Module 4 Project

Real Estate Investment Opportunities

...

Recommendations from a Time Series Analysis of Zillow Prices

Matthew Onstott

Table of Contents

Business Question

What are the best ZIP Codes to invest in?

- **Primary Goal**
  - Provide quality investment recommendations for **Gygax Real Estate** of the **Top 5** Best ZIP Codes
- **Secondary Goal**
  - Forecast real estate prices using Zillow data

Business Requirements

Gygax Real Estate operations and interests

1. **Operating Region**
   - The Southwest: Nevada, Utah, Arizona, Colorado & New Mexico
2. **Market Performance**
   - Refusal to invest in metro areas with *recent* low performance
3. **Recession Recovery**
   - Wary of ZIP Codes that struggled *following* the Great Recession
4. **Recession Volatility**
   - Wary of ZIP Codes that struggled *during* the Great Recession
5. **Final Selection**
   - Final ZIP Code selection is at the discretion of the investigating team

Operationalized Measures

State
Within Set

‘16 - ‘18 Growth
Top 10 Metros

Growth since JUN ‘09
Best 25%

Volatility from DEC ‘12 - JUN ‘09
Best 25%

Model Performance
Prediction RMSE

Right Location

Dataset

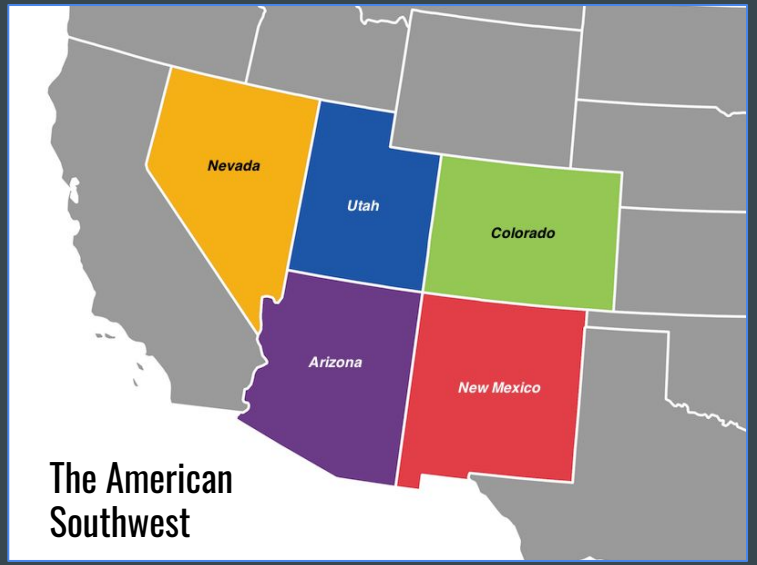
- ZIP Codes **14,723**
- Location Indicators
  - ZIP Code, City, Metro, County, State
- Timeline
  - APR ‘96 - APR ‘18 (22 years)
- Unique Values
  - City **7,554**
  - State **51**
  - Metro **701**
  - County **1,212**

