Project 2:
Ames Housing Price Prediction

Geh Si Rong
General Assembly DSI 20

Overview

1. Problem Statement

2. Exploratory Data Analysis

3. Feature Selection

4. Modelling

5. Conclusion

Problem Statement

- Booming real estate market in Ames

- Key features affecting house prices

- Best machine learning model to predict house prices

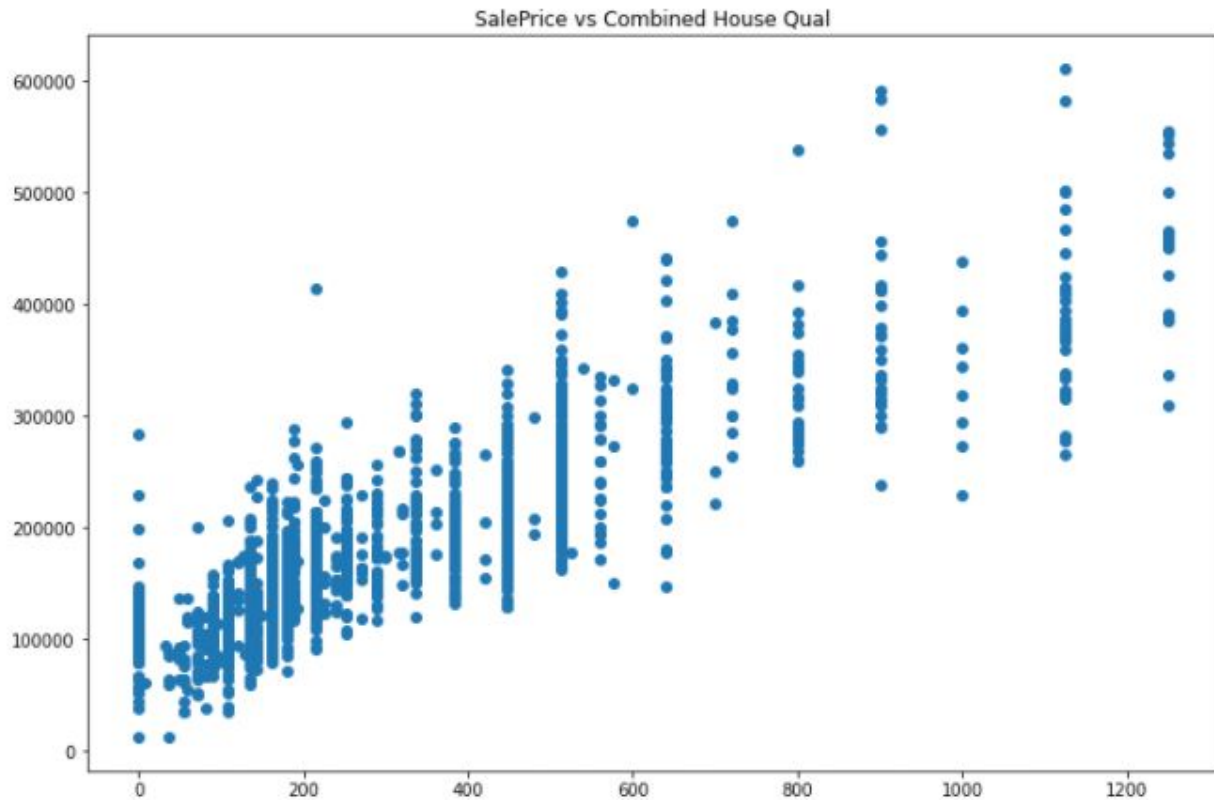- Secure best deals for investors and homeowners

