

jxn628 / dsc-phase-4-project Public

☆ 1 star 🍴 0 forks

★ Starred

👁 Unwatch ▾

<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

🔑 main ▾

...

👤 jxn628 deleted ...

11 seconds ago ⌚ 26

[View code](#)

☰ README.md



Phase 4 Project: Real Estate Time Series Zip Code Analysis

Summary

Business Problem FRC Investors is a nationwide real estate investment company. They buy homes all across the country and then sell them through their realty branch, FRC Homes. FRC's executive team is interested in maximizing their efforts and has tasked me with answering the question: **What are the top 5 best zip codes for us to invest in?**

FRC Investors is a nationwide real estate investment company. They buy homes all across the country and then sell them through their realty branch, FRC Homes. FRC's executive team is interested in maximizing their efforts and has tasked me with answering the question: **What are the top 5 best zip codes for us to invest in?**

Criteria: In my meeting with the executive team, I was given the following directions:

1. FRC Investors is looking to **invest 500,000 per home per zip code.**
2. They want to diversify by investing in **5 different zip codes, in 5 different states.**
3. FRC would like to **maximize their ROI** per home.
4. FRC would like to **minimize risk.** They want "sure thing" investments.
5. FRC is ready to invest now, and **wants forecasts for the next 3 years,.**

Therefore, I classify the "best" zip codes as having:

1. Median house prices around 500,000.
2. Highest forecasted ROI over the next 3 years (2018-2020)
3. Lowest chance of loss. (confidence index lower value)

