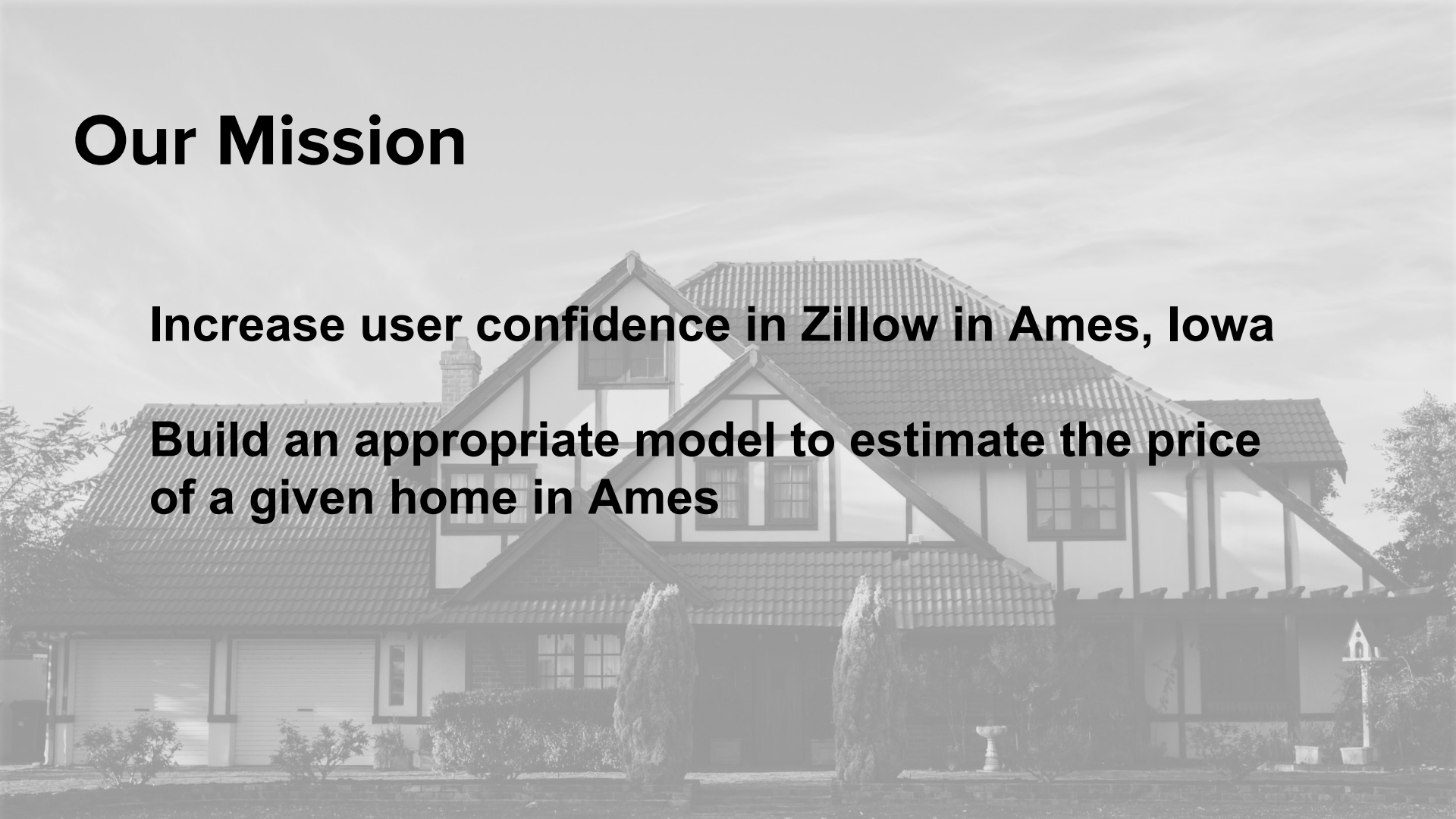
Ames Housing Analysis

Data Science | P2

DAN KIM

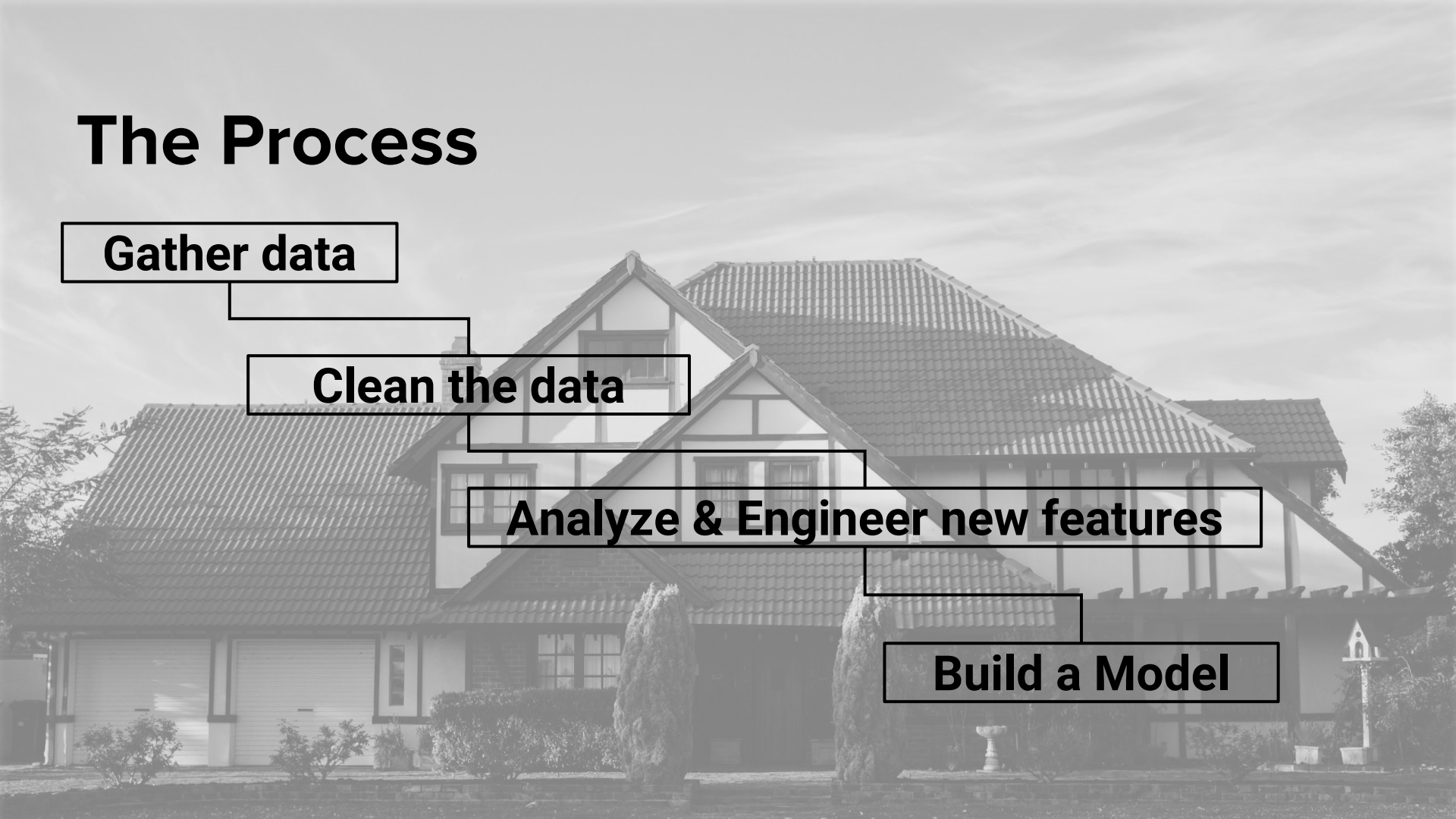
