

HOME PRICES BY TIM RAPPOLD

**IS ADDING THAT BEDROOM
WORTH IT?**



SHE'S JUST A FEW
MONTHS OLD THERE

**MY DAUGHTER
ARIA MARIE**

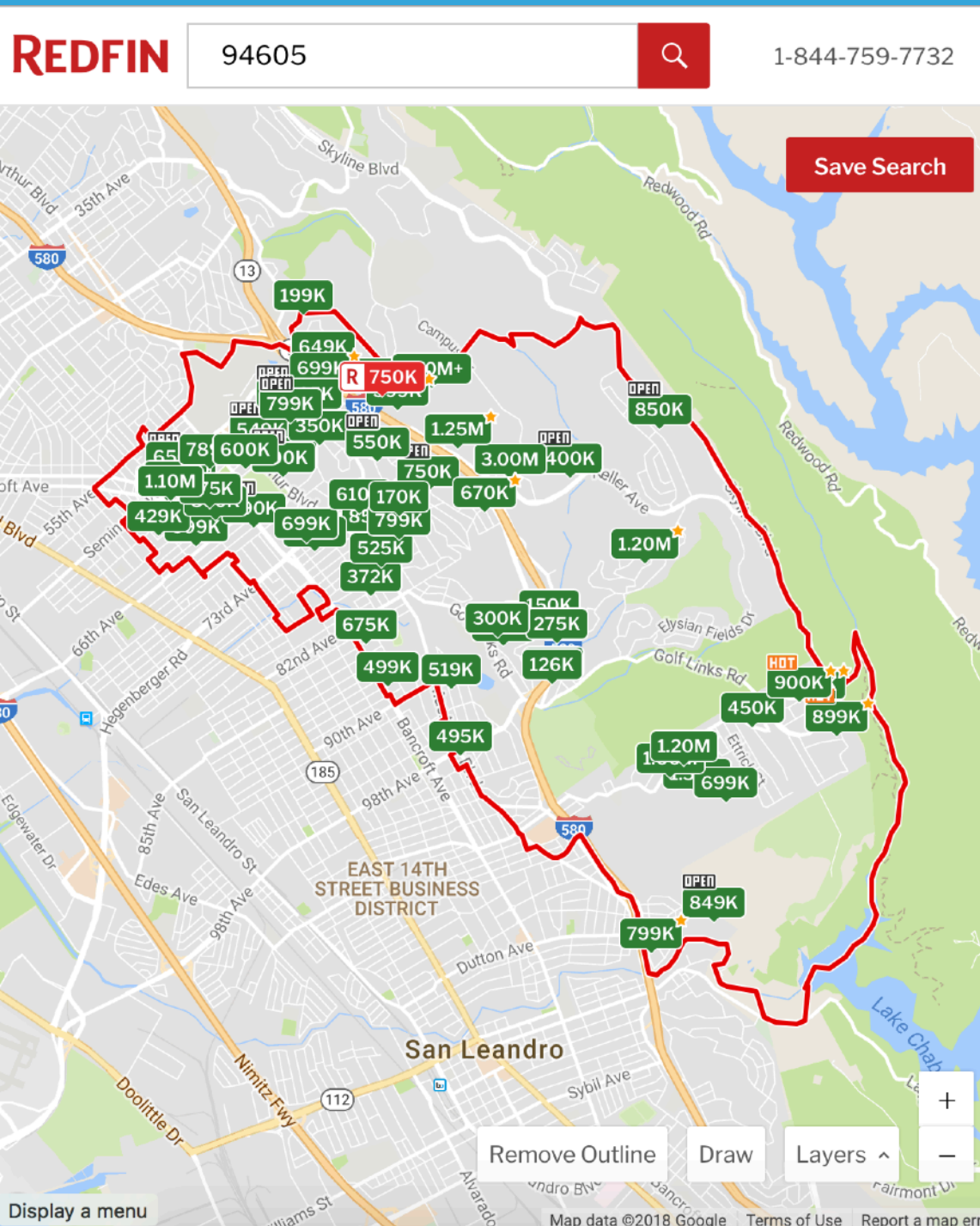


OUR HOUSE
2 BR. / 1 BATH

TOO SMALL

**BUILD A TOOL THAT HELPS ME
UNDERSTAND HOW CHANGES IN HOME
FEATURES AFFECT SALES VALUE.**

Me, thinking to myself.



