Time Series Analysis of Zillow Real Estate Prices

•••

Module 4 Project

Matthew Onstott

Table of Contents

Problem

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 Goals**
  - Test Time Series Analysis skills
  - Complete a common real-world Time-Series Modeling task
  - Forecast real estate prices using Zillow data

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

Business Issue Understanding

Gygax Real Estate operations, interests & requirements

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

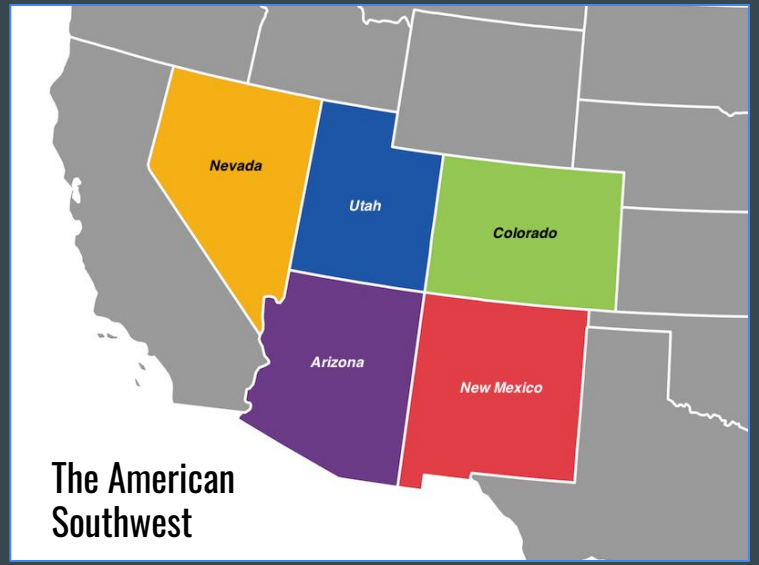
Data Understanding

Dataset

- ZIP Codes **14,723**
- Location Columns **5**
  - ZIP Code, City, Metro, County, State
- Identifier Columns **2**
  - Region ID, Size Rank
- Month-Year Columns **265**
  - APR ‘96 - APR ‘18
- Unique Values
  - City **7,554**
  - State **51**
  - Metro **701**
  - County **1,212**
- Missing Values
  - Columns with Any **220**
  - Columns with None **52**

