IST 718 – HOMEWORK 2
PREDICTING HOME PRICES

1 INTRODUCTION

Warren Buffet said Real estate is an asset "for a great many people," particularly for families that plan on being in the same location for many years. "If you know you're going to live in a given area or think it's very likely, for a considerable period of time, and you've got a family, the home is terrific."

We don't have to emphasize how important a house plays in our life. In this analysis will use the past historical house prices to answer the business questions on where to invest for the Syracuse Real Estate Investment Trust (SREIT).

2 ANALYSIS & MODELS

2.1 ABOUT THE DATA

We will be starting with the data set, collected and curated by Zillow for our analysis(http://files.zillowstatic.com/research/public/Zip/Zip_Zhvi_SingleFamilyResidence.csv) The dataset has house prices by zip code from March 1996 to March 202, shown in Figure 1.

RegionID	SizeRank	RegionName	RegionType	StateName	State	City	Metro	CountyName	1996-01-31	1996-02-29
61639	0	10025	Zip	NY	NY	New York	New York...	New York County	nan	nan
84654	1	60657	Zip	IL	IL	Chicago	Chicago-...	Cook County	364892.00000	364162.00000
61637	2	10023	Zip	NY	NY	New York	New York...	New York County	nan	nan
91982	3	77494	Zip	TX	TX	Katy	Houston-...	Harris County	200475.00000	200723.00000
84616	4	60614	Zip	IL	IL	Chicago	Chicago-...	Cook County	546663.00000	546231.00000
91940	5	77449	Zip	TX	TX	Katy	Houston-...	Harris County	97521.00000	97513.00000
61616	6	10002	Zip	NY	NY	New York	New York...	New York County	nan	nan
91733	7	77084	Zip	TX	TX	Houston	Houston-...	Harris County	97381.00000	97405.00000
93144	8	79936	Zip	TX	TX	El Paso	El Paso	El Paso County	82374.00000	82330.00000
84640	9	60640	Zip	IL	IL	Chicago	Chicago-...	Cook County	254388.00000	252774.00000
62037	10	11226	Zip	NY	NY	New York	New York...	Kings County	219465.00000	218301.00000
61807	11	10467	Zip	NY	NY	New York	New York...	Bronx County	239931.00000	239200.00000
92593	12	78660	Zip	TX	TX	Pflugerville	Austin-Ro...	Travis County	144867.00000	144705.00000

FIGURE 1

2.2 DATA PROCESSING

We will convert this wide-format data into a long format where each column date will be converted to a row to do a time-series forecast.

2.3 EXPLORATORY DATA ANALYSIS

We will look at the trend of house prices in four metros in Arkansas, shown in Figure 2. All four metros dipped in the 2007 recession except Fayetteville. And the overall market started rebound starting 2016.
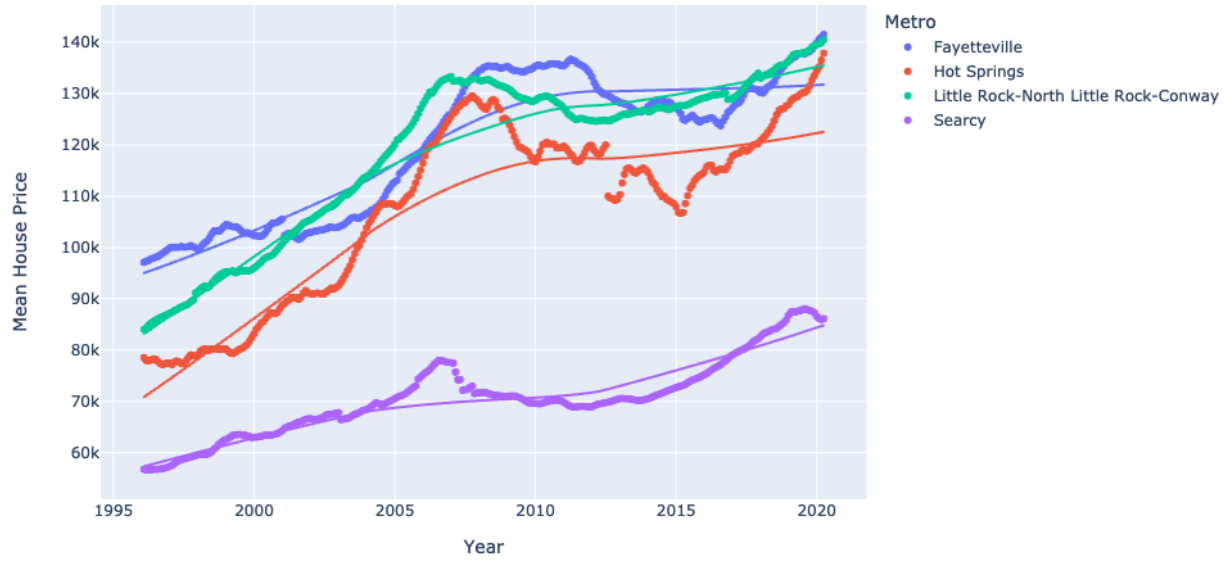
FIGURE 2

