

Just what makes for a good property value estimator?

Kelly Slattery
US-DSI-10
Jan. 17, 2020



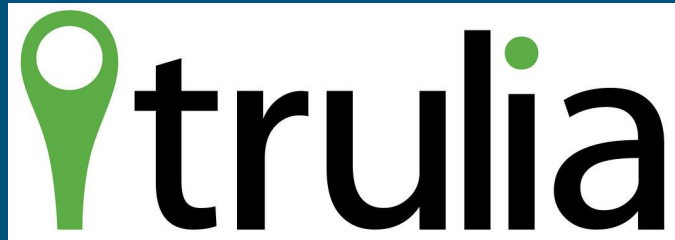
Automated Valuation Models (AVMs)

The Nextdoor logo consists of a white house icon with a checkmark inside, followed by the word "Nextdoor" in a white sans-serif font, all on a green rectangular background.

Nextdoor

The Redfin logo features a red square with a white stylized "R" inside, followed by the text "redfin.com" in a red sans-serif font, all on a white rectangular background.

redfin.com

The Trulia logo features a green location pin icon with a white circle inside, followed by the word "trulia" in a black sans-serif font, all on a white rectangular background.

trulia

The Zillow Zestimates logo features a stylized blue and green house icon, followed by the word "Zillow" in a blue sans-serif font and "Zestimates" in a smaller blue sans-serif font, all on a white rectangular background.

Zillow[®]
Zestimates



How does Zillow do it?

Features

Data

- Publicly available information (county and tax assessor records, listing services and brokerages)
- User-submitted

- Home facts (sq ft, location, bathrooms)
- Unique features (hardwood floors, granite countertops, landscaped backyard, etc.)
- Market conditions

Limitations

Median Error Rate

~1.9% on-the-market
~7.5% off-market

"We do not know about home values and remodels, unless they have been reported to the local tax assessor, so those values are not used in Zestimate calculations."

- Only look at a few factors
- Many outdated listings



How does Redfin do it?

Data

- Full access to the MLS
- Data based on recently-sold properties in the area
- List prices

Median Error Rate

~1.70% on-the-market
~5.95% off-market

Features

- Home facts (address, age, sq ft, bedrooms, baths, upgrades, school district)
- Photographs

Limitations

"Redfin is less likely to provide home value estimates for real estate appraised at more than \$2 million."



How does Trulia do it?

Data

- Publicly available information
- Agent-reported transactions

Median Error Rate

~5.3%

Features

- Location, lot size, square footage, number of bedrooms and bathrooms, etc.
- Property tax info: actual property taxes paid, exceptions to tax assessments, etc.
- Comparable recent sales of nearby homes

Limitations

- *No user-submitted data*
- *Fewest number of estimates*



How does Nextdoor do it?

Data

- HouseCanary
 - Proprietary Privacy Score

Median Error Rate

~2.9%

Features

- Sales prices of comparable houses
- Similar attributes of other nearby properties
- Home improvement history
- Local economic trends, consumer demand

Limitations

"HouseCanary's new portal ComeHome was questioned by MLS leaders for using data without proper authorization"

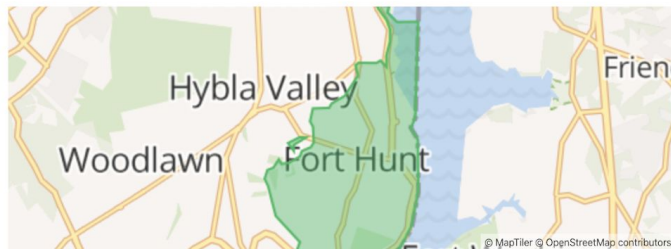


Public Agencies

[Help](#) · [Guidelines](#) · [Privacy](#)
[About](#) · [Jobs](#) · [Press](#) · [Blog](#)

© 2020 Nextdoor

Market overview in 22308



Median listing price

-0.2% from last month

Strong seller's market ⓘ

47 days average on market until sale



Get a professional valuation

5 nearby listings



\$593,300

8727 Bluedale St, Alexandria, VA 22308
5 beds · 3 baths · 2,648 sqft

Compare



\$378,000

613 Broad Creek Dr, Fort Washington, MD 20744
3 beds · 3 baths · 3,101 sqft

Compare

Real Estate

Edit Home Facts Only visible to you ⓘ



Square feet

Bedrooms

Lot Size (sqft)

Year Built

Full baths ⓘ

3/4 Baths ⓘ

1/2 Baths ⓘ

1/4 Baths ⓘ

Save changes

Is this a good time of year to sell?

What steps should I take to prepare my home for sale?

Home sale

This calculator
professional e

ⓘ Offer price

\$0

ⓘ Mortgage

\$10,000

ⓘ Real estate

6%

ⓘ Title, escrow

1%

ⓘ Staging a

\$10,000

Estimated s

\$0

How does my model do it?

