
Ames Housing Prices Prediction

Problem Statement

There are oftentimes mismatch of expectations between homeowners and home buyers. Homeowners look to maximize on the value of their house and home buyers look to get the best features on a tight budget. This project aims to answer some of these questions:

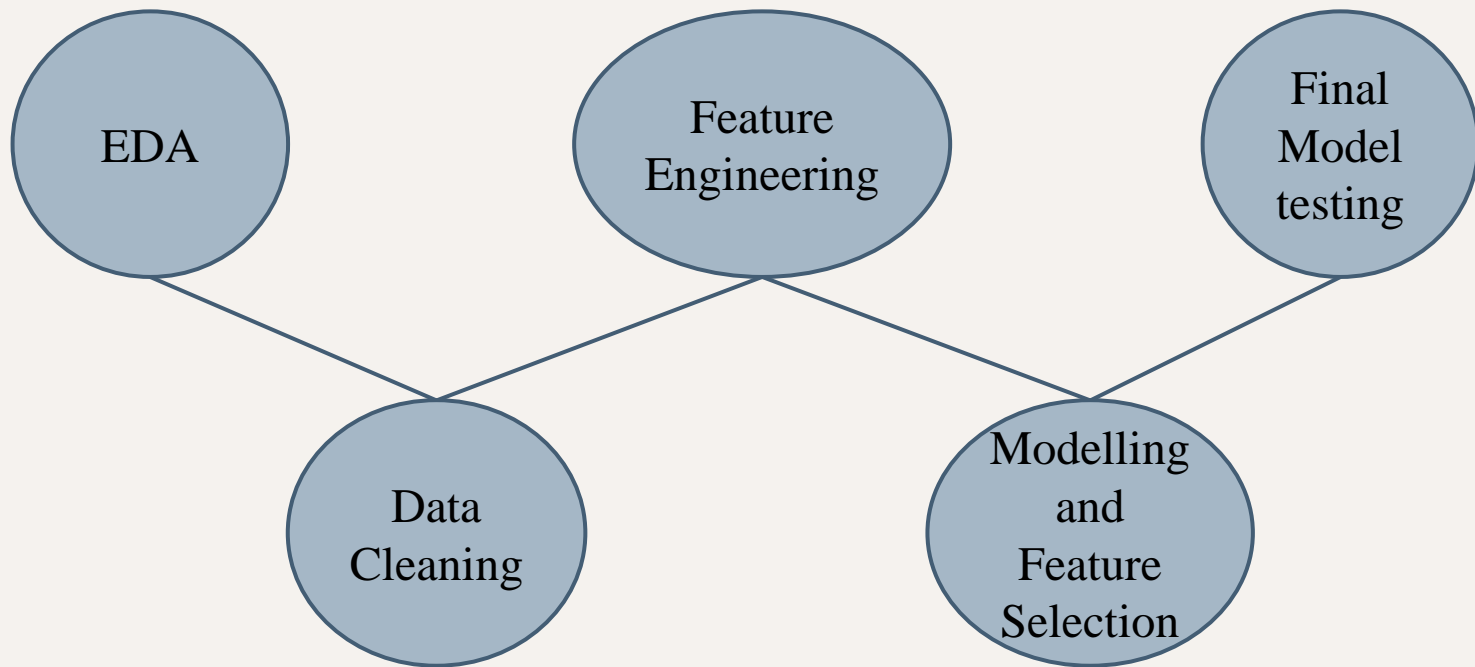
Homeowners

- How can homeowners maximize the value on their house?
- What are the features that add the most value to the house?

