

Estimating Home Values

September 15, 2021

A presentation by Steven J. Newton

How outliers are defined has a large impact on evaluation stage

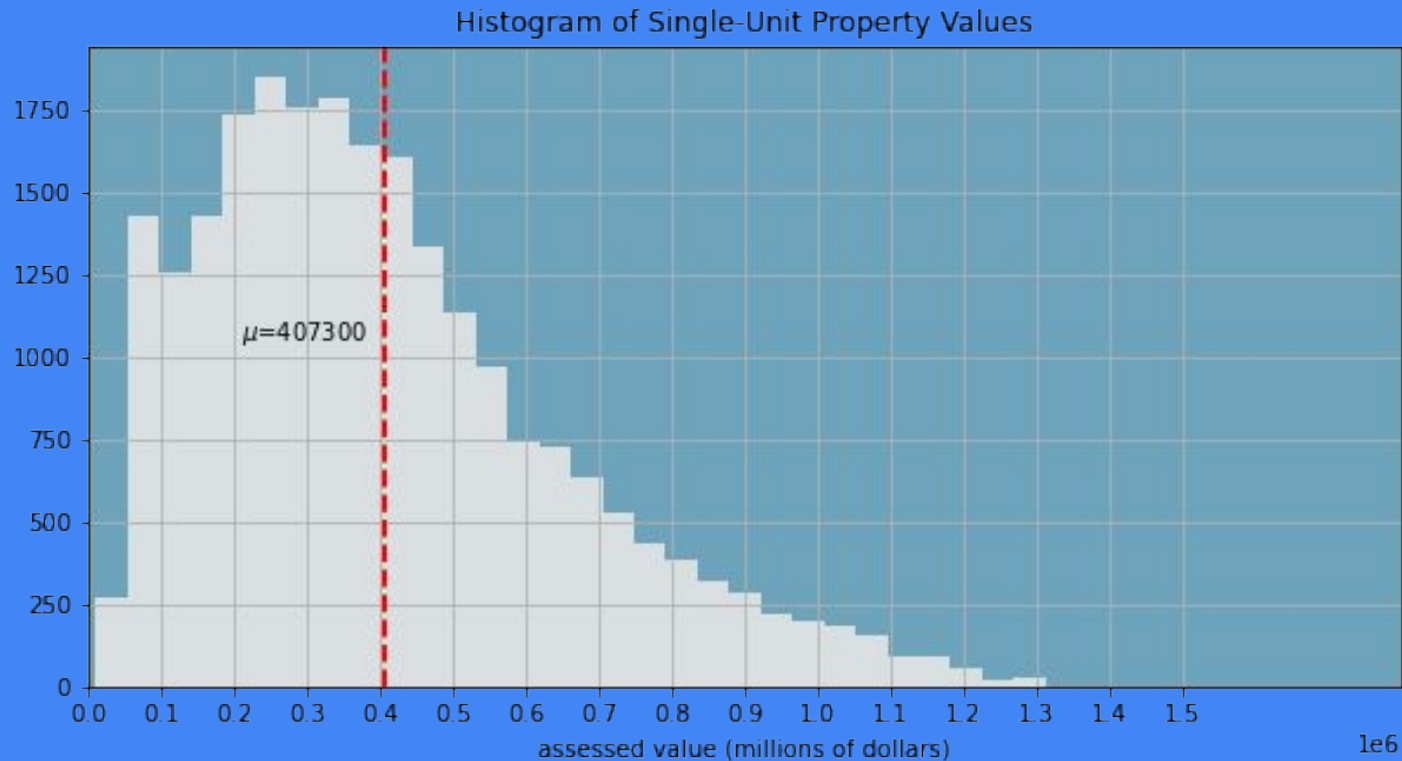
Incomplete cases: any property missing either **finished sqft.** or **tax value** → **114**

Outlier: a property with **tax value** or **tax amount** at least **2*IQR** beyond Q1 or Q3 → **5,132**

