Modeling Ames, Iowa

House Prices

Sean Bjork

Problem:

Can we develop an accurate* model for sale prices of houses in Ames, Iowa?

*Lower than $25,000 RMSE

About the data

Source: Ames, Iowa Assessor's Office

Shape:

2051 rows (houses) &

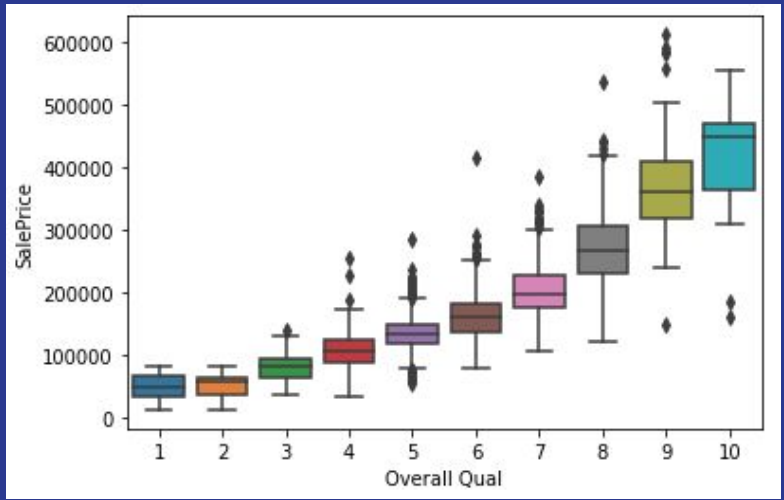
81 columns (variables)

Prominent Features:

Overall quality score and

Above-ground living area (Sq. Ft.)

Overall Quality vs. Sales Price

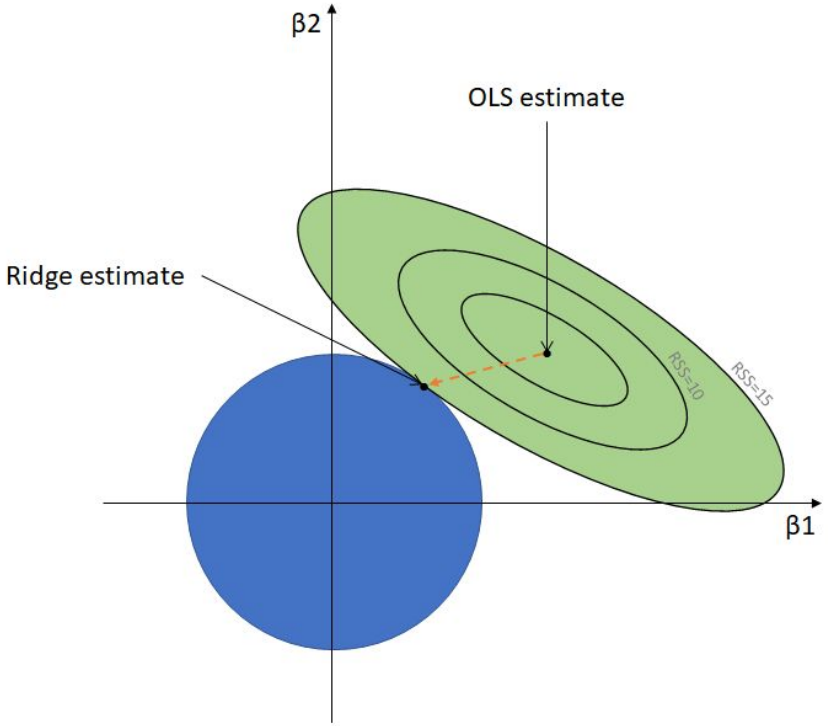


Sale Price vs. Selected Features

Polynomial Features

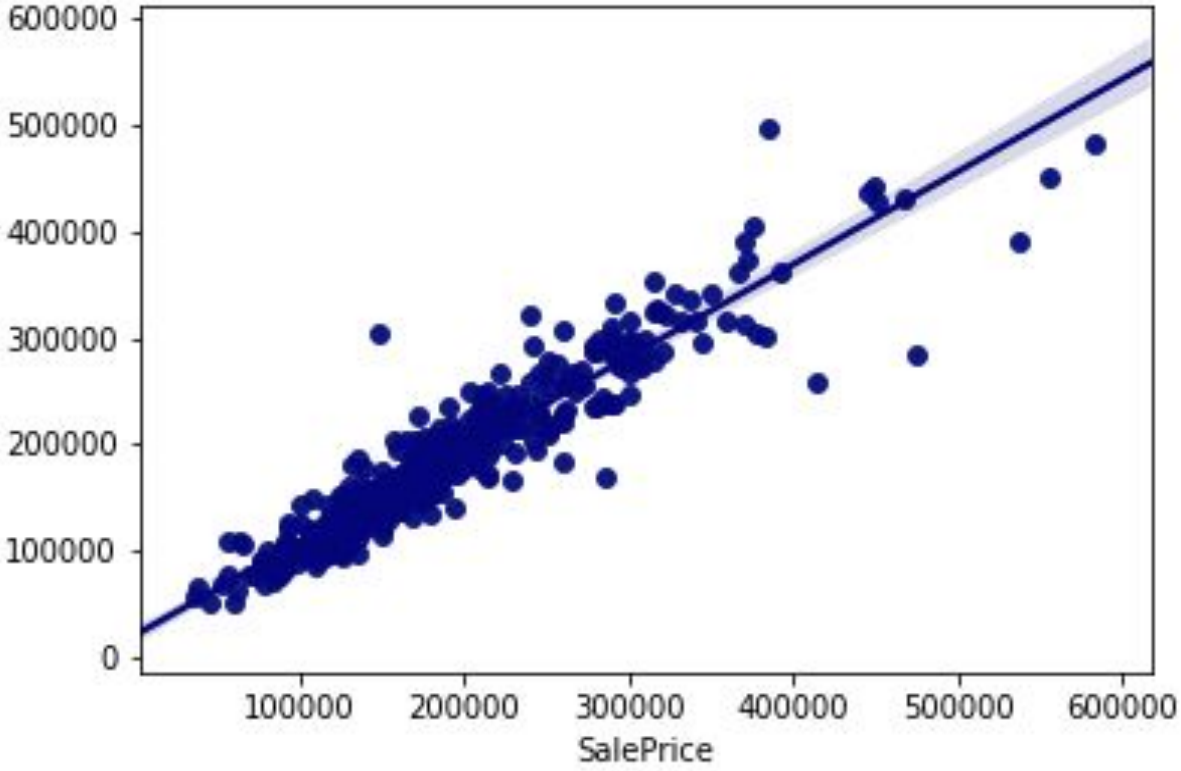
SalePrice	1.000000
Garage Yr Blt Neighborhood Projections	0.912233
Year Built Neighborhood Projections	0.912168
Year Remod/Add Neighborhood Projections	0.912152
Interaction Neighborhood Projections	0.911274
Neighborhood Projections	0.910895
Interaction TotRms AbvGrd Projections	0.898637
Overall Qual Neighborhood Projections	0.898630
Year Remod/Add Neighborhood * Total Rooms Projections	0.897904
Neighborhood Projections TotRms AbvGrd Projections	0.897777

Ridge Regression



$$ax^2+bx+c$$

True Target Values vs. Model Predictions



Training R2:	0.881
Test R2:	0.889
RMSE:	24586

Conclusions and Limitations

- We succeeded in developing a model with a root mean squared error under $25,000 ($24,586).

- In its current form, the model is not suitable for data without target column.
  - Tried to standardize model using historical data - was not successful.