Data

- Filling nulls
- Converting data types
- Handling outliers
- Creating dummy columns
- Creating new columns
- Creating interaction terms

Features (23)

- | | |
|---------------------------|---------------------------|
| - 'Overall Qual', | - 'Lot Area', |
| - 'Exter Qual', | - 'Paved Drive' |
| - 'Kitchen Qual', | - 'Bsmt Full Bath', |
| - 'Total Bsmt SF', | - 'Central Air', |
| - '1st Flr SF', | - 'Garage Cond', |
| - 'Garage Cars', | - 'Heating QC', |
| - 'Bsmt Qual', | - 'neighborhood_StoneBr', |
| - 'Year Built', | - '2nd Flr SF', |
| - 'Year Remod/Add', | - 'has_garage', |
| - 'Fireplace Qu', | - 'Bsmt Unf SF', |
| - 'neighborhood_NridgHt', | - 'zone_RM', |
| - 'Bsmt Exposure', | |

Why?

Junk in → Junk out



=



Data Cleaning

Filling nulls

Converting data types

Handling outliers

Creating new columns

- 17 categorical columns → ordinal values
 - Exter Qual
 - Bsmt Qual
 - Bsmt Exposure
 - Heating QC
 - Kitchen Qual
 - Fireplace Qu
 - Garage Cond
 - Paved Drive
- 64 dummy columns
 - MS Zoning
 - Neighborhood
- 3 additional calculated columns

Filling nulls

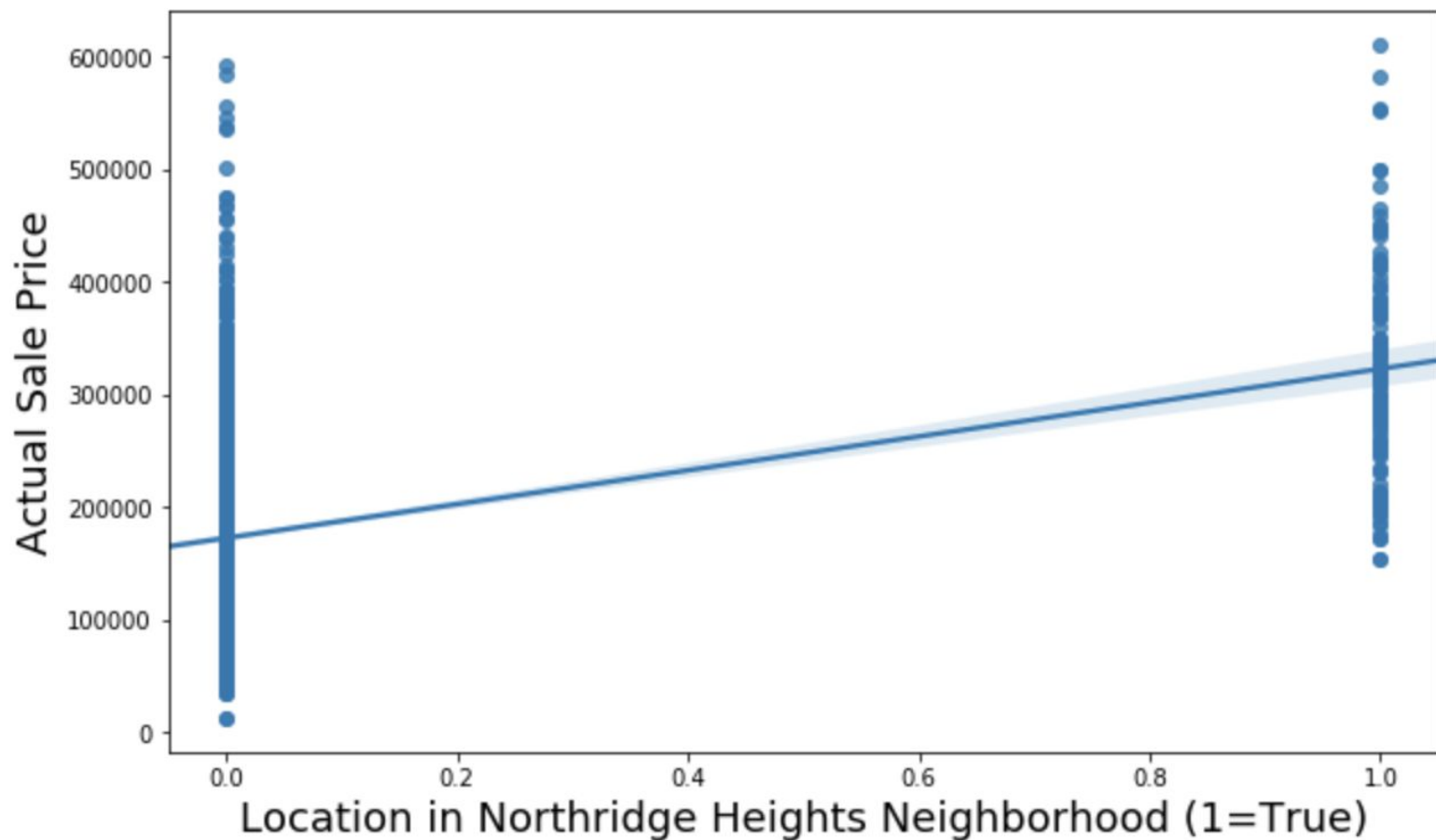
- 26 columns with missing values
- Deleted 5 rows containing multiple null values without clear meaning
- Filled NaN with 'N/A' or 0 where applicable to indicate lack of attribute
- Filled 'Garage Yr Blt' with mean year
- Filled 'Lot Frontage' with estimate based on 'Lot Area' and 'Lot Shape'
- Transformed and created binary columns into [0 , 1]
- Assigned ordinal values to 17 columns