Framework OSEMN

Model SARIMA

1st Filter - Region

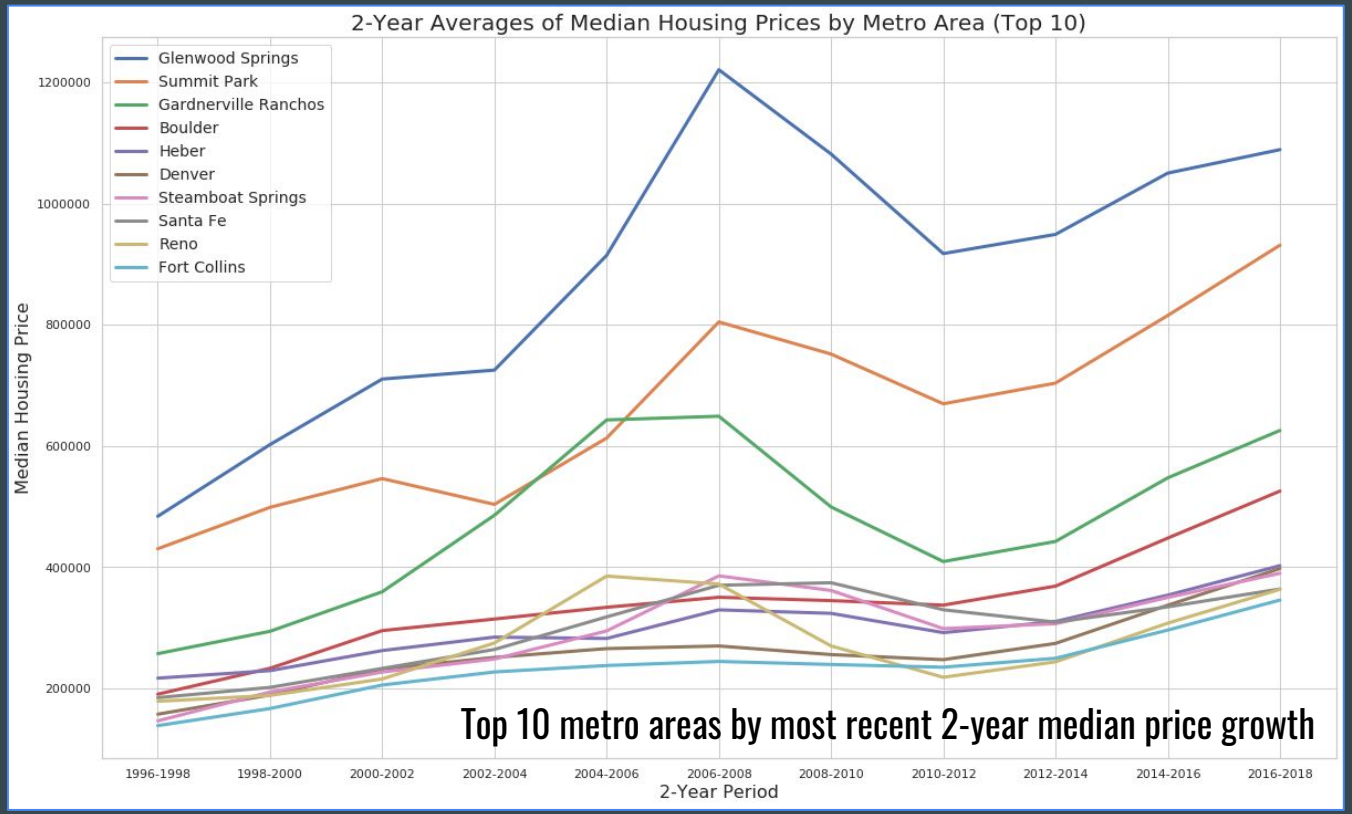
- ZIP Codes **763** (⬇94.82%)
  - CO **249**
  - AZ **230**
  - UT **121**
  - NV **103**
  - NM **60**



Recent Performance

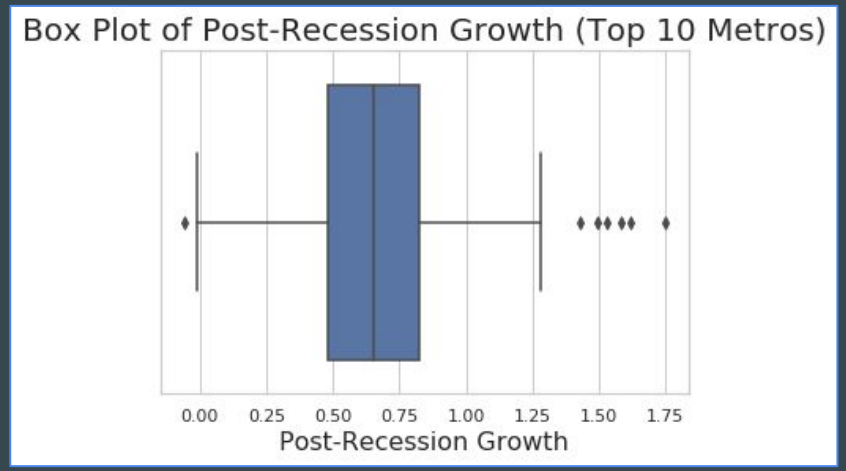
2nd Filter - Top 10 Metros

- ZIP Codes **188** (⬇98.72%)
  - CO **149**
  - AZ **0**
  - UT 4
  - NV **27**
  - NM **8**



