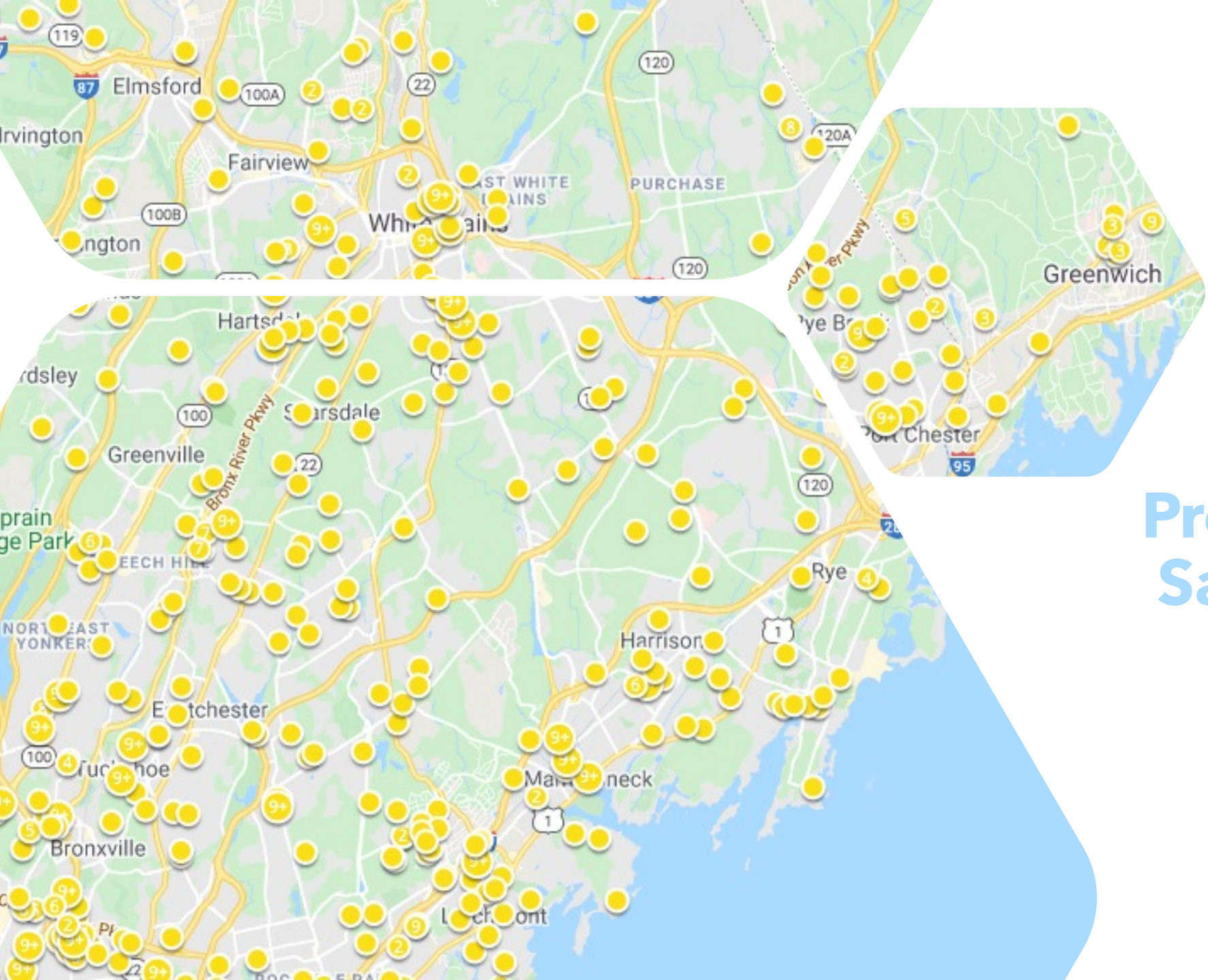


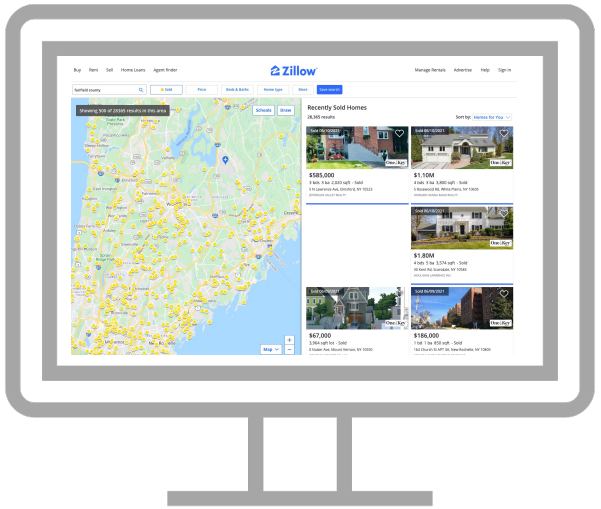
Predicting Home Sales in the NYC Suburbs

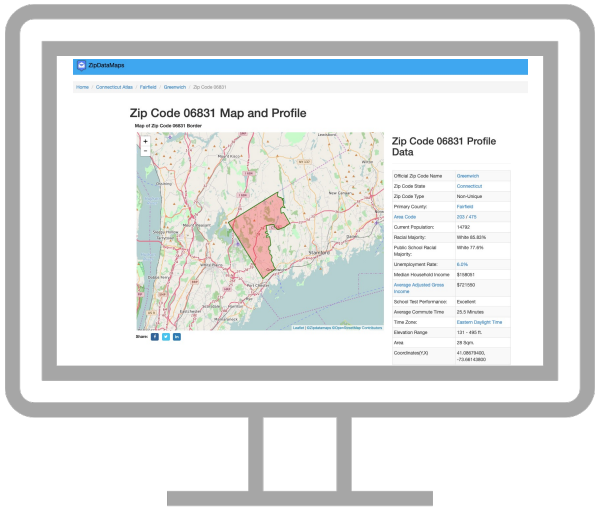
Jen Hilibrand

Why

- For most families, home equity is the largest component of household wealth
- US total real estate market is more than $35 Trillion dollars
- Post-pandemic desire for more space (demand for suburban versus urban homes)