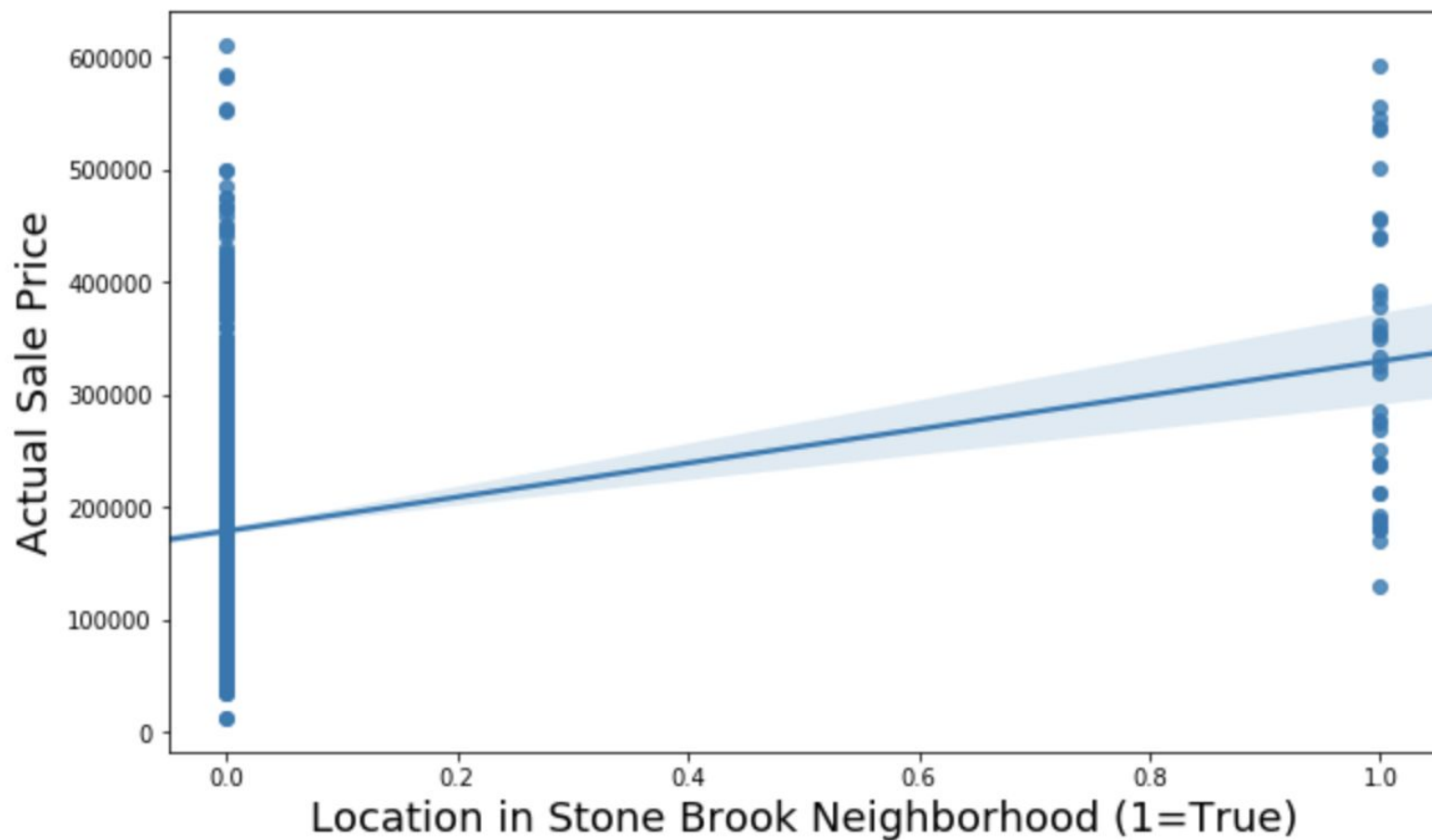
Engineered Features

Column Name	Column Description	% of Data	e^{\wedge} standardized coefficient
'neighborhood_NridgHt'	1 if house in Northridge Heights	~6%	~ 1.010
'neighborhood_StoneBr'	1 if house in Stone Brook	~2%	~ 1.004
'zone_RM'	1 if in a Residential Medium Density Area	~15%	~ 0.981
'has_garage'	1 if garage, 0 if no garage	~94%	~ 0.947