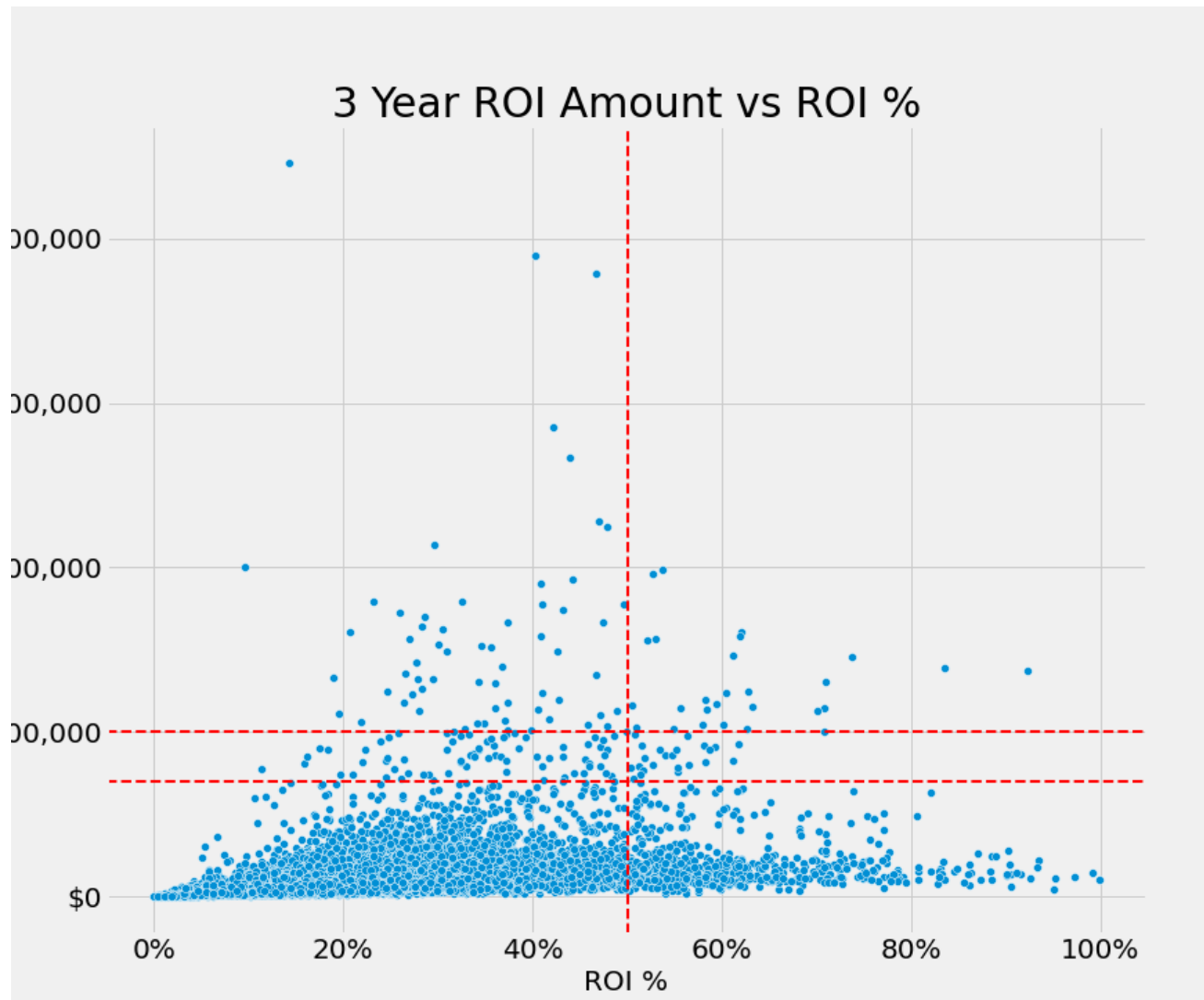
The Data

- I will be using the Zillow Median Home Price Research Dataset. It contains the median home prices for every Zip Code in the United States by month from January 1996 - April 2018.
- The data is fairly clean, although many of the Zip Codes don't have data towards the beginning of the dataset. My hypothesis is that these Zip Codes did not exist at that time. Either way, as this is all time series data, I will work with what I've got for each Zip Code.

My Process:

1. I searched the data provided for zip codes which have median housing prices near 500,000 at the end of the data (early 2018).
2. I then analyzed the performance of those zip codes over the last 3 years.
3. I ranked my findings based on ROI generated over those last 3 years.
4. I then transformed each of those zip codes into a unique time series for modeling.
5. I used SARIMAX modelling to forecast the future performance of each of my target zip Codes.
6. I used a parameter grid search to determine the optimal parameters for P, D, and Q.
7. I also withheld the last few years of data (usually 20% of dataset) as a test set to validate my model.
8. I validated my model by comparing the Root Mean Squared Error as a judge of model fit.
9. I also did a one-step forecast for each model and compared the prediction to the test set.
10. Once I determined the better performing model, I then fit that model on that zip code's entire time series to get a 3-year forecast.
11. I then analyzed the predicted mean and confidence intervals to determine how safe each zipcode would be to invest in over a three year period, as well as the potential Return on Investment.
12. I went through my target zip codes in the order of their ranked 3 year ROI from 2016-2018.
13. In order to preserve geographic integrity (per the Criteria listed above), once a Zip Code had been selected as one of the Top 5, I bypassed any future Zip Codes from that state and pursued zip codes from the remaining states.