Recession Proof

Percentage change in ZIP Code median housing price since the Great Recession



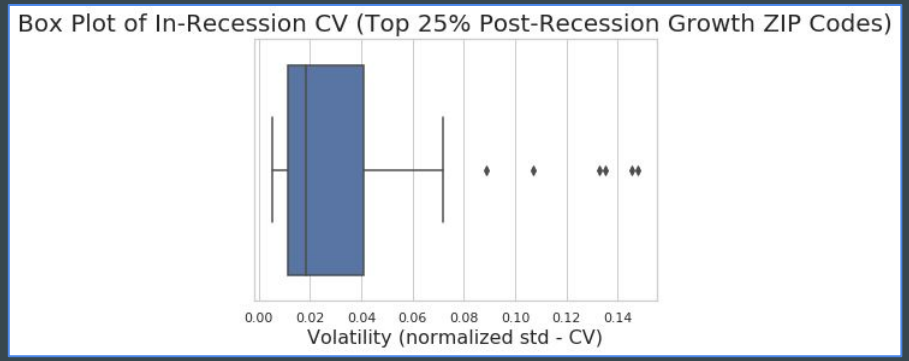
3rd Filter - Top 25% Growth

- ZIP Codes **47** (⬇99.68%)
  - CO **43**
  - NV **4**

4th Filter - Lowest 25% CV

- ZIP Codes **12** (⬇99.92%)
  - CO **12**
    - Denver **10**
    - Boulder **1**
    - Fort Collins **1**

Coefficient of variation for median housing price by ZIP Code during the Great Recession

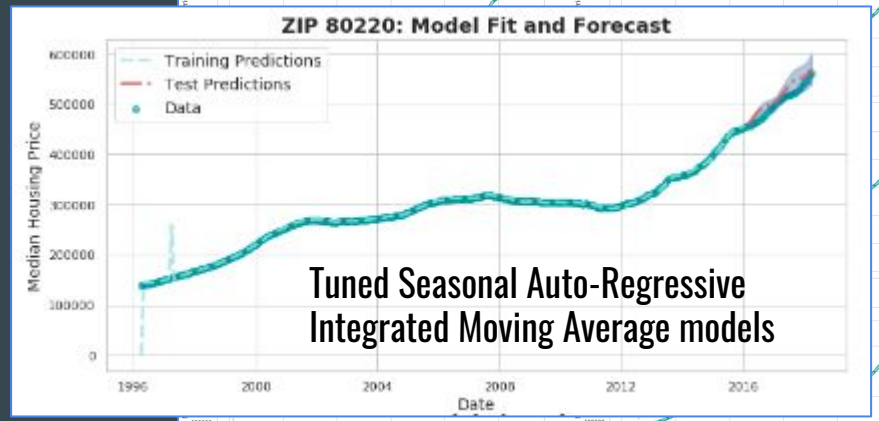


Robust Models

SARIMA models trained on 90% of the data fit observations well

5th Filter - Prediction Error
The deviation of results from testing set (10%) housing prices is used to identify the Top 5 ZIP Codes

