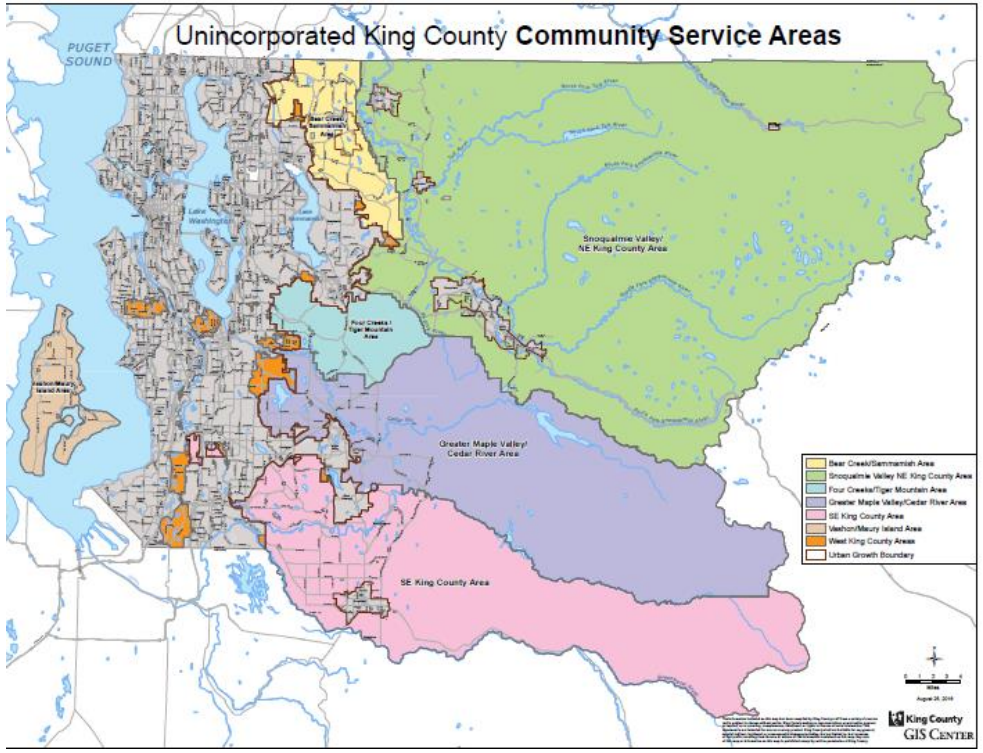
Phase 2 – Linear Regression Project

King County Housing Data



Phase 2 - Project

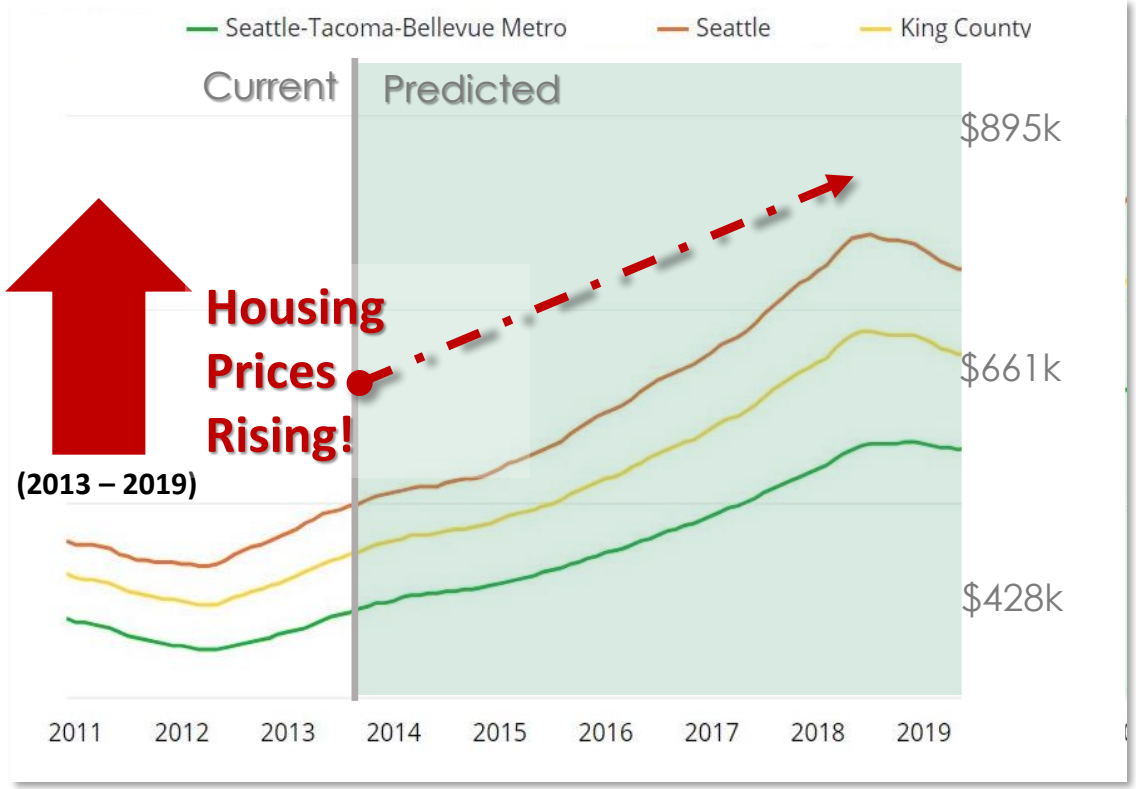
AGENDA

- Intro To Business & Opportunity
- Methodology
- 3 Recommendations
- Next Steps
- *Extra Credit (if time permits)*

