Predicting US Vacancy Rate by Zip Code

Capstone 2 Final Presentation

To dos

-

What is the goal? Who cares?

- The questions you asked
- The trends you investigated
- The resulting visualizations and conclusions

_____

- What questions are you answering?
- Who is your audience and why should they care about the information you're sharing?
- What are your major insights?
- What change do you want to bring about?

Also note the **key steps** to crafting any story itself:

- Introduce interesting characters
- Put them in a predicament
- Resolve the predicament
- Leave room for sequels

The lecturer also gives an incredibly useful piece of advice — useful regardless of the type of audience you have: *make the audience aware there's something important they didn't know they didn't know.*

Problem Identification Overview

What is the current vacancy rate in a certain zip code?

- Lack of specific AND current vacancy rate data
- Investors currently rely on “local” knowledge of an area
- What data do we have? What do we want?

Data Preprocessing Steps of Note

- AHS data - calculated new variable; vacancy rate
- Zillow Rent/Home Price Data - dropped rows with NaNs (each column had less than .2%)
- Time Series train/test split (5 splits)
- Created dummy variables for Categorical location variables

Model Description

Model Metrics -
`RandomForestRegressor()`

`Target Variable = 'VacancyRate%'`

final model features	parameters	hyperparameters

• **Zipcode**
• **RentPrice**
• **Year**
• **SizeRank**
• **HomePrice**
• **Dummy Variables**
  ◦ **State**
  ◦ **City**
  ◦ **Metro**
  ◦ **CountyName**
	*(n_estimators=100, *, criterion='mse', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0 , max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, ccp_alpha=0.0, max_samples=None)*	n/a

Model Performance

performance metrics

5 fold CV scores (training data):
Mean - 0.94, Std - 0.02

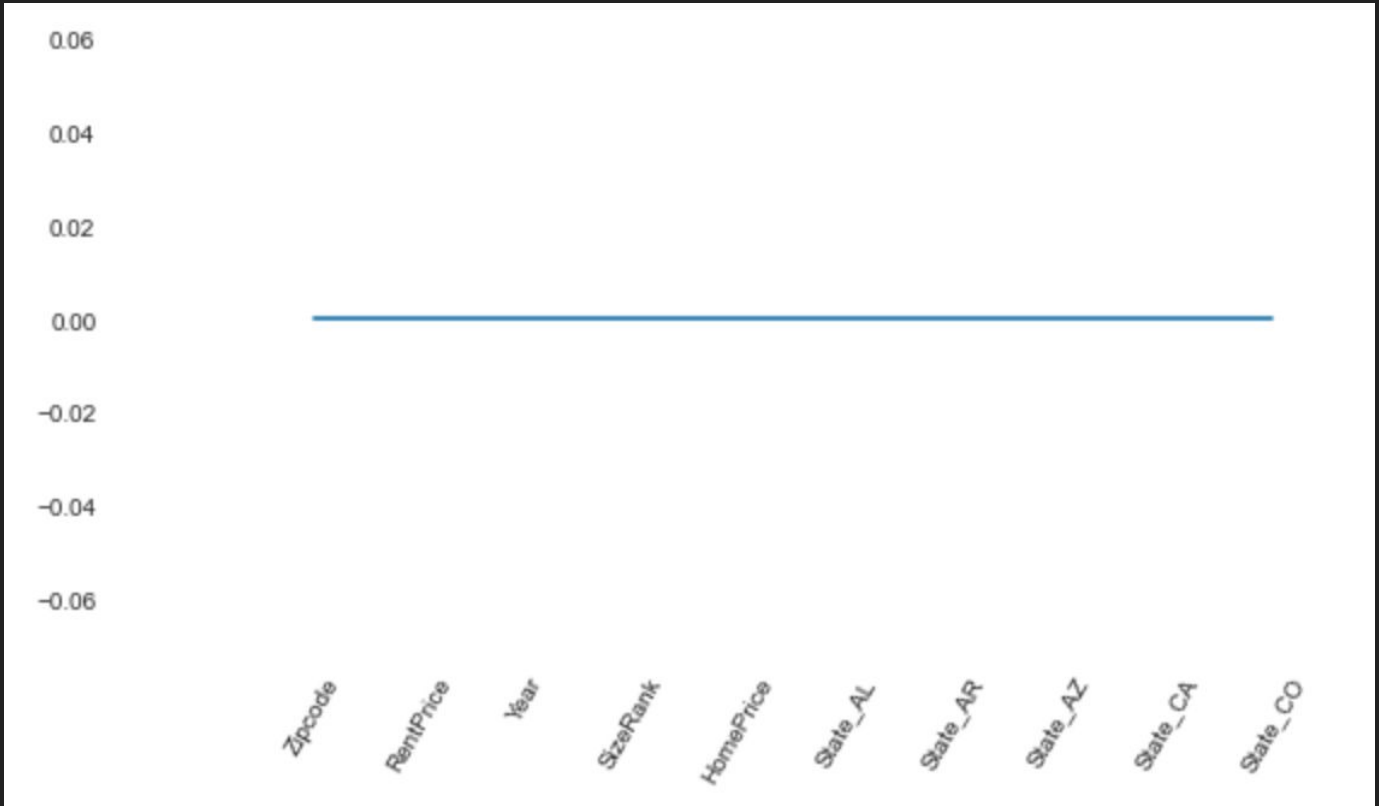
MAE (test set)
1.46, compared to DummyRegressor MAE of 5.18

