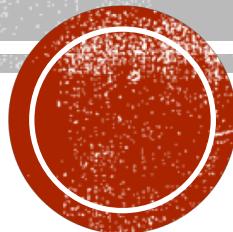


PRICING MODEL FOR AMES IOWA

Noah Monastersky / Data Scientist



AMES, IOWA

- Home to Iowa State University
- Population of over 65,000
- 30 miles away from Des Moines, the capital of Iowa

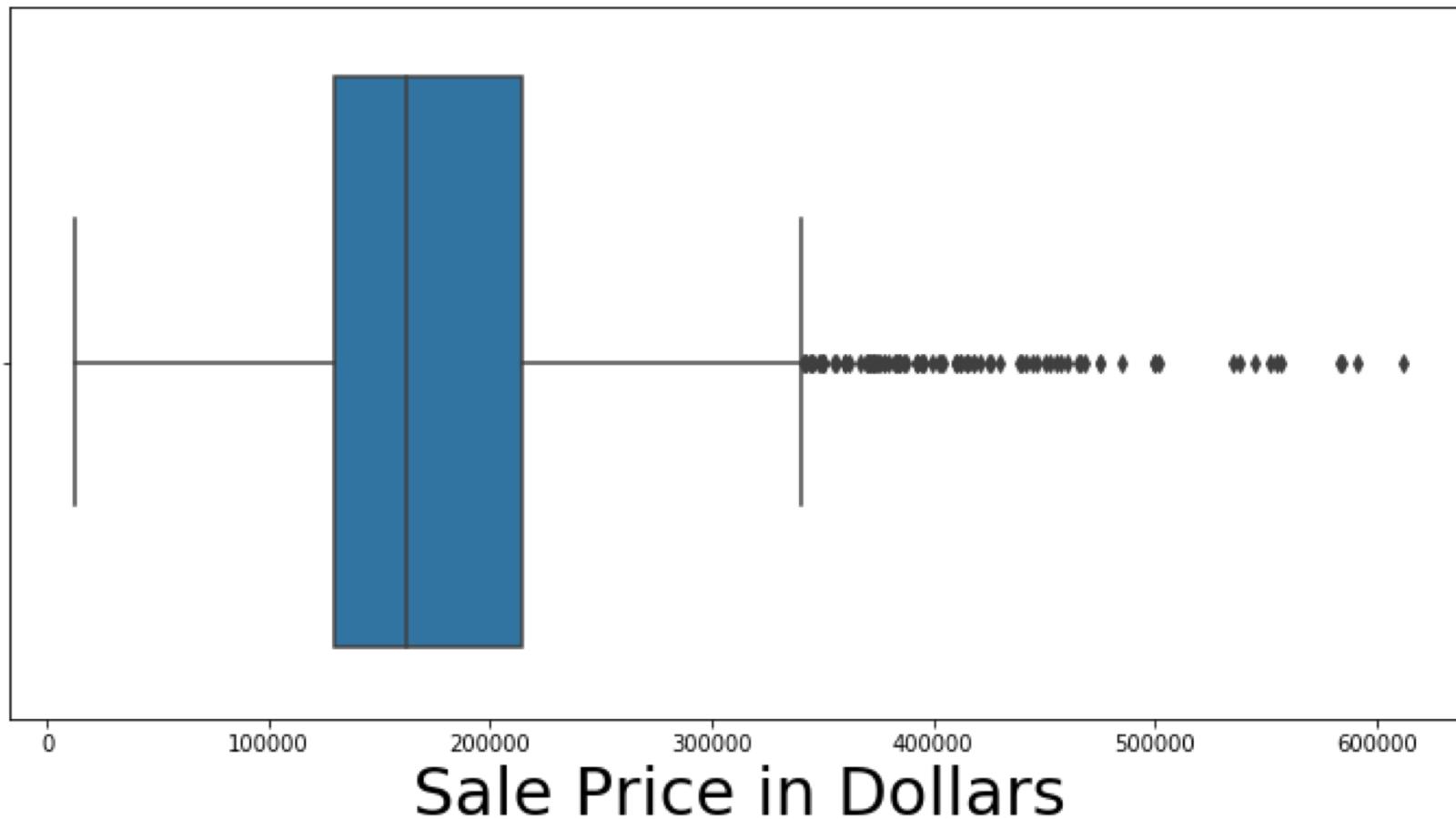


CLEANING THE DATA