When we look at the percent of price change from 1997 to 2020, on average, the house prices increased by 100%; they doubled in value, and in some zip codes, 1500%, that's a perfect investment opportunity.
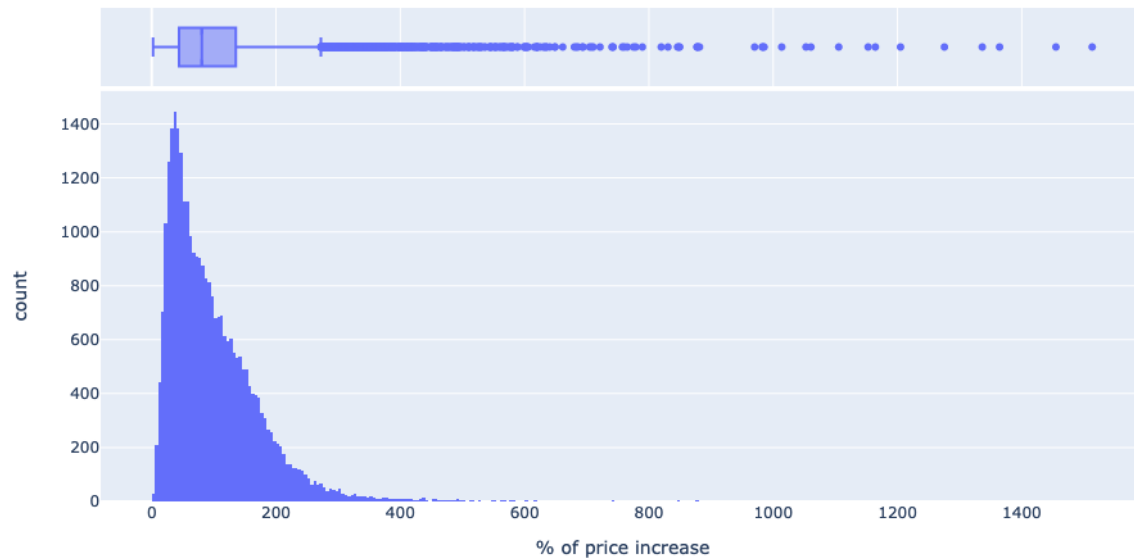