Our Mission



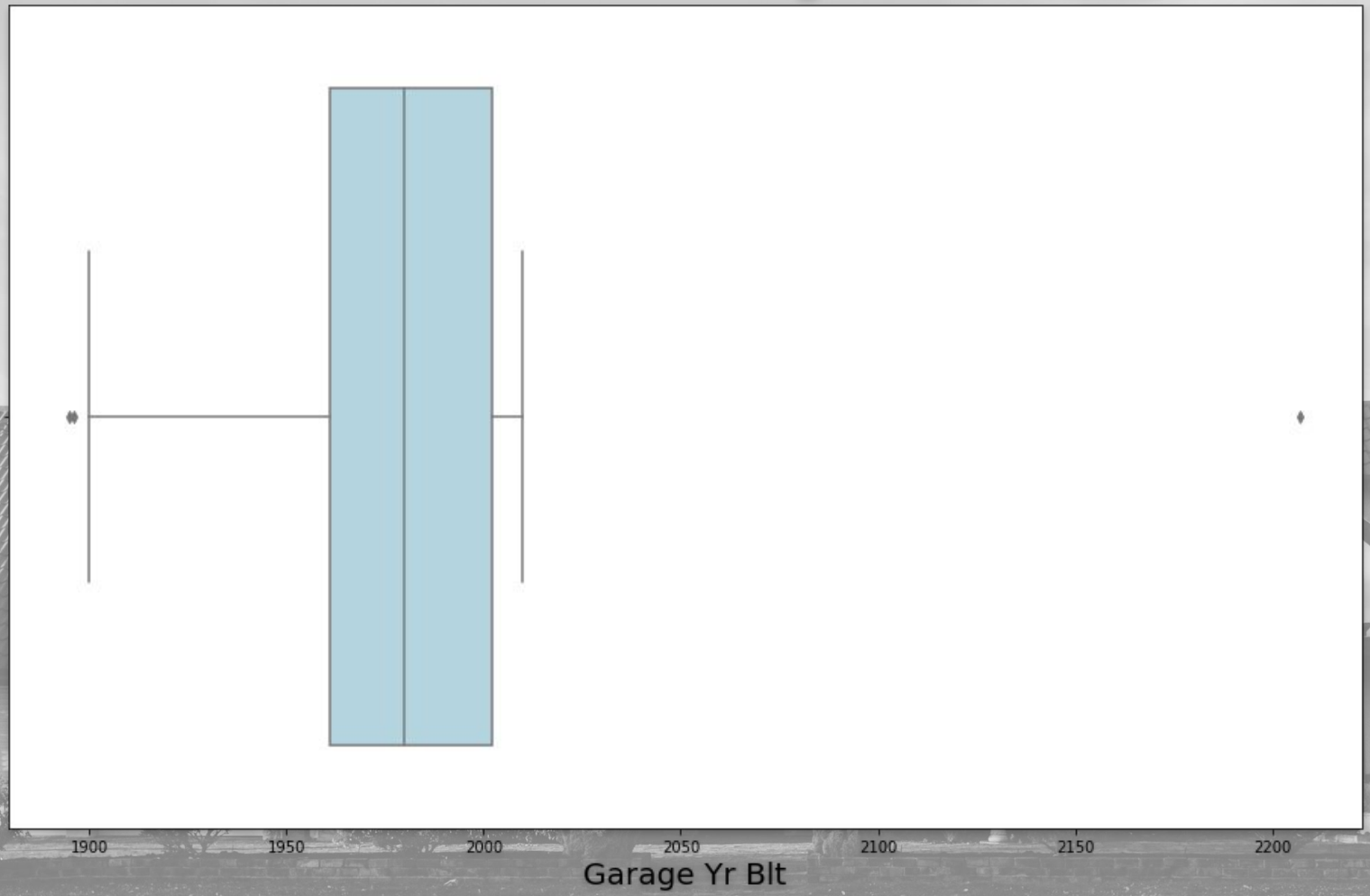
Increase user confidence in Zillow in Ames, Iowa