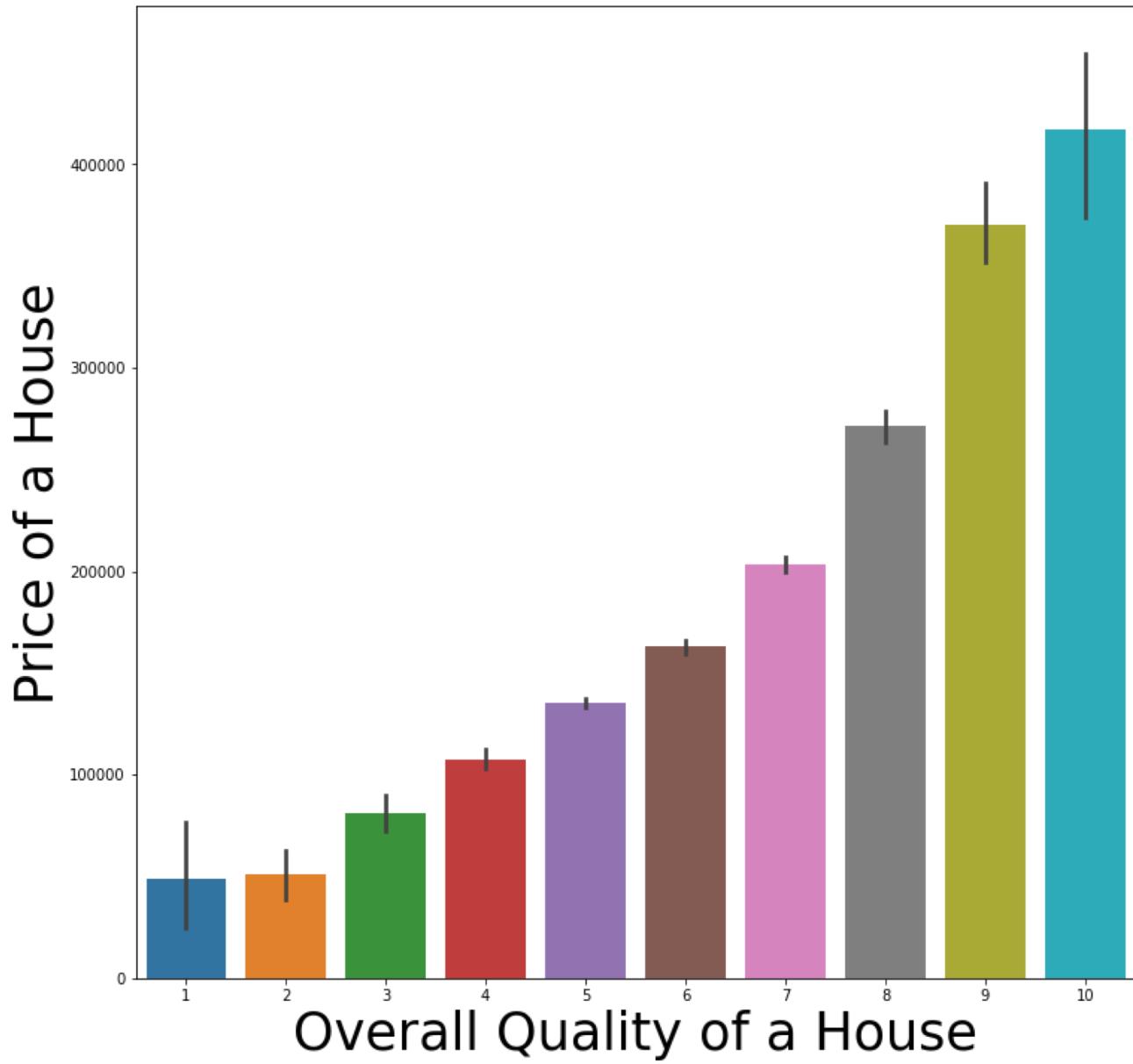
- Dummied Categorical nominal columns
- Dropped columns that had too many null values or was not sure how to handle null values
- Ordered categorical ordinal columns
- Filled in zero for numerical columns and none for categorical columns



