

An aerial photograph of a suburban neighborhood. In the foreground, a large house with a grey roof and white trim is visible, surrounded by green lawns and trees. A paved road runs diagonally across the middle ground, with a dark-colored car parked on the side. The background shows a dense residential area with many similar houses and trees under a clear blue sky.

Modeling & Analysis of Ames Housing Prices

Project 2: Desmond, Dora & Marianne

Most common phrase of real estate agents?



**“There are three things
that matter in property:
location, location,
location.”**

- Lord Harold Samuel, real estate tycoon in Britain



Location, Location, Location

2000 · Home improvement · 29 seasons

6.9/10

[TV.com](#)

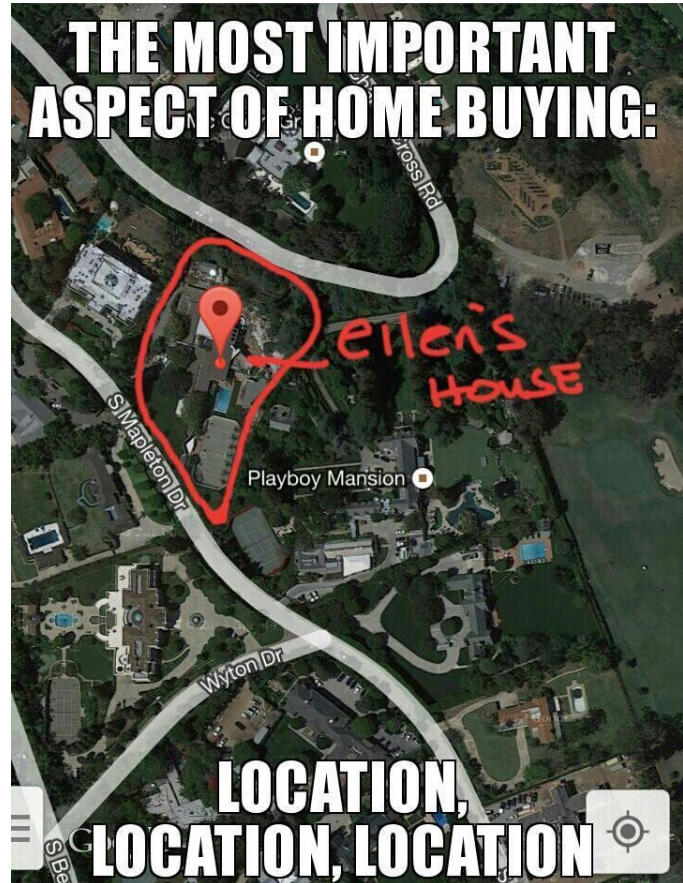
6.4/10

[IMDb](#)

85% liked this TV show

Google users

The country's choosiest home hunters are looking for just the right spot. They're also hoping hosts and real estate agents Kirstie Allsopp and Phil Spencer will help them secure the right home in their desired area. Nevertheless, sometimes sacrifices are required and the customer isn't always satisfied.





Problem Statement

What's the best model for predicting Sale Price for houses in Ames?

- Apart from the location, what features add the most value to the house, and which hurt the value the most?
- As a homeowner, is there anything you can do to your house to increase the value of your home?



Exploratory Data Analysis





Data Cleaning

Missing Values

- Cross-checked across related variables and replaced null values with 0 or “None”

