

Steve Shields II  
Project Director for Lotus Housing Solutions

## Steve Shields II

## Project Director for Lotus Housing Solutions

**01.**

## About the Data

Brief overview of Ames

**02.**

## Visualizations and EDA

Skewness and  
Correlations

**03.**

## Variables Used

Data Dictionary

**04.**

## Model Selection

Explanation of models  
used to test

**05.**

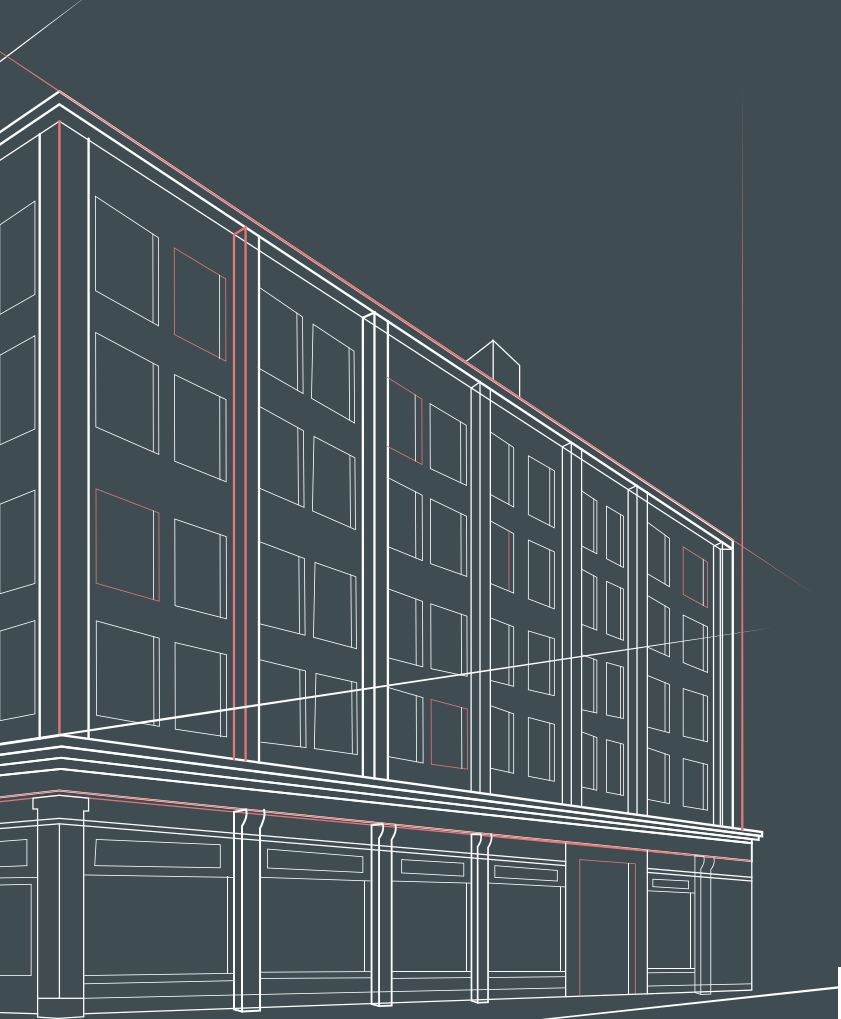
## Conclusion

What I found to be the  
best model

**06.**

## Questions?





# About the Data



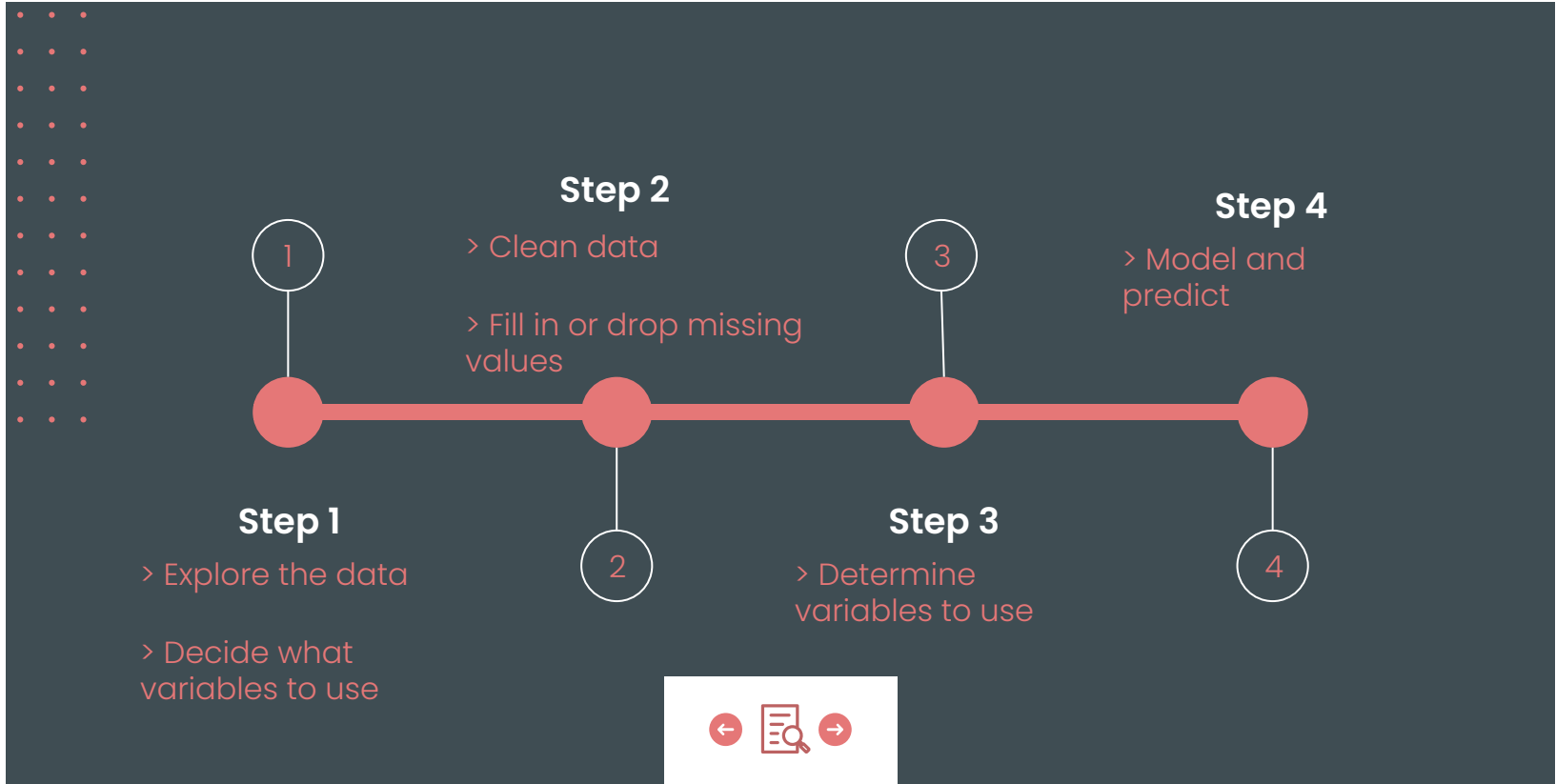


**The Ames, Iowa data set contains a comprehensive list of more than 80 variables that can be used to help determine sale prices for houses.**