R2 (test set)
0.92

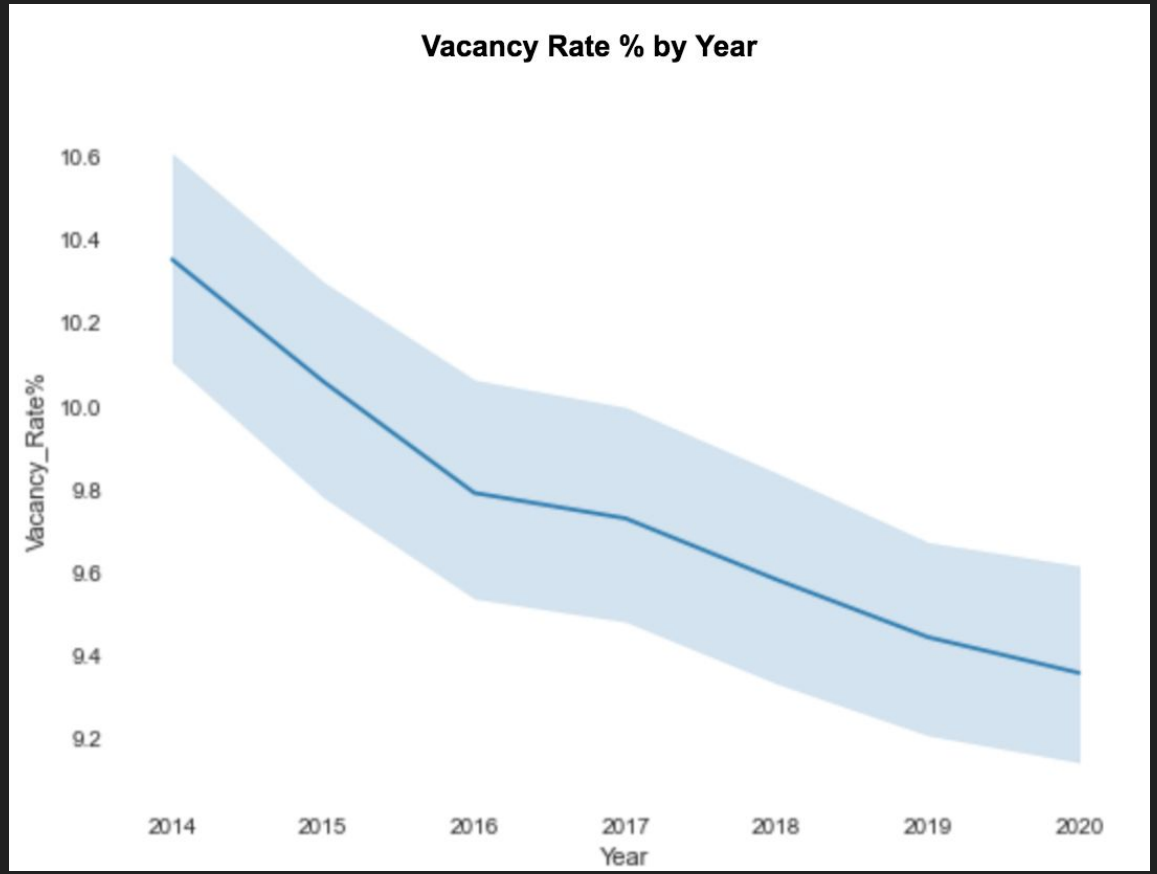
Adjusted R2 (test set)
0.78

Model Findings

Feature Importance



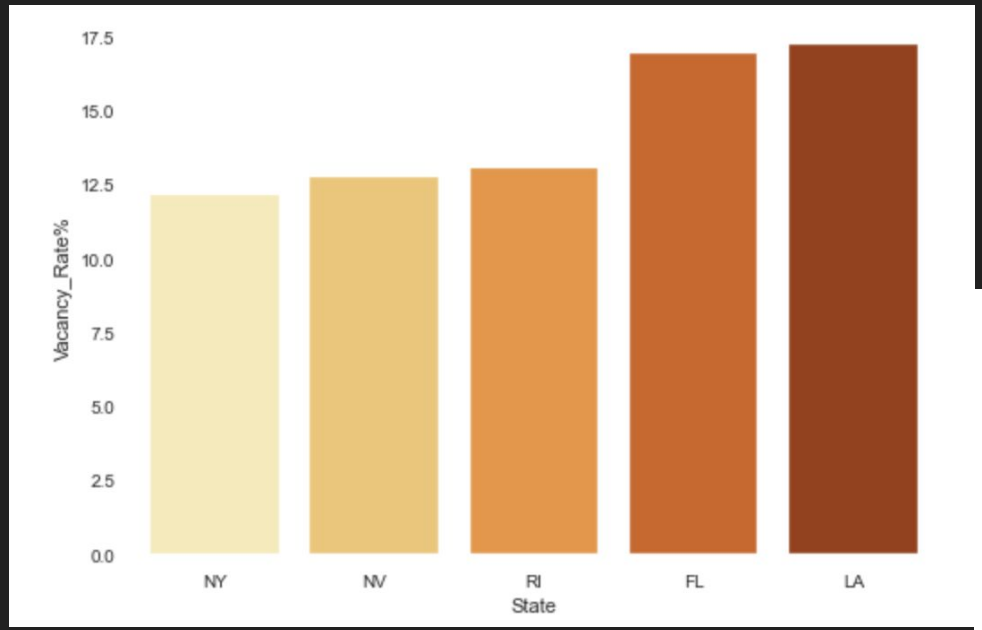
Model Findings

