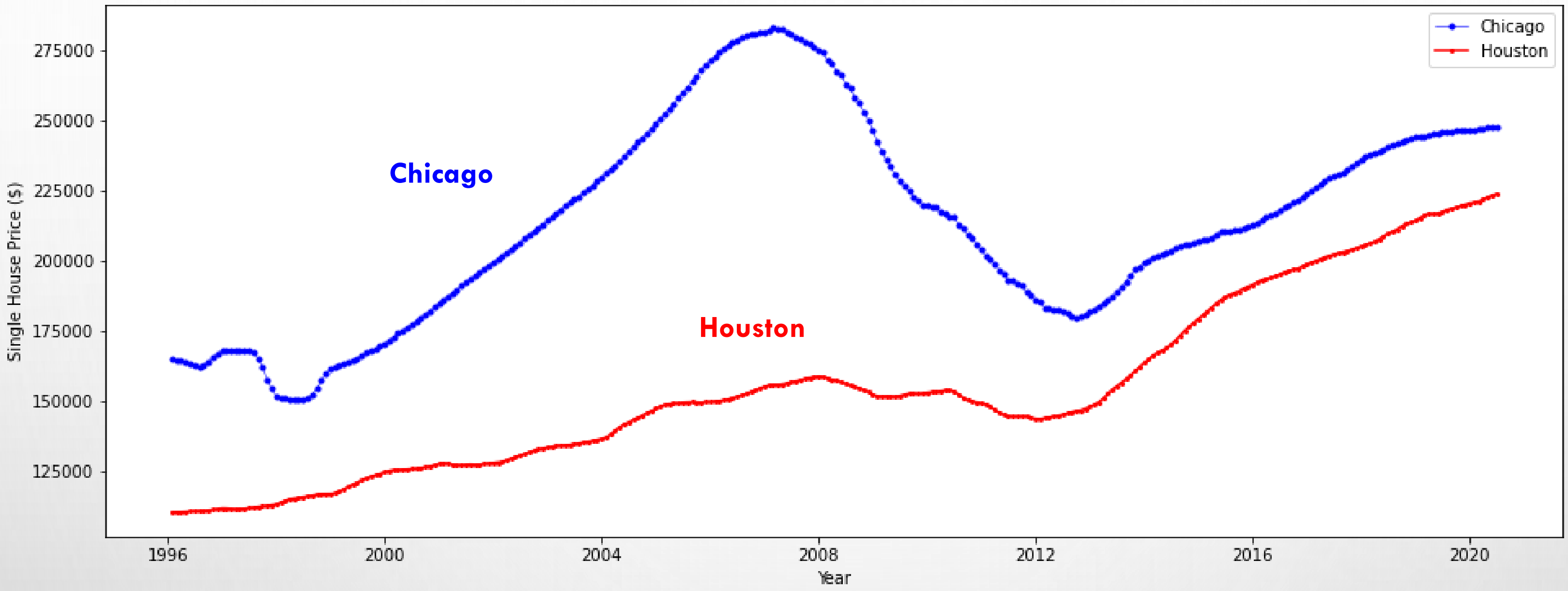
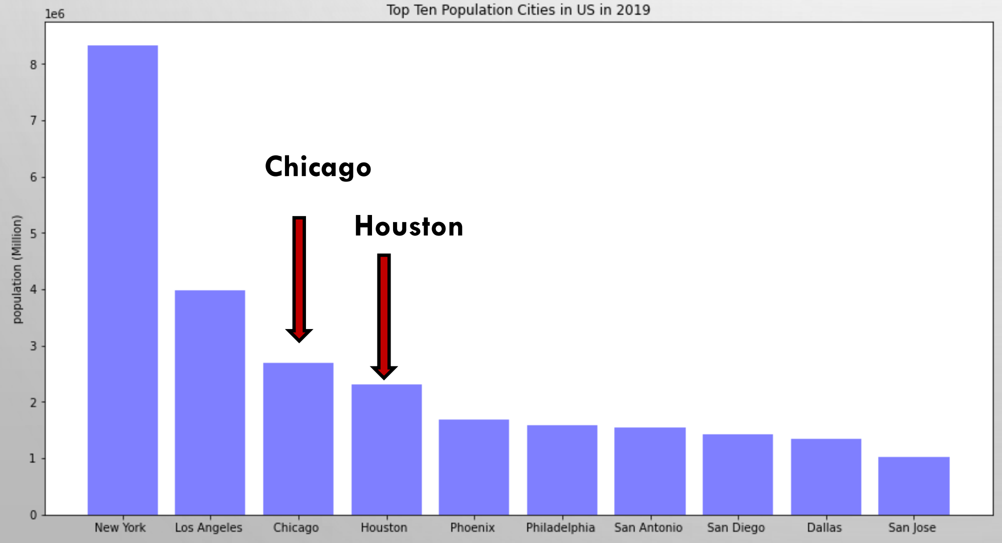
HOUSTON VS CHICAGO HOUSING PRICE

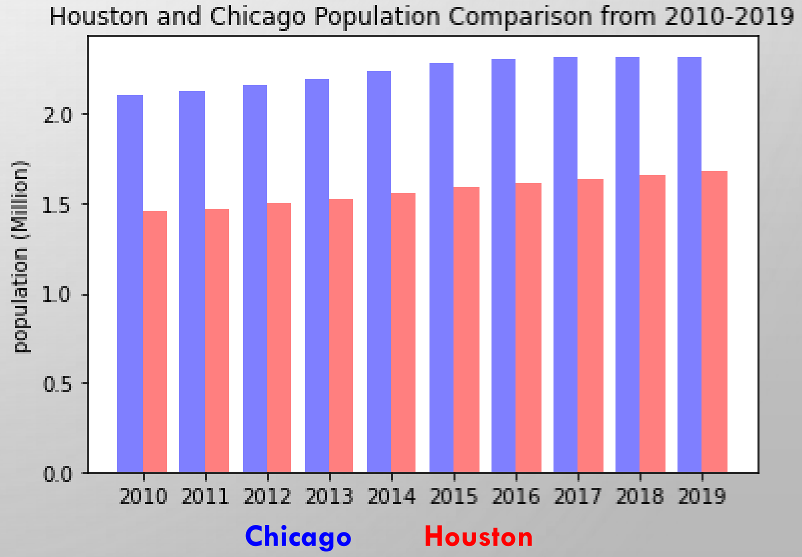
Grace Yu

8/18/2020

MOTIVATION


Chicago
Houston
Chicago
Houston
Single House Price ($)
275000
250000
225000
200000
175000
150000
125000
1996
2000
2004
2008
2012
2016
2020
Year