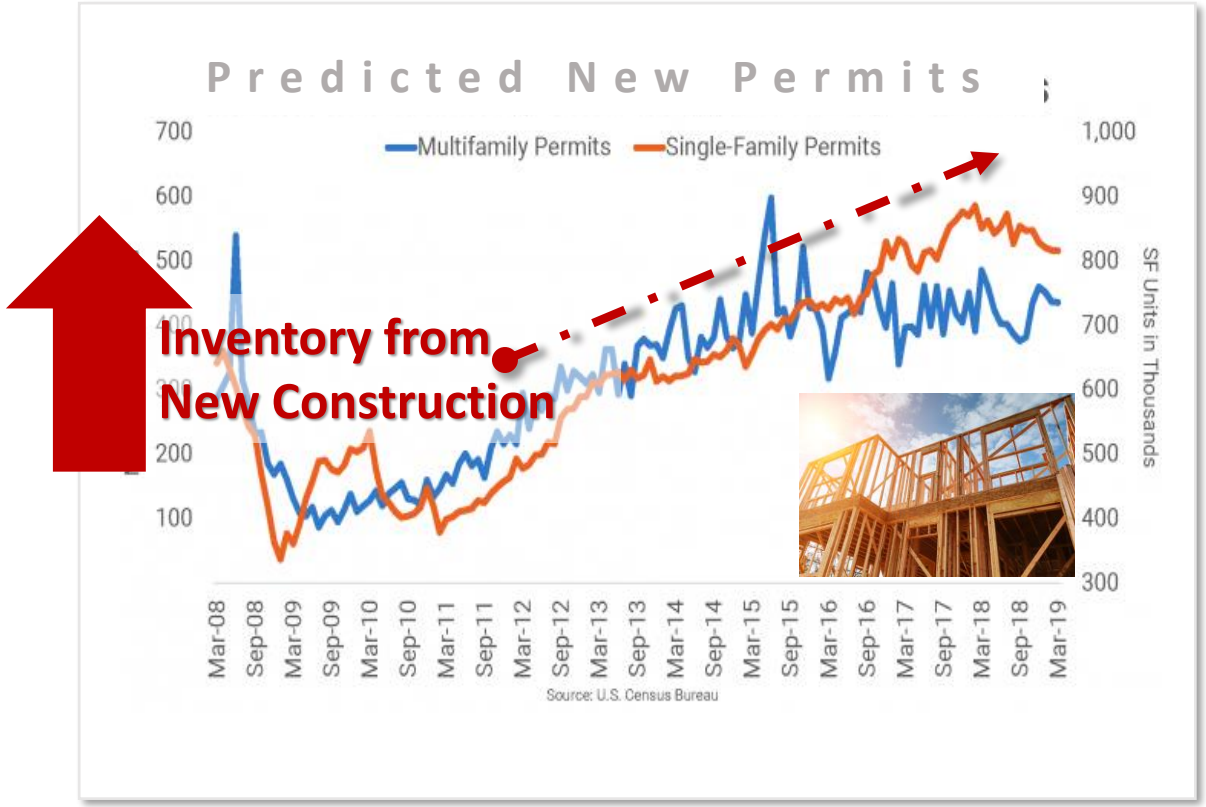
The Business

Opportunity:

Experts Predicting Sellers Market…



….But Inventory Up!



Goal: Increase Revenue From Seller Commissions **(Existing Homes)**

How?

1. Help Agents Identify "Most Valuable" Home Improvements Projects, for recommendations to sellers

2. Help Agents Set "Better" Prices... augment CMA process

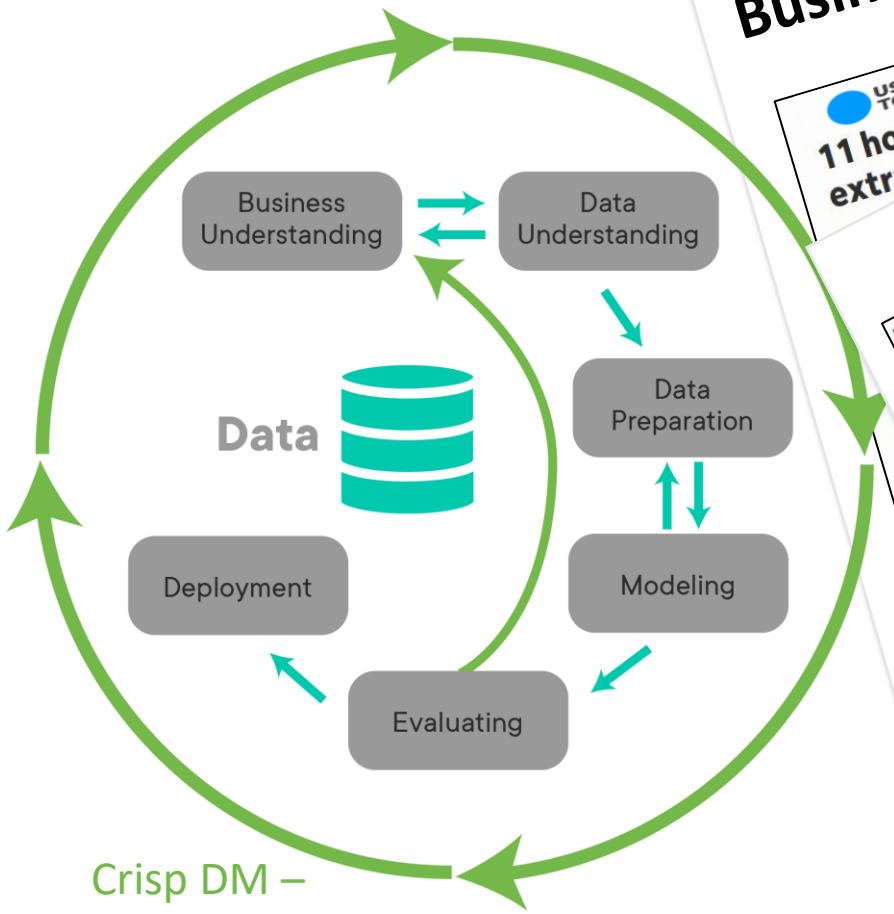
Business Value

"Smarter" Agents

I. Better Recommendations = Higher Prices = Commissions

II.Setting Better Prices = Getting True Market Value

Methodology



Crisp DM –

3 Recommendations

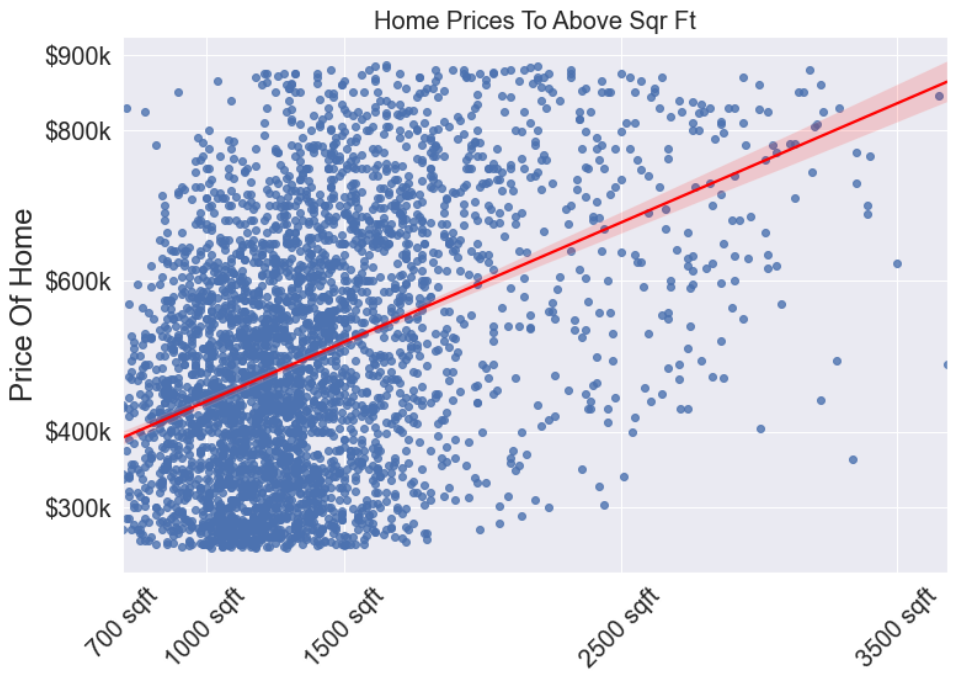
1 – Physical Factors - "controllable"

2 – Location Factor - "influenceable"