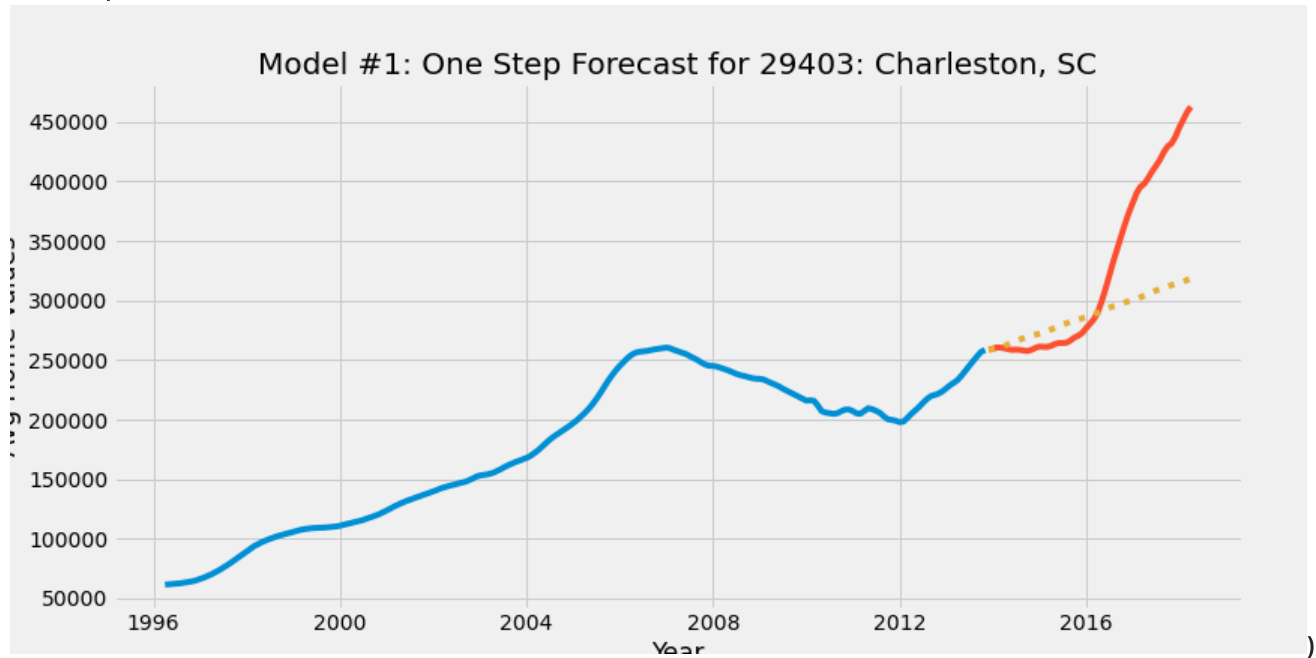
Target Zip Codes



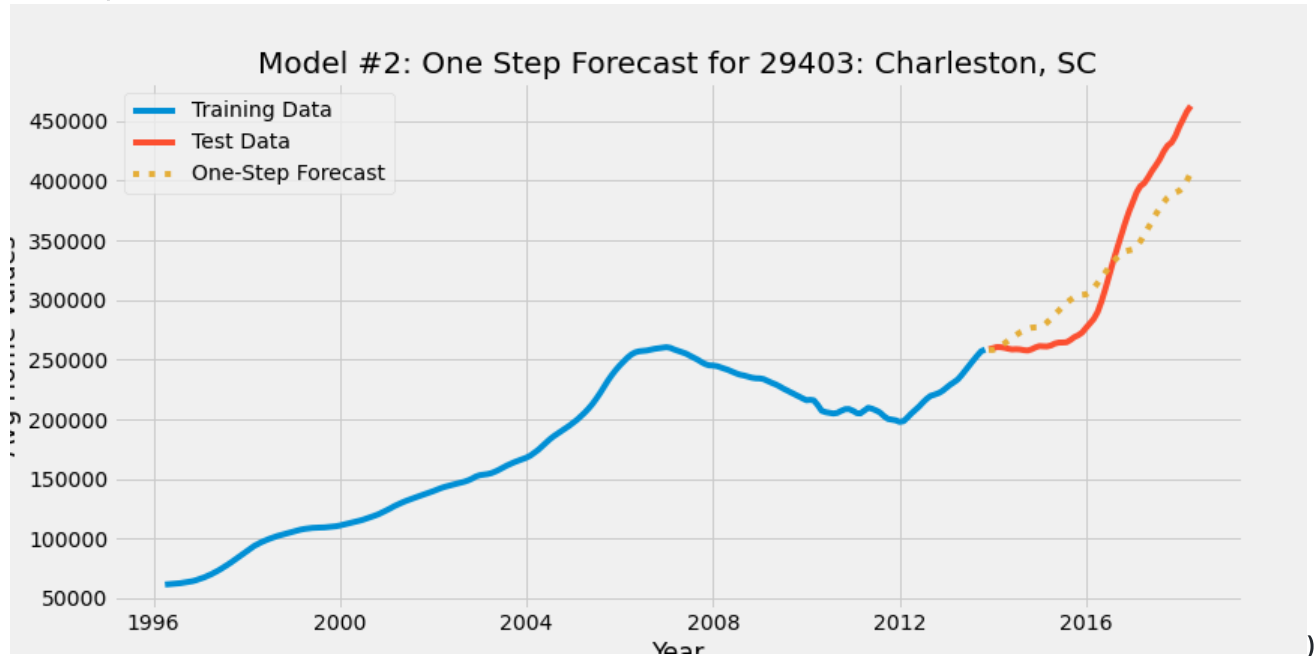
The red lines show where my target zip codes are. Per the criteria that I was given, the client wants to invest approximately \$500,000 per house per zip code. So I have set a range of 350,000 - 500,000 for home value. I also drew a vertical line at 50 Percent ROI as I am interested in the Zip Codes that are performing exceptionally well. The section of the scatterplot between the two horizontal red lines and to the right of the vertical line is where the Target Zip Codes are located.