Build an appropriate model to estimate the price of a given home in Ames

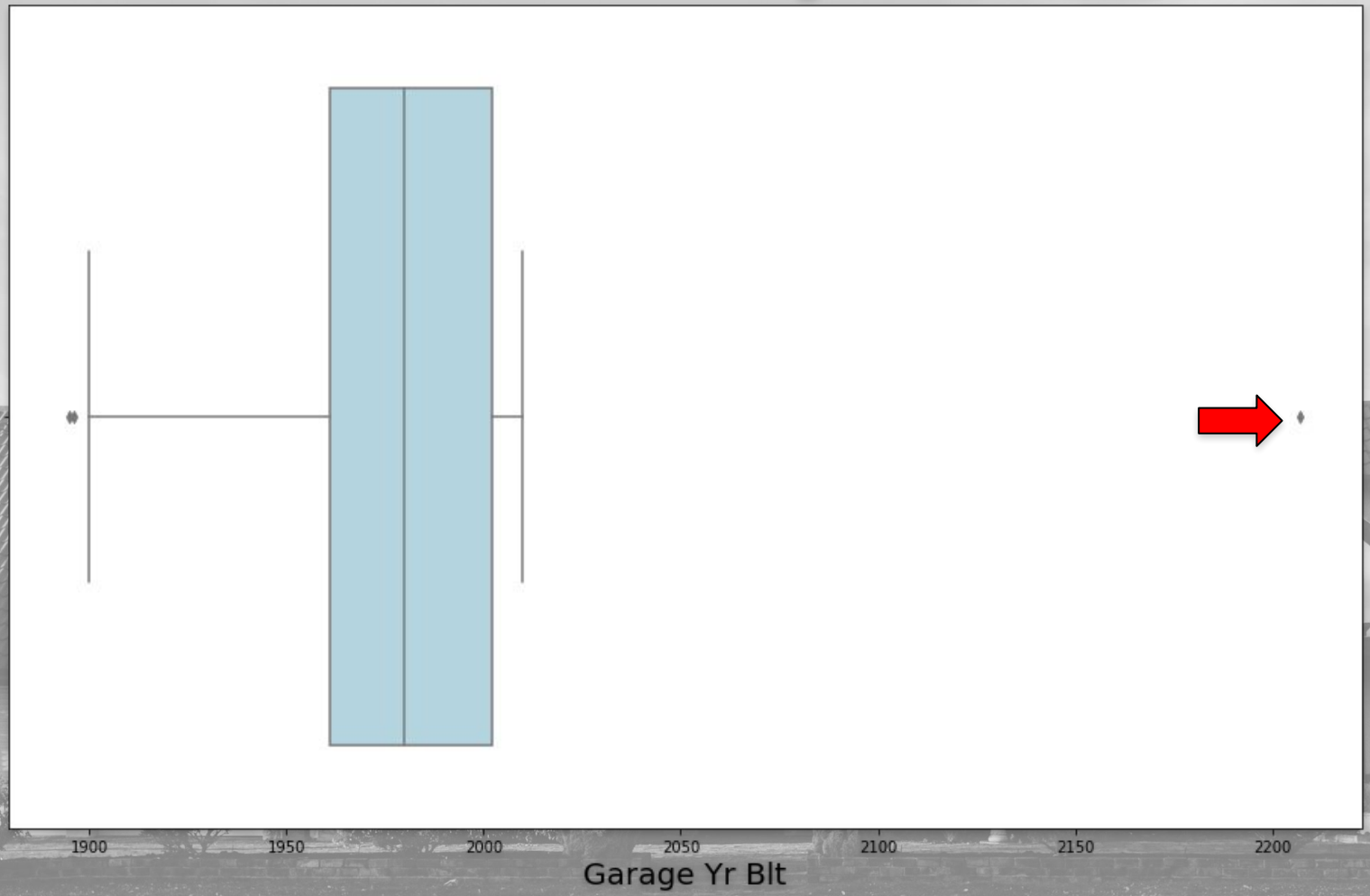
The Process


Gather data
Clean the data
Analyze & Engineer new features
Build a Model

Distribution of Garage Year Built



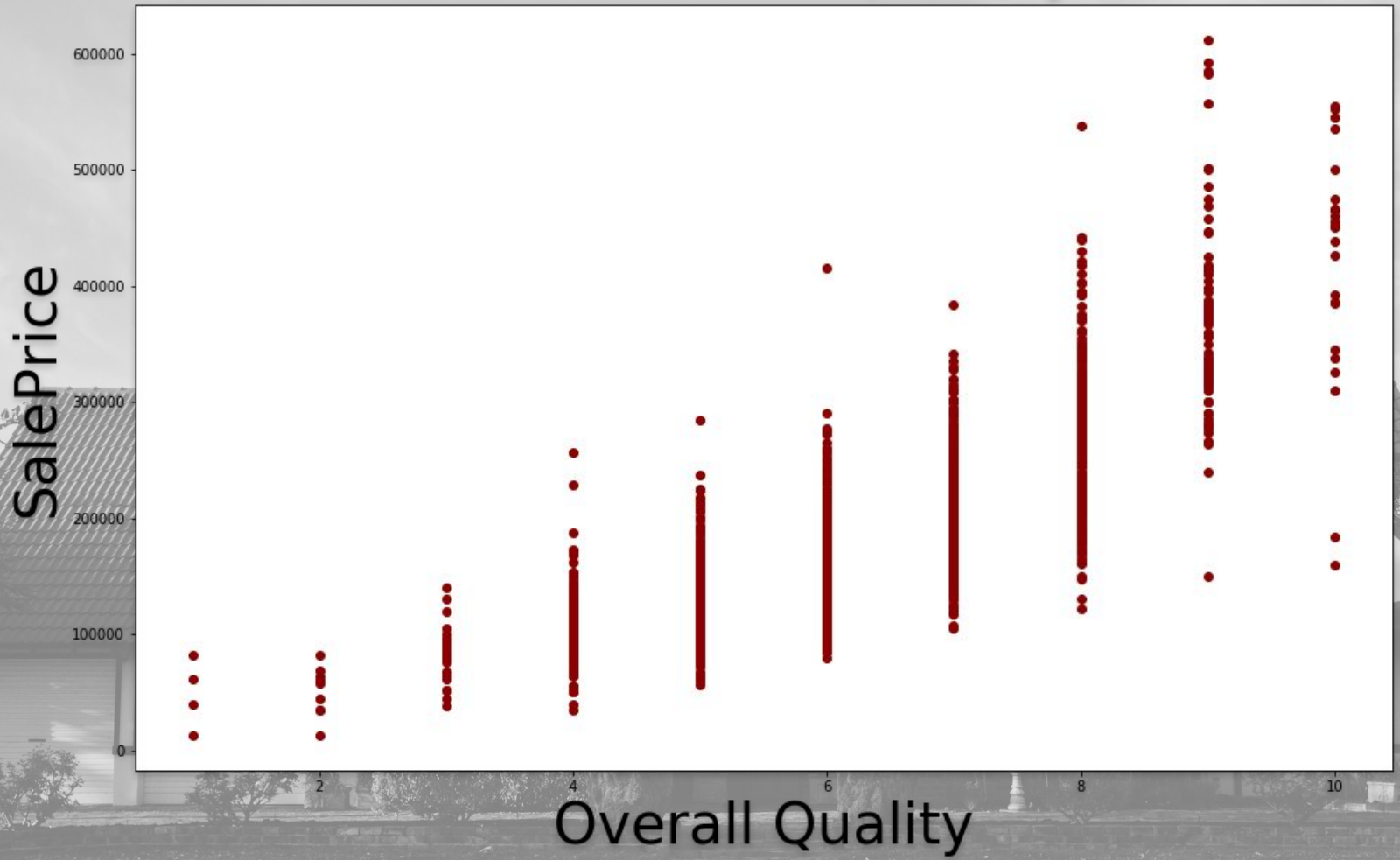
