



Ames Iowa Housing

David Capella • March 2019

Photo by [Andrew Seaman](#) on [Unsplash](#); Photo by [Drew Hays](#) on [Unsplash](#); Photo by [Drew Dau](#) on [Unsplash](#).

- We will be discussing the Ames Iowa Housing dataset.
- We will be training a model to predict the sale price in Ames, Iowa based on a few features from this dataset.
- Which features have the strongest effect on sale price adjustment?

	id	pid	ms_subclass	ms_zoning	lot_frontage	lot_area	street	alley	lot_shape	land_contour	...	screen_porch	pool_area	pool_qc	fence	misc_1
0	109	533352170	60	RL	72.774648	13517	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	
1	544	531379050	60	RL	43.000000	11492	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	
2	153	535304180	20	RL	68.000000	7922	Pave	NaN	Reg	Lvl	...	0	0	NaN	NaN	
3	318	916386060	60	RL	73.000000	9802	Pave	NaN	Reg	Lvl	...	0	0	NaN	NaN	
4	255	906425045	50	RL	82.000000	14235	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	

No data

- There was about 20.88% of either missing data or misinterpreted data. (Huge chunk)
- I decided to grab the mean of the missing data based on neighborhood or related feature.
- The other data I fixed so when it said null it really meant the house does not have said feature.
- Example: Masonry veneer area I filled with the mean based on the Masonry veneer type.

P Value

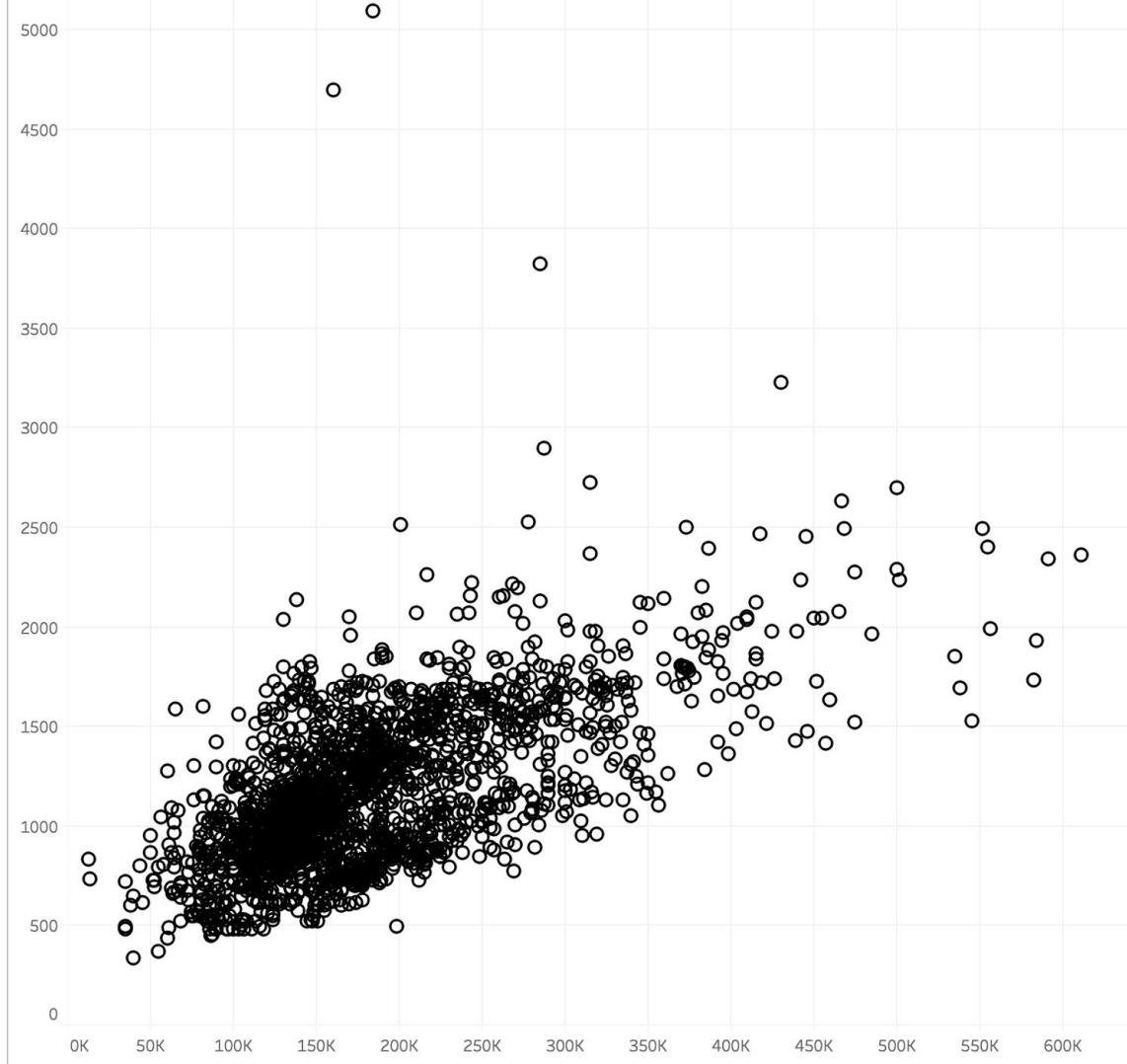
- Out of 234 feature columns, only 70 had a P value of less than .05.