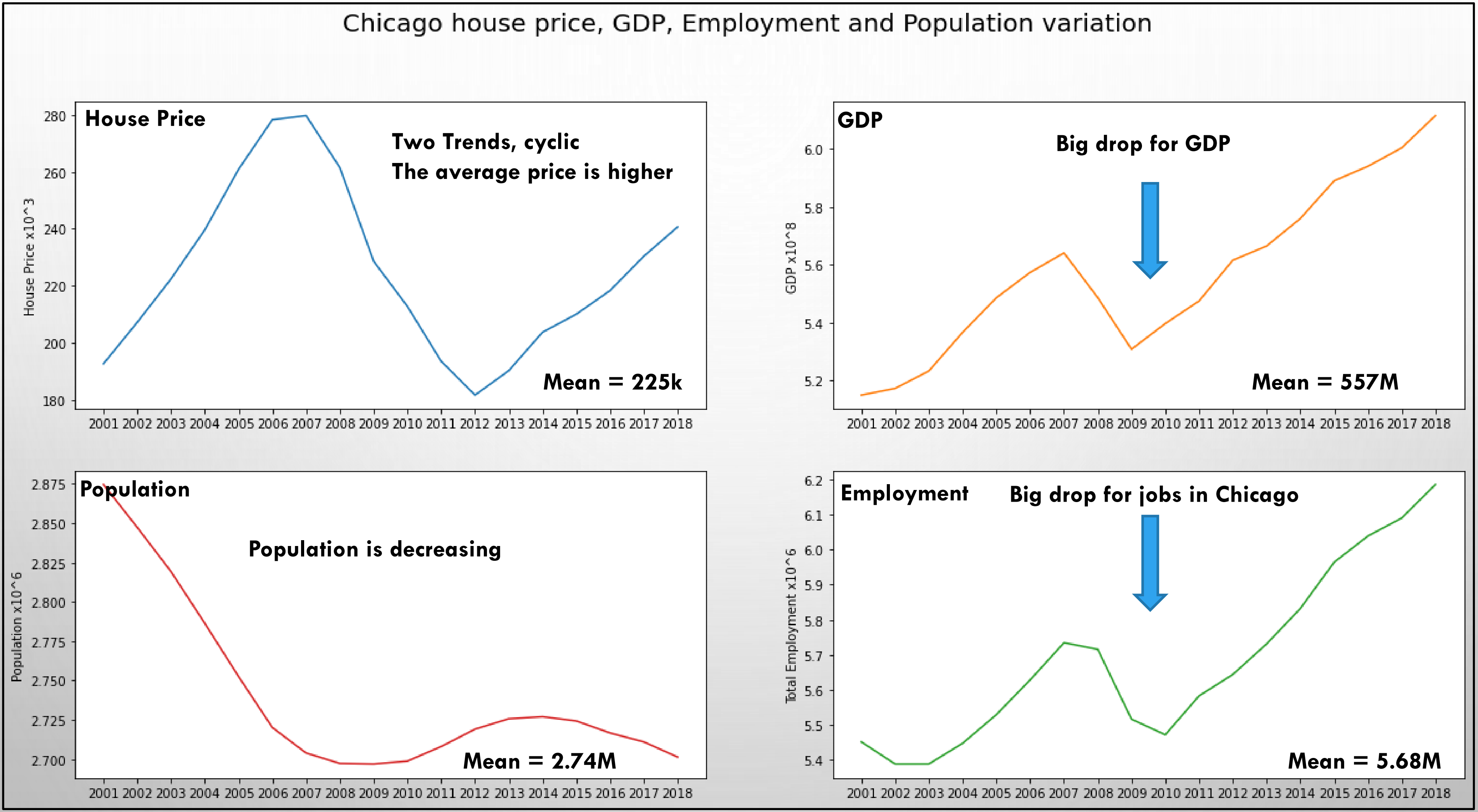

Top Ten Population Cities in US in 2019
1e6
population (Million)
8
7
6
5
4
3
2
1
0
Chicago
Houston
New York
Los Angeles
Chicago
Houston
Phoenix
Philadelphia
San Antonio
San Diego
Dallas
San Jose


Houston and Chicago Population Comparison from 2010-2019
population (Million)
2.0
1.5
1.0
0.5
0.0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Chicago
Houston

HIGHLIGHTS

1. Compared Chicago and Houston, the third and fourth largest city in population, housing price variation over the last two decades;
2. Bridge and weld different data sources. Build a complete dataset which includes house price, GDP, employment opportunities, and population for quantitative analysis;
3. Built multivariable and single variable linear regression model to predict house price;
4. Performed time series analysis to predict house price for the next 20 years.