1st Filter - Location

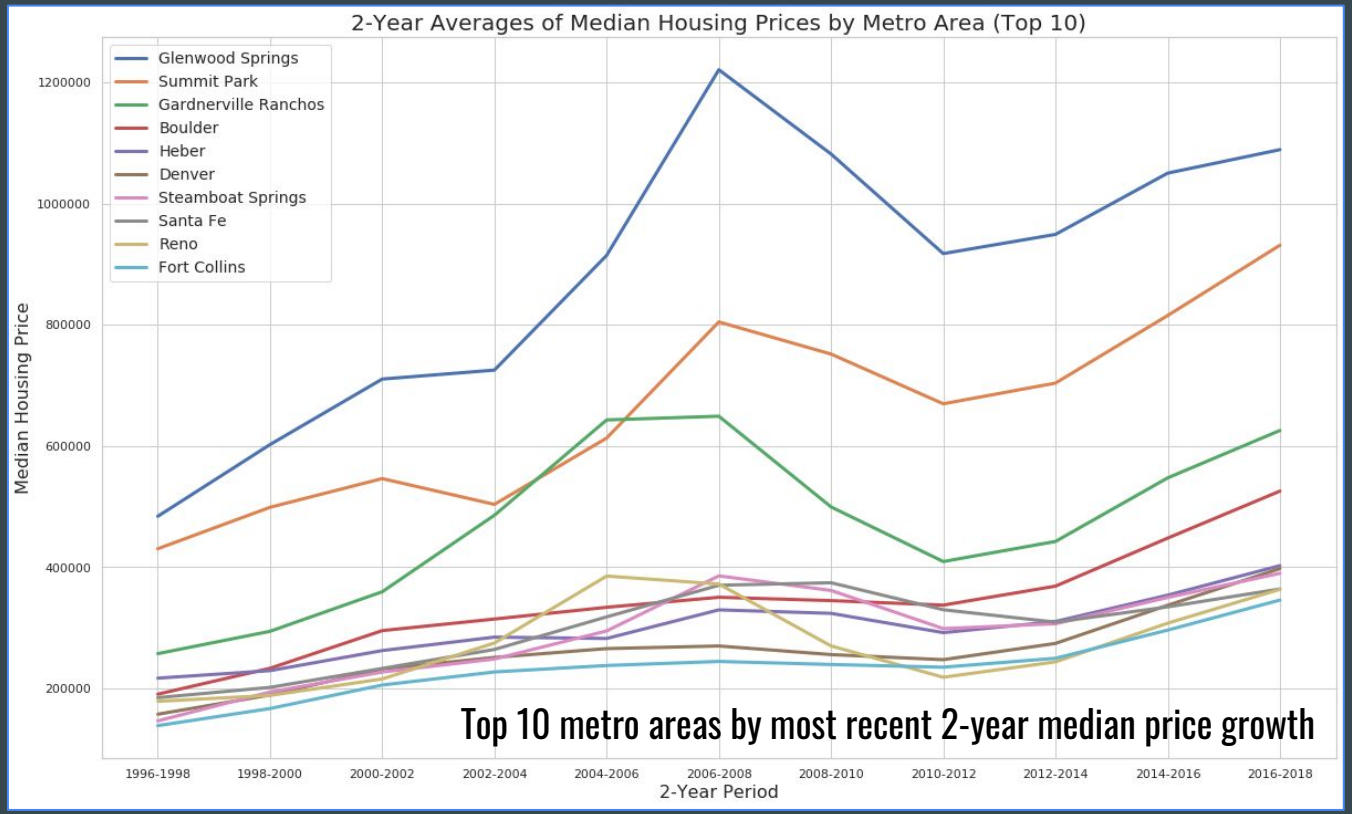
- Operations Region **763** (⬇94.82%)
  - CO **249**
  - AZ **230**
  - UT **121**
  - NV **103**
  - NM **60**



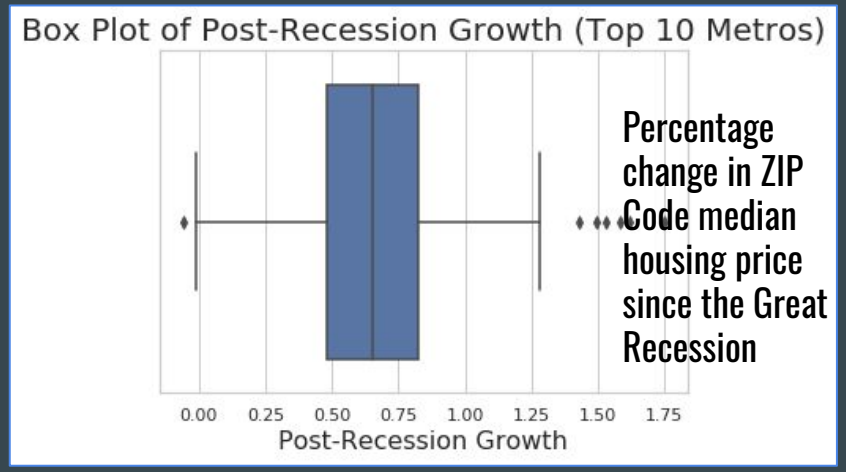
