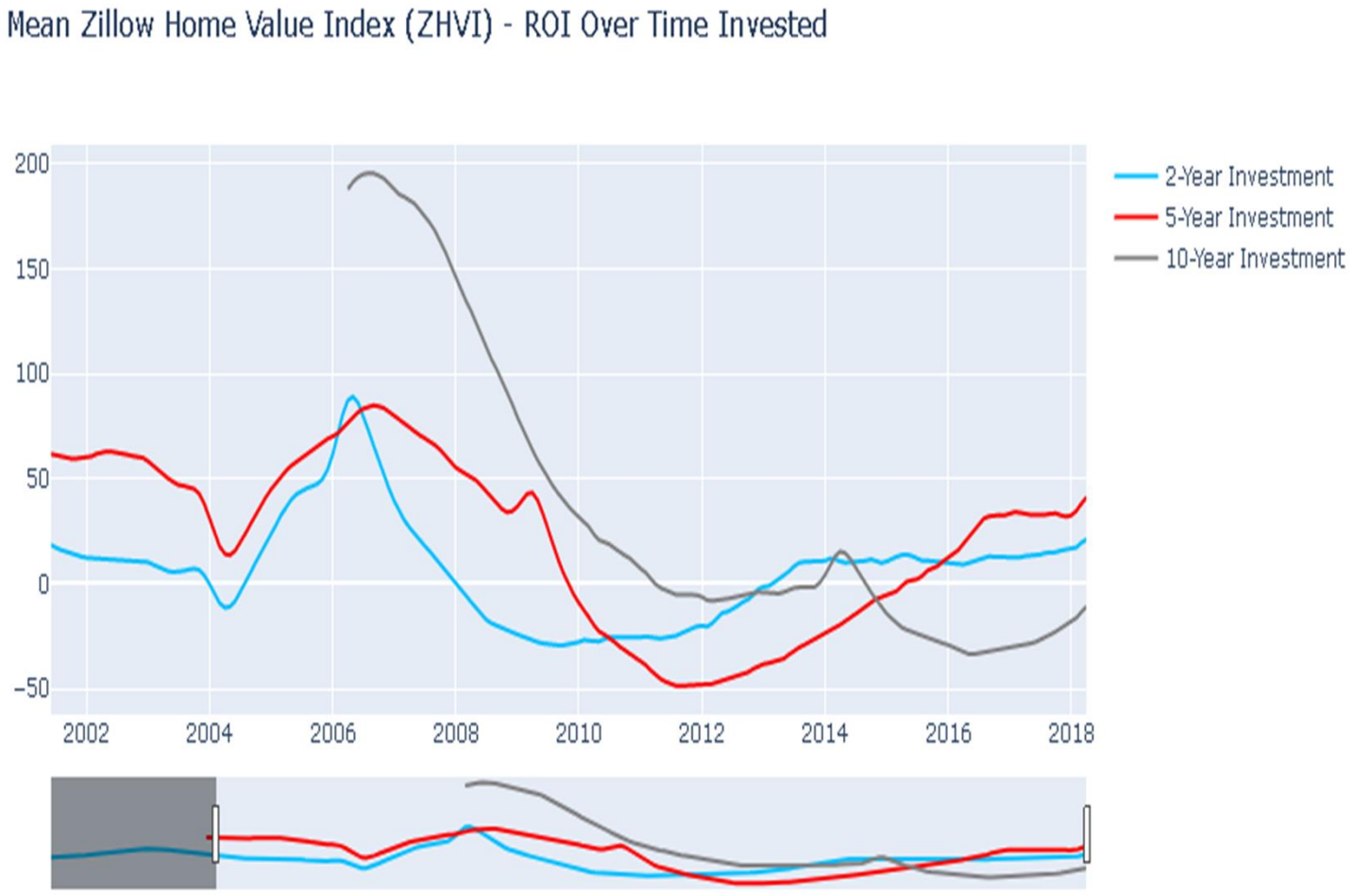
	34471	34474	34476	34480	34482
ARIMA(0,2,0) MSE Error:	102554872.4	25800662.37	45069038.72	15255684.17	1398333.333
ARIMA(0,2,0) RMSE Error:	10126.93796	5079.435241	6713.347803	3905.852553	1182.511452
ARIMA(3,2,2) MSE Error:	63903468.88	101996220.2	65411490.63	6533495.018	1028619.835
ARIMA(3,2,2) RMSE Error:	7993.964529	10099.31781	8087.737053	2556.07023	1014.20897
ARIMA(1,1,1) MSE Error:	47854858.51	13977248.28	26272254.82	4651892.133	706101.9612
ARIMA(1,1,1) RMSE Error:	6917.720615	3738.615824	5125.646771	2156.824548	840.2987333







For Future:


Mean Zillow Home Value Index (ZHVI) - ROI Over Time Invested
200
150
100
50
0
-50
2002
2004
2006
2008
2010
2012
2014
2016
2018
2-Year Investment
5-Year Investment
10-Year Investment

Thank you

- Github project link :

- https://github.com/ksis1st/Zillow_TimeSeries_Prediction

- Linkedin :

- https://www.linkedin.com/in/kishor-shankaranarayan-bab2a311/