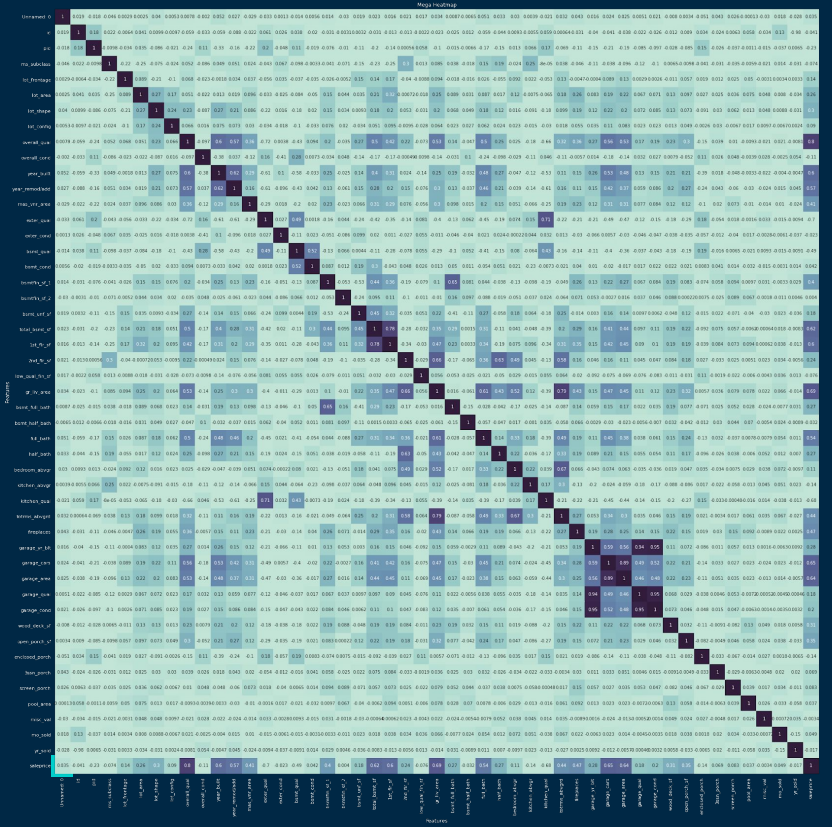
Ames Housing Modeling

Can we find a more concise model?

Problem Statement:

Can we create an efficient model for predicting housing prices and productionize it?

MEGA HEATMAP



Why?

We all know we wanted it. Let’s get it out of our systems.

Let’s make this useful

Overall quality house/land etc.

