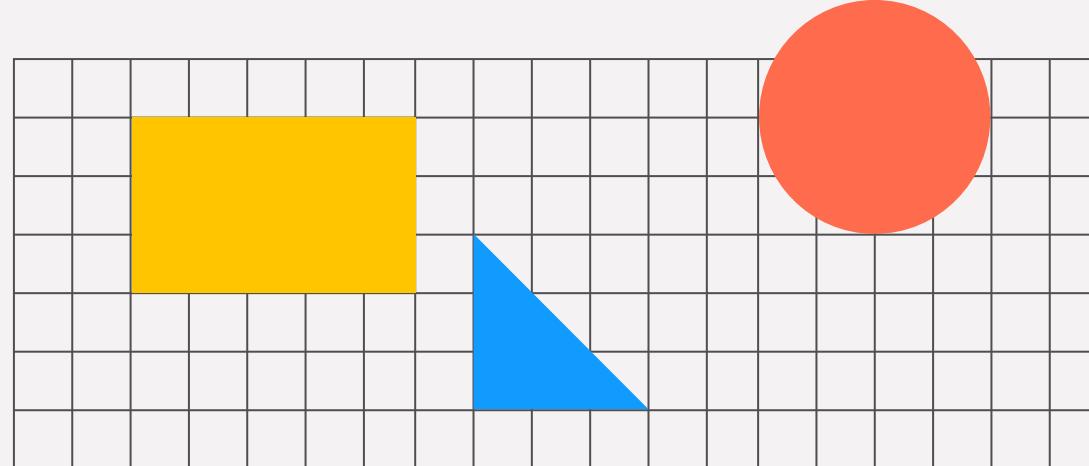


Pittsburgh Real Estate Market Analysis

Which parts of the Pittsburgh area are more and less expensive to buy a home, and why?



Data Cleaning

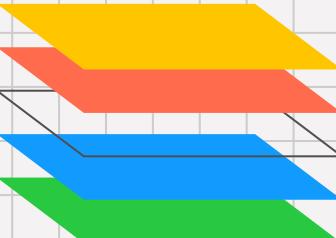


X_ID	PARD	PROPERTYHOUSENUM	PROPERTYTRACTNUM	PROPERTYADDRESS	PROPSCHOOLDIST
1	NA	NA	NA	NA	NA
2	12	0001000112000000	101	MARKET ST	NA
3	12	0001000112000000	101	MARKET ST	NA
4	12	0001000112000000	101	MARKET ST	NA
5	12	0001000112000000	101	MARKET ST	NA
6	12	0001000112000000	101	MARKET ST	NA
7	12	0001000112000000	101	MARKET ST	NA
8	12	0001000112000000	101	MARKET ST	NA
9	12	0001000112000000	101	MARKET ST	NA
10	12	0001000112000000	101	MARKET ST	NA
11	32	0001000112000000	101	MARKET ST	NA
12	32	0001000112000000	101	MARKET ST	NA
13	32	0001000112000000	101	MARKET ST	NA
14	32	0001000112000000	101	MARKET ST	NA
15	32	0001000112000000	101	MARKET ST	NA
16	32	0001000112000000	103	MARKET ST	NA
17	32	0001000112000000	103	MARKET ST	NA
18	32	0001000112000000	103	MARKET ST	NA
19	32	0001000112000000	103	MARKET ST	NA
20	32	0001000112000000	103	MARKET ST	NA
21	32	0001000112000000	103	MARKET ST	NA
22	32	0001000112000000	103	MARKET ST	NA
23	32	0001000112000000	103	MARKET ST	NA
24	32	0001000112000000	103	MARKET ST	NA
25	0001000112000000	103	MARKET ST	NA	

Parcels

Dataset from WPRDC
Assessor Records ~1972-2024

>500k observations
94 columns
338 MB .csv



Filtered by **residential** to exclude commercial space, vacant lots, etc.

Exclude sales prices **0** and **1**

Adjust for inflation:
Install quantmod and download CPI data. Extract year from sales date.
Merge CPI with dataset and calculate REALSALEPRICE

Filter for city of Pittsburgh, so exclude all rows where school district number is not 47

Breaking Down Sales Prices



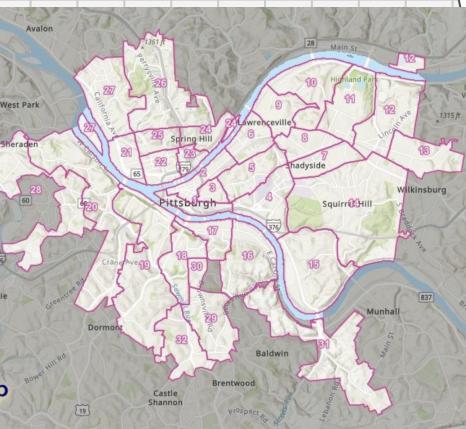
Geography

- 100

Find average sales price summarized by:

- Ward (MUNIDESC)
 - ZIP Code (PROPERTYZIP)
 - Neighborhood (NEIGHDESC)

In terms of geographic precision, it seems to be NEIGHDESC < PROPERTYZIP < WARD. Wards vary but they can be as small as one building, so let's use zip codes as a healthy medium.