FIGURE 3, PRICE CHANGE FROM 1996

Figure 4 & 5 Show the percent price change for the last five years from 2016. Northwest had an above-average growth, and it may be suited for our investment goals.
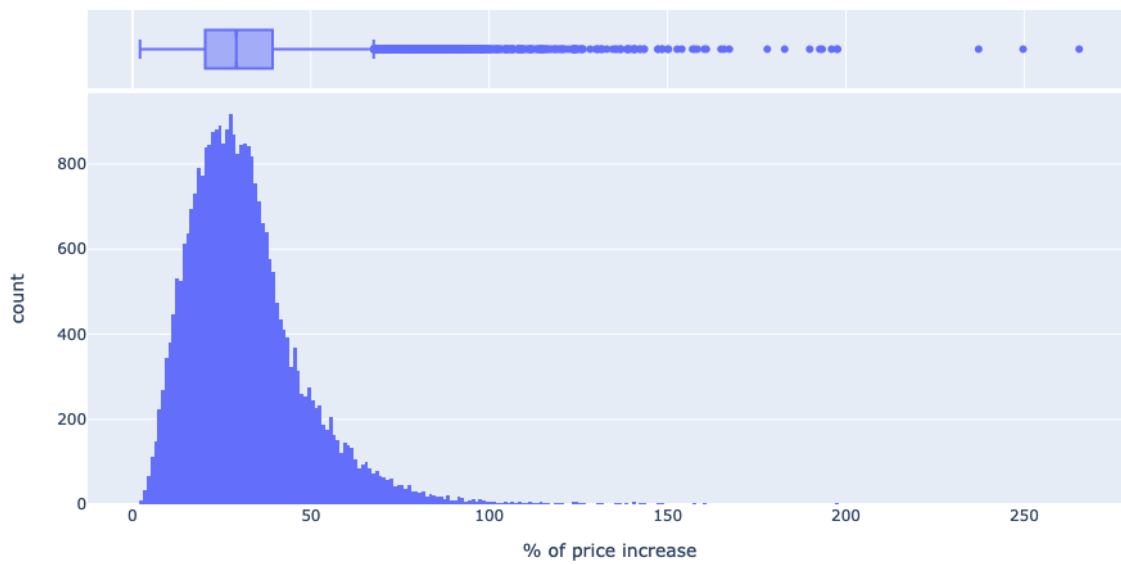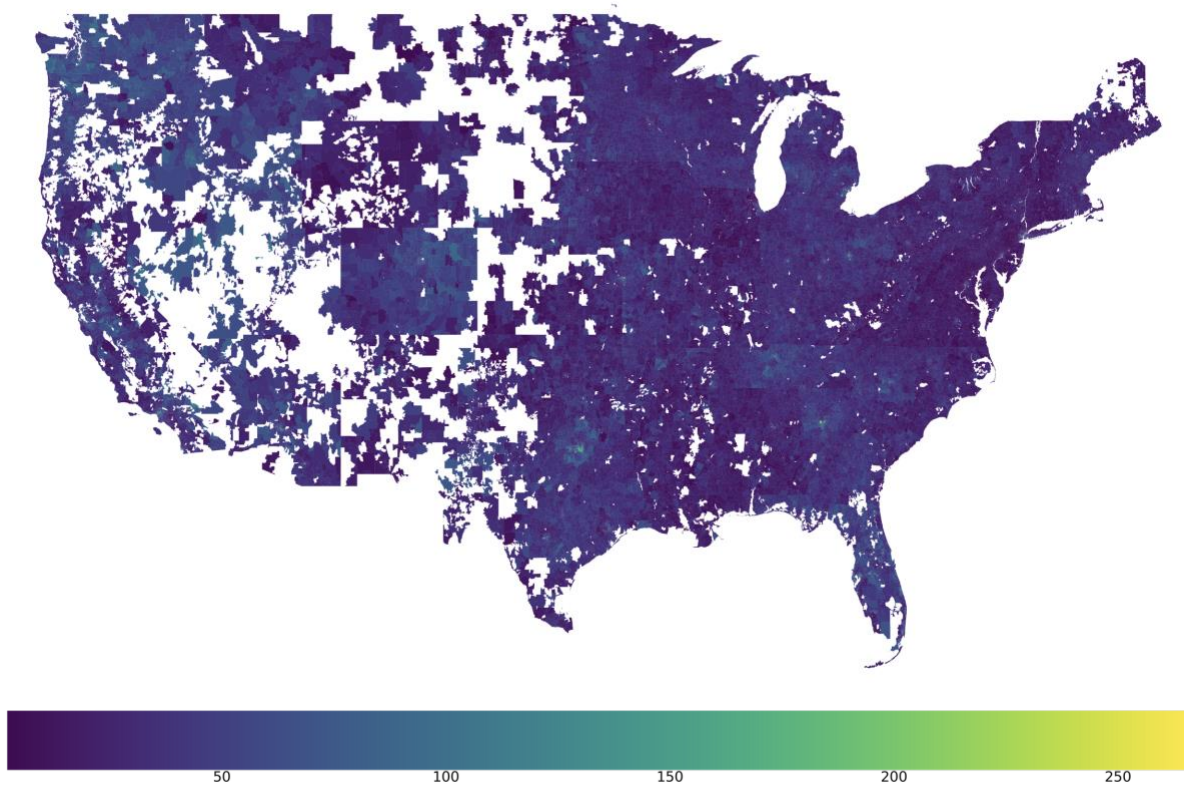

