

A person in a dark suit is holding a tablet computer. The tablet screen shows a candlestick chart with a yellow trend line. The background is a composite image featuring a city skyline (likely New York City) and a large, semi-transparent candlestick chart with green and red bars and a yellow trend line. The overall theme is financial analysis and real estate.

# Zillow real estate price analysis

# Zillow Real Estate Price Dataset

## # of price indices

~ 14 723 Region monthly price values  
from 1996-04 to 2018-04

## Hierarchy

- 7 554 cities
- 1 212 Counties
- 702 Metropolitans
- 51 State





# Key Questions

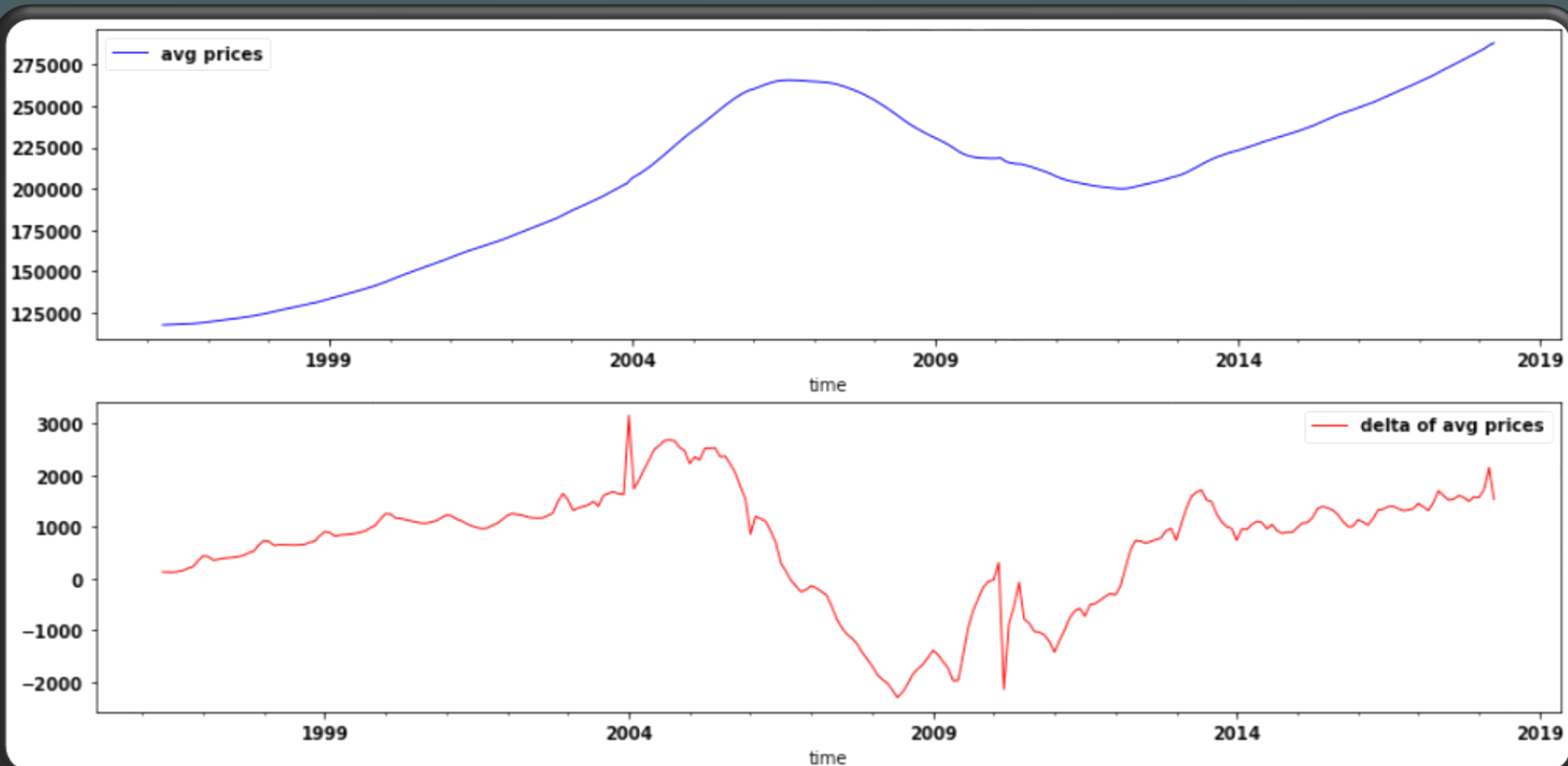
“What are the top 5 best zipcodes for us to invest in?”