IS ADDING THAT BEDROOM WORTH IT?

>10k listings from Redfin

27 Zip Codes

Los Angeles & Oakland

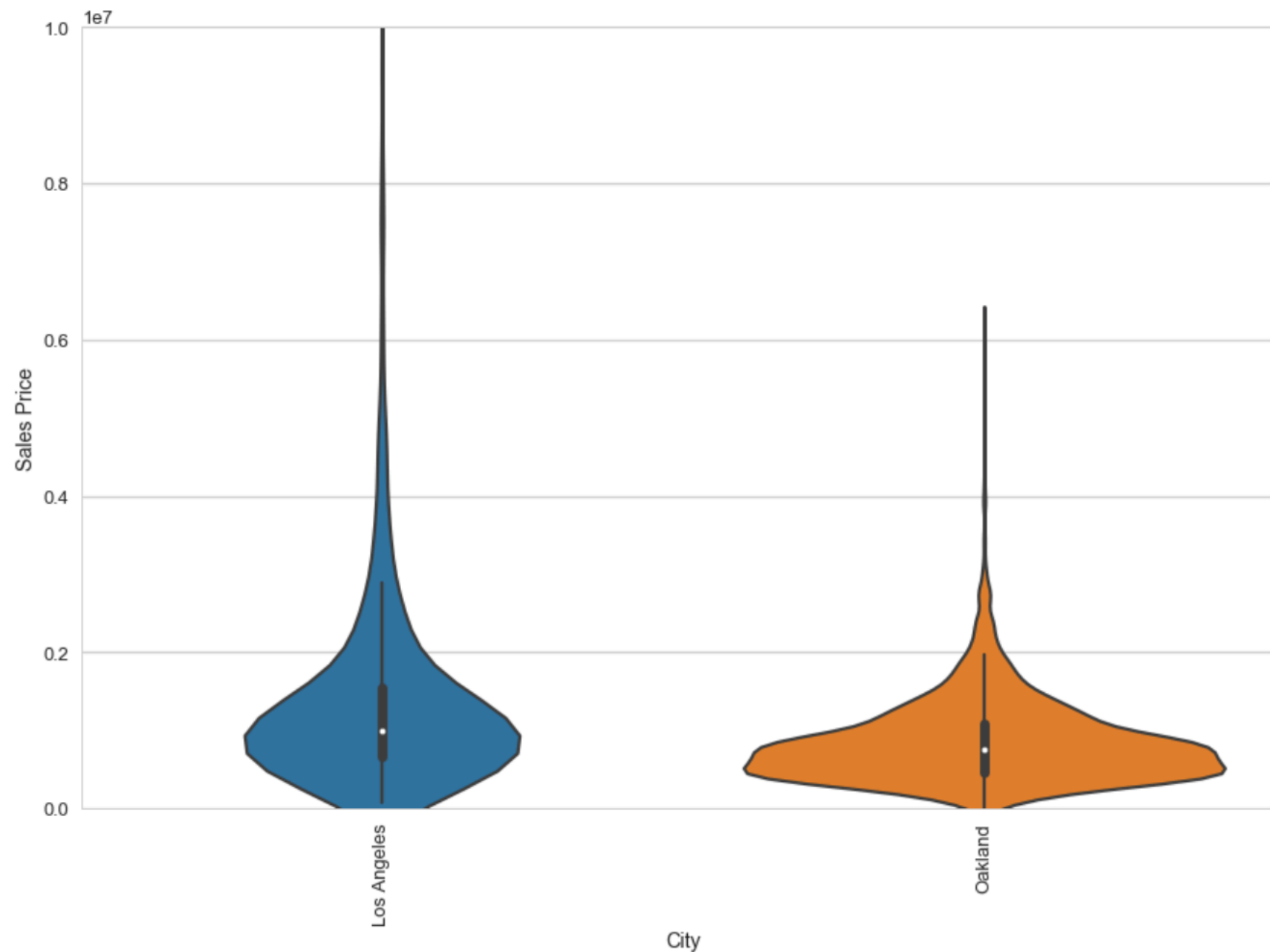
