Forecasting Future Home Prices

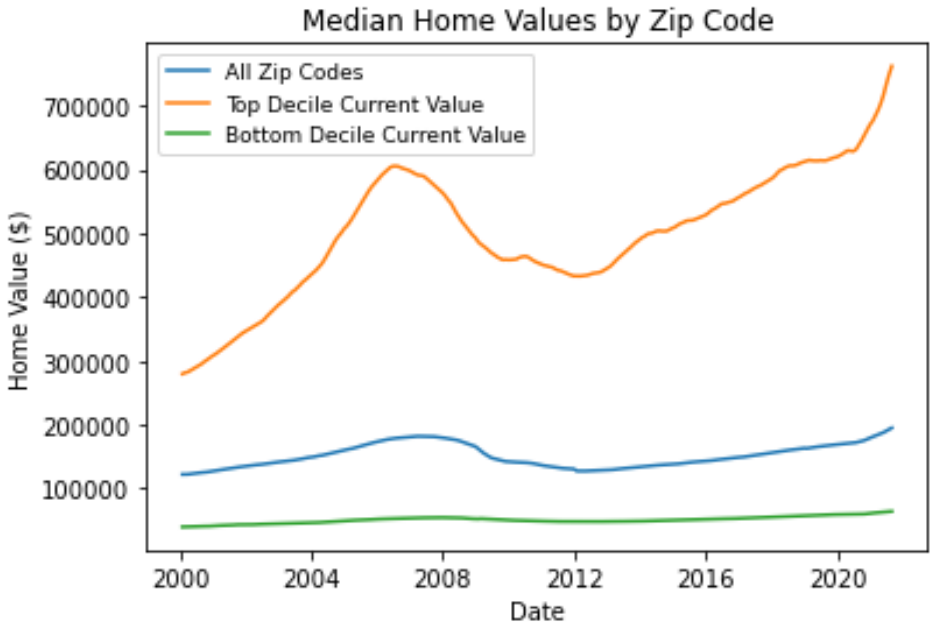
David Brehm
Fall 2021
https://github.com/dtbrehm/DSC680

Business Problem

The goal of this project was to explore time series data to obtain a better understanding of these types of problems. Housing prices are a common example of time series data, and I thought that would be interesting to examine to also understand historical housing prices. Zillow is one of the most well-known online real estate marketplaces, and they also have data available on their historical home values and rental rates. The data used from Zillow was their home values by zip code.

Exploratory Data Analysis

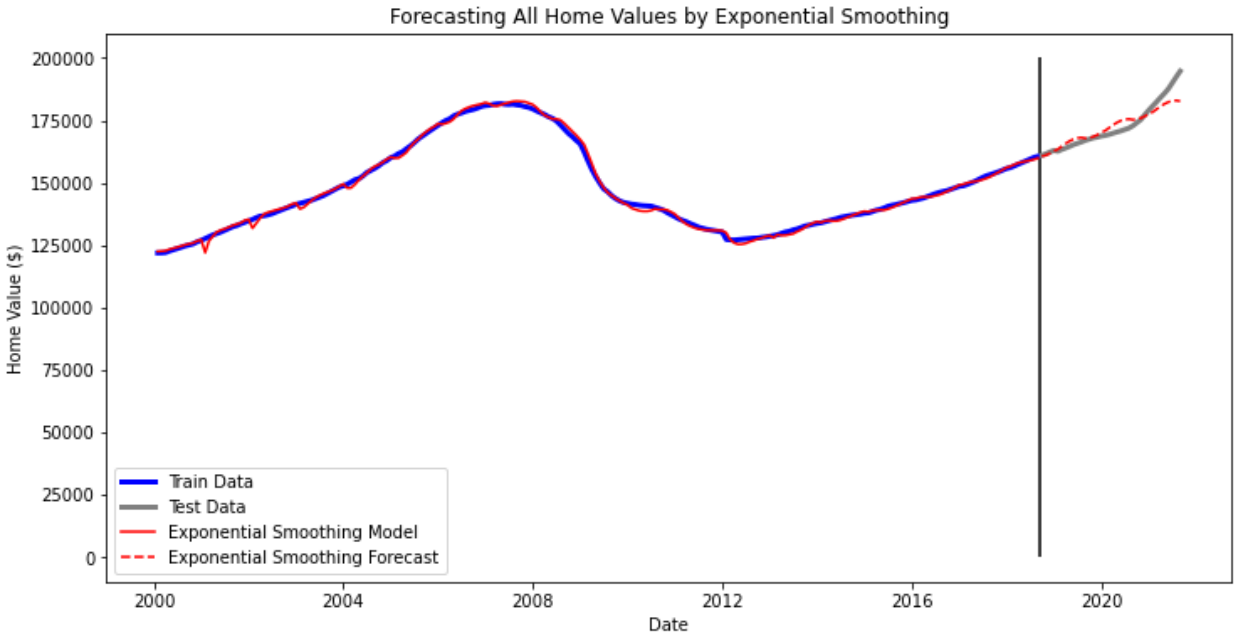
Apart from the forecasting side of this analysis, I also wanted to look into historical housing prices just in a general sense to gain a better grasp on how those have changed over time. I decided to break this out by all zip codes as well as by top and bottom decile of current home values. The graph below shows how these three subsets have changed since 2000. As we can see, the median home value of top decile zip codes has grown more than the overall median. The bottom decile, while being a lower value overall, also has a lower growth rate. The top decile has grown around 188%, while the bottom decile has only grown around 62%.