Data Understanding (2)

2nd Filter - Top 10 Metros

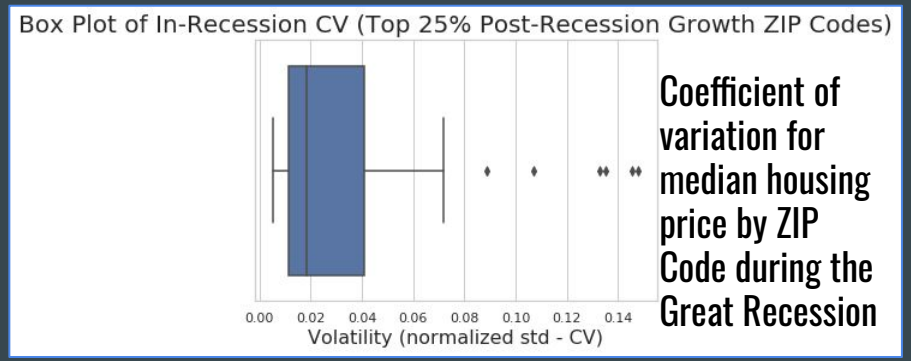
- Market Performance **188** (⬇98.72%)
  - CO **149**
  - AZ **0**
  - UT 4
  - NV **27**
  - NM **8**