We can understand sales price in terms of \$ per:

- Bedroom
 - Lot size (sqft)
 - Finished living area (sqft)

Preliminary Results



Average Sale Price by Ward

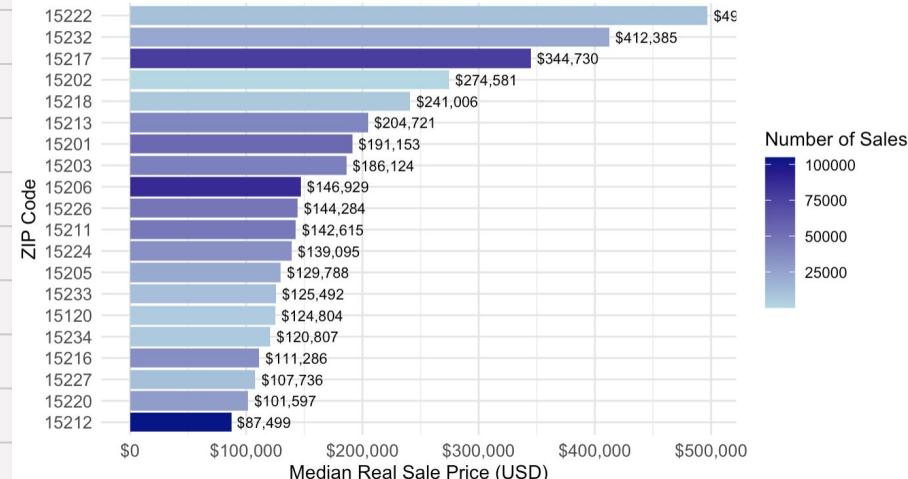
Sorted by Median Sale Price

Ward	Median Sale Price	Per Lot	Per Room	Per SqFt	Rank
2nd Ward - PITTSBURGH	\$466,863.20	\$584.07	\$370,603.66	\$429.83	1
14th Ward - PITTSBURGH	\$389,642.57	\$120.99	\$142,334.76	\$211.32	2
7th Ward - PITTSBURGH	\$385,168.56	\$218.04	\$190,788.89	\$258.18	3
22nd Ward - PITTSBURGH	\$272,914.17	\$224.42	\$129,283.44	\$179.64	4
11th Ward - PITTSBURGH	\$231,255.70	\$86.63	\$85,112.76	\$140.60	5
1st Ward - PITTSBURGH	\$218,588.55	\$1,182.67	\$161,947.49	\$202.01	6
9th Ward - PITTSBURGH	\$217,893.08	\$211.60	\$106,467.42	\$178.75	7
8th Ward - PITTSBURGH	\$207,593.14	\$109.15	\$86,648.89	\$133.55	8
17th Ward - PITTSBURGH	\$193,001.74	\$172.41	\$102,354.38	\$151.55	9
4th Ward - PITTSBURGH	\$185,107.11	\$365.35	\$102,746.19	\$154.69	10
6th Ward - PITTSBURGH	\$180,065.92	\$156.86	\$119,285.51	\$170.42	11



Top 20 Most Expensive ZIP Codes in Pittsburgh

Median Real Home Sale Prices with Transaction Volume



Adjusting for Zoning

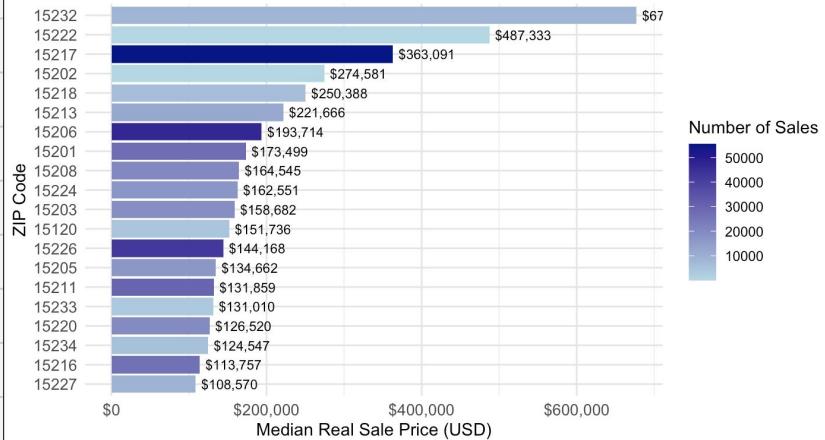
Average Sale Price by Ward (Single-Family)