Homebuyers

- What kind of house or features would homebuyers be able to afford based on their budget?
 - How much should buyers be paying for houses with certain features?
-

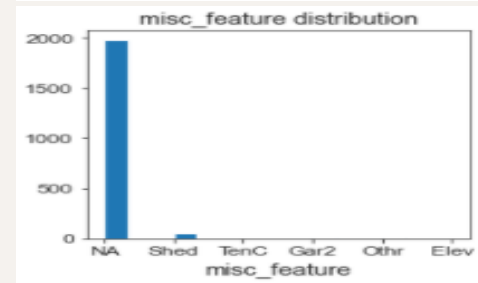
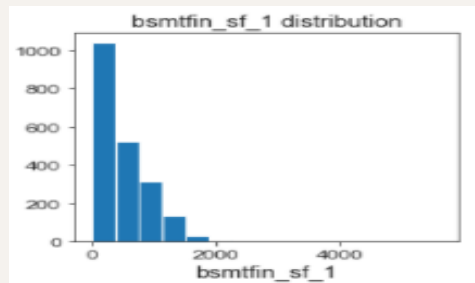
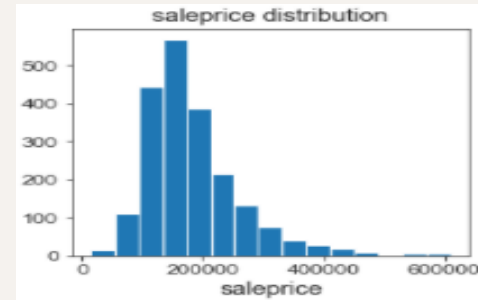
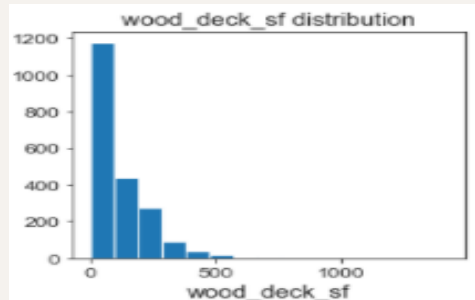
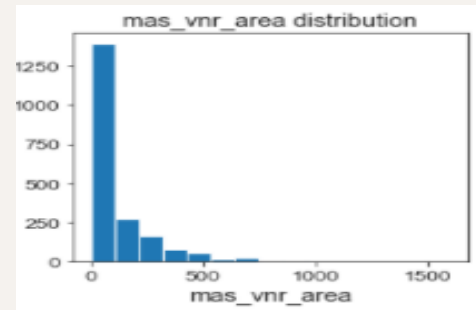
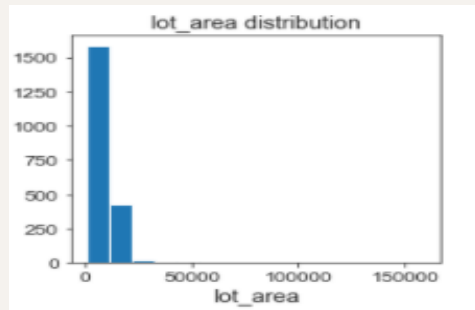
Workflow



