

# Predicting Home Sale Prices

Arink Bertrand

Fellow at General Assembly - DSI course

# The process

Define Problem

Explore Data

Model Data

Answer Problem



Can we predict the  
Home Sale Prices in  
Ames, Iowa based on  
certain features?

# The process

Define Problem

Explore Data

Model Data

Answer Problem



Can we predict the Home Sale Prices in Ames, Iowa based on certain features?

What to do with empty values, which features to choose to help predict the actual values.

# The process

Define Problem

Explore Data

Model Data

Answer Problem



Can we predict the Home Sale Prices in Ames, Iowa based on certain features?

What to do with empty values, which features to choose to help predict the actual values.

Which model can help best predict the closest Sale Price to the Actual, make changes to feature choices as needed.

# The process

Define Problem

Explore Data

Model Data

Answer Problem



Can we predict the Home Sale Prices in Ames, Iowa based on certain features?

What to do with empty values, which features to choose to help predict the actual values.

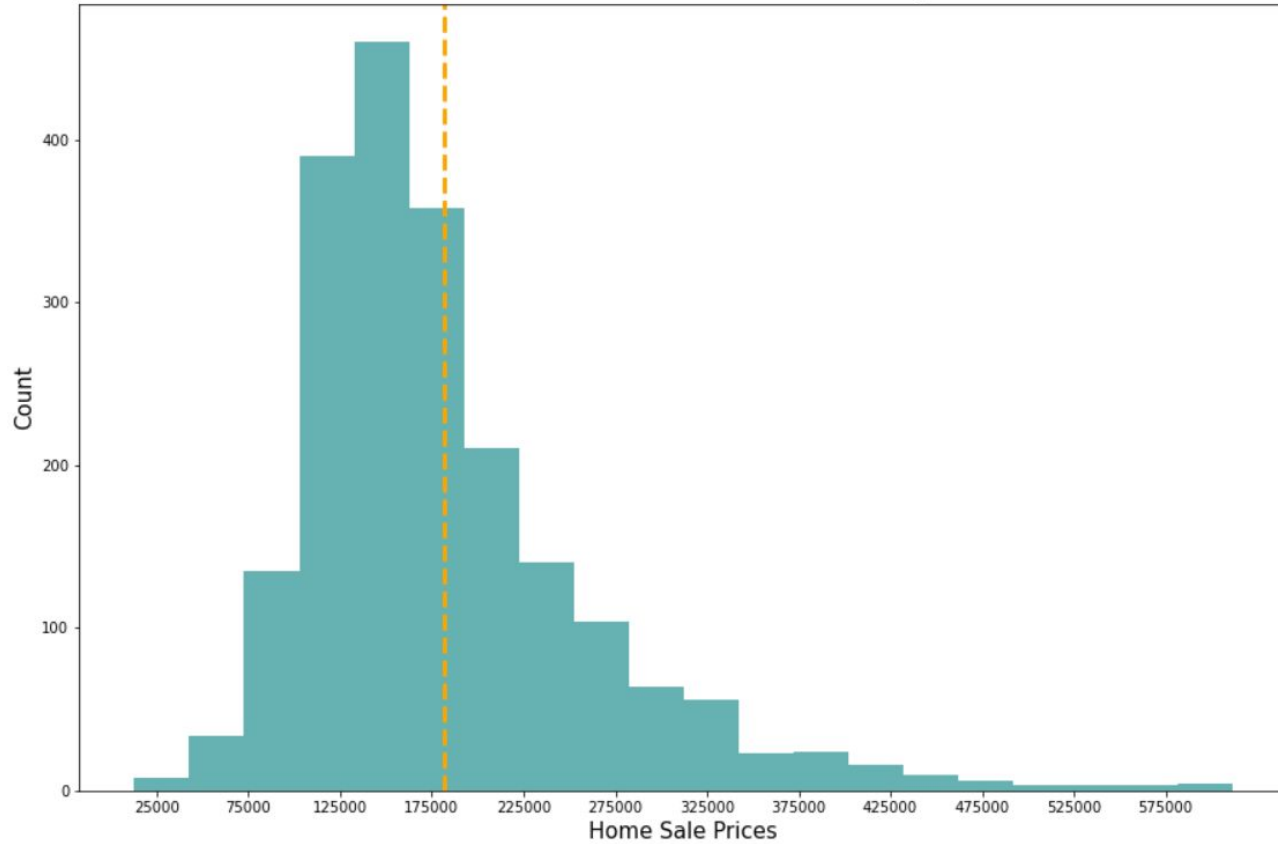
Which model can help best predict the closest Sale Price to the Actual, make changes to feature choices as needed.

