

SOLD MAY 31, 2017



Predicting house prices

The right price means a fast, profitable sale

Scrape Redfin
(BeautifulSoup, selenium)

The right price means a fast, profitable sale

Scrape Redfin
(BeautifulSoup, selenium)

3111 unique entries
11 features

The right price means a fast, profitable sale

Scrape Redfin
(BeautifulSoup, selenium)

Clean data
(pandas, seaborn)

The right price means a fast, profitable sale

Scrape Redfin
(BeautifulSoup, selenium)

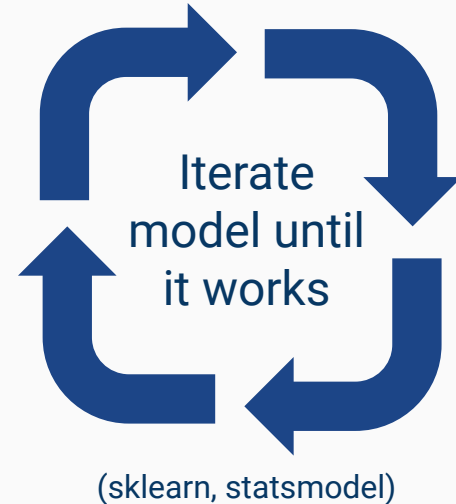
Clean data
(pandas, seaborn)

2050 usable entries
10 features

The right price means a fast, profitable sale

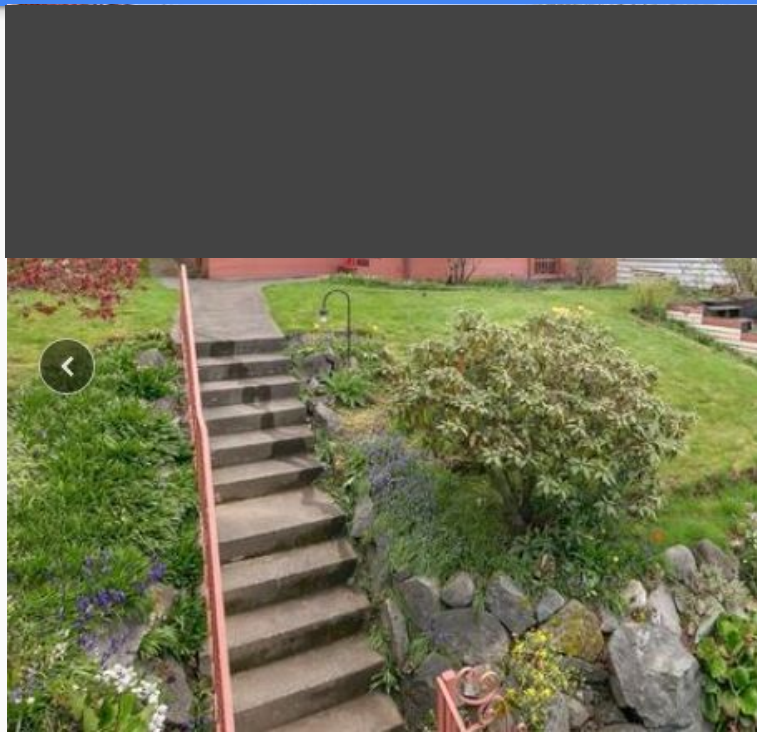
Scrape Redfin
(BeautifulSoup, selenium)

Clean data
(pandas, seaborn)



Features captured explain $\sim 70\%$ of variance

- ZIP code (quintile)
- Beds
- Baths
- Size (square feet)
- Lot size (square feet)
- Age (years)
- Sale date (years ago)
- Brokerage fee (%)
- Type (house, condo, townhouse)
- Basement



Model iteration doesn't improve predictive power enough

Polynomial overfit

LASSO eliminated # beds & basement

Linear regression with first order and a few second order features = baseline

LASSO eliminated # baths and ZIP



How much value does a bedroom add?

Linear regression on fresh split



How much value does a bedroom add?

Linear regression on fresh split

Model isn't sensibly interpretable

- Bedrooms are worth -\$20k
- Lot is worth -\$10/sq ft



80%

Features captured explain approximately 80% of the variance in house prices

There's a reason
realtors make
big paychecks!

