

# Ames Housing Project, Iowa



“ I like those ~~odds~~  
features.”

Prediction model and  
recommendations

# Problem Statement

Prospective homeowners are faced with a deluge of considerations when it comes to decisions regarding property purchase. Some of these considerations might include:

- 1) What is the condition of the property?
- 2) How old is the property?
- 3) Will the value of my property rise? Is this dependent on what kind of future development will take place near my area (e.g. Industrial/Commercial, etc.)?

We want to help prospective homeowners in Ames, Iowa to make informed decisions about which property they should consider purchasing for best value. We also want to help prospective sellers to determine how to maximise the value of their sales.

Specifically, our problem statement is as such:

**Which features of a property will have an effect on the sale price?**

# Our methodology

We use the dataset of residential property sold in Ames, Iowa from 2006 to 2010. The data dictionary can be found [here](#).

## Data Cleaning

Grouped the data into categories and filled the null values

## EDA

Performed exploratory data analysis to examine relationships between features. (using correlation and looking at their distributions)

## Feature Transform

Transformed other features to minimise multicollinearity by adding them together or dropping them

## Model Eval

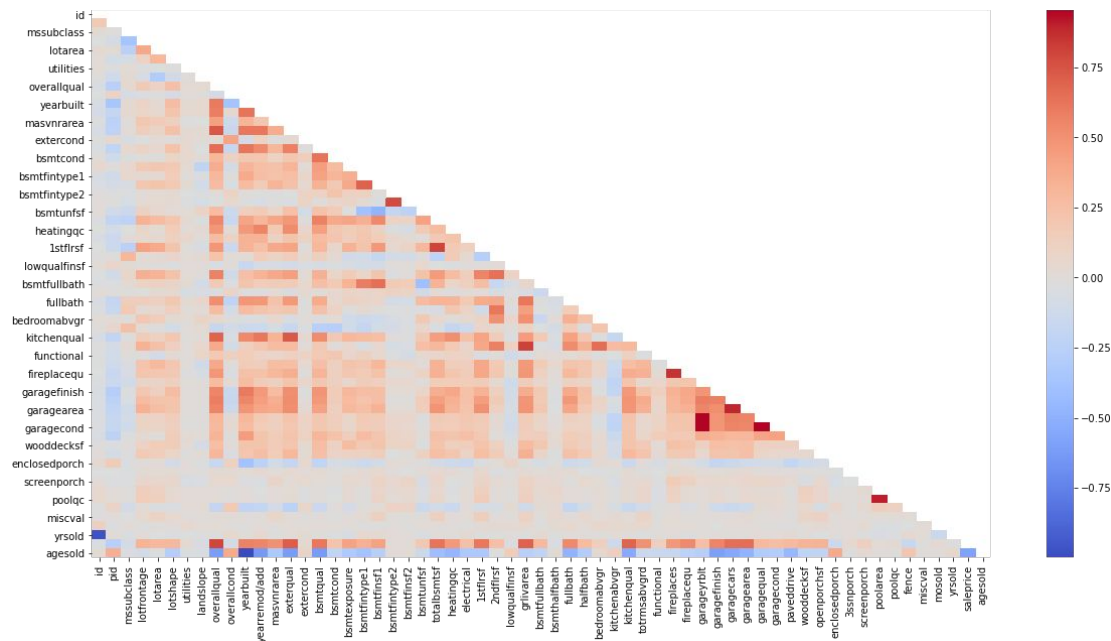
Built and evaluated Linear, Ridge and Lasso models using a preselected set of 23 features (117 after encoding categorical features)

## Production model

Built final production (Ridge) model using lassoed features.

# Findings

## EDA - a lot of strong correlation between independent variables



- ❖ exterqual and kitchenqual (around 0.73),
- ❖ 1stflrsf and totalbsmtsf (around 0.8)
- ❖ garagecars with garagearea (around 0.89)
- ❖ Implies multicollinearity!

Some features were combined, e.g. quality, living area related features while others were dropped.