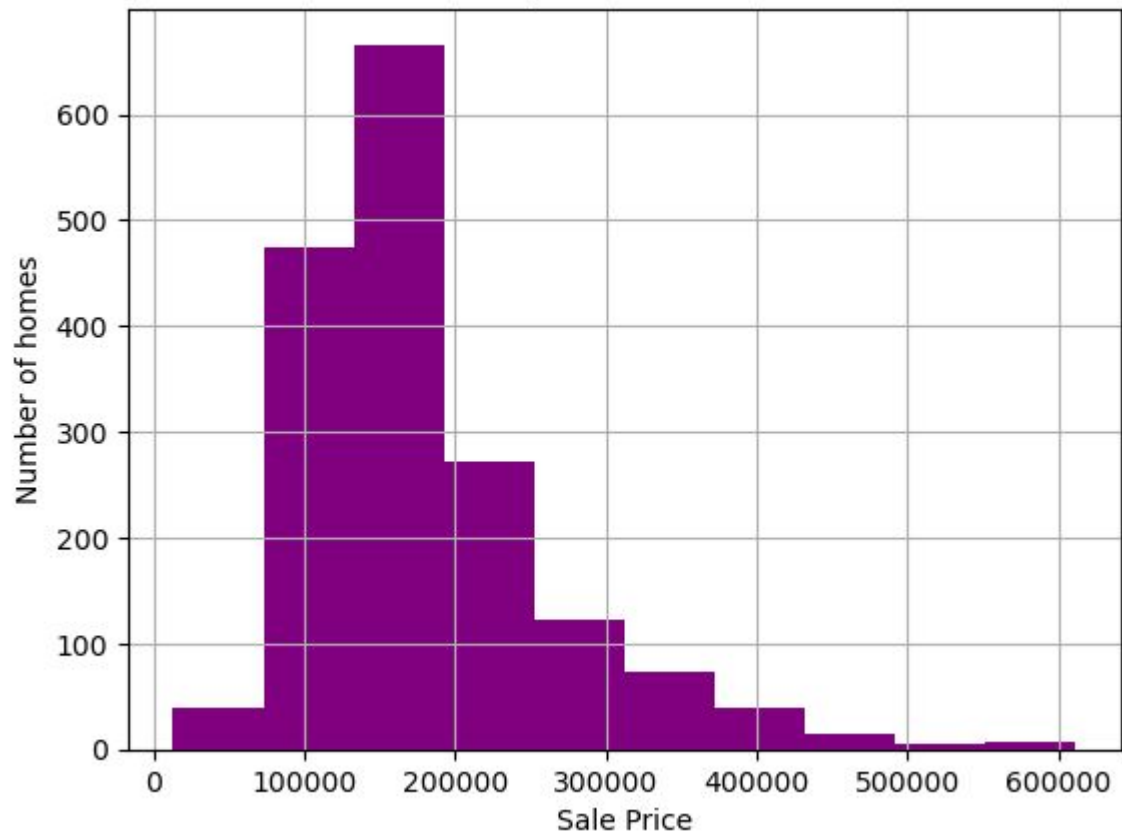


Feature	Type	Description
Half Bath	int64	Number of half baths
Lot Area	int64	Lot size in square feet
Wood Deck SF	int64	Deck area in square feet
Open Porch SF	int64	Porch area in square feet
Bsmt Fin SF1	int64	Rating of basement finished area
Fireplaces	int64	Number of fireplaces
Totrooms AbvGR	int64	Total rooms above ground
Garage YR built	float64	Year the garage was built
Full bath	int64	Number of full baths
Year Remod/Add	int64	Year of remodel or addition
Year built	int64	Year the house was built
1st flr SF	int64	First floor area in square feet
Garage Cars	float64	Number of cars that can fit in garage
Garage Area	float64	Area of garage in square feet
Gr Liv Area	int64	Above ground living area in square feet
Overall Qual	int64	Rates overall material and finish of house
SalePrice	int64	Last sale price of house

# Process



Sales price distribution



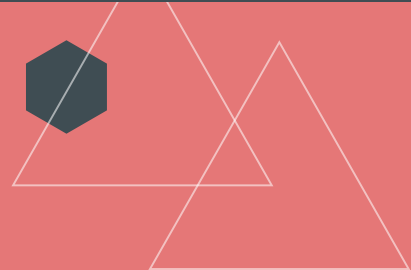
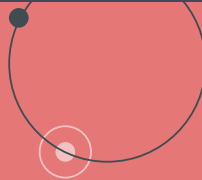


