DON'T REGRET YOUR DECISION.

You're moving to Ames?!

Presented by Nader Esmael

Discussion Points

Quick Overview

Problem Statement ft. Timmy
What We Know
First Features & Results
More Models and Benchmarks
Coefficients to Consider
Competition?
What I recommend for Timmy

This is Timmy

Background

My friend is moving to Ames, IA. However, he doesn't know where to start in looking for a house. Since it is his first home purchase, he wanted to look at the prices of previously sold homes in the area to help influence his decision. He was kind enough to "acquire" the data of sold homes in the area. However...

Problem

Some data is missing(test data)! My task is to figure out the prices of the homes of missing sale price using Linear, Lasso, Ridge and other types of models and fit it to the "train" data. I am to determine what are the best features for his first house purchase.

WHAT WE KNOW

Average price of a home in Ames

~$181469.70

Number of Given Features

81

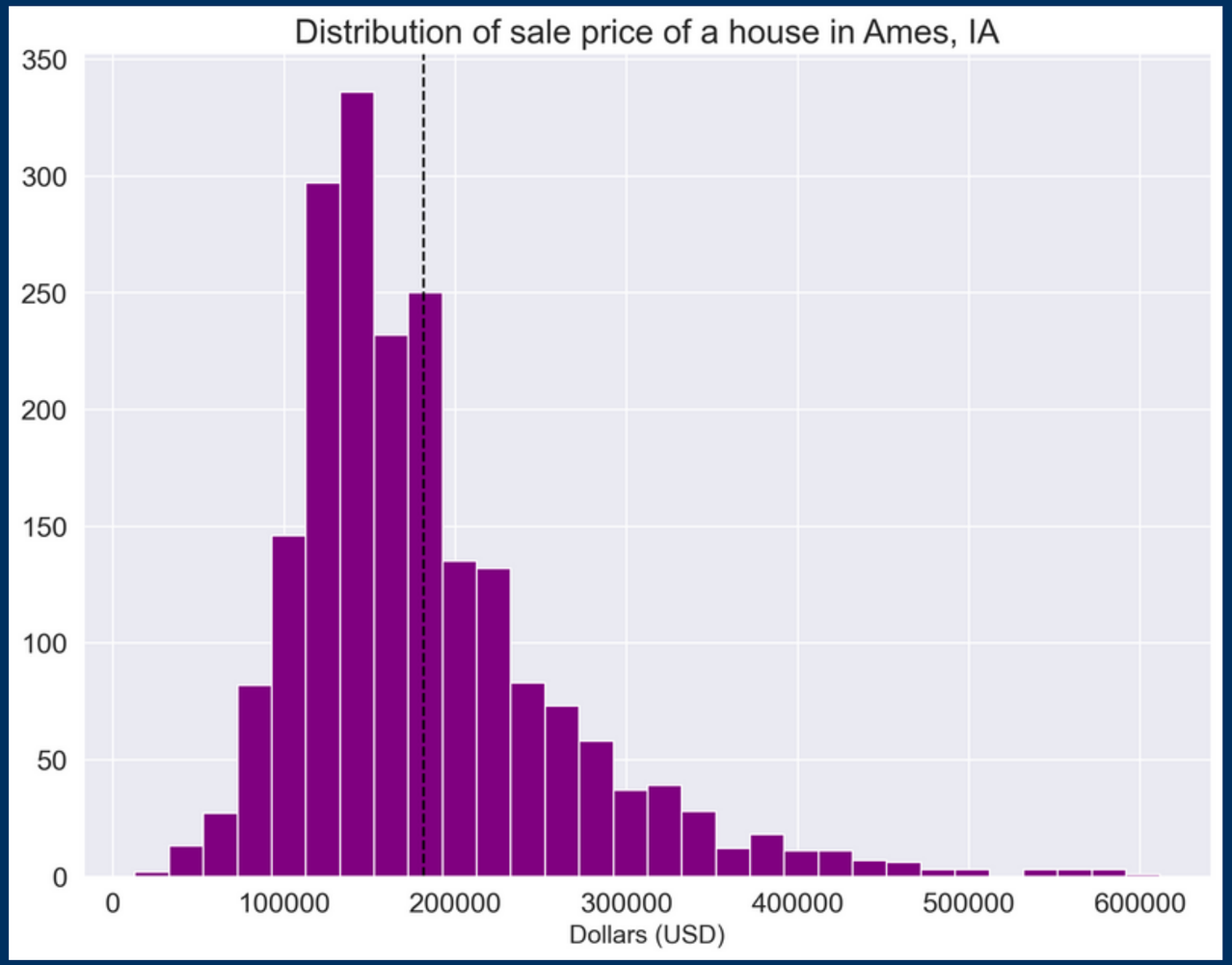
Sanity throughout this project

Hating Timmy

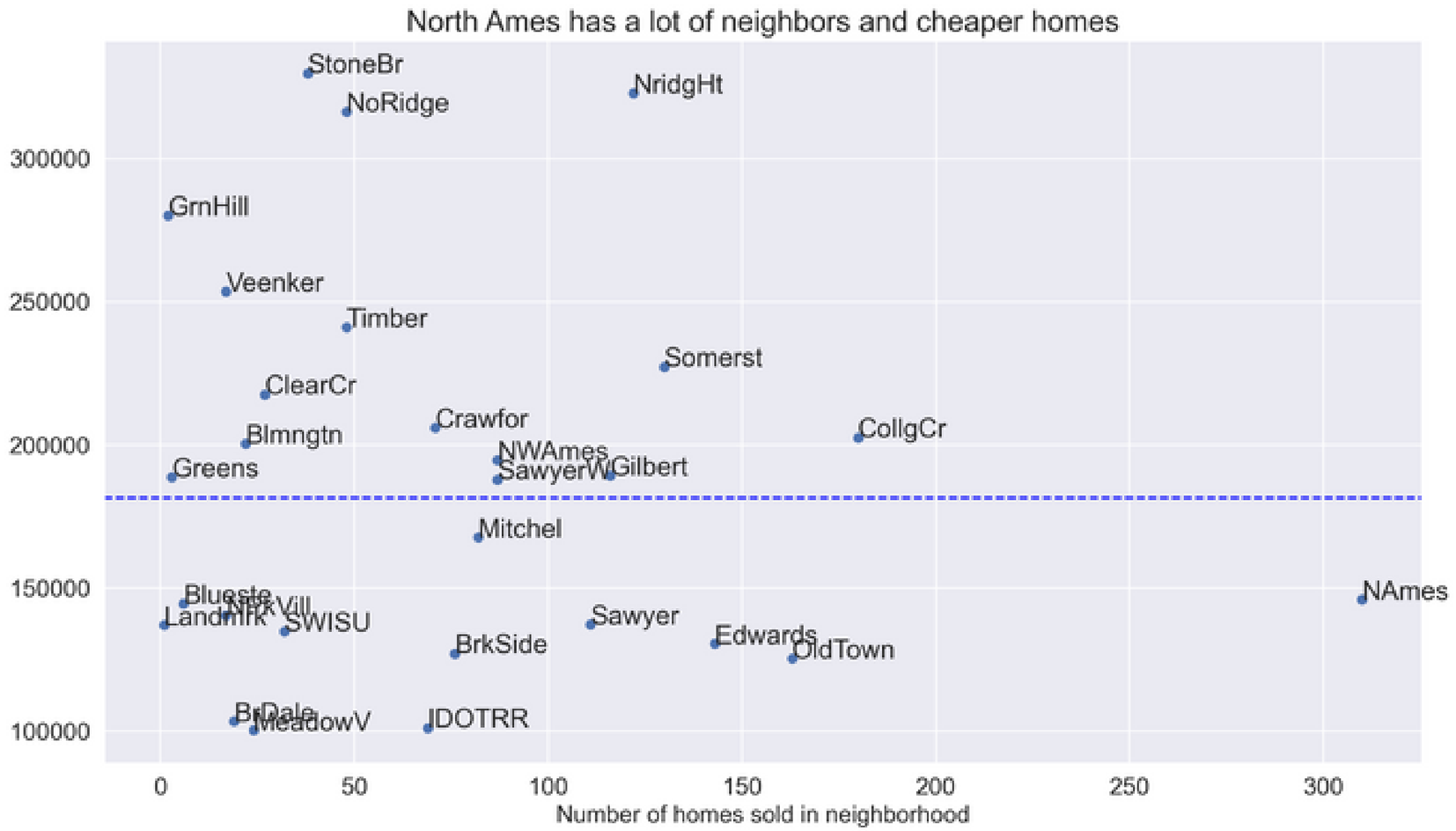
