

# Ames Iowa Housing Sale Prediction

Presented for:  
Iowa Housing Company

by Ian Stack



# Problem Statement

*I will use linear, ridge, and LASSO regression models to predict the sale price for properties in Ames Iowa housing dataset. The measurement I will use to predict the sale prices will be the root mean squared error (RMSE) from the regression models.*



# Flipping Houses

## WHAT IS FLIPPING HOUSES?

Flipping houses refers to the practice of buying a property, typically at a discounted price, and then renovating and reselling it for a profit. The goal is to quickly and efficiently turn around the property in order to make a profit.



- 01 • Root Mean Squared Error
- 02 • Train-Test Split
- 03 • Linear Regression
- 04 • Ridge Regression
- 05 • Lasso Regression

# Analytic Tools

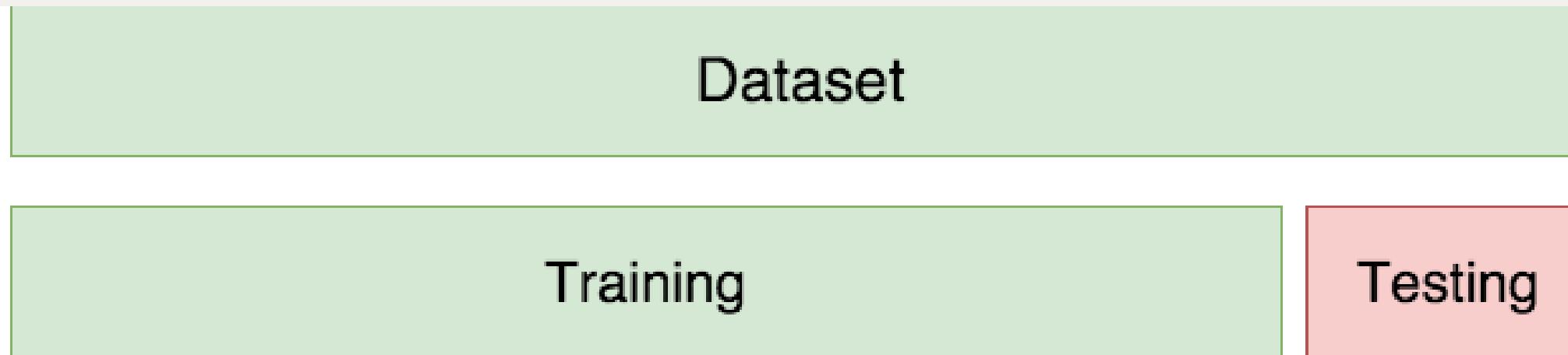


# Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

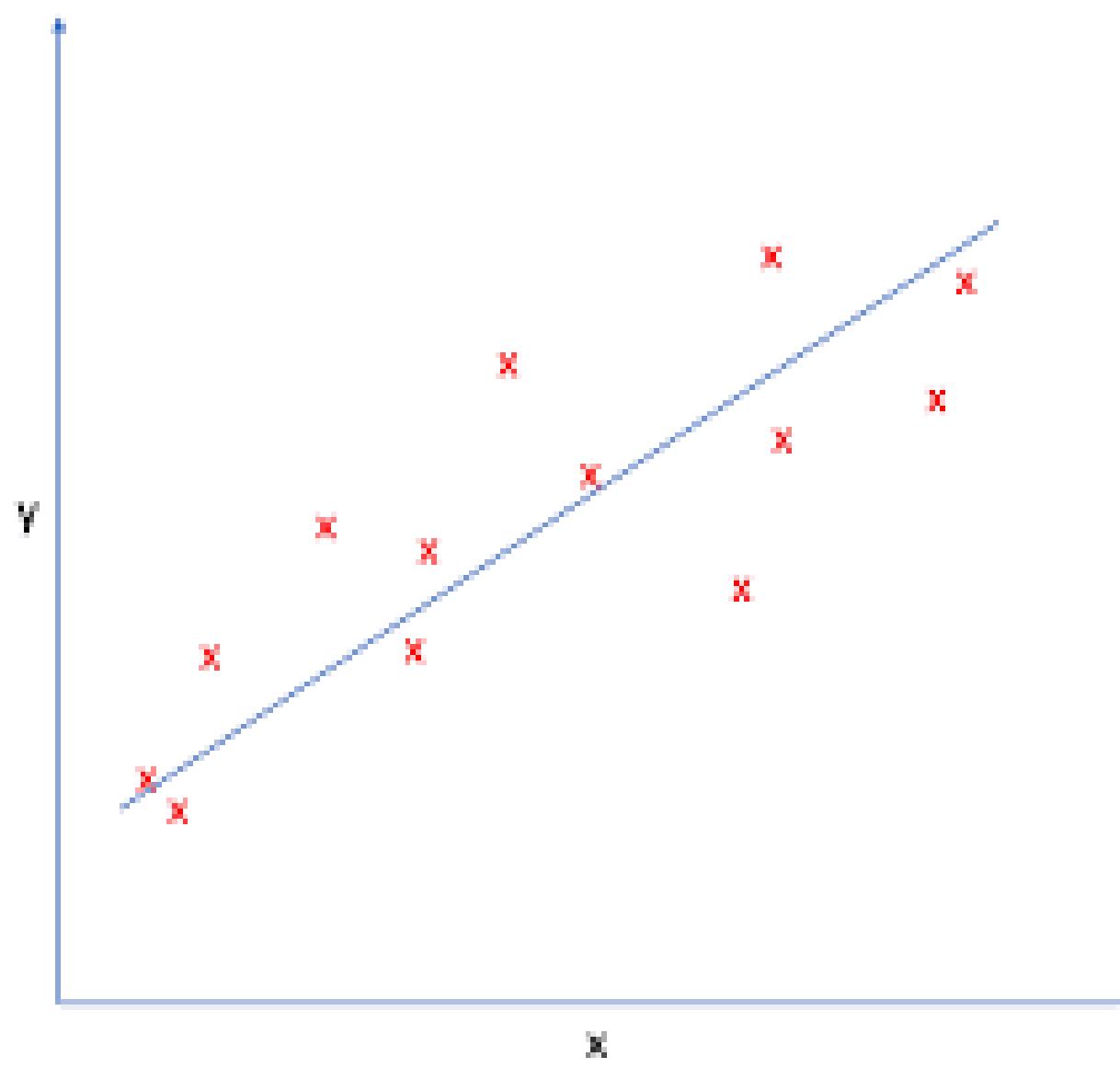
- shows how far predictions fall from measured true values

# Train-Test Split



- Splitting the dataset to help measure performance of models.