CHICAGO - ALL FEATURES OVERVIEW



Chicago house price, GDP, Employment and Population variation

House Price

Two Trends, cyclic
The average price is higher

Mean = 225k

GDP

Big drop for GDP

Mean = 557M

Population

Population is decreasing

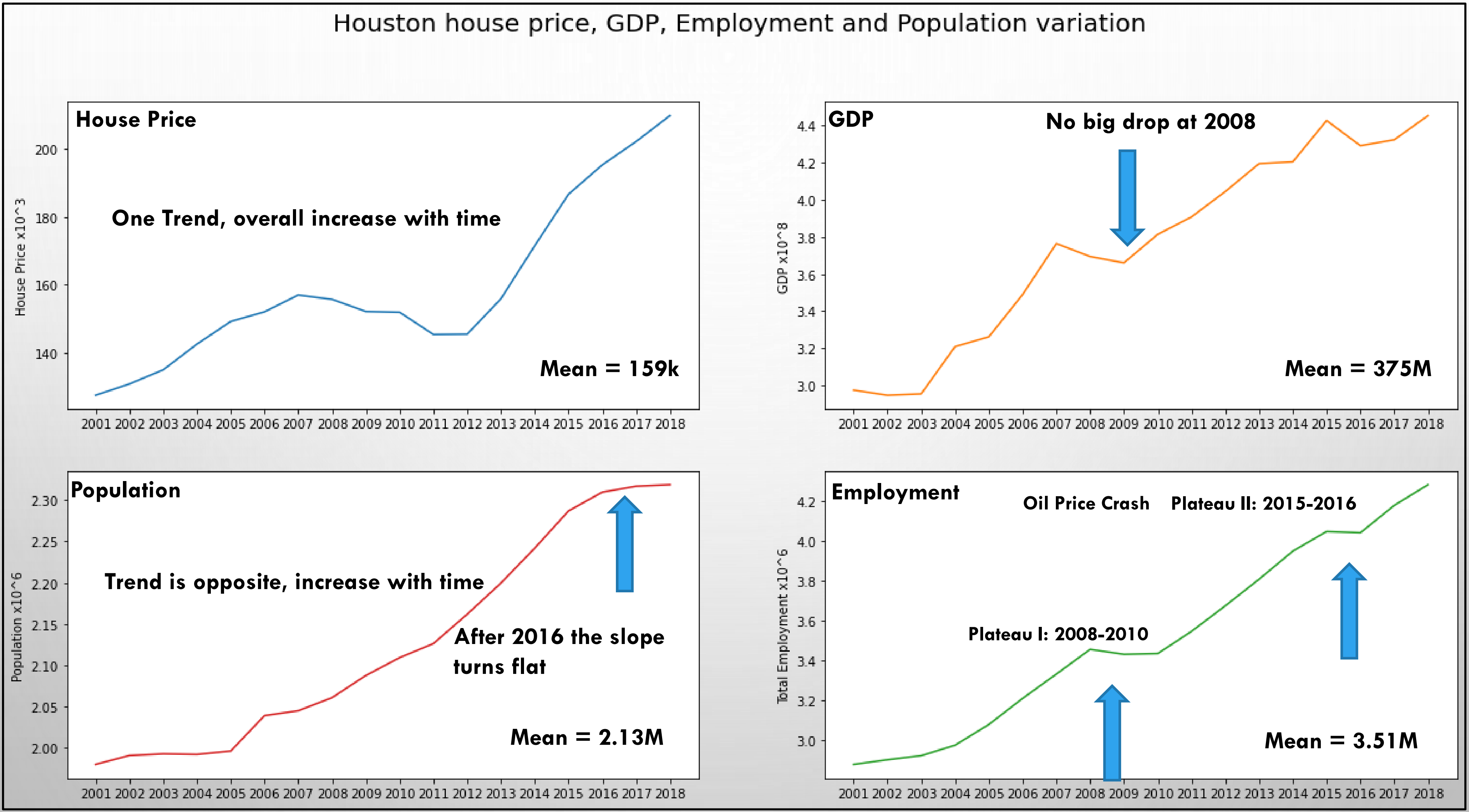
Mean = 2.74M

Employment

Big drop for jobs in Chicago

Mean = 5.68M

HOUSTON - ALL FEATURES OVERVIEW



Houston house price, GDP, Employment and Population variation

House Price

One Trend, overall increase with time

Mean = 159k

GDP

No big drop at 2008

Mean = 375M

Population

Trend is opposite, increase with time

After 2016 the slope turns flat

Mean = 2.13M

Employment

Oil Price Crash

Plateau II: 2015-2016

Plateau I: 2008-2010

Mean = 3.51M

CHICAGO - ALL FEATURES CROSS PLOT

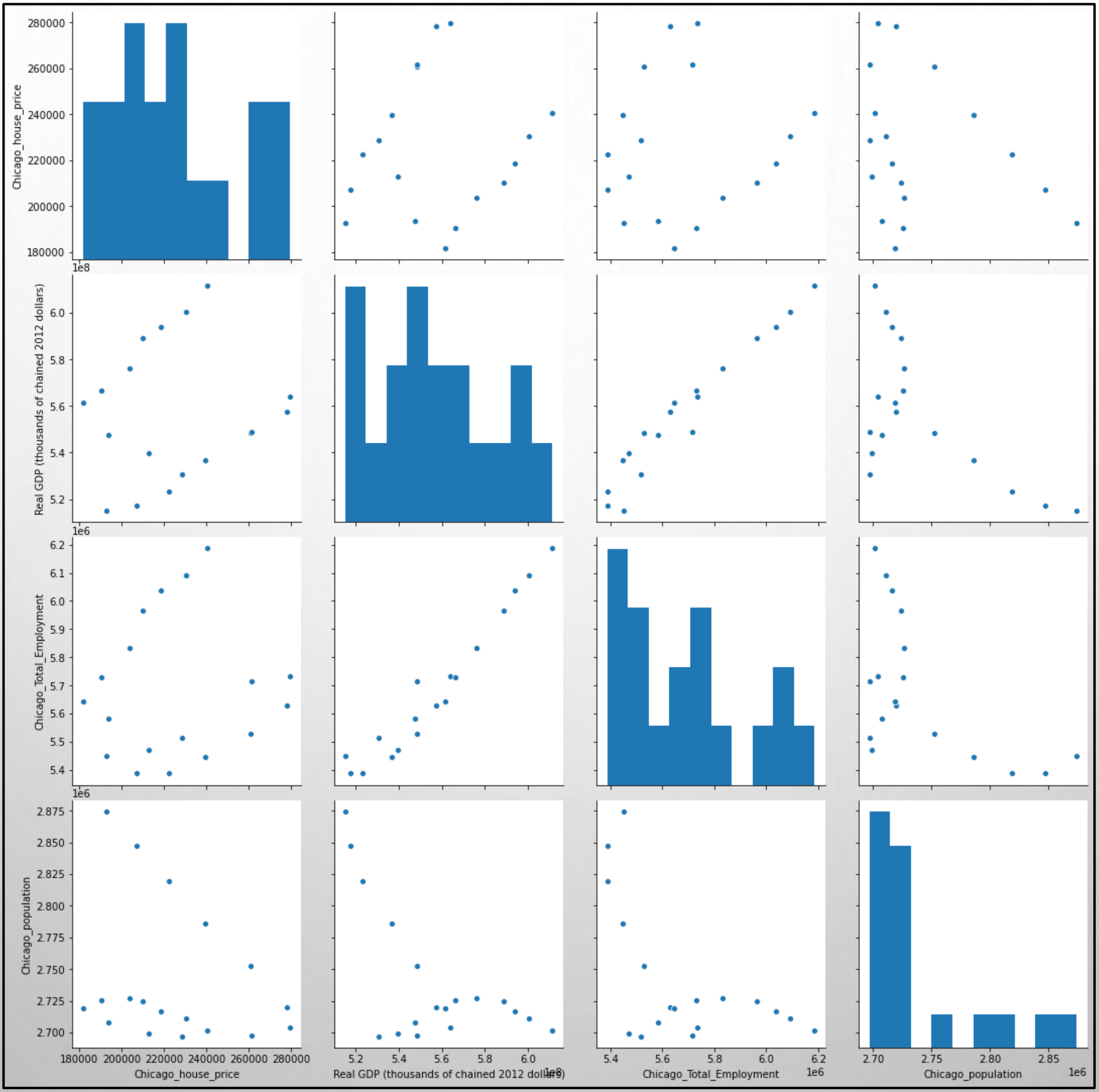
House Price GDP Employment Population

House Price

GDP

Employment

Population



HOUSTON - ALL FEATURES CROSS PLOT

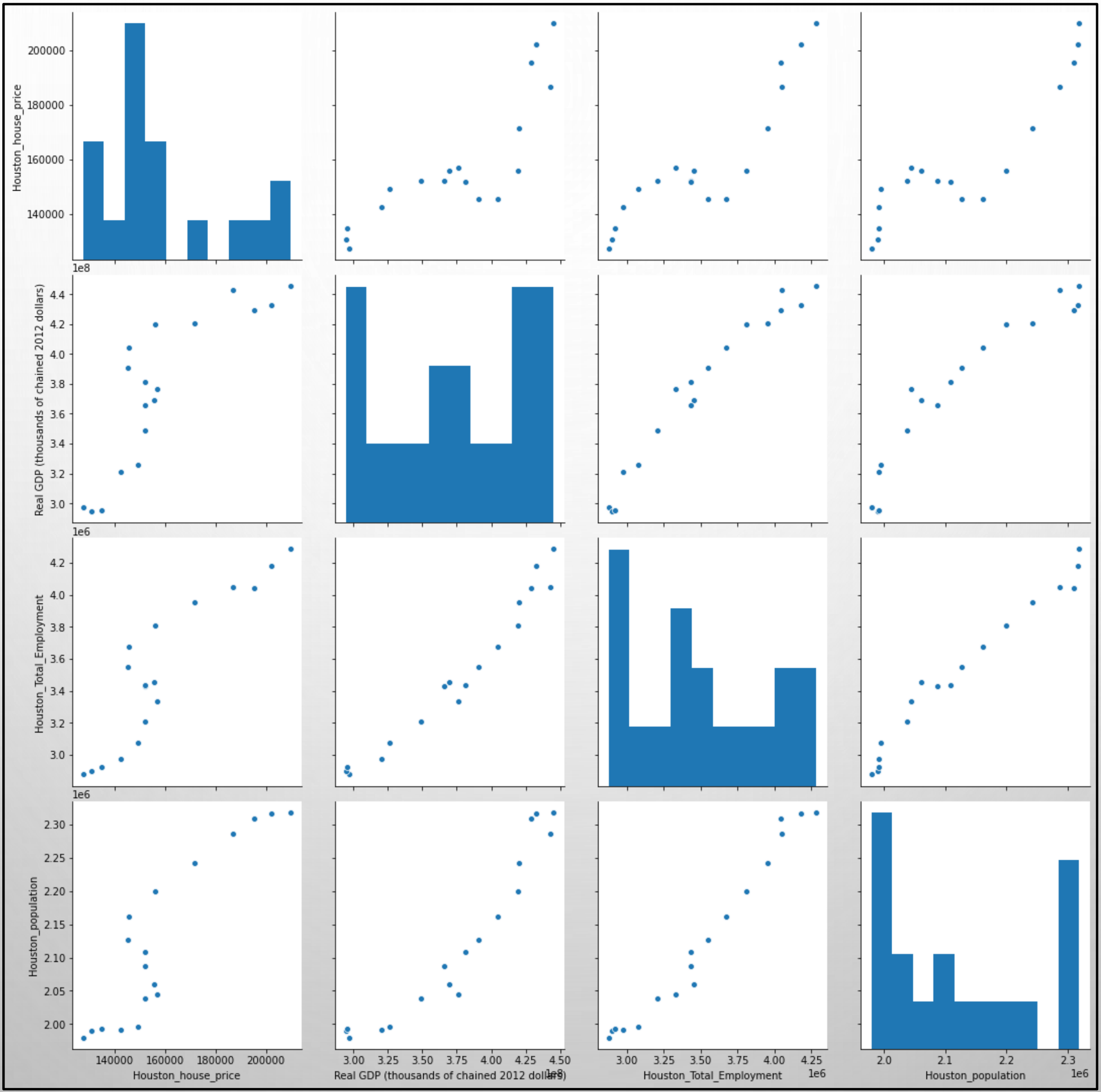
House Price GDP Employment Population

House Price

GDP

Employment

Population



CHICAGO - ALL FEATURES CORRELATION COEFFICIENT

	House Price	Real GDP (thousands of chained 2012 dollars)	Total Employment	Population
House Price	1.000000	0.100000	0.090000	-0.260000
Real GDP (thousands of chained 2012 dollars)	0.100000	1.000000	0.970000	-0.630000
Total Employment	0.090000	0.970000	1.000000	-0.560000
Population	-0.260000	-0.630000	-0.560000	1.000000

- House price has very low correlation with GDP, total employment and population.
- GDP has very high correlation with total employment
- GDP has negative relationship with population