Data Understanding (3)



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



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

3rd Filter - Top 25%

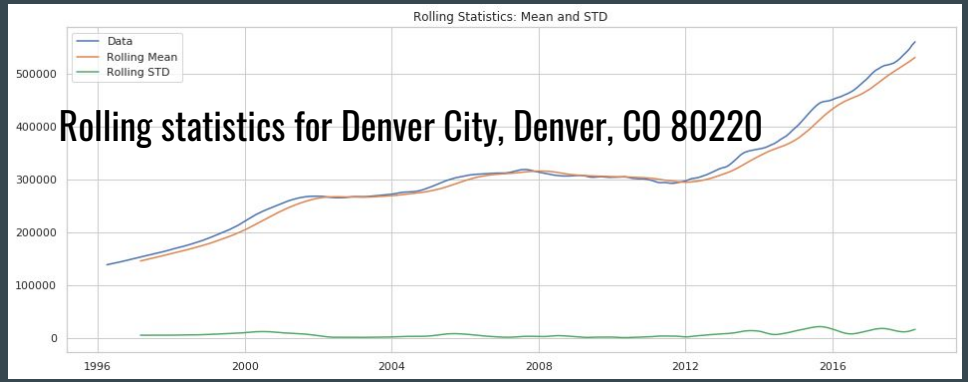
- Recession Recovery **47** (⬇99.68%)
  - CO **43**
  - NV **4**

4th Filter - Lowest 25%

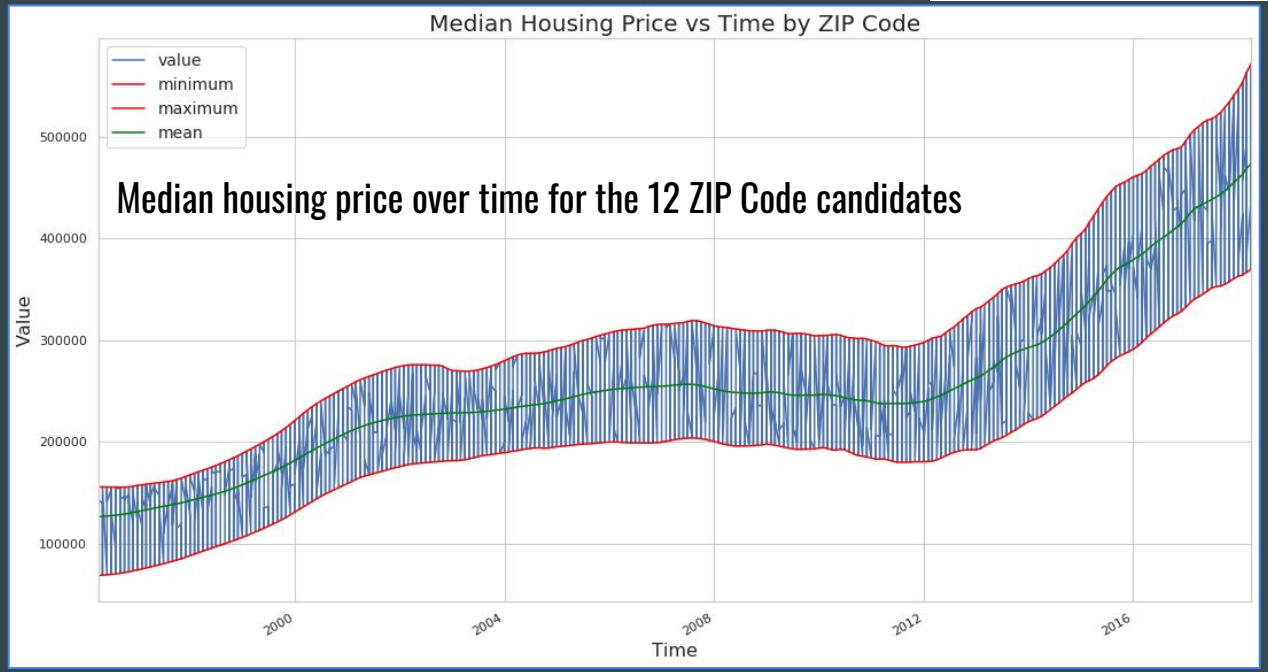
- Recession Volatility **12** (⬇99.92%)
  - CO **12**
    - Denver **10**
    - Boulder **1**
    - Fort Collins **1**