Sorted by Median Sale Price

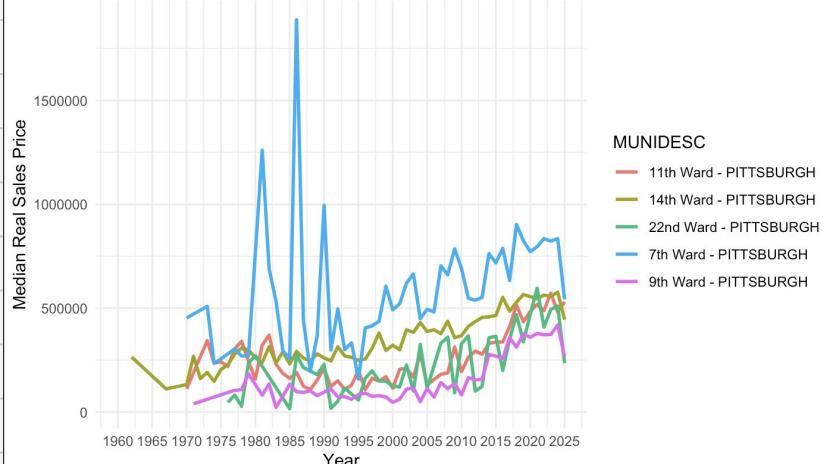
Ward	Median Sale Price	Per Lot	Per Room	Per SqFt	Rank
7th Ward - PITTSBURGH	\$676,346.95	\$164.18	\$199,770.22	\$259.30	1
14th Ward - PITTSBURGH	\$425,295.01	\$92.21	\$141,938.87	\$213.69	2
22nd Ward - PITTSBURGH	\$306,964.73	\$243.38	\$129,868.72	\$167.92	3
11th Ward - PITTSBURGH	\$306,141.94	\$87.58	\$97,907.72	\$159.93	4
9th Ward - PITTSBURGH	\$244,205.98	\$200.77	\$111,444.62	\$179.73	5
8th Ward - PITTSBURGH	\$204,193.36	\$95.29	\$84,267.97	\$131.01	6
4th Ward - PITTSBURGH	\$176,126.35	\$84.60	\$77,238.68	\$129.65	7
6th Ward - PITTSBURGH	\$161,142.08	\$126.05	\$85,241.14	\$133.79	8
10th Ward - PITTSBURGH	\$160,020.46	\$65.61	\$67,634.81	\$132.34	9
17th Ward - PITTSBURGH	\$159,143.81	\$132.63	\$82,144.91	\$126.53	10
15th Ward - PITTSBURGH	\$145,860.00	\$51.83	\$59,737.82	\$119.73	11

Top 20 Most Expensive ZIP Codes in Pittsburgh (Single-Family)

Median Real Home Sale Prices with Transaction Volume

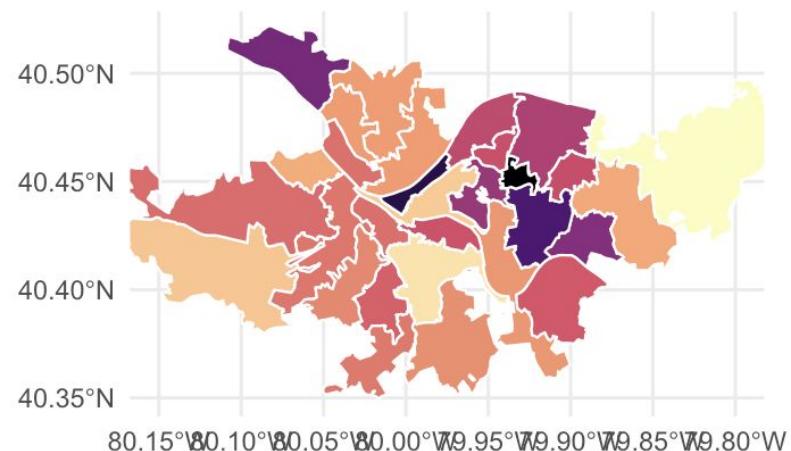


Top 5 Most Expensive Neighborhoods - Price Trends (Single-Family)