## Null values and Outliers?



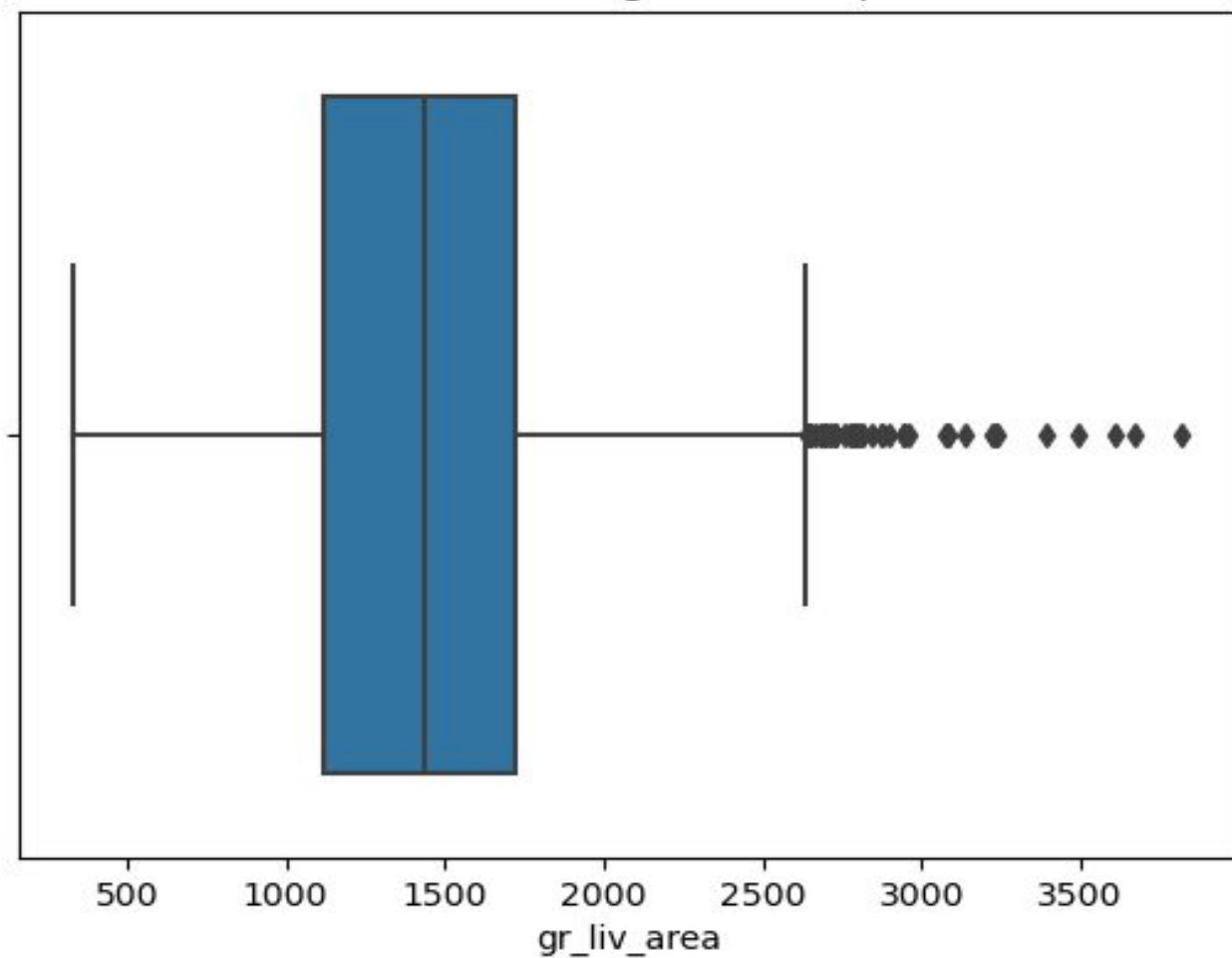
Null Values – Were either filled with 0's or the mean (Garage Year Built)

Outliers – Were dropped, especially from the Sale Price column. In order to get a better baseline for model selection.