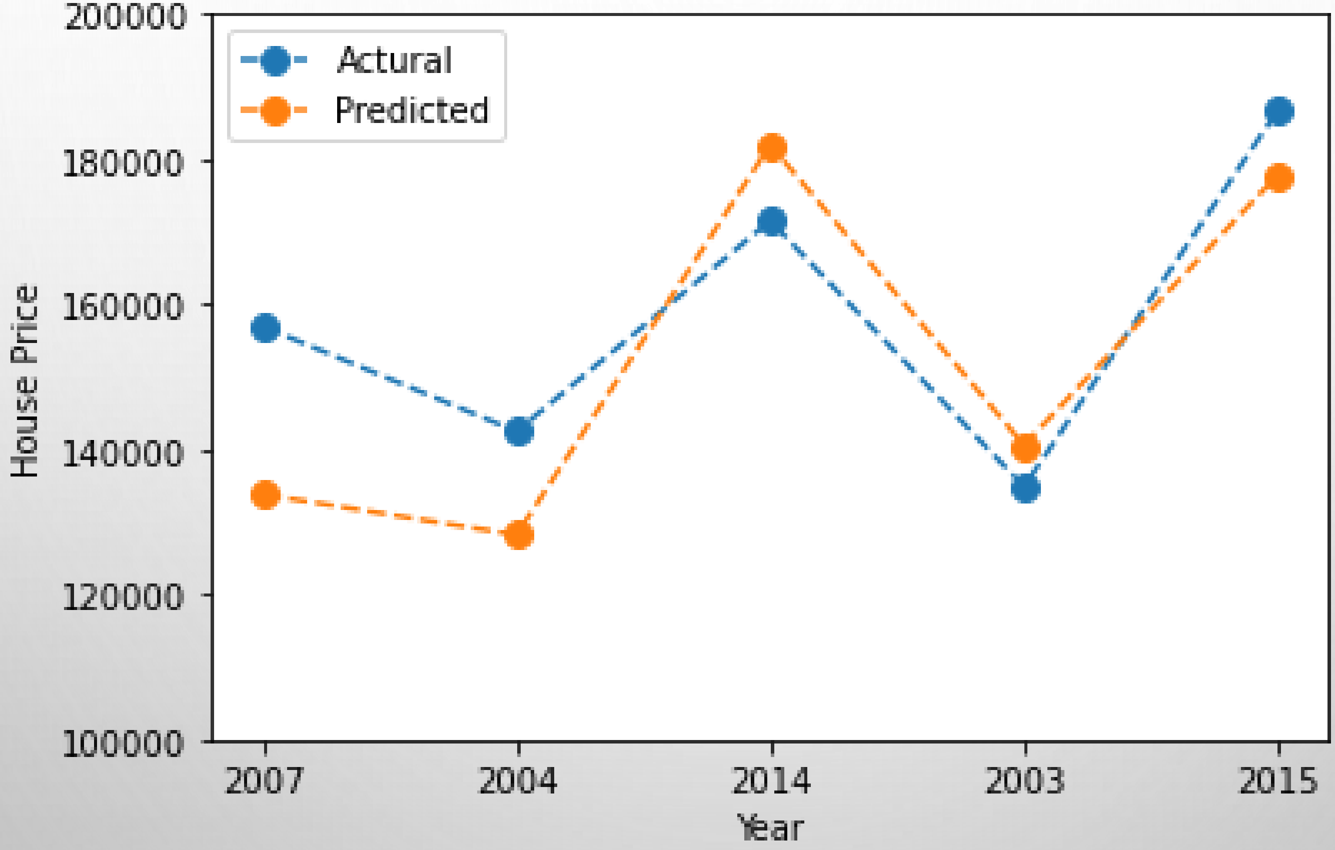
HOUSTON - ALL FEATURES CORRELATION COEFFICIENT

	House Price	Real GDP (thousands of chained 2012 dollars)	Total Employment	Population
House Price	1.000000	0.840000	0.900000	0.900000
Real GDP (thousands of chained 2012 dollars)	0.840000	1.000000	0.980000	0.940000
Total Employment	0.900000	0.980000	1.000000	0.980000
Population	0.900000	0.940000	0.980000	1.000000

- Instead, House price has very high correlation with GDP, total employment and population.
- GDP, total employment and population has very high correlation with each other, so we can drop features like total employment and population, decrease the prediction model order to one.
- For Linear Regression model both multi-variable and single variable models are tested.

HOUSTON Model1 – Multi-variable Linear Regression

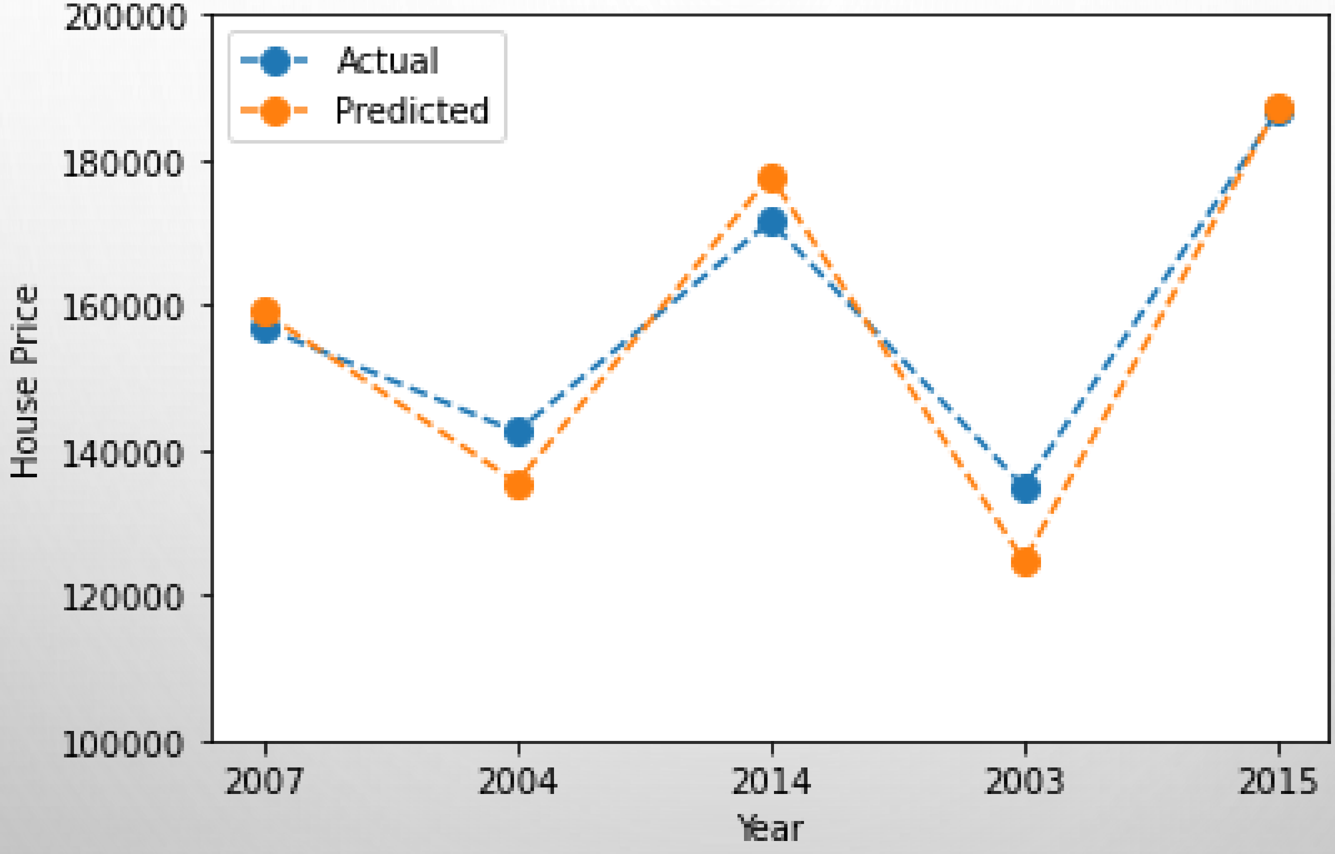
HOUSE PRICE = -38699.16*GDP+61567.69*TOTAL_EMPLOYMENT-1527.66*POPULATION



Data sources: 17 data points , Training: 75%, instances (12), Testing: 25%, instances (5)