## Key aspects

- Profit metrics
- Risk metrics
- Profit and Risk trade-off
- Forecast period



# Average prices

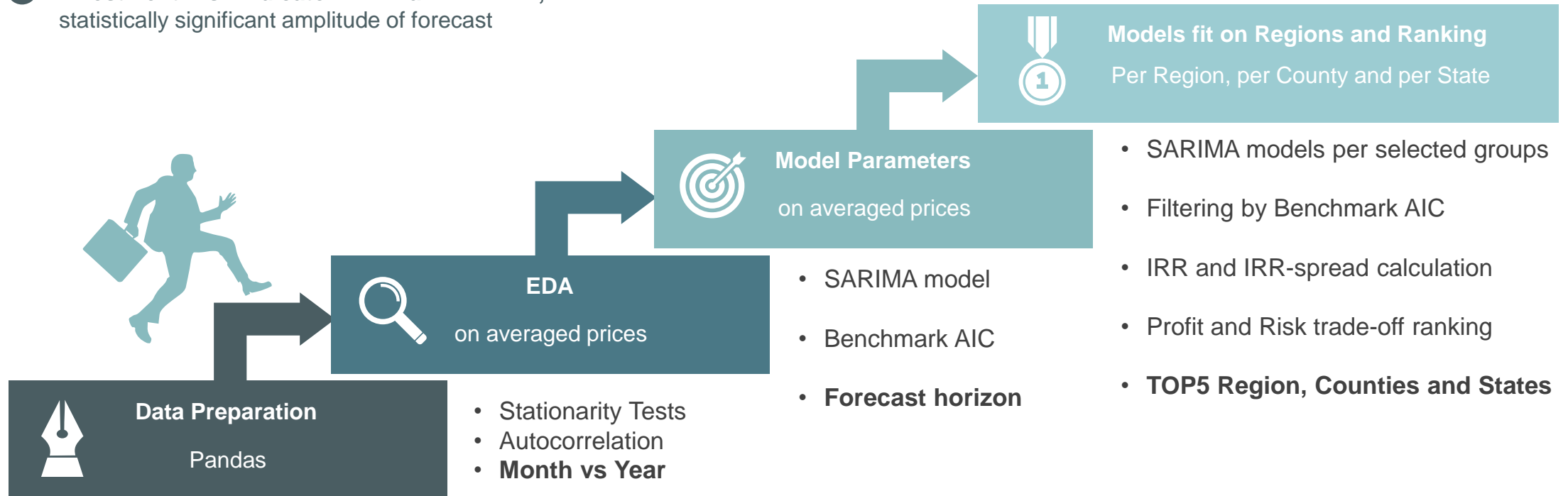


- Global cycles (15+ years)
- No significant intra-year fluctuations
- Some intra-year seasonality on price changes

# Forecast Methodology and Metrics

➤ Investment Profitability Indicator: forecasted IRR

➤ Investment Risk Indicator:  $IRR_{max} - IRR_{min}$ , statistically significant amplitude of forecast



# Profit vs Risk Trade-off

A green arrow pointing to the right, representing the direction of increasing return.

Highest Rate of Return

A red arrow pointing to the left, representing the direction of decreasing risk.