Rows of Houses Missing Sale Price

878

TIMMY SHOULD HAVE AN EASY TIME FINDING A HOME



LET'S TAKE A PEEK AT FUTURE NEIGHBORS

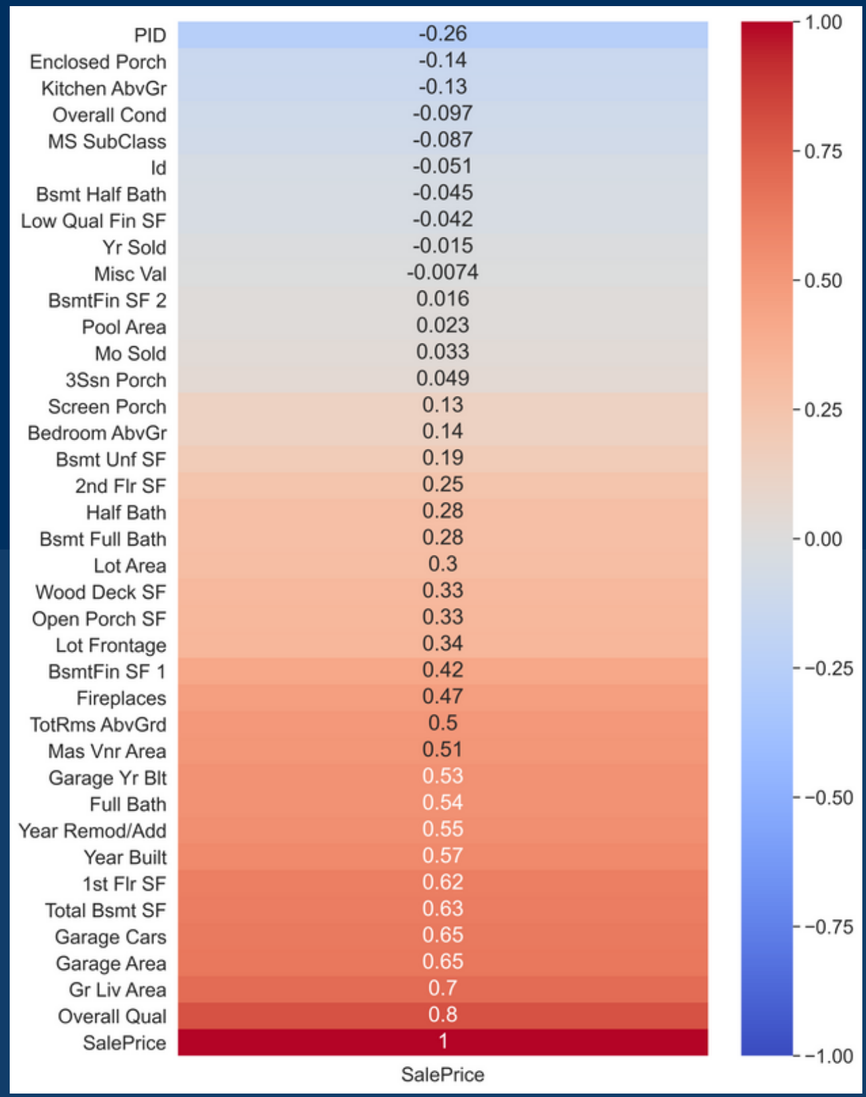


First Features

Corr > .5

First assumption

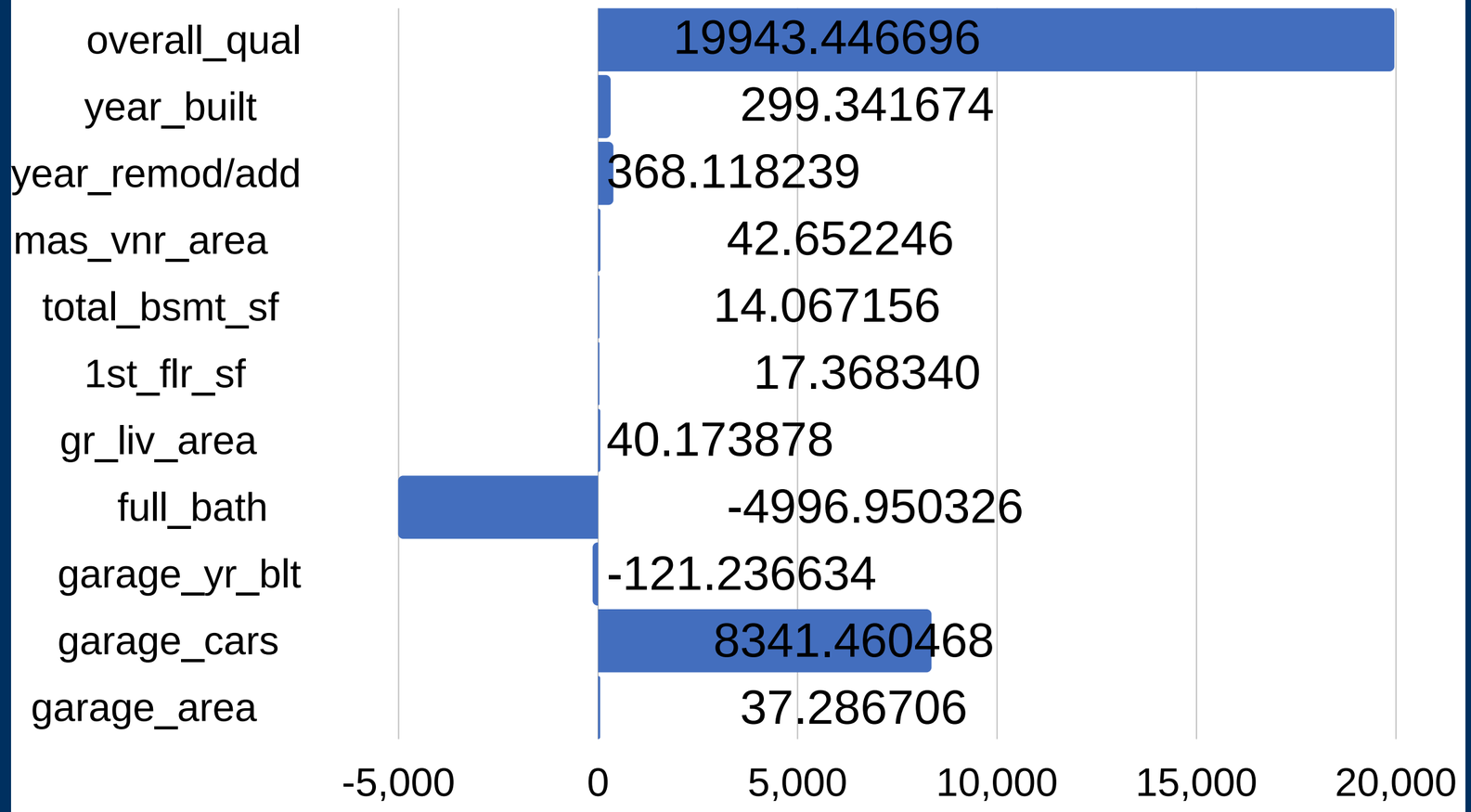
There were 11 features that had a high correlation with Sale Price in the "train" data.