3 – Consider a new Pricing Factor

Recommendation 1: Augment Physical Features

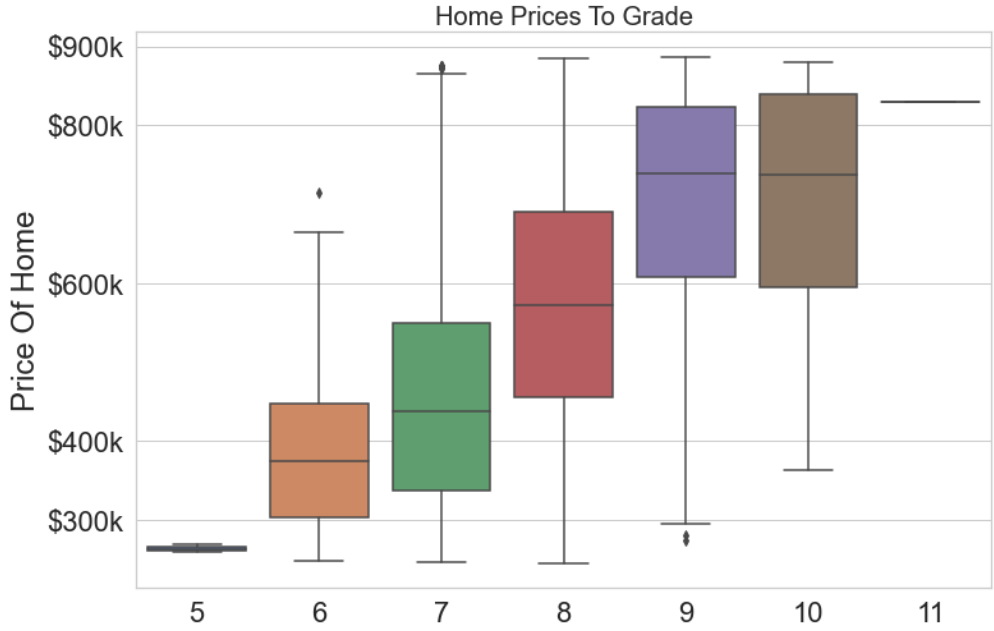
1. # Main-Level Finished Square Feet



+100 Above Sqft ➡ **+ ~ $4K**

2. Grade "Specialness" of Home

Construction and Design, 7 = Average older sub-divisions



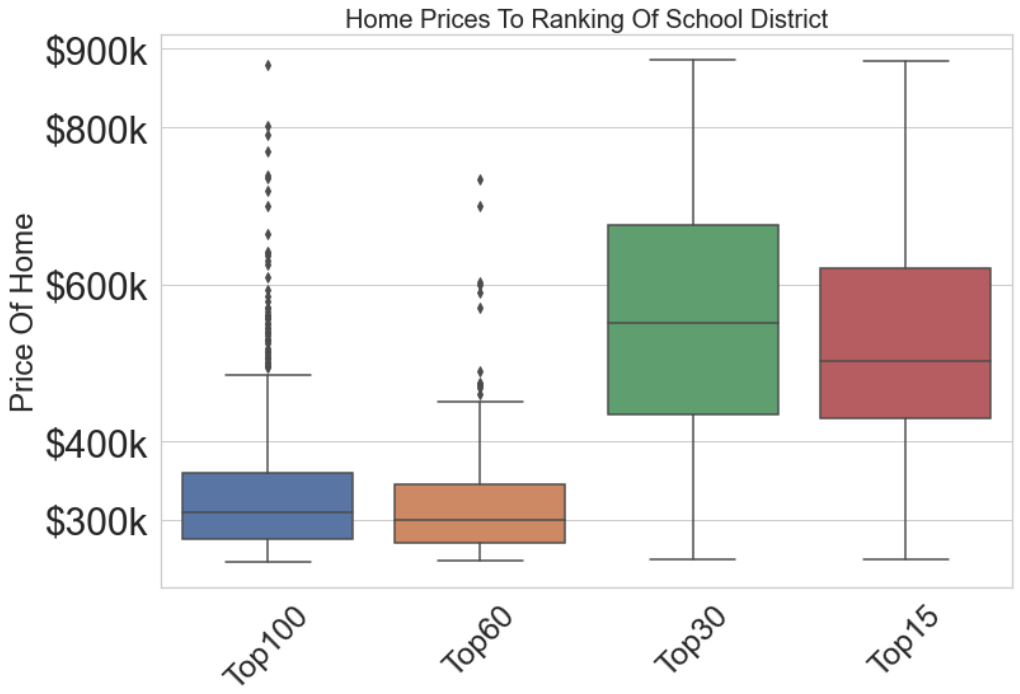
Grade 7 ➡ **+ ~$22K**

*+1 Grade**

*Grade 8+ = better materials in both the exterior and interior finish work, architectural design, solid woods, bathroom fixtures and more luxurious options.

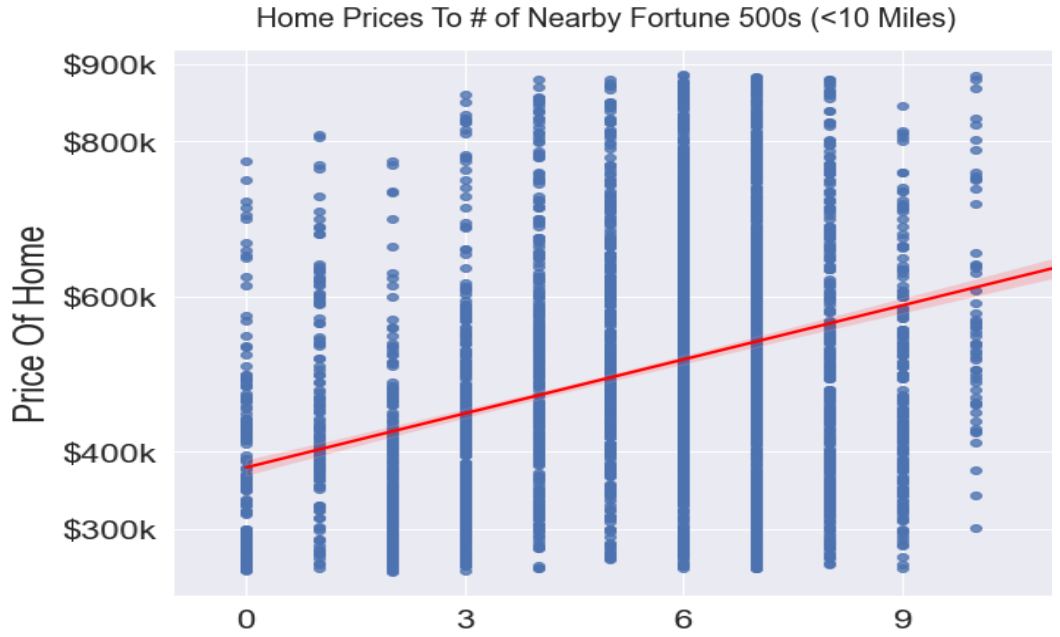
Recommendation 2: Work to Impact "Influenceable" Factors