Sold in last 3 years

Linear Regression and Lasso

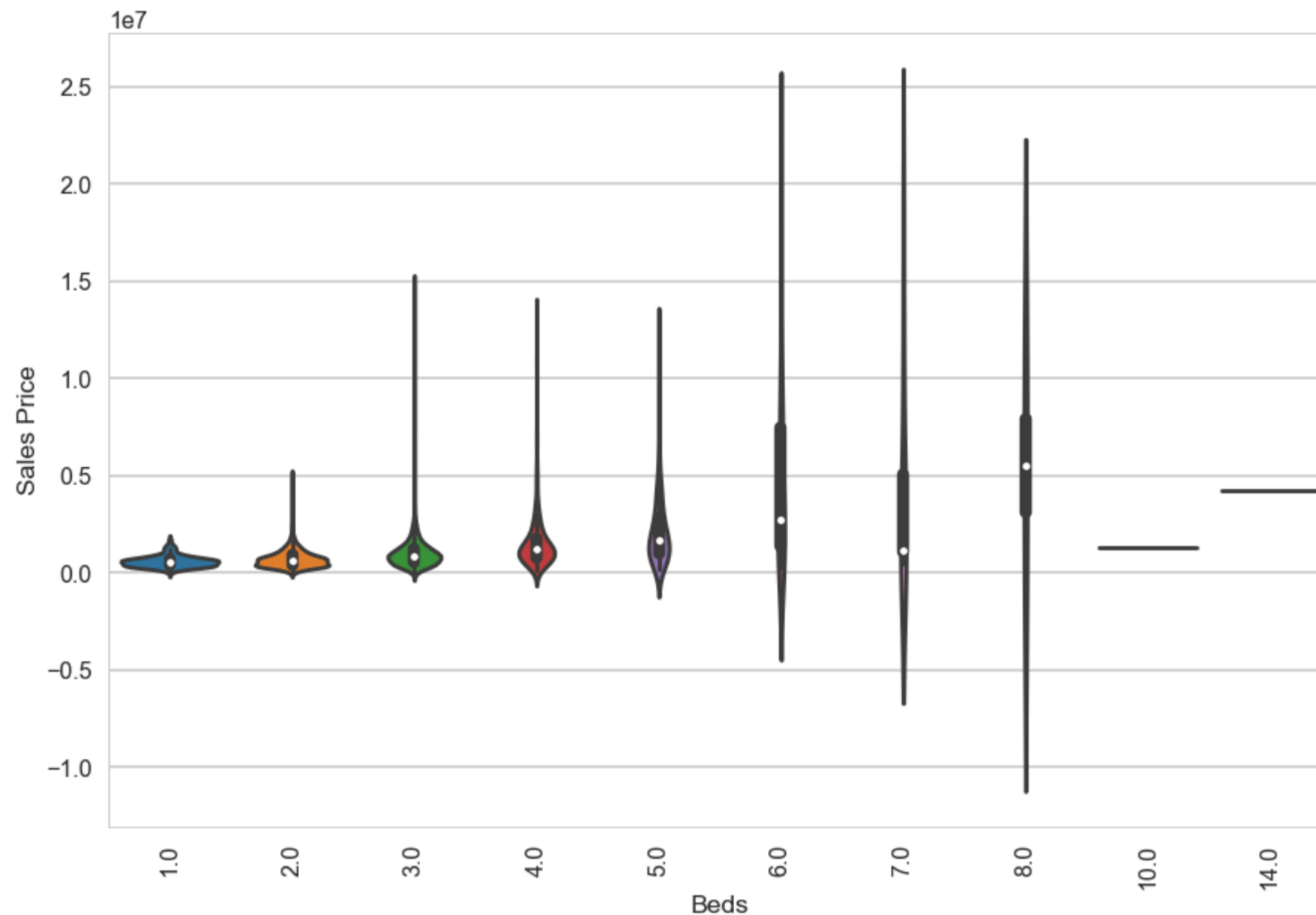
FEATURES CONSIDERED:

- ▶ Bedrooms
- ▶ Bathrooms
- ▶ Square footage
- ▶ Lot size
- ▶ Zip Codes
- ▶ Style (Single family, condo, etc)
- ▶ Two cities: Oakland and LA
- ▶ And more...

A TALE OF TWO CITIES

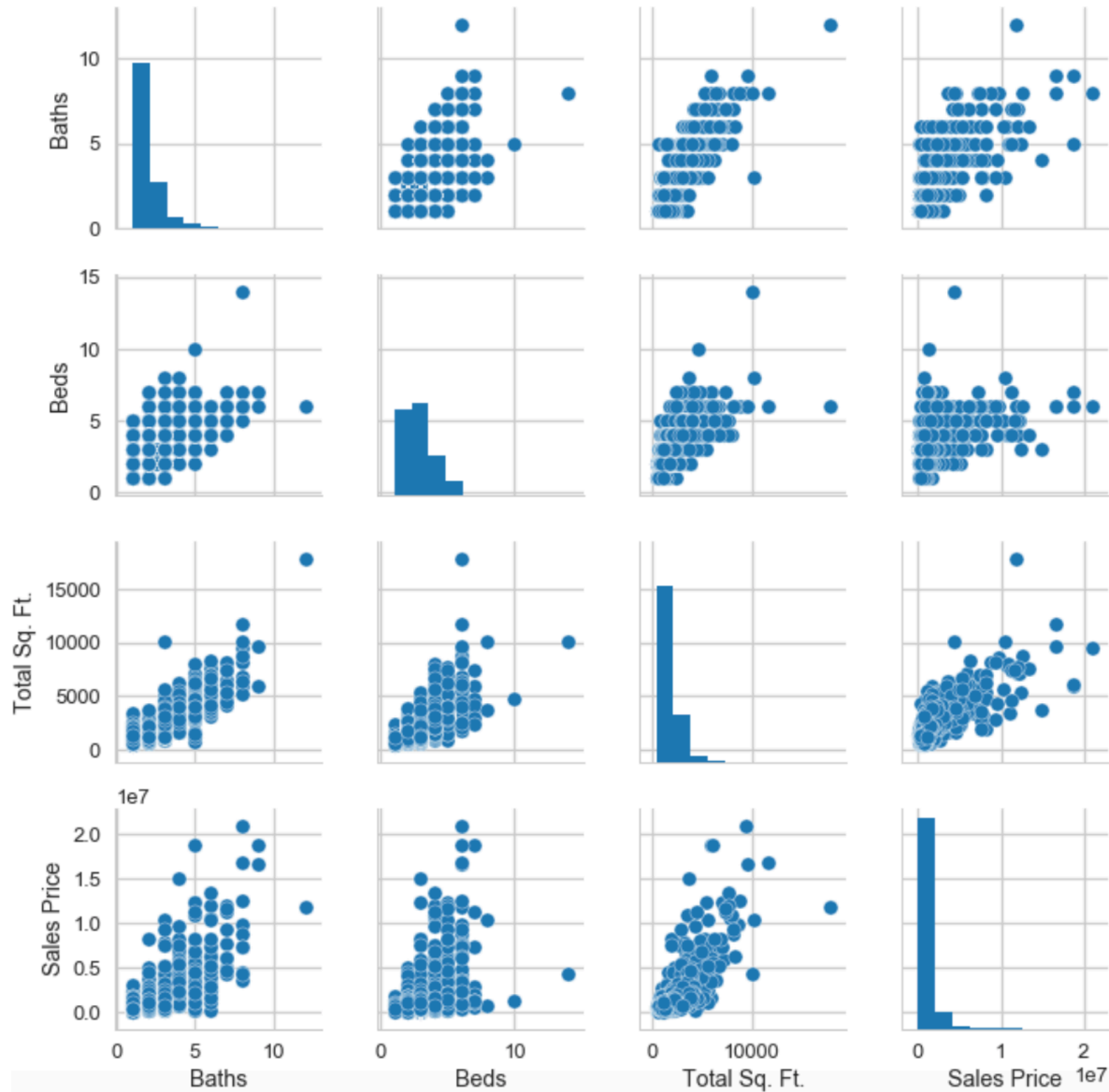


INDIVIDUAL FEATURE RELATIONSHIPS ARE COMPLICATED



ADDING **BEDROOMS** DOESN'T HAVE A SIMPLE EFFECT ON SALES PRICE

MULTICOLLINEARITY: ISSUE OF ATTRIBUTION