Now that we have our predictions, what are our next steps?

# Why do we need to predict Sale Prices?

- Buying or Selling Homes
- Remodeling
- Investing

Home Sales Price Distribution in Ames, Iowa

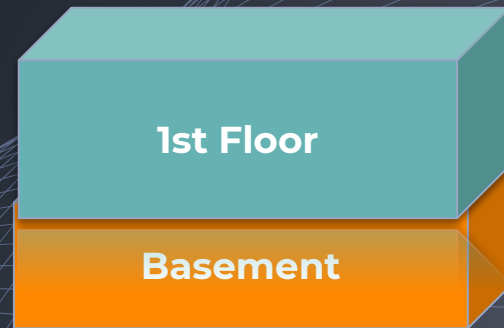


# Taking care of **Missing** Values

	Pool Quality
GD	4
TA	2
Fa	2
EX	1
NaN	2042



	Has Pool
0	2042
1	9







# chosen Features

SalePrice  
100%

Overall Qual  
80%

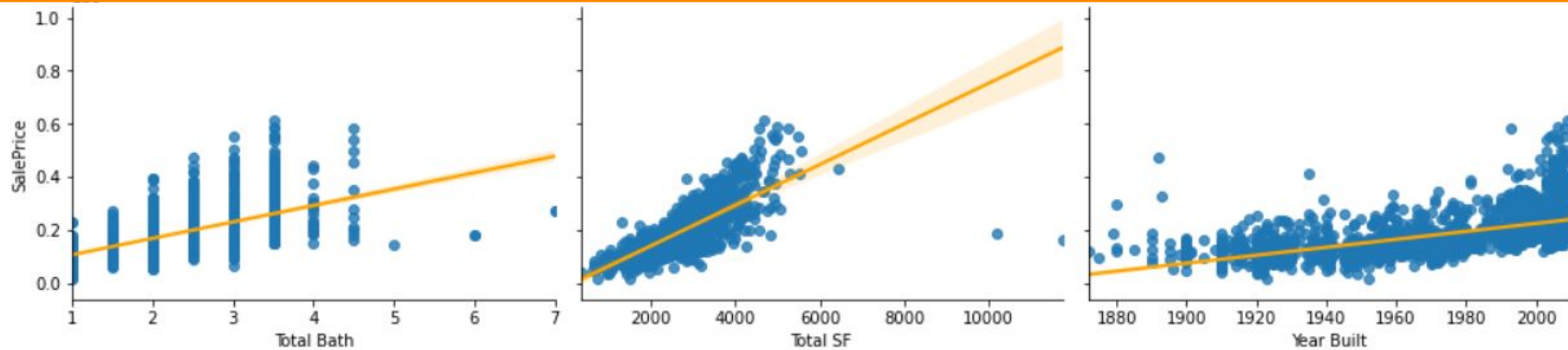
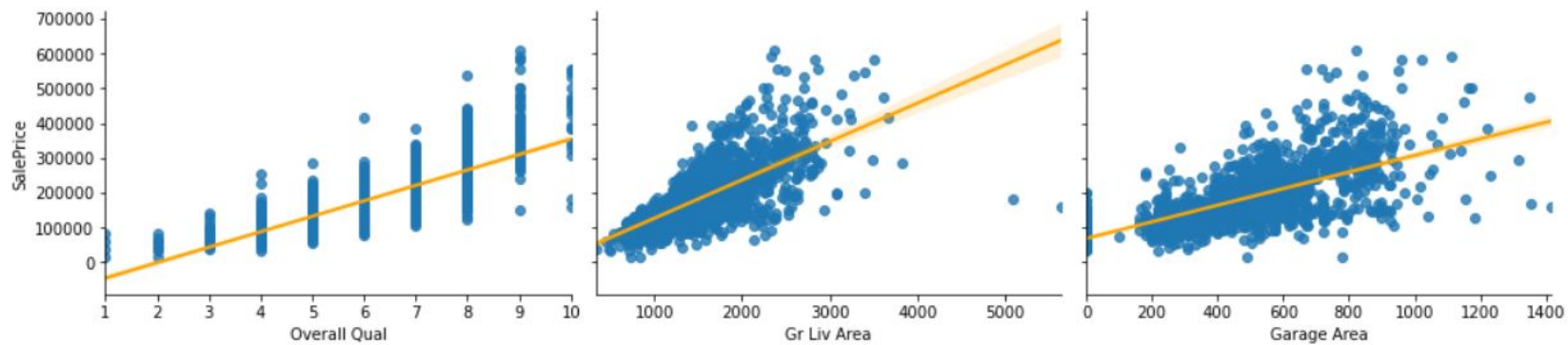
Total SF  
78%