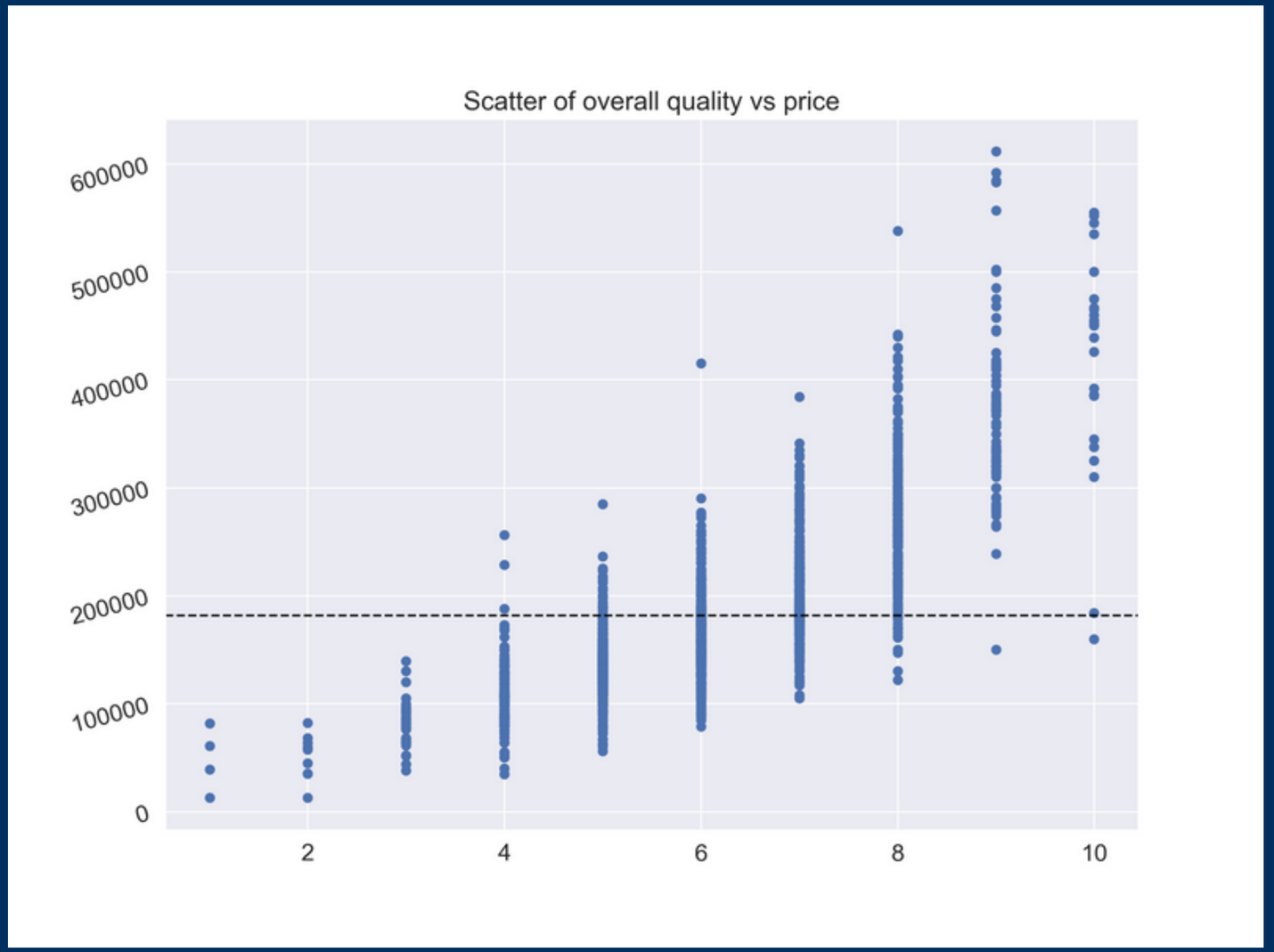


What were the results of the first linear model?

Obviously, hard to interpret here.

One statement you can say is that holding all else constant, as overall quality increases, the price of the home increases by about $20k.




Scatter of overall quality vs price
600000
500000
400000
300000
200000
100000
0
2
4
6
8
10

There's more than one model

As time went on, I started to add and subtract features...

Disclaimer: Most models were Linear Regressions because they produced the best results.

Modeling Benchmarks

R2 Scores (Linear Regression)

Feature Set 1

Training score:
0.7925242267944131
Test score:
0.8224348663093287

Feature Set 2

Training score:
0.8287125336393979
Test score:
0.8345724494443683

Feature Set 3

Training score:
0.8674697211050928
Test score:
0.817215695487534

Feature Set 4

Training score:
0.8898429449547527
Test score:
0.8324550014648933

Modeling Benchmarks

RMSE (Linear Regression)

Feature Set 1	*Feature Set 2*	*Feature Set 3*	*Feature Set 4*
31941.92581264128	30830.897683095838	34045.718222792704	32595.58532312685