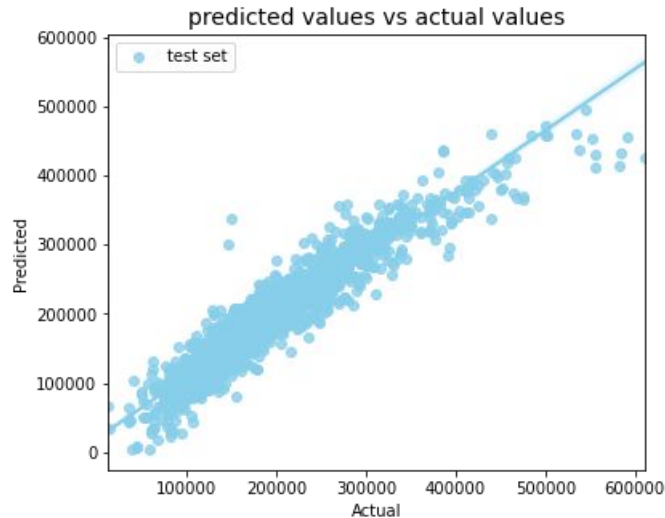
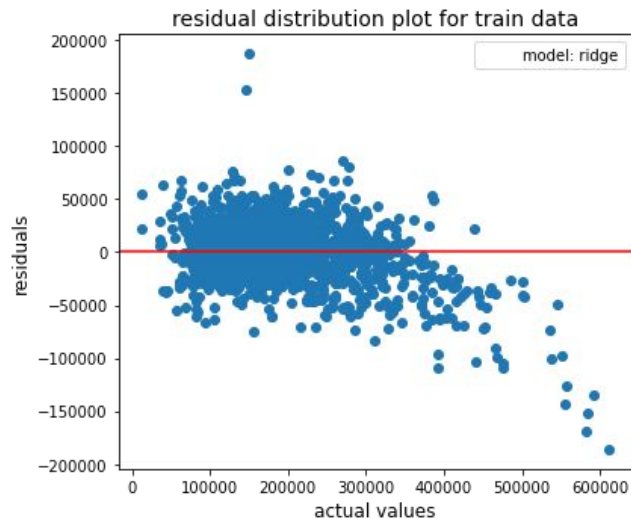
# Findings

**Model Eval - Ridge had lowest RMSE and highest r2 scores**

<b>Model</b>	<b>R2</b>	<b>RMSE</b>
Linear(train)	0.88628	5.178110585812386e+16
Linear(test)	0.88997	27556.2398
Lasso(train)	0.88938	27809.8671
Lasso(test)	0.88636	27546.1351
Ridge (train)	0.89351	27625.5983
Ridge (test)	0.89087	26994.0547
Dummy(baseline)	-0.0003877	81732.4755

# Findings

**Production Model – Our final ridge regression model seems pretty robust**



- ❖ Data seems well distributed, tending to cluster towards the middle of the plot. Implies homoscedasticity.
- ❖ Actual vs predicted values scatter plot shows strong correlation between predictions and actual results.
- ❖ Implies model does not seem to be overfitting, and should be able to generalise sale prices well.

# Conclusions and Recommendations -1

**Production Model - Our final ridge regression model can generalise reasonably well on unseen data**

```
In [135]: #RMSE train  
          rmse(ridge, X_final, y)
```

```
Out[135]: ('RMSE:', 27343.120587491725)
```

Your most recent submission

Name	Submitted	Wait time	Execution time	Score
submission_predict.csv	just now	1 seconds	0 seconds	30895.07650

Complete

[Jump to your position on the leaderboard](#) ▼

RMSE vs Kaggle submission RMSE

# Conclusions and Recommendations -2

To address the problem statement, some features that affect sale price of a property are:

- 1) **Neighborhood**: The neighborhood the property is located in.
- 2) **Quality**: The overall finishing and material of the house.
- 3) **Square Feet**: The size of the property.
- 4) **Materials e.g. Foundation, Roof** : The materials used to construct the property

Surprisingly, the age of the property did not influence the sale price as much as we initially assumed. It was zeroed out by lasso.

To ensure that future homeowners purchase property that is value for money, we would recommend them to consider these 4 factors before making a purchase.





THANK YOU.

A rustic wooden sign with the words "THANK YOU." painted in bright yellow, mounted on a wall of corrugated metal and wood. The sign is made of two horizontal wooden planks. The background wall is composed of vertical wooden planks and sections of corrugated metal. Two rectangular openings in the wall, framed by wood, are visible above the sign. The foreground shows a dark, textured surface, possibly asphalt or concrete, with some green grass visible at the base of the wall.