PRICE OUTLIERS

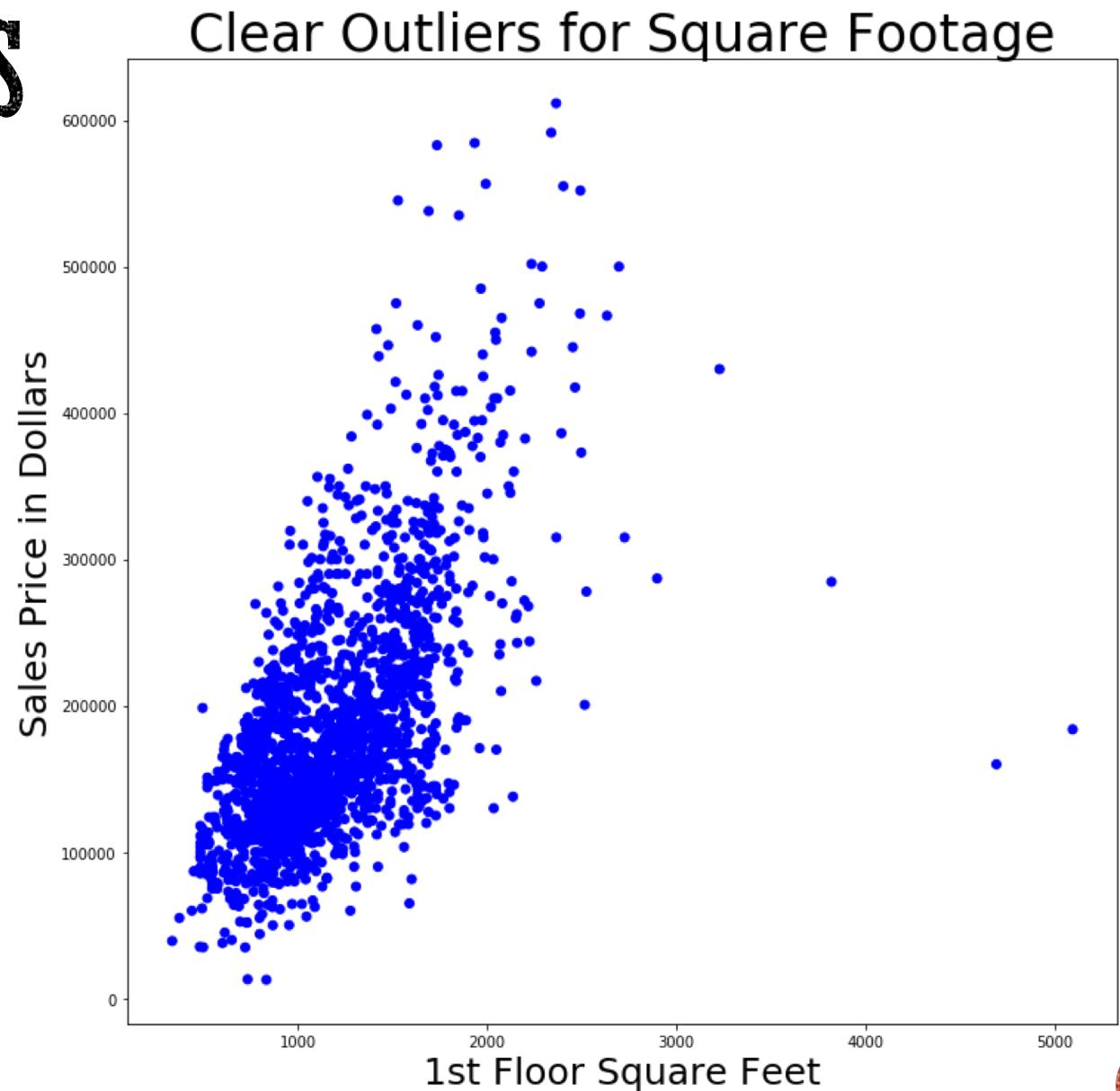


**OVERALL
QUALITY OF
HOUSE
INCREASES AS
PRICE
INCREASES**

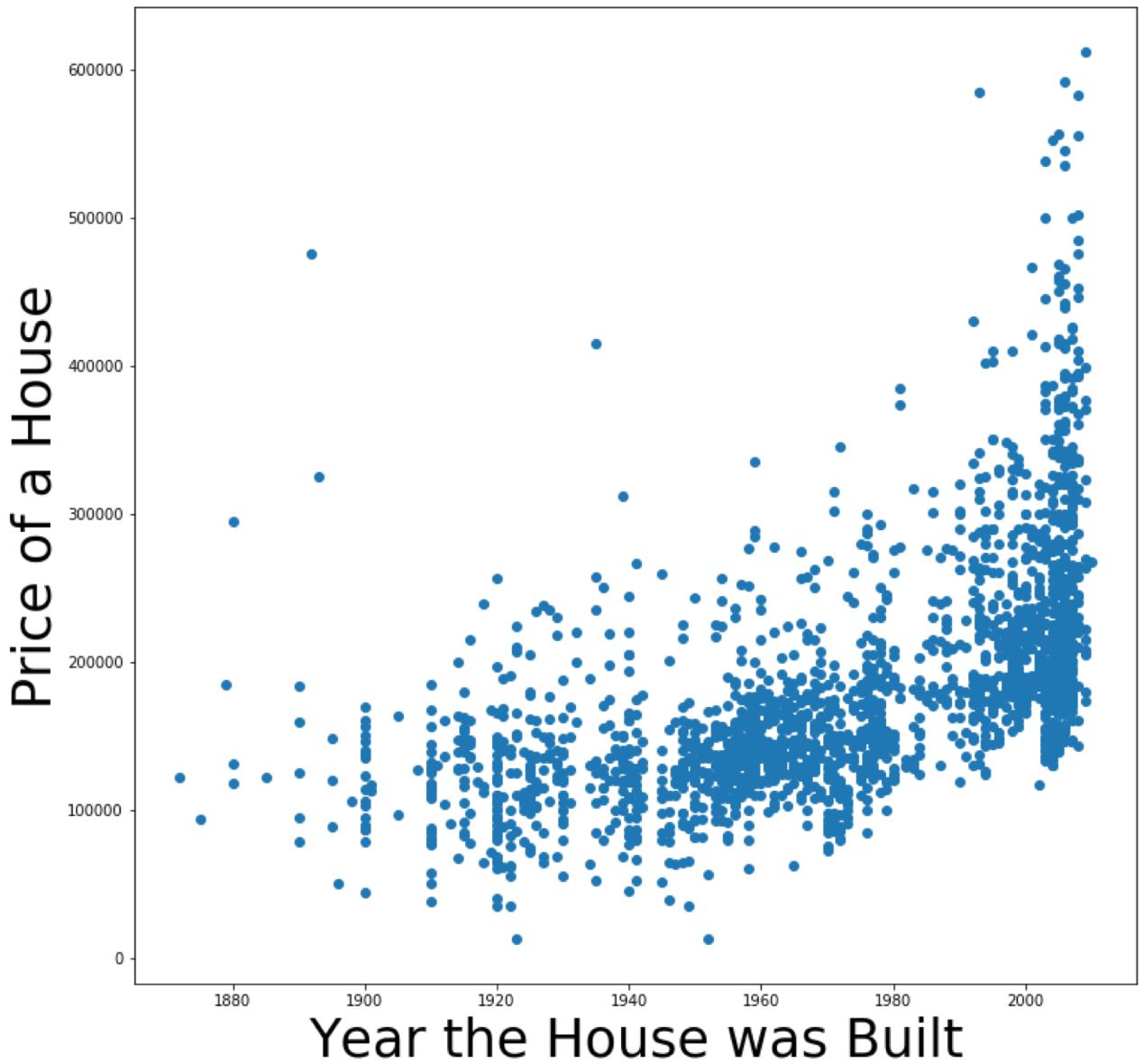


HANDLING OUTLIERS

- Confirm that there are outliers for sq footage
- Don't appear to be outliers for price though



MUCH MORE
HOUSES ARE
BUILT
RECENTLY



CREATING NEW COLUMNS

- For sq feet columns that were outliers I removed outliers and made the outliers equal to the mean of those columns
- For all houses that have a 1st floor sq footage of over 3800 I made a dummy column named huge houses
- Interaction column for area for entire house



MODELING

- Used a power transformer to fit my mode
 - scales and normalizes my data
- Decided to use lasso over ridge to model my data
 - lasso R2 was 90.61 while Ridge R2 was 89.367
 - Lasso zeros out irrelevant variables