Gr Liv Area  
70%

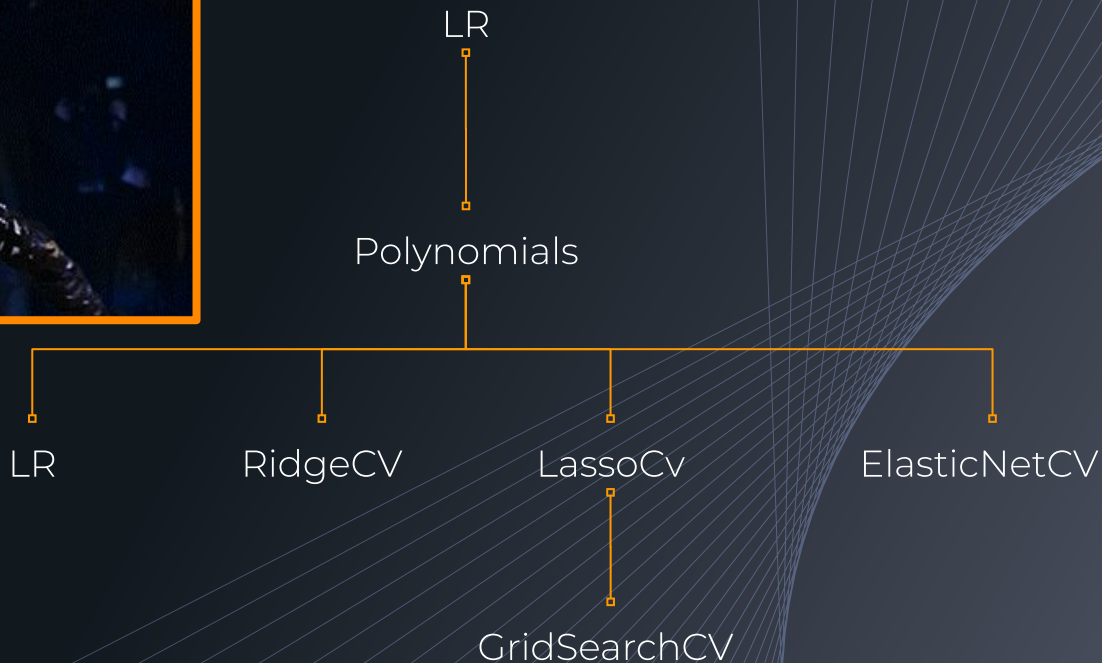
Garage Area  
65%

Total Bath  
63%

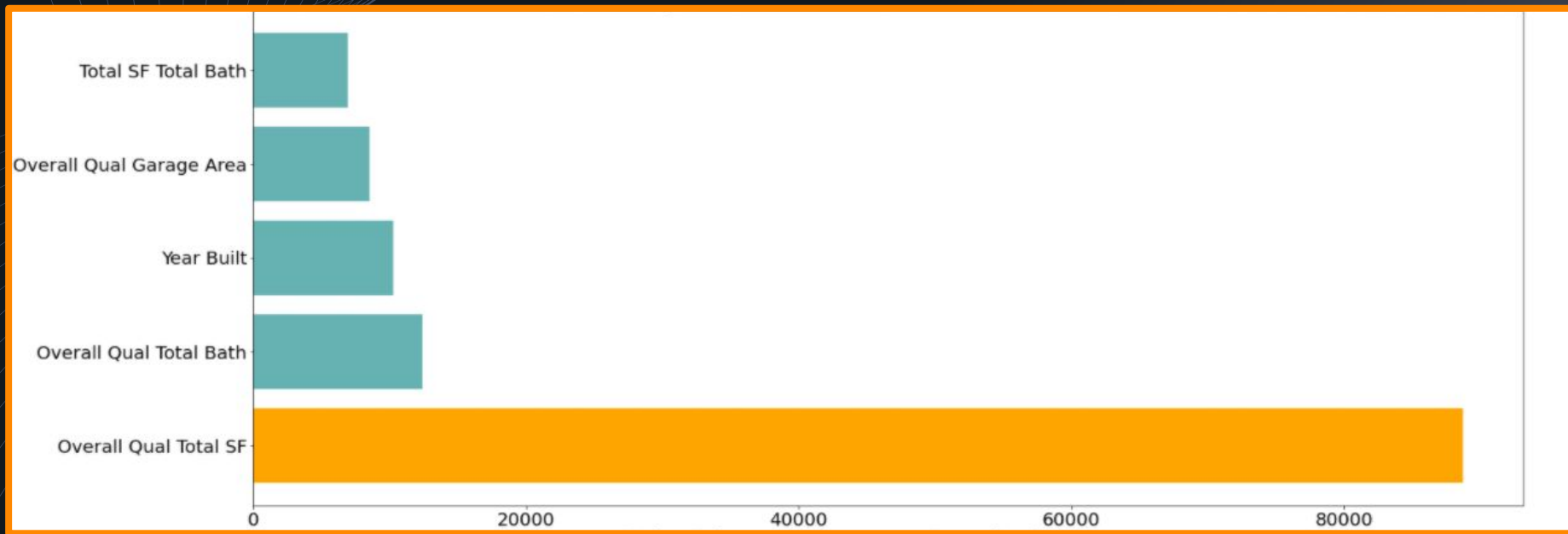
Year Built  
57%



# Let's put our model to work



# Top 5 Feature Predictors



**\$25, 182.79**

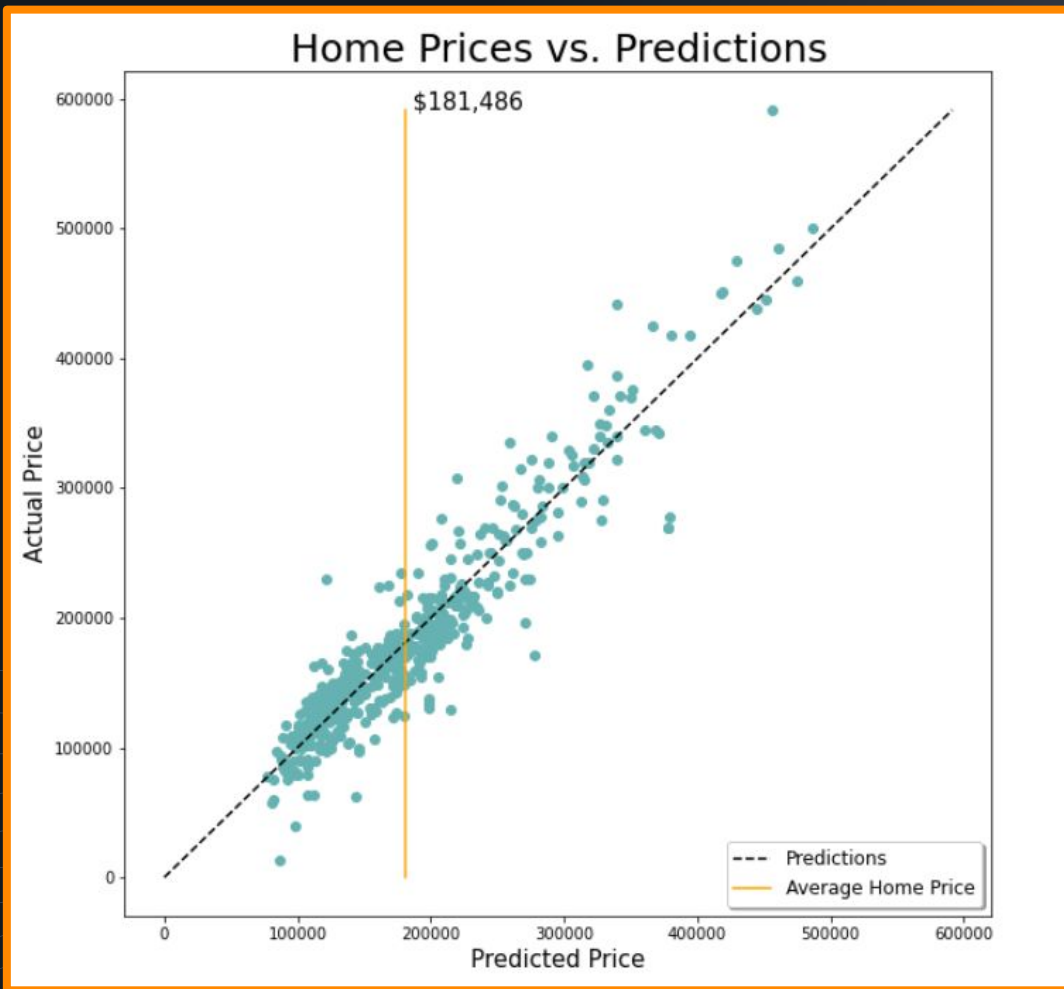
To Actual Price