The top 5 ZIP Codes are **80002**, **80521**, **80222**, **80121** & **80033**

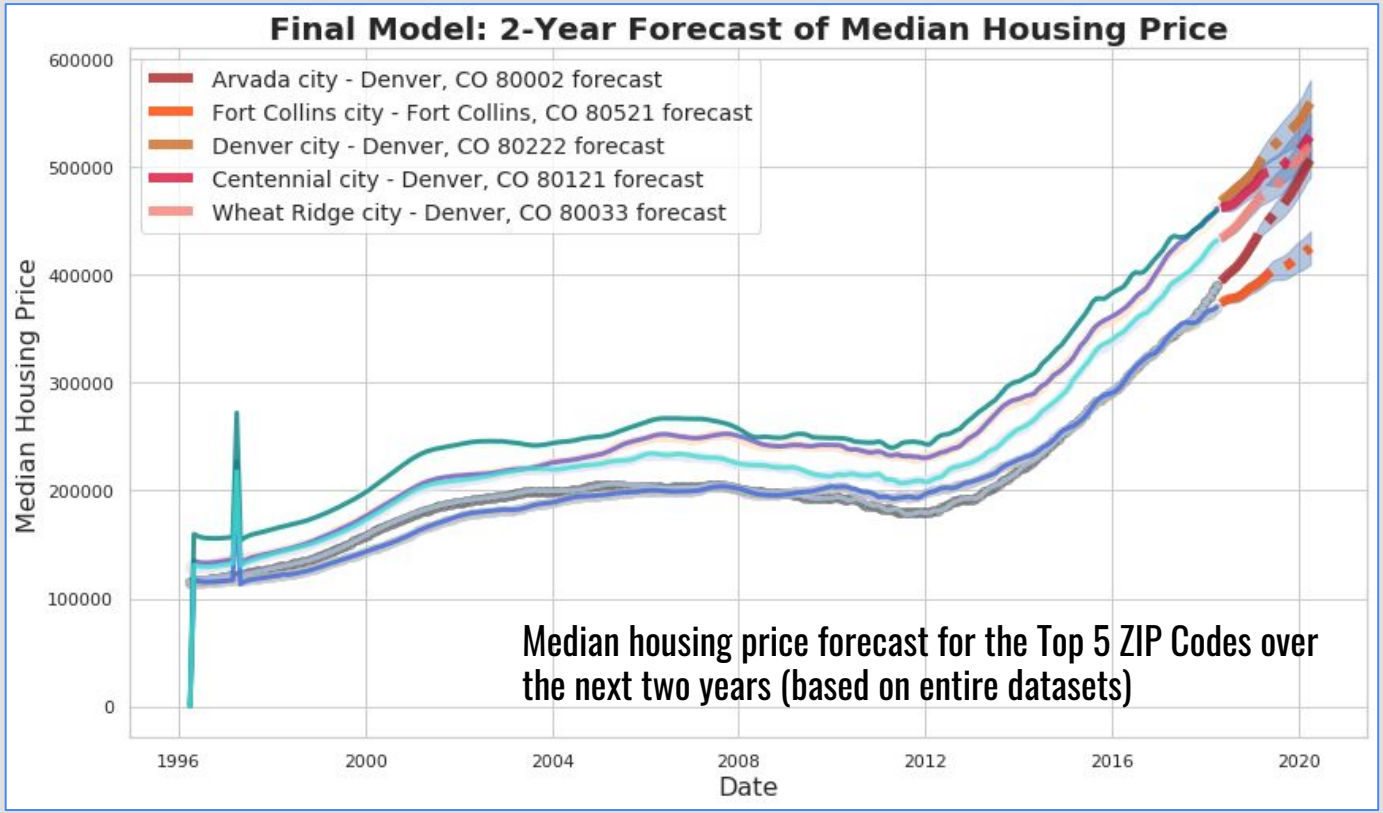


Recommendations

Average predicted price over the next 2 years:

1. **Denver City (80222)**
   - $510,217 ± $8,414.60
2. **Centennial City (80121)**
   - $491,279 ± $8,280.73
3. **Wheat Ridge City (80033)**
   - $474,621 ± $11,287.80
4. **Arvada City (80002)**
   - $445,946 ± $12,041.40
5. **Fort Collins City (80521)**
   - $397,523 ± $9.454.27

Low risk, high growth options



Median housing price forecast for the Top 5 ZIP Codes over the next two years (based on entire datasets)

Future Work

- Considerations
  - Investigate other time frequencies
  - Increase hyperparameters grid search
  - Incorporate additional data sources
    - Income
    - Taxes
    - Demographics
    - Environment
    - Education
    - Quality of Life
    - Cost of Living