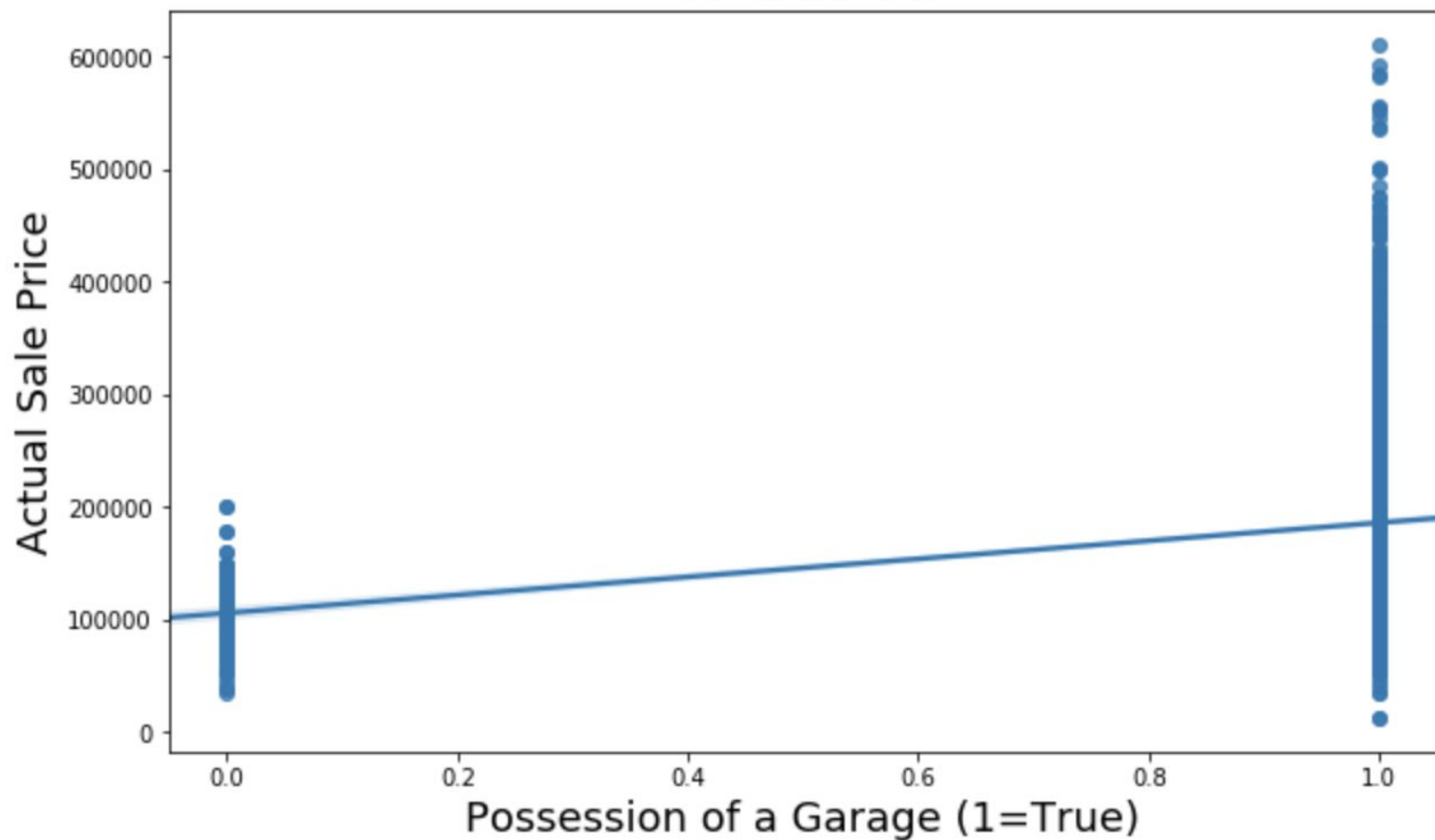
Effect of Location in Northridge Heights Neighborhood
on Sale Price