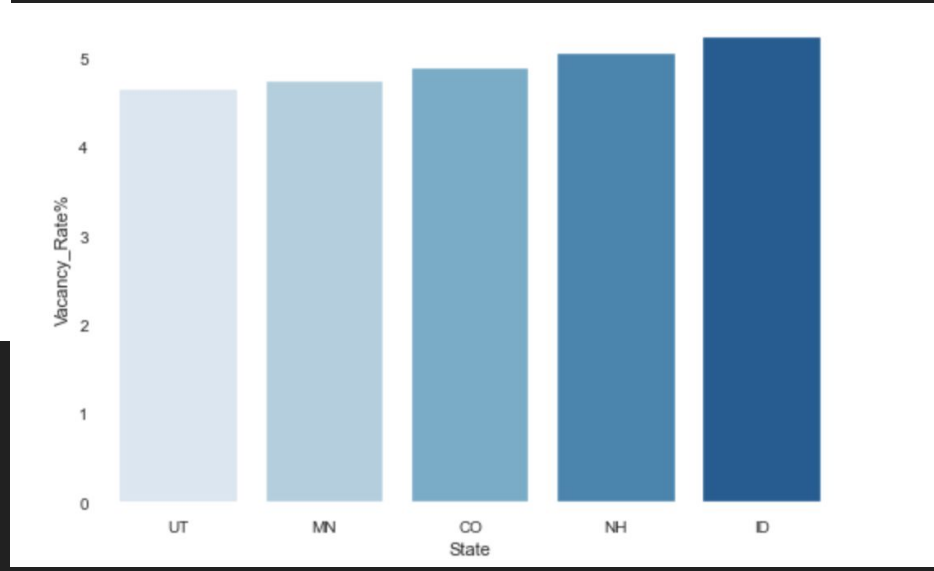


Model 2020 Predictions

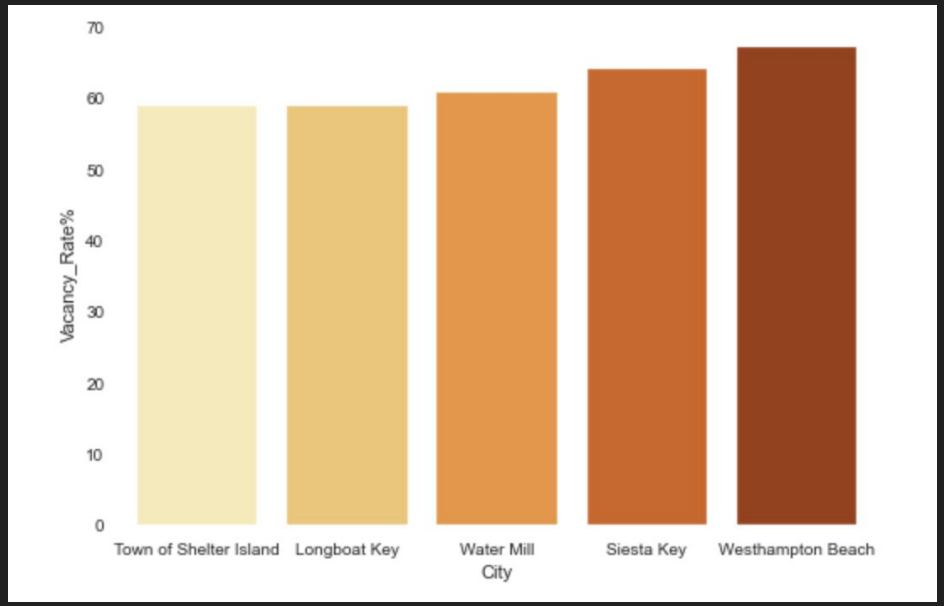
Vacancy Rate

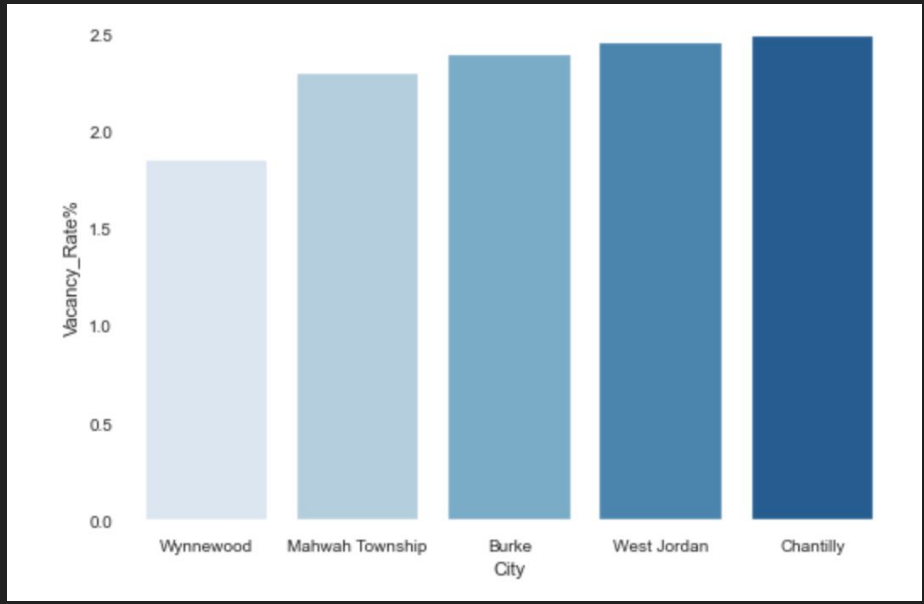
2020 Vacancy Rates - Highest/Lowest by State