Overall Quality

0.8

0.4

Basement

Square Footage of basement

Above Ground Living square footage

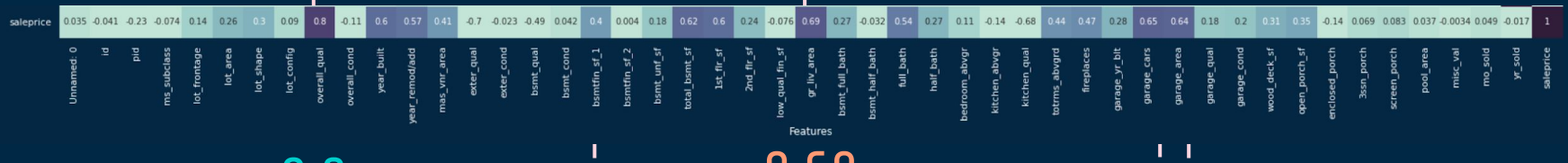
Living Area

0.69

0.64–0.65

Garage Space

Both # of cars that can fit in the garage and overall square footage



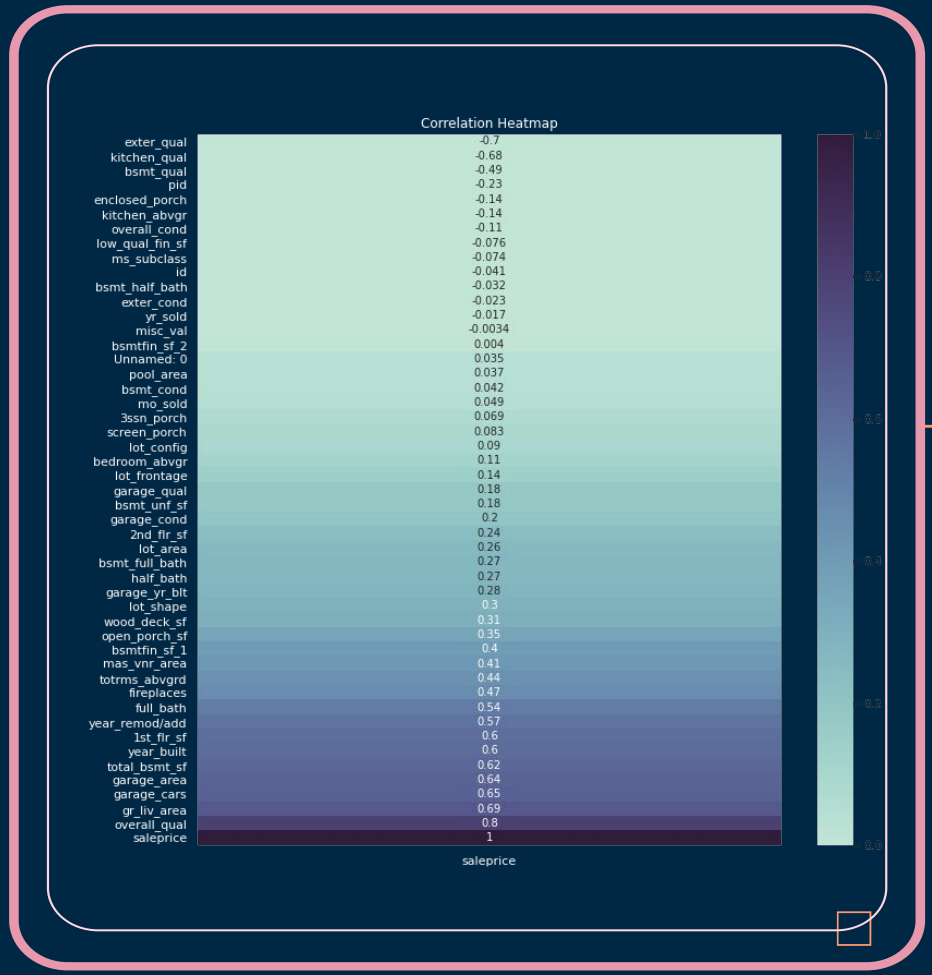
Better Tools

Fortunately There are
better tools in our toolbox

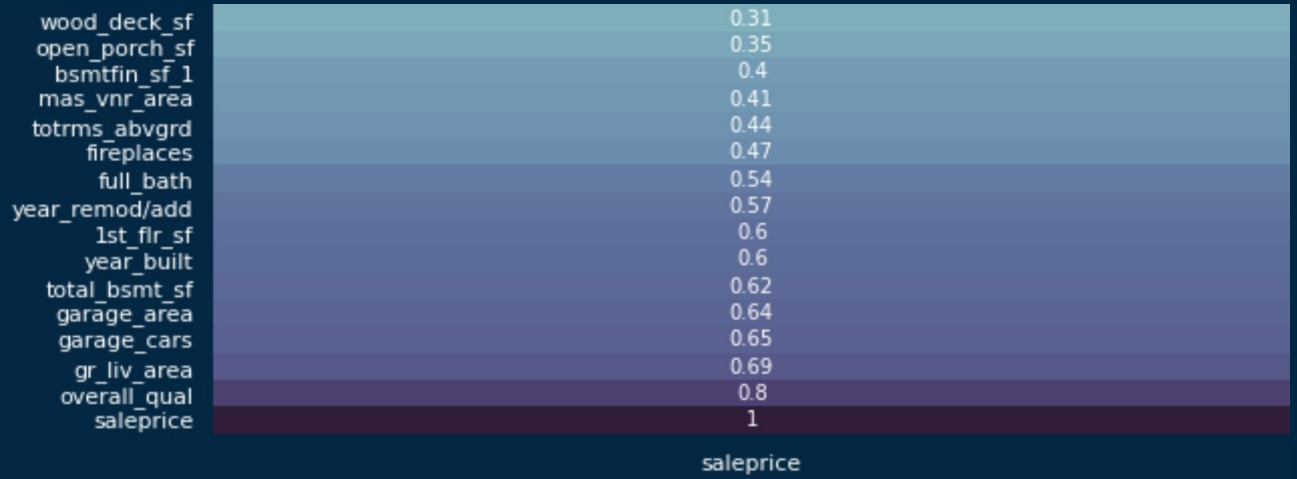
Correlation Heatmap

All features and their correlation level with Sale Price.