1. School District



Top 60 → + ~ $88K
+30

2. # Nearby Fortune 500's*



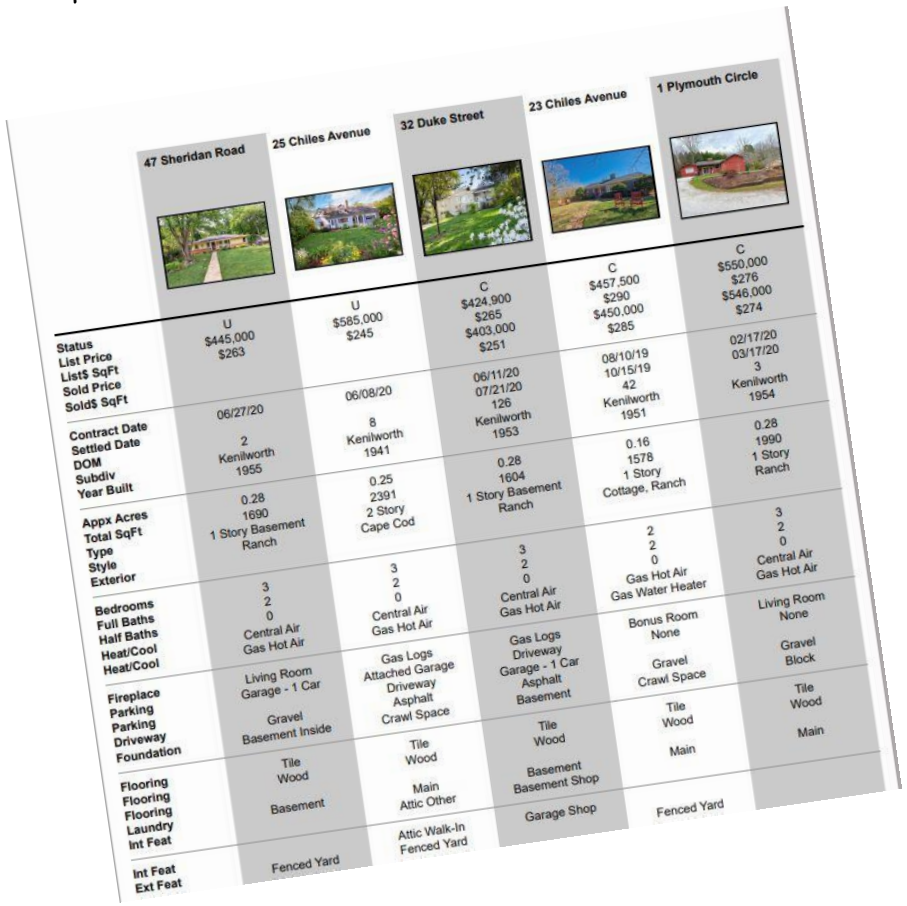
0 F500s → + ~ $53K
+1-3

*Nearby = < 10 Miles From Home

Recommendation 3: Setting "Better" Prices...Augmenting CMA

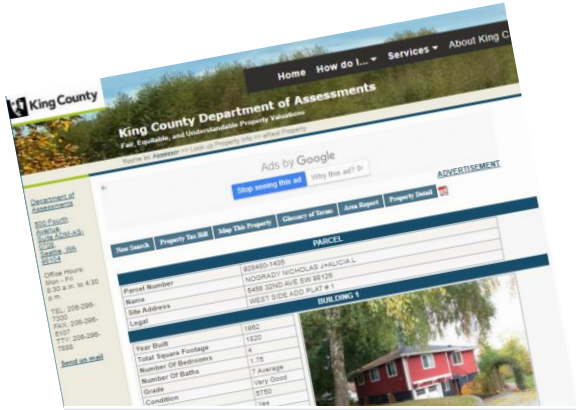
Comparative Market Analysis (CMA)

- Home Features, Condition, Location, Competition



	47 Sheridan Road	25 Chiles Avenue	32 Duke Street	23 Chiles Avenue	1 Plymouth Circle
Status List Price List$ SqFt Sold Price Sold$ SqFt	U $445,000 $263	U $585,000 $245	C $424,900 $265 $403,000 $251	C $457,500 $290 $450,000 $285	C $550,000 $276 $546,000 $274
Contract Date Settled Date DOM Subdiv Year Built	06/27/20 2 Kenilworth 1955	06/08/20 8 Kenilworth 1941	06/11/20 07/21/20 126 Kenilworth 1953	08/10/19 10/15/19 42 Kenilworth 1951	02/17/20 03/17/20 3 Kenilworth 1954
Appx Acres Total SqFt Type Style Exterior	0.28 1690 1 Story Basement Ranch	0.25 2391 2 Story Cape Cod	0.28 1604 1 Story Basement Ranch	0.16 1578 1 Story Cottage, Ranch	0.28 1990 1 Story Ranch
Bedrooms Full Baths Half Baths Heat/Cool Heat/Cool	3 2 0 Central Air Gas Hot Air	3 2 0 Central Air Gas Hot Air	3 2 0 Central Air Gas Hot Air	2 2 0 Gas Hot Air Gas Water Heater	3 2 0 Central Air Gas Hot Air
Fireplace Parking Parking Driveway Foundation	Living Room Garage - 1 Car Gravel Basement Inside	Gas Logs Attached Garage Driveway Asphalt Crawl Space	Gas Logs Driveway Garage - 1 Car Asphalt Basement	Bonus Room None Gravel Crawl Space	Living Room None Gravel Block
Flooring Flooring Flooring Laundry Int Feat	Tile Wood Basement	Tile Wood Main Attic Other	Tile Wood Basement Basement Shop	Tile Wood Main	Tile Wood Main
Int Feat Ext Feat	Fenced Yard	Attic Walk-In Fenced Yard	Garage Shop	Fenced Yard	

Adding Assessor Appraisals



TAX ROLL HISTORY

Valued Year	Tax Year	Appraised Land Value ($)	Appraised Imps Value ($)	Appraised Total ($)	Appraised Imps Increase ($)	Taxable Land Value ($)	Taxable Imps Value ($)	Taxable Total ($)
2020	2021	232,000	450,000	682,000	0	232,000	450,000	682,000
2019	2020	231,000	436,000	667,000	0	231,000	436,000	667,000
2018	2019	236,000	416,000	652,000	0	236,000	416,000	652,000
2017	2018	197,000	338,000	535,000	0	197,000	338,000	535,000
2016	2017	176,000	288,000	464,000	0	176,000	288,000	464,000
2015	2016	233,000	277,000	510,000	0	233,000	277,000	510,000
2014	2015	129,000	192,000	321,000	0	129,000	192,000	321,000
2013	2014	129,000	126,000	255,000	0	129,000	126,000	255,000
2012	2013	129,000	123,000	252,000	0	129,000	123,000	252,000
2011	2012	126,000	154,000	280,000	0	126,000	154,000	280,000
2010	2011	137,000	166,000	303,000	0	137,000	166,000	303,000
2009	2010	135,000	164,000	299,000	0	135,000	164,000	299,000
2008	2009	145,000	226,000	371,000	0	145,000	226,000	371,000
2007	2008	122,000	169,000	291,000	0	122,000	169,000	291,000
2006	2007	101,000	160,000	261,000	0	101,000	160,000	261,000
2005	2006	88,000	148,000	236,000	0	88,000	148,000	236,000
2004	2005	76,000	142,000	218,000	0	76,000	142,000	218,000