Modelling 2 One Step Forecast SARIMA Models will be done on each Target Zip Code. Example: 29403:Charleston, SC

One-Step-Forecast: Model 1

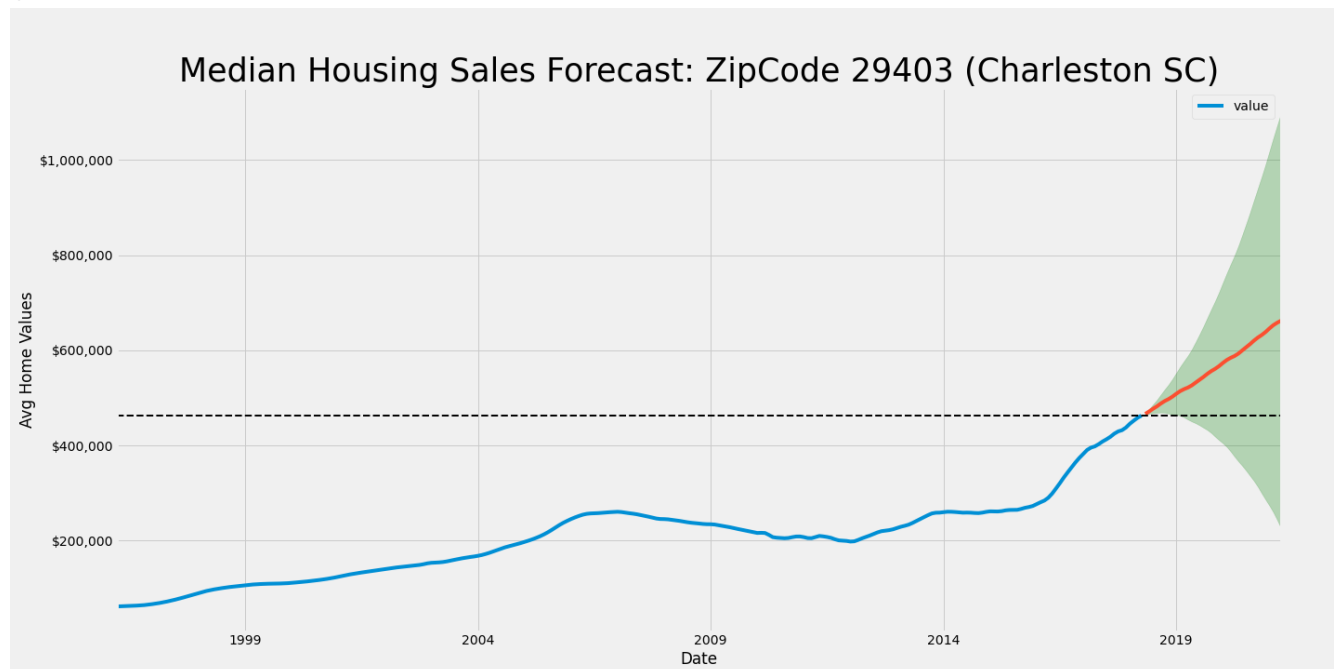


One-Step-Forecast: Model 2



Model Comparison & Selection Model 2's forecast is a better match for the test data. I will use this to make my predictions.