Workflow





Zillow®

ZipDataMaps

BeautifulSoup
Selenium

- Price
- Time on market
- Square Feet
- Beds
- Baths
- Address/location
- Distance to midtown

BeautifulSoup

Per Zipcode

- Median household income
- Average gross household income
- Public school test performance
- Average commute time

Pandas, geopy, sklearn, matplotlib, seaborn

Data

Challenges:

- Zillow's advanced web scrapping detection
  - Added request headers to mask HTML scrape from search page
  - Focused on features from aggregate search page as opposed to individual home pages

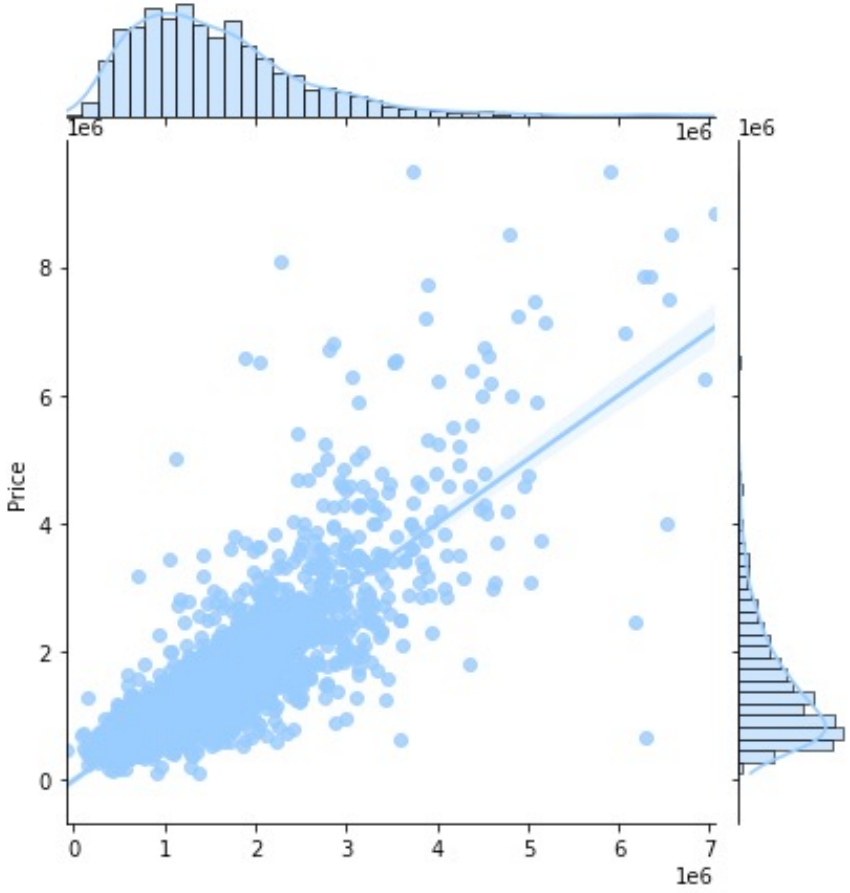
- Outliers
  - Dropped homes sold at greater than $10mm to reduce variance in model
  - Dropped towns with less than 10 homes

Data focus:

- Sold homes (as opposed to currently listed homes)
  - ~3.5k homes over the past 6 months across 10 towns

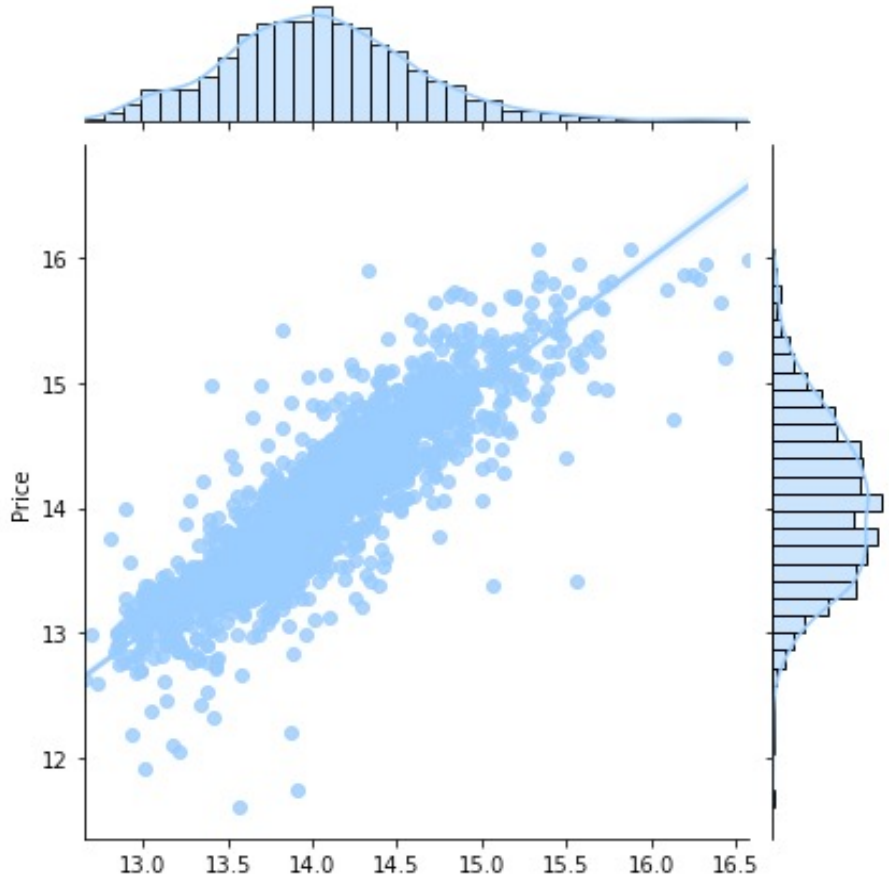
Step One: Basic OLS Regression

- Categorical features such as "Town" and "Zipcode" not taken into account
- Skewed distribution of prices
- Cross validation average score R^2 of .66



Step Two: Feature engineering and transformation

- Feature engineering:
  - Polynomial features: Beds^2
  - Multiplicative features: School*Space
  - Formulaic features:
    - Average of commute time and distance to midtown
    - Total sum of square feet, beds, and baths
- Dummy variable creation for categorical features:
  - Zipcode dummies
  - Town dummies
- Transformation of target
  - Log transform on Y (Price) to adjust for skew
- Cross validation average R^2 score of .75