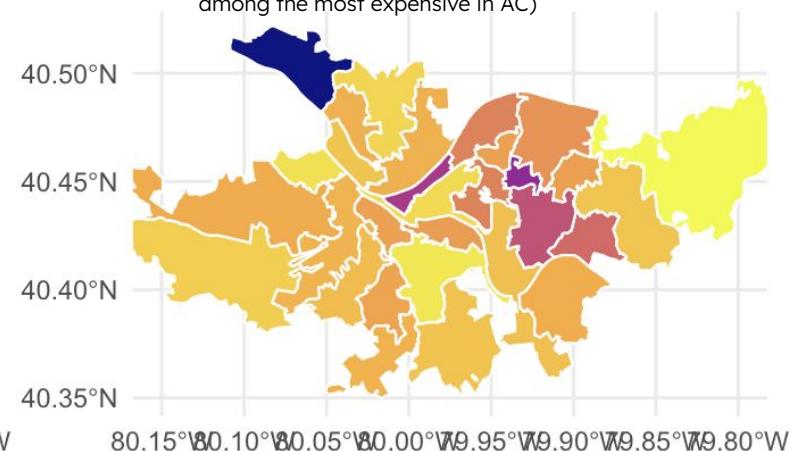


Pittsburgh Real Estate Heatmap

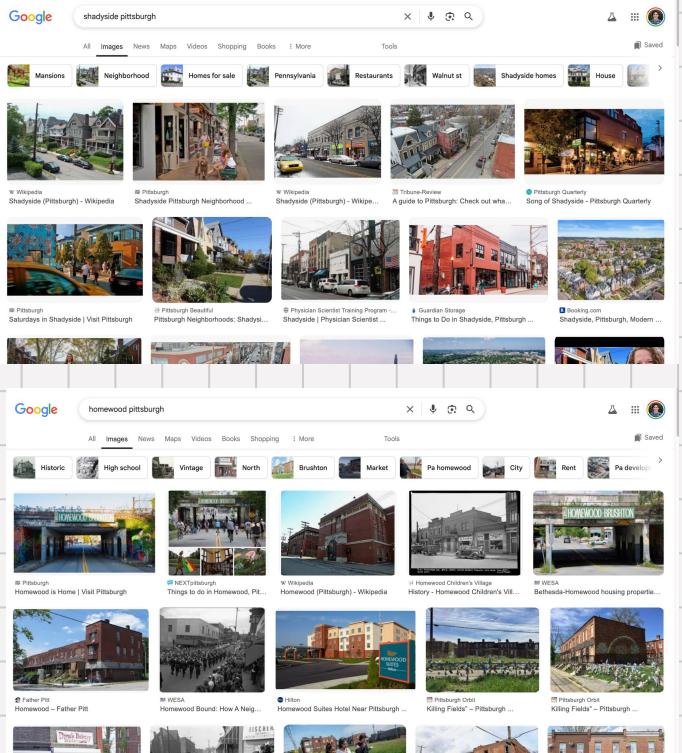
Median Sale Price by ZIP Code



Average Sale Price Per SqFt by ZIP Code



The disparities



Difference in single-family homes in **7th Ward (Shadyside)** and **13th Ward (Homewood/East Hills)**

- Median Real Sales Price: **\$642,993**
- Per sqft of lot: **\$148.44**
- Per bedroom: **\$182,054**
- Per sqft of living area: **\$222.60**

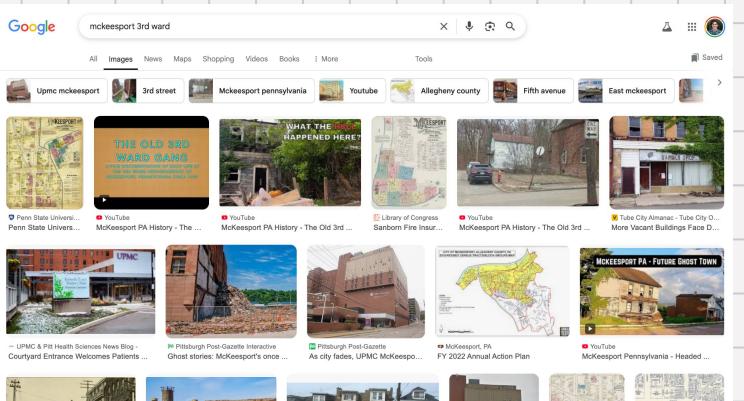
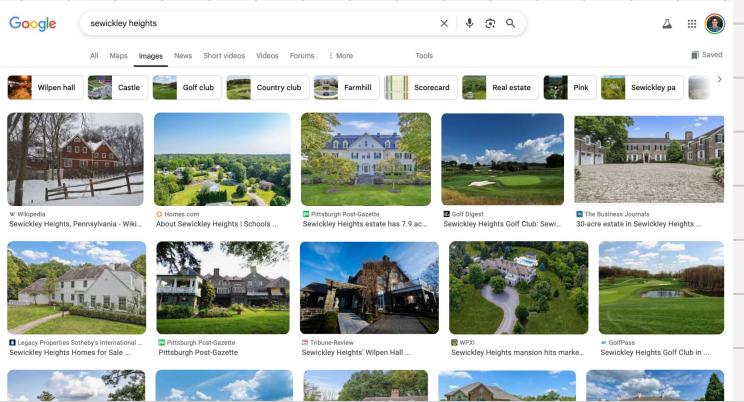
Rough affordability:

- 7th: **\$4,591.03**
mortgage+monthlies*
12*4= \$220,369 income
- 13th: **\$662**
mortgage*12*4=
\$31,776 income
Almost 7x more

Average Sale Price by Ward (Single-Family)					
Ward	Median Sale Price	Per Lot	Per Room	Per SqFt	Rank
7th Ward - PITTSBURGH	\$676,346.95	\$164.18	\$199,770.22	\$259.30	1
14th Ward - PITTSBURGH	\$425,295.01	\$92.21	\$141,938.87	\$213.69	2
22nd Ward - PITTSBURGH	\$306,964.73	\$243.38	\$129,868.72	\$167.92	3
11th Ward - PITTSBURGH	\$306,141.94	\$87.58	\$97,907.72	\$159.93	4
9th Ward - PITTSBURGH	\$244,205.98	\$200.77	\$111,444.62	\$179.73	5
8th Ward - PITTSBURGH	\$204,193.36	\$95.29	\$84,267.97	\$131.01	6
4th Ward - PITTSBURGH	\$176,126.35	\$84.60	\$77,238.68	\$129.65	7
6th Ward - PITTSBURGH	\$161,142.08	\$126.05	\$85,241.14	\$133.79	8
10th Ward - PITTSBURGH	\$160,020.46	\$65.61	\$67,634.81	\$132.34	9
17th Ward - PITTSBURGH	\$159,143.81	\$132.63	\$82,144.91	\$126.53	10
15th Ward - PITTSBURGH	\$145,860.00	\$51.83	\$59,737.82	\$119.73	11
23rd Ward - PITTSBURGH	\$97,428.43	\$116.13	\$64,379.56	\$95.20	22
26th Ward - PITTSBURGH	\$96,660.98	\$25.33	\$40,030.88	\$74.83	23
1st Ward - PITTSBURGH	\$95,513.16	\$175.74	\$95,927.78	\$100.94	24
25th Ward - PITTSBURGH	\$87,381.50	\$189.35	\$106,508.41	\$198.51	25
16th Ward - PITTSBURGH	\$77,469.05	\$55.50	\$48,706.95	\$82.67	26
24th Ward - PITTSBURGH	\$76,519.94	\$40.10	\$40,387.49	\$69.81	27
5th Ward - PITTSBURGH	\$73,088.42	\$32.81	\$37,715.00	\$69.62	28
Mt. Oliver	\$61,087.84	\$24.68	\$30,080.96	\$56.16	29
18th Ward - PITTSBURGH	\$55,275.14	\$32.49	\$36,673.12	\$65.86	30
30th Ward - PITTSBURGH	\$43,053.32	\$21.65	\$24,504.68	\$42.76	31
12th Ward - PITTSBURGH	\$37,976.01	\$17.58	\$20,364.91	\$41.32	32
13th Ward - PITTSBURGH	\$33,353.71	\$15.74	\$17,716.22	\$36.69	33
NA	NA	NaN	NaN	NaN	34

Data Source: WPRDC Assessor Records

Optional: Beyond city limits



Difference in single-family homes in **Sewickley Heights** and 3rd Ward of **McKeesport**

- Median Real Sales Price: **\$1,095,881.76**
- Per sqft of lot: ~equal (SH lot sizes larger?)
- Per bedroom: **\$295,542.55**
- Per sqft of living area: **\$268.85**

Rough affordability:

- SH: **\$5,635**
mortgage*12*4= \$270,480 income
- McK: **\$46**
mortgage*12*4= \$2,160 income
- Over 125x more