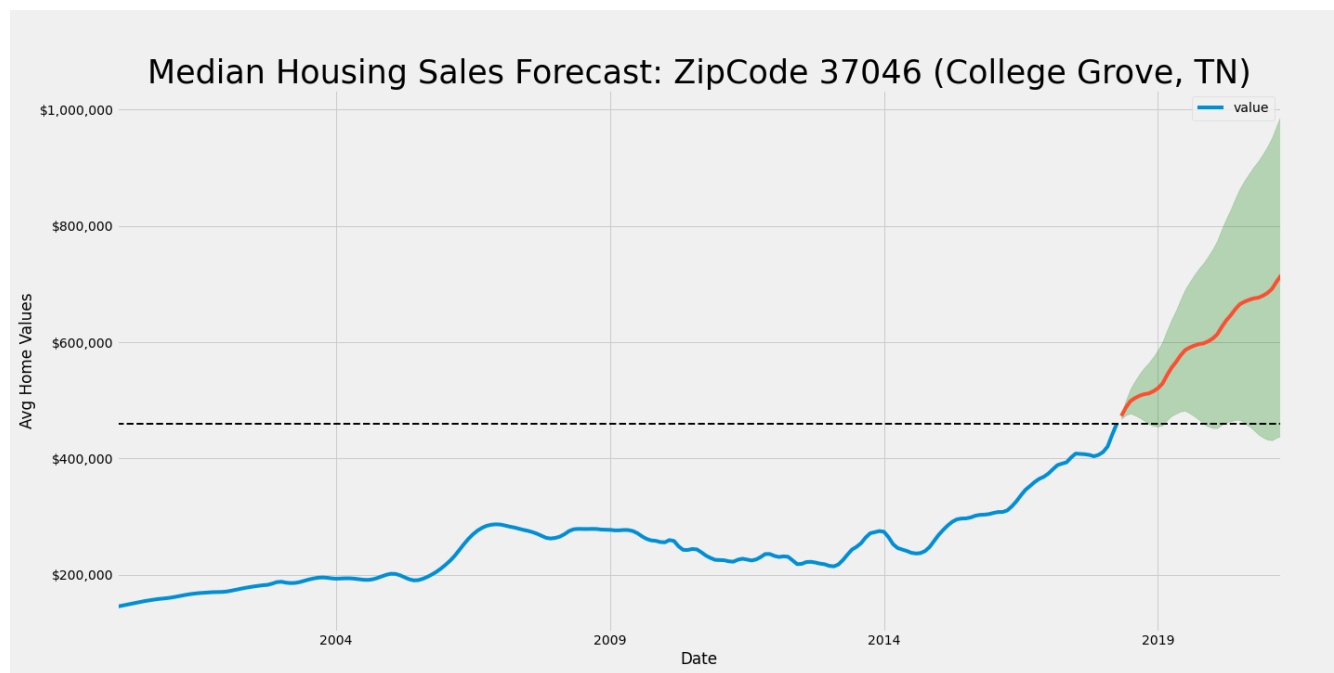
Charleston Forecast



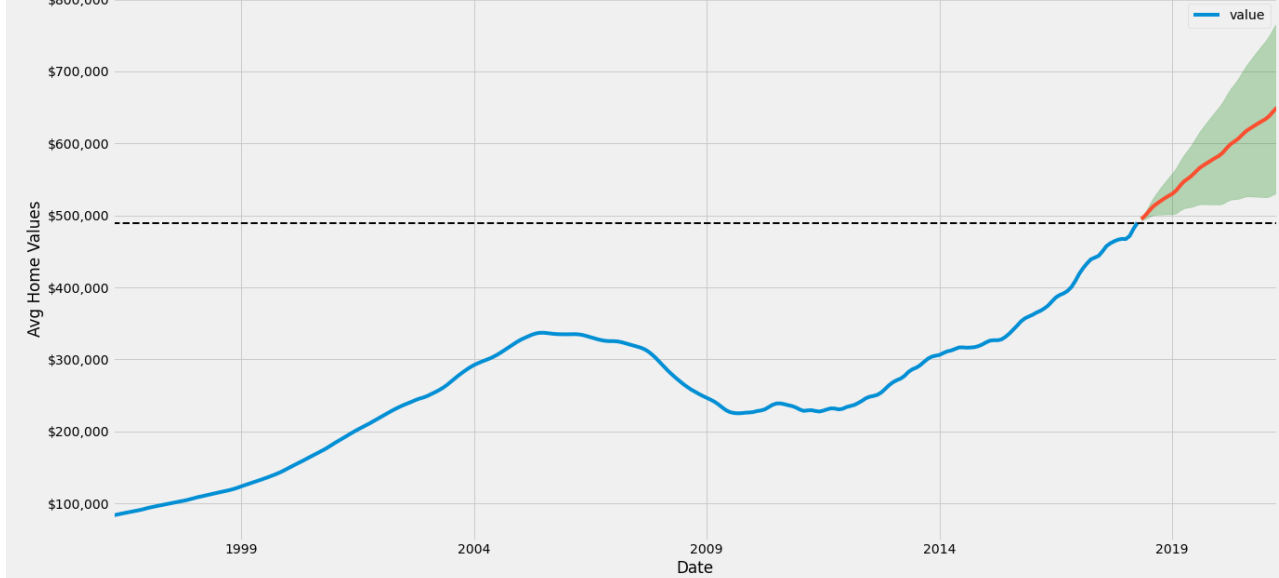
Final Analysis:

- While 1-2 Years out looks good, there is too much uncertainty in Forecast Year 3 to recommend Charleston as one of the Top 5 Zip Codes.
- That said, if you reduced the investment window to 1-2 years, Charleston is a good choice and therefore should be considered with that caveat.
- I will recommend this Zip Code as an alternative recommendation with the caveat that it should be re-evaluated during the 1-2 year range.

