Analyzing Real Estate Data with ARIMA

Derek VanBriesen

Business Case

A real estate investment firm has asked the following question:

What are the five best ZIP codes for us to invest in?

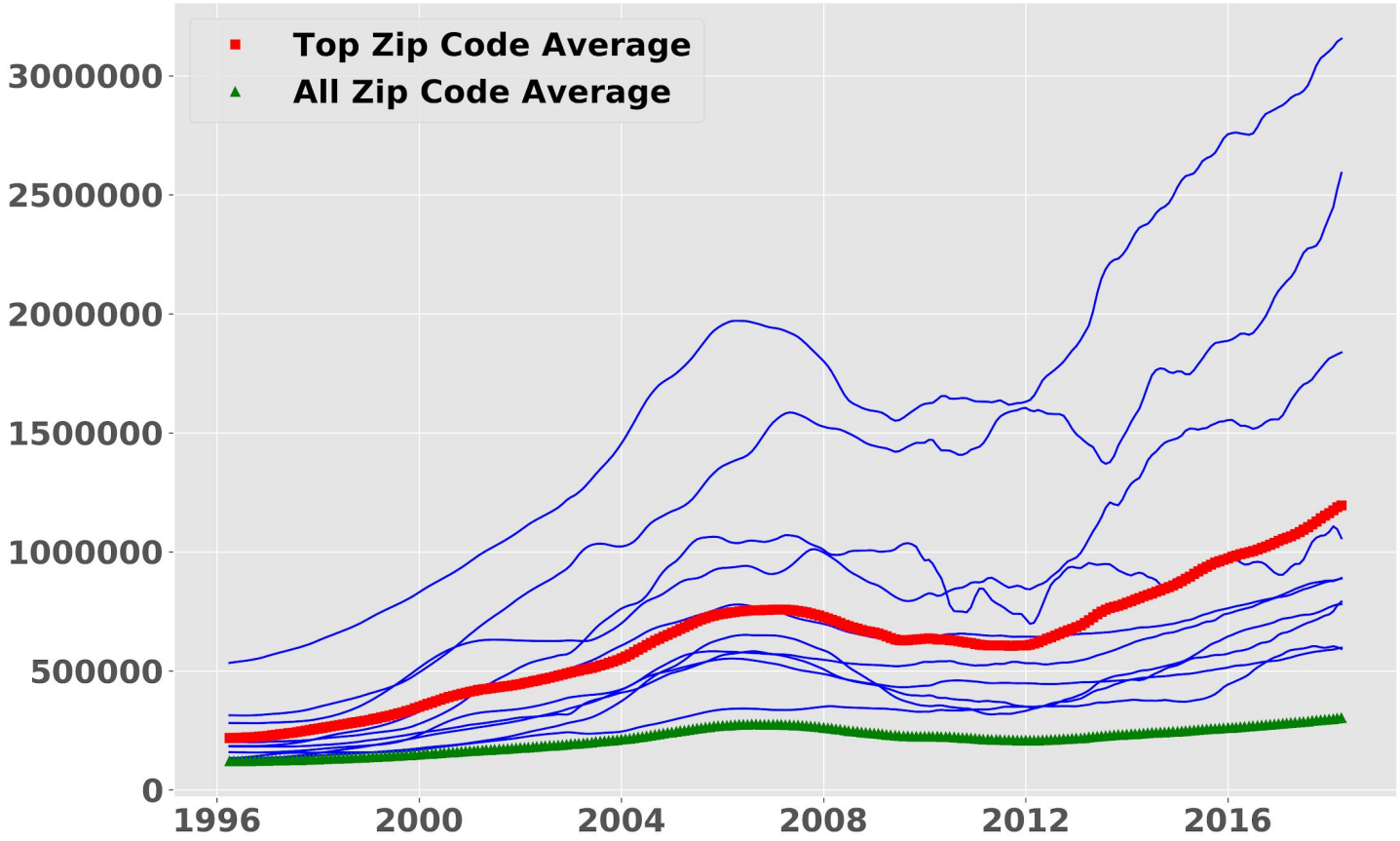