HOUSTON Model2 – Single-variable Linear Regression

HOUSE PRICE = 0.00042*GDP-695.3363



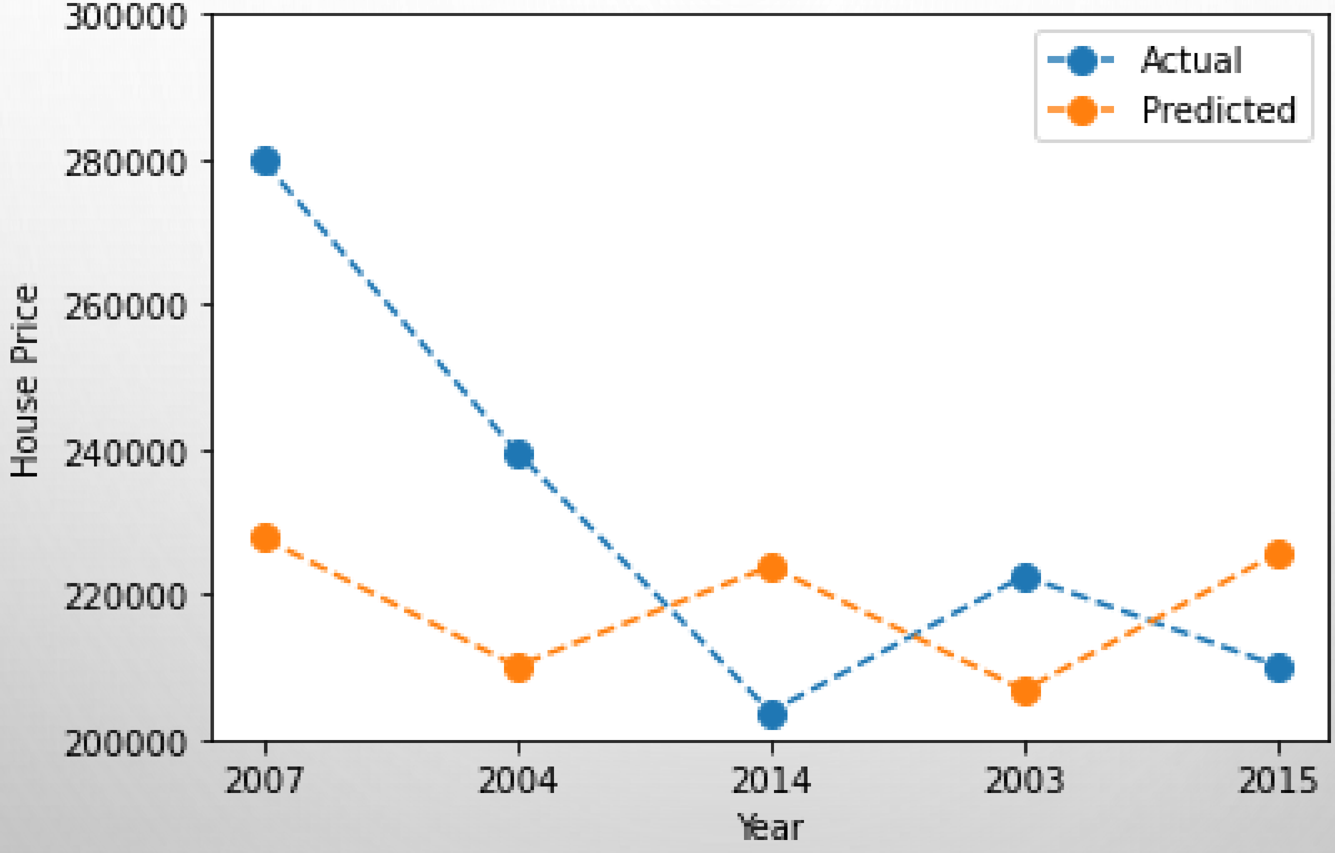
Better Fit

The simpler the better

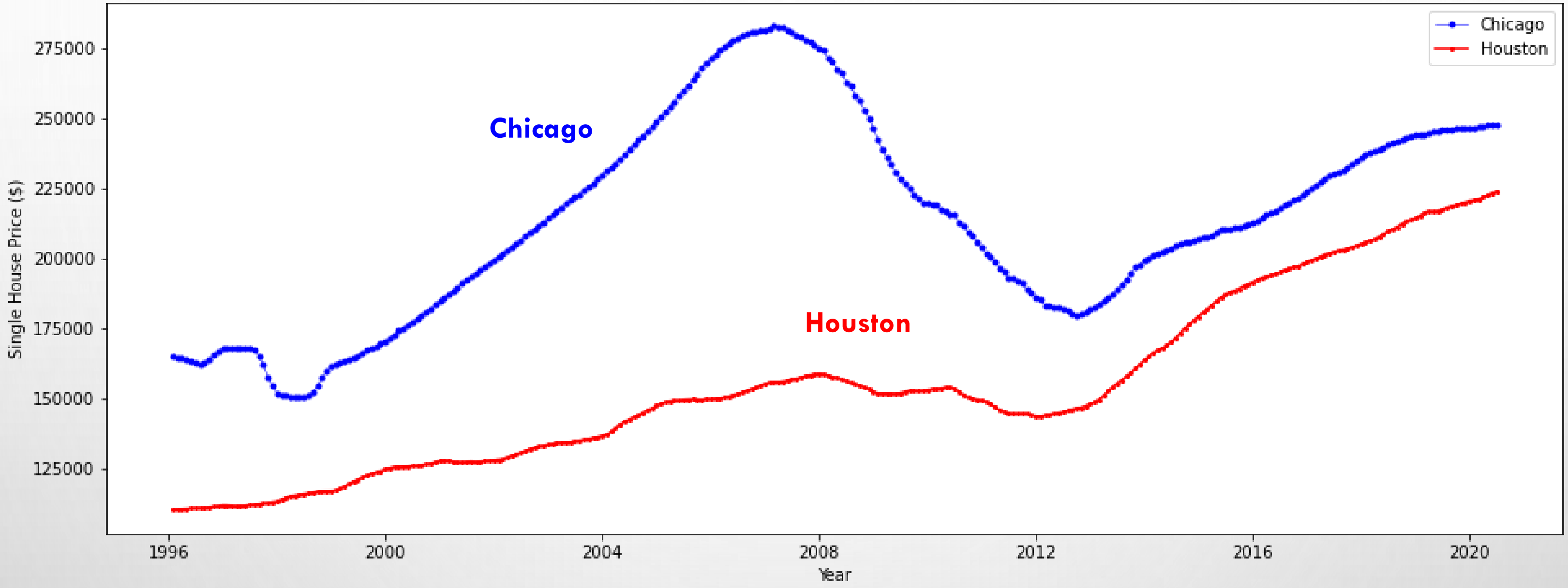
CHICAGO

Model1 – Multi-variable Linear Regression

HOUSE PRICE = -11176.31*GDP+12039.75*TOTAL_EMPLOYMENT-9225.55*POPULATION



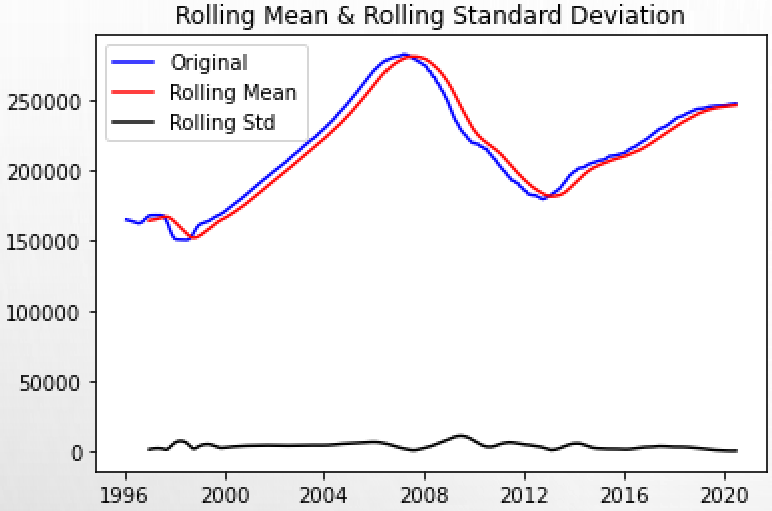
Time series analysis

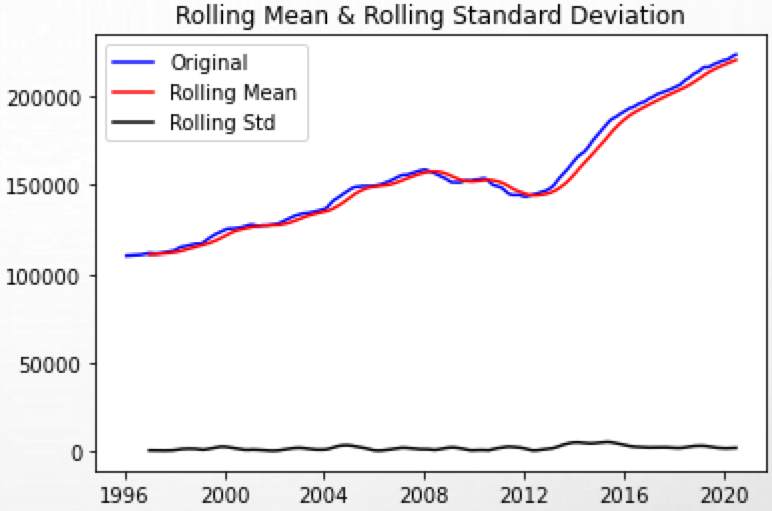


- ➢ No seasonality
- ➢ There is a low frequency trend

DATA STATIONARY OR NON-STATIONARY

Rolling Statistics





Augmented Dickey-Fuller Test

```
ADF Statistic: -2.743111756487362
p-value: 0.0668705107489899
Critical Values: 1%: -3.4541800885158525
                 5%: -2.872031361137725
                 10%: -2.5723603999791473
```

```
ADF Statistic: -0.2864468062292527
p-value: 0.927402056623277
Critical Values: 1%: -3.4541800885158525
                 5%: -2.872031361137725
                 10%: -2.5723603999791473
```

Stationary?