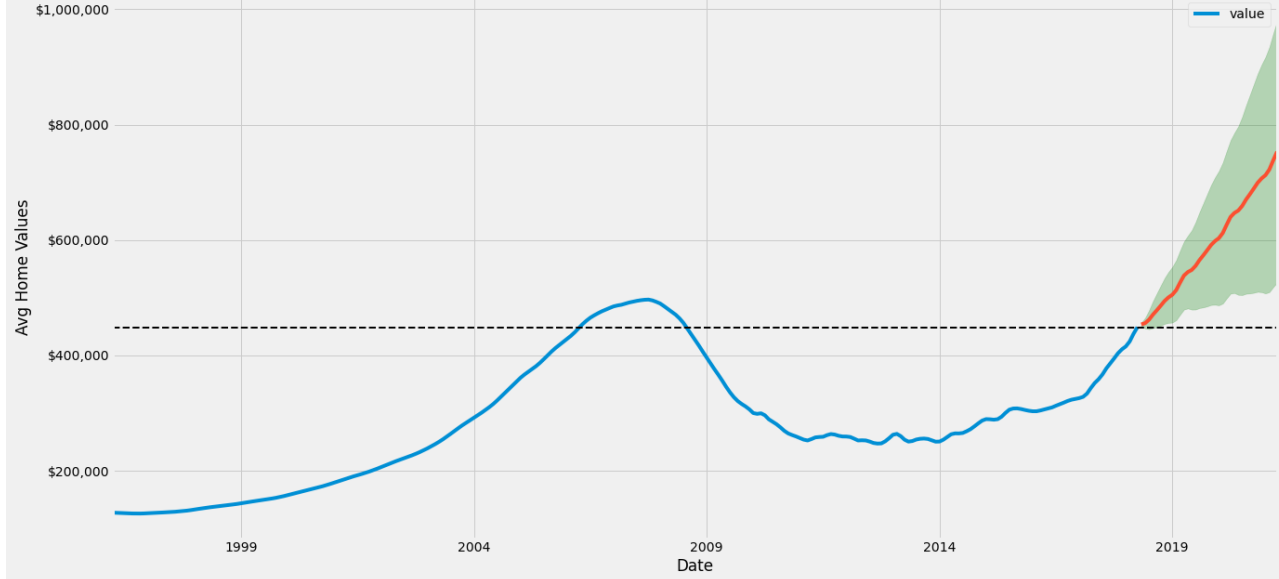
After Modeling 15 Zip Codes, These are the 5 that I recommended



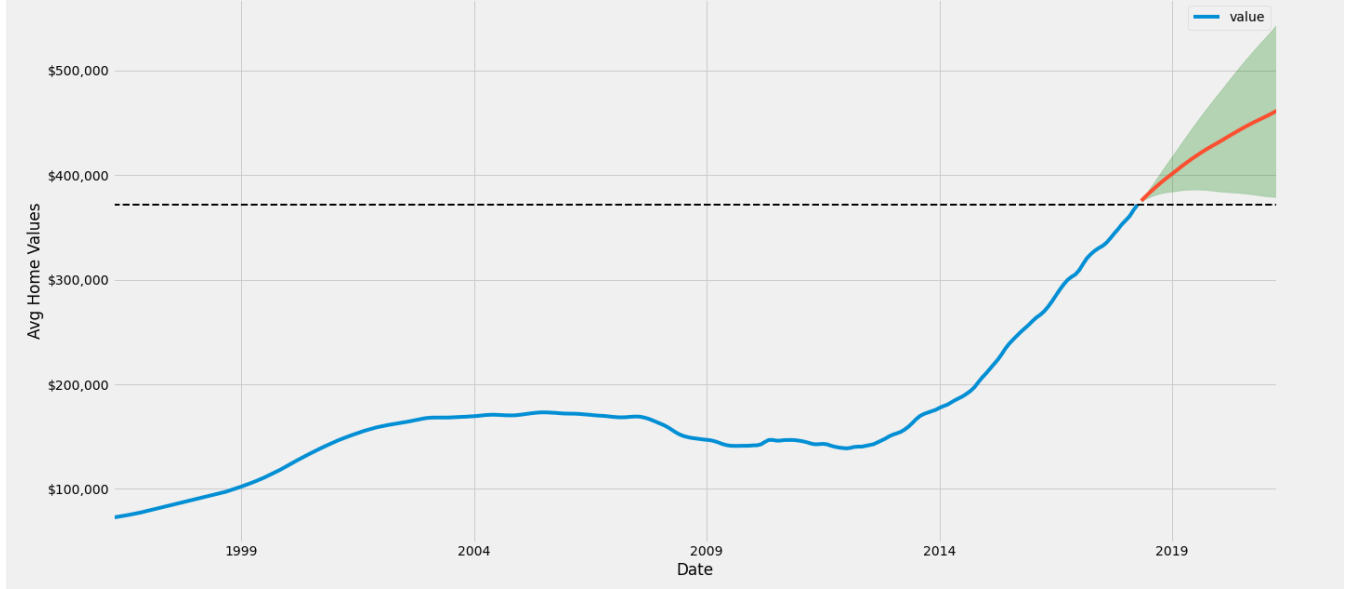
Median Housing Sales Forecast: ZipCode 02128 (Boston MA)