OLS Regression Results

Dep. Variable:	saleprice	R-squared:	0.916
Model:	OLS	Adj. R-squared:	0.906
Method:	Least Squares	F-statistic:	87.12
Date:	Thu, 21 Mar 2019	Prob (F-statistic):	0.00
Time:	07:36:05	Log-Likelihood:	-23501.
No. Observations:	2051	AIC:	4.746e+04
Df Residuals:	1821	BIC:	4.876e+04
Df Model:	229		
Covariance Type:	nonrobust		

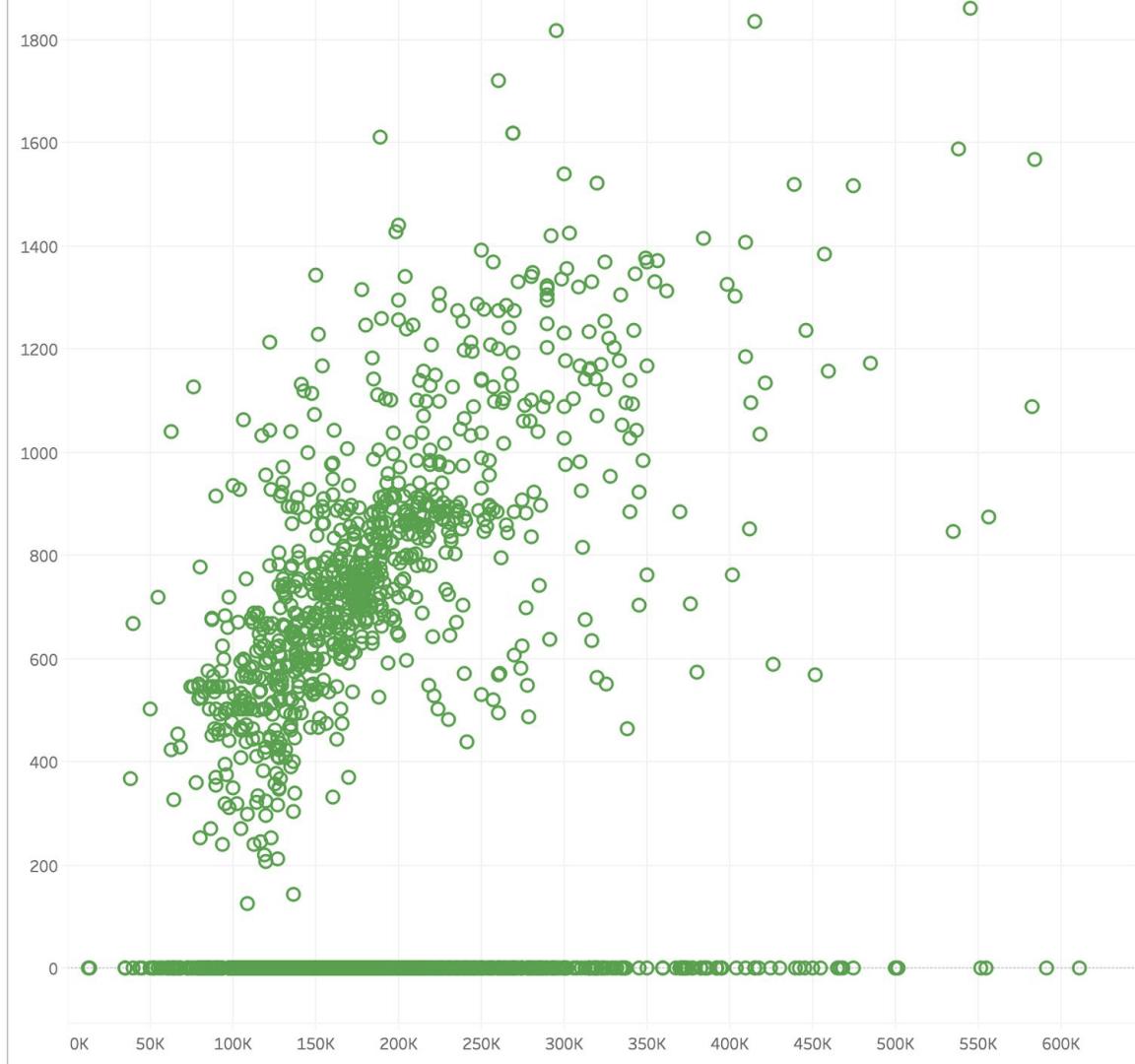
EDA

1st Floor SF. Vs. Sale Price

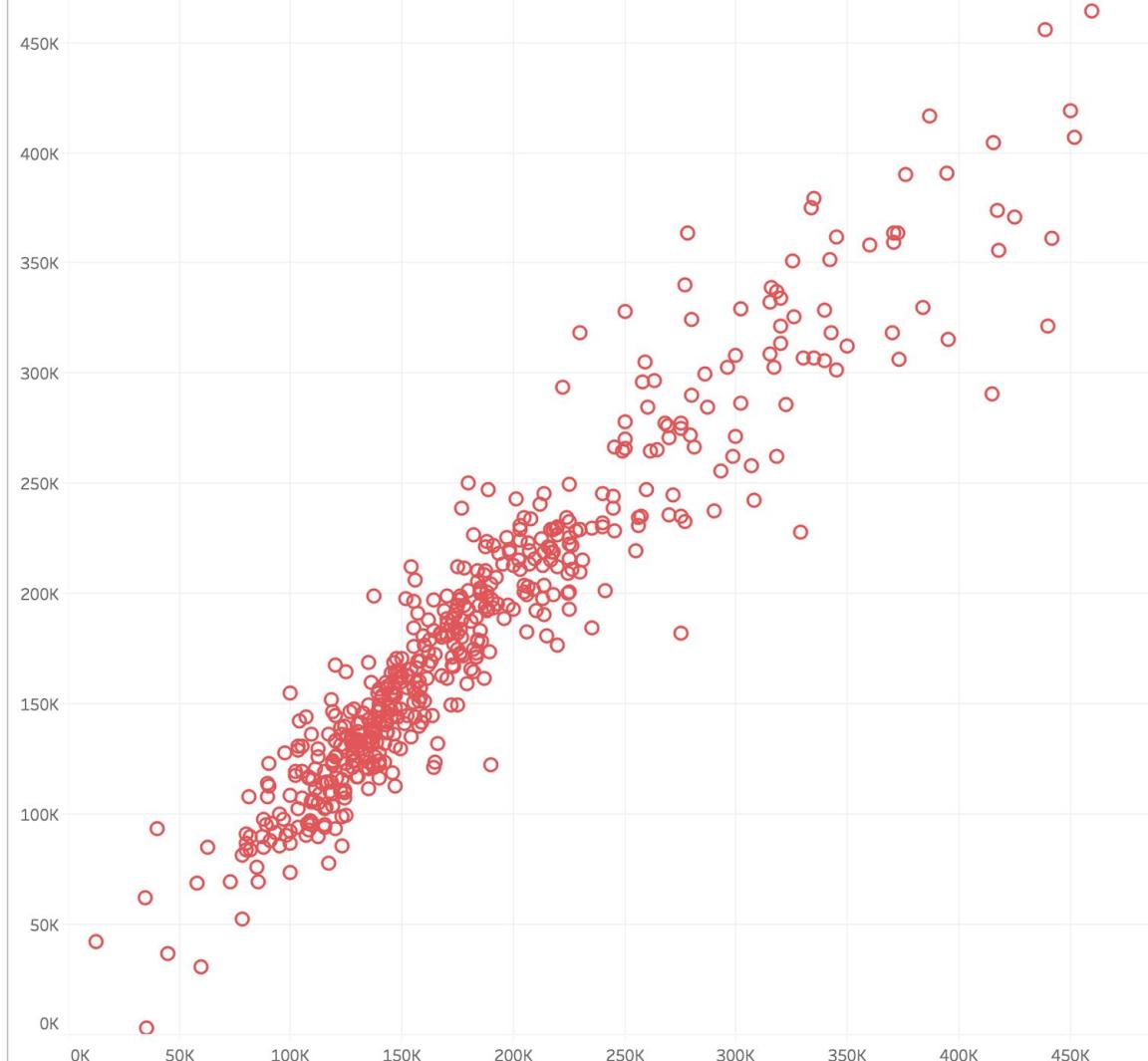


EDA

2nd Floor SF. Vs. Sale Price



Model

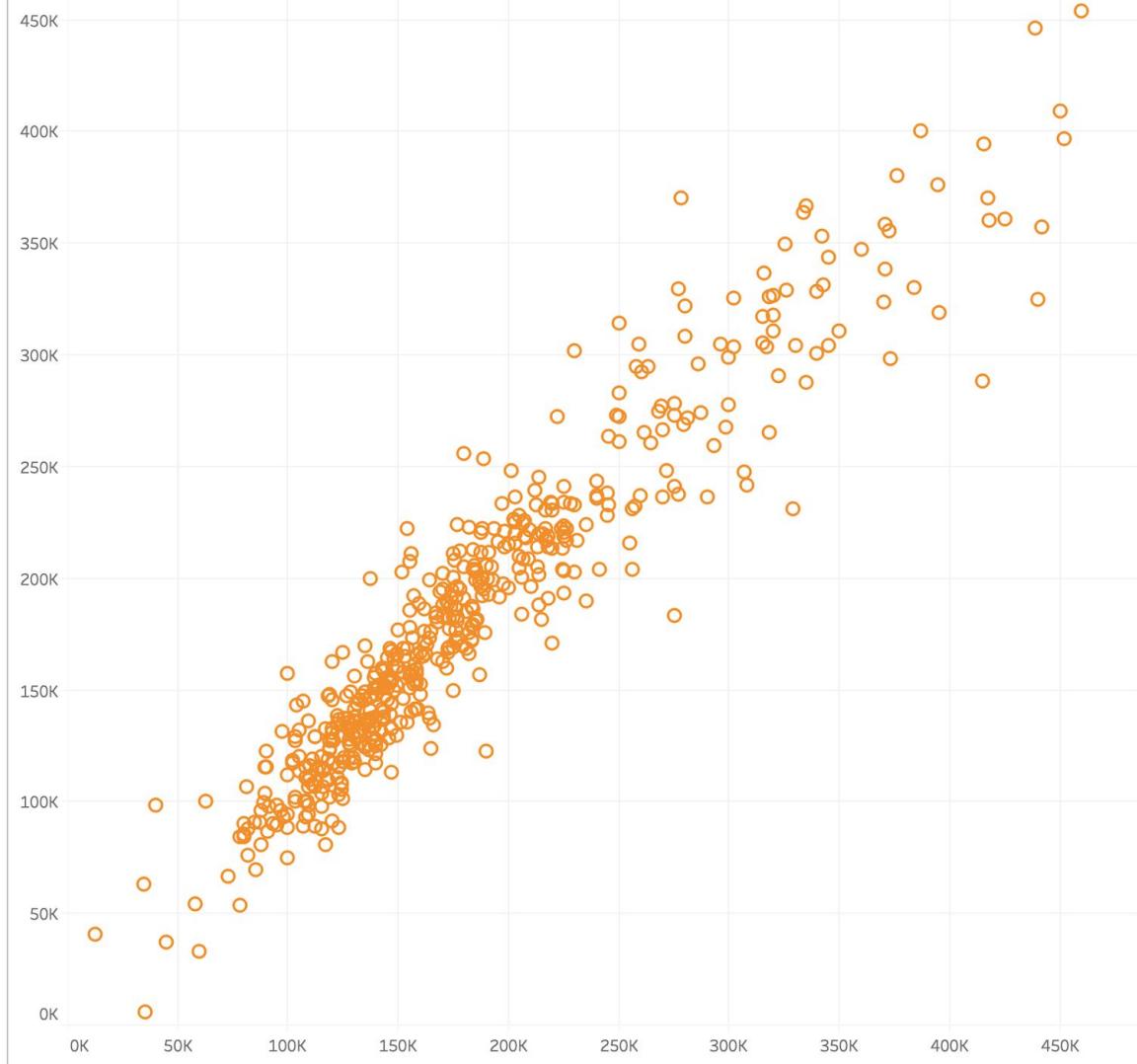


Linear Regression

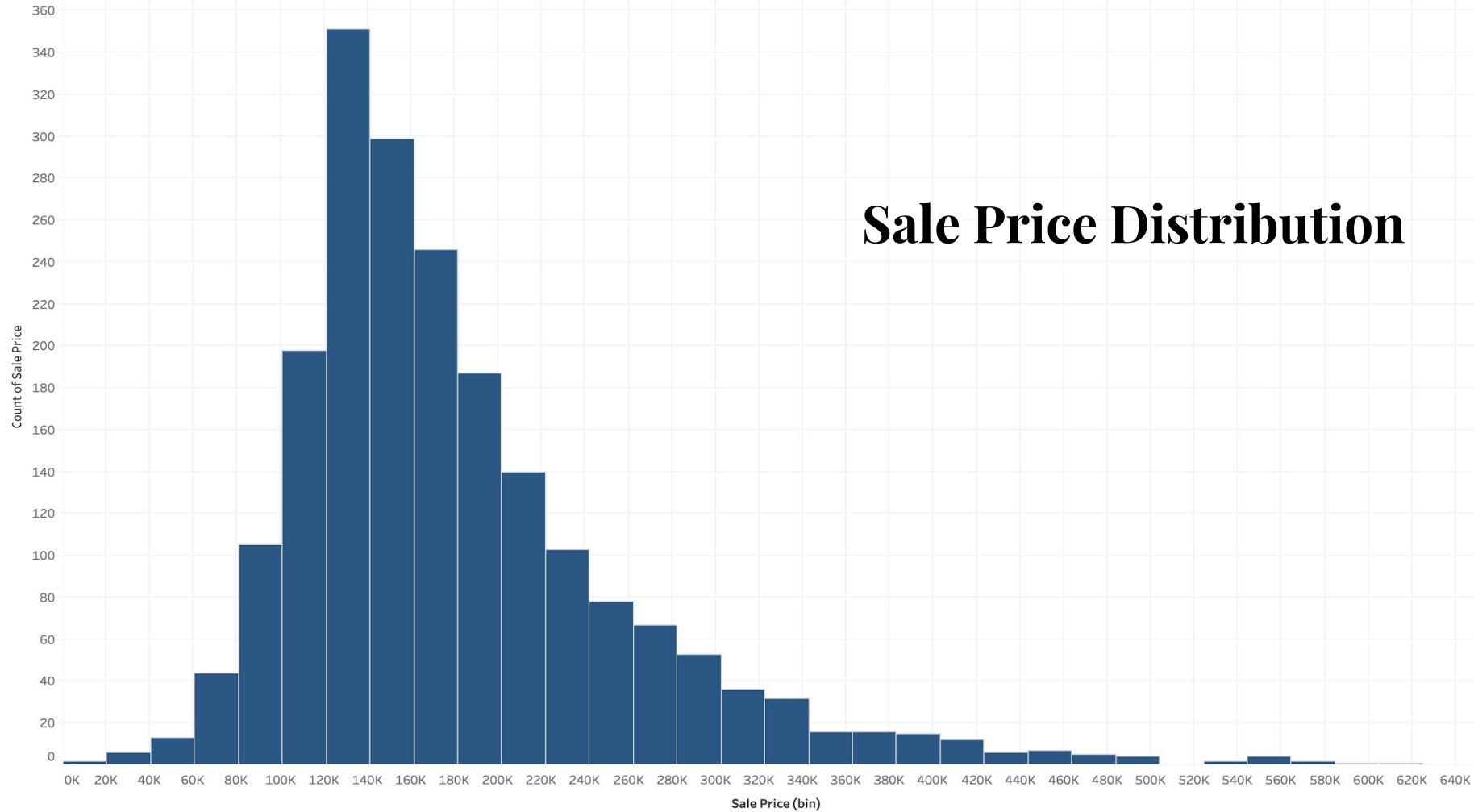
Predicted vs. True Sale Price

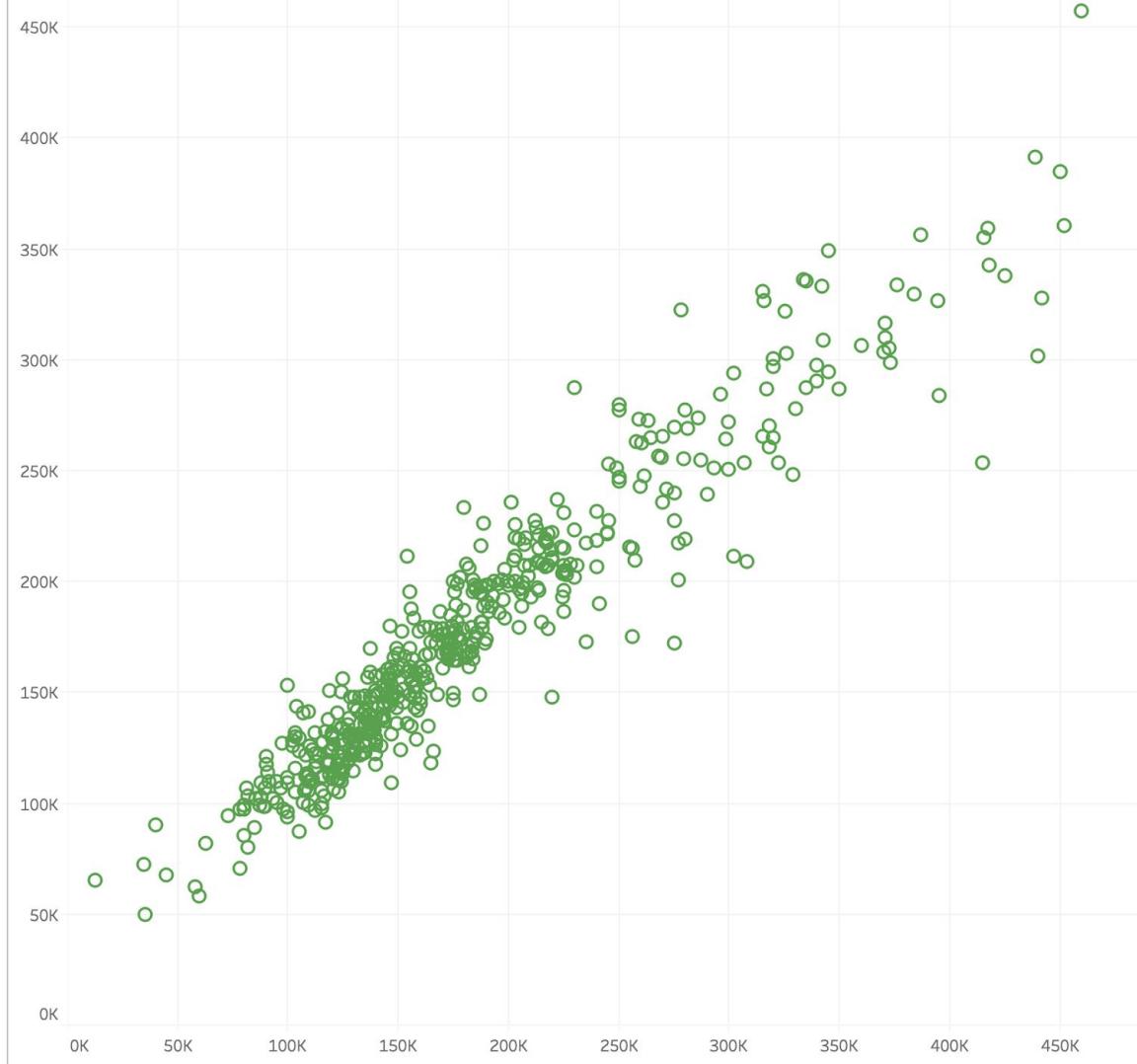
Lasso CV