Questions to Answer

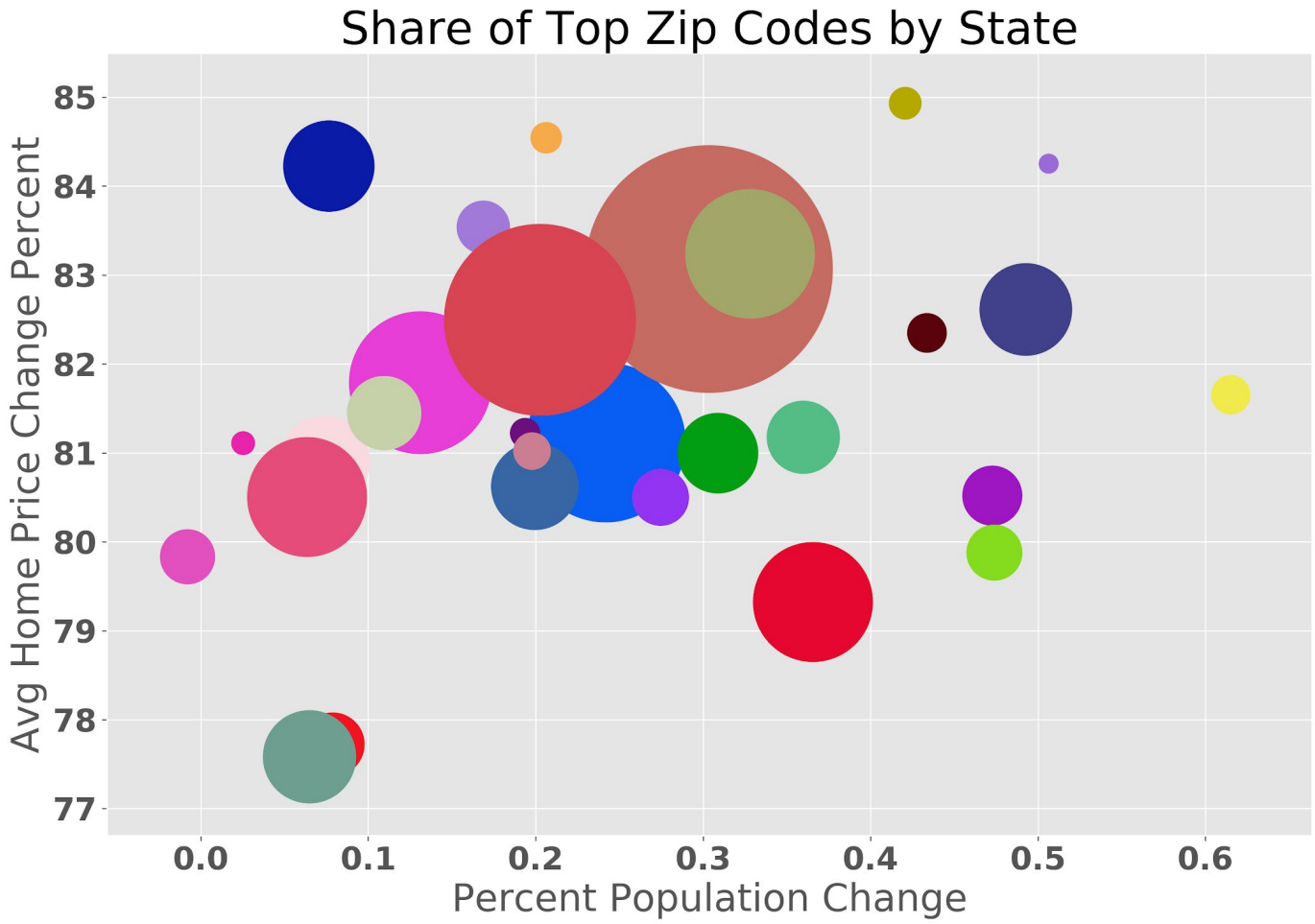
1. What criteria should be used to evaluate the investments?
2. What criteria should be used to evaluate the models?
3. What are the five best ZIP codes for investment?