Single-unit properties sold May-Aug. 2017 in dataset → **33,376**

Range of tax values (**target variable**) in training set → **\$11k - \$1.31M**

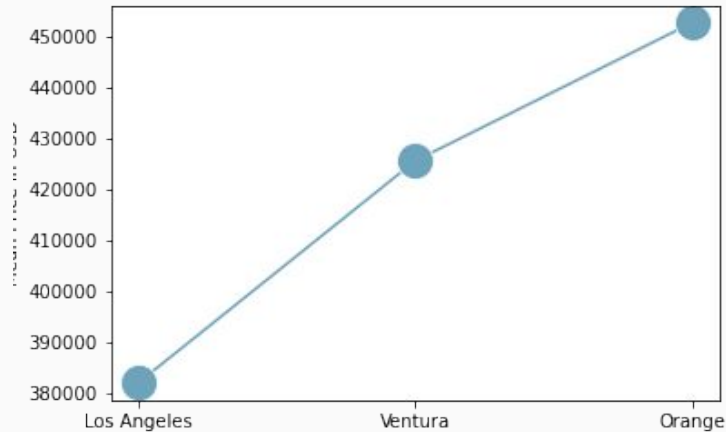
A Right-Skewed Target Variable



Mann-Whitney U and Pearson Correlation Tests

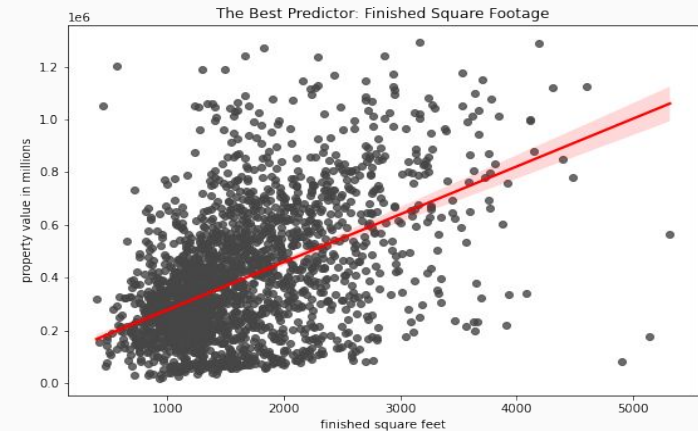
Is value related to county?

Yes, across all counties, $p < .01$.



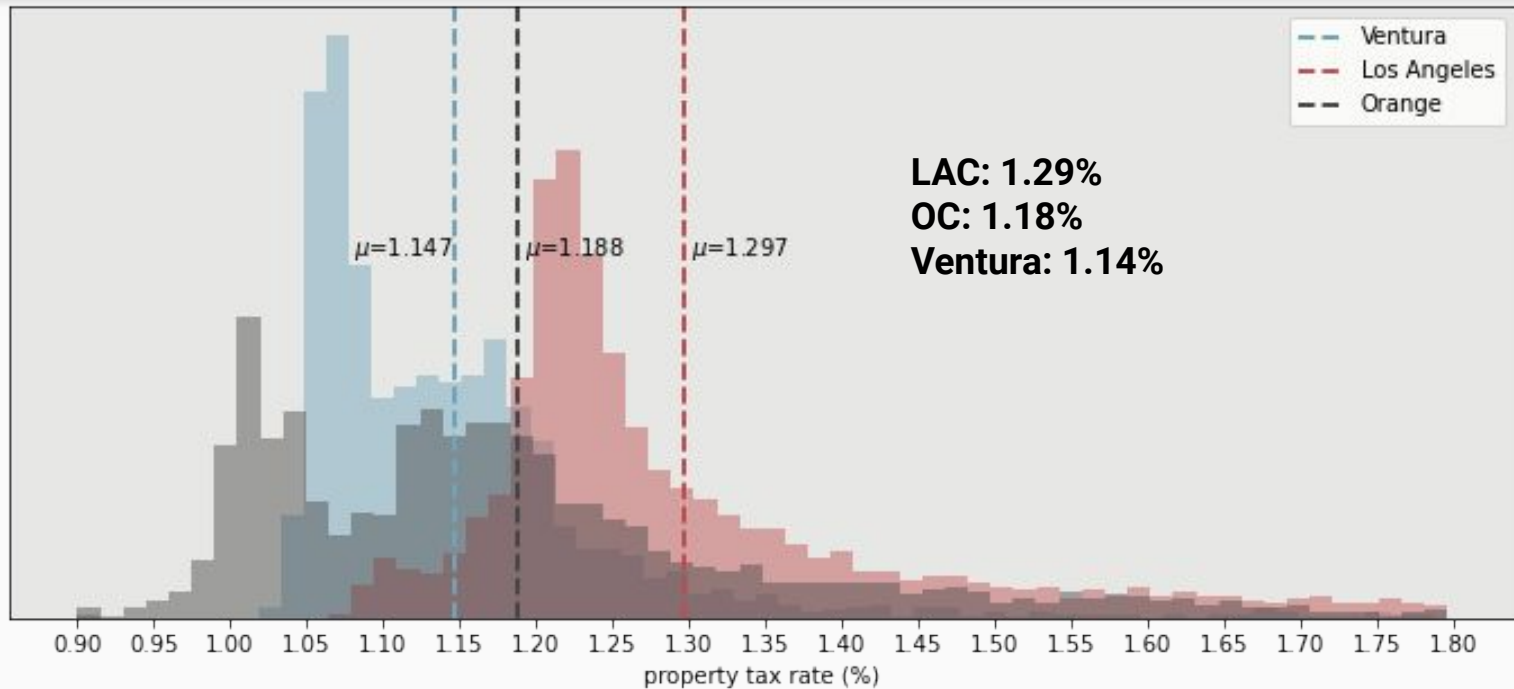
What numerical variables are strong predictors of the target?

