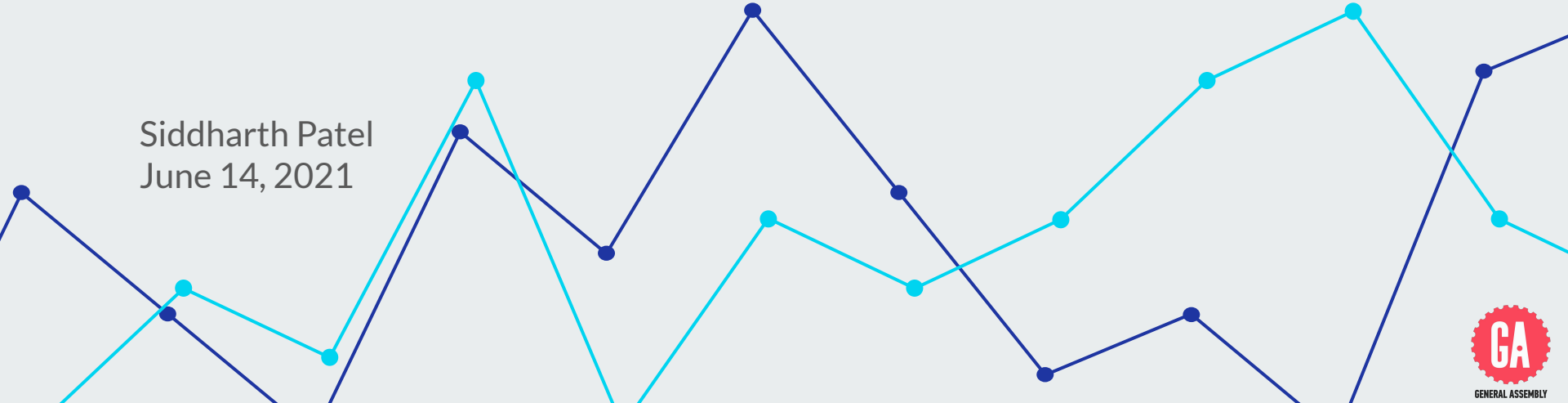
Ames Housing Price Analysis

Siddharth Patel
June 14, 2021



GENERAL ASSEMBLY

Overview

Housing prices can be affected by

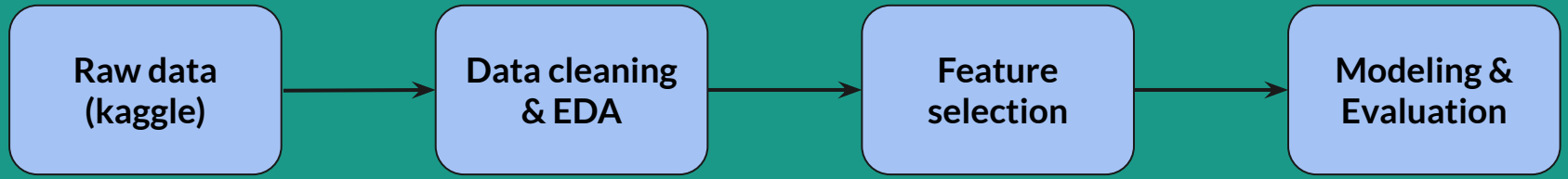
- Extrinsic factors: policies, tax, market [not in control]
- Intrinsic factors: plot area, amenities, beds & baths, quality [can be quantified]

Problem Statement

As a data analyst we are tasked with the challenge to use this quantified housing data, and make analytical predictions for the property sale prices in the city, using simple machine learning methods.

At our agency, we use this data to predict house prices for any seller that intends to make a sale for their property based on the house's features.