Skewness

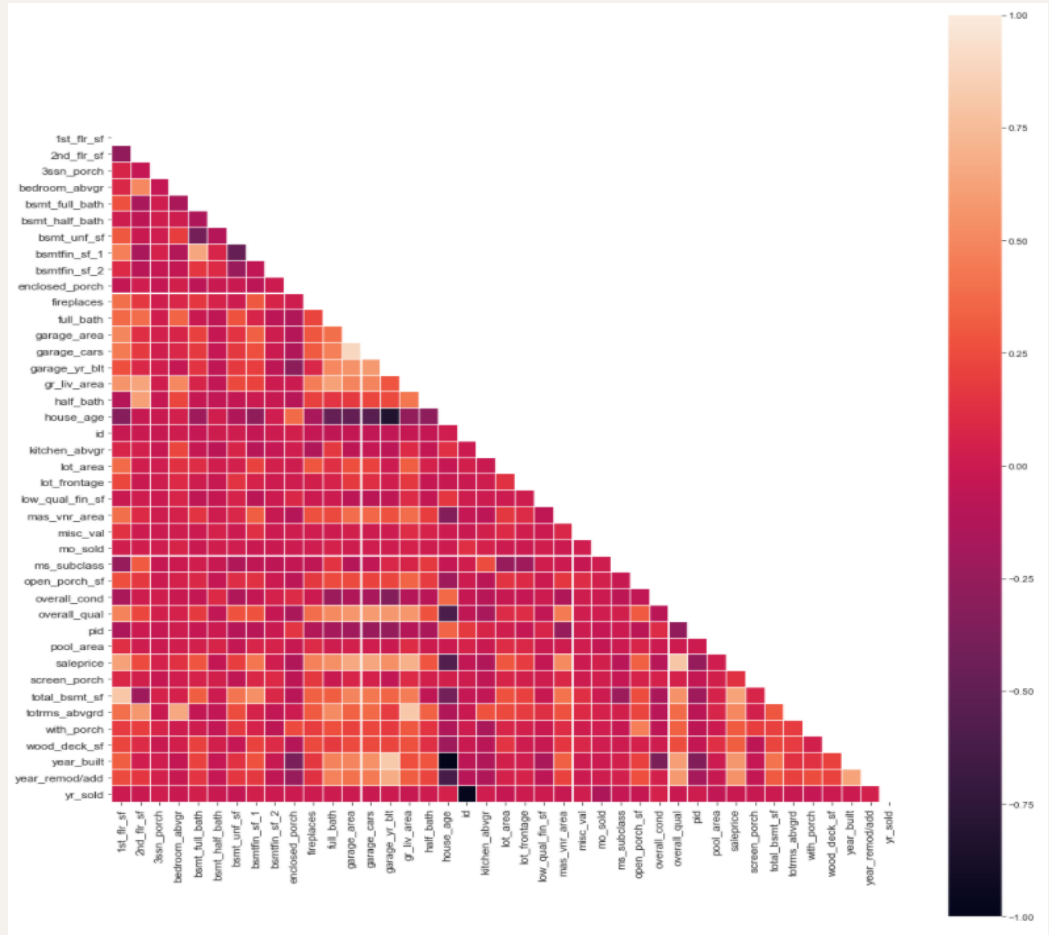
Some features were observed to be heavily skewed.

Sale Price was also observed to be right skewed.



Collinearity

Some level of collinearity were observed between the features eg. number of garage cars and garage area.



Null values

Some features
observed to have high
percentage of null
values.

A null value in the
corresponding feature
likely indicates
absence of the feature
eg. lack of pool, lack
of alley

Feature	Percentage of null values
pool_qc	0.995612
misc_feature	0.988308
alley	0.931741
fence	0.804973
fireplace_qu	0.487567
lot_frontage	0.160897
garage_finish	0.055583
garage_qual	0.055583
garage_yr_blt	0.055583
garage_cond	0.055583
garage_type	0.055095
bsmt_exposure	0.028279
bsmtfin_type_2	0.027304
bsmtfin_type_1	0.026816
bsmt_cond	0.026816
bsmt_qual	0.026816
mas_vnr_area	0.010726
mas_vnr_type	0.010726
bsmt_half_bath	0.000975
bsmt_full_bath	0.000975
garage_area	0.000488
total_bsmt_sf	0.000488
bsmt_unf_sf	0.000488
bsmtfin_sf_2	0.000488
bsmtfin_sf_1	0.000488
garage_cars	0.000488

Data Cleaning / Feature Engineering

Data Cleaning

1. Null values : Imputed to '0', or 'NA'.
2. Removing of outliers

Feature Engineering

1. Dropped columns with high single value counts and high collinearity
 2. Created columns by combining features of the same category eg. 'with_porch' to indicate presence of porch
 3. Converted ordinal data to numerical
 4. Log transformed features with high skewness
 5. Scaling of features due to unequal scales
-