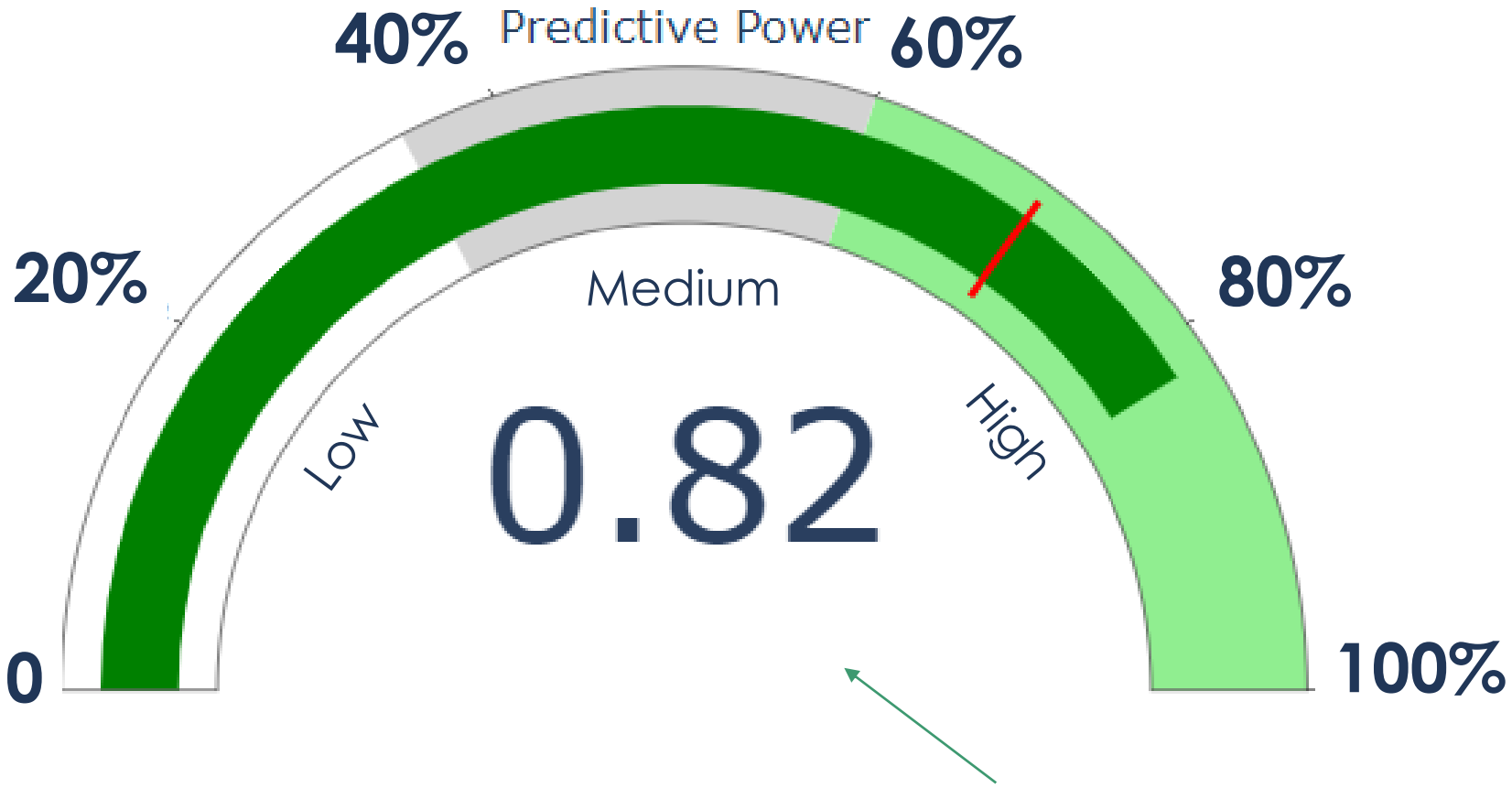
Assessor Appraisal Feature

200% >

Next Feature

Predictive Power

= Models ability to explain variability in the data being predicted



Increase after adding
Previous Yr Assessor Appraisals

Summary of Recommendations

- Increase Seller Commissions By Coaching Sellers To:

1. "Controllable"

- *Above Sqr ft*
- *Grade*

2. "Influenceable"

- *School District*
- *F 500s*

- Look To Non-Traditional Sources When Setting Prices

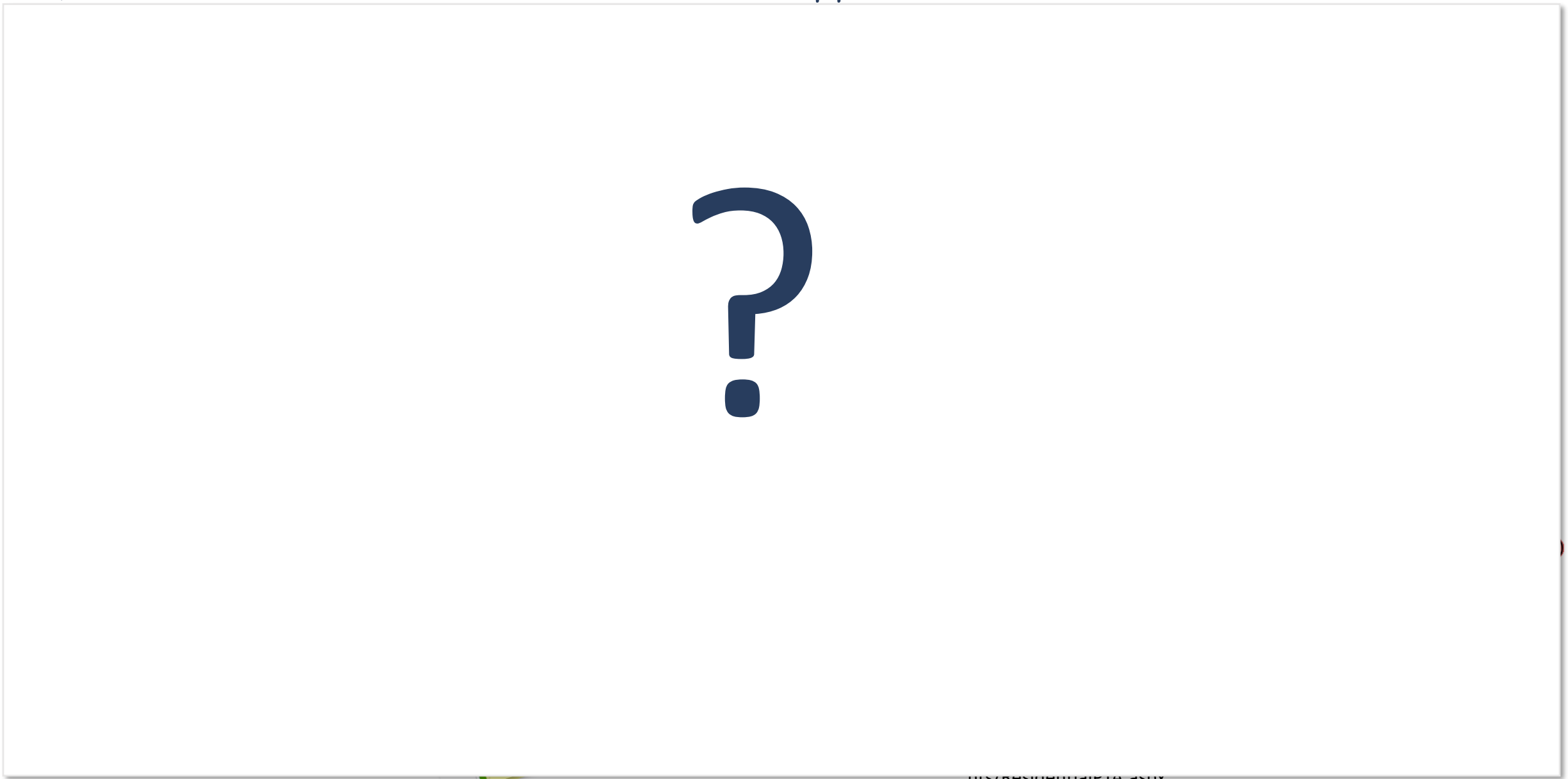
3. Tax Assessor Records

Future Work/ Next Steps

- Augment current model
  - Study homes with greatest variance between actual & predicted
    - Add additional interaction
      - # of Attractions, Inflation Factors, Other
- Develop additional models for more niche homes
  - (<$200k, >$900k, Waterfront, Other)
- **Deployment!**