Distribution of Garage Year Built



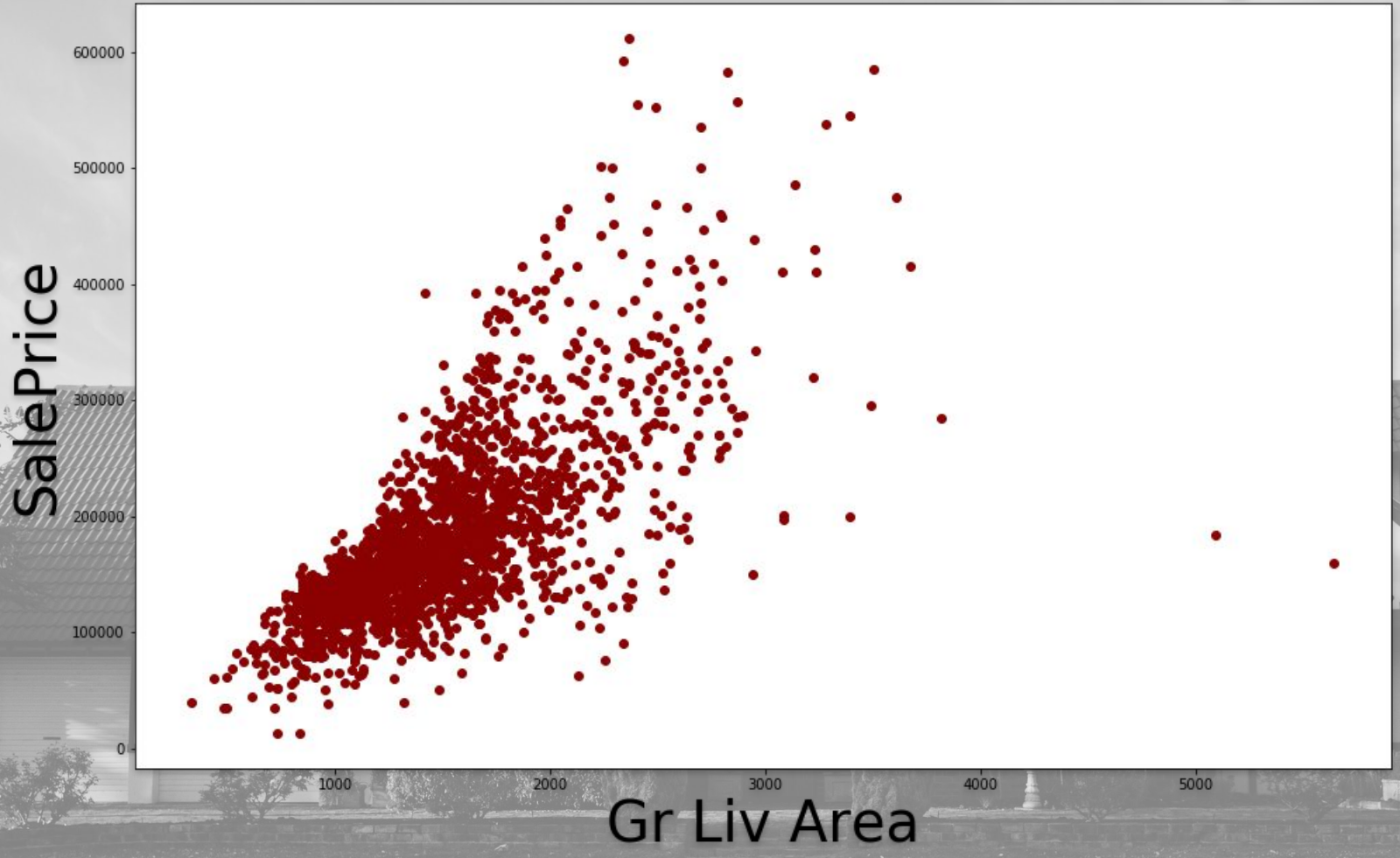
Initial Analysis

- Overall Quality
- Above Grade (Ground) Living Area
- Garage Size (by cars and in sqft.)
- Basement Size (in sqft.)
- 1st Floor Size (in sqft.)
- Number of Full Bathrooms