1 = Highest corr
0 = Lowest corr



Our top contenders:



Overall Quality

Overall Quality of the house has largest impact on sale price

Above Ground Living Area

Living area in square ft.

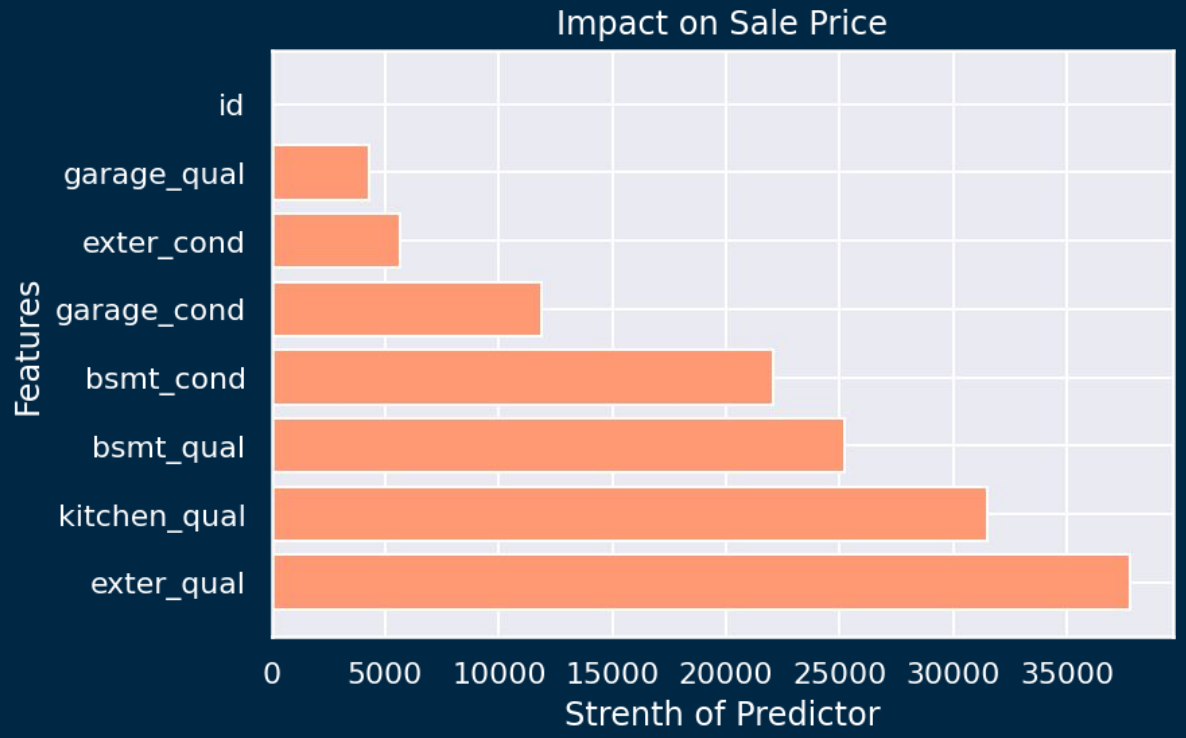
Car Garage

Number of cars that can fit in the garage

Confidence Predictors 1st set

Chosen Features:

Garage Quality
Exterior Quality
Garage Condition
Basement Condition
Basement Quality
Kitchen Quality
Exterior Quality



Can we do better?