Appendix

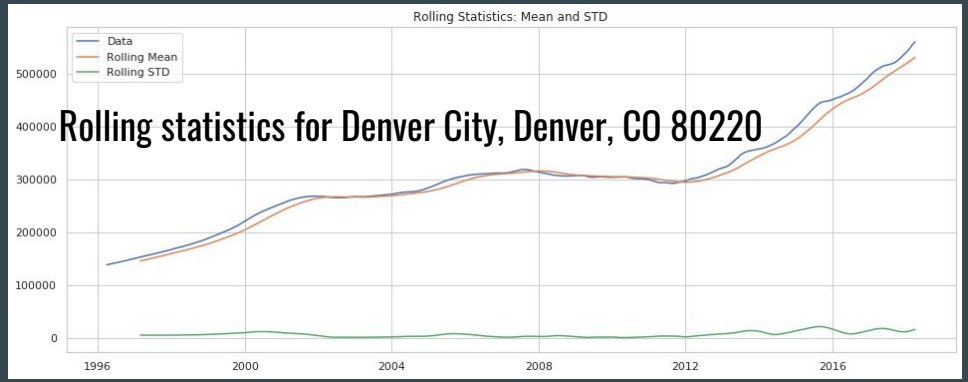
Framework Approach

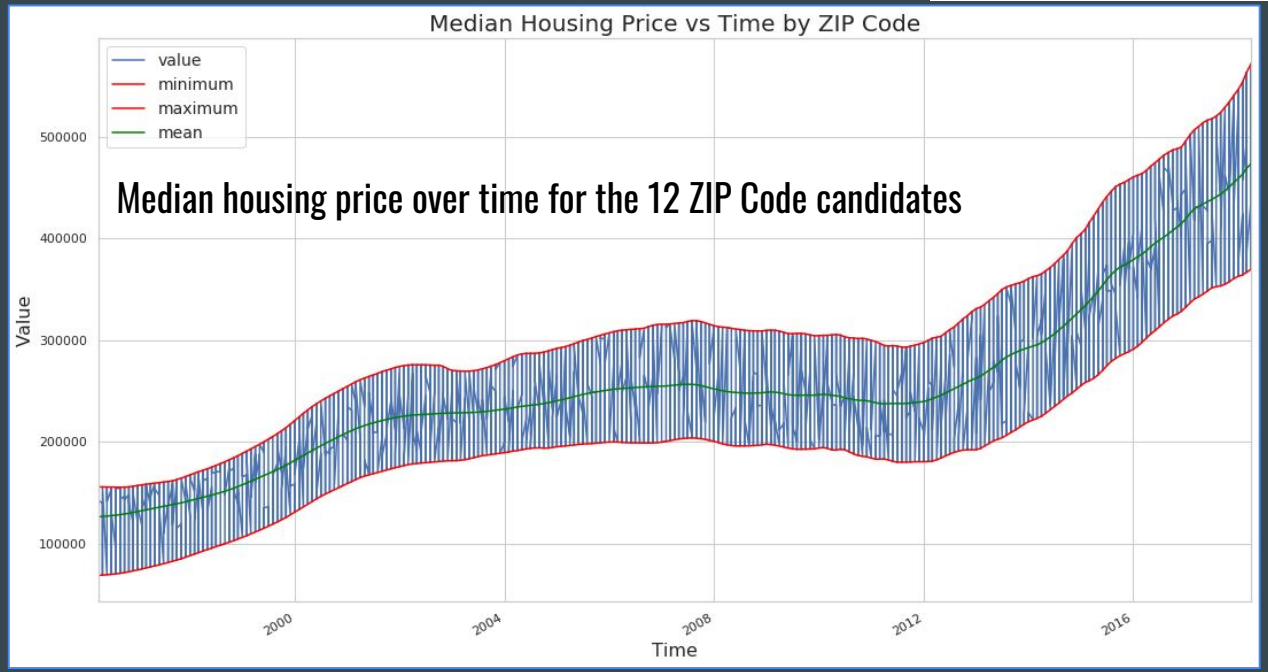
CRoss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining (CRISP-DM)

- **Business Issue Understanding**
  - Review business requirements
- **Data Understanding**
  - Import libraries, load files, inspect contents, create features, filter dataset
- **Data Preparation**
  - Drop unnecessary features, handle missing values, melt dataset, check trends, test stationarity, assess differencing & correlations
- **Modeling**
  - Tune to training sets, predict with testing sets, compare performance, select Top 5
- **Validation**
  - Review summaries, seasonal decomposition, plot diagnostics, validate coefficients
- **Recommendation**
  - Apply full data, forecast prices, plot observations with model fit & predictions