FIGURE 4, PRICE CHANGE FROM 2016

FIGURE 5, PRICE CHANGE FROM 2016

2.4 MODELS

2.4.1 MODEL TRAINING AND TESTING STRATEGY

We will hold data for 2020 and beyond for our time series modeling to test our model accuracy. And we will train a model individually for each zip code since the data we had is zip code specific. So, we will train a total of 30464 models.

And for our time series, we will be using the prophet library. The prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. (Facebook Prophet, n.d.)

2.4.2 TIMESERIES MODELING USING PROPHET LIBRARY

We selected a random zip and trained our initial model, which gave us the predictions shown in   Figure 6. This model didn't give us a good forecast since the model does not see the COVID19 trend.
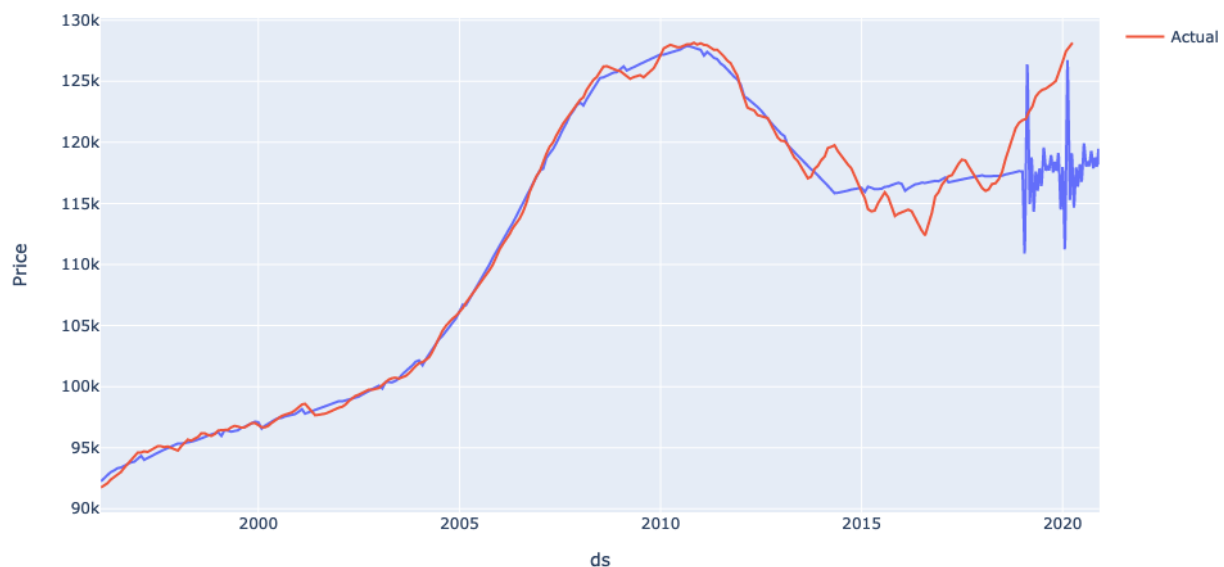