Data Preparation

The time series are not stationary



Rolling statistics for Denver City, Denver, CO 80220

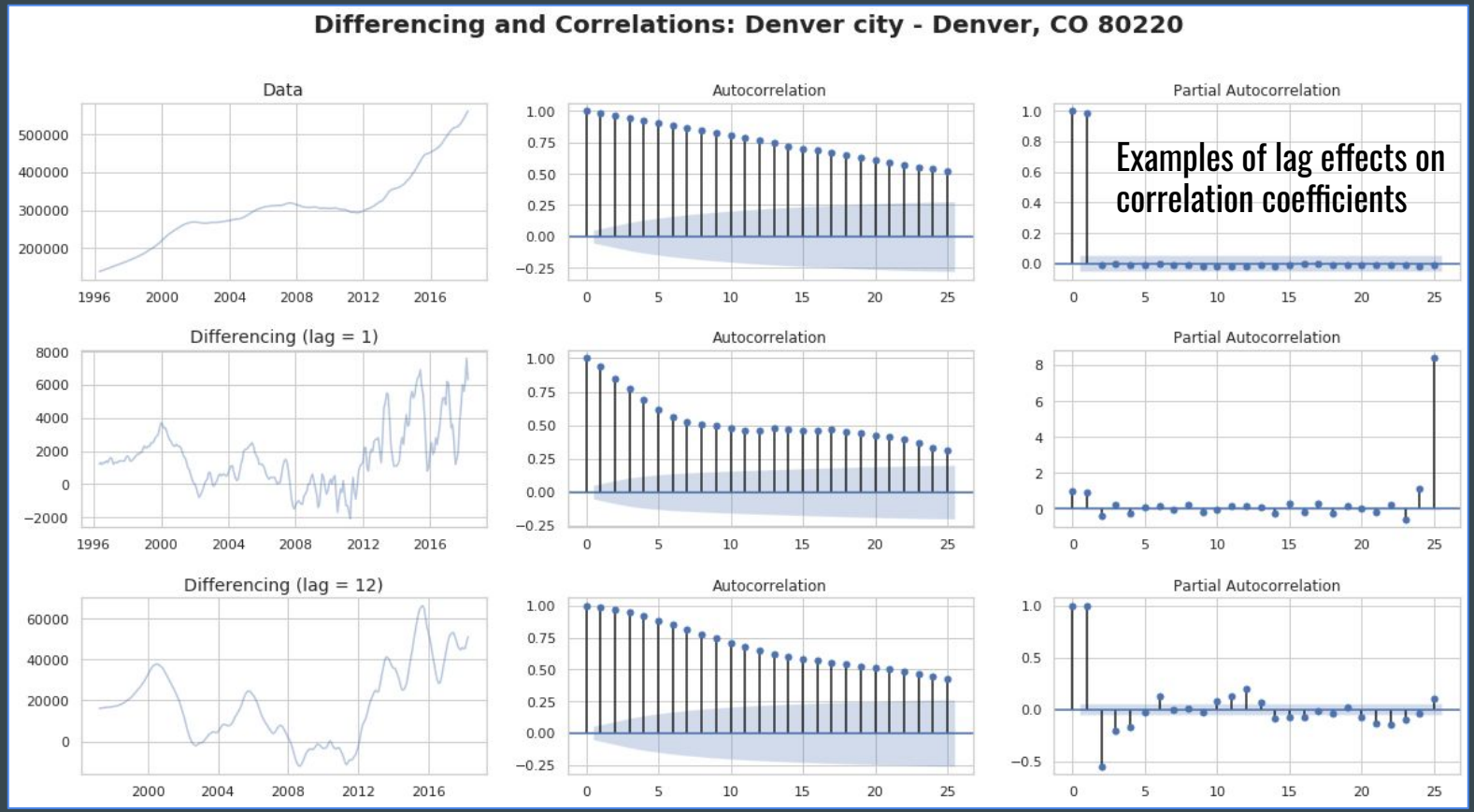


Median housing price over time for the 12 ZIP Code candidates

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

Data Preparation (2)