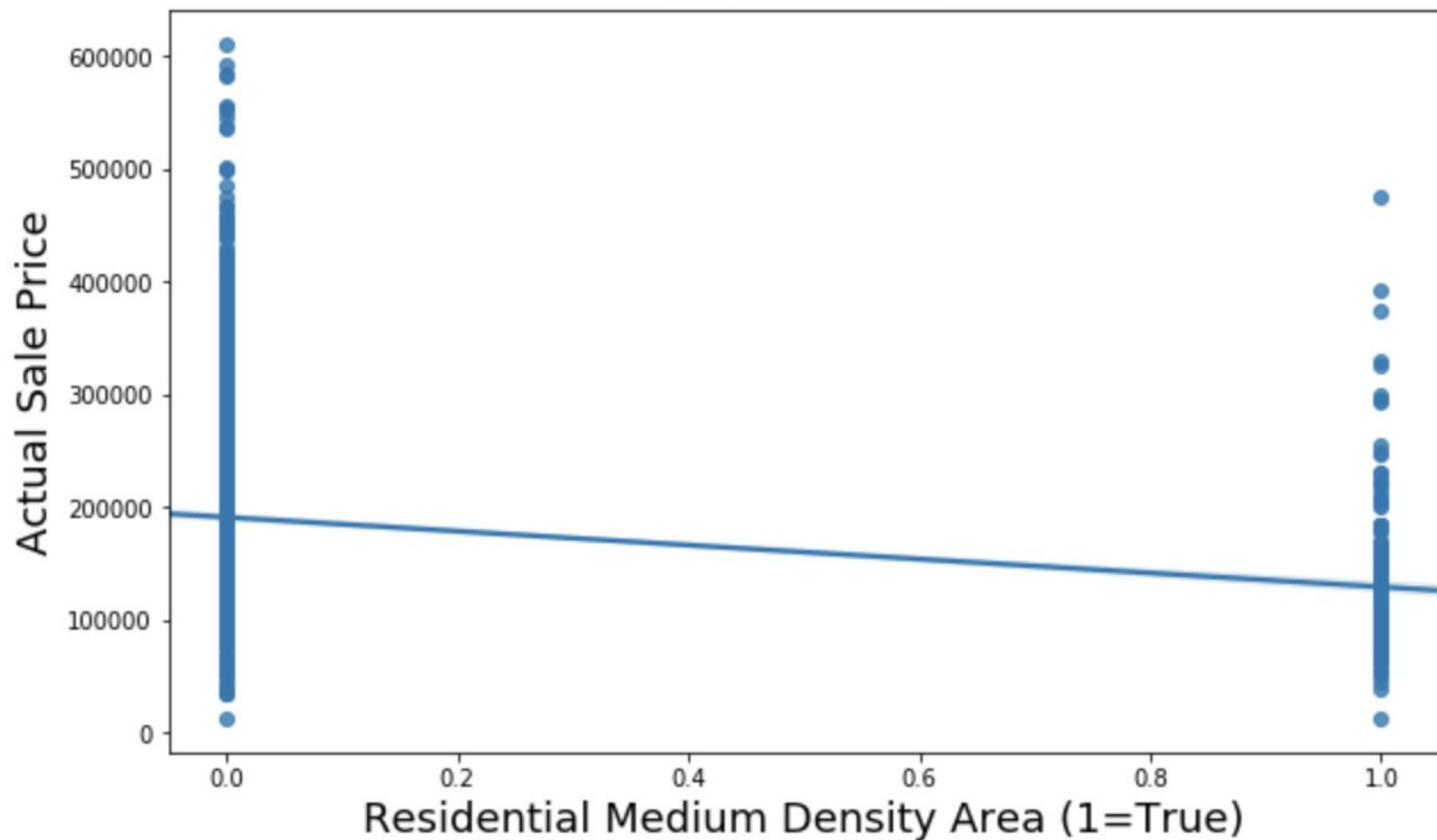
Effect of Location in Stone Brook Neighborhood
on Sale Price



Effect of Possessing a Garage on Sale Price



Effect of Location in a Residential Medium Density Area
on Sale Price



Significance of Features

	coef	std err	t	P> t	[0.025	0.975]		coef	std err	t	P> t	[0.025	0.975]
const	6.3523	0.454	14.006	0.000	5.463	7.242	Bsmt Unf SF	-8.294e-05	9.92e-06	-8.357	0.000	-0.000	-6.35e-05
neighborhood_StoneBr	0.0464	0.023	2.035	0.042	0.002	0.091	Total Bsmt SF	0.0002	1.66e-05	9.577	0.000	0.000	0.000
neighborhood_NridgHt	0.0368	0.014	2.569	0.010	0.009	0.065	Heating QC	0.0245	0.004	6.173	0.000	0.017	0.032
zone_RM	-0.0702	0.009	-7.484	0.000	-0.089	-0.052	Central Air	0.0978	0.014	7.168	0.000	0.071	0.125
has_garage	-0.2249	0.044	-5.091	0.000	-0.312	-0.138	1st Flr SF	0.0003	1.65e-05	16.703	0.000	0.000	0.000
Lot Area	4.05e-06	5.07e-07	7.992	0.000	3.06e-06	5.04e-06	2nd Flr SF	0.0002	9.03e-06	27.668	0.000	0.000	0.000
Overall Qual	0.0759	0.004	18.756	0.000	0.068	0.084	Bsmt Full Bath	0.0193	0.008	2.460	0.014	0.004	0.035
Year Built	0.0003	0.000	2.019	0.044	9.95e-06	0.001	Kitchen Qual	0.0323	0.007	4.450	0.000	0.018	0.047
Year Remod/Add	0.0017	0.000	7.817	0.000	0.001	0.002	Fireplace Qu	0.0123	0.002	6.129	0.000	0.008	0.016
Exter Qual	0.0203	0.009	2.266	0.024	0.003	0.038	Garage Cars	0.0265	0.006	4.110	0.000	0.014	0.039
Bsmt Qual	0.0111	0.006	1.926	0.054	-0.000	0.022	Garage Cond	0.0922	0.014	6.420	0.000	0.064	0.120
Bsmt Exposure	0.0136	0.003	4.103	0.000	0.007	0.020	Paved Drive	0.0834	0.014	6.079	0.000	0.057	0.110

Strongest Predictors (ranked)