Predicted vs. True Sale Price



Sale Price Distribution





Lasso CV W/Transform

Predicted vs. True Sale Price

Above grade living area

- Above grade living area is in square feet.
- As it increases by 1 square foot, the sale price increases by 29¢.
- As it decreases by 1 square foot, then in turn the sale price decreases by 29¢, naturally.

Overall Quality

- Quality has rating of 1 - 10; 1 being worst.
- As Quality increases by 1 rating, the sale price increases by about 21¢.
- As Quality decreases by 1 rating, then in turn the sale price decreases by 21¢, naturally.

Year Built

- The year the house was built.
- As the house becomes younger by each year, the sale price raises by about 14¢.
- As the house becomes older by each year, the sale price drops by about 14¢.

Overall Condition

- The overall condition rating from 1-10; 1 being worst.
- As the house increases its rating by 1 in overall condition, the sale price raises by about 11¢.
- As the house decreases its rating by 1 in overall condition, the sale price drops by about 11¢.

In conclusion there are many factors that affect the sale price in Ames, Iowa.



Photo by [Paul Hanaoka](#) on [Unsplash](#); Photo by [Arnel Hasanovic](#) on [Unsplash](#); Photo by [Ragnar Vorel](#) on [Unsplash](#)

Some of the more important features were *Above Grade Living Area*, *Overall Quality*, *Year Built*, and *Overall Condition*.