EDA - Key Features

SalePrice vs Combined Area
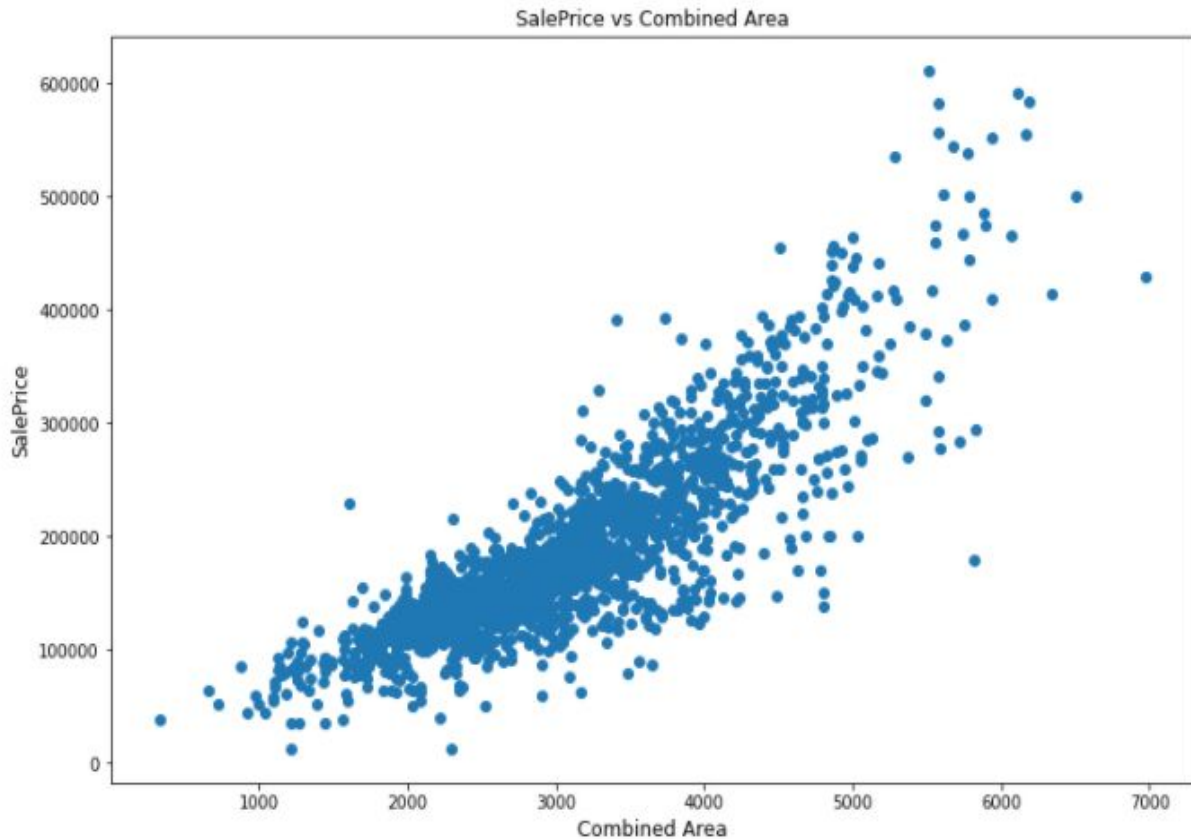
Combined Features

- Gr Liv Area

- Total Bsmt SF

- Garage Area

EDA - Key Features

SalePrice vs Combined House Qual

Combined Features

- Overall Qual
- Exter Qual
- Bsmt Qual
- Kitchen Qual

EDA - Key Features

SalePrice vs Total Rooms Above Grade
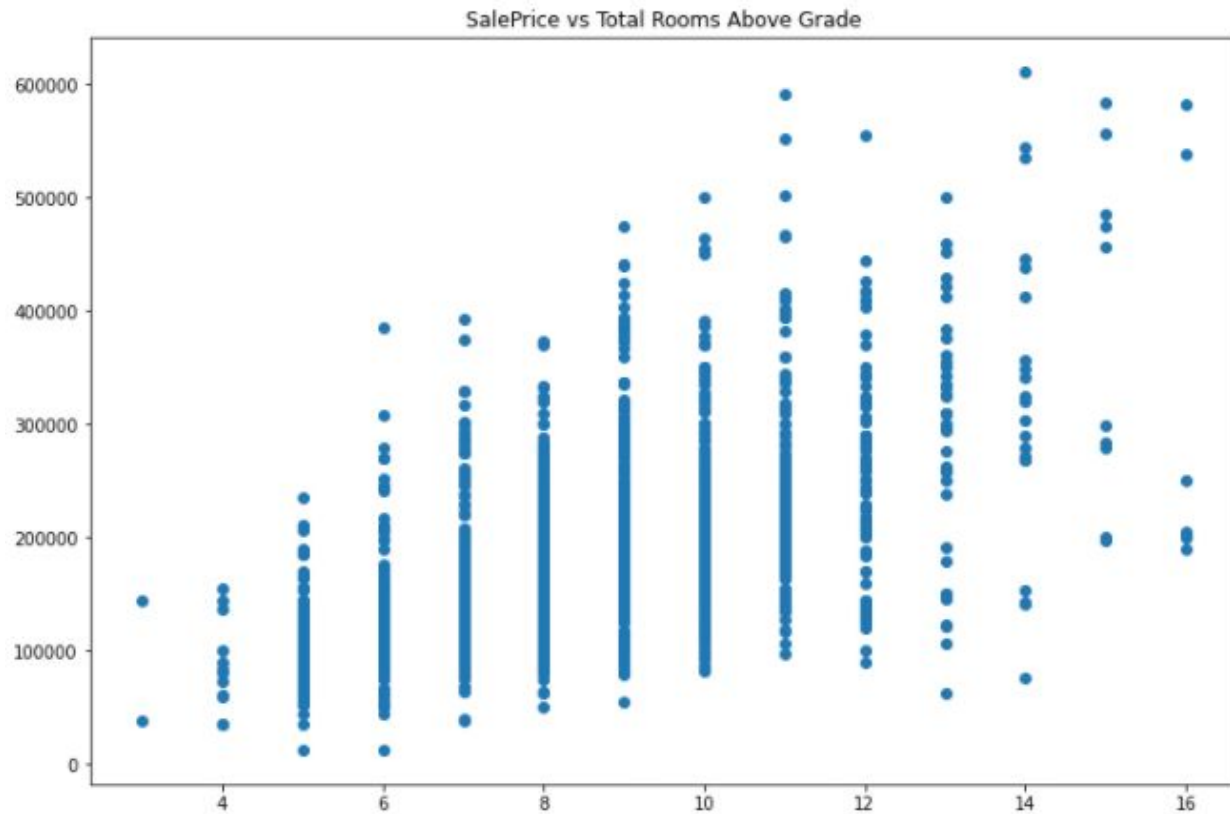
Combined Features

