Housing Prices of Aimes, Iowa

Aakash Sharma

How modeling a regression can predict prices of housing developments

Problem

We are given a dataset that is relatively complete. The data set contains multiple errors that can corrupt the data & skew the graphs, making it difficult for us to find patterns or correlations to a variable we would like to predict. Our variable sales price can be used to sell someone on a dataset for considering changing or purchasing a house in a certain neighborhood.

We want to clean the data & model it using linear, ridge & lasso regressions to predict the sales price of the house. We can pass in features of the data frame to help with predicting this or help correlate a relationship.

Steps

EDA/Data Cleaning

EDA

- Look at distributions
- Look at correlations
- Look at relationships to target

Data Cleaning

- How to impute null value.
- How to handle outliers
- Want to combine any features
- **Want to have interaction terms**
- Want to manually drop collinear features

Model Prep

Train/Test Split

- Look at distributions

Scale the Model

- Standardize the data
- Useful for data which has negative values
- Arranges the data in a normal distribution

Instantiate the Model

Cross Validation

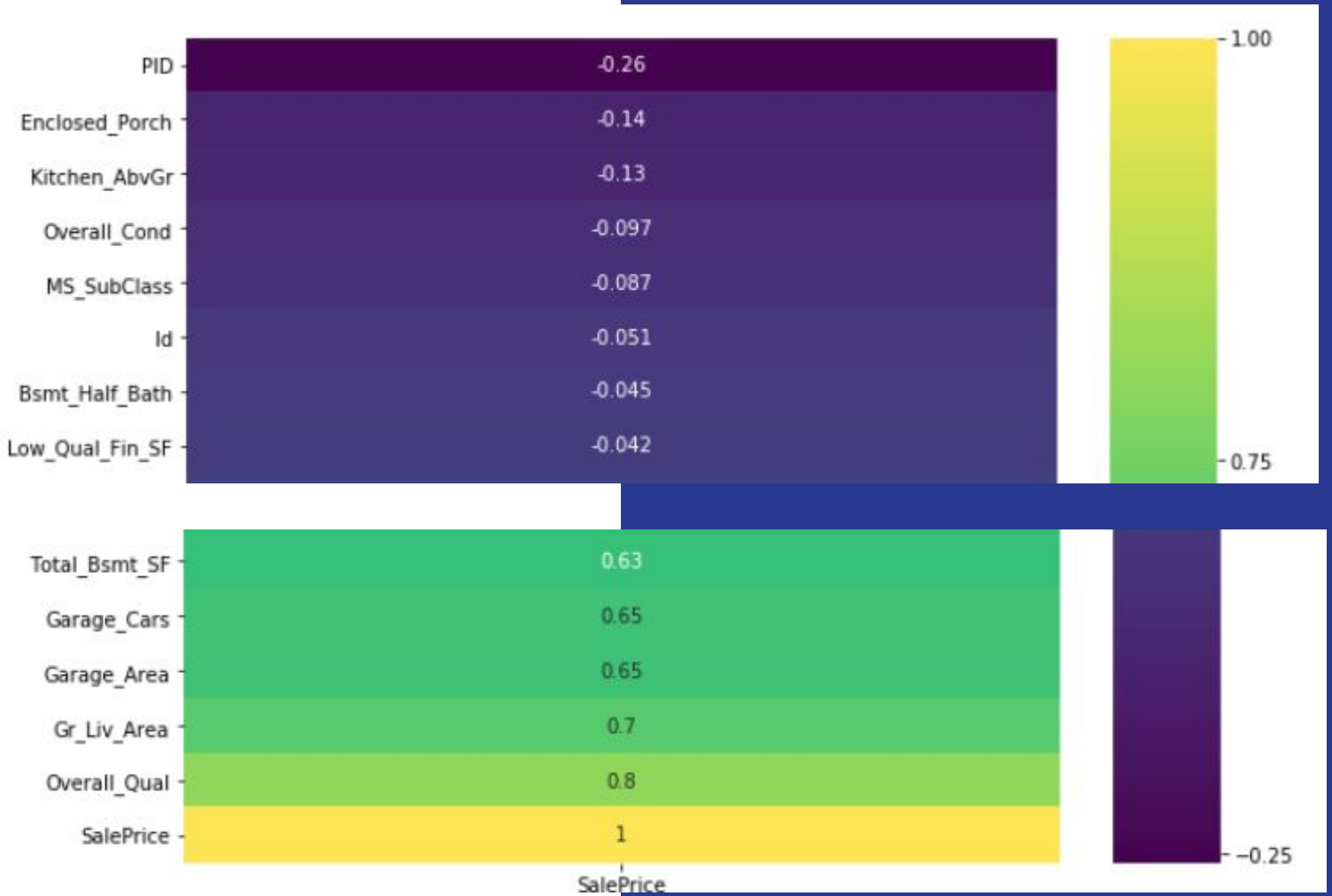
- On the 3 models to see which model is the best for our data
- Applied to more subsets.