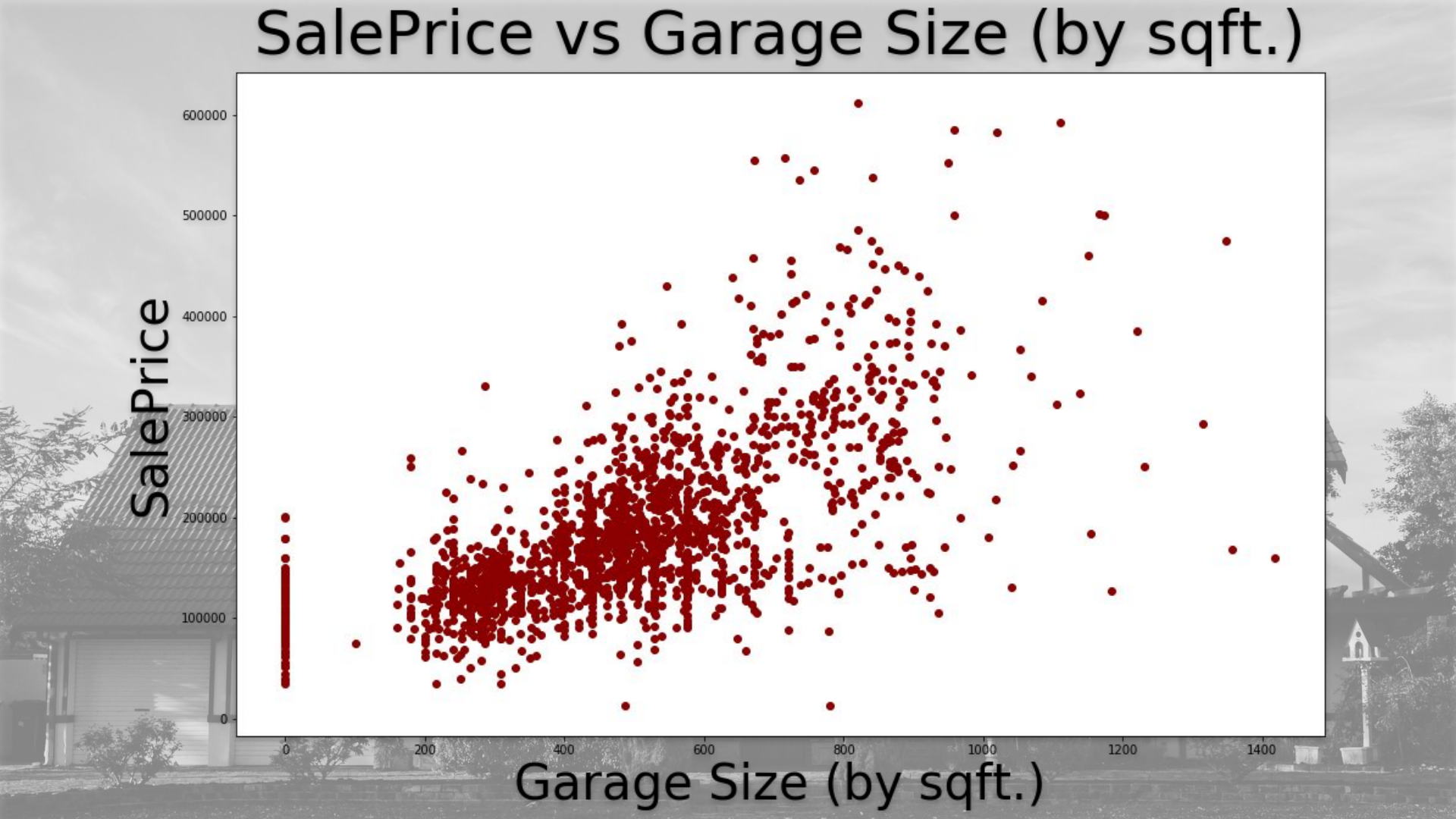
SalePrice vs Overall Quality of House



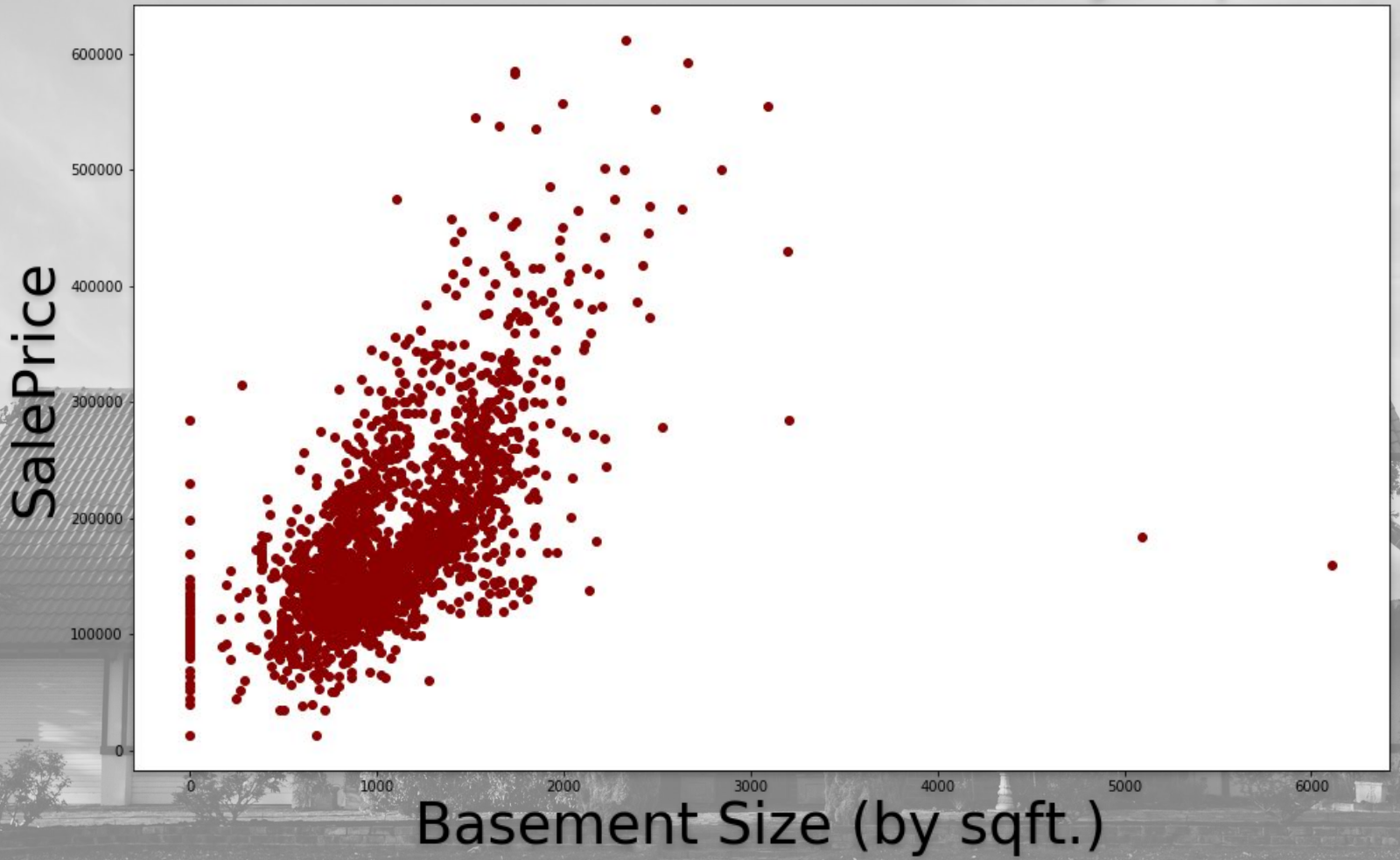
SalePrice vs Above Ground Living Area



