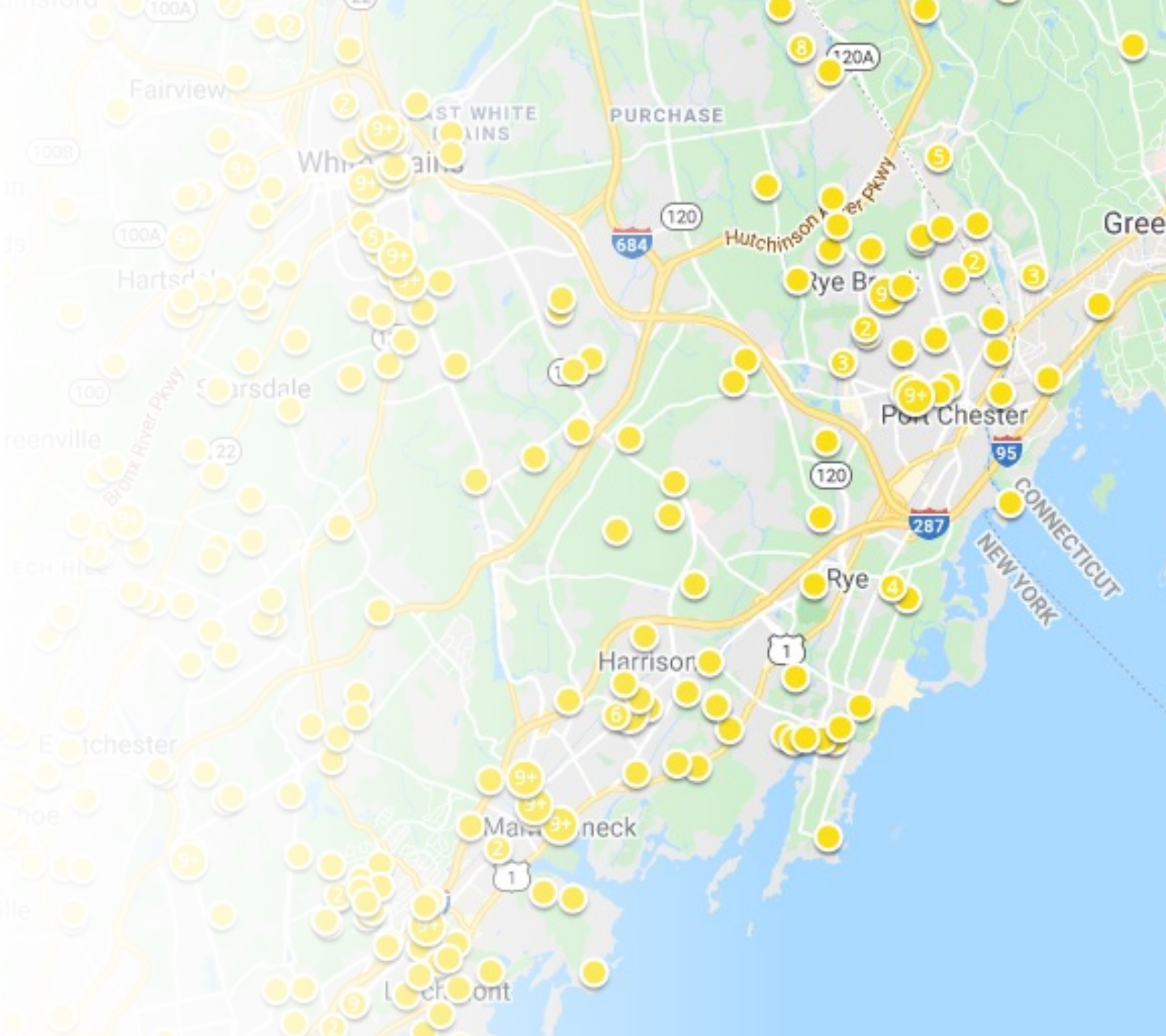
Building an Application to Predict Home Valuations

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)