Step 3: Regularized models

- Standardize numeric features
- Solve for optimal alphas for lasso and ridge regularization
- Observe largest absolute value coefficients:
  - Ridge
    - Beds (+0.249)
    - Baths (+0.191)
    - Sqft (+.103)
  - Lasso
    - Average Gross Income (+5.602)
    - Zipcode 10573 (-2.617)
    - Sqft (+0.264)

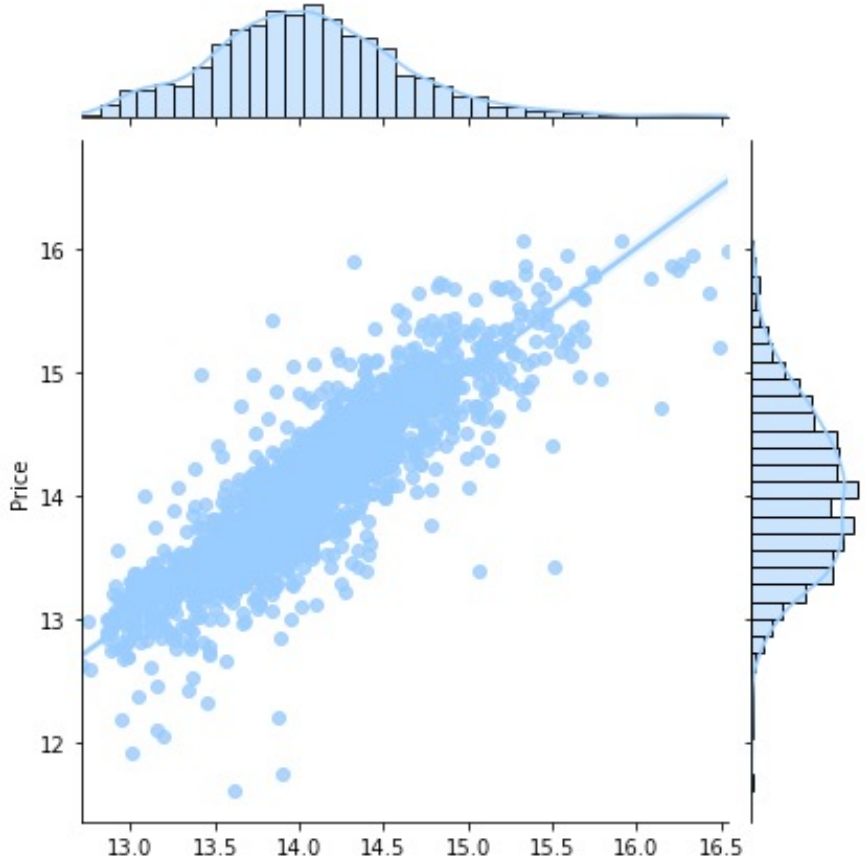
Step 3: Regularized models

- Standardize numeric features
- Solve for optimal alphas for lasso and ridge regularization
- Observe largest absolute value coefficients:
  - Ridge
    - Beds (+0.249)
    - Baths (+0.191)
    - Sqft (+.103)
  - Lasso
    - Average Gross Income (+5.602)
    - Zipcode 10573 (-2.617) ← **Taxes!**
    - Sqft (+0.264)