2020 Vacancy Rates - Highest/Lowest by City





Model 2020 Predictions

Vacancy Adjusted Rent/Price Ratio

Rent/Price Ratio

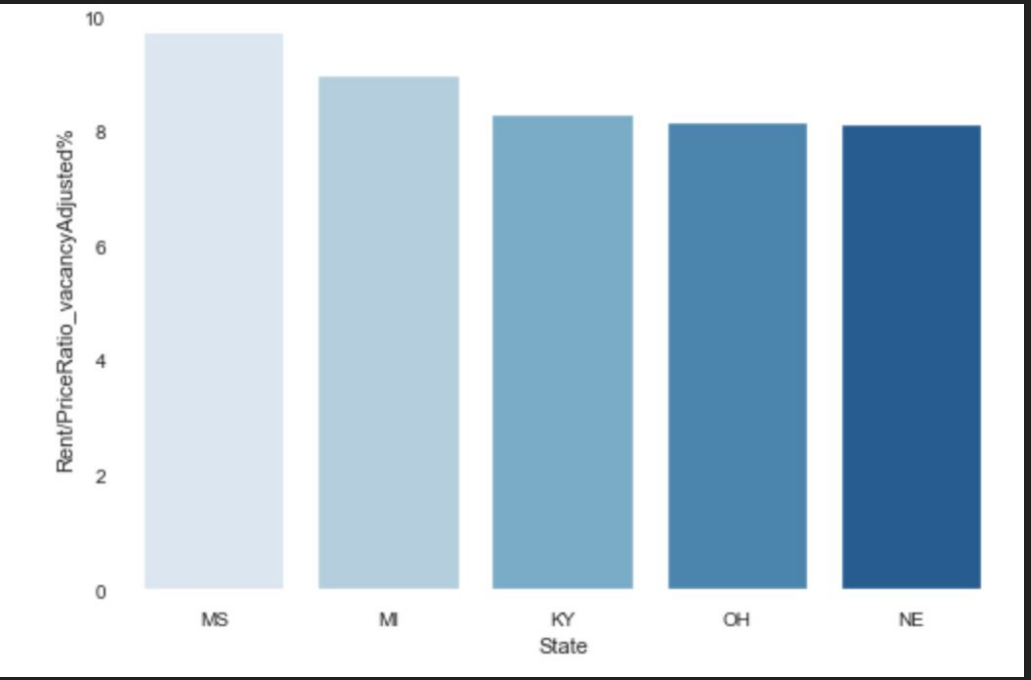
Monthly rent * 12 / home price

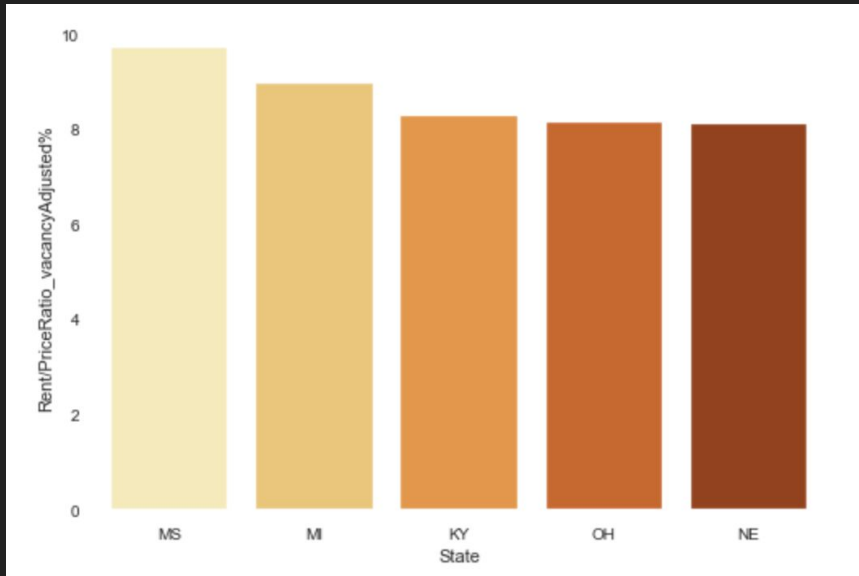
12%

Vacancy Adjusted..

((Monthly rent * 12) * (1 - vacancy rate)) / home price

States - Potential Rental Real Estate Investment





Places to potentially invest

States: Mississippi, Michigan, Kentucky, Ohio, Nebraska

Counties: Lucas County, OH; Luzerne County, PA; Wayne County, MI; Baltimore City, MD; DeSoto County, MS

Cities: Jennings, MO; Detroit, MI; Hampton Bays, NY; Northwoods, MO; Park Forest, IL