# Linear Regression



- Linear Relationship between two variables using a best fit line

# Ridge Regression

*Regularization*

Adds a penalty

Magnitude of coefficients

# Lasso Regression

*Regularization*

Adds a penalty

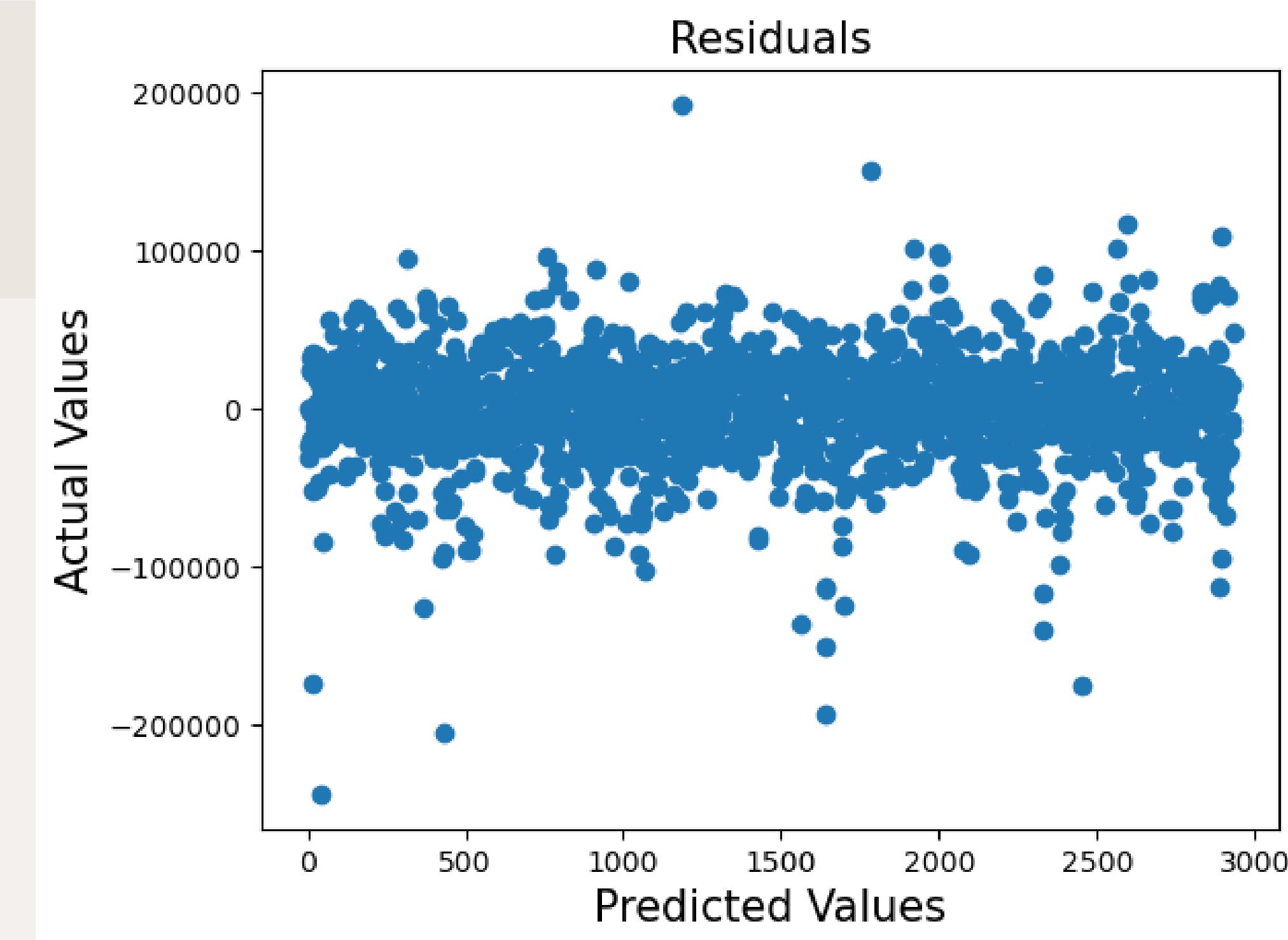
squares coefficients

**Helps with overfitting and  
reduces complexity**