Reviewing Trends

The time series are not stationary

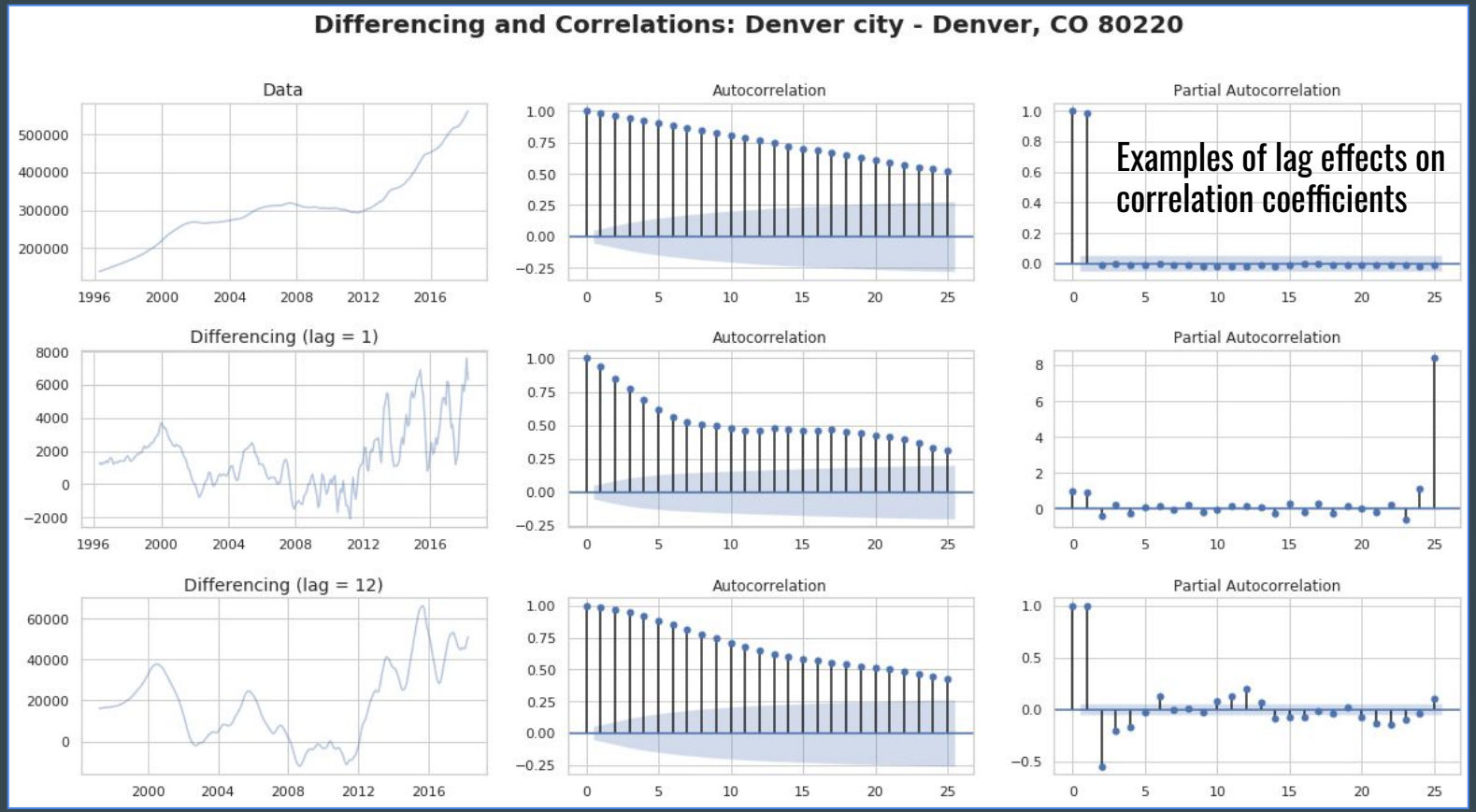




- Observations **3,180**
  - ZIP Codes **12**
  - M/Y Dates **265**
- Missing Values **0**