Model Performance

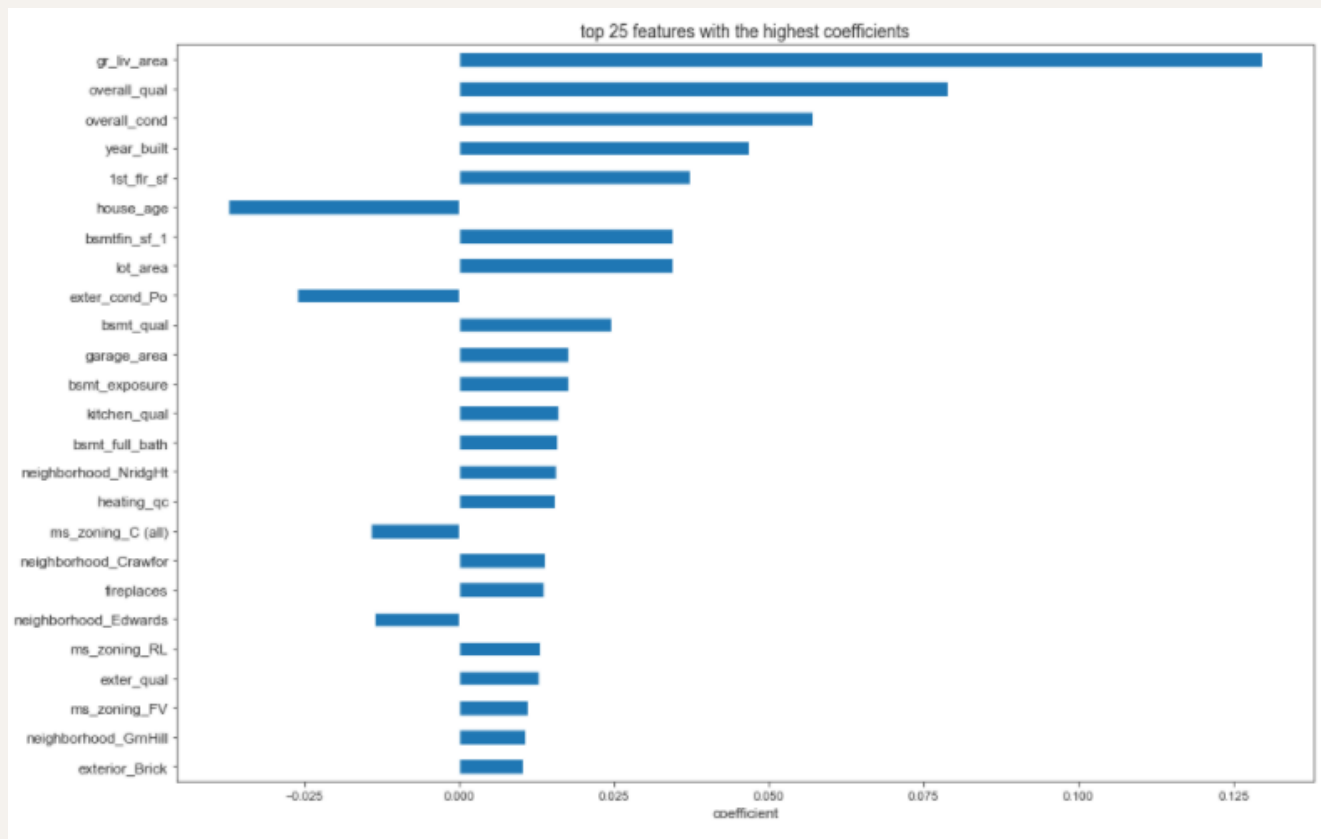
Elastic net was selected for the highest R2 score and the lowest RMSE.

	R2	RMSE
Baseline	-4.263	0.3977
Linear Regression	-4.4210	264433330
Ridge Regression	0.9145	0.1162
Lasso Regression	0.9164	0.1149
Elastic Net Regression	0.9166	0.1148

Final features

From the most important features, we see that many are related to area, quality and location of the house.

The older the house and the poorer the condition, the more the sale price of the house decreases.



Recommendations

- Square feet living area, location and quality of the house are deemed the most important factors in deciding house prices
 - House buyers should look to purchase property in Northridge Heights, Crawford but not Edwards.
 - To maximize the sale price, house owners looking to sell their property should do it sooner rather than later.
 - People looking to sell should do it sooner rather than later
 - To increase the value of a home:
 - Repaint/remodel the interior and exterior finish
 - Renovate the kitchen
 - Add a fireplace and full bath (if not already present)
 - Renovate the house if it had been severely damaged
-