Lasso v Ridge training data

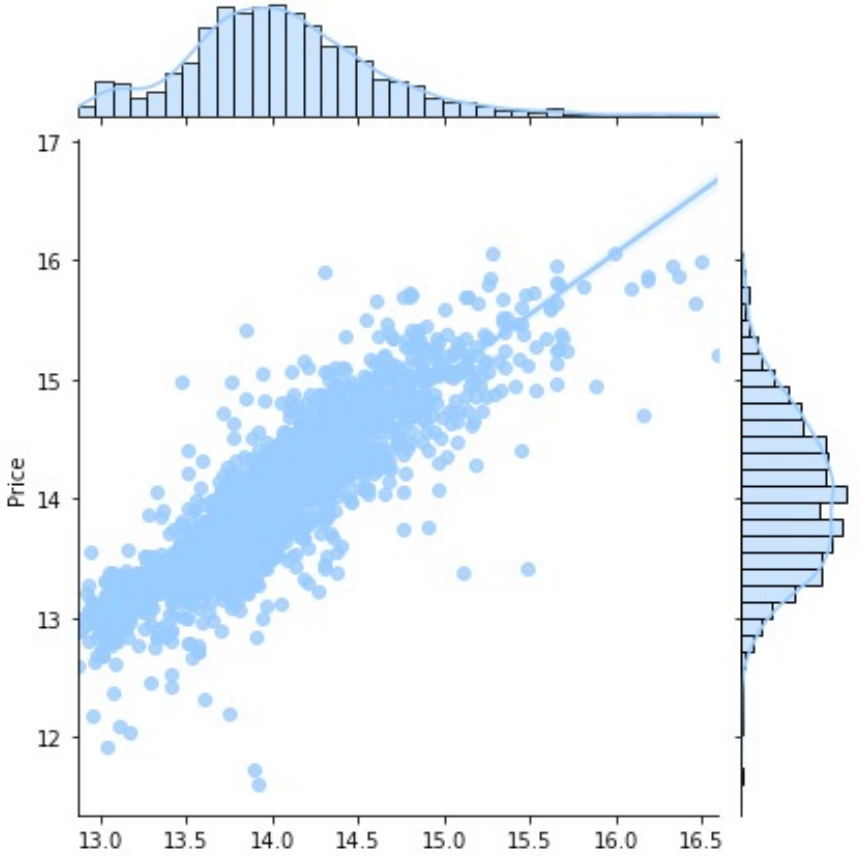
Ridge

CV average R^2 .754



Lasso

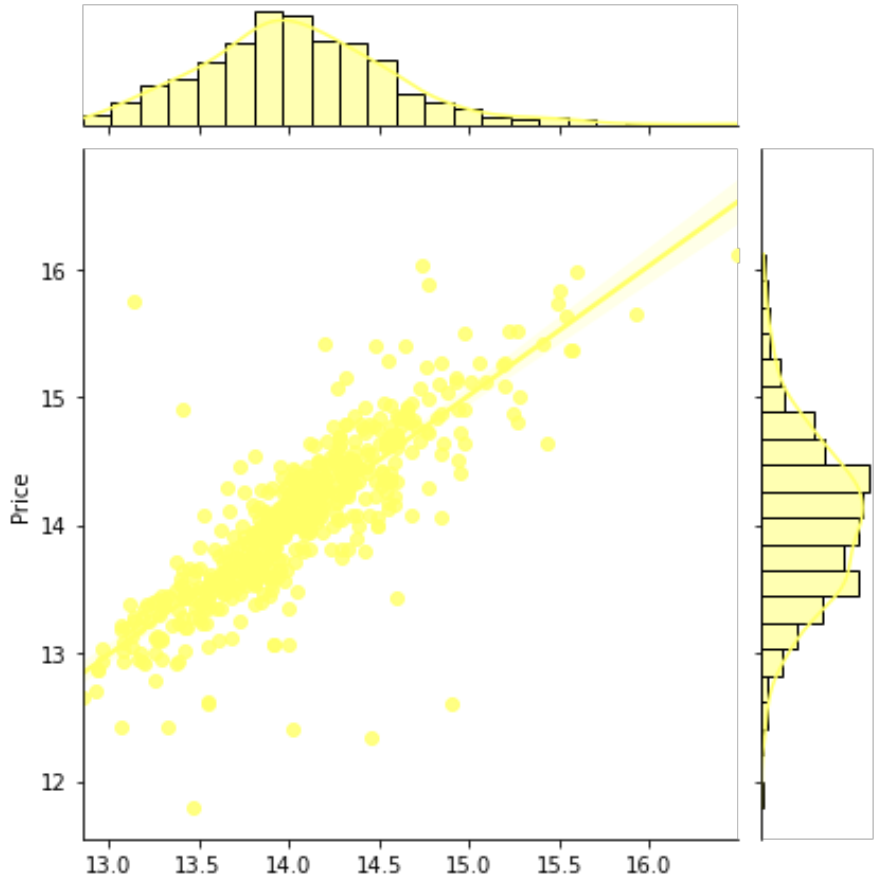
CV average R^2 .749



Lasso v Ridge test data