Lowest Risk



**TOP5**

INTERSECTION of 2 SORTED LISTS

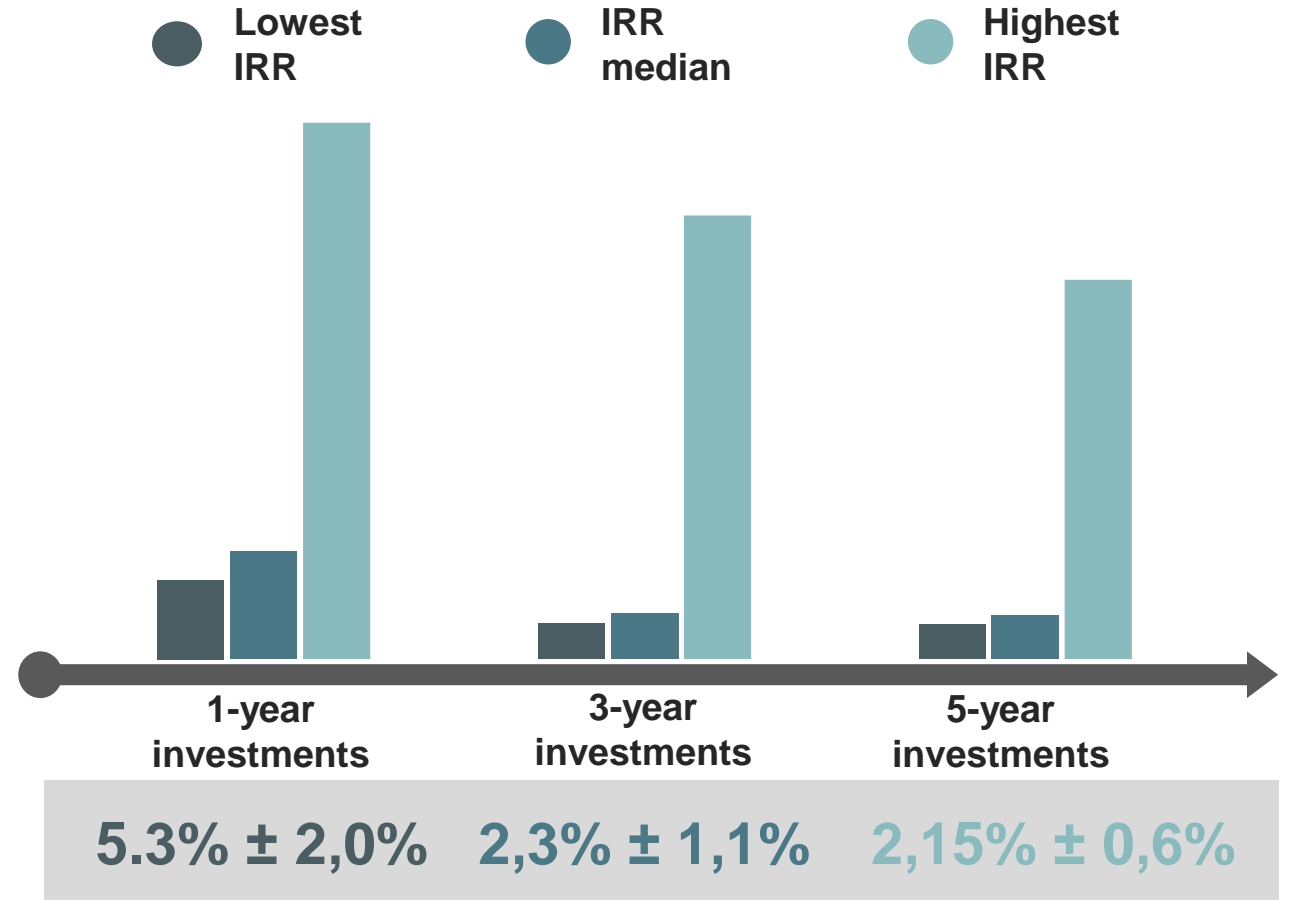
# TOP 5 Best Regions

## ZIPCODES for 1-year horizon

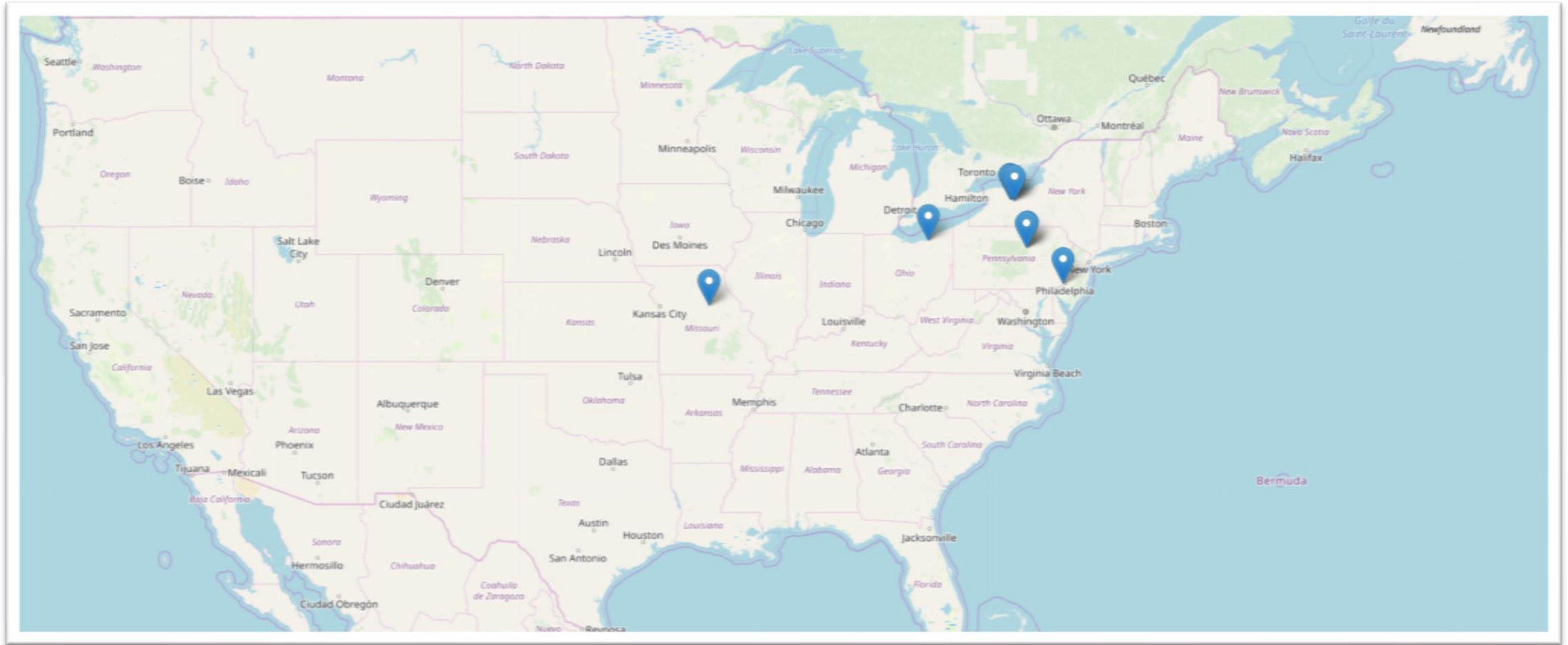
- 14414 (Avon, NY)
- 14423 (Caledonia, NY)
- 14485 (Lima, NY)
- 19023 (Darby, PA)
- 44102 (Cleveland, OH)