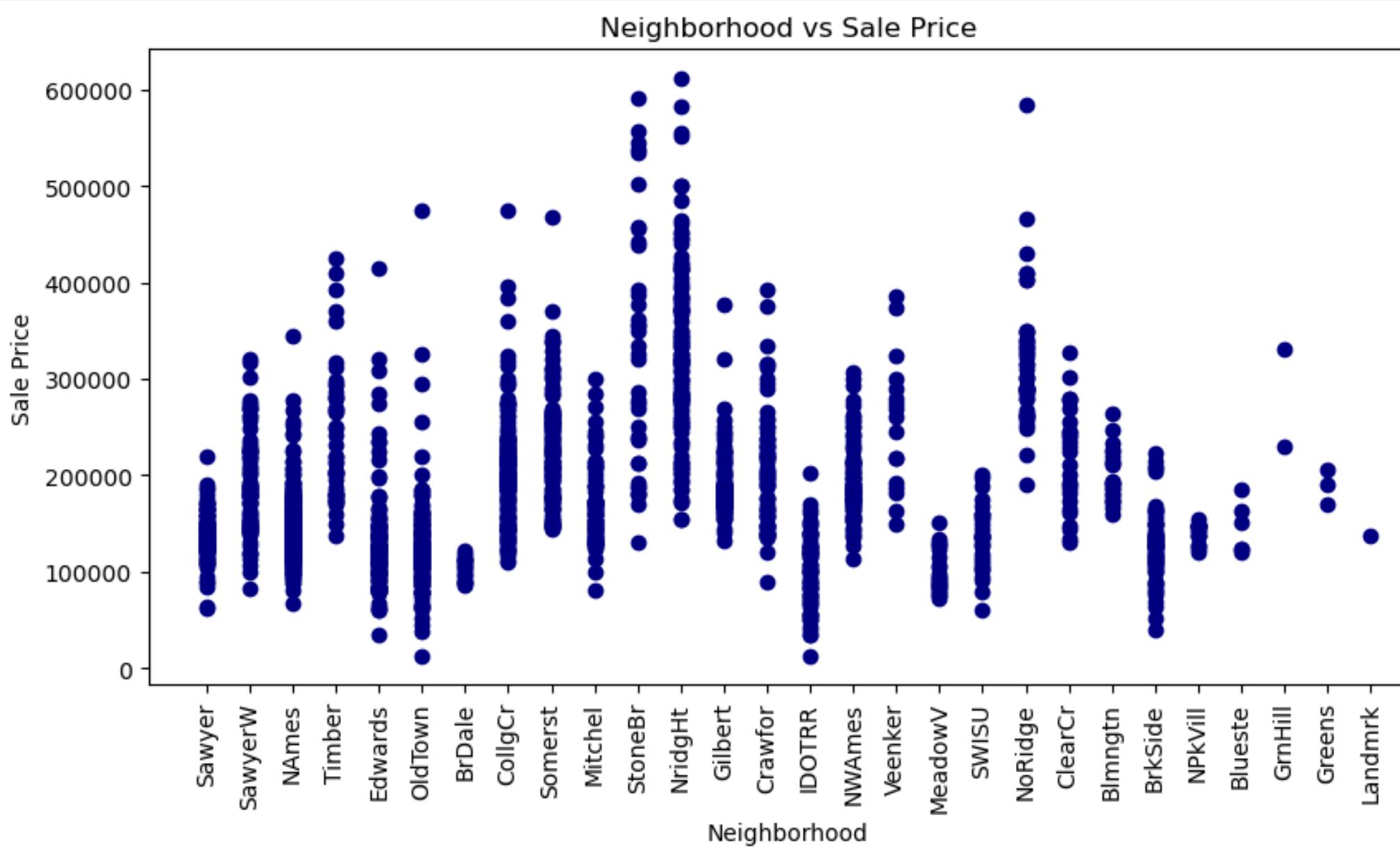
# Baseline Linear Regression Model

Train RMSE: 34078.61

Test RMSE: 34084.74



# Categorical Feature: Neighborhood:

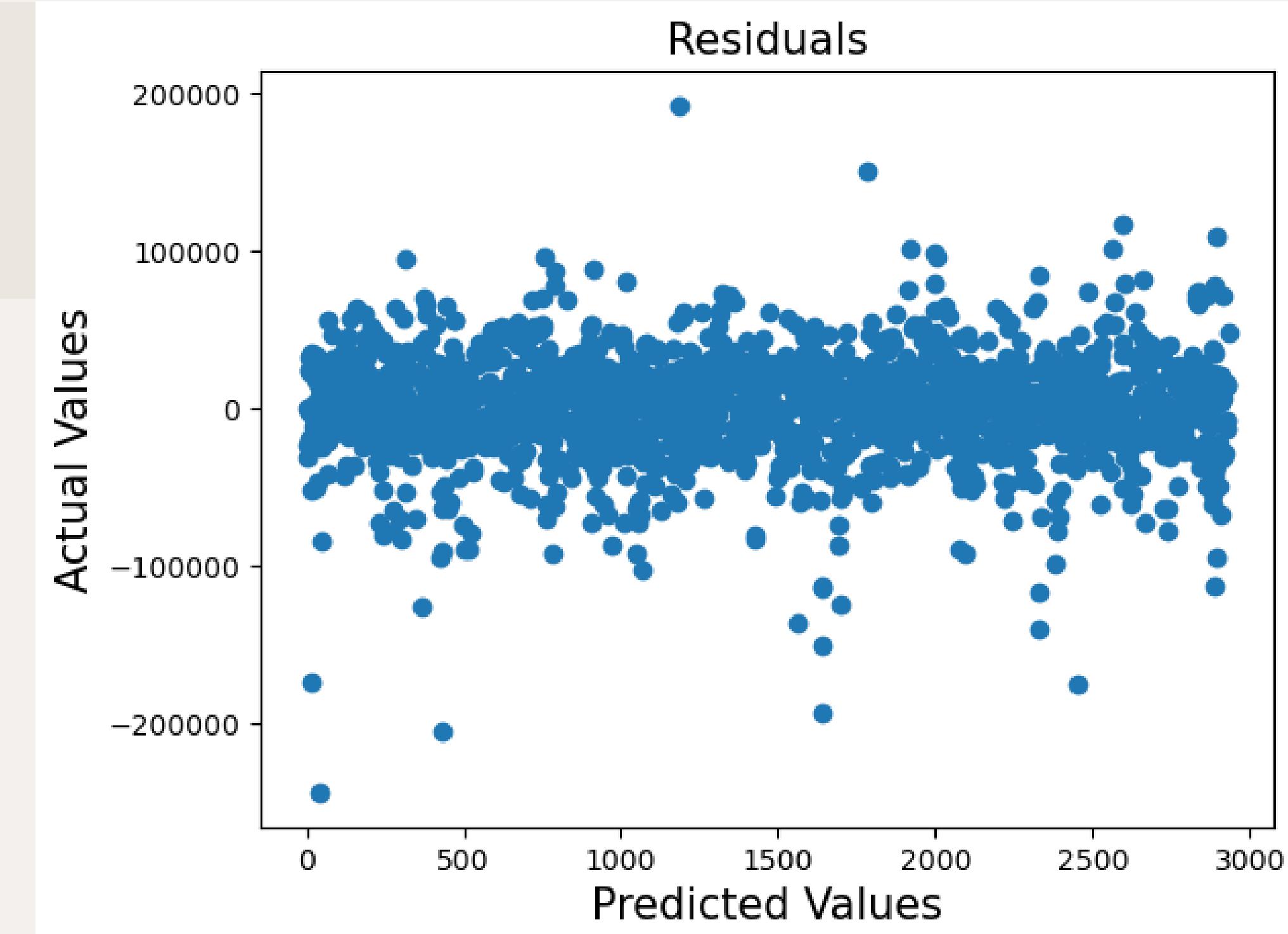


*Adding Neighborhood into model*

# LASSO Regression Model 1

Train RMSE: 28267.13

Test RMSE: 26828.42

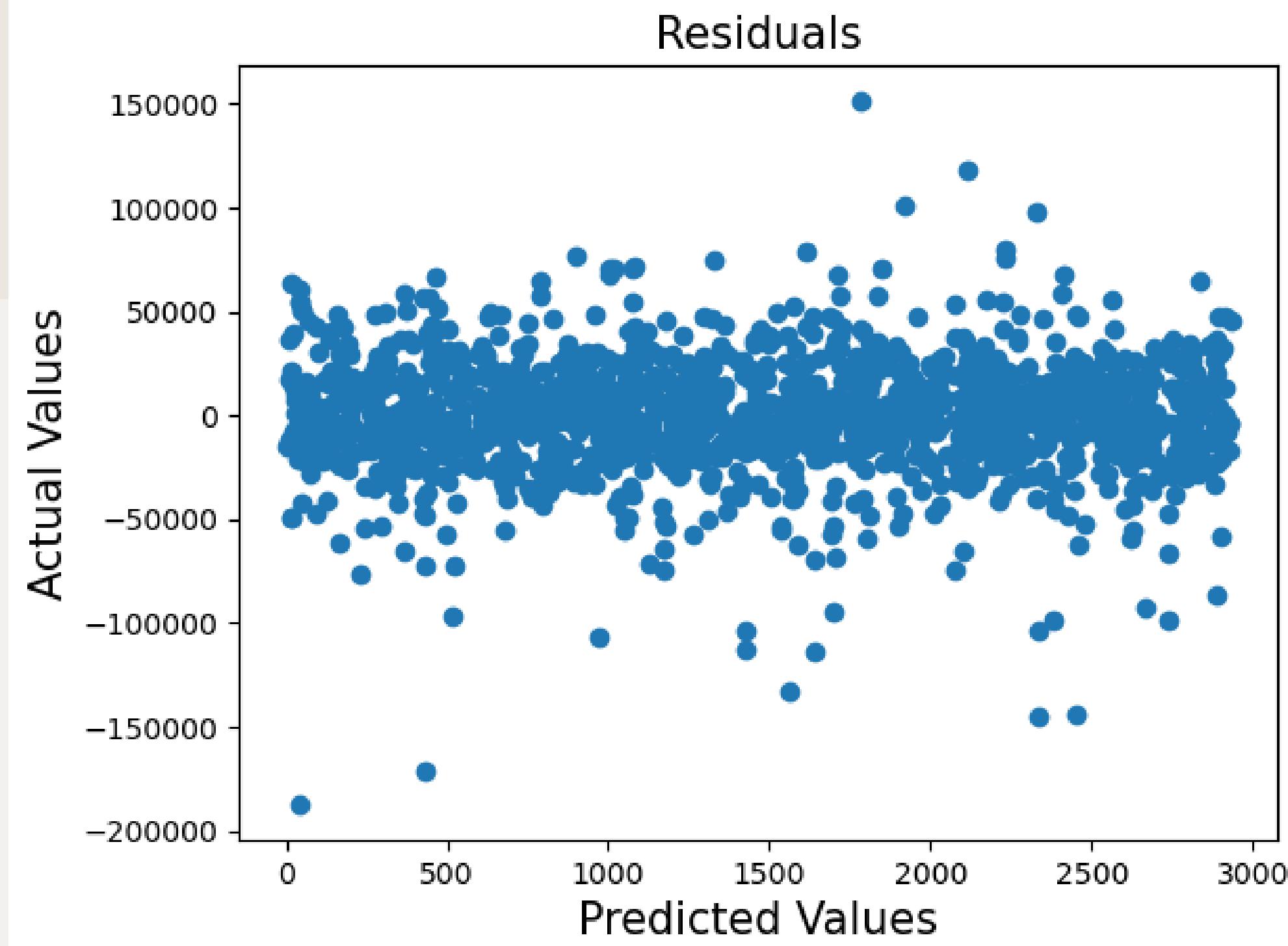