**\$181,486**

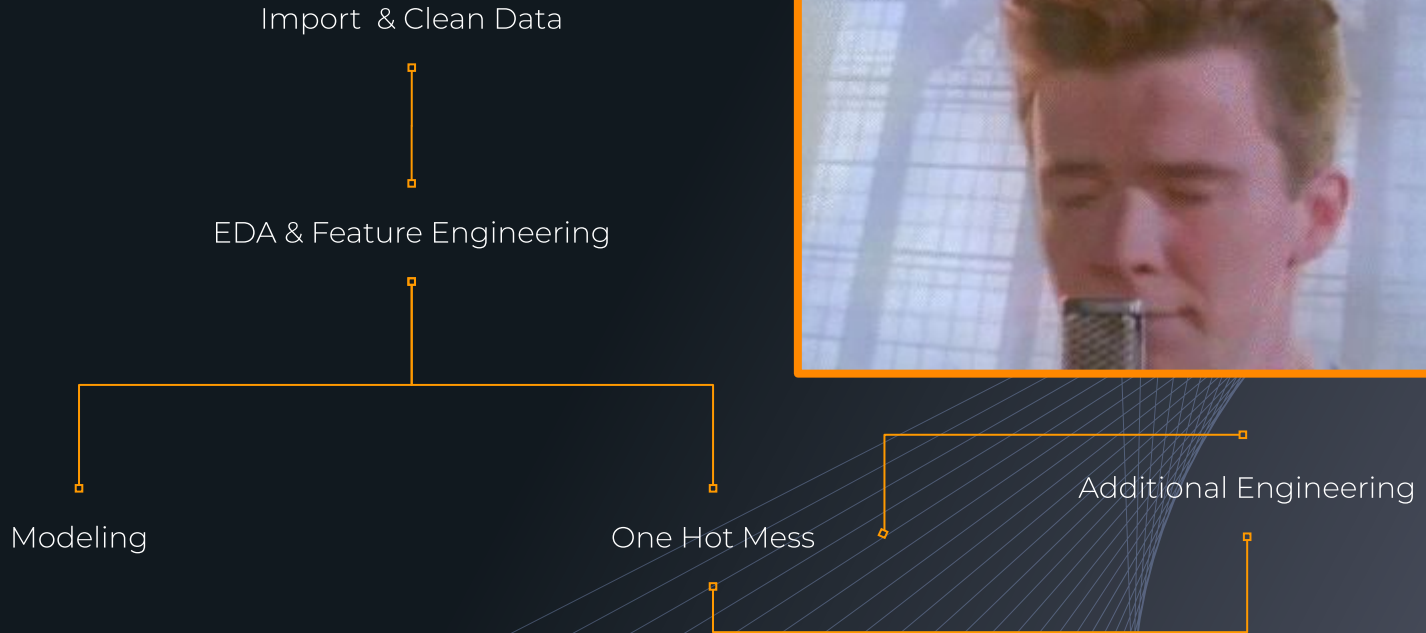
Average Home Price

**88.95%**

Accountability



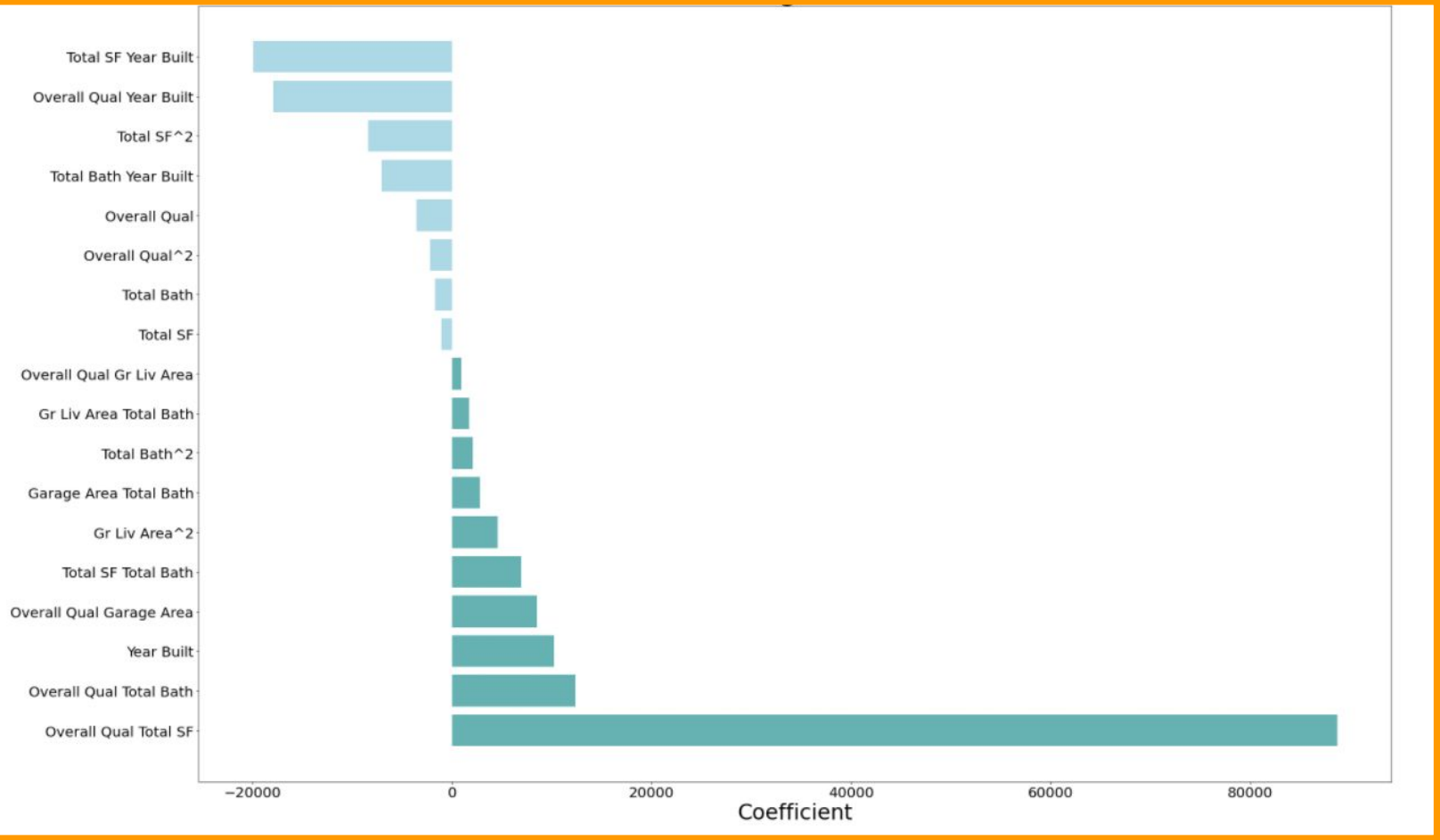
# Notebook Process





## Model Aim

Train Score	<b>0.9383</b>
Test Score	<b>0.9295</b>
Train RMSE	<b>19746</b>
Test RMSE	<b>20805</b>





# Model Score Comparisons

	Train RMSE	Kaggle RMSE	Notes
LR	<b>32336.27</b>	<b>37027.42</b>	<b>Overfit</b>
LR w. Poly	<b>26417.70</b>	<b>29072.82</b>	<b>Negative Coef_</b>
Ridge CV	<b>26305.96</b>	<b>29620.10</b>	
Lasso CV	<b>27384.13</b>	<b>297409.22</b>	
Elastic CV	<b>66293.64</b>	-	
LassoGS	<b>26410.38</b>	<b>26347.73</b>	<b>Final 25182.79</b>

	Train RMSE	Kaggle RMSE	Notes
LR	-	-	
LR w. Poly	<b>22624.51</b>	-	<b>Overfit</b>
Ridge CV	<b>22398.57</b>	-	<b>Overfit</b>
Lasso CV	<b>21016.95</b>	-	<b>Exciting!</b>
Elastic CV	-	-	
LassoGS	<b>20805.78</b>	<b>34390.38</b>	<b>WHAT?!?</b>

Thank You !