FIGURE 6, TRAINING DATA UNTIL 2019

So, we will include all of our data set in our training, which resulted in better forecasting, as shown in Figure 7
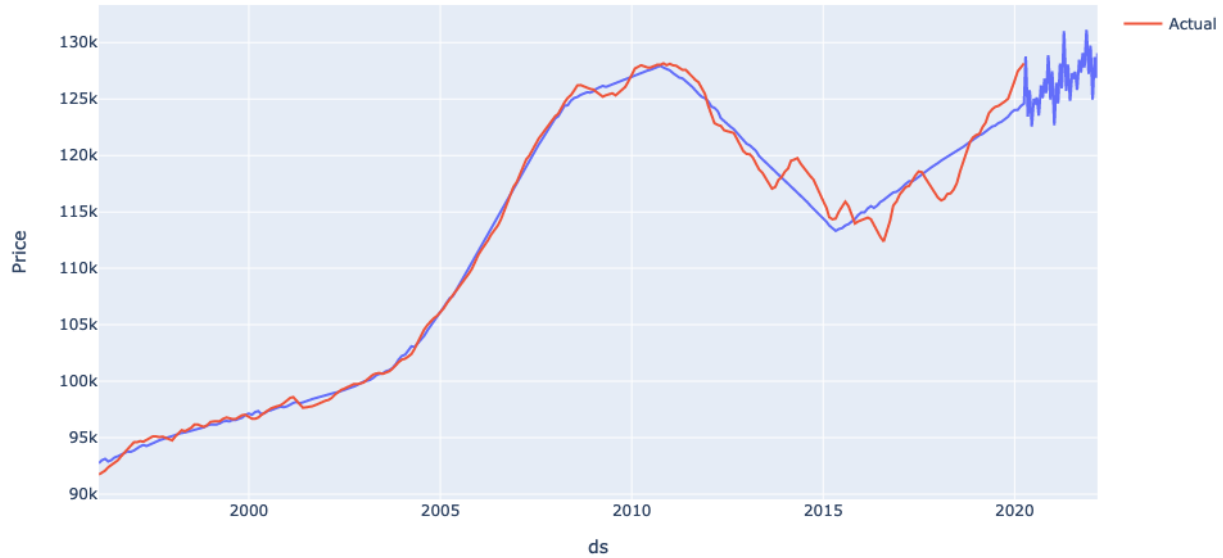
2.4.3 TIMESERIES MODELING OPTIMIZATION

Our model is amplifying the seasonality, not in our dataset.
Figure 8, shows the cross-validation of our model. The prediction error is at 1% for a month and 3% for a year out.