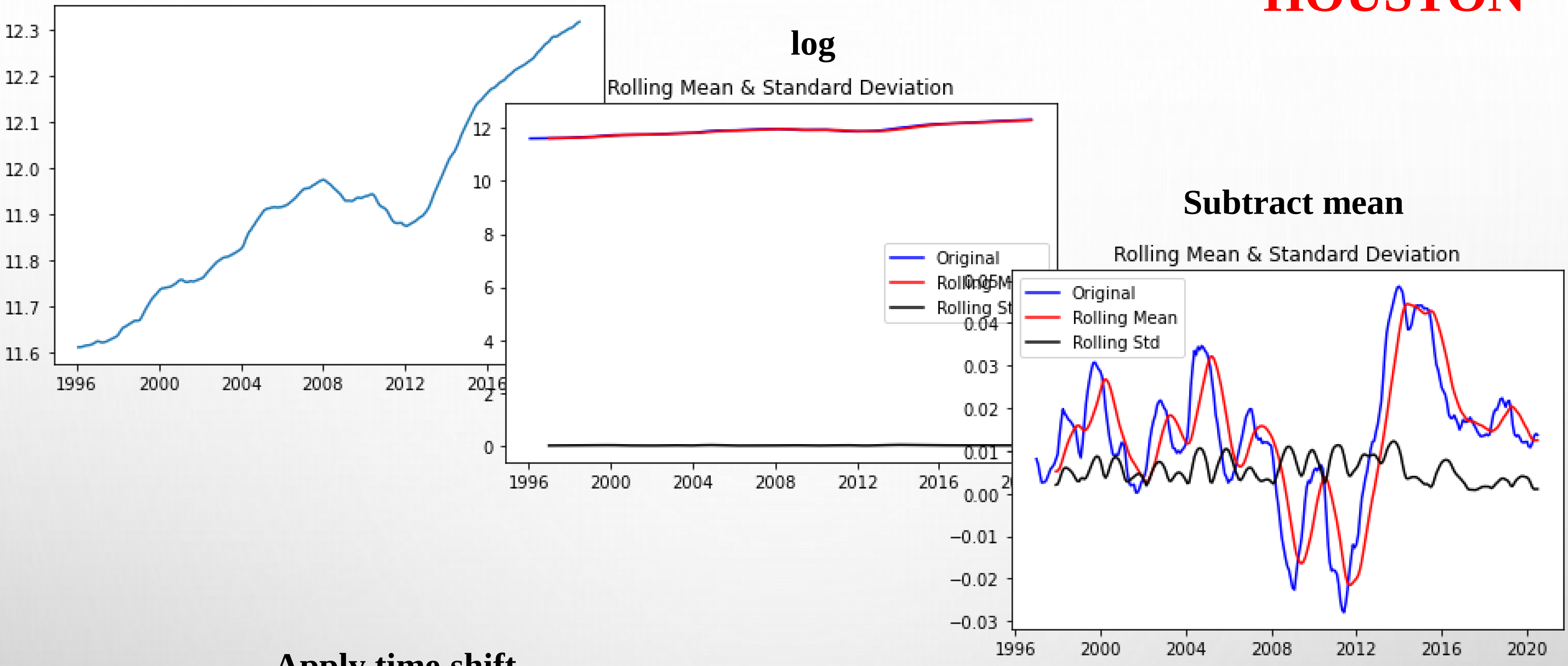
Non-Stationary?