Average Sale Price by Ward (Single-Family) Sorted by Median Sale Price					
Ward	Median Sale Price	Per Lot	Per Room	Per SqFt	Rank
Sewickley Heights	\$1,104,897.13	\$6.24	\$300,820.11	\$280.15	1
Fox Chapel	\$776,459.60	\$18.00	\$218,959.33	\$241.89	2
Edgeworth	\$738,884.47	\$52.14	\$233,144.76	\$270.81	3
7th Ward - PITTSBURGH	\$676,346.95	\$164.18	\$199,770.22	\$259.30	4
Pine	\$579,167.25	\$25.85	\$154,826.56	\$193.66	5
Marshall	\$493,544.42	\$25.04	\$142,572.79	\$184.18	6
Franklin Park	\$483,710.58	\$22.19	\$129,900.79	\$182.31	7
Ohio	\$478,352.53	\$61.31	\$139,638.39	\$192.68	8
Glen Osborne	\$465,550.32	\$35.45	\$149,285.22	\$201.99	9
Sewickley Hills	\$438,984.63	\$12.73	\$150,342.01	\$204.31	10
13th Ward - PITTSBURGH	\$33,353.71	\$15.74	\$17,716.22	\$36.69	163
10th Ward - McKEESPORT	\$33,048.86	\$11.65	\$19,572.89	\$36.49	164
3rd Ward - DUQUESNE	\$31,526.67	\$10.90	\$17,673.09	\$36.80	165
North Braddock	\$30,837.69	\$16.99	\$22,856.60	\$46.66	166
7th Ward - McKEESPORT	\$28,036.39	\$34.63	\$44,602.98	\$79.44	167
2nd Ward - DUQUESNE	\$27,431.50	\$14.42	\$16,518.03	\$33.01	168
Braddock	\$27,122.73	\$15.32	\$14,915.65	\$28.71	169
2nd Ward - McKEESPORT	\$19,703.34	\$8.98	\$7,284.20	\$14.92	170
6th Ward - McKEESPORT	\$18,225.31	\$24.00	\$26,904.94	\$45.47	171
5th Ward - McKEESPORT	\$14,167.04	\$14.41	\$13,841.42	\$23.89	172
3rd Ward - McKEESPORT	\$9,015.37	\$6.63	\$5,277.56	\$11.30	173
NA	NA	NaN	NaN	NaN	174