## ZIPCODES for 3 and 5-year horizons

- 17701 (Williamsport, PA)
- 17702 (South Williamsport, PA)
- 19023 (Darby, PA)
- 44102 (Cleveland, OH)
- 65255 (Hallsville, MO)



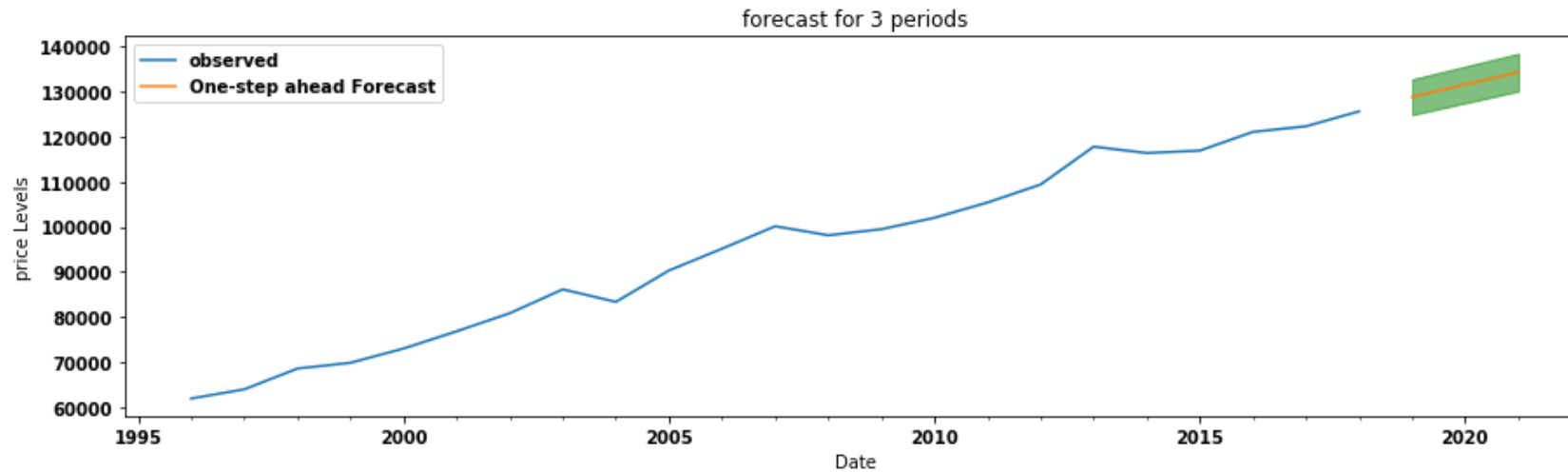
# TOP 5 Best Regions: map



**TOP ZIPCODES: 44102, 17701, 17702, 19023, 65255, 14414, 14423, 14485**

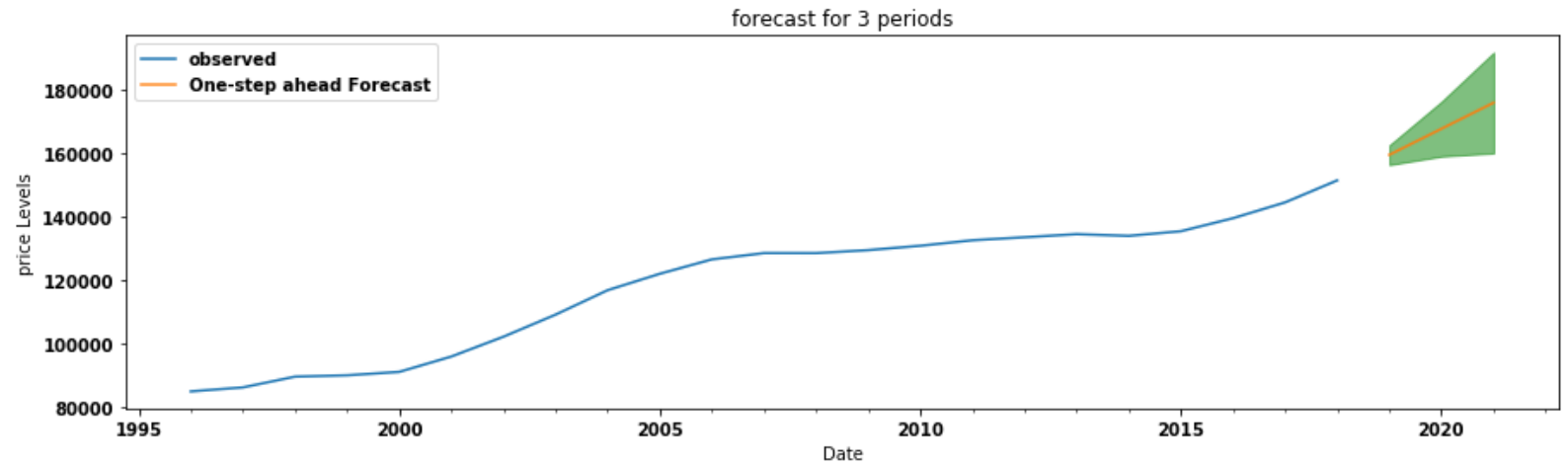