Null RMSE = 79239.33504161824

Modeling Benchmarks

R2 Scores (Linear Regression) With Logged Features

Feature Set 1

Training score:
0.8172897846675002
Test score:
0.831930845976321 5

Feature Set 2

Training score:
0.842127118493724 2
Test score:
0.825120654173575 4

Feature Set 3

Training score:
0.880319962666246 3
Test score:
0.822716611185609 2

Feature Set 4

Training score:
0.89842194218709
Test score:
0.838960420120263 8

Modeling Benchmarks

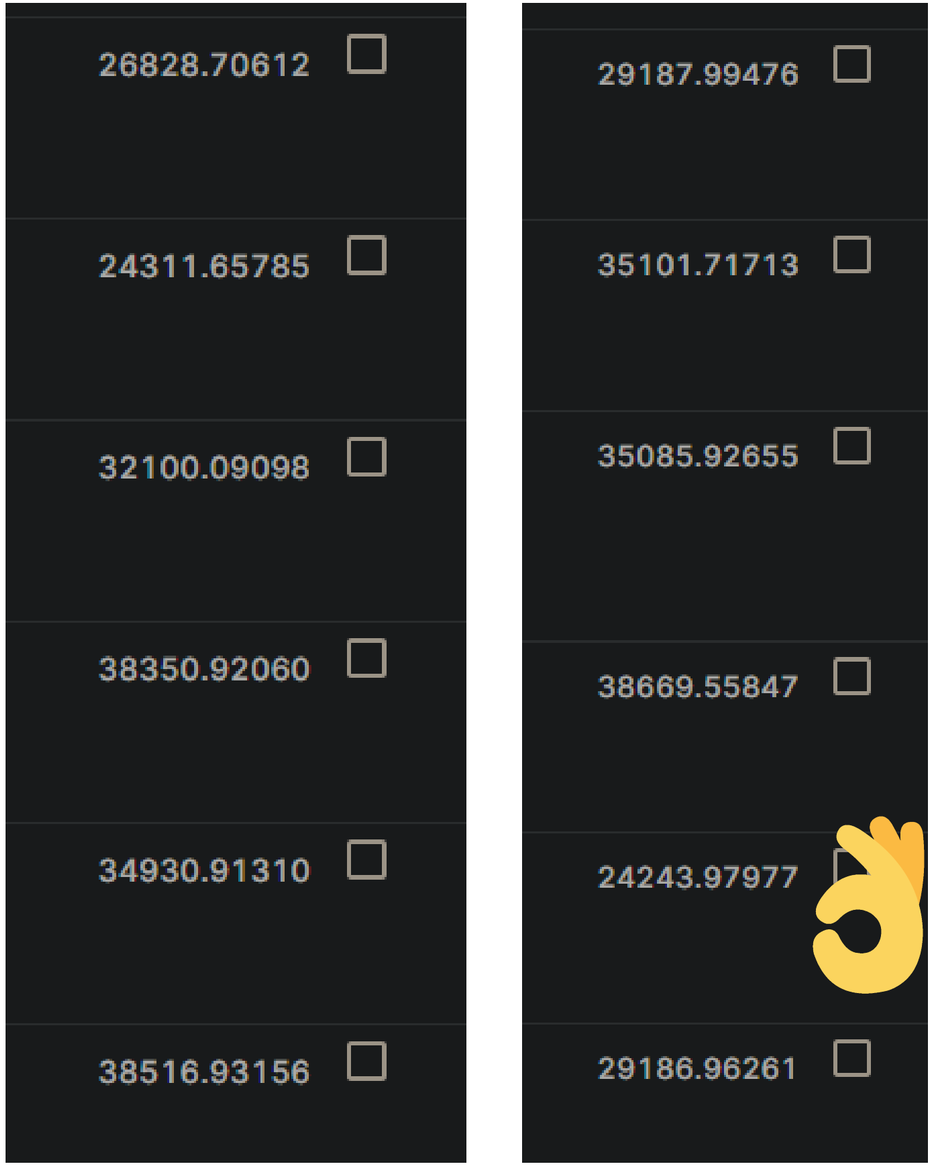
RMSE (Linear Regression) with Logged Features

Feature Set 1	*Feature Set 2*	*Feature Set 3*	*Feature Set 4*
27375.72404094178	28743.13855801938	33130.49181361535	41477.858180177944

Null RMSE = 79239.33504161824

Sometime down the road, my friend made this a competition. These were the RMSE scores I was able to achieve.

GAME ON



In the end...