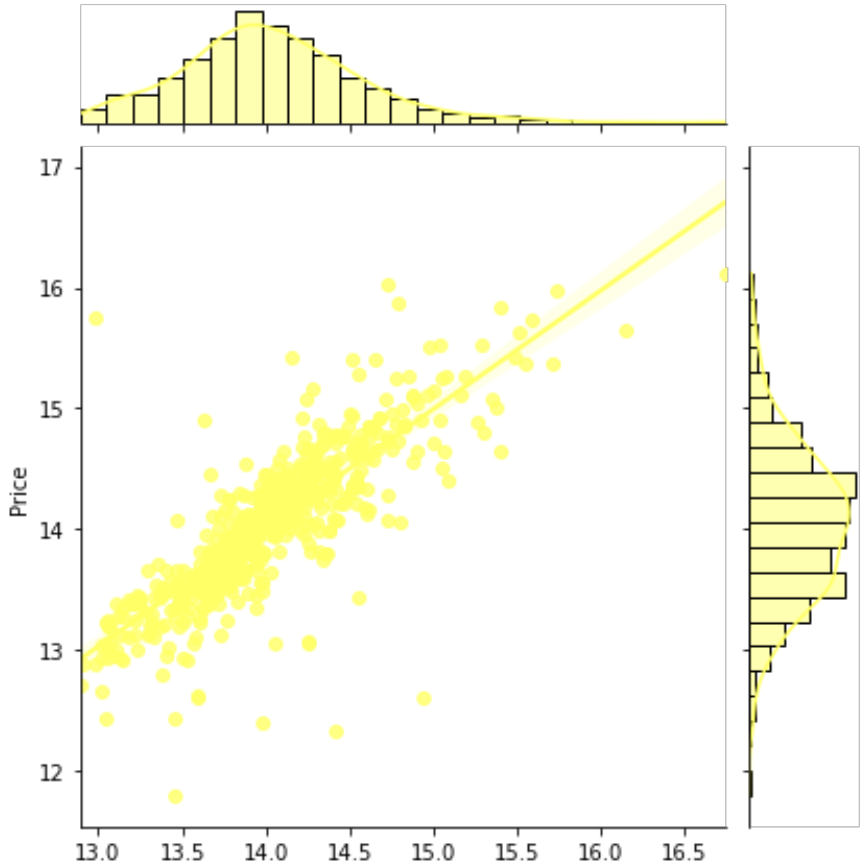
Ridge

R^2 of .685



Lasso

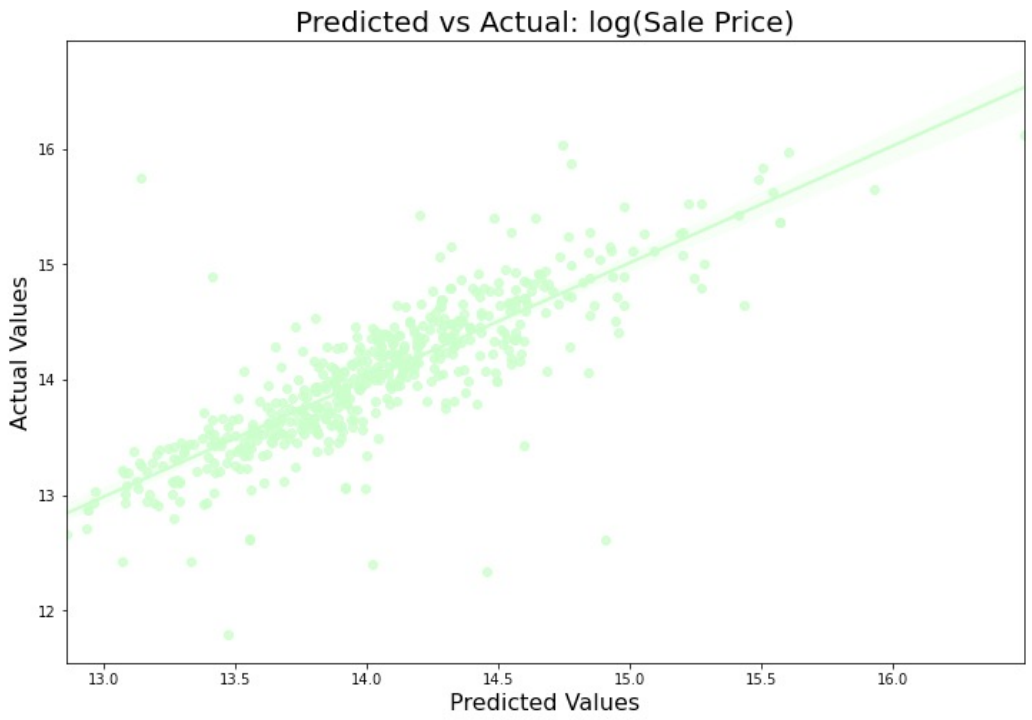
R^2 of .664



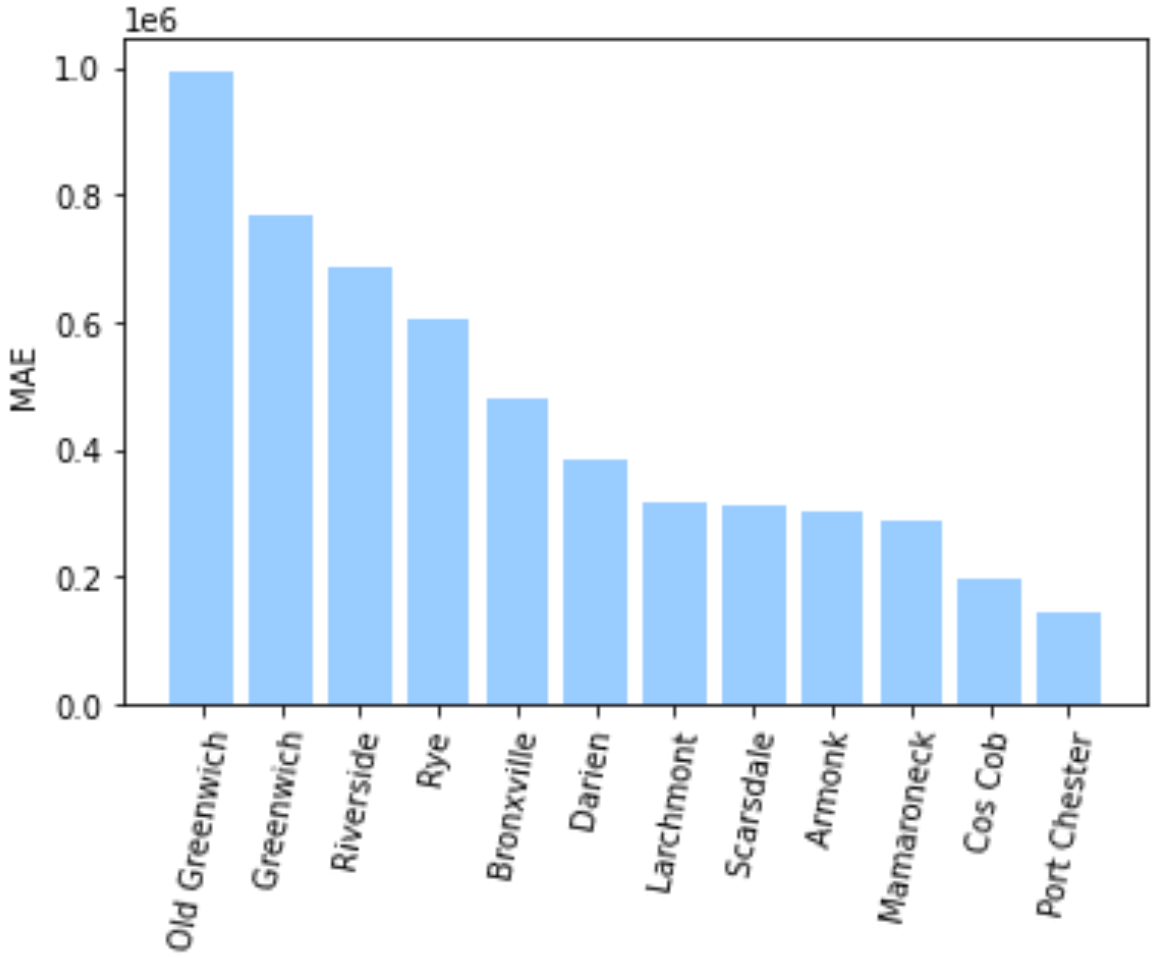
Are we overfitting…?