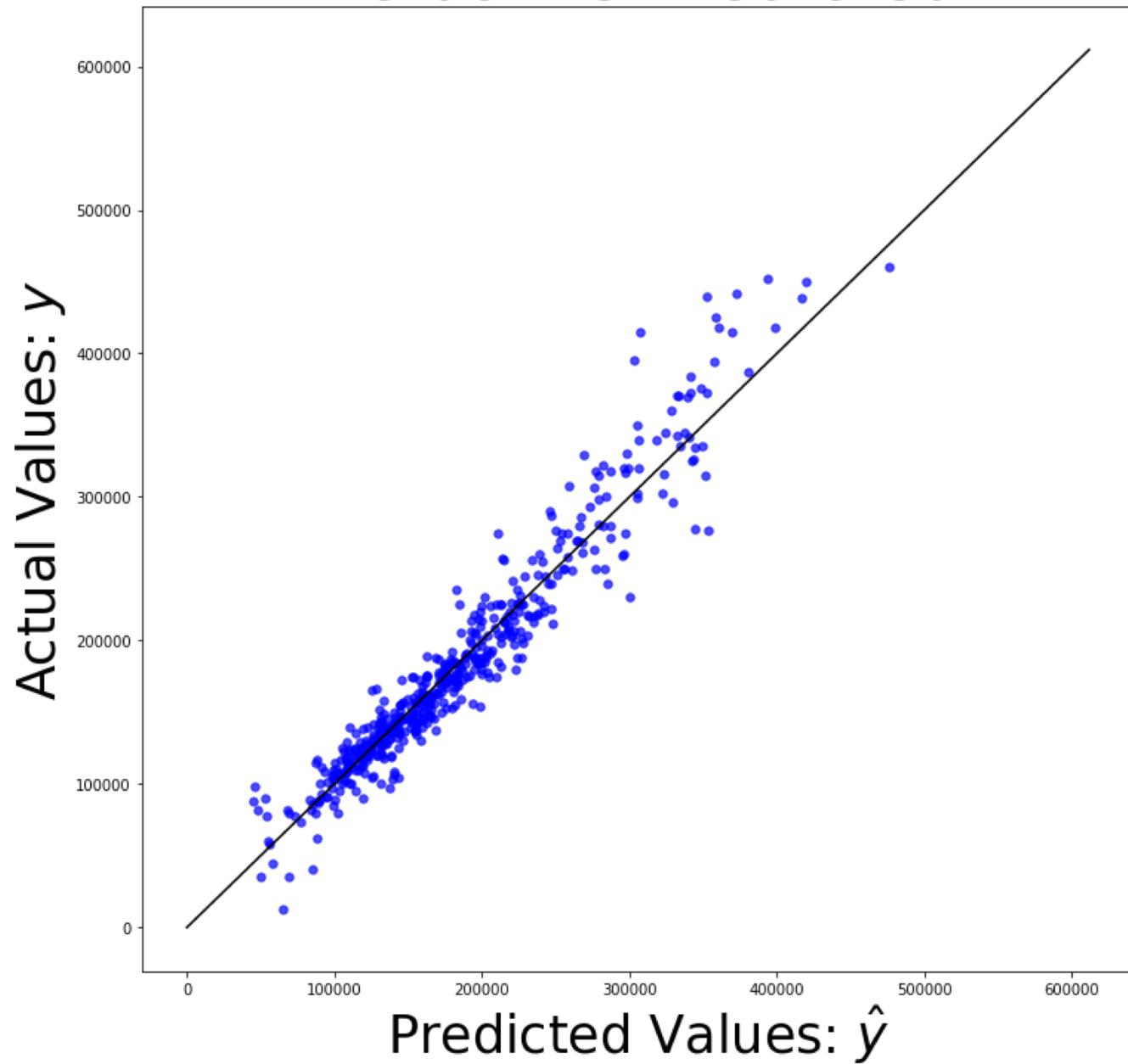


FEATURES OF THE MODEL

- Above ground living area (sq feet)
- Material and finish of the house
- Year of house built
- Lot size in square feet
- Overall condition of the house



Actual vs Predicted



MODEL ACCURACY

- R²: 93.0857
- RSME: 20,604.312
- Possible drawbacks on this model
 - training score is 5 points higher than my testing score.
 - Harder than linear regression to understand the relationship between variables

