

REAL ESTATE INVESTING

Consultant: Harrison Gu



Business Case:

Our private wealth management firm recently acquired a high value client, who is looking to invest in real estate, with the goal of passing the investments to their kids. The first priority is risk aversion, followed by % returns.

Project objective:

Our aim is to build and optimize time series models to predict median house prices across various zip codes 20-30 years in the future.

Data

Our data was taken from [zillow.com](https://www.zillow.com), and contains monthly average median house prices of 14,273 zip codes from across the country, from April 1996 to April 2018.

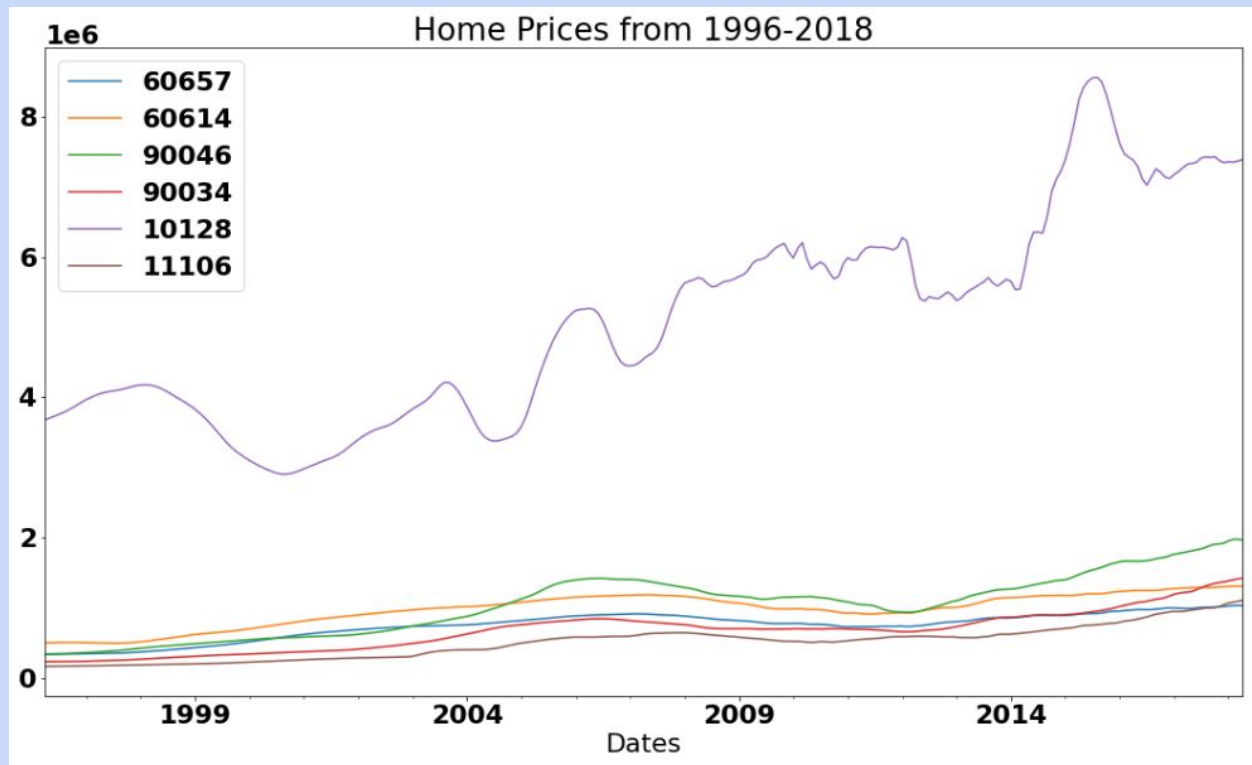
Because our client values risk aversion, we are only going to look at the largest zip codes in the 3 most populated cities:

- Chicago
- Los Angeles
- New York (Manhattan)

	RegionID	RegionName	City	State	Metro	CountyName	SizeRank	1996-04	1996-05	1996-06
0	84654	60657	Chicago	IL	Chicago	Cook	1	334200.0	335400.0	336500.0
1	84616	60614	Chicago	IL	Chicago	Cook	4	498100.0	500900.0	503100.0
2	96027	90046	Los Angeles	CA	Los Angeles-Long Beach-Anaheim	Los Angeles	31	340600.0	341700.0	343000.0
3	96015	90034	Los Angeles	CA	Los Angeles-Long Beach-Anaheim	Los Angeles	76	231300.0	231700.0	232100.0
4	61703	10128	New York	NY	New York	New York	22	3676700.0	3704200.0	3729600.0
5	62009	11106	New York	NY	New York	New York	1066	164600.0	165300.0	165900.0

2017-07	2017-08	2017-09	2017-10	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04
1005500	1007500	1007800	1009600	1013300	1018700	1024400	1030700	1033800	1030600
1289800	1287700	1287400	1291500	1296600	1299000	1302700	1306400	1308500	1307000
1839800	1861100	1888600	1903900	1907500	1922100	1952400	1974500	1975900	1966900
1287500	1310000	1329900	1347200	1360300	1368500	1382700	1398000	1411400	1419100
7410100	7422400	7417600	7427300	7371400	7342700	7353300	7350300	7363000	7386600
987400	995600	997600	1000800	1017900	1043900	1067100	1079200	1091000	1104300