TWO MODEL APPROACHES

OLS,
MANUAL
MODEL
SELECTION
(P-VALUE AND VIF INSPECTION)

LASSO WITH CROSS-TERMS,
5-FOLD CV,
GRID SEARCH,
BACKWARD SELECTION

TWO MODEL APPROACHES

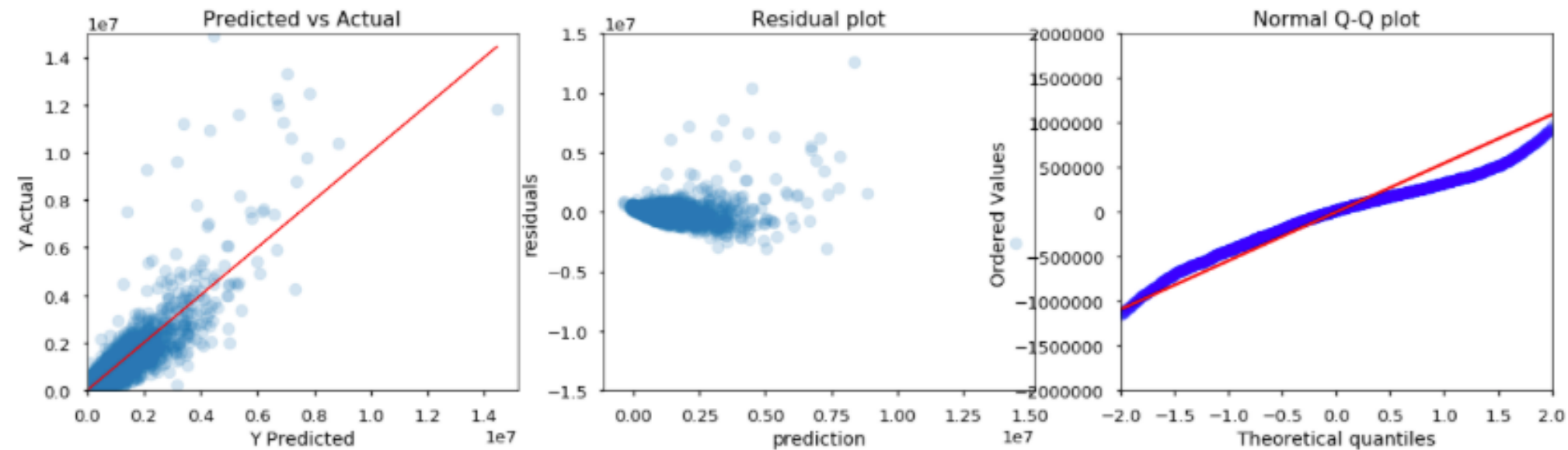
21 MANUALLY SELECTED FEATURES
EXCELLENT STATISTICAL SIGNIFICANCE ($P < 5\%$)
MINIMAL COLLINEARITY ($VIF < 5$)
YET STILL POORLY CONDITIONED
POOR R^2 VALUE: 0.67

