Ames Housing Predictions

Mitchell Meislin

Problem Statement

1. How close can we get to predicting the final sale price of a house in the Ames data set?
2. What features are most important in predicting sale price of a house in the Ames data set?
3. How is neighborhood associated with final sale price of a house in the Ames data set?

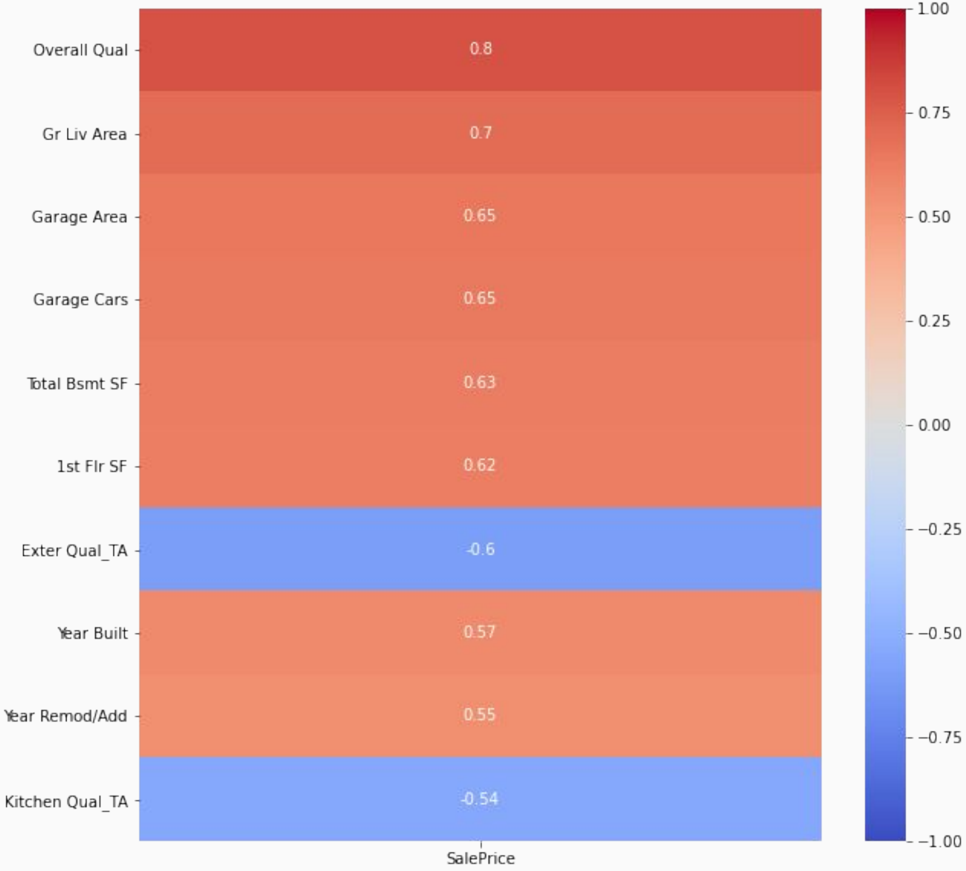
Why Does It Matter?

1. This model can be used to give buyers a more accurate estimate of housing prices.

2. This model can help homeowners figure out the value of their home.

3. This can show homeowners how to best increase the value of their home

Strongest Correlates to Sale Price



Procedure:

- Find all column with data type object in training and testing data
- Turn all object data into dummies variables
- Match all columns from training and testing data (keep sale price)
- Fill all null numerical values with the mean
- Now you have a cleaned data set and can start initial modeling!!

Creating your models!!

- Set your X variable as all columns except Sale Price. Set your y variable as Sale Price.
- Run train test split, call linear regression and then fit your training data to your regression
- Check R2 scores, cross valuation scores and most importantly RMSE!
- Have a look at your top 10 correlated variables (from the heatmap) and their coefficients sorted by value!