Time Dependency

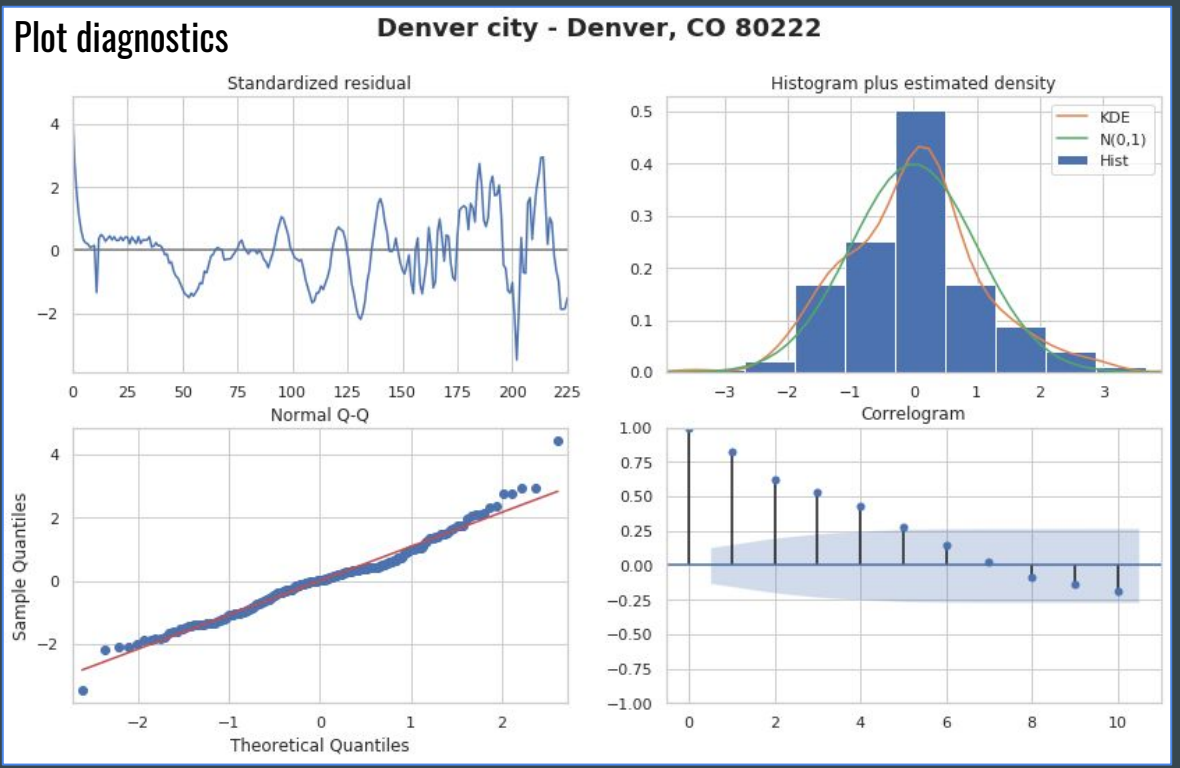
Observations are dependent on prior points in time

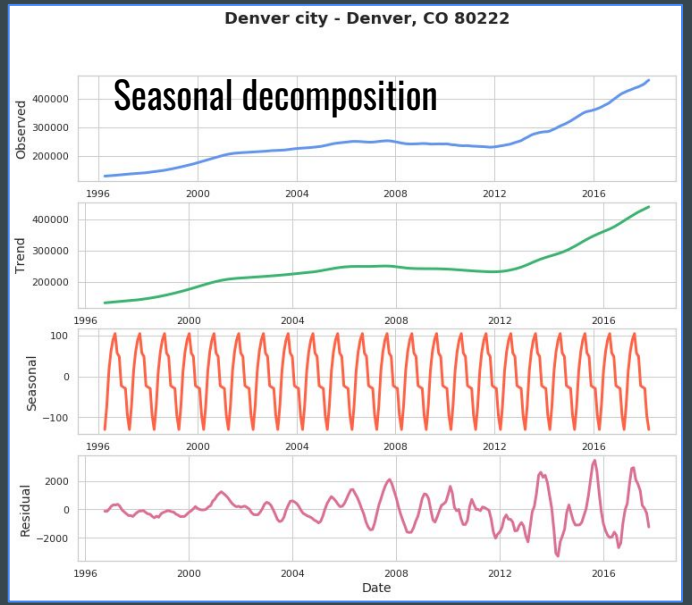


Data Inspection

There is a long-term trend increasing prices & a short-term fluctuation with a frequency of once per year

Plot diagnostics





Prediction variation is not consistent over time; this is at least partly due to the inherent non-normality in housing price data