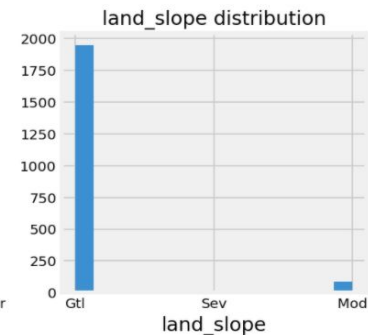
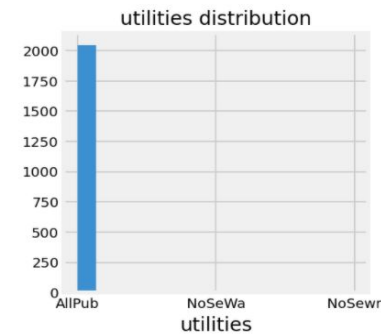
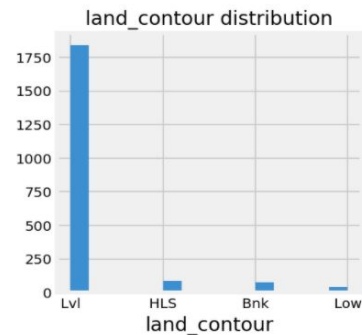
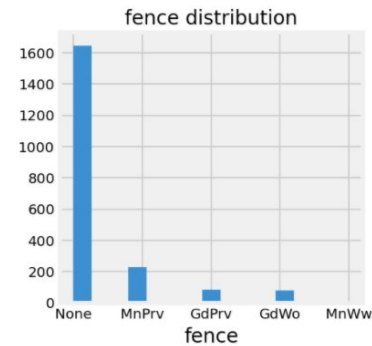
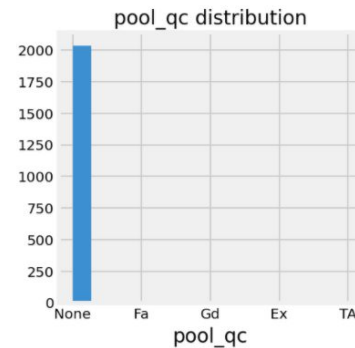
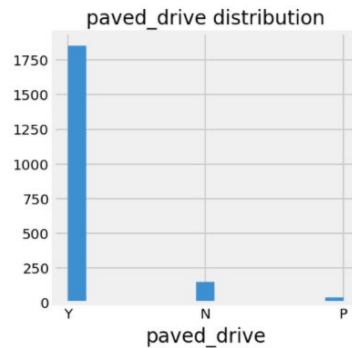
Removal