HOUSTON

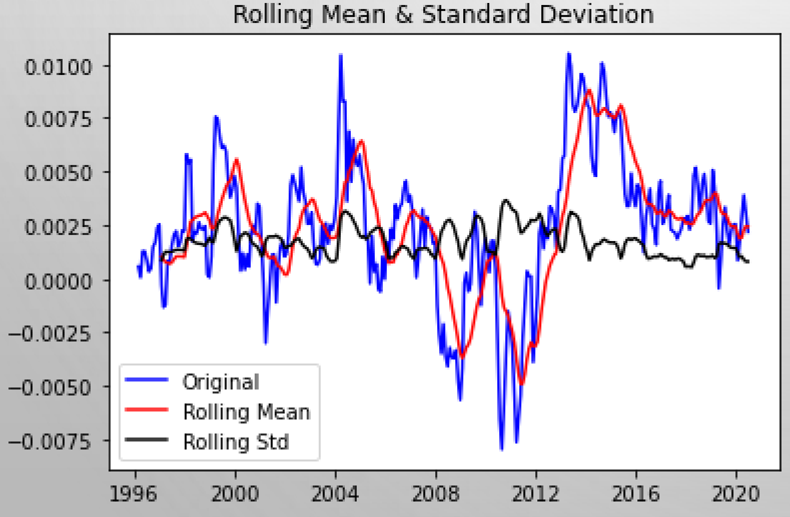
log

log

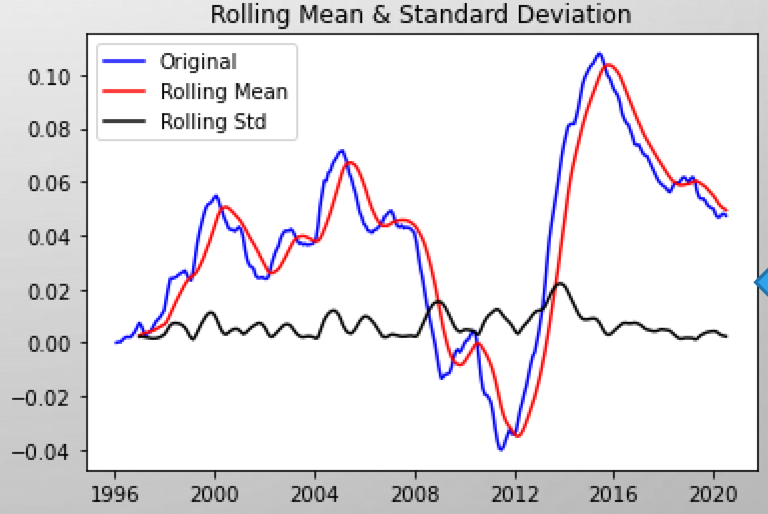
Subtract mean



Apply time shift



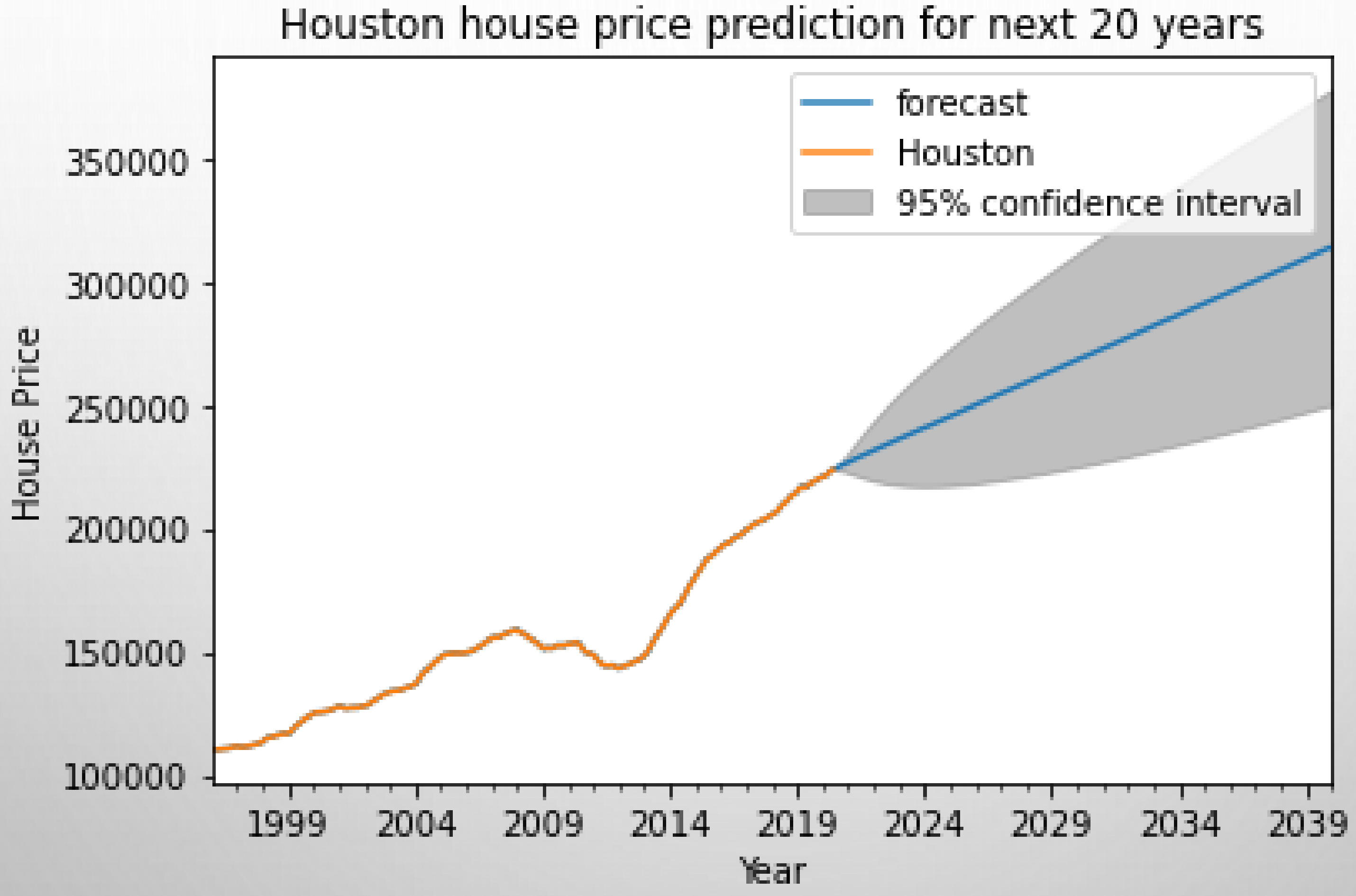
Apply exponential decay



The best

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL

HOUSTON

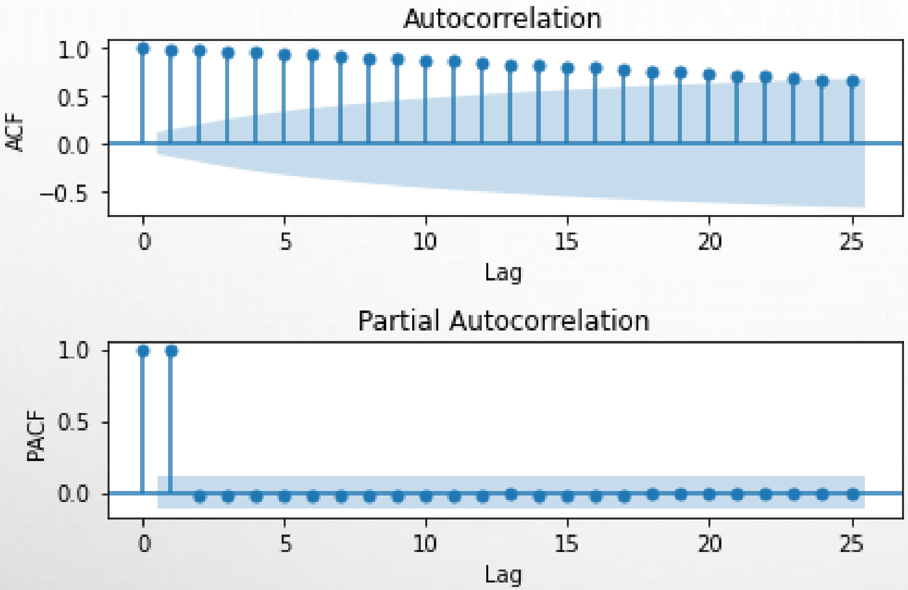


ARIMA Model (1,1,1)