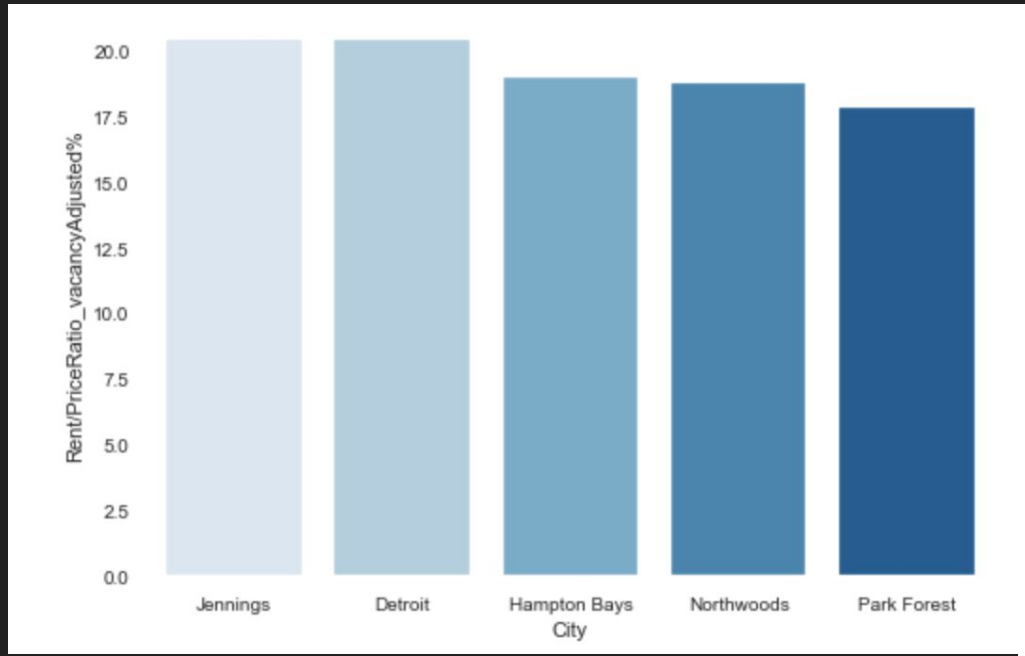
Places to you may to avoid when investing

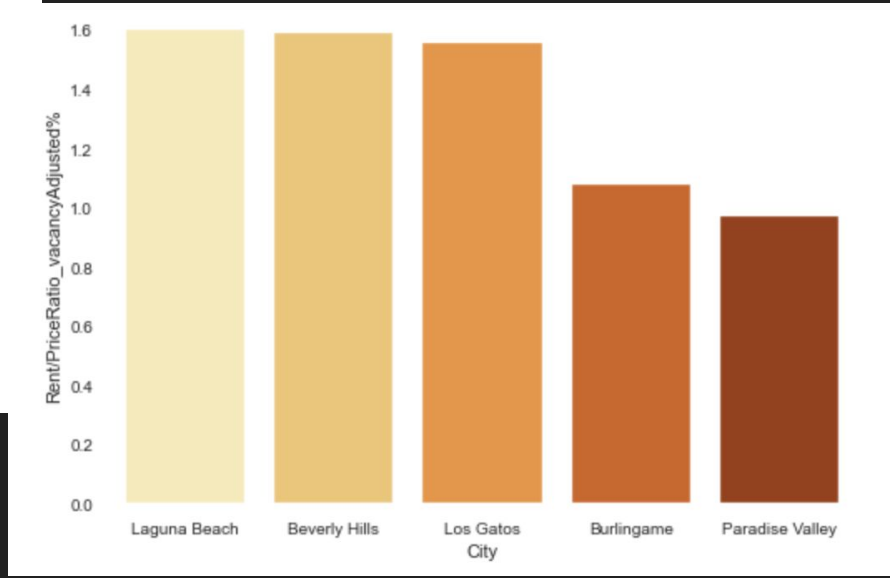
States: Hawaii; Washington, DC; California; Washington; Oregon

Counties: El Dorado County, CA; Santa Clara County, CA; Marin County, CA; San Francisco County, CA; San Mateo County, CA

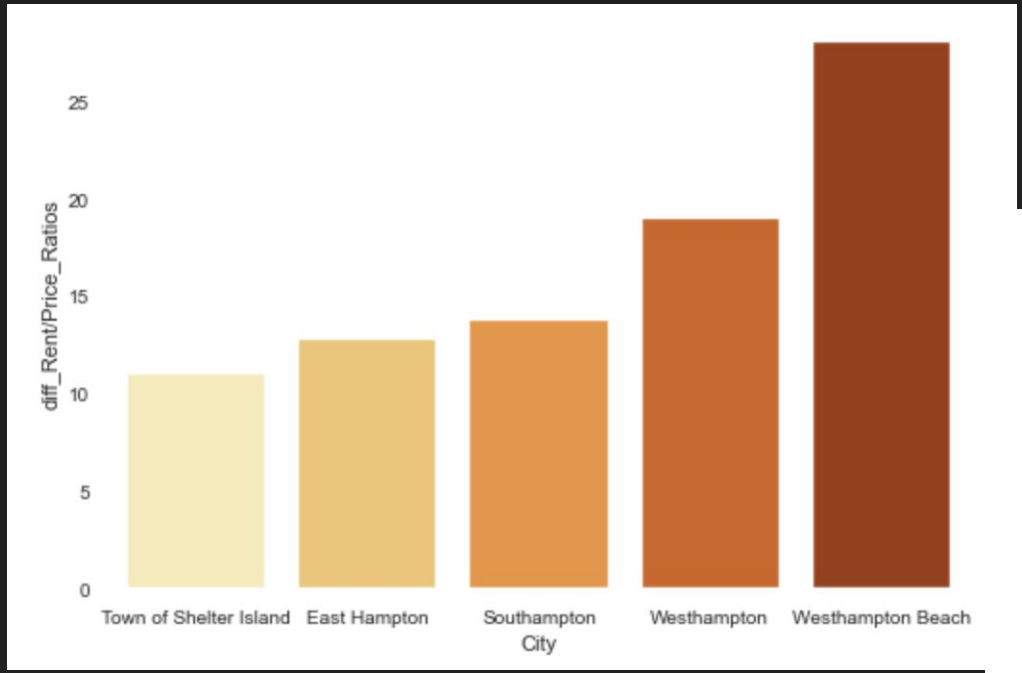
Cities: Cupertino, CA; Los Alamitos, CA; Redwood City, CA; Burlingame, CA, Wynnewood, PA