Building from Previous Work

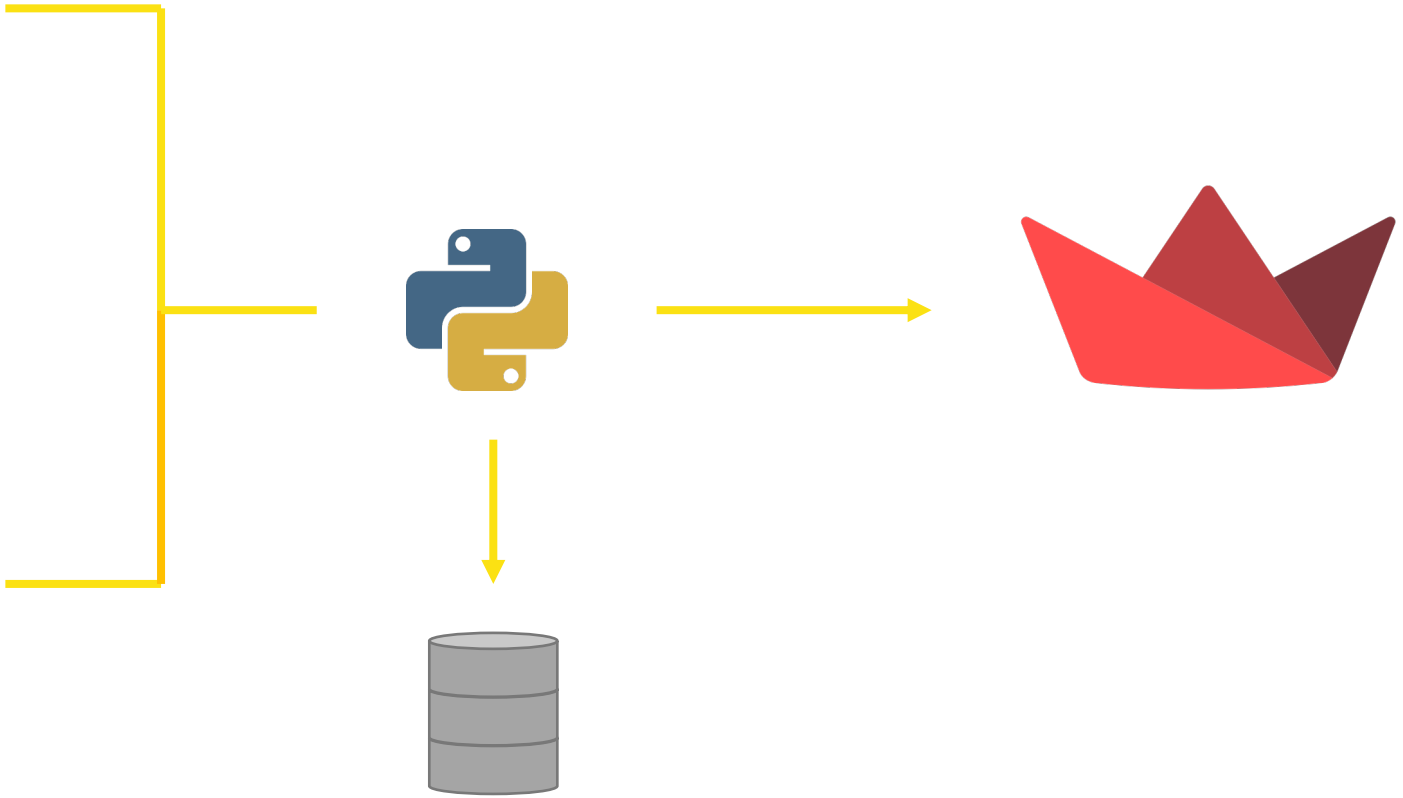
- Previous Zillow webscrapping to answer a question specific to northern NYC suburbs

- Convert this question into a flexible tool that could run a basic analysis for a house in any area

Data Pipeline

- Square Footage
- Beds
- Baths
- Year Built
- Lot Size
- Geography
- Zestimate
- Tax Assessment
- Days on Zillow

- State income tax brackets and rates



Zillow Page

Zillow Save Share More

$1,625,000 5 bd | 4 ba | 3,540 sqft

105 Middlesex Rd, Darien, CT 06820

Est. payment: $8,322/mo Get pre-qualified

Contact Agent | Take a Tour

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

Facts and features

Type:	Single Family Residence	**Cooling:**	Central Air
Year built:	1950	**Parking:**	2 Attached Garage spaces
Heating:	Forced Air, Oil	**Lot:**	0.59 Acres
		Price/sqft:	$459

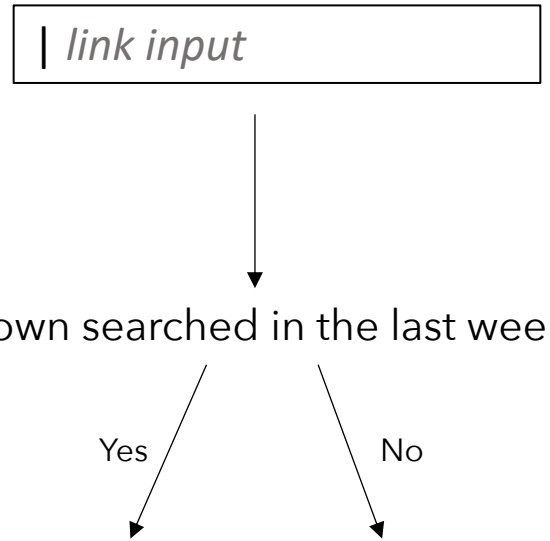
Public tax history

Year	Property Taxes	Tax Assessment
2020	$14,182(-0.9%)	$868,490

Application Flow

Other Key problems

- Data acquisition and storage
- Consistency in regression with variable inputs



- Speed *(solution: timestamp dictionary)*

Pull from DB

- Grab all saved home sales recently scraped for this town from our DB

Scrape Process

- Generate list of unique sale links from a town
- Check this new link list against saved sales; drop data already stored
- Scrape remaining new data, recombine with stored data

Application Output

- Beds, baths, lot size, sqft (with percentile ranks)
- Listed price
- Predicted price from OLS
- Spread (between predicted and listed price)
- Tax assessment
- Number of homes in comparison set
- Most similar homes (in features, not price)
- R2 and regression/density plot of OLS
- State income tax brackets