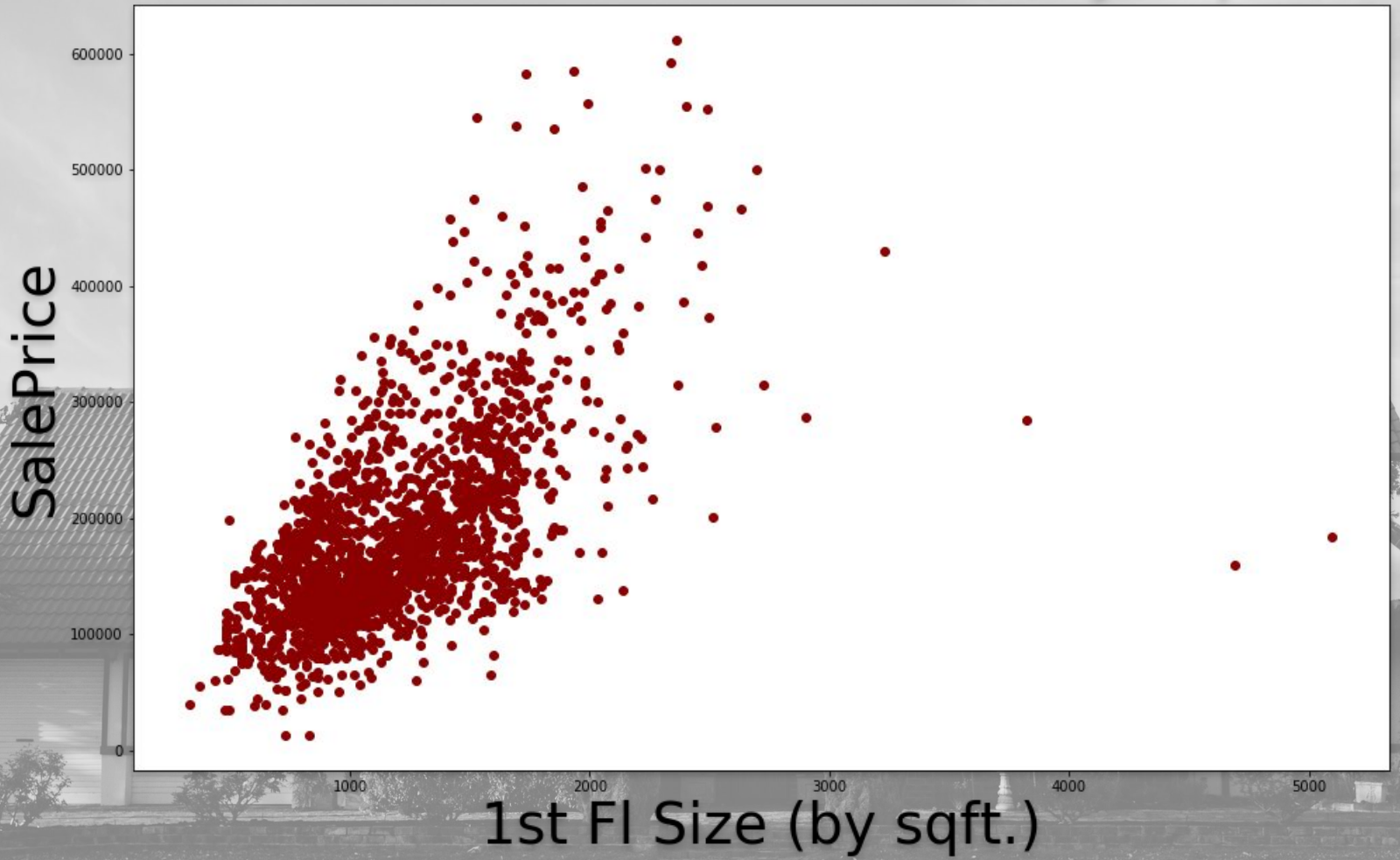
SalePrice vs Garage Size (by sqft.)



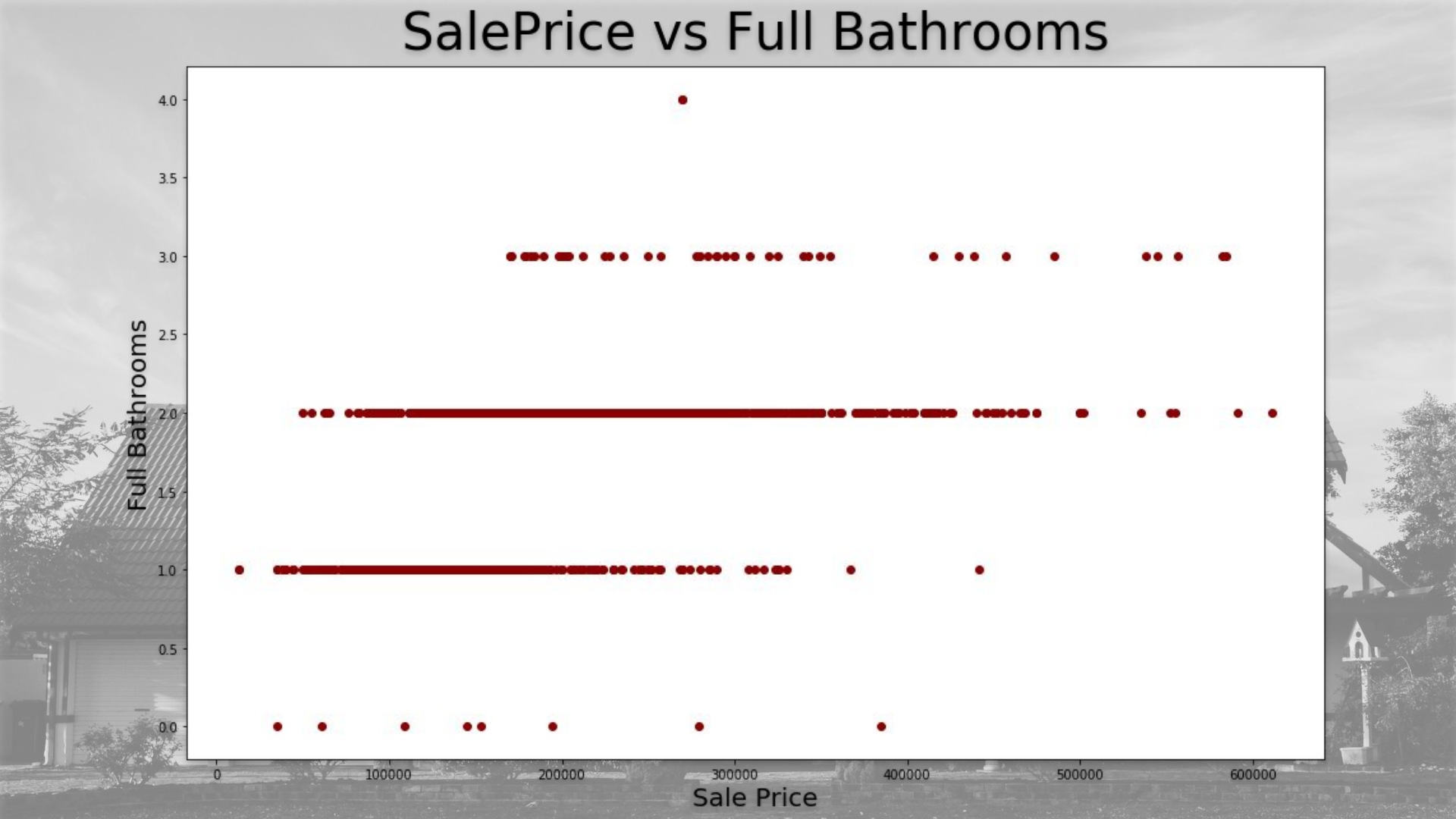
SalePrice vs Basement (by sqft.)