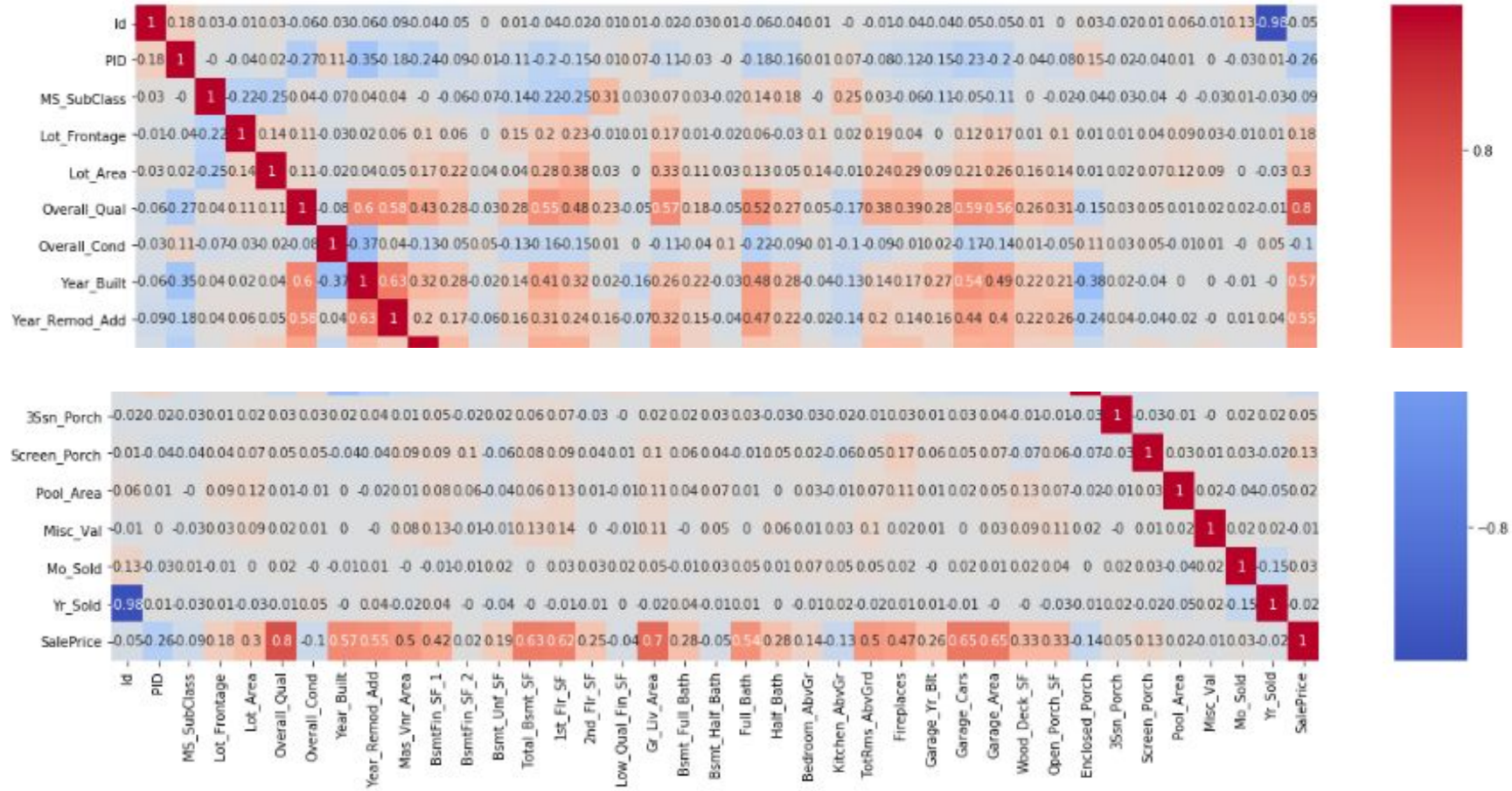
The Model

Linear Regression

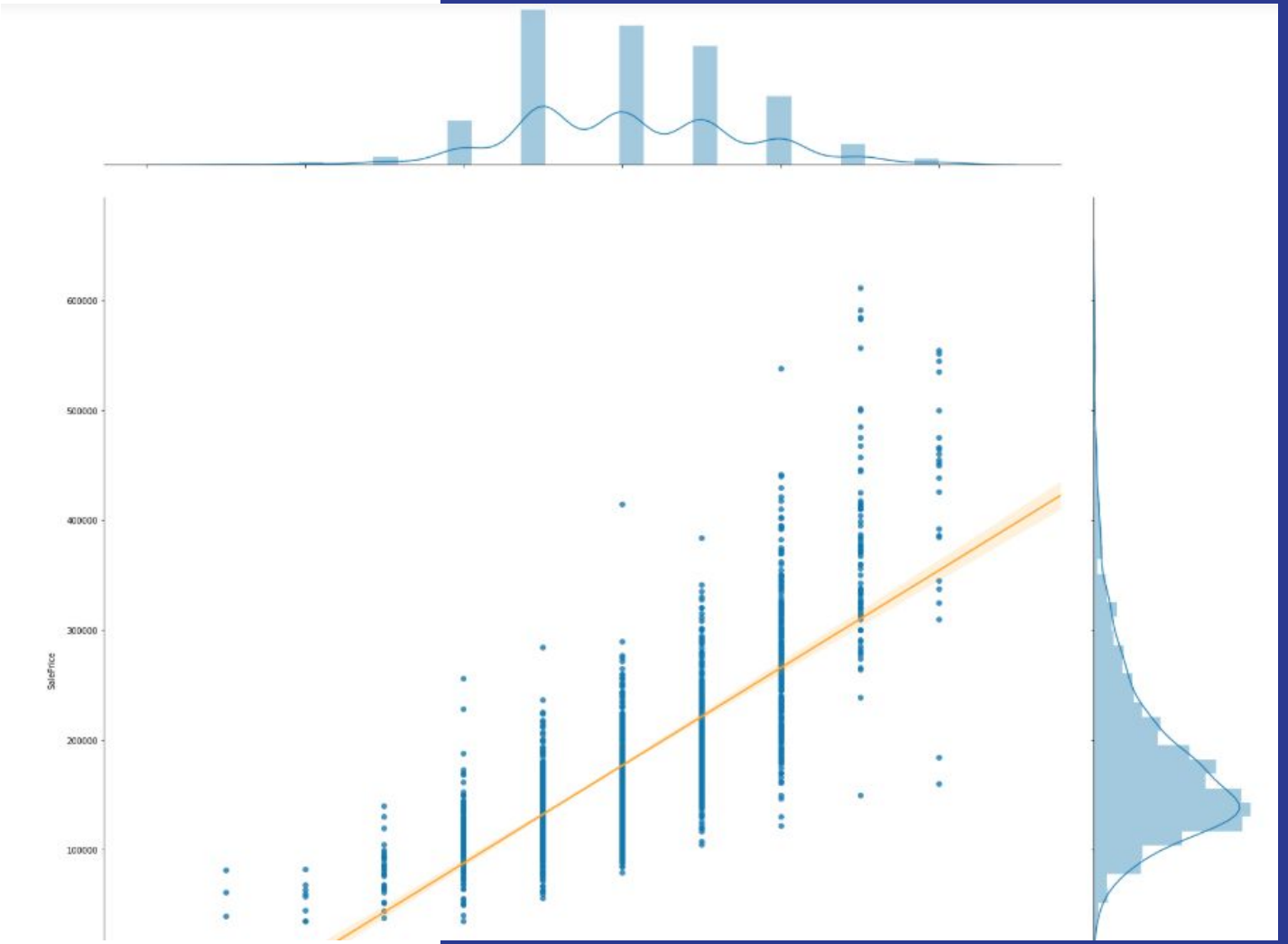
Lasso Regression - The Best

Ridge Regression - Runner Up


PID
Enclosed_Porch
Kitchen_AbvGr
Overall_Cond
MS_SubClass
Id
Bsmt_Half_Bath
Low_Qual_Fin_SF
-0.26
-0.14
-0.13
-0.097
-0.087
-0.051
-0.045
-0.042
1.00
0.75
Total_Bsmt_SF
Garage_Cars
Garage_Area
Gr_Liv_Area
Overall_Qual
SalePrice
0.63
0.65
0.65
0.7
0.8
1
−0.25
SalePrice


Id
PID
MS_SubClass
Lot_Frontage
Lot_Area
Overall_Qual
Overall_Cond
Year_Built
Year_Remod_Add
3Ssn_Porch
Screen_Porch
Pool_Area
Misc_Val
Mo_Sold
Yr_Sold
SalePrice
0.8
−0.8

Impact



Business Recommendations

- Which features appear to add the most value to a home?

  Square Feet, Overall Quality, Garage Area/Cars, Beds & Baths

- Which features hurt the value of a home the most?

  Having a kitchen above the garage, the condition of the home

- What are things that homeowners could improve in their homes to increase the value?

  Homeowners could renovate the home, make sure the overall condition is on the same level as the overall quality

- What neighborhoods seem like they might be a good investment?

  Ames, Iowa was ranked the #1 choice based on median monthly rent of $785 according to “The Crazy Tourist.” The best neighborhoods would be those that fit in that area & according to the data, that would

Business Recommendations Cont.

- Do you feel that this model will generalize to other cities? How could you revise your model to make it more universal OR what date would you need from another city to make a comparable model?

  I would add other features that could correlate to an increase in a sales price of a house. Maybe even use some interaction based columns to help see correlations between 2 or multiple columns.

- Do you feel that this model will generalize to other cities? How could you revise your model to make it more universal OR what date would you need from another city to make a comparable model?

  The model would not generalize to other cities based on the sample of an entire population. We would have to use every possible variation of independent variable which differs tremendously to measure every variation of dependent variables.