I found the best model I could possibly produce.

ms_subclass	-168.447166
overall_qual	9576.490185
overall_cond	5586.831510
year_built	374.672628
year_remod/add	80.150149
total_bsmt_sf	14.624137
1st_flr_sf	-2.965972
gr_liv_area	49.146851
full_bath	4499.193567
half_bath	5146.743645
bedroom_abvgr	-3535.562464
totrms_abvgrd	-163.833591
fireplaces	7051.975308
garage_yr_blt	-116.274295
garage_cars	9475.924157
garage_area	16.384405

mo_sold	-32.446462
yr_sold	-418.633301
street_Grvl	-5426.056922
street_Pave	5426.056922
alley_Grvl	2227.058808
alley_Pave	2541.547685
bldg_type_1Fam	6393.956473
bldg_type_2fmCon	26177.358722
bldg_type_Duplex	-12135.687213
bldg_type_Twnhs	-12744.443327
bldg_type_TwnhsE	-7691.184655
heating_qc_Ex	1879.553094
heating_qc_Fa	-1128.844207
heating_qc_Gd	1782.052296
heating_qc_TA	-2532.761182

garage_qual_Fa	-14096.912195
garage_qual_Gd	22746.743574
garage_qual_Po	-27974.911158
garage_qual_TA	-11722.352005
garage_cond_Ex	-24300.094373
garage_cond_Fa	-2674.557463
garage_cond_Gd	-10443.001238
garage_cond_Po	2738.027774
garage_cond_TA	3632.193516
electrical_Fuse	-4601.341135
electrical_Mix	9505.879919
electrical_Sbrkr	-4904.538784
kitchen_qual_Ex	29693.132335
kitchen_qual_Fa	-10946.366909
kitchen_qual_Gd	-7241.294456
kitchen_qual_TA	-11505.470970

house_style_1.5Fin	-3271.593267
house_style_1.5Unf	-6469.359290
house_style_1Story	2794.655145
house_style_2.5Fin	-6108.923227
house_style_2.5Unf	2292.508487
house_style_2Story	-5386.486151
house_style_SFoyer	13041.312481
house_style_SLvl	3107.885823
bsmt_qual_Ex	19839.984542
bsmt_qual_Fa	-5544.037519
bsmt_qual_Gd	-4928.289036
bsmt_qual_Po	-3520.722840
bsmt_qual_TA	-6376.590587
bsmt_cond_Avg	-1149.773158
bsmt_cond_Gd	620.117718

neighborhood_Blmngtn	1745.555244
neighborhood_Blueste	-408.481653
neighborhood_BrDale	7538.967933
neighborhood_BrkSide	-12037.184229
neighborhood_ClearCr	15898.467102
neighborhood_CollgCr	-7927.000045
neighborhood_Crawfor	7162.072798
neighborhood_Edwards	-17417.889979
neighborhood_Gilbert	-13358.552062
neighborhood_Greens	11607.014111
neighborhood_IDOTRR	-17514.403693
neighborhood_MeadowV	5369.175003
neighborhood_Mitchel	-10450.67896
neighborhood_NAmes	-10521.000965
neighborhood_NPkVill	750.362581
neighborhood_NWAmes	-16395.041744
neighborhood_NoRidge	27921.833925
neighborhood_NridgHt	32350.807866
neighborhood_OldTown	-22229.124956
neighborhood_SWISU	-17361.065351
neighborhood_Sawyer	-11690.064915
neighborhood_SawyerW	-11883.128332
neighborhood_Somerst	2846.855855
neighborhood_StoneBr	53284.836931
neighborhood_Timber	286.044094
neighborhood_Veenker	2431.623444

So Where is Timmy moving to?

Location, Location

Avoid Stone Brook (53284.836931) and Northridge Heights (32350.807866).

You don't need...

- A basement
- A garage
- A chef's kitchen
- or a Contract with 15% Down payment on regular terms

Overall Quality & Condition

He should not skip out on the quality (9576.490185) nor the overall condition (5586.831510).

The amount of rooms

Room space isn't an issue (-163.833591) but he should be concerned about the types of bathrooms. (full=4499.193567, half=5146.743645)

Avoid Fireplaces

Just pursue Central Air conditioning because you don't need a fireplace, unless you wanna risk looking like parts of the West Coast (oof, too soon; 7051.975308)

These guys can help

Geraldine Lim

Licensed Realtor

Roger Smith

Finance Specialist

Liana Webber

Consultant

Thank you!

*And to Timmy, your house better look like a house from **MTV Cribs** after all the work I went through!*