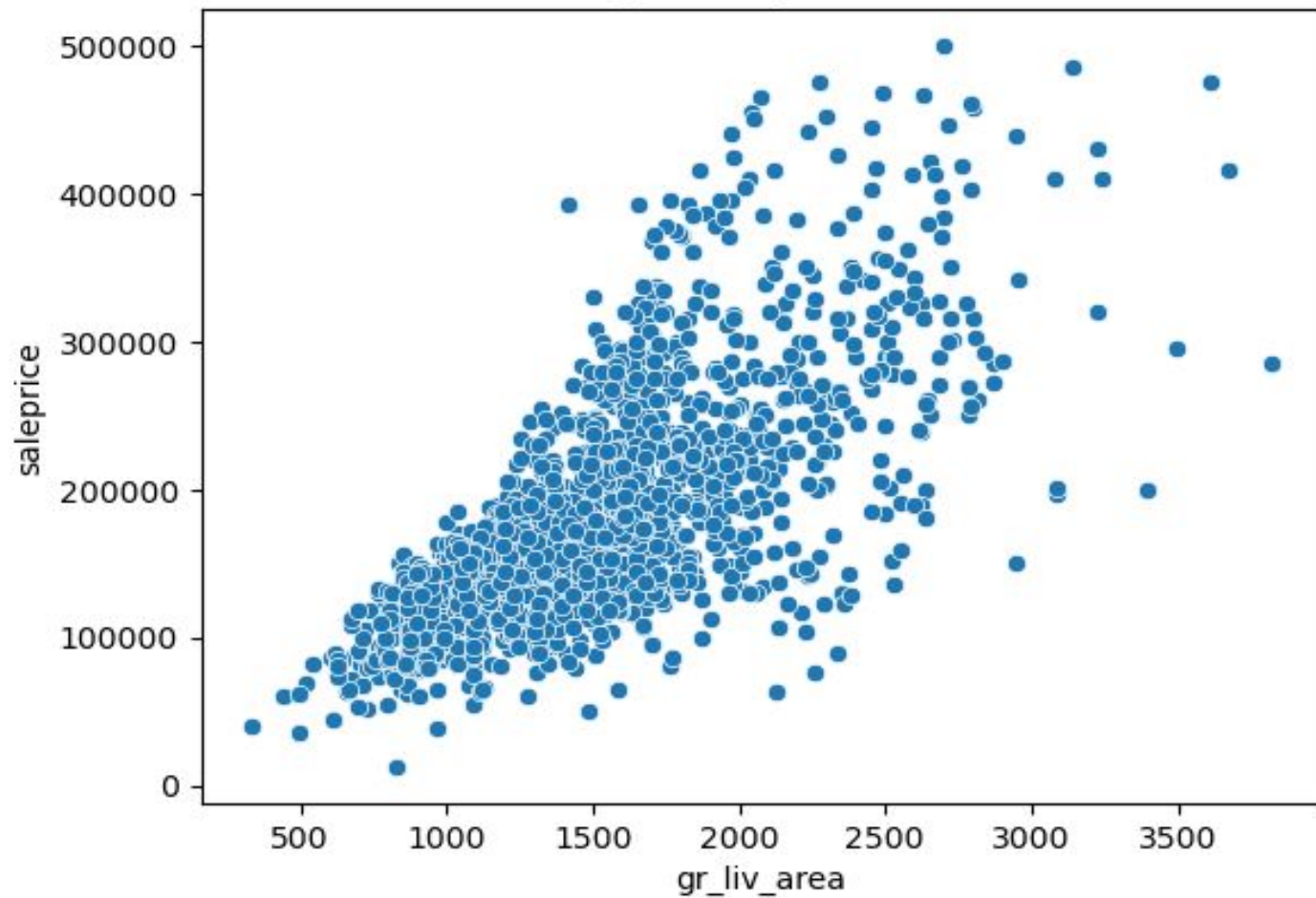




Above Ground Living Area in Sqare Feet



Living Area by Sale Price



## Model Selection

**Baseline : \$178,182.59**



**OLS**

$R^2$ : .900

Mean: \$178,991.33

RMSE: \$23,240.90



**Ridge Model(SS)**

$R^2$ : .906

Mean: \$178,546.60

RMSE: \$22,885.58



**LassoCV (SS)**

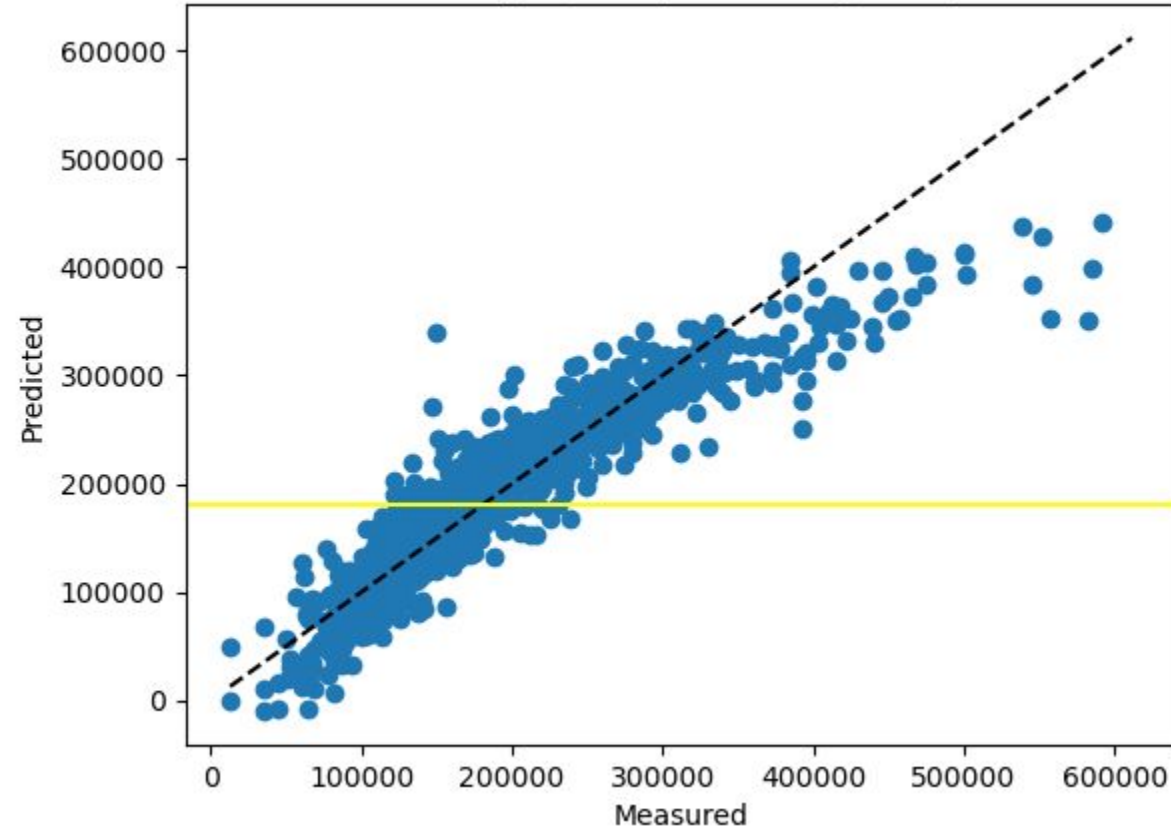
$R^2$ : .906

Mean: \$178,546.61

RMSE: \$22,882.49



Lasso\_cv(black) vs. Mean(yellow)



**Black dotted line**

Predicted sales  
price

**Yellow line**

Mean sales price





# Questions?

Sources: De Cock, Dean. "Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project." Journal of Statistics Education 19.3 (2011).