# 3-year Forecast Charts



**<= 17701 (Williamsport, PA)**

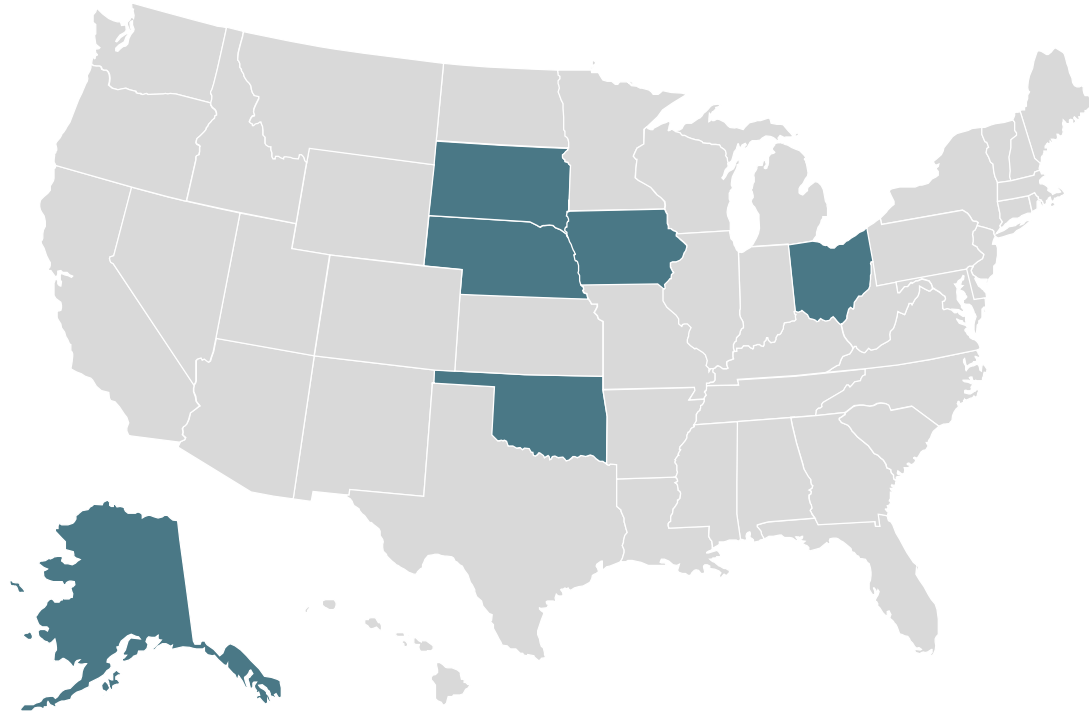
**14414 (Avon, NY) =>**



# TOP5 Best Counties and States

COUNTIES		STATES	
1-year	3-year	1-year	3-year
Allegheny	Chemung	Iowa	Iowa
Erie	Erie	Nebraska	Nebraska
Maury	Maury	Ohio	Ohio
Pontonoc	Pontonoc	Oklahoma	Alaska
West Baton Rouge	Dubuque	South Dakota	South Dakota
4.4% ± 2.4%	4.1% ± 2.3% <sub>ot</sub>	4.0% ± 2.3%	3.8% ± 3.3%