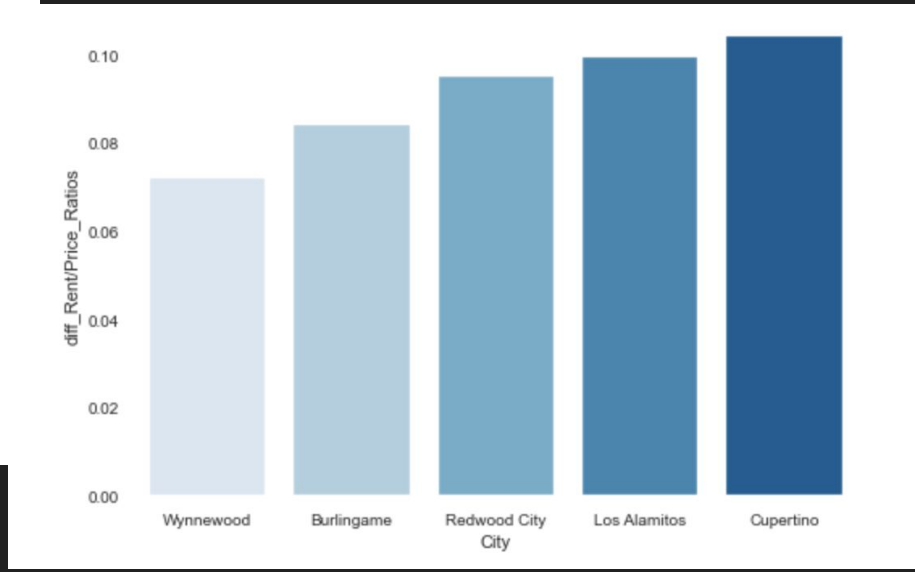
Cities - Potential Rental Real Estate Investment





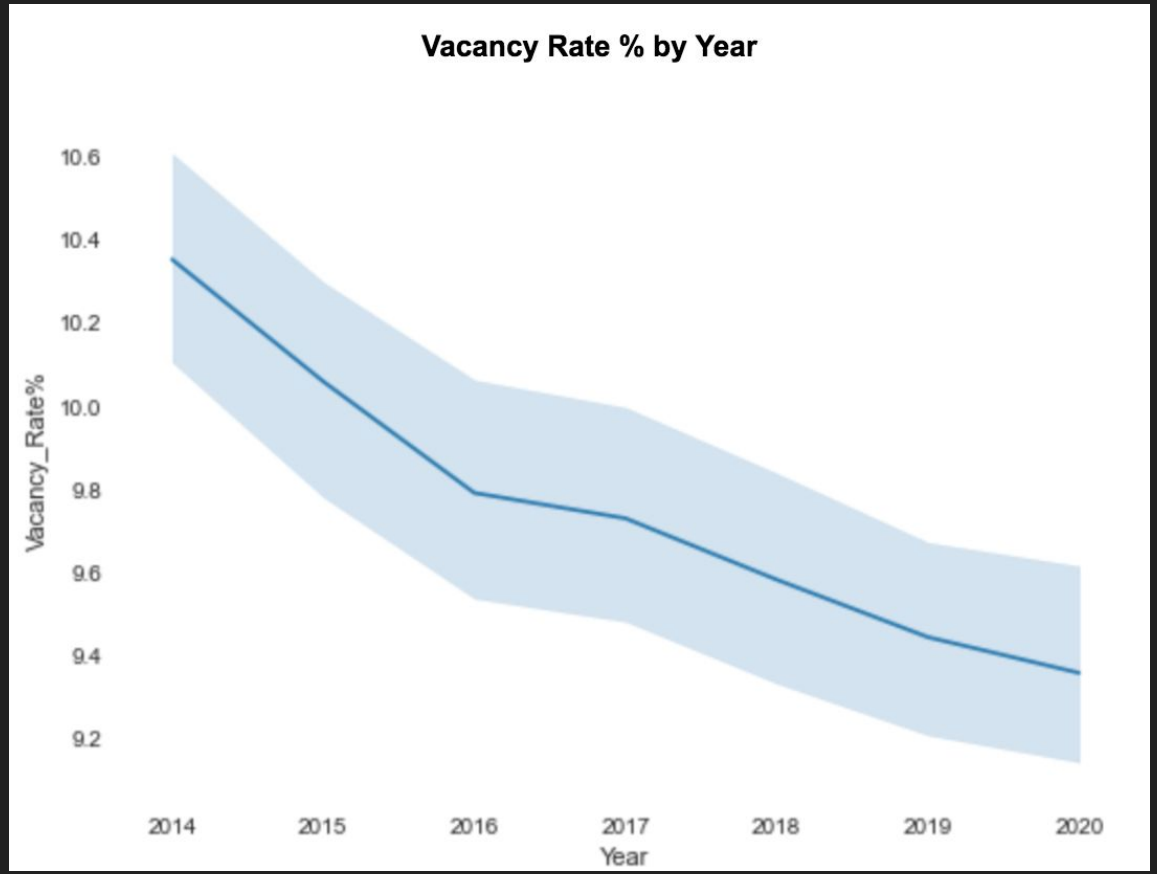
Cities where Rent/Price Ratios most/least impacted by vacancy