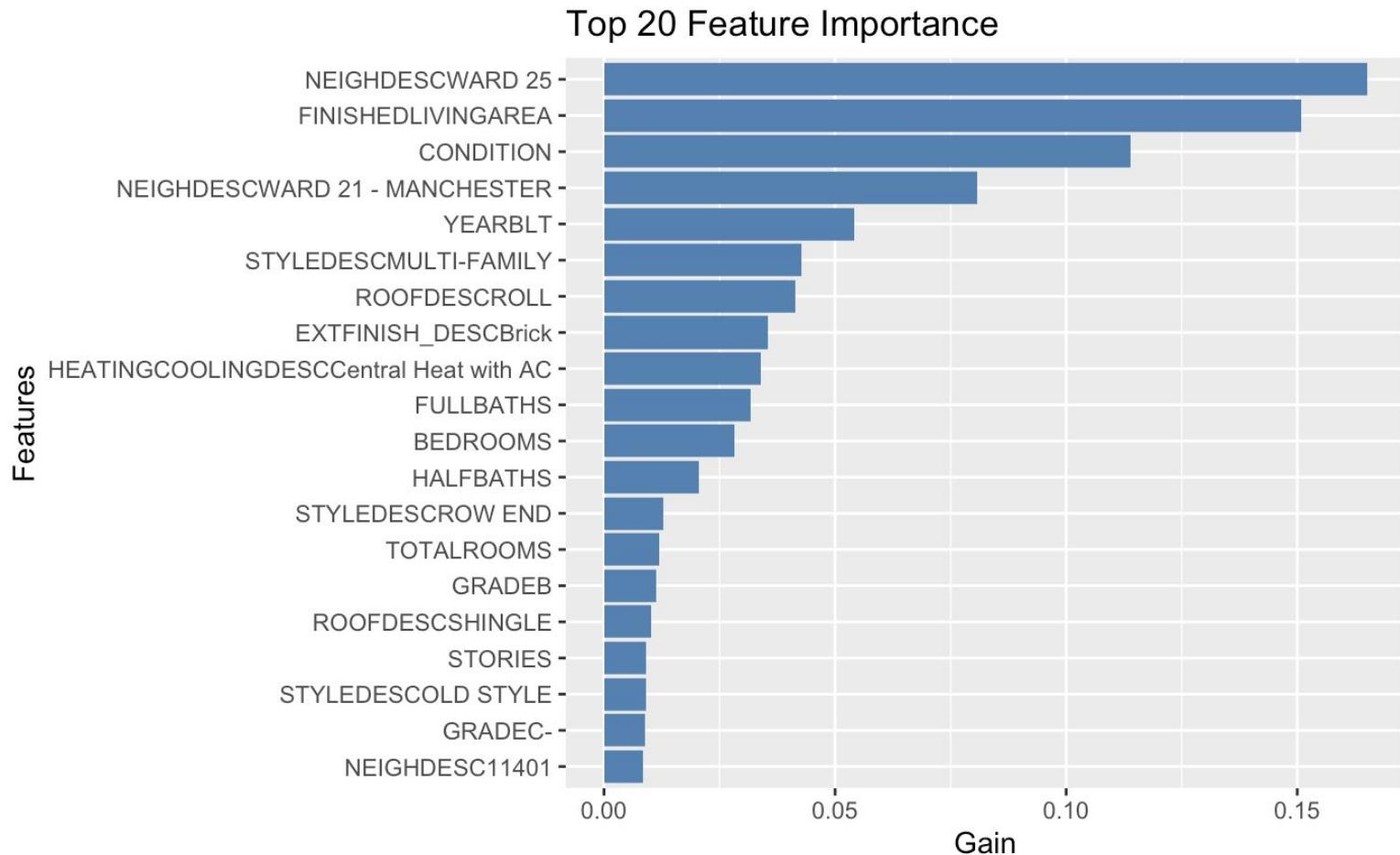
Data Source: WPRDC Assessor Records

XGBoost

70-30
Decision trees
for large
datasets with
many
categories -
100 rounds

RMSE: 189804.5
R-squared:
0.8085214

Location and square
footage are large
determinants of
value, but that's not
really the question –
how can we account
for size to capture the
inherent value of the
property and the land
it sits on?

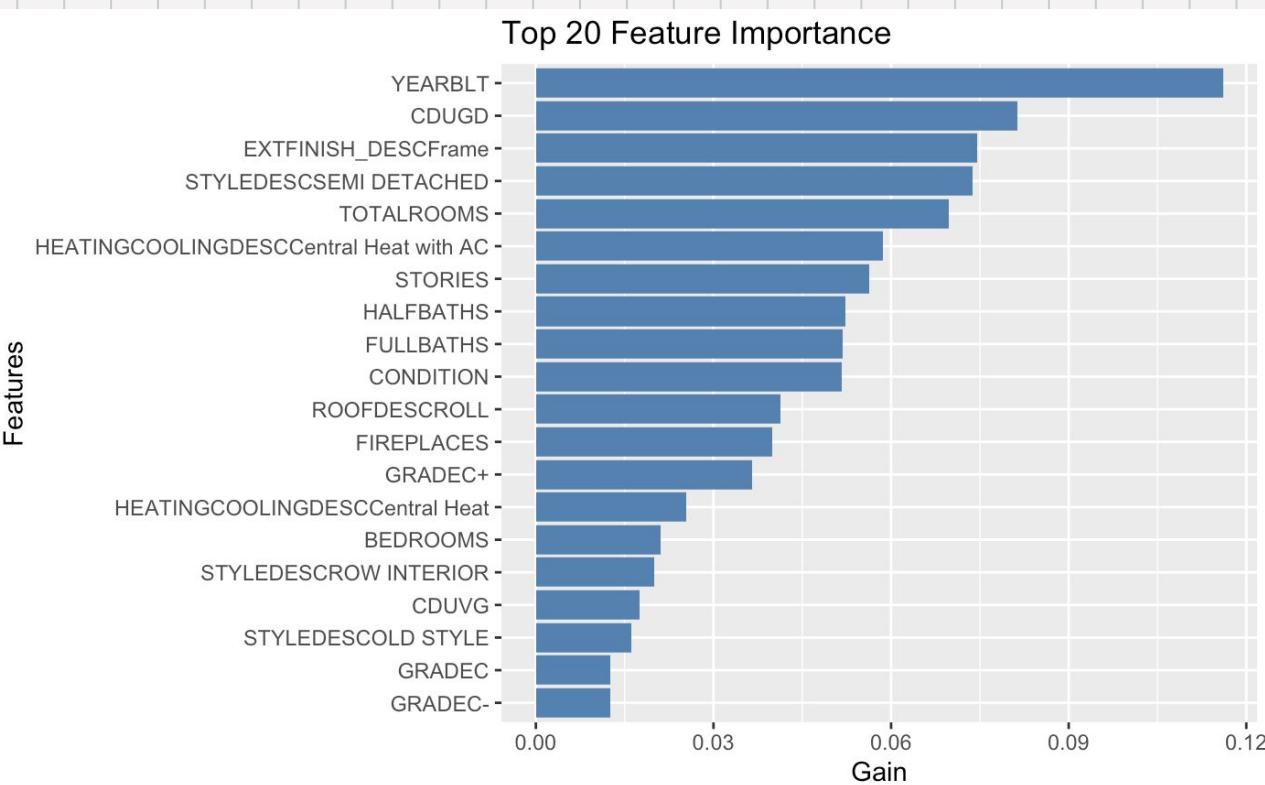


Re-run instead using REALSALEPRICE_PER_SQFT as target variable
and excluding FINISHEDLIVINGAREA and NEIGHDESC