1. 'Overall Qual'
2. '2nd Flr SF'
3. '1st Flr SF'
4. 'Garage Cond'
5. 'Total Bsmt SF'
6. 'Year Remod/Add'
7. 'Lot Area'
8. "Paved Drive"
9. 'Central Air'
10. 'Kitchen Qual'

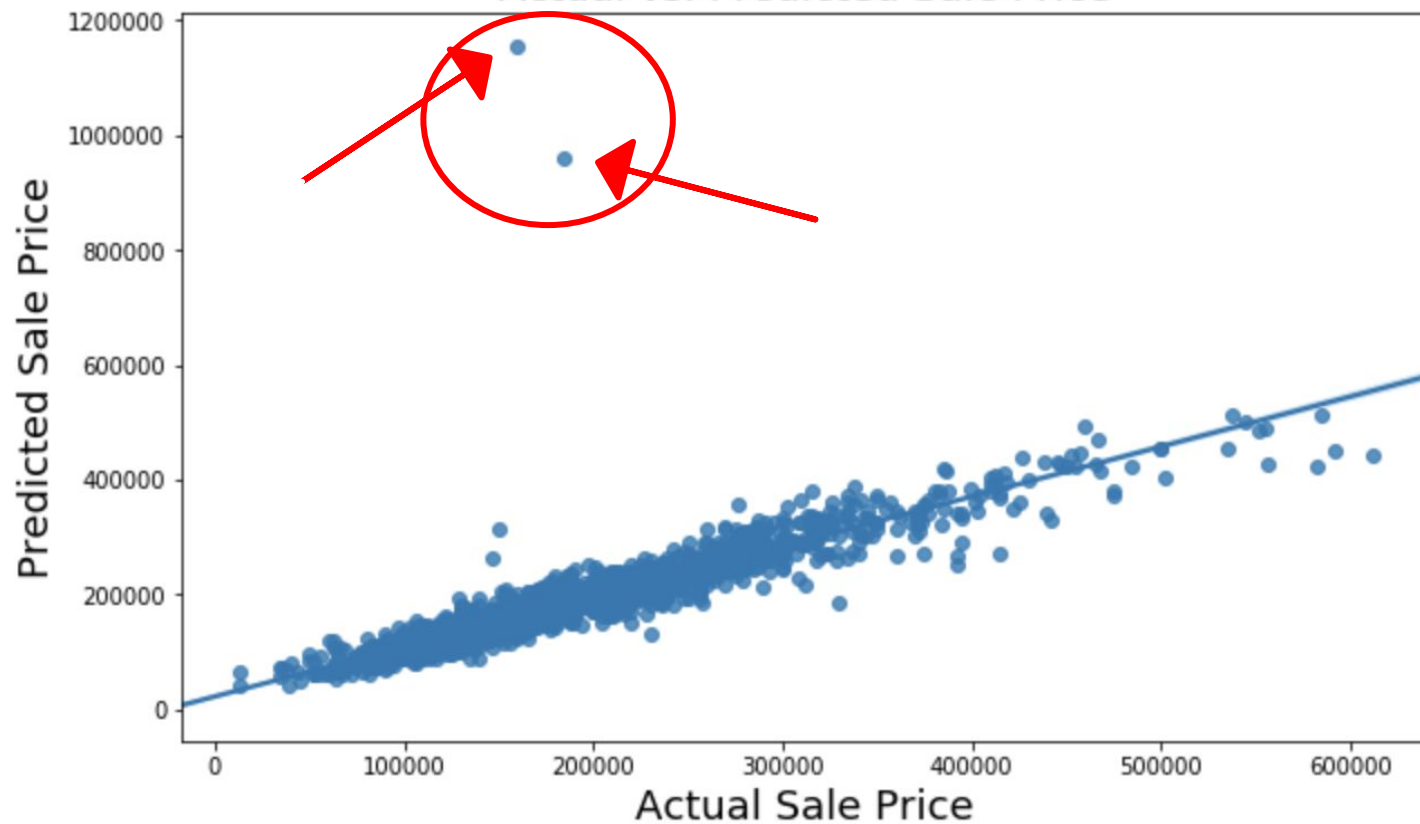