Data workflow



train.csv
2051 observations
80 features
1 target

test.csv
879 observations
80 features

train
1230 observations

valid
821 observations

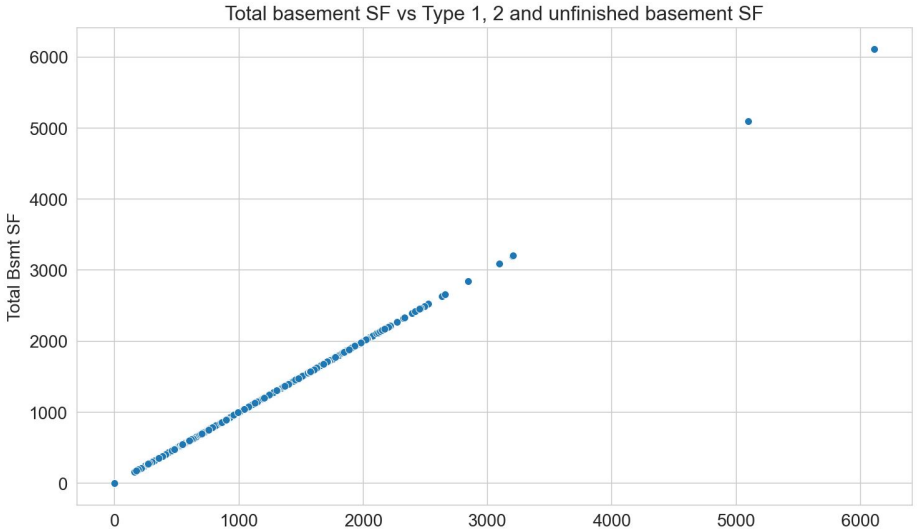
test
879 observations

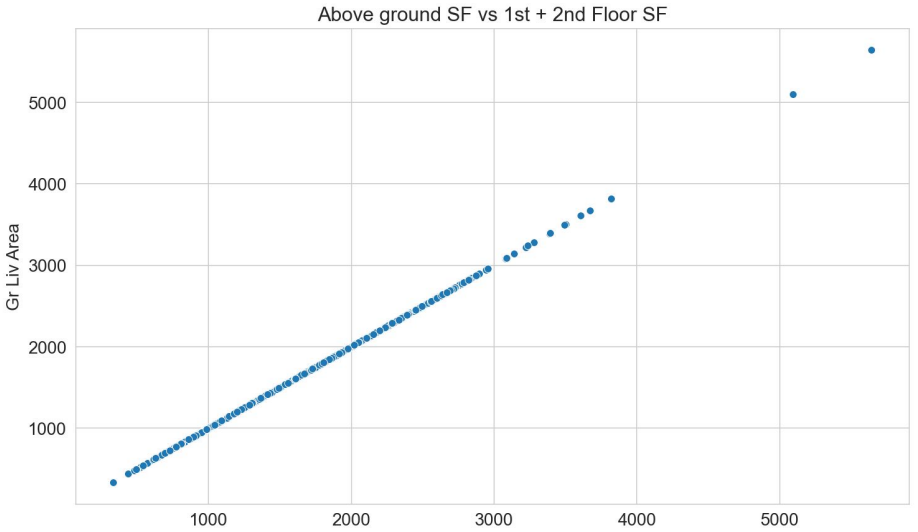
train, valid, test
26 features

Data Exploration

- Dropped columns
- Added interaction columns
- Replace NaNs with NA/mode/mean
- Binarized columns
- One-hot encoded ordinal features
- Outlier threshold