Ridge fit

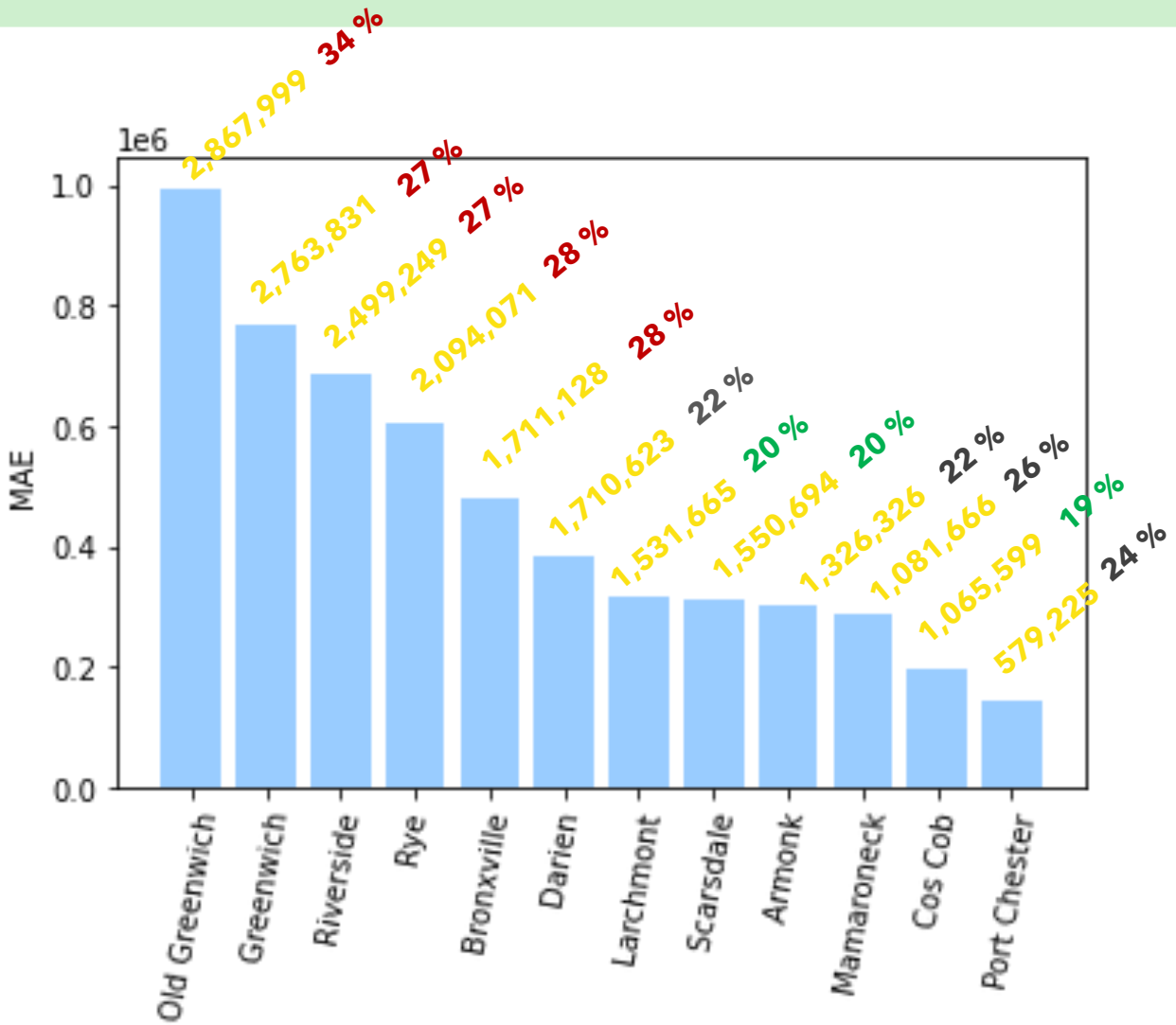
Ridge SSE: 68.87
Ridge RMSE: 7667,73.72
Ridge MAE: 392,364.06