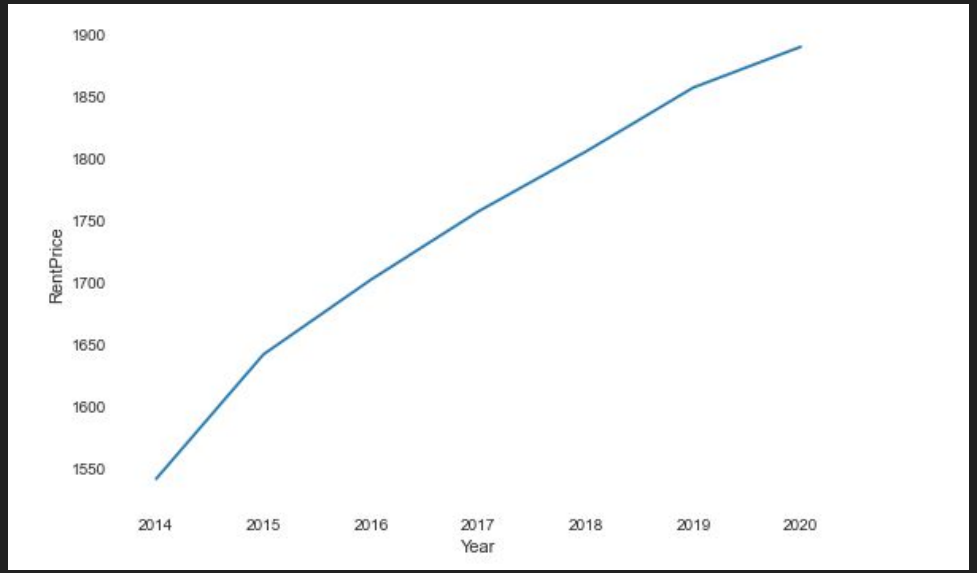


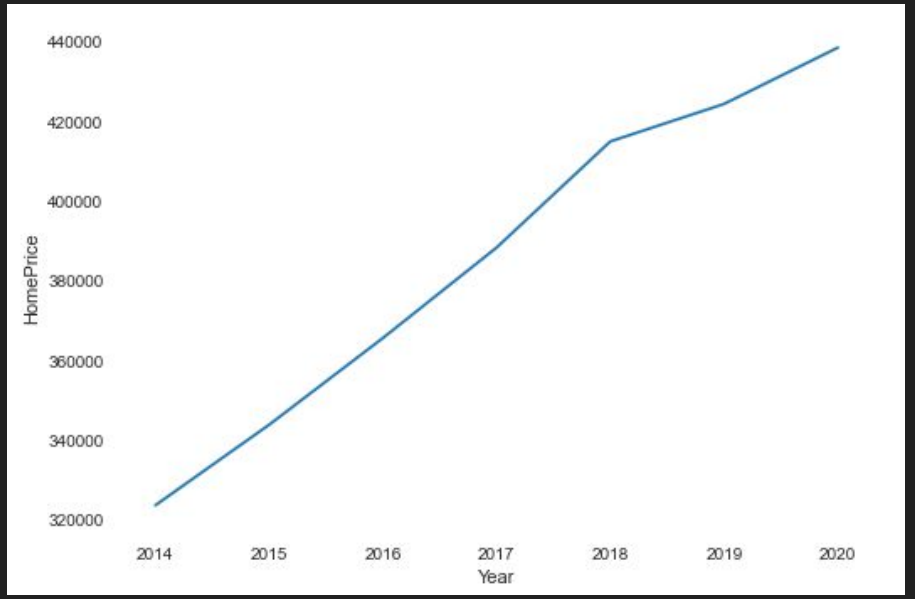
2014 - 2020 Trends

Model Findings

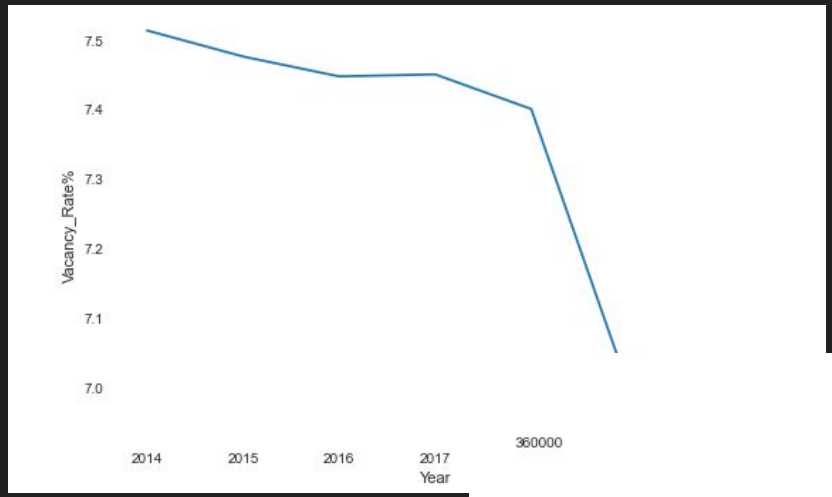


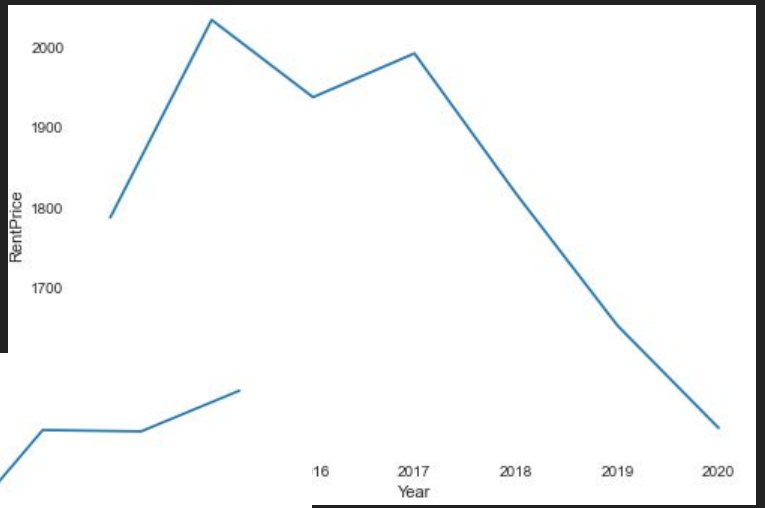
Avg. Rent Prices & Home Prices, Over Time

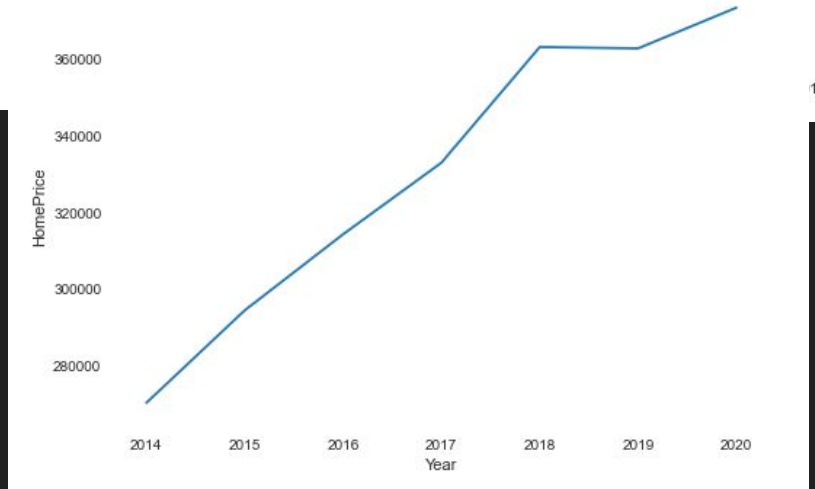




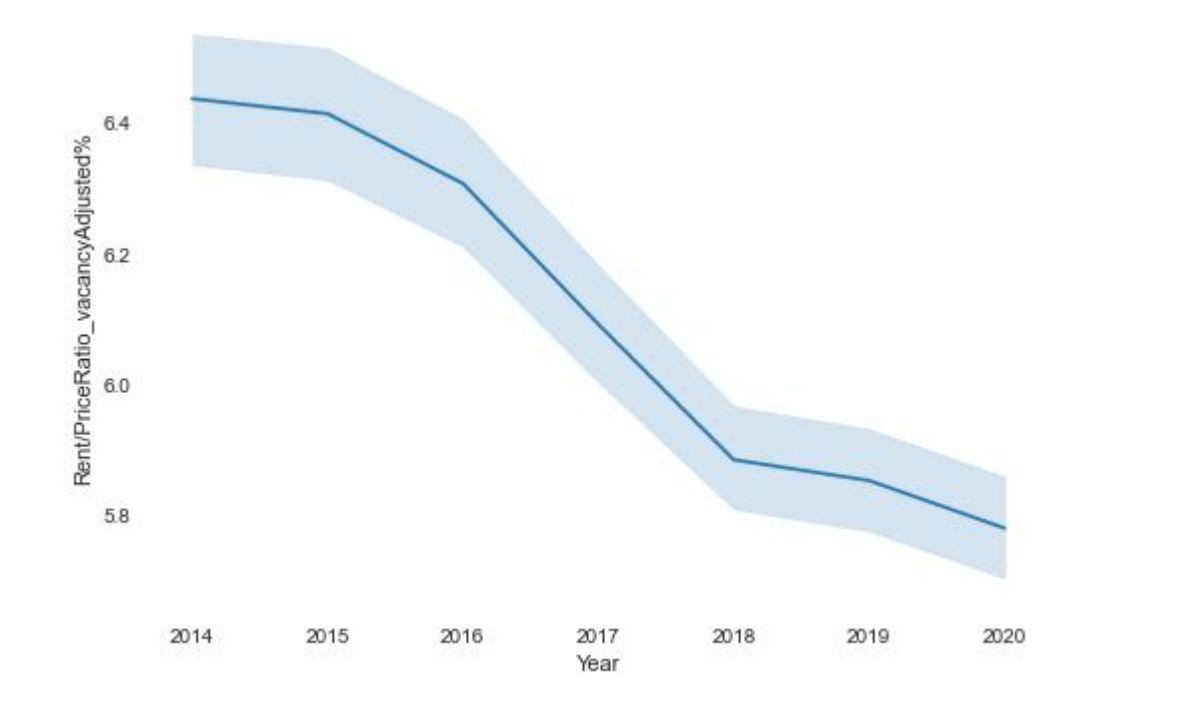
Std. Dev Over Time - Vacancy Rate, Rent & Home Prices







US Getting Harder for Real Estate Investing



Next Steps

- Get more zip codes represented in model
  - Get rental data by zip code (predict and/or use ACS)
  - Deal with NaNs differently
  - Drop rent prices from the model
- Try models with more/less variables
- Test more hyperparameters/models
- Build website

Future Ideas

- Create “investability” index
- Test predictions as real data comes out
- Create regional models


THANK
YOU