EDA - columns that add upto another column



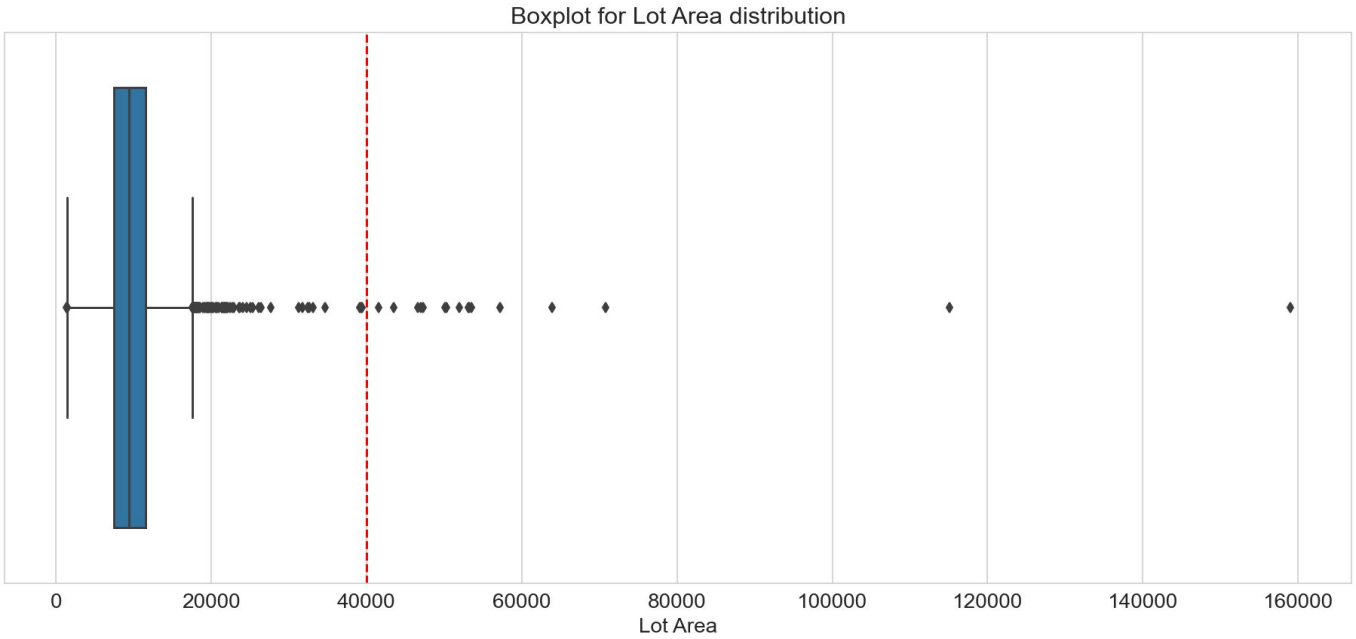


Total Bsmt SF = BsmtFin SF 1 + BsmtFin SF 2 + Bsmt Unf SF

Gr Liv Area = 1st Flr SF + 2nd Flr SF + Low Qual Fin SF

GENERAL ASSEMBLY

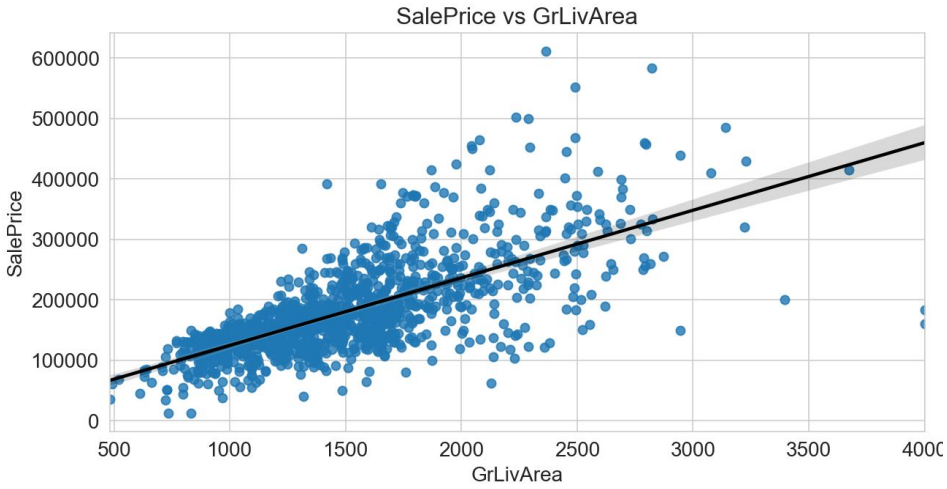
EDA - outliers need to go



Lot Area box plot showing outliers > 40000

GENERAL ASSEMBLY

Modeling & Feature Selection

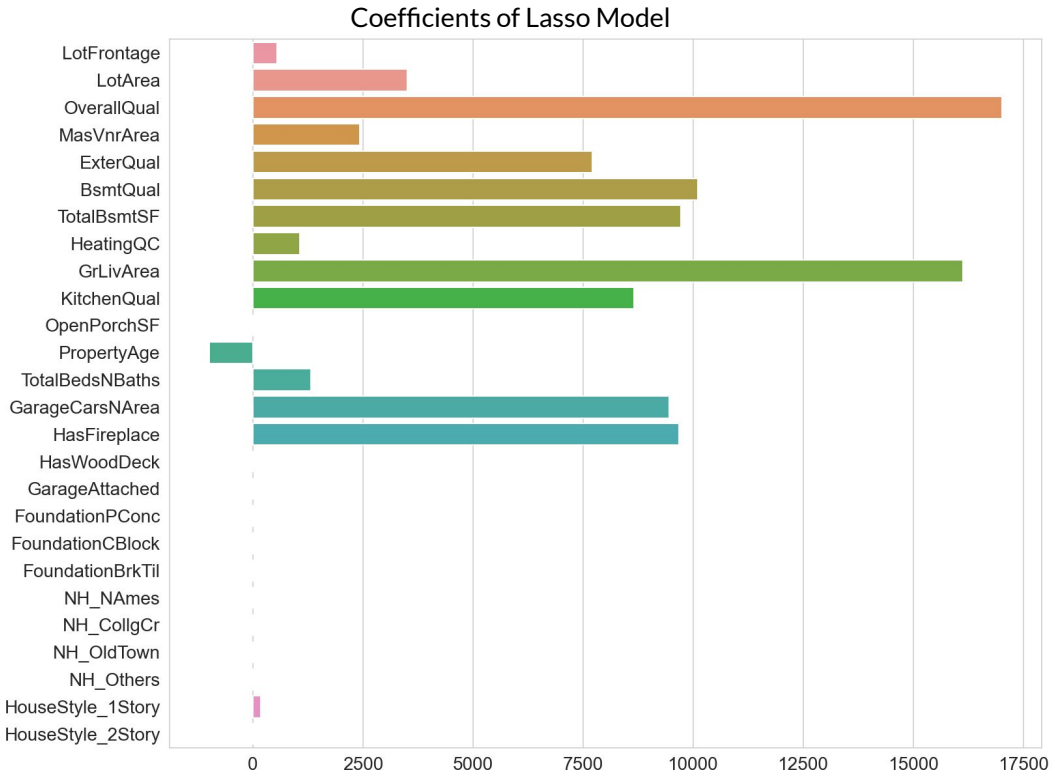


SalePrice increases with Ground Living Area

SalePrice decreases with PropertyAge

GENERAL ASSEMBLY

Modeling & Feature Selection



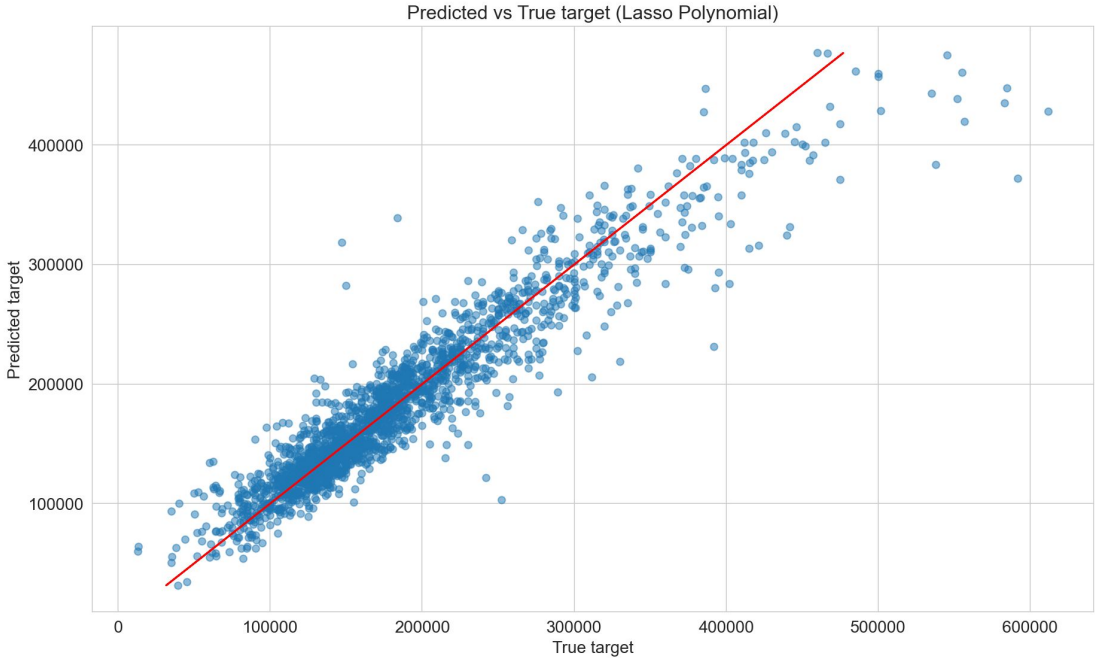
Models built:

1. SLR
2. MLR
3. Ridge
4. Lasso
5. ElasticNet

Lasso has the capability to zero the coefficients that are not significant.

GENERAL ASSEMBLY

Results - Lasso



RMSE summary

Model	Train	Test
Baseline	$77014.97	$80259.83
Production	$24470.15	$28435.67

GENERAL ASSEMBLY

Conclusions

- 80 → 26 features selected
- Scaling and polynomial feature interactions
- Production model: Lasso regularized regression with polynomial features
- 306 / 351 polynomial interaction terms zeroed
- Most positive impact features:
    - Basement Quality
    - Exterior Quality
    - Kitchen Quality
    - Overall Quality
    - Has Fireplace

Thank You!