SalePrice vs 1st Fl Size (by sqft.)



SalePrice vs Full Bathrooms



Modeling - Linear Regression

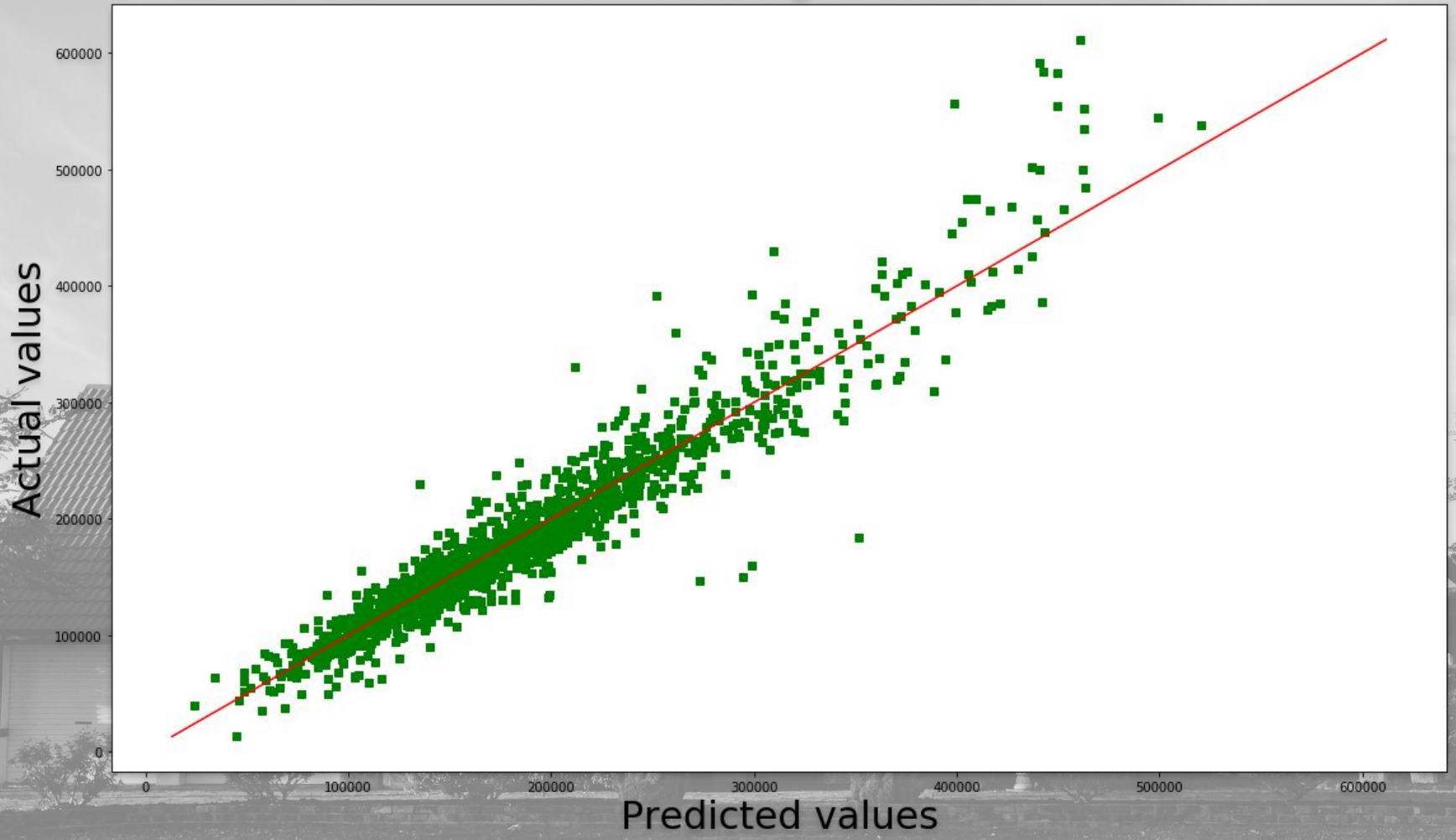
Cross Value Score	R2 Score
0.763333	0.810221

Modeling - Lasso Regression

Moderate Correlation Threshold (0.5)