- Removed outliers from the training set
- Dropped rows with inconsistent values

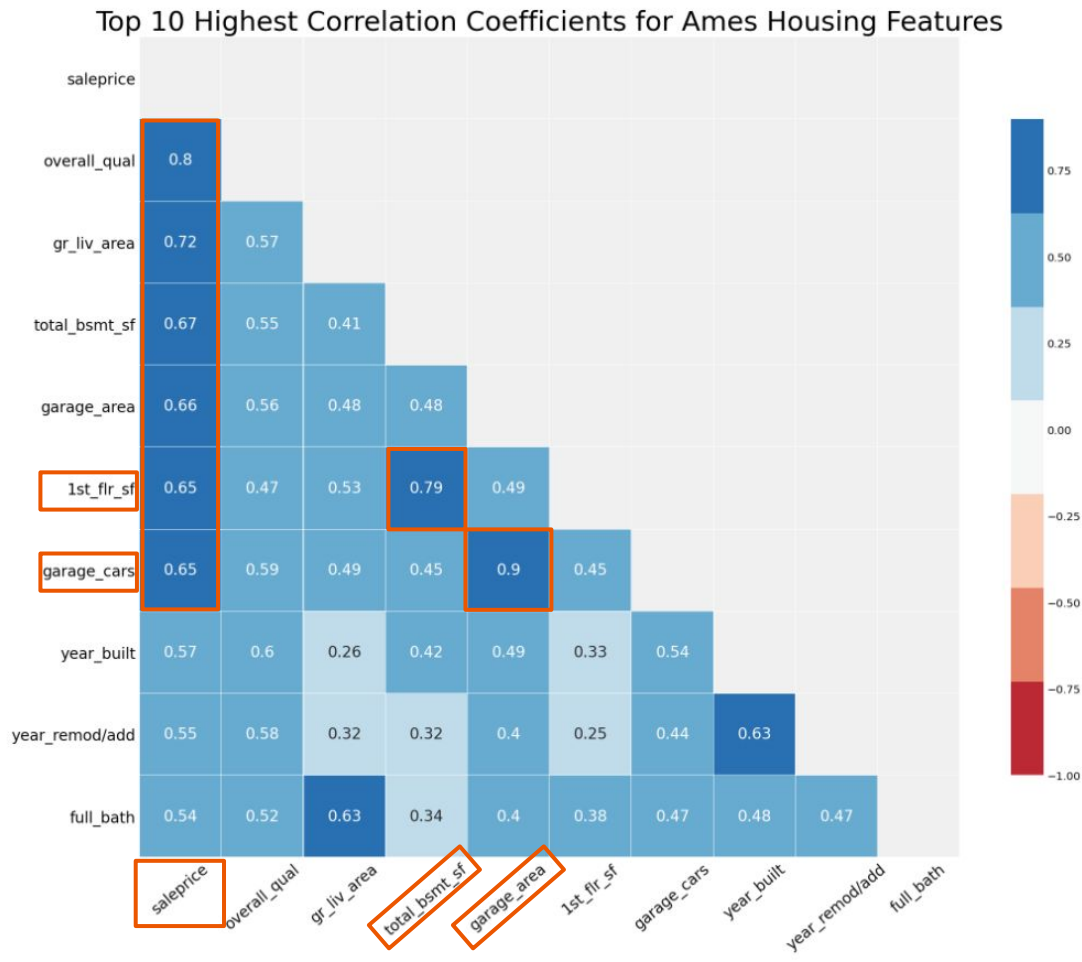
Modification

- Dummy encoded categorical variables
- Converted ordinal categories to numerical variables