# TOP5 Best States



## TOP 5 STATES:

- Alaska
- Oklahoma
- Ohio
- Iowa
- South Dakota
- Nebraska

In total 6 states as union of 1-year and 3-year lists

# Business Recommendations

- Prices on the given dataset **could be forecasted** with **high and moderate** level of accuracy
- Recommended forecast **length: 5 years or less**
- Top5 **zipcodes** for investments:
  - for **1-year** maturity: **14414, 14423, 14485, 19023, 44102;**
  - for **3-5-year** maturity: **17701, 17702, 19023, 44102, 65255**
- Median **IRR** for TOP5 **zipcodes**: **5,3% (1-year), 2,3% (3-year), 2,1% (5-year)**
- Top5 counties for investments (3-year): **Erie, Keweenaw, Maury, Pontonoc, Dubuque**
- Top5 states for investments (3-year): **Alaska, Iowa, Nebraska, Ohio, South Dakota**



# Future Work

1. Add advanced financial KPI for profit and risk measurements and verify ranking results
2. Add macroeconomic data (GDP growth, rise of population, etc) as additional predictors to Prices
3. Increase forecast power by finding ways of optimizing calculations complexity for monthly based index





# Appendix



# Appendix A. Detailed Calculation Steps

## B. Benchmark SARIMA

1. Prepare TS with avg-prices
2. Find optimal SARIMA-params **(0,1,1)(0,1,1,1)**
3. Calculate benchmark AIC (**376**)

## D. Models Fit

1. Run all TS with 3 DF with fixed SARIMA params (from B)
2. Filter results with Benchmark AIC (see in B) and collect to DF
3. Calculate IRR and IRR-spread for each filtered row and add to DF

## E. Advanced tests

1. Calculate and compare with forecast charts from E a set of models with different optimal SARIMA params for each row
2. Run and compare Monthly Based model and forecast with results from prev.steps (annually based)



## A. EDA

1. Check stationarity (ADF test)
2. Check ACF and PACF
3. Check monthly ts vs rolling ts with 12month
4. Choose resize option (**YEAR**) and range for SARIMA params **(0,1)** – as pqr and **s=1**



## C. Prepare data for model

1. For original question: row ~ Region
2. Additional 1: row ~ County, prices – averaged
3. Additional 2: row ~ State, prices – averaged
4. Time period for (1)-(3) - **Year**