EXTRA CREDIT –

1. How Do Assessor Establish Appraisals?
2 How Well Does our Model Predict Assessor Appraisals?



Thanks For Your Time and Consideration!

Appendix

Model Results:

"Does This Mean?"

OLS Regression Results

Dep. Variable:	log_price	R-squared:	0.829
Model:	OLS	Adj. R-squared:	0.825
Method:	Least Squares	F-statistic:	254.4
Date:	Tue, 24 Nov 2020	Prob (F-statistic):	0.00
Time:	05:49:22	Log-Likelihood:	663.58
No. Observations:	1288	AIC:	-1277.
Df Residuals:	1263	BIC:	-1148.
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.3818	0.021	576.995	0.000	12.340	12.424
AssessorAppraisals_x	1.2368	0.041	30.371	0.000	1.157	1.317
Sch_d_Top15	0.2257	0.012	18.274	0.000	0.202	0.250
Sch_d_Top30	0.2060	0.014	14.620	0.000	0.178	0.234
Under10	0.1524	0.033	4.577	0.000	0.087	0.218
Over20	-0.3516	0.049	-7.149	0.000	-0.448	-0.255
view_1	0.0598	0.026	2.301	0.022	0.009	0.111
view_2	0.0538	0.018	3.009	0.003	0.019	0.089
view_4	0.1124	0.056	2.000	0.046	0.002	0.223
condition_4	0.0350	0.009	3.778	0.000	0.017	0.053
condition_5	0.1143	0.014	8.205	0.000	0.087	0.142
grade_6	-0.0682	0.020	-3.442	0.001	-0.107	-0.029
grade_8	0.0483	0.010	4.942	0.000	0.029	0.067
sls_mnth_3	0.0449	0.019	2.326	0.020	0.007	0.083
sls_mnth_4	0.0665	0.018	3.647	0.000	0.031	0.102
sls_mnth_5	0.0690	0.018	3.843	0.000	0.034	0.104
sls_mnth_6	0.0859	0.018	4.679	0.000	0.050	0.122
sls_mnth_7	0.0672	0.019	3.529	0.000	0.030	0.105
sls_mnth_8	0.0724	0.019	3.815	0.000	0.035	0.110
sls_mnth_9	0.0508	0.020	2.573	0.010	0.012	0.090
sls_mnth_10	0.0649	0.019	3.413	0.001	0.028	0.102
sls_mnth_11	0.0658	0.020	3.261	0.001	0.026	0.105
sls_mnth_12	0.0620	0.020	3.053	0.002	0.022	0.102
sqft_basement	0.1204	0.029	4.149	0.000	0.063	0.177
sqft_above	0.3468	0.038	9.209	0.000	0.273	0.421

Omnibus:	15.828	Durbin-Watson:	2.014
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21.376
Skew:	-0.142	Prob(JB):	2.28e-05
Kurtosis:	3.563	Cond. No.	20.1

Adj. R-squared: 0.828

Measures - goodness-of-fit - 82% of our predicted price can be explained by our model features, 18% may be cause by random error

Model Progression

Model1

	Measure	Value
0	train_R2	0.401
1	test_R2	0.372
2	train_mse	0.0689523
3	test_mse	0.0712373
4	intercept	12.7574
5	waterfront_1	0.191622
6	view_1	0.125222
7	view_2	0.0629582
8	view_3	0.0328921
9	view_4	0.173902
10	condition_4	0.0960266
11	condition_5	0.154242
12	grade_6	-0.165765
13	grade_8	0.200531
14	grade_9	0.317696
15	grade_10	0.272693
16	grade_11	-5.55112e-17
17	ZipFirst3_981	0.141611
18	sls_mnth_2	0.0156561
19	sls_mnth_3	0.0961082
20	sls_mnth_4	0.0803946
21	sls_mnth_5	0.0535677
22	sls_mnth_6	0.0242366
23	sls_mnth_7	0.0422843
24	sls_mnth_8	0.0632581
25	sls_mnth_9	0.0374608
26	sls_mnth_10	0.0439298
27	sls_mnth_11	0.0139689
28	sls_mnth_12	0.027499
29	lf_renovated_1	0.121325
30	sqft_lot	-0.248111
31	sqft_basement	0.123596
32	sqft_above	0.523408

Model2

	Measure	Value
0	train_R2	0.684
1	test_R2	0.678
2	train_mse	0.0363533
3	test_mse	0.0365513
4	intercept	12.5296
5	Sch_d_Top15	0.359512
6	Sch_d_Top30	0.357542
7	Sch_d_Top60	-0.070847
8	Under10	0.160892
9	Over20	-0.703032
10	sqft_lot	-0.102805
11	sqft_basement	0.105728
12	sqft_above	0.584546
13	view_1	0.086734
14	view_2	0.0941699
15	view_3	0.0934741
16	view_4	0.264415
17	condition_4	0.087975
18	condition_5	0.140688
19	grade_6	-0.157626
20	grade_8	0.154334
21	grade_9	0.261723
22	grade_10	0.252175
23	sls_mnth_2	0.0128365
24	sls_mnth_3	0.0678127
25	sls_mnth_4	0.063692
26	sls_mnth_5	0.030454
27	sls_mnth_6	-0.00721413
28	sls_mnth_7	0.00466603
29	sls_mnth_8	0.00364861
30	sls_mnth_9	0.00401947
31	sls_mnth_10	-0.00177026
32	sls_mnth_11	-0.010404
33	sls_mnth_12	-0.00733327
34	lf_renovated_1	0.106718

Model3

	Measure	Value
0	train_R2	0.830
1	test_R2	0.800
2	train_mse	0.0207198
3	test_mse	0.023869
4	intercept	12.4181
5	AssesorAppraisals_x	1.19366
6	Sch_d_Top15	0.224482
7	Sch_d_Top30	0.197072
8	Sch_d_Top60	-0.0203361
9	Under10	0.132568
10	Over20	-0.378409
11	view_1	0.0593785
12	view_2	0.0561976
13	view_3	0.0182722
14	view_4	0.106432
15	condition_4	0.0389325
16	condition_5	0.116402
17	grade_6	-0.0749406
18	grade_8	0.0605497
19	grade_9	0.0363376
20	grade_10	0.0929671
21	sls_mnth_2	0.0119643
22	sls_mnth_3	0.0529574
23	sls_mnth_4	0.0754875
24	sls_mnth_5	0.0737535
25	sls_mnth_6	0.0897722
26	sls_mnth_7	0.0730218
27	sls_mnth_8	0.0778737
28	sls_mnth_9	0.0564489
29	sls_mnth_10	0.0698463
30	sls_mnth_11	0.0694779
31	sls_mnth_12	0.0666932
32	lf_renovated_1	0.0213601
33	sqft_lot	-0.0401348
34	sqft_basement	0.120024
35	sqft_above	0.310434