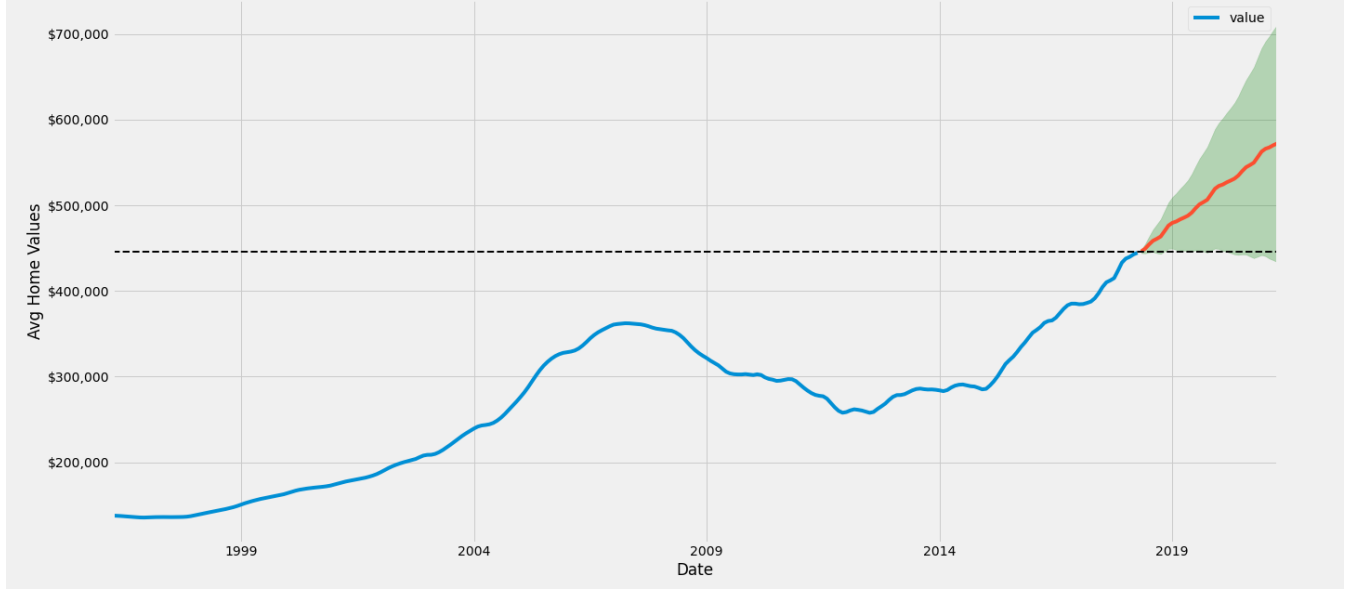
Median Housing Sales Forecast: ZipCode 07087 (Union City NJ)