Cross Value Score	R2 Score
0.837509	0.912826

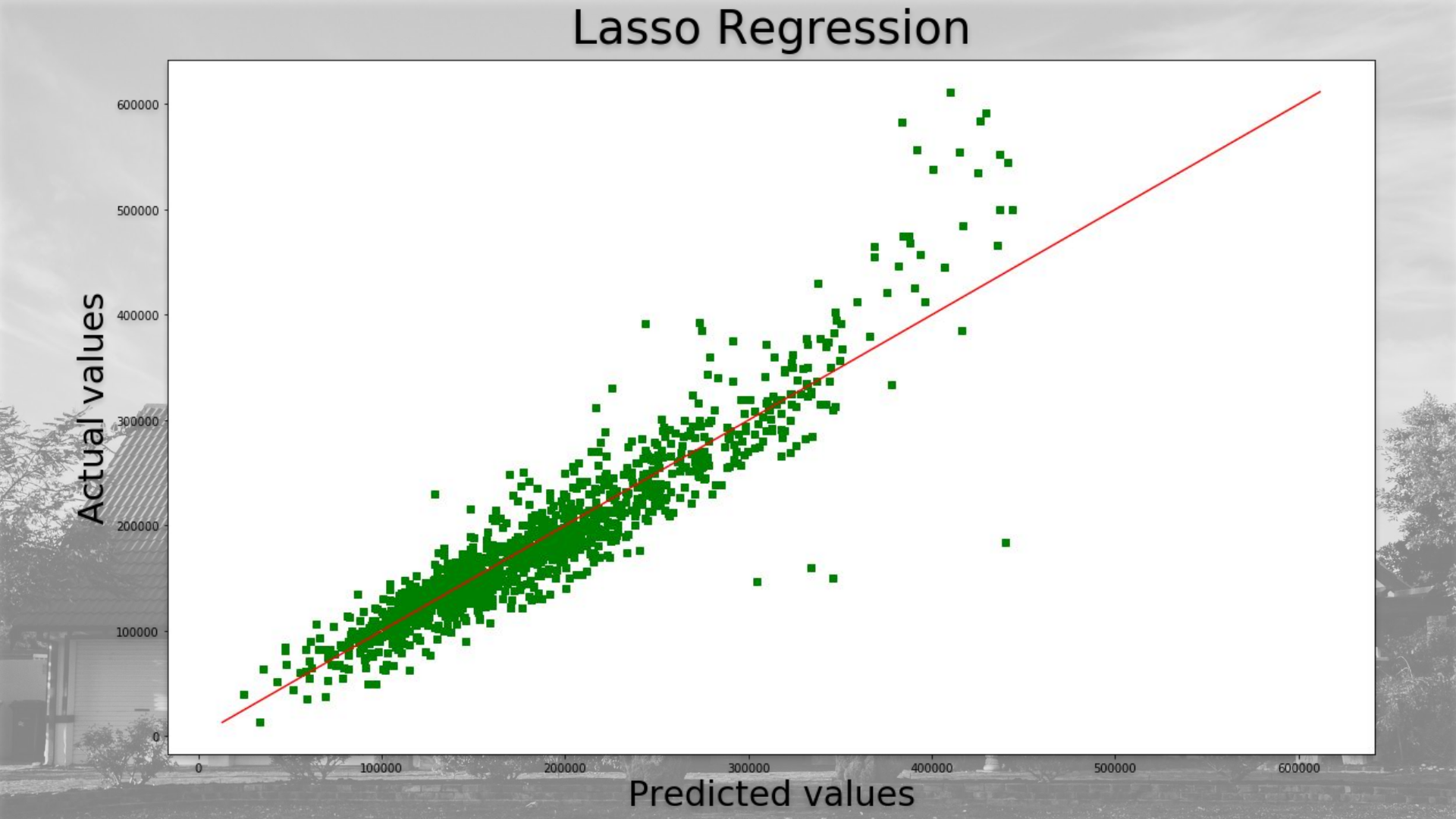
Lasso Regression



Modeling - Lasso Regression

Higher Correlation Threshold (0.6)

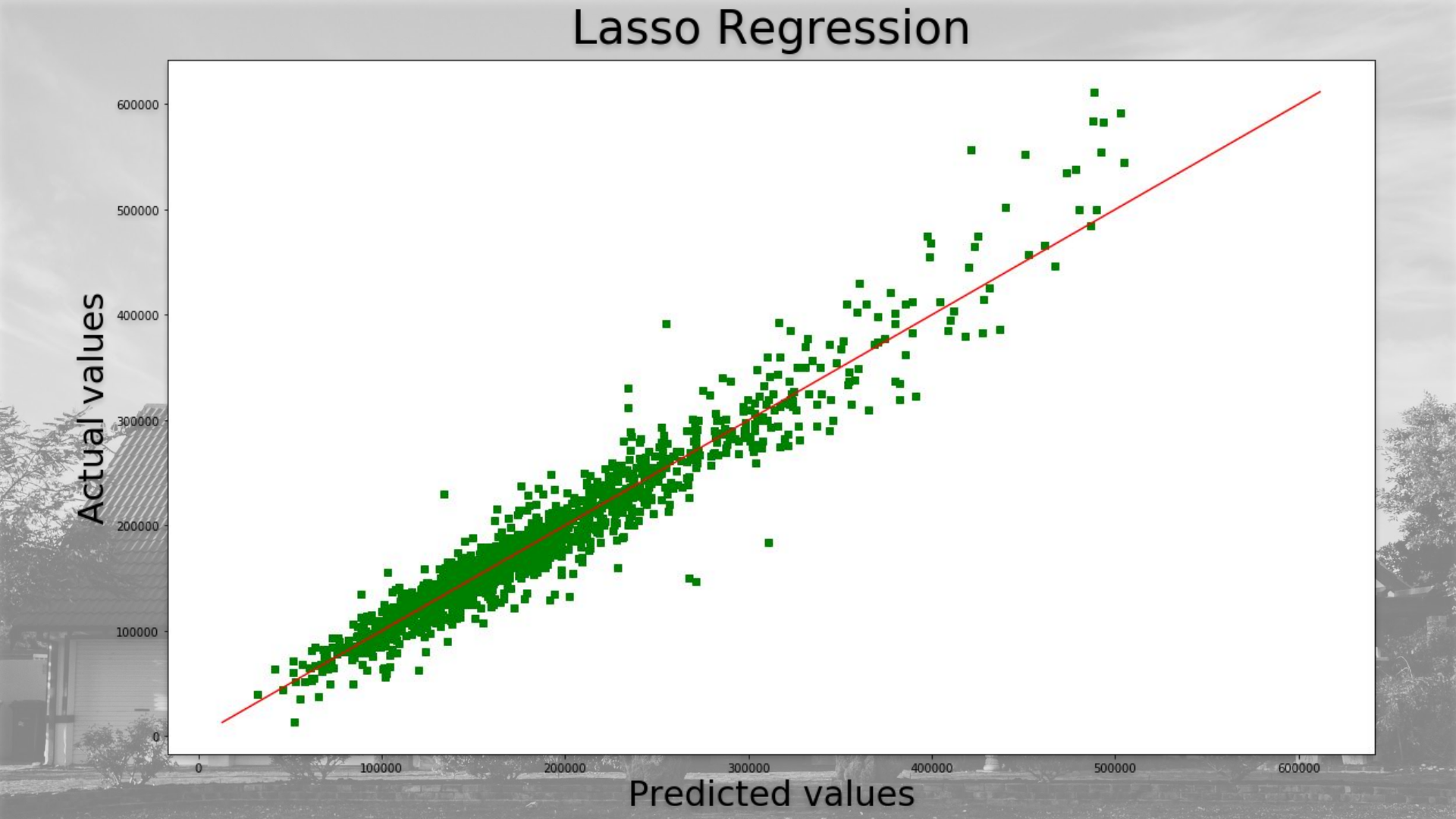
Cross Value Score	R2 Score
0.809514	0.885090


Lasso Regression
Actual values
Predicted values
600000
500000
400000
300000
200000
100000
0
0
100000
200000
300000
400000
500000
600000

Modeling - Lasso Regression

Lower Correlation Threshold (0.4)

Cross Value Score	R2 Score
0.834811	0.916678


Lasso Regression
Actual values
Predicted values
600000
500000
400000
300000
200000
100000
0
0
100000
200000
300000
400000
500000
600000

Findings

Top 5 Positive Influencers

1. Overall Quality & Above Ground Living Area in Sq.Ft.
2. Overall Condition & Above Ground Living Area in Sq.Ft.
3. Masonry Veneer Area in Sq.Ft. & Basement Full Bathrooms
4. Lot Area in Sq.Ft. & Height of Basement in inches
5. Year Built & Year Remodeled

Only if we had more data...


Thank you!
Questions?