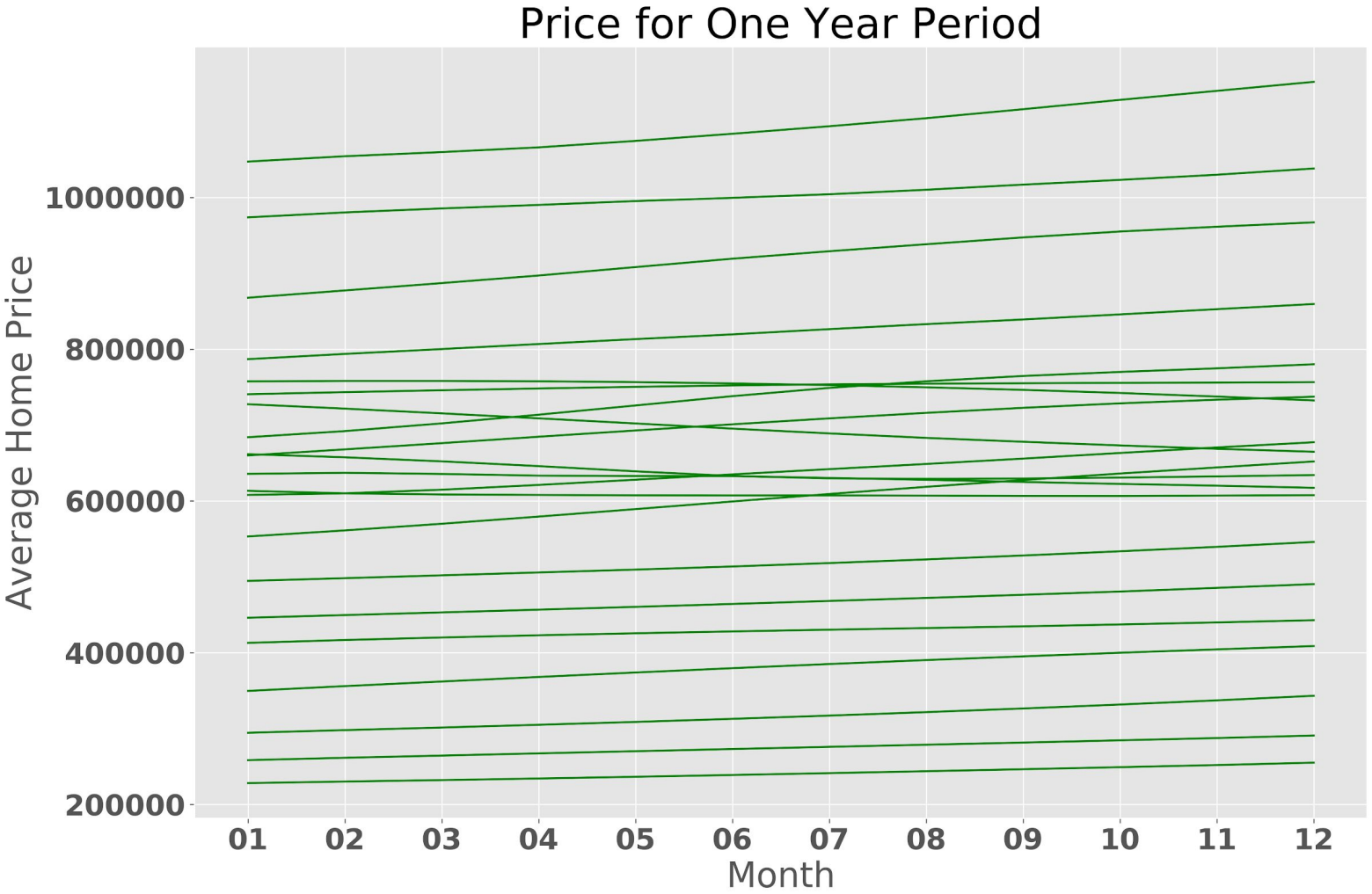
The Data

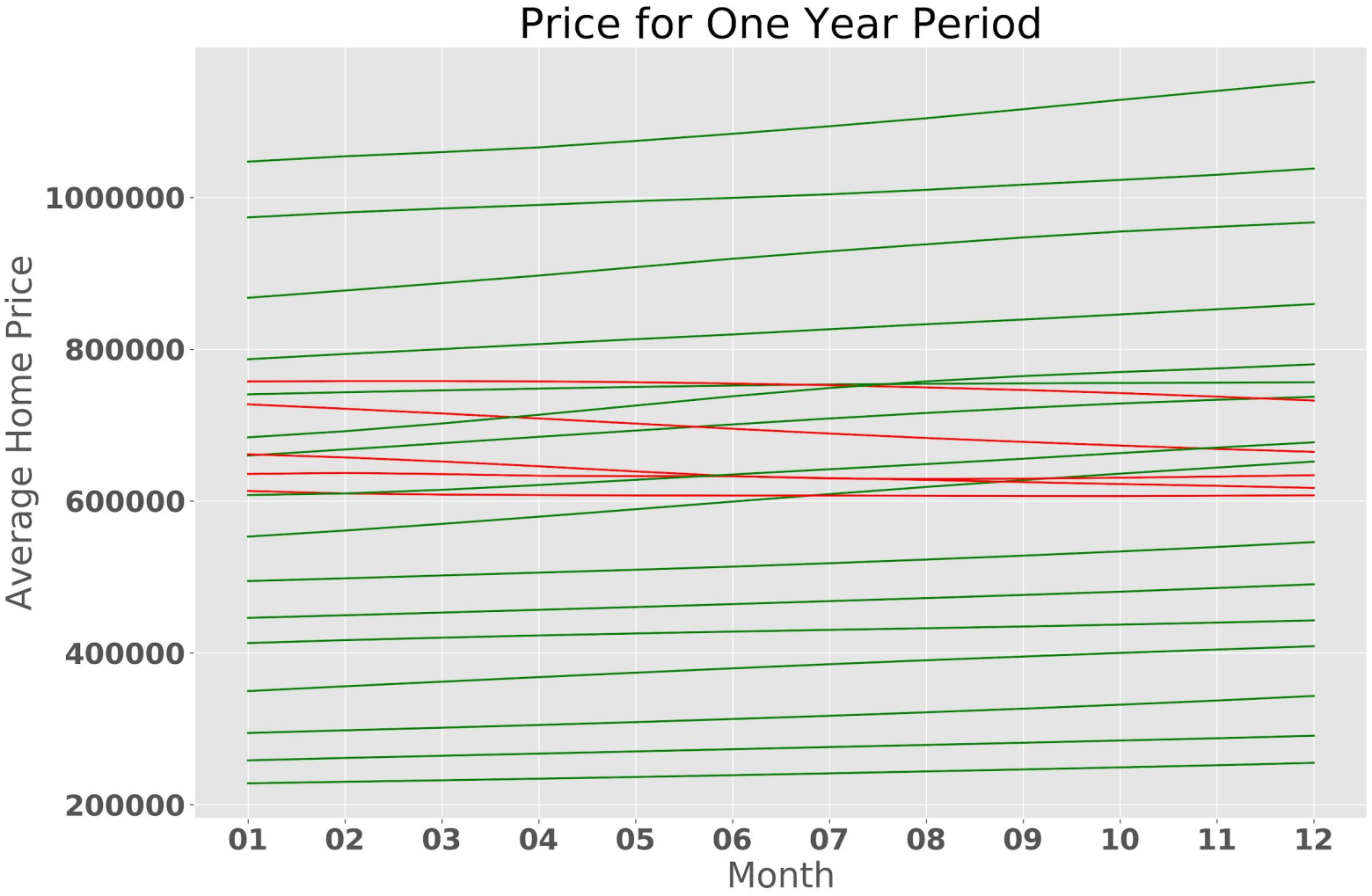
- Zillow Data Set
  - Time series data on home prices between April 1996 and April 2018
  - Contains an average representing each ZIP code
  - This was filtered down using the overall ROI from 1996-2018
- Census Data
  - State by state population estimates from the Census Bureau
  - Covers the beginning and end dates of the time series (1996 & 2018)