OLS,
MANUAL
MODEL
SELECTION (P-VALUE
AND VIF INSPECTION)

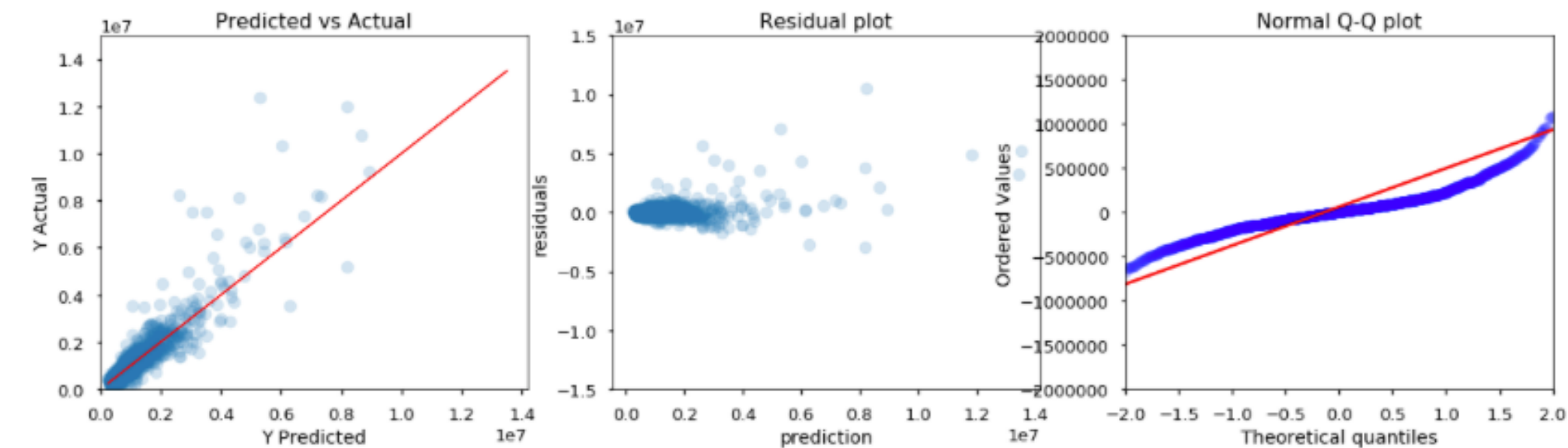
690 FEATURES ALGORITHMICALLY REDUCED TO 64
SQRT SCALING
HIGH COLLINEARITY ($VIF \rightarrow \infty$)
HIGH R^2 VALUE ON TEST SET: 0.86

LASSO WITH
CROSS-TERMS,
5-FOLD CV,
GRID SEARCH,
BACKWARD SELECT

TWO MODEL APPROACHES



**OLS,
MANUAL
MODEL
SELECTION (P-VALUE
AND VIF INSPECTION)**



**LASSO WITH
CROSS-TERMS,
5-FOLD CV,
GRID SEARCH,
BACKWARD SELECT**

RESULTS: WHAT'S A GOOD BUDGET FOR ADDING

- ▶ 1 BEDROOM?
- ▶ 1 BED + 1 BATH?

Adding Rooms in Oakland (Eastmont Hills, 94605)

	Baths	Beds	Total Sq. Ft.	Predicted Price	Δ Predicted Price
1	1	2	1700	\$793,459	\$0
2	1	3	1900	\$837,277	\$43,818
3	2	3	2100	\$882,274	\$88,815

RESULTS: WHAT IF YOU DID THE SAME THING IN A DIFFERENT ZIP CODE?

- ▶ 1 BEDROOM?
- ▶ 1 BED + 1 BATH?