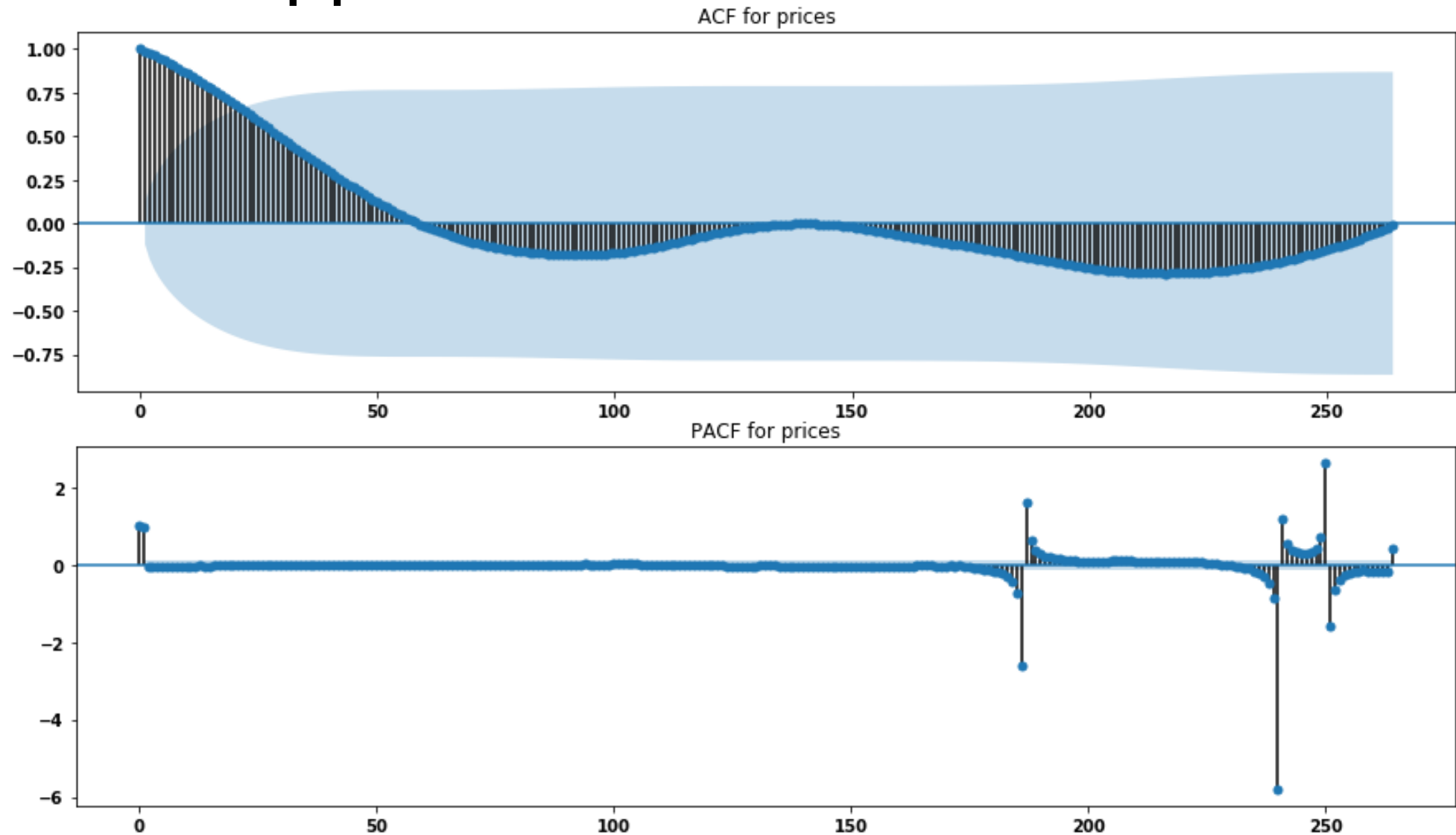


## E. Ranking

1. Rank all rows (Regions, Counties and States) in 3 DF by IRR and IRR\_spread
2. Collect top best (5) results and put to resulting DF