Initial Findings: Highly Skewed Variables



Initial Findings: Collinearity

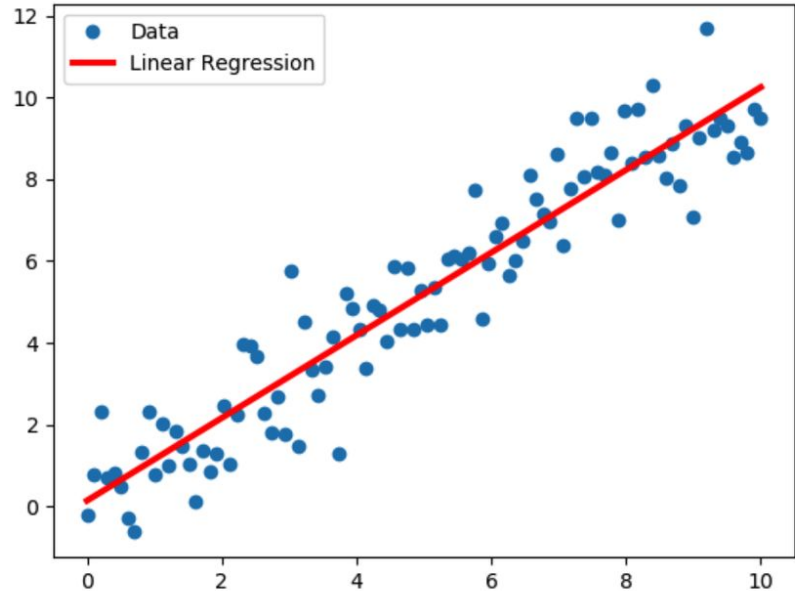




Modelling and Fine-Tuning

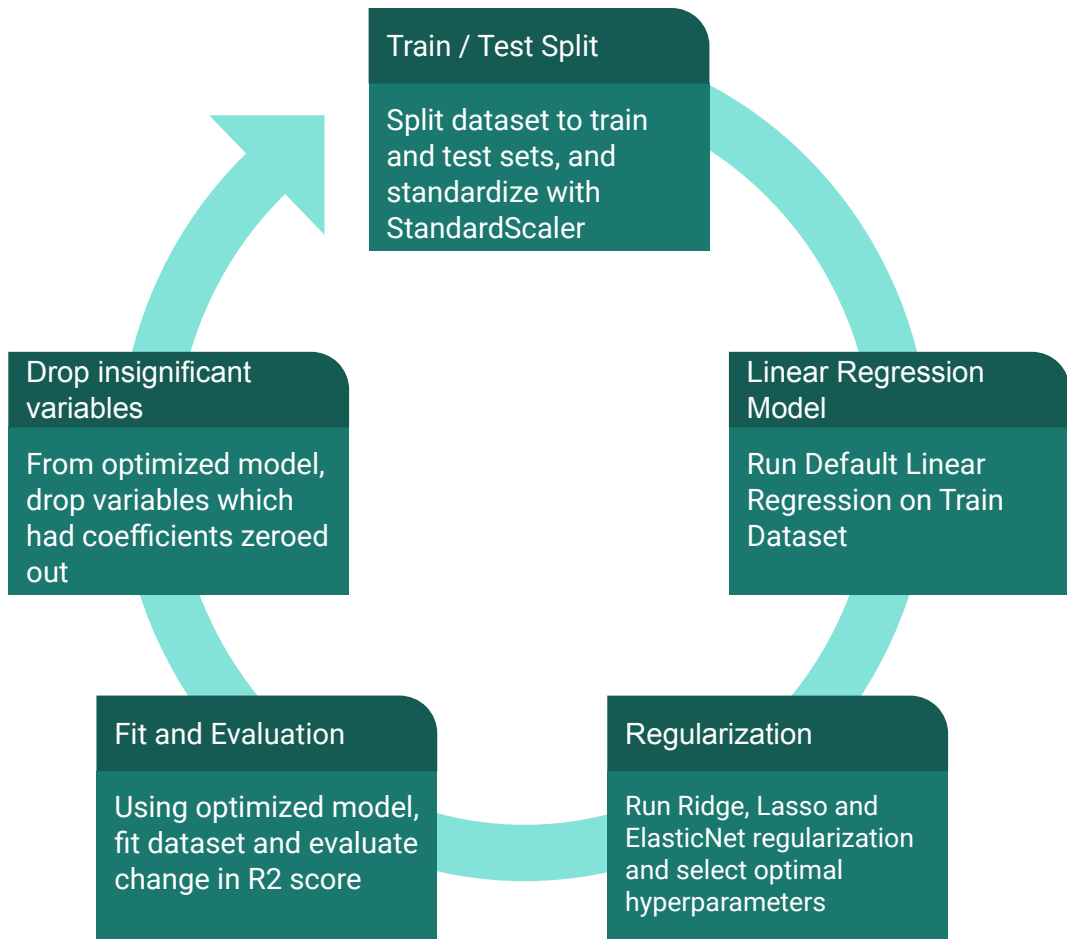
Model: Linear Regression

- **Target variable:** “SalePrice”, which is a numerical variable
- **Predictor variables:** All variables (numerical and dummy variables)



Model Fine-Tuning Process

- In total, we ran through this process up to 3 cycles to obtain our best model





Model Selection

- Default Linear Regression Model displayed signs of overfitting and multicollinearity
- Addressed with Regularization
- Selected model with highest cross-validated R2 score
- **Optimal model: Lasso**

Default Linear Regression Model:
The root mean squared error is 6697088579349205.0
The R2 score is $-6.744328791420067e+22$

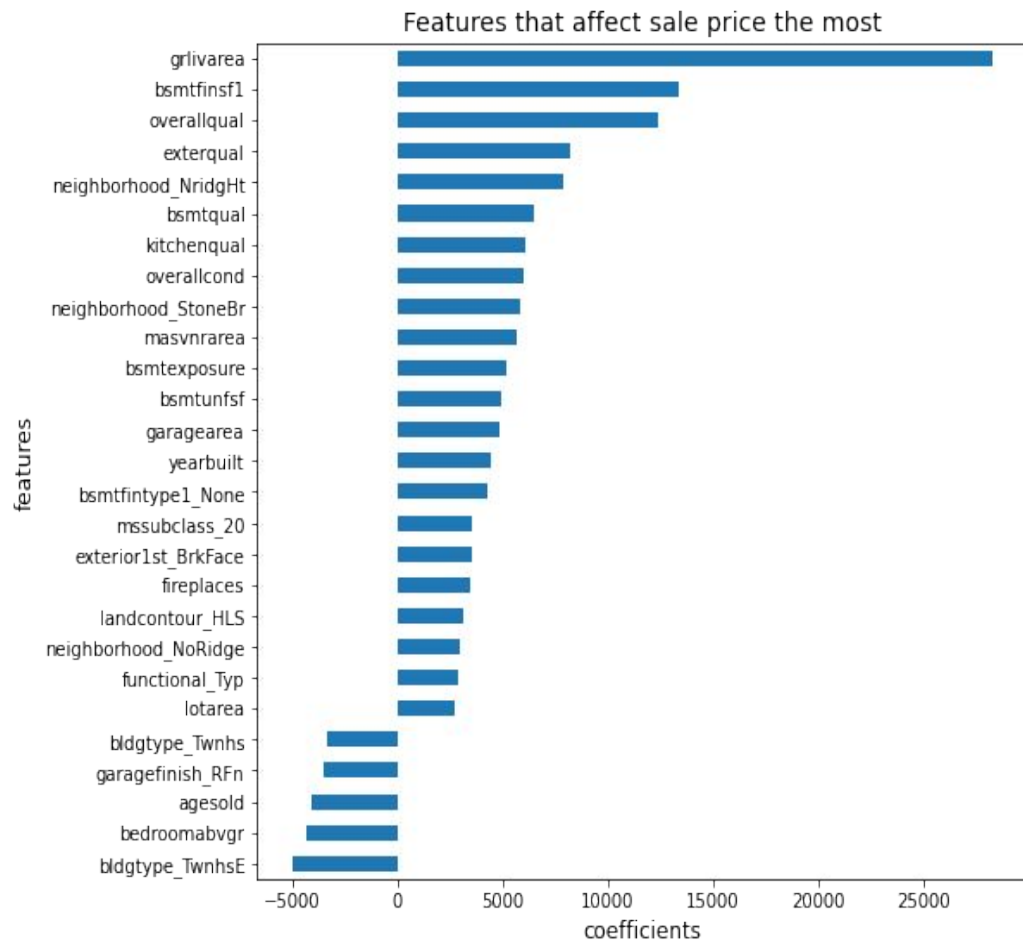
Optimized Ridge Regression Model:
The root mean squared error is 22031.19983392274
The R2 score is 0.9182713100332537