	Durbin-Watson
1	1.970781
2	1.986095
3	2.025427

Model 4 Optimized Using #Backward Elimination

OLS Regression Results

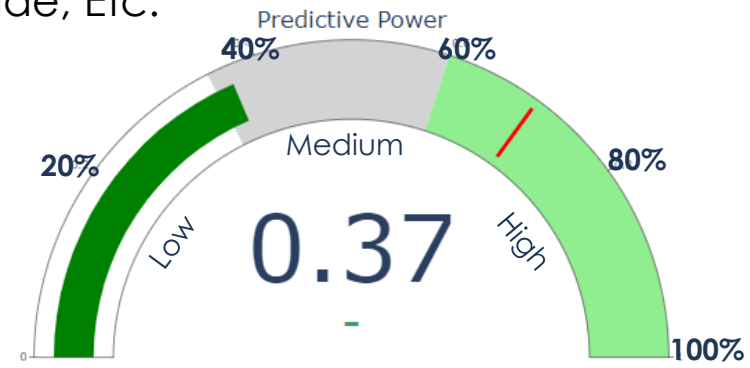
Dep. Variable:	log_price	R-squared:	0.829
Model:	OLS	Adj. R-squared:	0.825
Method:	Least Squares	F-statistic:	254.4
Date:	Tue, 24 Nov 2020	Prob (F-statistic):	0.00
Time:	05:49:22	Log-Likelihood:	663.58
No. Observations:	1288	AIC:	-1277.
Df Residuals:	1263	BIC:	-1148.
Df Model:	24		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.3818	0.021	576.995	0.000	12.340	12.424
AssesorAppraisals_x	1.2368	0.041	30.371	0.000	1.157	1.317
Sch_d_Top15	0.2257	0.012	18.274	0.000	0.202	0.250
Sch_d_Top30	0.2060	0.014	14.620	0.000	0.178	0.234
Under10	0.1524	0.033	4.577	0.000	0.087	0.218
Over20	-0.3516	0.049	-7.149	0.000	-0.448	-0.255
view_1	0.0598	0.026	2.301	0.022	0.009	0.111
view_2	0.0538	0.018	3.009	0.003	0.019	0.089
view_4	0.1124	0.056	2.000	0.046	0.002	0.223
condition_4	0.0350	0.009	3.778	0.000	0.017	0.053
condition_5	0.1143	0.014	8.205	0.000	0.087	0.142
grade_6	-0.0682	0.020	-3.442	0.001	-0.107	-0.029
grade_8	0.0483	0.010	4.942	0.000	0.029	0.067
sls_mnth_3	0.0449	0.019	2.326	0.020	0.007	0.083
sls_mnth_4	0.0665	0.018	3.647	0.000	0.031	0.102
sls_mnth_5	0.0690	0.018	3.843	0.000	0.034	0.104
sls_mnth_6	0.0859	0.018	4.679	0.000	0.050	0.122
sls_mnth_7	0.0672	0.019	3.529	0.000	0.030	0.105
sls_mnth_8	0.0724	0.019	3.815	0.000	0.035	0.110
sls_mnth_9	0.0508	0.020	2.573	0.010	0.012	0.090
sls_mnth_10	0.0649	0.019	3.413	0.001	0.028	0.102
sls_mnth_11	0.0658	0.020	3.261	0.001	0.026	0.105
sls_mnth_12	0.0620	0.020	3.053	0.002	0.022	0.102
sqft_basement	0.1204	0.029	4.149	0.000	0.063	0.177
sqft_above	0.3468	0.038	9.209	0.000	0.273	0.421

Omnibus:	15.828	Durbin-Watson:	2.014
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21.376
Skew:	-0.142	Prob(JB):	2.28e-05
Kurtosis:	3.563	Cond. No.	20.1

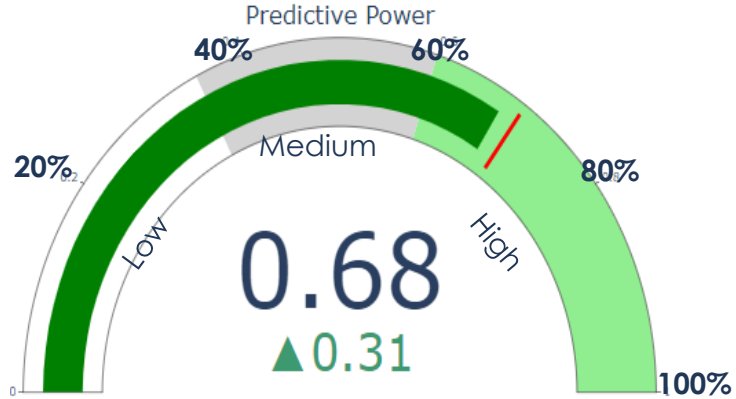
Assessor Appraisals Have Strong Predictive Strength

M1. Using physical "features"…Sqr Footage, Grade, Etc.