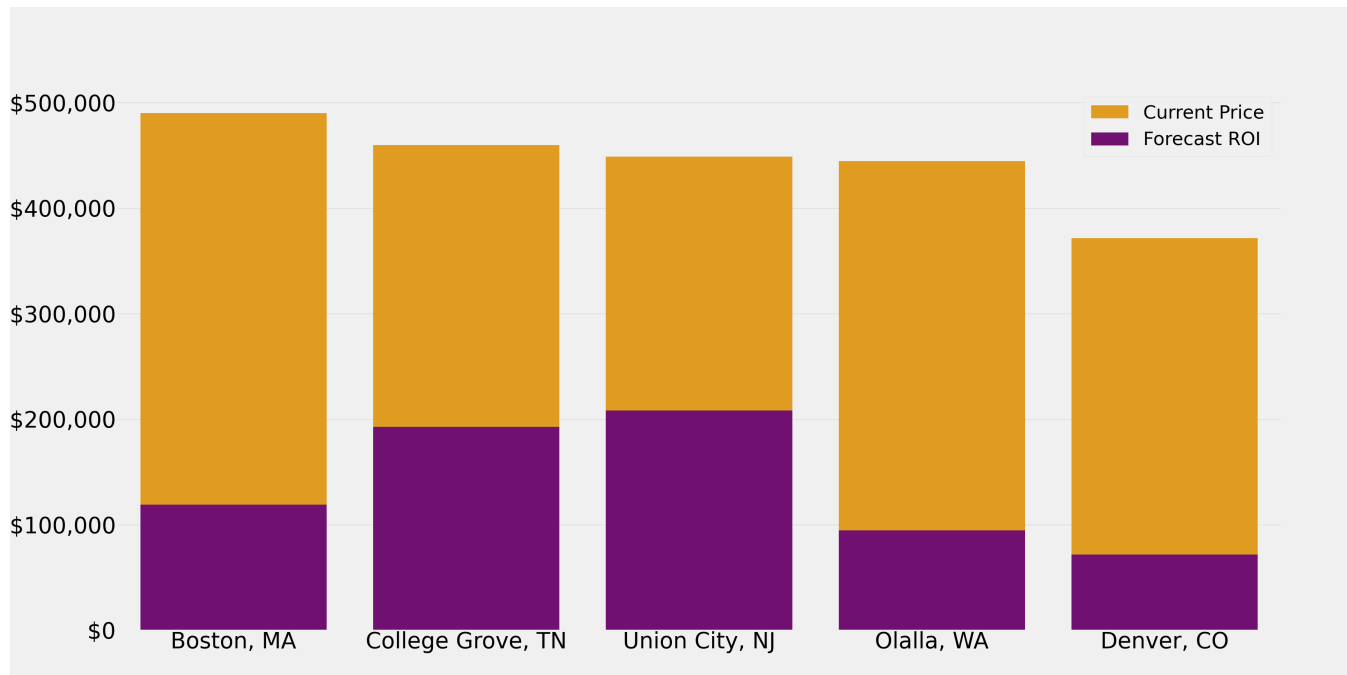
Median Housing Sales Forecast: ZipCode 80204 (Denver CO)



Median Housing Sales Forecast: ZipCode 98359 (Olalla WA)



Final Analysis



Recommendations

The Top 5 Zip Codes completely meet the criteria that I was given:

- Cost of Entry: Average Home Values around 500,000 -- (currently 370,000 - 489,000)
- Geographic Diversity: Each is in a different State.
- High ROI: Each is forecasted to have a high return on investment over the next 3 years.

-- Mean Forecast Values for Forecast Year 3: 1.2% ROI - 1.6% ROI.

-- Mean Forecast ROI Amounts for Forecast Year 3: 87,000 - 282,000.

- Low Risk: Each has been forecast with a low risk of losing money over the next 3 years.

-- Maximum Forecast Loss at Forecast Year 3 : -5% ROI

- Timing: All forecasts were for 3 years.

Recommendation #2: ZipCodes for Shorter investment ZipCodes that showed a good return with low risk for 1 or 2 years.

- i. 29403 Charleston SC
- ii. 98043 Mountlake Terrace WA

Conclusion These Zip Codes still show a lot of potential, but need to be watched more closely. They all show great potential for one or even two years of investment, but then the third year forecast becomes more uncertain.

I am in confident in recommending both of these Zip Codes during that shorter investment window, and then running models again as more data comes in.

- Analyzing the past performance over the last 3 years proved to be a good way to identify zip codes that have a high probability to continue performing well 2-3 years into the future.
- While not all Zip Codes that I identified and targetted ended up being recommended, I only had to model 15 Zip Codes in order to get my 5 "best" Zip Codes to recommend. It also led to me discovering 2-3 Zip Codes that nearly met all criteria and became a strong alternative recommendation.

- The approach taken throughout this entire process is replicatable and can and should be done often to monitor zip codes where FRC is focusing as well as identifying new markets to enter into or increase their presence in. If FRC continues to do this, they will have a distinct competitive advantage in the markets where they have a presence and will be able to increase the value of their investments for both the Company and it's shareholders.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%