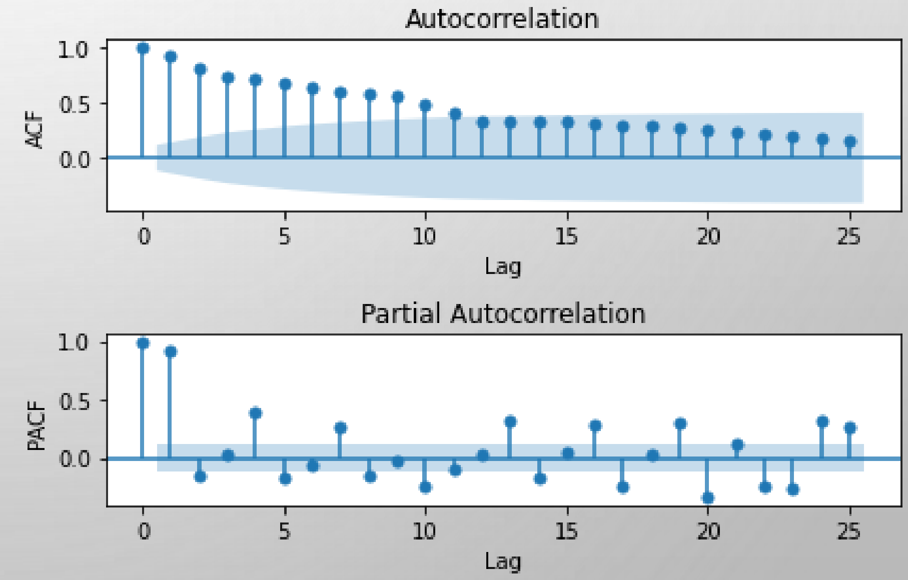
ARIMA MODEL PARAMETER SELECTION

HOUSTON

Original Data



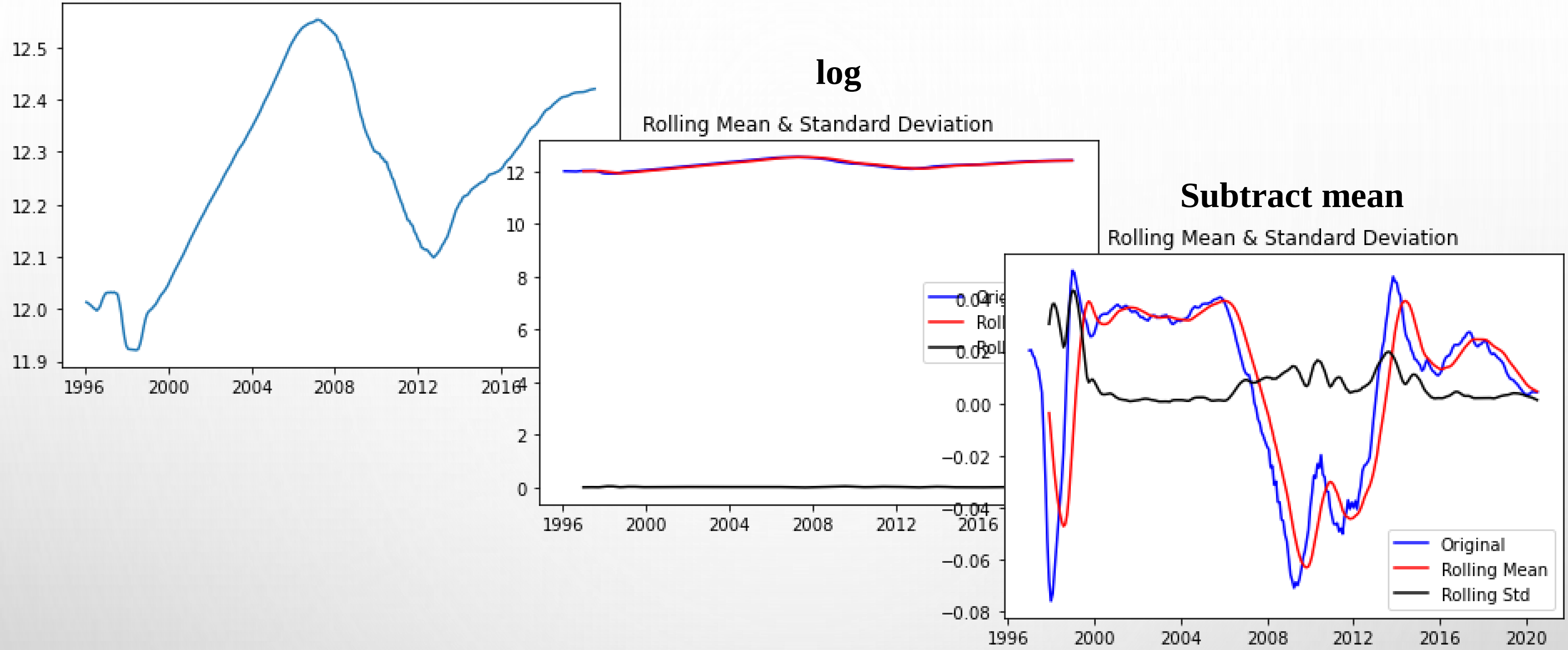
1st difference



CHICAGO

None of this better than the original data

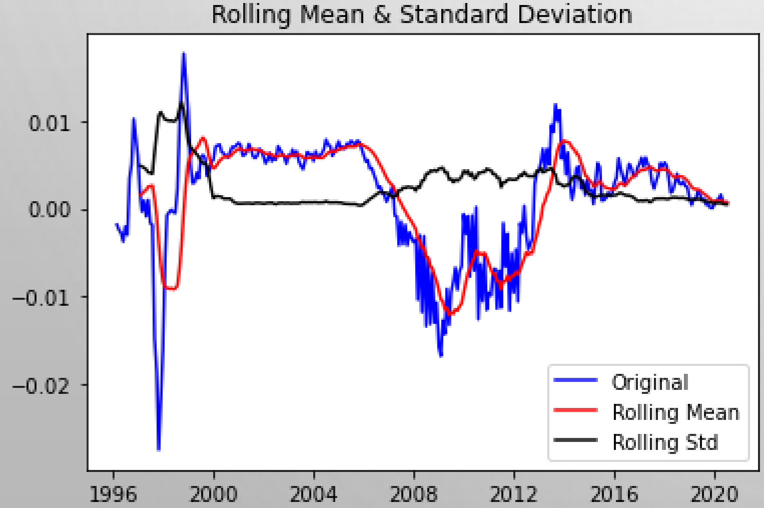
log



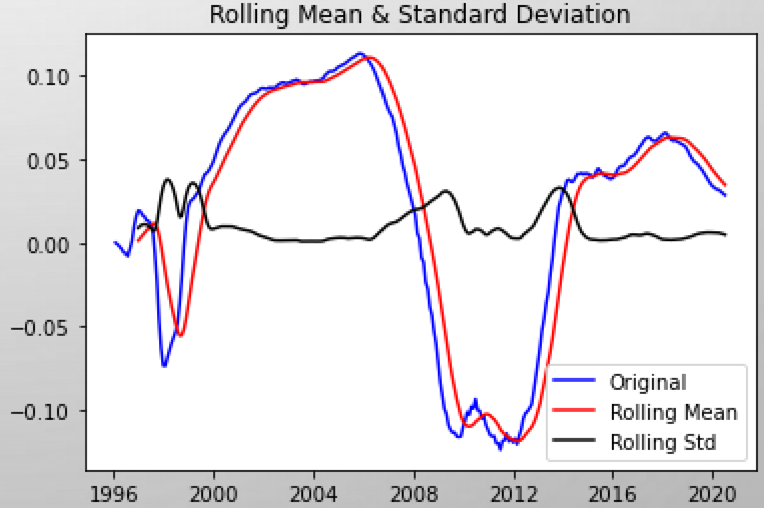
log

Subtract mean

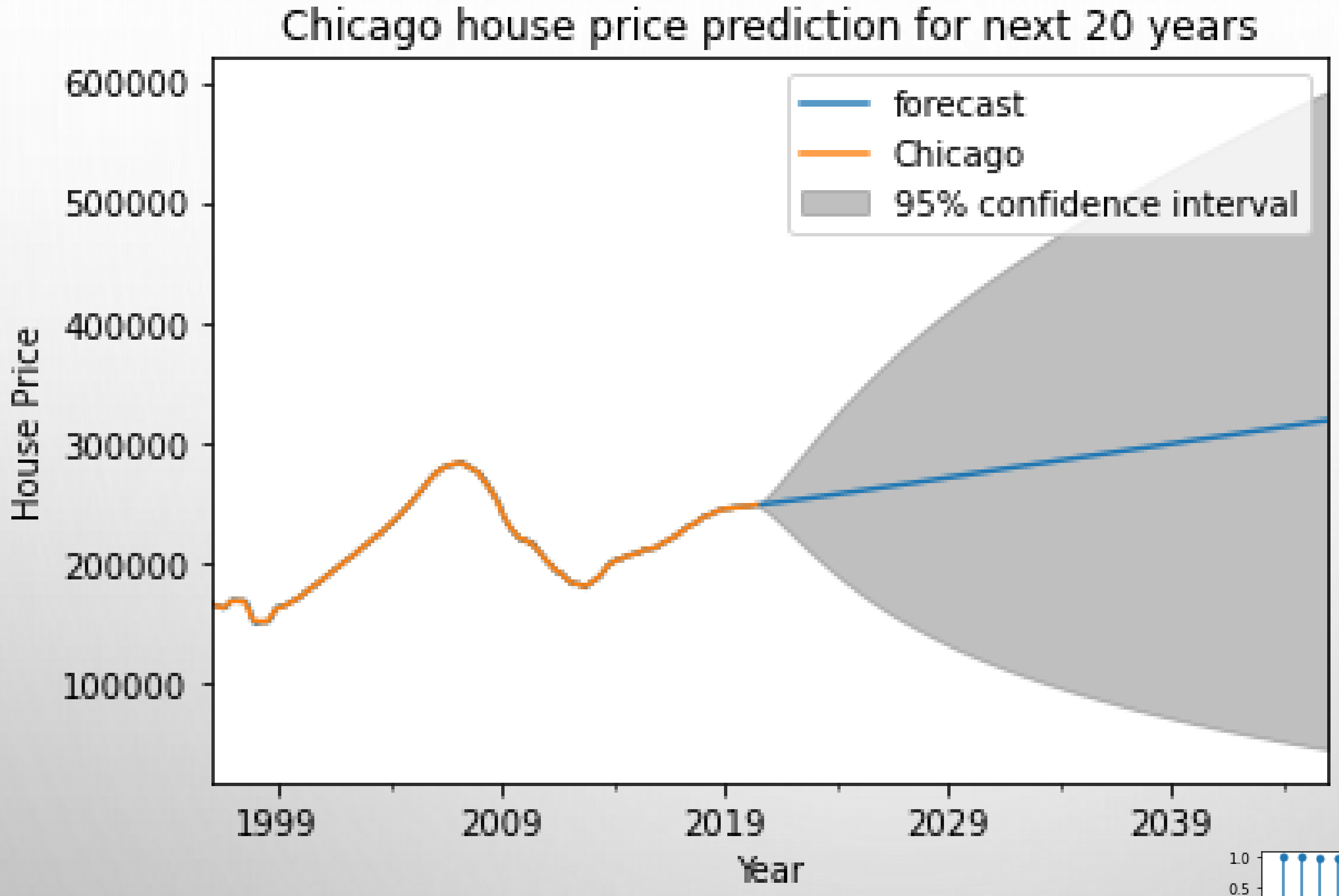
Apply time shift

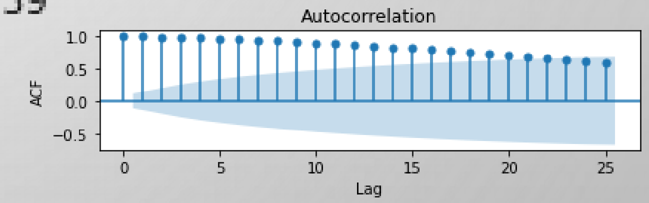


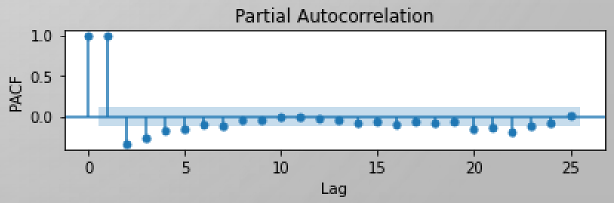
Apply exponential decay



AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL - CHICAGO







ARIMA Model (1,1,1)

CONCLUSION

- **Linear regression model was built for Chicago and Houston house price;**
- **Houston house price is highly related to GDP;**
- **Chicago house price is not linear to GDP, total employment, and population;**
- **Time series analysis show very dynamic range for Chicago house price. It could go up or go down with big error range;**
- **Time series analysis for Houston show an overall increase model.**

FUTURE WORK

- **Add more features and longer years of observations;**
- **Try more complicated machine learning algorithm.**

DATA SOURCES

House Price (single-family house)

- https://www.kaggle.com/moezabid/zillow-all-homes-data

Population

- https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-cities-and-towns.html

GDP and Employment

- https://www.bea.gov/data/economic-accounts/regional