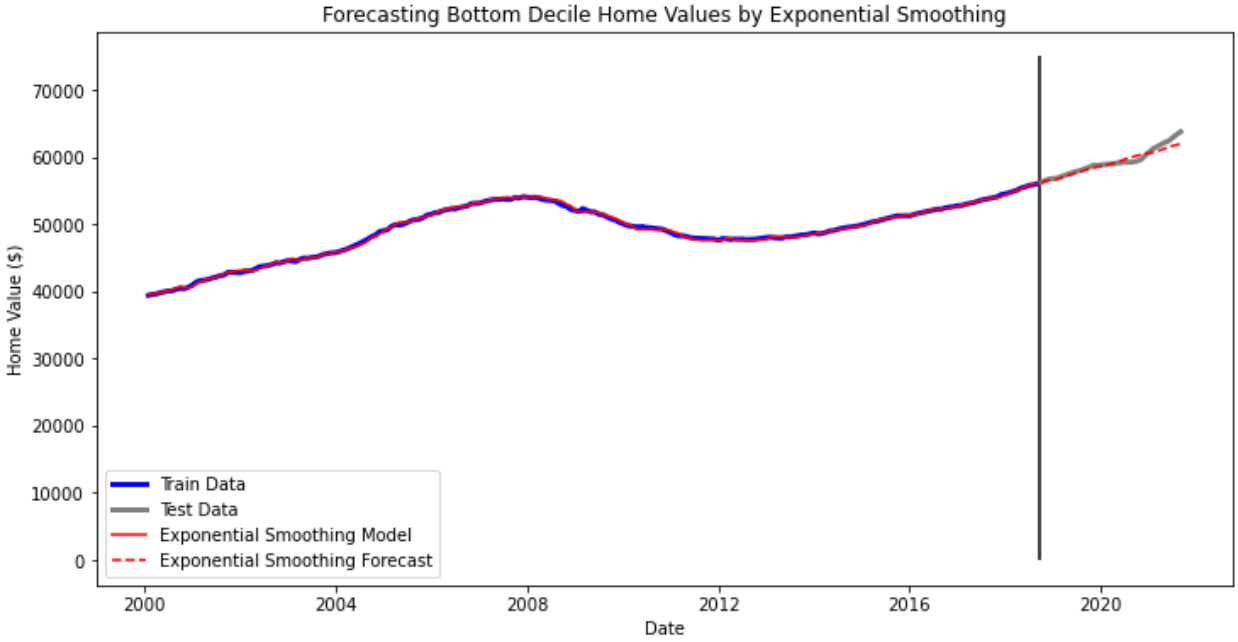


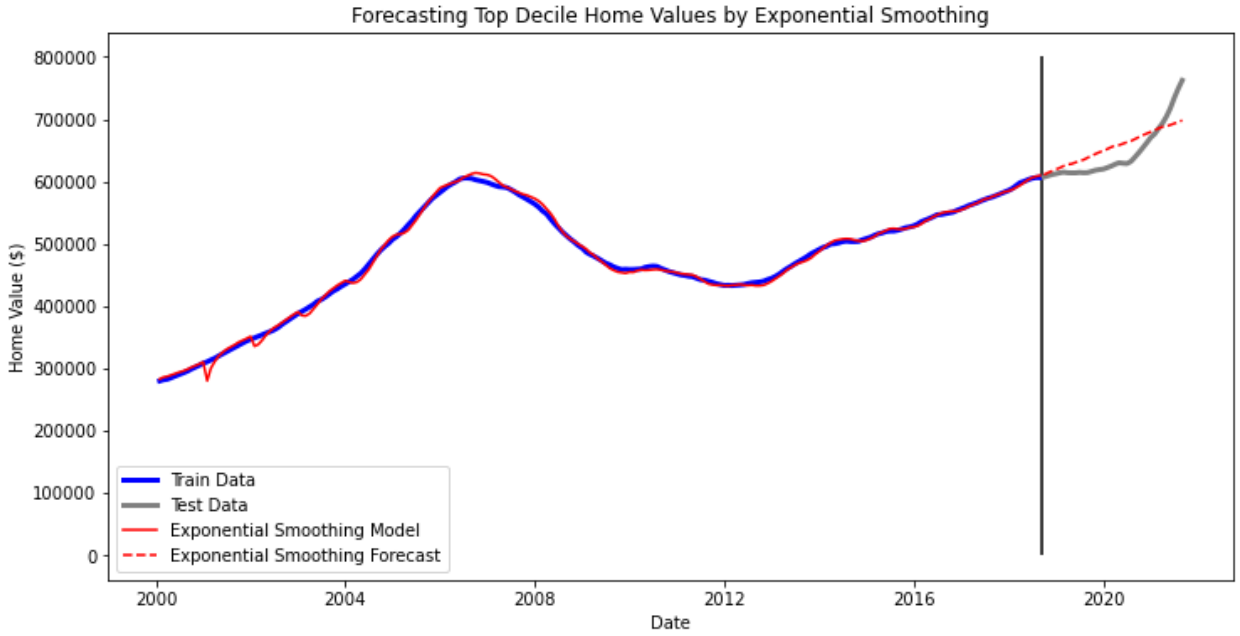

```
Median percent change, all zip codes:                          120.53 %
Median percent change, top decile current home value:          187.67 %
Median percent change, bottom decile current home value:       61.9 %
```

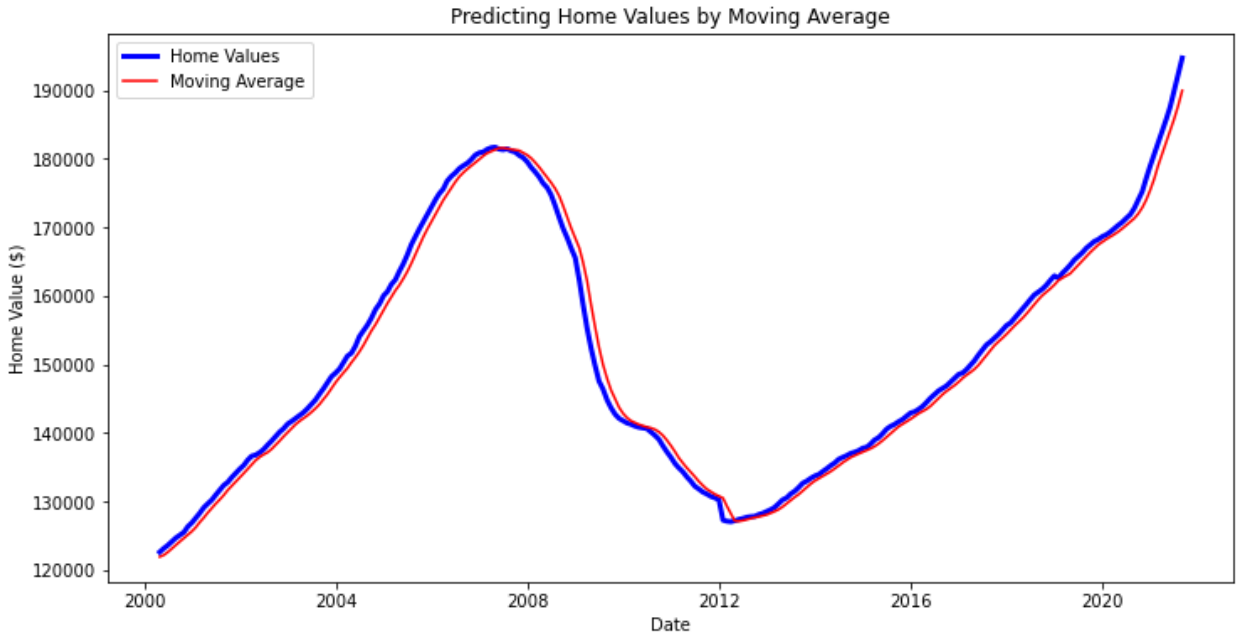
Forecasting

I decided to look into two methods of estimating time series data, Holt-Winters exponential smoothing and a basic moving average. For the exponential smoothing, the most recent three years is used as test data while all of the data prior to that is used as training data. For the moving average, the window was set to three months.









Results

The bottom decile exponential smoothing performed the best by mean square error, while the top decile performed the worst.

```
Exponential Smoothing MSE, All Zip Codes :          13286870.12
Exponential Smoothing MSE, Top Decile :             661850462.39
Exponential Smoothing MSE, Bottom Decile :          377245.08
Moving Average MSE:                                 3077186.86
```

References

Zillow Housing Data:

Bowers, M. (2021, March 25). Housing Data. Zillow Research. https://www.zillow.com/research/data/