- TotRms AbvGrd
- Full Bath
- Half Bath

Feature Selection

1. Combined Area
2. Combined House Qual
3. House Age
4. Garage Age
5. Years Since Remodel/Add
6. Total Rooms AbvGrd
7. Garage Finish
8. Garage Cars
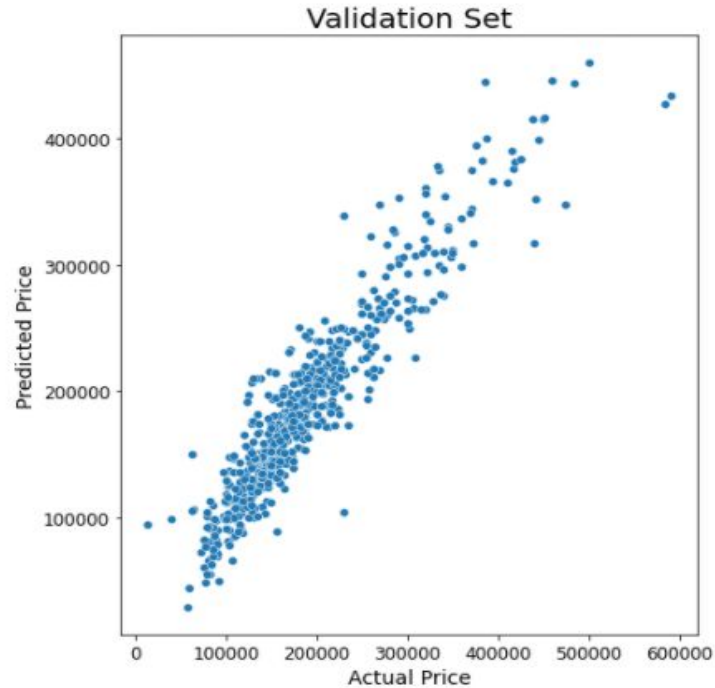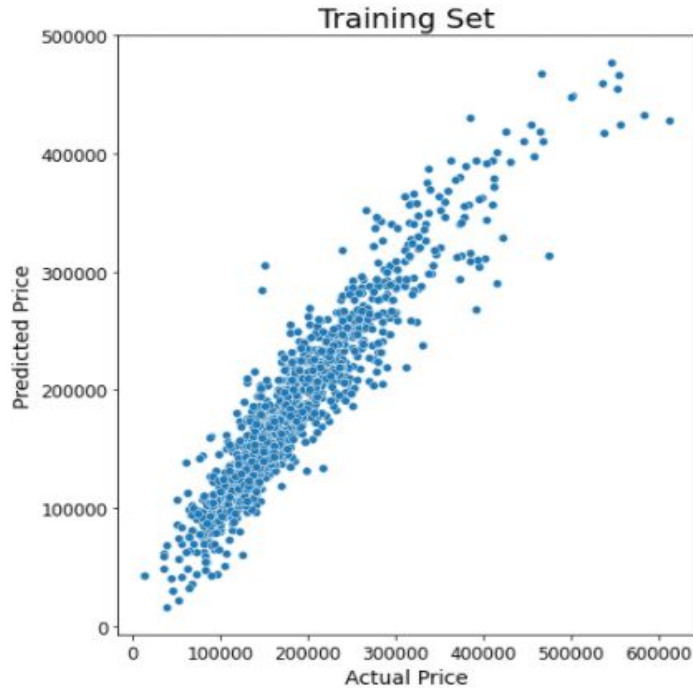9. Neighborhood (one-hot encoded)