“Joe Biden” Effect

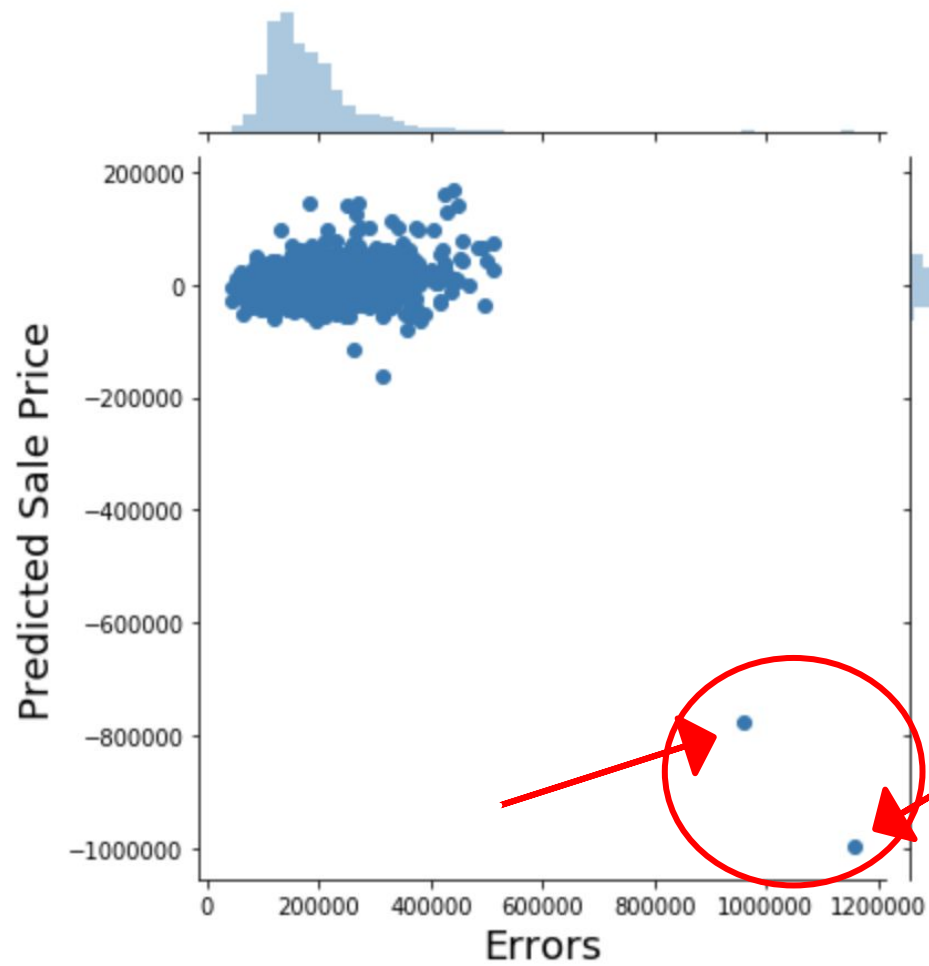


Though there will
always be
irreducible error...

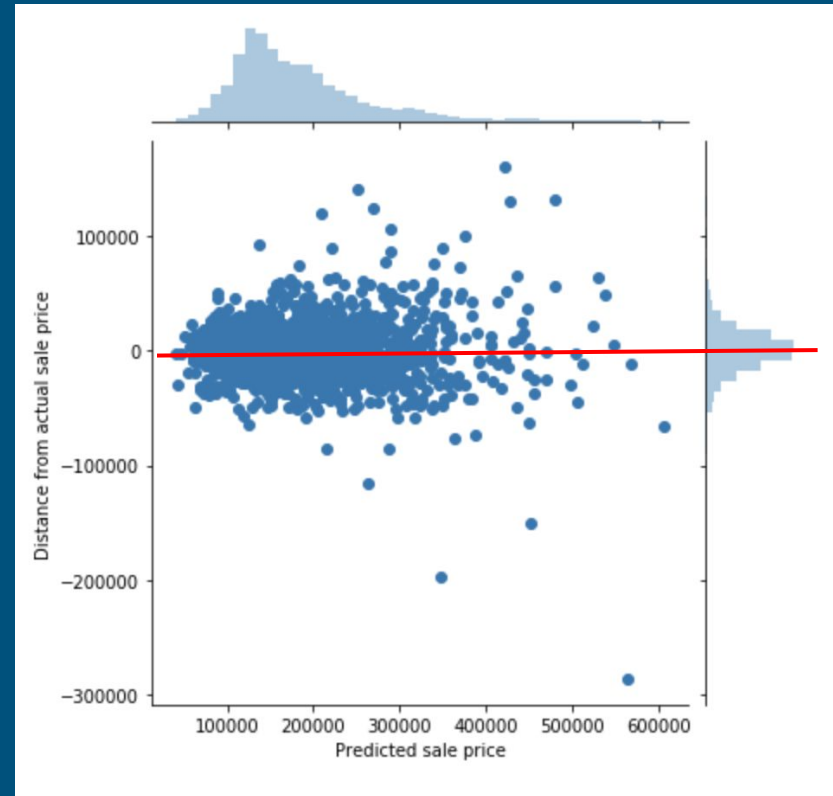
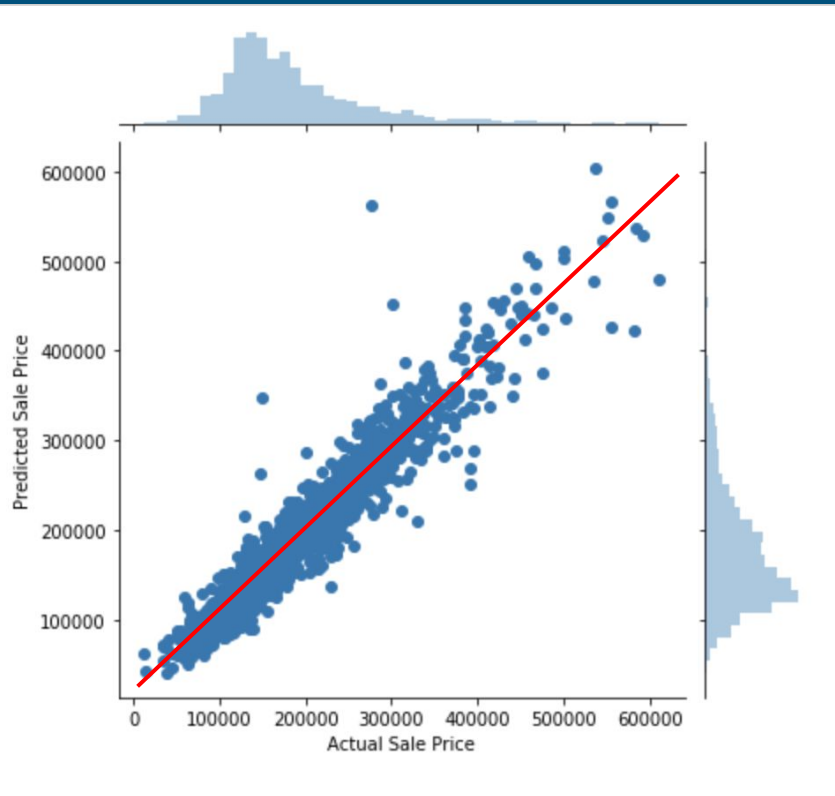
some data points
are just noise