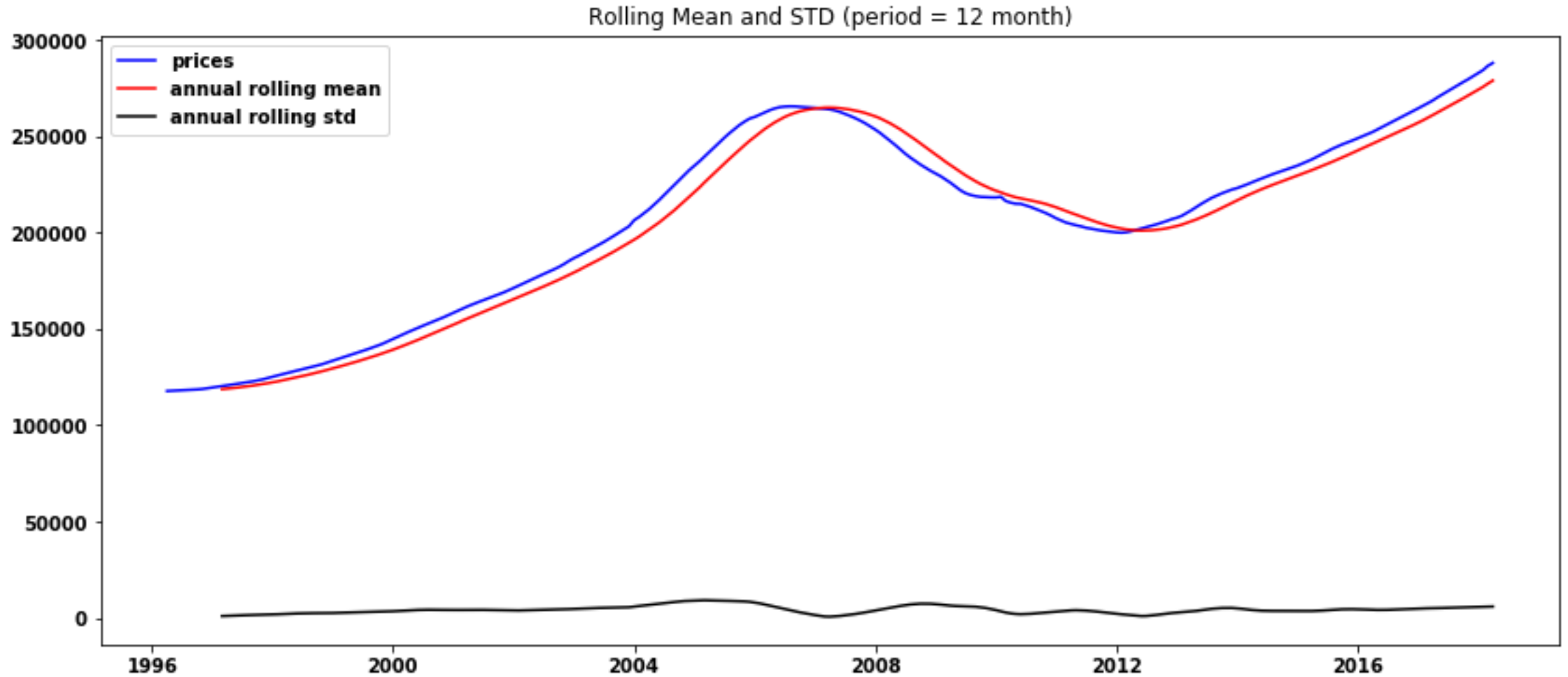


# Appendix B. ACF and PACF



The charts above calculated on average prices

# Appendix C. Month vs Year

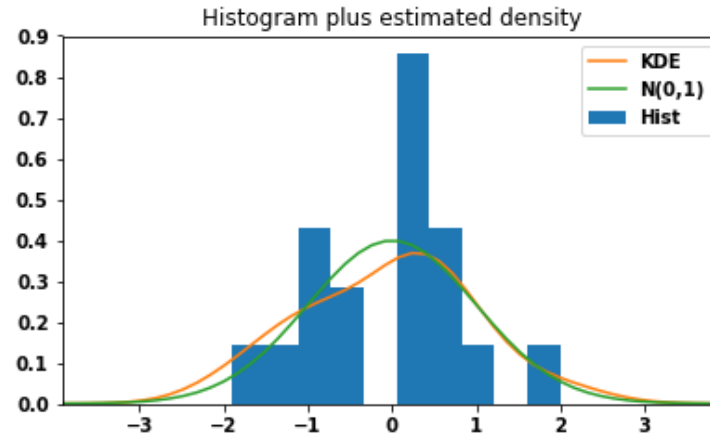


The charts above calculated on average prices

# Appendix D. Benchmark model and forecast

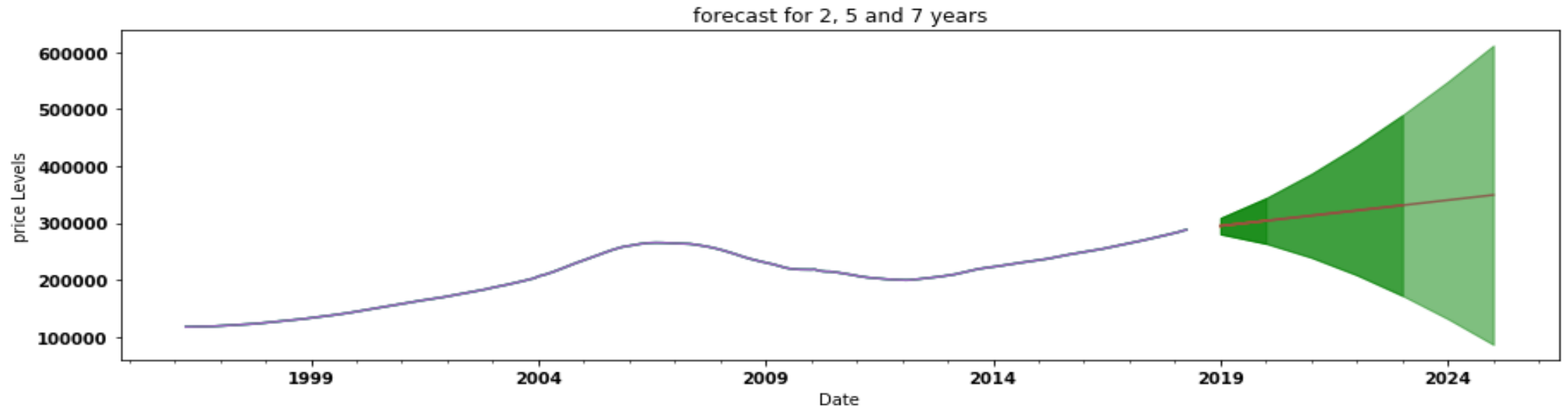
## Best Model Params:

- Data: **average price p.a.**
- pdq: **(0,1,1)**
- PDQs: **(0,1,1,1)**
- AIC: **376**



## Residual params and Validation:

- P-values: **less than 1%**
- Skew: **-0.03**
- Kurtosis: **2.56**



The chart above calculated on average prices