Project 2

Houses Price Prediction using Linear Regression

Trevor Chang

Overview

In this project, I will use the Ames housing data to create a multiple linear regression model that predicts the price of houses in Ames, IA.

Data

Predictors:

All kinds of characteristics of houses, most of them are categorical variable

Response:

Sale price

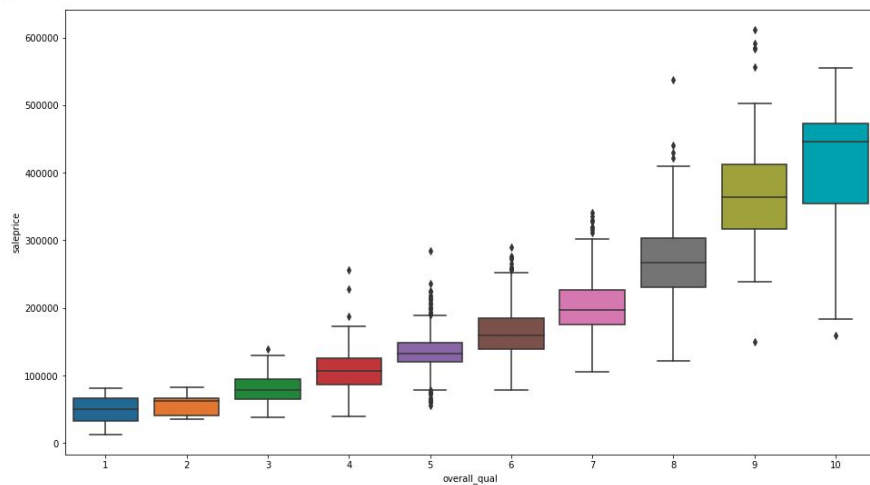# Interesting Features

Overall_Qual:

Already highly correlated with sale_price

+   Exterior Qual
+   Bsmt Qual
+   Kitchen Qual

      Even more correlated (0.83+)

Average room area and total baths

From my past experiences:

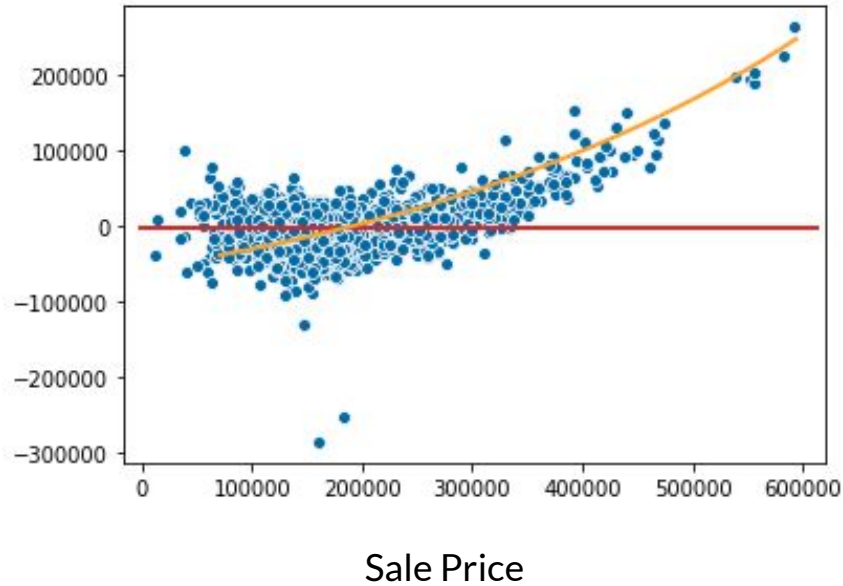Average rooms and total baths are import

However:

Baths need to be weighted
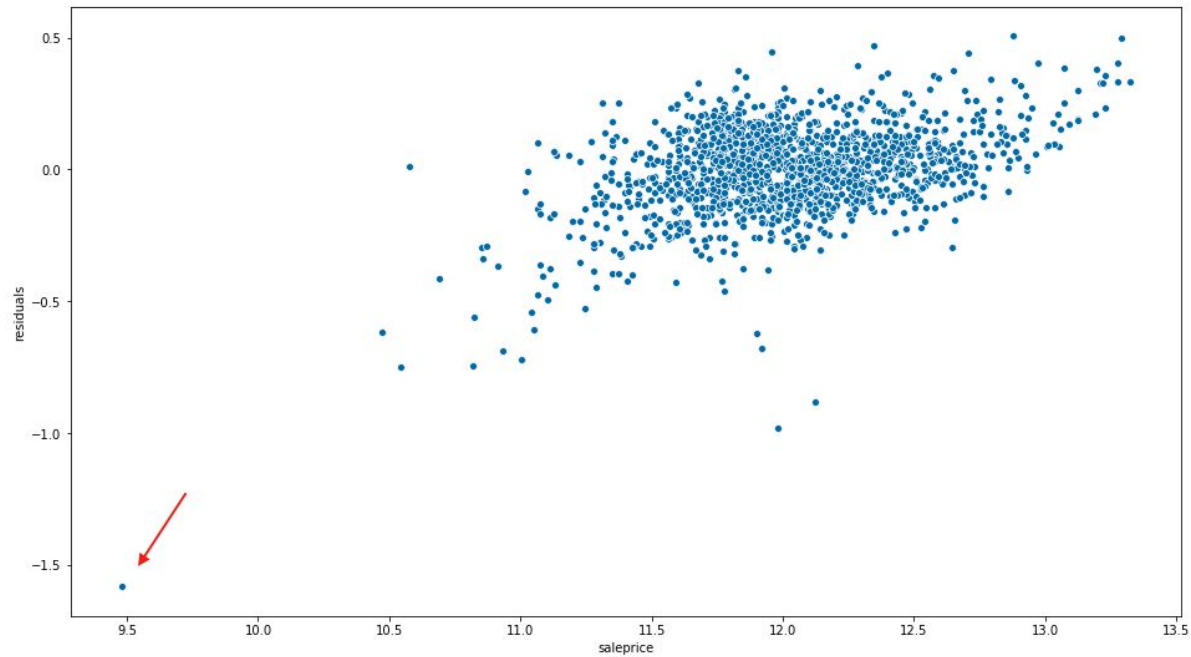
Half_baht = 0.5 * full_bath

Residuals Plot

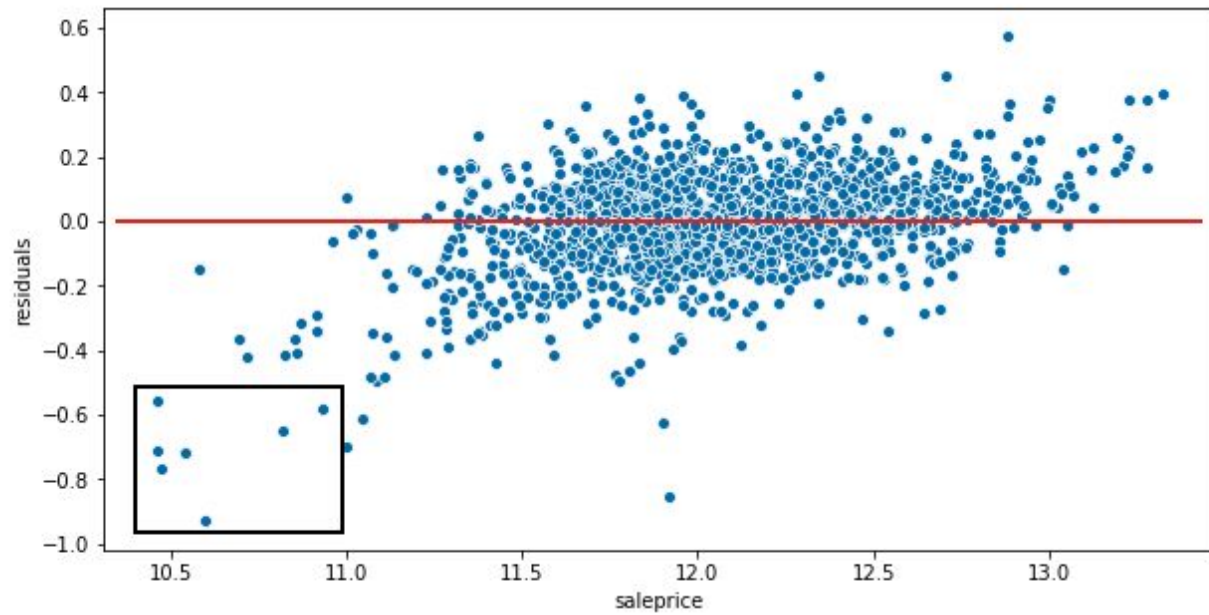Problem 1:

Residuals

Sale Price

Residuals Plot

Problem 2:

Fixed

Result

Training data: R2: 0.8693

Testing data: R2 0.8553