Optimized Lasso Regression Model:
The root mean squared error is 21926.616719628048
The R2 score is 0.9190128564627861

Optimized ElasticNet Regression Model:
The root mean squared error is 21926.69298383955
The R2 score is 0.9190136235314507

















Important Variables

- Largest coefficients = Important variables
- Top 30 coefficients (absolute values) were presented
- Model was able to zero out more than half of the initial variables included



Model Performance

- Private score: 28415
- Our model performance was just outside of top 15 on leaderboard

#	Δpub	Team Name	Notebook	Team Members	Score ?	Entries	Last
1	—	Stack_S			0.00000	2	2y
2	▲ 38	Mason Childress			22200.10...	9	2y
3	▲ 34	KevinWA			22217.46...	27	2y
4	▲ 27	Varan Satchi			23253.82...	2	2y
5	▲ 10	Alireza Karimi			26130.00...	16	2y
6	▲ 54	Kyle Sokolis			26838.80...	4	2y
7	▼ 4	Eamon Fleming			27116.10...	23	2y
8	▲ 55	Tri_NN			27134.54...	7	2y
9	▲ 33	Cameron Bronstein			27768.64...	21	2y
10	▲ 46	Sarah Scolnik			27924.67...	8	2y
11	▲ 27	sung lee			28040.46...	14	2y
12	▲ 27	Chris Shaw			28289.95...	7	2y
13	▲ 19	Johnny McGregor			28332.33...	21	2y
14	▲ 13	Drew Hoppes			28332.96...	11	2y
15	▲ 15	Shawn M			28350.12...	31	2y
16	▲ 48	Ryan Dorris			28539.63...	4	2y



Insights & Recommendations



Insights

POSITIVE effect on sale price

- Size
- Condition and quality
- Functional
- Single storey homes
- Recently built
- Located in Northridge Heights, Stone Brook, Northridge
- Hillside property
- Presence of fireplace, brick exterior

NEGATIVE effect on sale price

- Townhouses
- Too many bedrooms
- Unfinished features
- Poor quality



Recommendations for homeowners

Improving quality of home features

Improve condition and
quality of finishing of
kitchen, basement,
garage and external areas

Ensure house is in finished condition

Ideally, house should be
in move-in condition
when shown to
prospective homebuyers

Installing features to keep house warm

Install fireplaces and
brick exterior for house

An aerial photograph of a suburban neighborhood. In the foreground, there are two large houses with grey roofs and green lawns. A paved road runs diagonally from the bottom right towards the center. A dark-colored car is parked on the side of the road. The background shows a dense residential area with many smaller houses and trees under a clear blue sky.

Thank you
Any questions?