	Variable Name	Coefficient Value
0	Overall Qual	6882.710389
1	Year Built	174.325393
2	Year Remod/Add	44.067666
3	Total Bsmt SF	7.641452
4	1st Flr SF	11.502933
5	Gr Liv Area	27.246149
6	Garage Cars	8619.707201
7	Garage Area	2.650013
8	Exter Qual_TA	-32375.395191
9	Kitchen Qual_TA	-26282.449721

Reiterate your model!

- Start by scaling features in your current model for an apples to apples comparison
- Next run a lasso function and try out many alphas to find a more optimal fit!
- Fit your model!
- Now check R2 scores, cross valuation scores and most importantly RMSE!
- This lasso function just slightly helped my model.

Feature Engineering

- Make a list of features to drop which have the smallest coefficients in my model!
- These features have been scaled and can be used for an apples to apples comparison.
- This function can change upper and lower bounds to cut off more or less features.

```
my_coefficients= list(zip(X.columns, lasso_cv.coef_))
#print([my_coeficients])
my_list_of_deletes = []
for i in range(0, len(my_coefficients)):
    my_coefficients[i][1]
    if my_coefficients[i][1] < 3000 and my_coefficients[i][1] > -3000:
        my_list_of_deletes.append(my_coefficients[i][0])
my_list_of_deletes
```

Fit The New Data To Have Polynomial Features

- Set your X variable as all columns except Sale Price. Set your y variable as Sale Price.
- Add polynomial features to your X!
- Check the new shape is (w, z) where w>z!
- Run train test split, call linear regression and then fit your training data to your regression
- Check R2 scores, cross valuation scores and most importantly RMSE!
- This model is overfit!

Reiterate your model!

- Start by scaling features in your current model for an apples to apples comparison

- Next run a lasso function and try out many alphas to find a more optimal fit!

- Fit your model!

- Now check R2 scores, cross valuation scores and most importantly RMSE!

- This model got me to my final prediction with an RMSE of 23082.15.

Conclusions:

- This model can help predict the price of a home in the ames data with error of roughly $23082.15. This model can help homeowners and interested home buyers asses the value of a home!

Recommendations

1. All neighborhoods and their associated prediction for sale price of a home (all else being held equal)

	Neighborhood Name	Coefficient Value
22	Neighborhood_StoneBr	33091.187974
16	Neighborhood_NridgHt	21509.021303
15	Neighborhood_NoRidge	17750.146644
21	Neighborhood_Somerst	-1132.742310
13	Neighborhood_NPkVill	-4191.973771
8	Neighborhood_Greens	-6495.283524
1	Neighborhood_BrDale	-7233.413614
24	Neighborhood_Veenker	-9033.538074
5	Neighborhood_Crawfor	-14190.583482
23	Neighborhood_Timber	-23557.201438
0	Neighborhood_Blueste	-23613.409059
20	Neighborhood_SawyerW	-23759.311360
7	Neighborhood_Gilbert	-23859.888164
10	Neighborhood_MeadowV	-24798.064956
3	Neighborhood_ClearCr	-25467.362571
4	Neighborhood_CollgCr	-25477.644798
14	Neighborhood_NWAmes	-26214.152334
12	Neighborhood_NAmes	-28086.176463
2	Neighborhood_BrkSide	-28643.230386
11	Neighborhood_Mitchel	-29079.413927
19	Neighborhood_Sawyer	-30149.989479
9	Neighborhood_IDOTRR	-33977.940798
18	Neighborhood_SWISU	-34907.252093
17	Neighborhood_OldTown	-35063.578194
6	Neighborhood_Edwards	-42558.069846

Recommendations

- Top ten correlated features to sale price

- All else held equal, each extra unit for every one of these variable results in the associate increase in sale price

	Variable Name	Coefficient Value
0	Overall Qual	6882.710389
1	Year Built	174.325393
2	Year Remod/Add	44.067666
3	Total Bsmt SF	7.641452
4	1st Flr SF	11.502933
5	Gr Liv Area	27.246149
6	Garage Cars	8619.707201
7	Garage Area	2.650013
8	Exter Qual_TA	-32375.395191
9	Kitchen Qual_TA	-26282.449721

THANKS FOR LISTENING!!!