## LASSO Regression Model 2

Train RMSE: 27675.73

Test RMSE: 26172.21

Adding: [Masonry Veneer Area in sq. ft] Feature



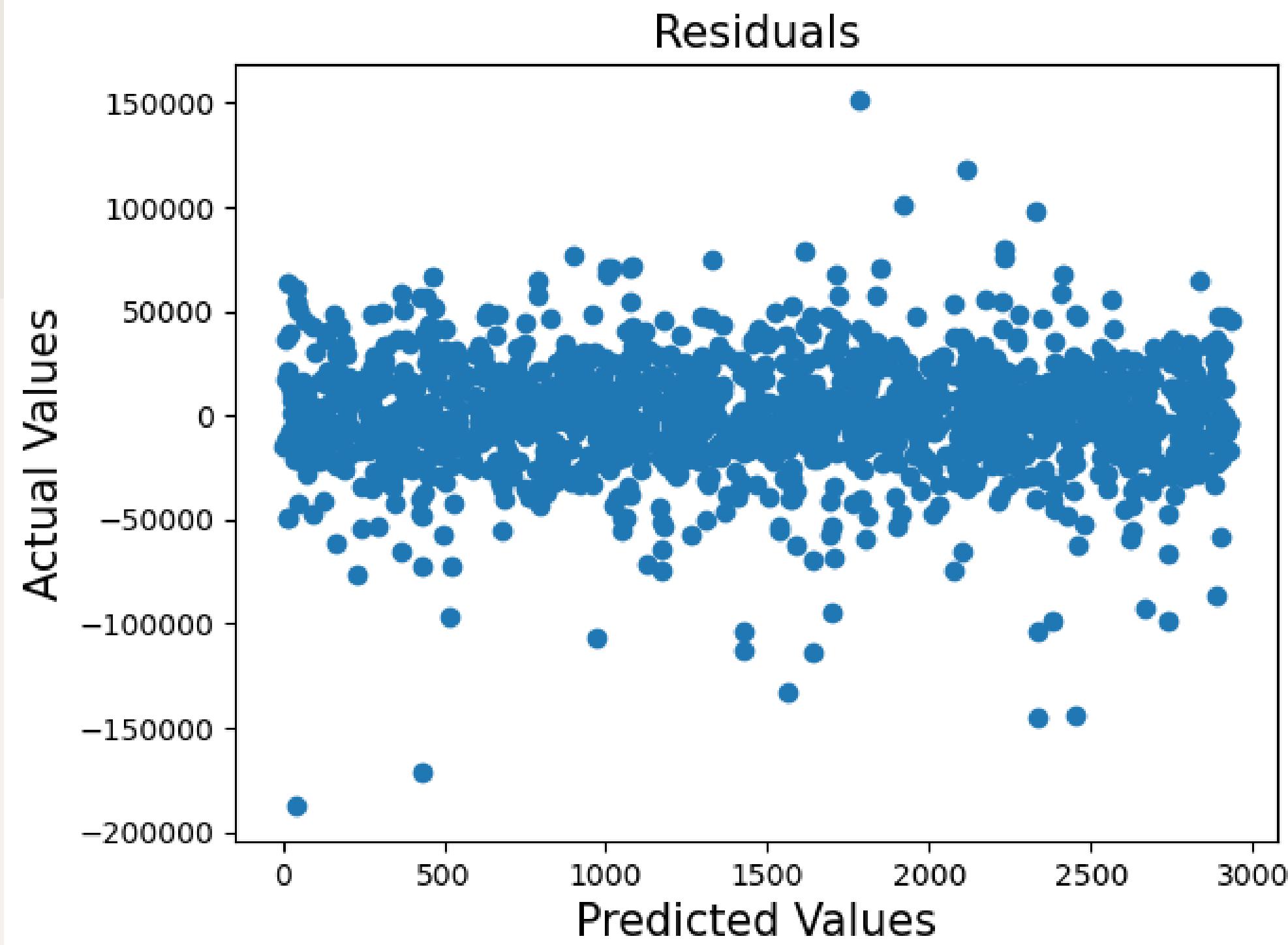
# LASSO Regression Model 3

Train RMSE: 24825.39

Test RMSE: 23366.27

Adding:

- [Kitchen Quality] Feature
- [Exterior Quality] Feature



# Recommendations:

## Features that will increase the Sale Price

- Overall Material & Quality is: Very Good or Excellent
  - Neighborhoods: Stone Brook, Northridge Heights, and Northridge
  - Size of Living Room and Basement sq. footage
- 

## Features that will decrease the Sale Price

- Kitchen quality : Average or Typical
- Exterior quality: Average or Typical

# Conclusion



LASSO REGRESSION

*Eliminate insignificant coefficients*

LASSO > LINEAR

*LASSO out-performed all other models*

LASSO RESIDUALS

LASSO residual scattered

Potential Overfit

# Sources

<https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>

<https://medium.com/@tzjy/10-regression-metrics-data-scientist-must-know-python-code-included-c6ca40320504>

<https://www.jeremyjordan.me/linear-regression/>