Observations are dependent on prior points in time

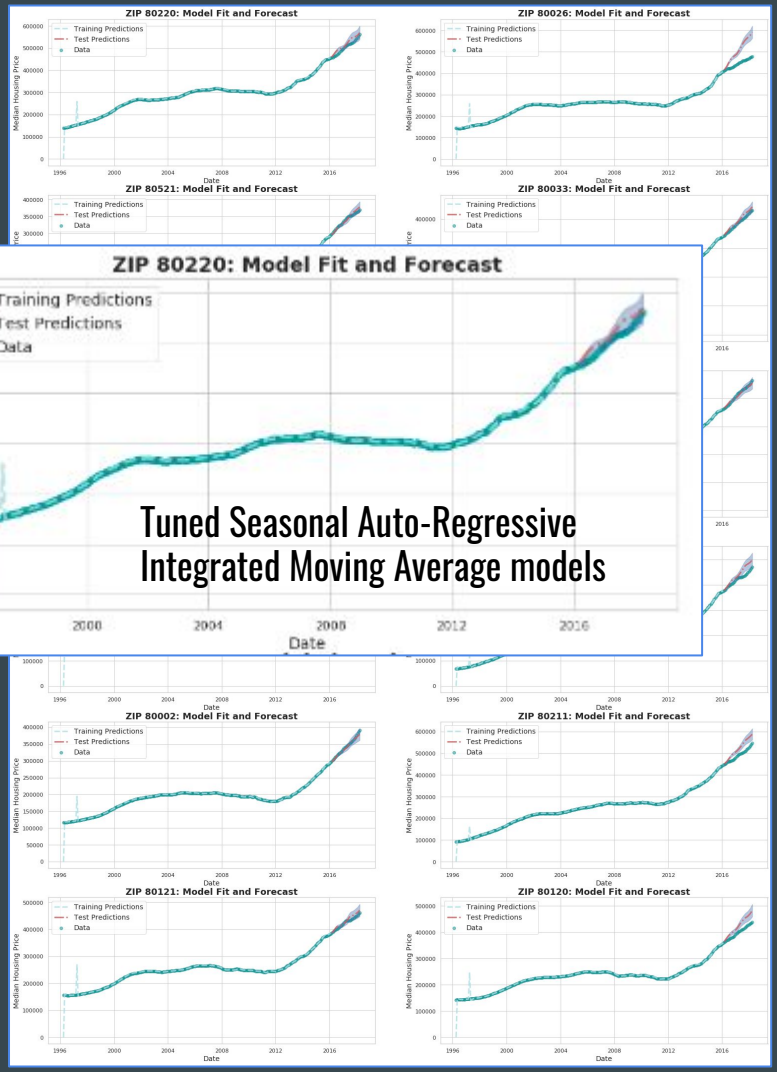


Modeling

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

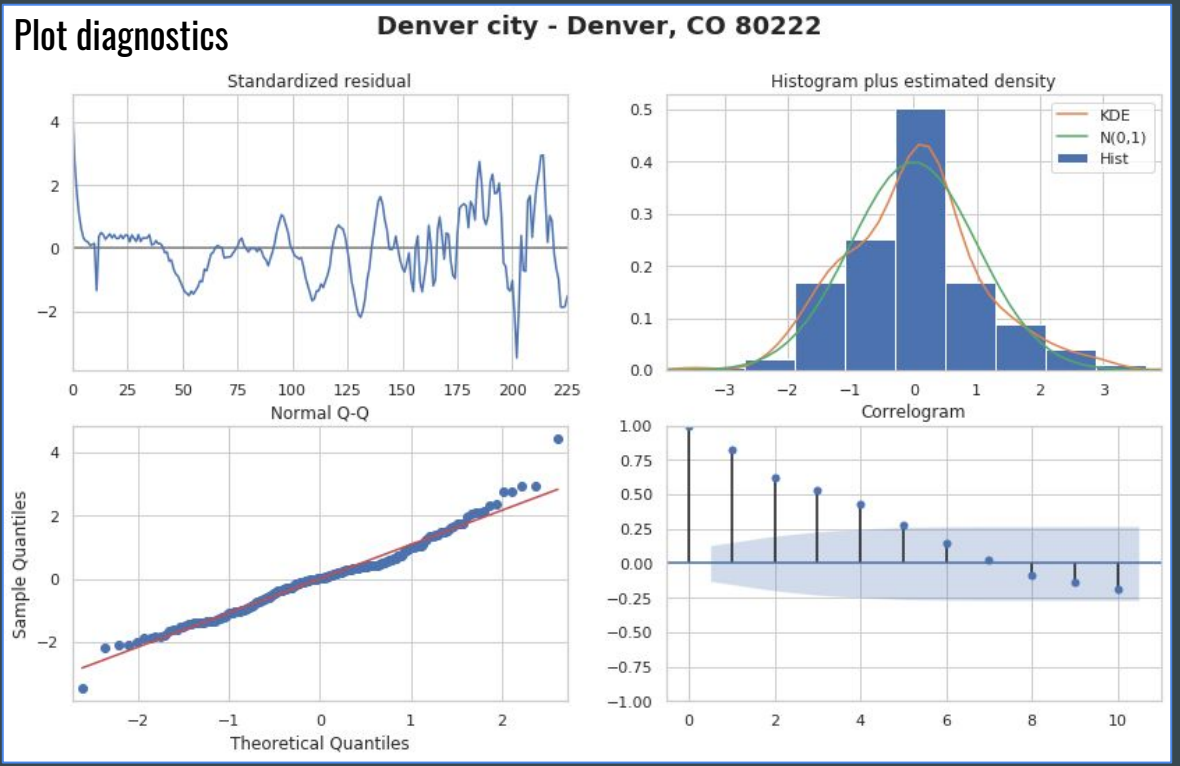
The top 5 ZIP Codes are **80002**, **80521**, **80222**, **80121** & **80033**