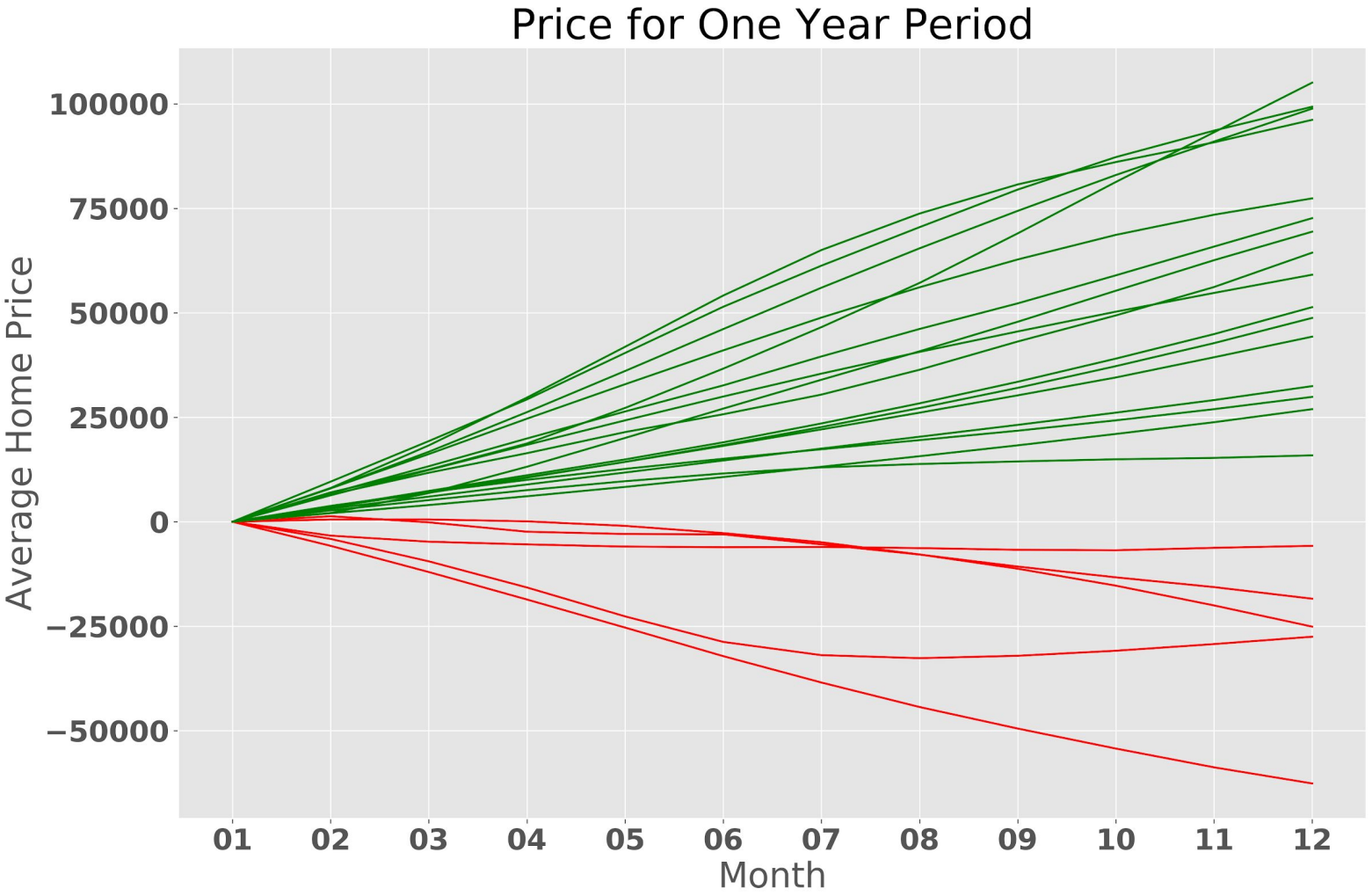
Exploratory Data Analysis


Top Zip Code Average
All Zip Code Average
3000000
2500000
2000000
1500000
1000000
500000
0
1996
2000
2004
2008
2012
2016


Share of Top Zip Codes by State
Avg Home Price Change Percent
85
84
83
82
81
80
79
78
77
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Percent Population Change


Price for One Year Period
Average Home Price
1000000
800000
600000
400000
200000
01
02
03
04
05
06
07
08
09
10
11
12
Month


Price for One Year Period
Average Home Price
1000000
800000
600000
400000
200000
01
02
03
04
05
06
07
08
09
10
11
12
Month


Price for One Year Period
Average Home Price
100000
75000
50000
25000
0
−25000
−50000
01
02
03
04
05
06
07
08
09
10
11
12
Month

Modeling Process

Modeling Steps

1. Create a model that fits the average time series
2. Check the model against all of the ZIP codes
3. Select the ZIP codes with the highest predicted values
4. Adjust parameters and model each of these separately
5. Rank the ZIP codes according to my investment metrics

Investment Metrics

- Return on investment
- Adjust for risk
- Five year period

Why?

- ROI helps find the greatest profit potential
- Adjusting for risk ensures that the downside doesn't outweigh the benefits
- Five years is a fairly standard investment period, but not too far out

Modeling Criteria

- Akaike information criterion (AIC)
  - This checks for overfitting and underfitting
  - Used when tuning models
- Mean Squared Error
  - This was used when adjusting the ROI for risk

Results

Top Five ZIP Codes

1. 90057
2. 11959
3. 94063
4. 95133
5. 95128

Future Work

- Test whether different filtering methods provide better results
- More time to test model parameters
- Fully model more individual ZIP codes

Thank you for listening!

Citations

1. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html

2. https://www.zillow.com/research/data/

3. https://www2.census.gov/programs-surveys/popest/tables/1990-2000/state/totals/st-99-03.txt

4. https://worldpopulationreview.com/states/state-abbreviations