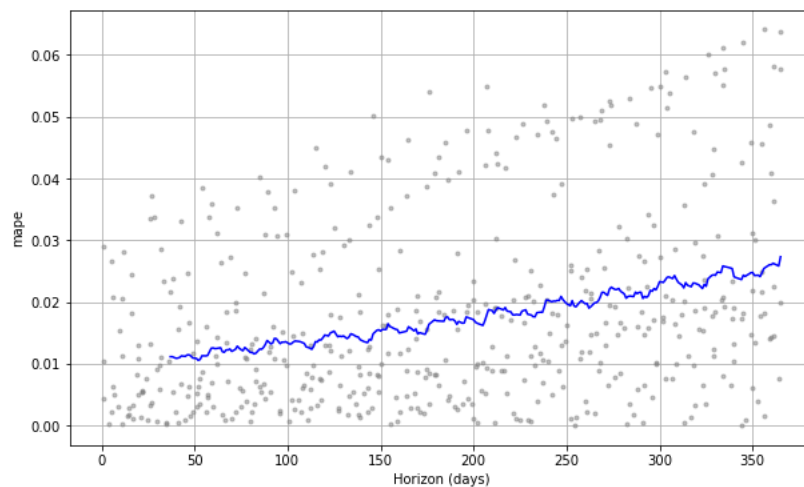FIGURE 8

To tune the model, we will set up a grid search to tune the hyperparameters changepoint_prior_scale and seasonality_prior_scale. The optimal parameters we arrived at are changepoint_prior_scale: 0.1, seasonality_prior_scale: 0.01. Figure 9, shows the final optimized model, which is a slight improvement over our previous model.
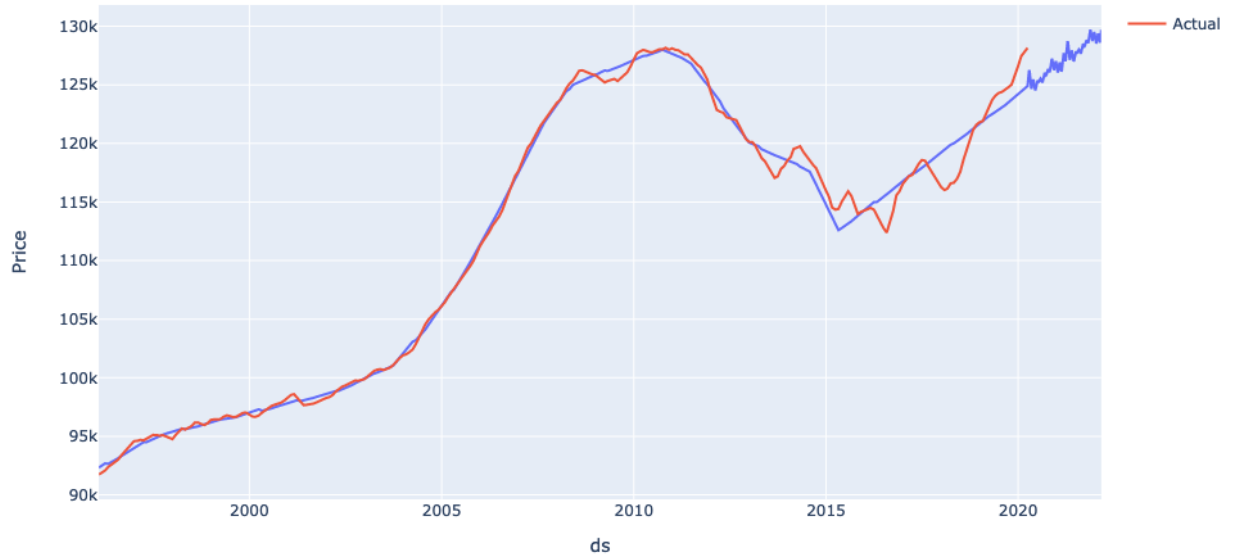
2.4.4 FINAL MODELS

It's time we set up a batch job to build the models for each of our zip codes. We used the python threading library to parallelize the model building on a 16 CPU/60Gb server. It took 4 hours to build all our models. The models saved on disk took around 60GB of space.

3 RESULTS

3.1.1 FORECAST COMPARISON

We compared the percent change in price from March 2020 to March 2021for the zip codes and tabulated the top and bottom 10 locations as shown in Table 1. If we must choose the top 3 zips to invest in, we can simply pick the top 3 locations from the table.

Given Syracuse is a Non-Profit Organization, what would be our best interest to invest in. Should we target college towns to improve student housing? Table 2, summarized the top 10 college towns by their percentage of price change in a year.

StateName	State	City	Metro	CountyName	zip	Price Difference	Percentage
AL	AL	Franklin	NaN	Monroe County	36444	15694.84724	0.300248
AR	AR	Sweet Home	Little Rock-North Little Rock-Conway	Pulaski County	72053	9992.805815	0.26008
TX	TX	Hargill	McAllen-Edinburg-Mission	Hidalgo County	78549	16444.48753	0.257017
MO	MO	Fremont	NaN	Carter County	63941	25015.47144	0.247186
ND	ND	Hunter	Fargo	Cass County	58048	33710.68844	0.243262
CO	CO	Yoder	Colorado Springs	El Paso County	80864	66840.86089	0.238368
OH	OH	Youngstown	Youngstown-Warren-Boardman	Mahoning County	44510	4672.135507	0.237949
PA	PA	Freeport Tov	NaN	Greene County	15352	17342.96338	0.23778
TX	TX	Pyote	NaN	Ward County	79777	13265.3387	0.231228
NC	NC	Dunn	Dunn	Harnett County	28547	39440.30107	0.22791
NY	NY	New York	New York-Newark-Jersey City	New York County	10282	-260509.5484	-0.212584
WV	WV	Auburn	NaN	Ritchie County	26325	-10374.35805	-0.213787
UT	UT	Duck Creek V	NaN	Kane County	6350	-40329.40253	-0.216601
AR	AR	Beedeville	NaN	Jackson County	72014	-6986.074939	-0.220478
VA	VA	Roanoke	Roanoke	Roanoke City	24011	-20441.10724	-0.222587
PA	PA	Toboyne Tov	Harrisburg-Carlisle	Perry County	17071	-10499.16408	-0.234749
TX	TX	Channing	NaN	Hartley County	79018	-90292.93182	-0.241894
LA	LA	Leesville	Fort Polk South	Vernon Parish	71474	-30033.10182	-0.275475
IN	IN	Gary	Chicago-Naperville-Elgin	Lake County	46406	-11666.54671	-0.278657
MI	MI	Detroit	Detroit-Warren-Dearborn	Wayne County	48207	-25414.65227	-0.280061