Actual vs. Predicted Sale Price





Model Evaluation without Outliers



My Ames Housing Market Pricing Model

Data

- Publicly available dataset from Ames Assessor's Office for individual residential properties sold in Ames, Iowa from 2006 - 2010

Median Error Rate

~6.8%

Features

- Basic home information
- 2 neighborhoods, 1 zone

Limitations

- *No market conditions considered*
- *No geohashing / unique neighborhood modeling*
- *No nuances accounted for*
- *Iowa State University not considered*

What could improve my model?

Recommendations
based on popular AVMs
(Zillow, Redfin, Trulia, NextDoor)

- Include more data
 - Market conditions (Ex. Did any new large employers come to Ames 2006-2010?)
 - Location-specific criteria (Ex. schools, crime rates, proximity to public transport, stores, water)
 - Neighborhood-specific models
 - Compare to sale prices of comparable houses nearby
 - Expertise knowledge
 - Real estate property appraisal is location-specific and nuanced
 - Expertise advisable for all fields
-

Key takeaways

1. Simplicity is key for generalization
2. Determine if outliers are part of “irreducible error” noise or the “Joe Biden Effect”
3. A good model starts with good data
4. In real estate, nuance is key

Food for thought...

inman

Agent

Brokerage

Agent Appreciation Month

Connect New York

More ▾

Search 🔍

Log In

SUBSCRIBE

TECHNOLOGY

Zestimate will be a 'live offer for every house' Zillow CEO says

In addition, the Zestimate has also generated some controversy in the past; the tool is [currently the subject of a lawsuit](#) alleging Zillow conspired with real estate brokerages to hide its valuations for certain properties. [Tweaks to the tool](#) that led to valuation changes have raised eyebrows, and questions, in the recent past as well.



Sources

Images:

<https://www.inman.com/2019/06/25/nextdoor-launches-your-home-valuation-how-much-worth/>

<http://press.redfin.com/logos/>

<https://www.flickr.com/photos/truliavisuals/5182312285>

<https://about.nextdoor.com/us-media/>

<https://chicagoagentmagazine.com/2018/07/30/cook-county-used-zillows-zestimate-tool-official-assessments-really/>

<https://faithway.church/junk-in-junk-out/>

<https://www.washingtonpost.com/nation/2019/05/17/joe-biden-onion-campaign-meme/>

Information:

<https://www.chicagotribune.com/real-estate/ct-re-property-valuation-0113-story.html>

<https://www.inman.com/2019/06/25/nextdoor-launches-your-home-valuation-how-much-worth/>

<https://joshflagg.com/post/are-online-home-valuation-estimates-accurate>

<https://www.inman.com/2019/04/02/zestimate-will-be-a-live-offer-for-every-house-zillow-ceo-says/>

<https://www.trulia.com/info/trulia-estimates/>

<https://www.homebay.com/tips/which-home-value-estimator-is-better-redfin-or-zillow/>

<https://nextdoor.com/real-estate/home-dashboard?is=tab>

<https://listwithclever.com/real-estate-blog/zillow-vs-redfin-vs-trulia/>

https://www.washingtonpost.com/realestate/which-firm-has-more-accurate-online-estimates-for-homes-zillow-or-redfin/2019/02/19/7a1f6436-346b-11e9-a400-e481bf264fdc_story.html?noredirect=on

<https://www.zillow.com/zestimate/>

<https://www.thebalance.com/what-is-mls-1798770>

https://help.nextdoor.com/s/article/about-property-value-estimates?language=en_US

<https://www.housecanary.com/real-estate-data-api/>

<https://www.housecanary.com/resources/housecanary-data-points-part-i-valuation/>

<https://www.inman.com/2019/01/25/about-face-new-zillow-competitor-removes-all-listings/>