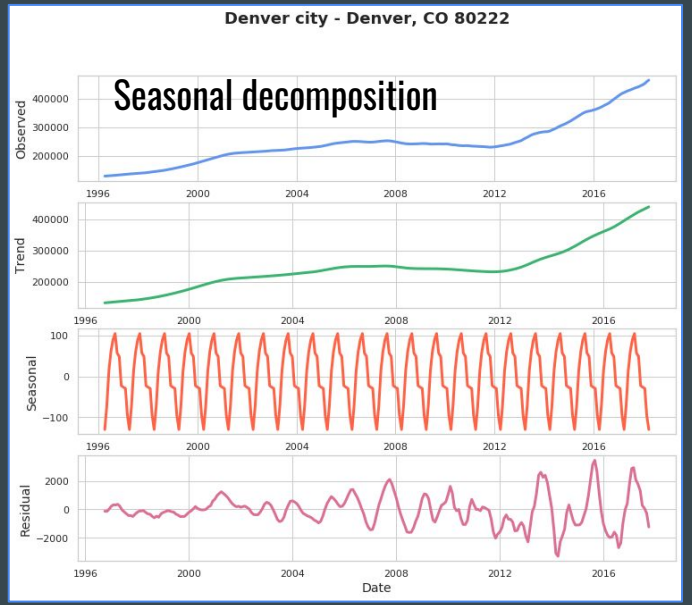


Validation

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

Plot diagnostics



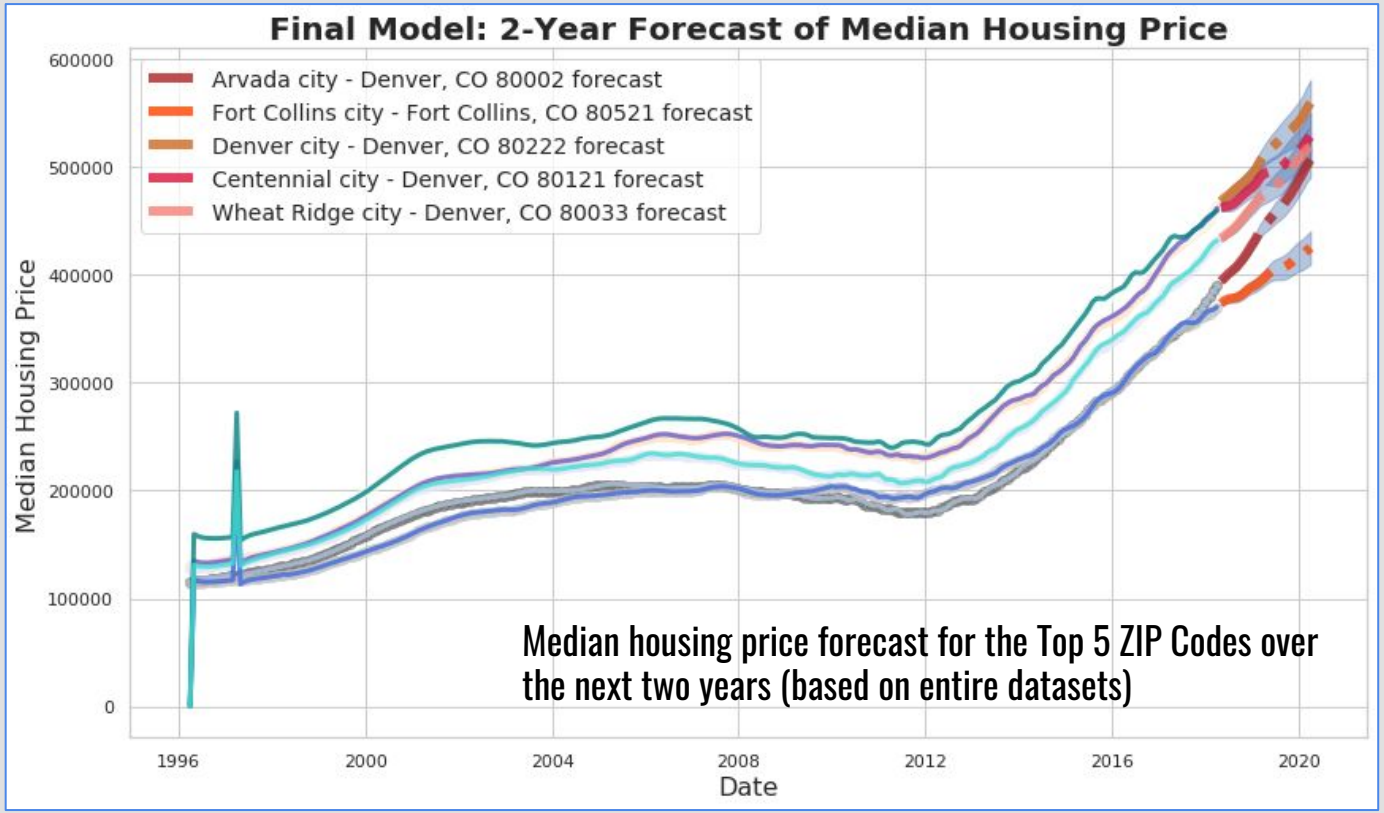


Prediction variation is not consistent over time; this is at least partly due to the inherent non-normality in housing price data

Recommendations

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

;



Future Work

- Considerations
  - Remove or reduce heteroscedasticity
  - Increase hyperparameters grid search
  - Incorporate additional data sources
    - Income
    - Taxes
    - Demographics
    - Environment
    - Education
    - Quality of Life
    - Cost of Living