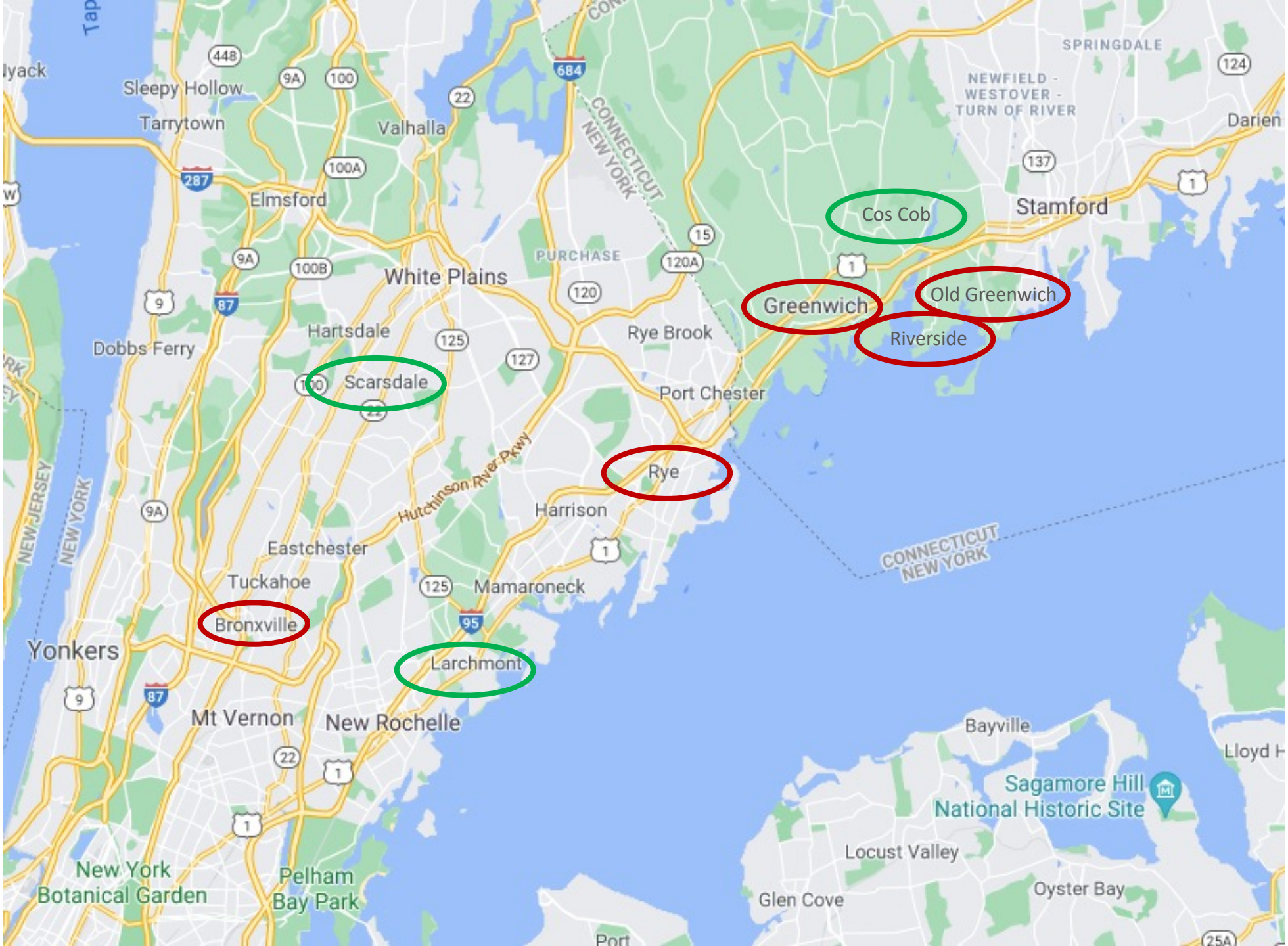


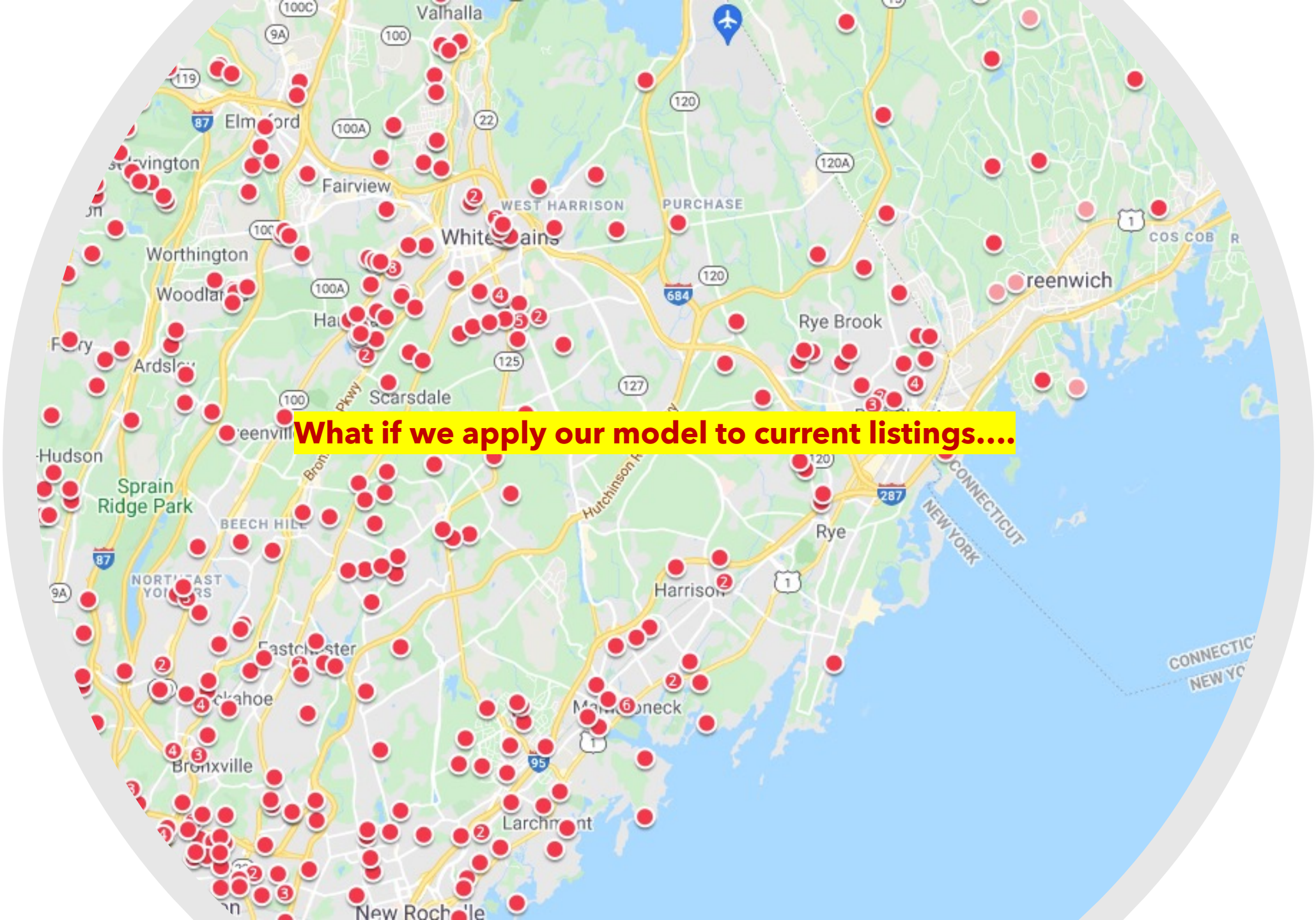
Mean Absolute Error: Town by Town Comparison



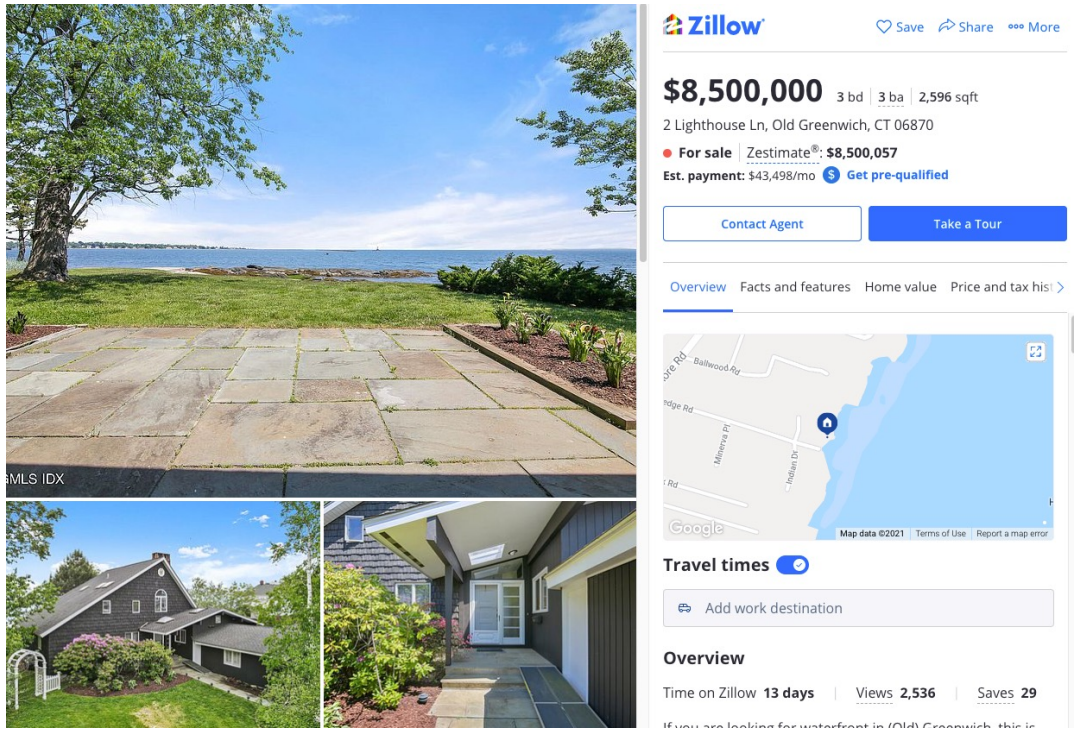
Mean Absolute Error: Town by Town Comparison