Modelling

- Linear RMSE: 28030

- **Ridge: RMSE: 27958**

- Lasso RMSE: 28047
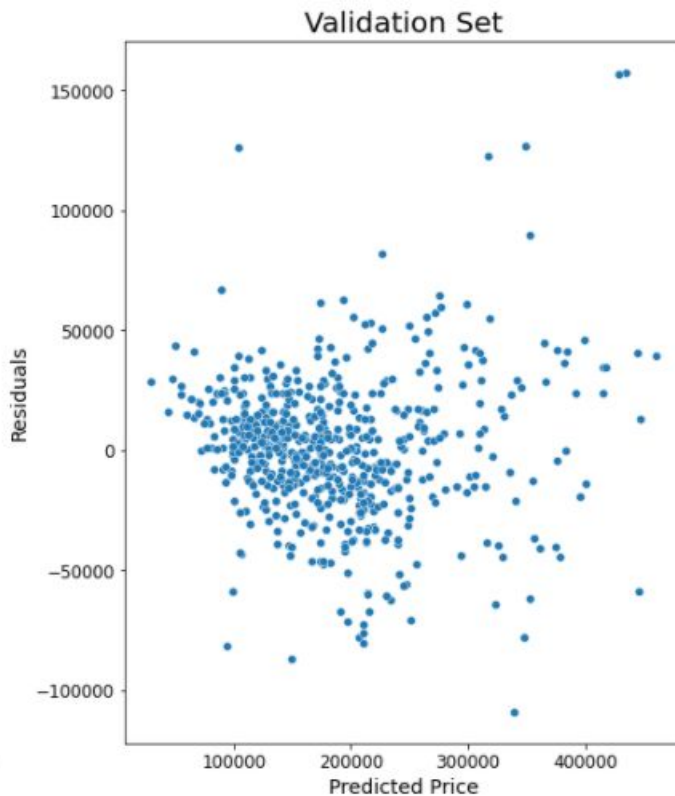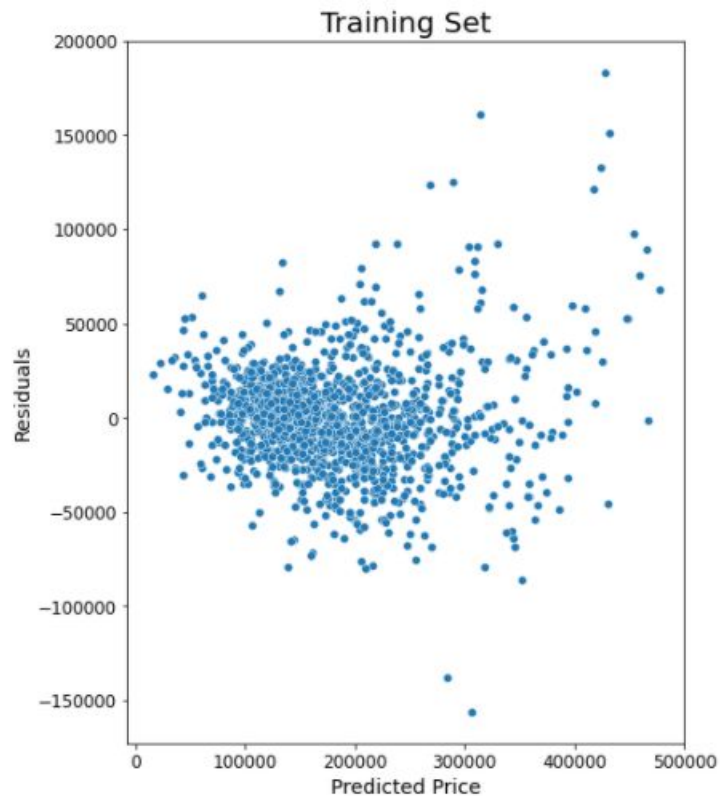
Model Predictions