Finished sqft. [corr.=.51] & **bathrooms** [corr.=.39]



Note heteroskedasticity

Property Tax Rates by County



Model Evaluation

<i>metric/model</i>	Mean	MVP	LinearReg	LassoLars
RMSE	245000	211000	206000	190000
MAPE	.90	.76	.73	.66
R ²	0	.26	.29	.40

RMSE: root mean squared error; units are same as target variable (USD)

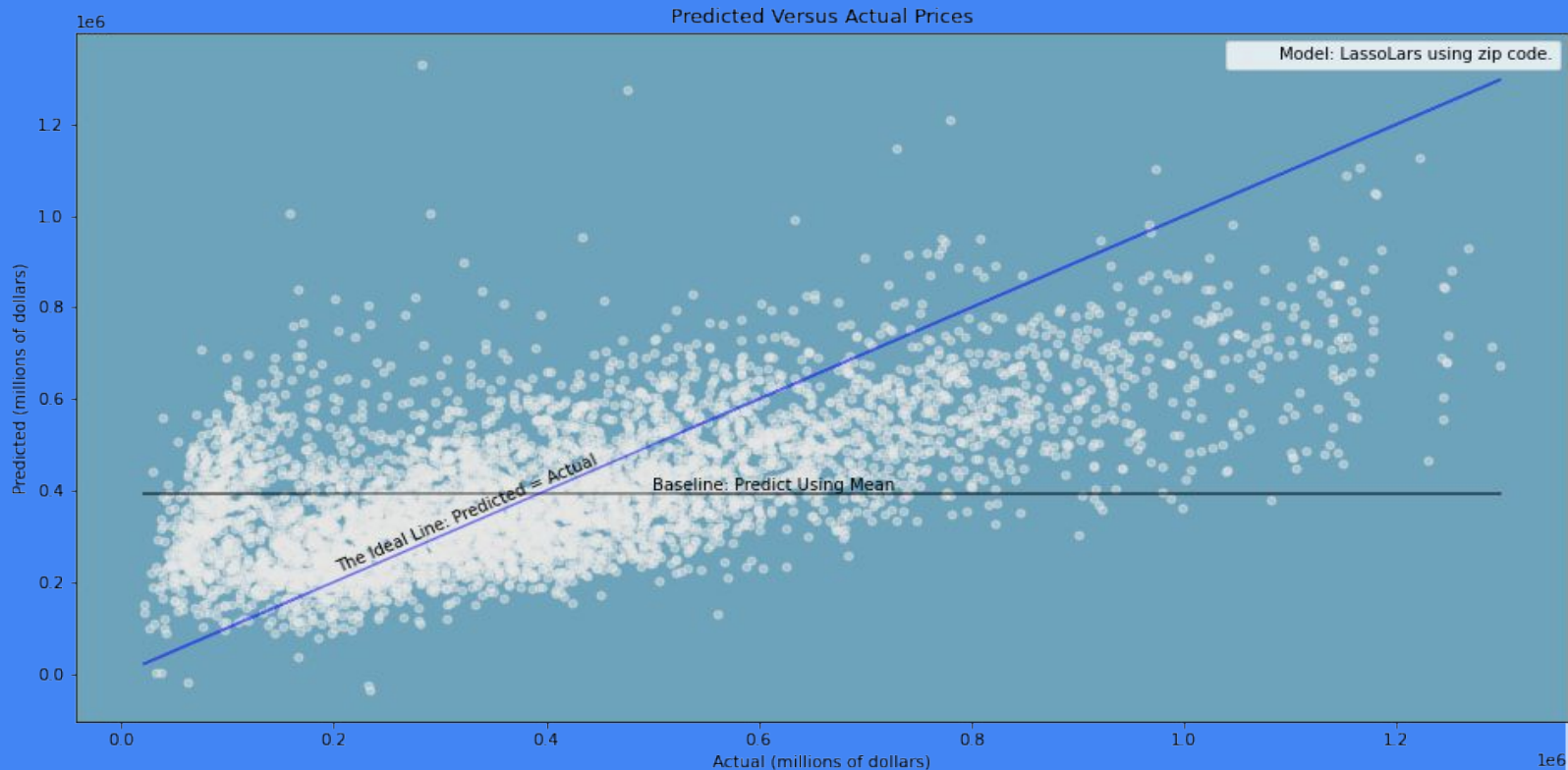
MAPE: mean absolute percentage error; from the Zillow Prize FAQ:

“Home sale prices have a right-skewed distribution and are also strongly heteroscedastic so we need to use a relative error metric instead of an absolute metric to ensure valuation models are not biased towards expensive homes.”
Additional usage note: MAPE puts a heavier penalty on negative errors and is therefore biased towards a method whose forecasts are too low.

R²: proportion of variance in target variable explained by model

	LassoLars on test set
RMSE	188000
MAPE	.63
R ²	.40

Residuals from LASSO+LARS: Is this model valid?



Where are these houses anyway?

Latitude-longitude and FIPS data placed the houses in three counties: **Orange**, **Ventura**, and **Los Angeles**. But the same houses' zip codes placed them in **Northern California** or elsewhere. Such a discrepancy becomes a show-stopper when testing on real data.




Flask App Example 1



4 bd | 3 ba | 2,689 sqft

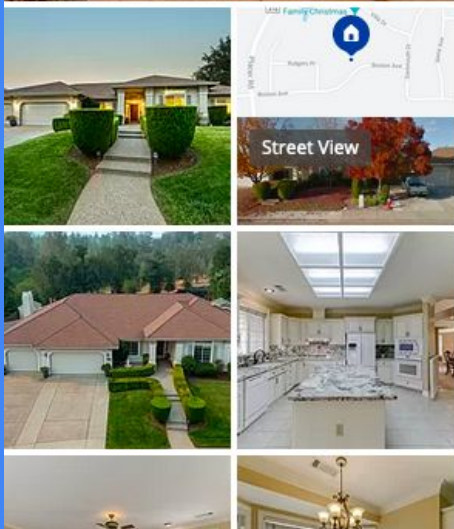
4156 Boston Ave, Redding, CA 96001

● **Sold: \$705,000** | Sold on 09/14/21 | Zestimate®: **\$650,000**

Est. refi payment: \$3,013/mo  [Refinance your loan](#)

[Home value](#) | [Owner tools](#) | [Home details](#) | [Neighborhood details](#)

Home value



Zestimate

\$650,000



Zestimate range

\$618,000 - \$683,000



Last 30-day change

+ \$56,097 (+9.4%)



Zestimate per sqft

\$242

Flask App Example 2



3 bd | 2 ba | 1,421 sqft

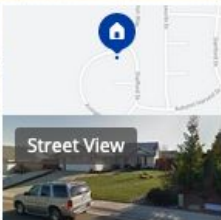
3217 Avington Way, Shasta Lake, CA 96019

● **Sold: \$353,000** | Sold on 09/10/21 | Zestimate®: **\$331,200**

Est. refi payment: \$1,493/mo  [Refinance your loan](#)

[Home value](#) | [Owner tools](#) | [Home details](#) | [Neighborhood details](#)

Home value



Zestimate

\$331,200



Zestimate range

\$315,000 - \$348,000



Last 30-day change

- \$31,700 (-8.7%)



Zestimate per sqft

\$233