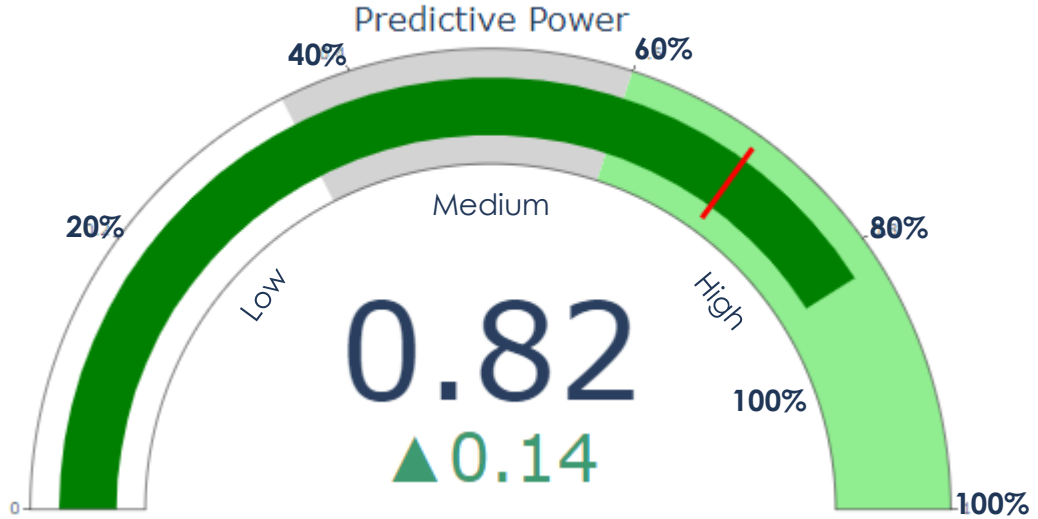


M2. Adding influenceable "features"…

School Districts, Fortune 500's



M3. Adding Previous Yr Assessor Appraisals



*

Influence = 1.3

200% Higher Than Next Factor*

**vs. Next Highest Sqr Ft Above = .42*

Assessor Model

	Measure	Value
0	train_R2	0.737
1	test_R2	0.733
2	train_mse	0.0374024
3	test_mse	0.0386394
4	intercept	12.3369
5	Sch_d_Top15	0.313072
6	Sch_d_Top30	0.439637
7	Sch_d_Top60	-0.0320625
8	Under10	-0.153349
9	Over20	-0.845506
10	sqft_lot	-0.0787835
11	sqft_basement	0.0470247
12	sqft_above	0.841512
13	view_1	0.119014
14	view_2	0.1136
15	view_3	0.168162
16	view_4	0.167149
17	condition_4	0.102763
18	condition_5	0.0617159
19	grade_6	-0.21732
20	grade_8	0.188465
21	grade_9	0.306399
22	grade_10	0.371521
23	sls_mnth_2	-0.00452489
24	sls_mnth_3	0.0324349
25	sls_mnth_4	0.0378942
26	sls_mnth_5	-0.0442387
27	sls_mnth_6	-0.134375
28	sls_mnth_7	-0.118092
29	sls_mnth_8	-0.126045
30	sls_mnth_9	-0.157024
31	sls_mnth_10	-0.151455
32	sls_mnth_11	-0.141337
33	sls_mnth_12	-0.162813
34	lf_renovated_1	0.0527851

Average Home - Exact Numbers Used For Calculations

Main-Level Finished Square Feet

1382 A_sft
$362,807

+100 Sqft →

1482 A_sft
$366,606

~ $4K

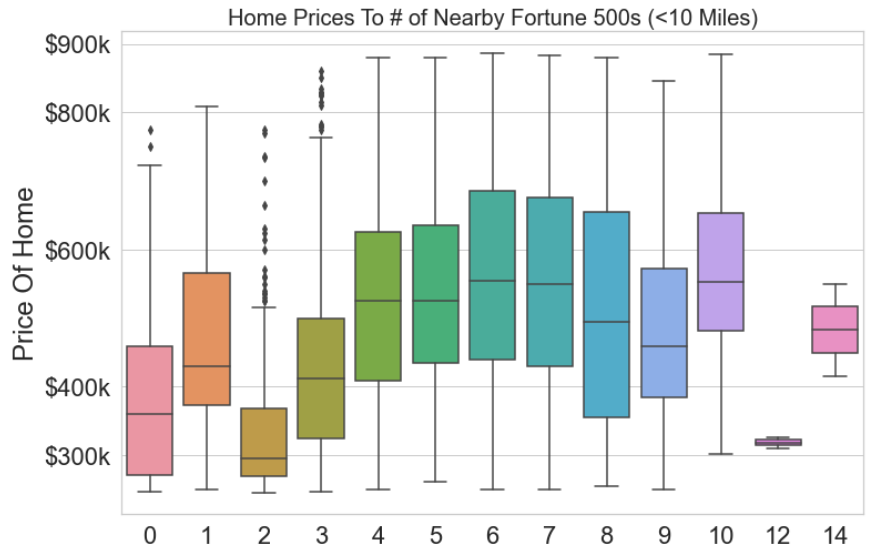
Grade "Specialness" of Home

Grade 7
$362,807

+1 Grade →

Grade 8
$385454

~$22K



Definition of Average Home

	a_bedrooms	a_bathrooms	a_sqft_living	a_sqft_lot	a_sqft_basement	a_sqft_living15	a_sqft_lot15	a_floors	a_sqft_above	a_AgeOfHome	a_ttl_rooms
0	3.555344	2.062061	2099.061788	6932.400218	716.332879	1851.636314	7027.358779	1.16753	1382.748909	60.865867	7.689477

terfront_1	a_view_1	a_view_2	a_view_3	a_view_4	a_condition_4	a_condition_5	a_grade_6	a_grade_8	a_grade_9	a_grade_10	a_grade_11	a_ZipFirst3_581
0	0	0	0	0	0	0	0	0	0	0	0	0

e_mnth_2	...	a_sle_mnth_6	a_sle_mnth_7	a_sle_mnth_8	a_sle_mnth_9	a_sle_mnth_10	a_sle_mnth_11	a_sle_mnth_12	a_lf_renovated_1	a_sc_bedrooms	a_sc_
0	...	0	0	0	0	0	0	0	0	0.319418	

a_sc_bathrooms	a_sc_sqft_living	a_sc_sqft_lot	a_sc_sqft_basement	a_sc_sqft_living15	a_sc_sqft_lot15	a_sc_floors	a_sc_sqft_above	a_sc_AgeOfHome	a_sc
0.270515	0.342204	0.488733	0.330107	0.346178	0.213695	0.083765	0.22911	0.48579	

a_sc_sqft_above	a_sc_AgeOfHome	a_sc_ttl_rooms	a_Sch_d_Top15	a_Sch_d_Top30	a_Sch_d_Top60	a_sc_Under10	a_sc_Over20	a_sc_AssessorAppraisals_x
0.22911	0.48579	0.169354	0	0	1	0.37204	0.068402	0.180597

BUILDING GRADE

Represents the construction quality of improvements. Grades run from grade 1 to 13. Generally defined as:

1-3 Falls short of minimum building standards. Normally cabin or inferior structure.

4 Generally older, low quality construction. Does not meet code.

5 Low construction costs and workmanship. Small, simple design.

6 Lowest grade currently meeting building code. Low quality materials and simple designs.

7 Average grade of construction and design. Commonly seen in plats and older sub-divisions.

8 Just above average in construction and design. Usually better materials in both the exterior and interior finish work.

9 Better architectural design with extra interior and exterior design and quality.

10 Homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage.

11 Custom design and higher quality finish work with added amenities of solid woods, bathroom fixtures and more luxurious options.

12 Custom design and excellent builders. All materials are of the highest quality and all conveniences are present.

13 Generally custom designed and built. Mansion level. Large amount of highest quality cabinet work, wood trim, marble, entry ways etc.

https://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r#g

BUILDING CONDITION

Relative to age and grade. Coded 1-5.

1 = Poor- Worn out. Repair and overhaul needed on painted surfaces, roofing, plumbing, heating and numerous functional inadequacies. Excessive deferred maintenance and abuse, limited value-in-use, approaching abandonment or major reconstruction; reuse or change in occupancy is imminent. Effective age is near the end of the scale regardless of the actual chronological age.

2 = Fair- Badly worn. Much repair needed. Many items need refinishing or overhauling, deferred maintenance obvious, inadequate building utility and systems all shortening the life expectancy and increasing the effective age.

3 = Average- Some evidence of deferred maintenance and normal obsolescence with age in that a few minor repairs are needed, along with some refinishing. All major components still functional and contributing toward an extended life expectancy. Effective age and utility is standard for like properties of its class and usage.

4 = Good- No obvious maintenance required but neither is everything new. Appearance and utility are above the standard and the overall effective age will be lower than the typical property.

5= Very Good- All items well maintained, many having been overhauled and repaired as they have shown signs of wear, increasing the life expectancy and lowering the effective age with little deterioration or obsolescence evident with a high degree of utility.