Actual vs Predicted Price

- Strong linear relationship
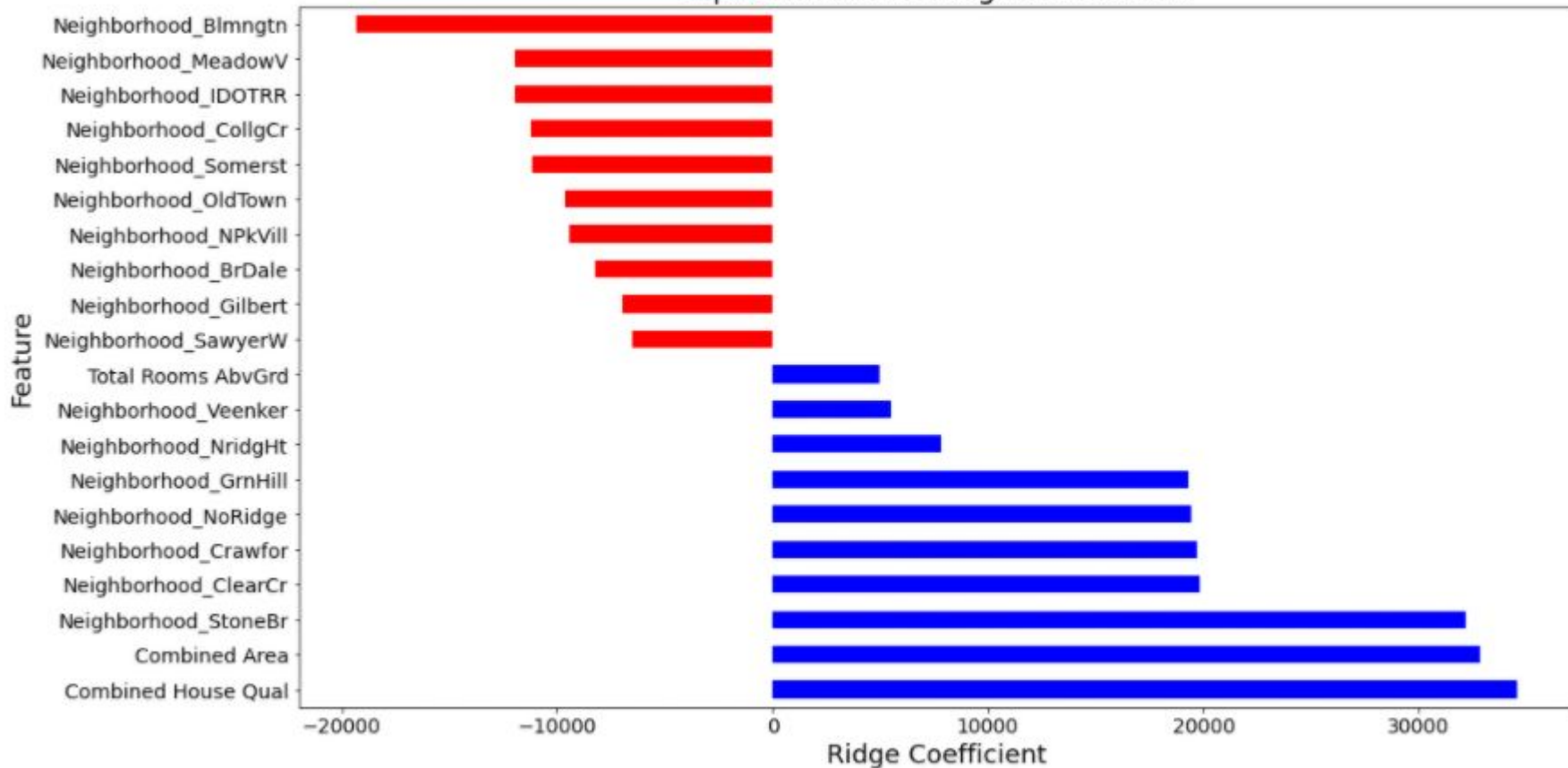- Deviations above $300,000

Model Predictions

Residuals Plot

Key Predictors

Top & Bottom 10 Ridge Coefficients

Key Predictors

- Combined house quality

- Combined area

- Neighborhood (Stonebrook)

Conclusion

Factors influencing home prices

- House size and quality

- Location

Important considerations excluded from data set

- Economic conditions

- Housing market bubble 2007-2008

Thank you!