Cos Cob
Greenwich
Old Greenwich
Riverside
Scarsdale
Rye
Bronxville
Larchmont
Stamford
White Plains
Rye Brook
Port Chester
Harrison
Mamaroneck
New Rochelle
Mt Vernon
Yonkers
Eastchester
Tuckahoe
Hartsdale
Dobbs Ferry
Elmsford
Valhalla
Tarrytown
Sleepy Hollow
Darien
Bayville
Sagamore Hill
National Historic Site
Locust Valley
Oyster Bay
Glen Cove
New York
Botanical Garden
Pelham
Bay Park
CONNECTICUT
NEW YORK
PURCHASE
Hutchinson River Pkwy
NEWFIELD -
WESTOVER -
TURN OF RIVER
SPRINGDALE


What if we apply our model to current listings....
Valhalla
Fairview
WEST HARRISON
PURCHASE
Worthington
Rye Brook
Scarsdale
Hudson
Sprain Ridge Park
BEECH HILL
Rye
Harrison
Bronxville
Larchmont
New Rochelle
COS COB
CONNECTICUT
NEW YORK

Most “over-priced” homes



Zillow Save Share More

$8,500,000 3 bd | 3 ba | 2,596 sqft

2 Lighthouse Ln, Old Greenwich, CT 06870

For sale | Zestimate®: $8,500,057

Est. payment: $43,498/mo Get pre-qualified

Contact Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist

Travel times

Add work destination

Overview

Time on Zillow **13 days** | Views **2,536** | Saves **29**

Zillow Save Share More

$8,995,000 5 bd | 5 ba | 4,964 sqft

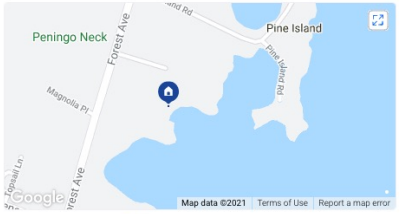
945 Forest Ave, Rye, NY 10580

For sale | Zestimate®: $8,955,961

Est. payment: $48,254/mo Get pre-qualified

Contact an Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist



Travel times

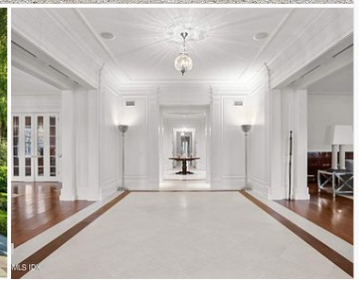
Add work destination

Overview

Time on Zillow **62 days** | Views **3,290** | Saves **35**

The “waterfront effect”

“Value” homes



Zillow Save Share More

$7,700,000 7 bd | 12 ba | 16,802 sqft

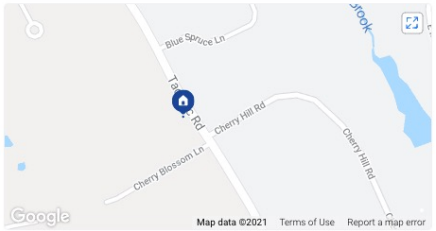
4 Cherry Blossom Ln, Greenwich, CT 06831

For sale | Zestimate®: **$7,700,015**

Est. payment: $39,087/mo Get pre-qualified

Contact Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist



Travel times

Add work destination

Overview

Time on Zillow **17 days** | Views **5,281** | Saves **290**

Majestic Richard Granoff designed stone and clapboard

Zillow Save Share More

$8,995,000 6 bd | 12 ba | 15,773 sqft

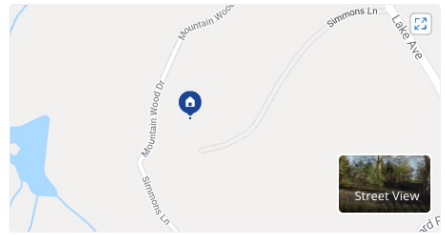
12 Mountain Wood Dr, Greenwich, CT 06830

For sale | Zestimate®: **$8,993,402**

Est. payment: $45,785/mo Get pre-qualified

Contact Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist



Travel times

Add work destination

Overview

Time on Zillow **76 days** | Views **1,403** | Saves **32**

“Value” homes

Zillow Save Share More

$7,700,000 7 bd | 12 ba | 16,802 sqft

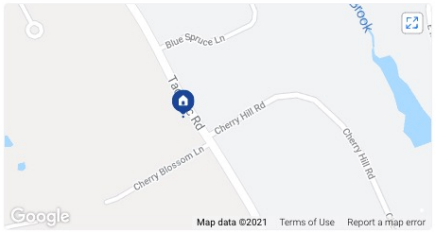
4 Cherry Blossom Ln, Greenwich, CT 06831

For sale | Zestimate®: **$7,700,015**

Est. payment: $39,087/mo Get pre-qualified

Contact Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist



Travel times

Add work destination

Overview

Time on Zillow **17 days** | Views **5,281** | Saves **290**

Majestic Richard Granoff designed stone and clapboard

Zillow Save Share More

$8,995,000 6 bd | 12 ba | 15,773 sqft

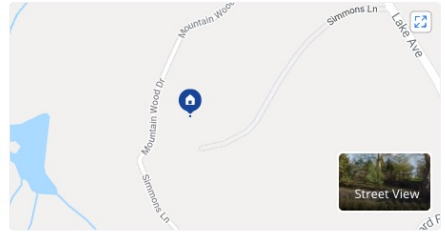
12 Mountain Wood Dr, Greenwich, CT 06830

For sale | Zestimate®: **$8,993,402**

Est. payment: $45,785/mo Get pre-qualified

Contact Agent | Take a Tour

Overview | Facts and features | Home value | Price and tax hist



Travel times

Add work destination

Overview

Time on Zillow **76 days** | Views **1,403** | Saves **32**

Diminishing marginal value of square footage + “backcountry effect”

Future Work

- Add more towns to my model
- Overlay tax conditions
- Build out selenium pipeline for more specific house data