TABLE 1, TOP 10 AND BOTTOM TEN ZIP'S BY PRICE CHANGE

State	City	Metro	CountyName	zip	Price Difference	Percentage
AL	Franklin	NaN	Monroe County	36444	15694.84724	0.300248
MO	Fremont	NaN	Carter County	63941	25015.47144	0.247186
OH	Youngstown	Youngstown-Warren-Boardman	Mahoning County	44510	4672.135507	0.237949
MI	Detroit	Detroit-Warren-Dearborn	Wayne County	48217	6701.8025	0.206824
LA	Baton Rouge	Baton Rouge	East Baton Rouge Parish	70812	16939.06729	0.202258
AR	Evansville	Fayetteville-Springdale-Rogers	Washington County	72729	28062.94325	0.202025
AL	Birmingham	Birmingham-Hoover	Jefferson County	35221	5334.578941	0.185607
GA	Atlanta	Atlanta-Sandy Springs-Roswell	Fulton County	30314	32466.35781	0.179406
MI	Detroit	Detroit-Warren-Dearborn	Wayne County	48238	4963.255122	0.174736
IA	Davenport	Davenport-Moline-Rock Island	Scott County	52801	13486.7574	0.171208

TABLE 2, TOP 10 COLLEGE TOWNS

4 CONCLUSIONS

We explored the time-series housing data to predict our best investment opportunities, below are our recommendations

4.1.1 WHAT TECHNIQUE/ALGORITHM/DECISION PROCESS DID YOU USE TO DOWNSAMPLE?

We did not down-sample the data; we increased the compute power to process all data.

We recommend the below two sets of options based purely on their value appreciation in a year. If we choose one, Franklin, Alabama, will be our top choice; it grew faster and is also a college town.

Top Overall		Top College Town	
AL	Franklin	AL	Franklin
MO	Fremont	AR	Sweet Home
OH	Youngstown	TX	Hargill

4.2 APPENDIX – BONUS

DEVELOP A GEOGRAPHIC VISUALIZATION THAT IN YOUR VIEW BEST DEPICTS THE DATA AND RECOMMENDATION

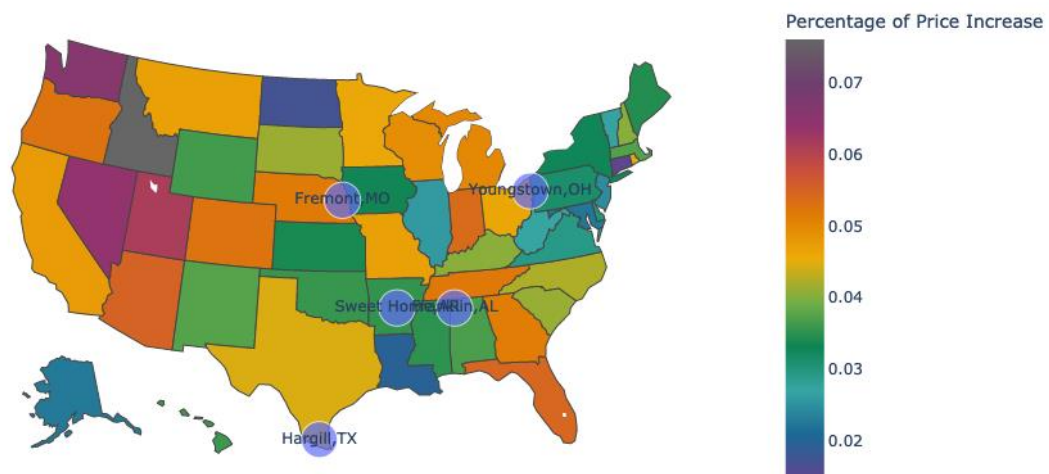Figure 10 shows the average price increase we predicted by state and the top recommendations.

FIGURE 10, PRICE PREDICTED FOR 2021