Data

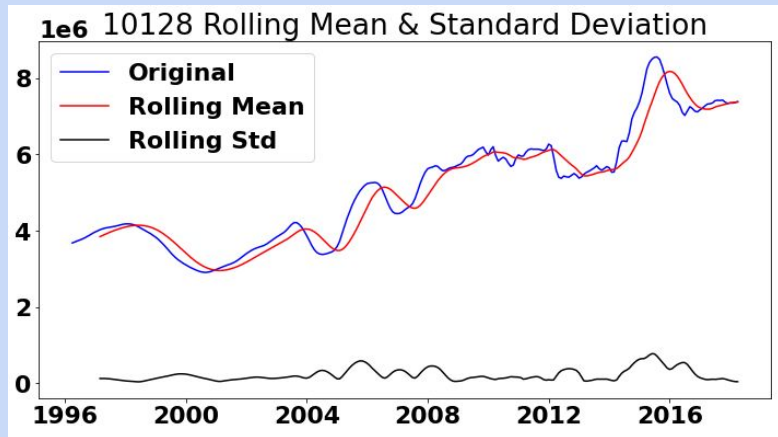


EDA

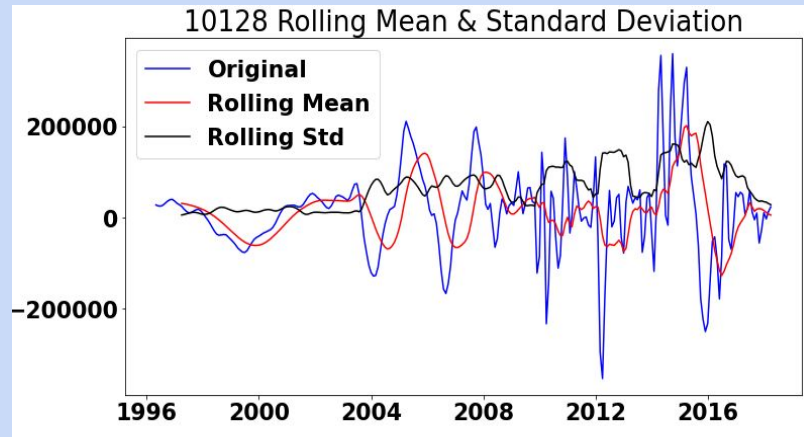
To detrend each of the 6 datasets, we used differencing (various periods), log scaling, or a combination of the 2.

In this example, we used differencing by 1 month.

Before



After



ARIMA

We ran baseline ARIMA models for all of the zip codes, using parameters determined by our EDA. We then used auto ARIMA and auto ARIMA dynamic forecasting to try to obtain better results.

Zip Code	Baseline		Auto ARIMA		Auto ARIMA Dynamic	
	order (p,d,q)	AIC	order (p,d,q)	AIC	order (p,d,q)	AIC
60657	1,0,19	-1987	2,1,4	-2017	2,1,4	-1957
60614	1,0,19	-1920	1,1,1	-1945	1,1,1	-1926
90046	0,3,12	-1938	1,1,8	-1980	1,1,8	-1887
90034	0,2,12	-1904	1,1,7	-2020	1,1,7	-1927
10128	5,1,5	5357	4,0,1	4774	4,0,1	4731
11106	2,3,7	-1706	1,1,5	-1798	1,1,5	-1737

FB Prophet

We also ran models for all of the zip codes using FB Prophet. Because FB Prophet does not generate an AIC score, and the preprocessing for FB Prophet is different than that for ARIMA model, it is difficult for us to compare with ARIMA at this moment.

10128	Train RMSE	Test RMSE
Auto ARIMA Dynamic	267597	1486386
FB Prophet	332096	1025583

Predictions

ARIMA

Best Zip Codes:

1. 10128
2. 90034
3. 11106

Worst Zip Codes:

1. 90046
2. 60657
3. 60614

Zip Code	Price Today	Price in 20 Years	ROI %
60657	\$1,030,600	\$1,000,243	-2.94%
60614	\$1,306,999	\$1,269,479	-2.78%
90046	\$1,966,900	\$1,828,195	-7.05%
90034	\$1,419,009	\$1,631,686	14.98%
10128	\$7,386,600	\$11,905,953	61.18%
11106	\$1,104,300	\$1,251,926	13.68%

FB Prophet

Best Zip Codes:

1. 11106
2. 90034
3. 90046

Worst Zip Codes:

1. 60614
2. 10128
3. 60657

Zip Code	Price Today	Price in 20 Years	ROI %
60657	\$1,030,600	\$1,856,913	80.18%
60614	\$1,306,999	\$2,137,908	63.57%
90046	\$1,966,900	\$5,317,138	170.33%
90034	\$1,419,009	\$3,853,978	171.57%
10128	\$7,386,600	\$12,981,069	75.73%
11106	\$1,104,300	\$3,205,375	190.26%

Next Steps

1. Deep dive into comparing ARIMA and FB Prophet models.
2. Incorporate seasonality and exogenous variables, using SARIMAX, to create better predictive models.
3. Select neighborhoods within the chosen zip codes using the same method.

THANK YOU for your time and attention!

Project Repository: [Github repo link](#)

Authors today can be reached
using the following information:

Harrison Gu:

[*harrison.s.gu@gmail.com*](mailto:harrison.s.gu@gmail.com)