RMSE: 103.0189

R-squared: 0.5664012

Predictably, square footage and location are very strong predictors of sales price, but when we remove that and predict on a per sqft basis, we get some interesting but weaker results



Compare to McCullough
(2021) analysis of
Pittsburgh housing sales
prices at the College of
Wooster, focused on
green space proximity –

Distance from green spaces is significantly negatively associated with ln(salesprice), as well as age. Positively and significantly associated were floors, bathrooms, bedrooms, and garage

McCullough 67

Variable	Table 4: Regression Results		OLS		S2SLS (LAG)		S2SLS (SARMA)	
	Reduced Model	Full model	Reduced Model	Full Model	Reduced Model	Full Model	Reduced Model	Full Model
<i>LNLotArea</i>	0.1302832*	0.1134114*	0.1349763*	0.1183388*	0.1535432*	0.1447697*		
	(.0060679)	(.0061256)	(.0061117)	(.0061673)	(.0064051)	(.0064102)		
<i>LNArea</i>	0.7364534*	0.7191867*	0.7374818*	0.7174655*	0.6671893*	0.6349905*		
	(.0171145)	(.0182965)	(.0170214)	(.0181884)	(.0155088)	(.0162772)		
<i>NumFloors</i>	-0.0458779*	-0.0274019*	-0.0469953*	-0.0298104*	-0.0359495*	-0.0296295*		
	(.0085867)	(.0088044)	(.0085427)	(.0087661)	(.0075006)	(.0076530)		
<i>Age</i>	-0.0037811*	-0.0032442*	-0.0038265*	-0.0036271*	-0.0054439*	-0.0068101*		
	(.0001827)	(.0006812)	(.0001821)	(.0006783)	(.0001892)	(.0006737)		
<i>NumBaths</i>	0.1676161*	0.1518793*	0.16489*	0.1492634*	0.1159399*	0.1072775*		
	(.0060705)	(.0061113)	(.0060337)	(.0060674)	(.0050442)	(.0050384)		
<i>NumBedrooms</i>	0.0577963*	0.0529214*	0.0577468*	0.0533392*	0.0371143*	0.036967*		
	(.0075699)	(.0075621)	(.0074981)	(.0074885)	(.0061870)	(.0061643)		
<i>Age2</i>		-0.00000426		-0.0000019		0.0000092		
		(5.10e-06)		(0.0000051)		(0.0000048)		
<i>NumFire</i>		0.1106865*		0.1097374*		0.0733866*		
		(.0068108)		(.0067770)		(.0056384)		
<i>Garage</i>		0.02261*		0.0221874*		0.0165643*		
		(.0041974)		(.0041923)		(.0036513)		
<i>GreenDist</i>	-3.106815*	-2.258265*	-2.8999949*	-2.1333877*	-2.4038193*	-2.3404086*		
	(.7366672)	(.7403561)	(.7339994)	(.7366525)	(.8924741)	(.8973546)		
<i>LNSALE_LAG</i>			0.0092641*	0.0089326*	0.0091012*	0.0090906*		
			(.00008201)	(.00008117)	(.00009812)	(.00009746)		
<i>LAMDA</i>					0.5482674*	0.5480349*		
					(.0088611)	(.0089347)		
<i>CONST</i>	5.133945*	5.347221*	4.9989031*	5.2454276*	5.606361*	5.9284311*		
	(.1098134)	(.1253773)	(.1100525)	(.1252759)	(.1081926)	(.1195326)		
R2 (spatial R2)	0.6131	0.6207	0.6188	0.626	0.6114	0.6175		

Note: * indicates significance at $\alpha < 0.01$, ** indicates significance at $\alpha < 0.05$, *** indicates significance at $\alpha < 0.1$. Robust Standard errors are in parentheses.

McCullough, Will, "The Great Outdoors, Indoors:
An Evaluation of Green Spaces and Housing
Prices in Pittsburgh, PA" (2021). Senior
Independent Study Theses. Paper 9589.