Confidence Predictors 4th set

Chosen Features:

1st Floor sft

Total Basement sft

Garage Area sft

Ground Level area

Basement Year Built

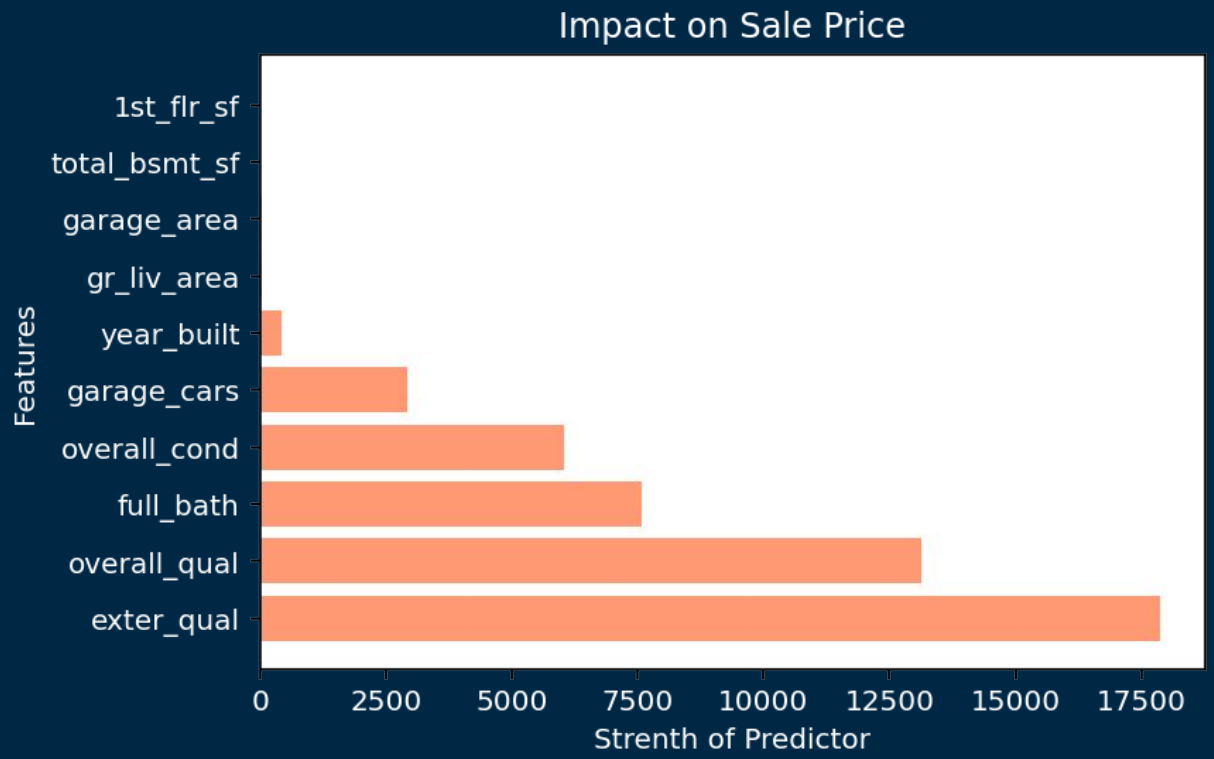
Cars in garage

Overall Condition

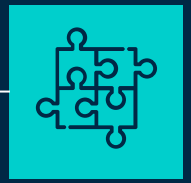
Full Bath

Overall Quality

Exterior Quality



Ridge Scores



Qual

0.596

Here you could describe the topic of the section

Lot

0.171

Here you could describe the topic of the section

Min

0.759

Here you could describe the topic of the section

Corr

0.820

Here you could describe the topic of the section

Product Process

Step 1

Input Data

Takes housing data and cleans it including removing outliers

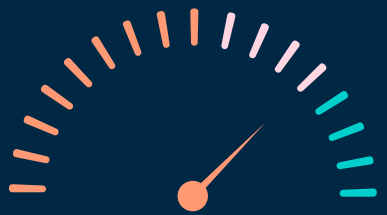
Step 2

Run the model

Determines the best metrics to use to determine housing pricing predictions

Step 3

Receive predictions!

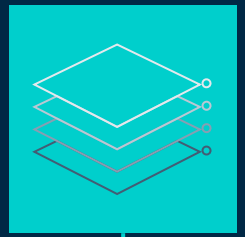


Presents predicted material

0.820

Using our heatmap helped us find the best correlated data for our model.

Ranking



14	Kira Helm		35326....

Conclusions

Is it accurate enough?

Having a test ridge regression score of over 80% is pretty good for this model

Consider Efficiency

I think this score could be improved, but the real question is how efficient we can make it

More funding needed

To improve and ship product.

THANKS


Questions?