Adding Rooms in Oakland (Eastmont Hills, 94605)

	Baths	Beds	Total Sq. Ft.	Predicted Price	Δ Predicted Price
1	1	2	1700	\$793,459	\$0
2	1	3	1900	\$837,277	\$43,818
3	2	3	2100	\$882,274	\$88,815

Adding Rooms in LA (Ladeira Hts., 90056)

	Baths	Beds	Total Sq. Ft.	Predicted Price	Δ Predicted Price
4	1	2	1700	\$1,215,749	\$0
5	1	3	1900	\$1,269,849	\$54,100
6	2	3	2100	\$1,325,126	\$109,377

CONCLUSIONS AND NEXT STEPS

- ▶ Developed a location-sensitive tool to probe the relationship between home features and home prices
- ▶ Insignificant: Time
- ▶ Future work ought to focus on assessing statistical significance of model, confidence intervals, and a window of accuracy



APPENDIX FOR OLS

OLS Regression Results						
=====						
Dep. Variable:	Sales Price	R-squared:	0.671			
Model:	OLS	Adj. R-squared:	0.669			
Method:	Least Squares	F-statistic:	296.3			
Date:	Fri, 27 Apr 2018	Prob (F-statistic):	0.00			
Time:	07:26:20	Log-Likelihood:	-45643.			
No. Observations:	3073	AIC:	9.133e+04			
Df Residuals:	3051	BIC:	9.146e+04			
Df Model:	21					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	2.978e+04	3.63e+04	0.821	0.412	-4.13e+04	1.01e+05
Total Sq. Ft.	719.8426	14.389	50.028	0.000	691.630	748.055
Lot Size Sq. Ft.	1.8794	0.457	4.113	0.000	0.983	2.775
Townhouse	-5.15e+05	5.95e+04	-8.661	0.000	-6.32e+05	-3.98e+05
90024	5.98e+05	1.18e+05	5.052	0.000	3.66e+05	8.3e+05
90026	-1.224e+05	6.43e+04	-1.905	0.057	-2.49e+05	3615.998
90041	-2.59e+05	5.62e+04	-4.609	0.000	-3.69e+05	-1.49e+05
90049	1.486e+06	7.66e+04	19.408	0.000	1.34e+06	1.64e+06
90056	-8.211e+05	6.7e+04	-12.259	0.000	-9.52e+05	-6.9e+05
90221	-5.876e+05	5.63e+04	-10.434	0.000	-6.98e+05	-4.77e+05
90301	-3.276e+05	7.45e+04	-4.397	0.000	-4.74e+05	-1.82e+05
90403	9.006e+05	8.69e+04	10.359	0.000	7.3e+05	1.07e+06
94601	-5.245e+05	6.45e+04	-8.126	0.000	-6.51e+05	-3.98e+05
94602	-3.362e+05	5.66e+04	-5.938	0.000	-4.47e+05	-2.25e+05
94605	-6.202e+05	6.09e+04	-10.187	0.000	-7.4e+05	-5.01e+05
94606	-4.197e+05	6.79e+04	-6.180	0.000	-5.53e+05	-2.87e+05
94607	-3.791e+05	1.13e+05	-3.361	0.001	-6e+05	-1.58e+05
94609	-1.576e+05	6.99e+04	-2.253	0.024	-2.95e+05	-2.04e+04
94610	-1.986e+05	7.39e+04	-2.688	0.007	-3.43e+05	-5.37e+04
94611	-4.41e+05	5.76e+04	-7.650	0.000	-5.54e+05	-3.28e+05
94618	-1.499e+05	6.11e+04	-2.455	0.014	-2.7e+05	-3.02e+04
94619	-4.058e+05	5.56e+04	-7.304	0.000	-5.15e+05	-2.97e+05
=====						
Omnibus:	3645.475	Durbin-Watson:	2.038			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	870881.657			
Skew:	5.913	Prob(JB):	0.00			
Kurtosis:	84.619	Cond. No.	3.39e+05			
=====						

	VIF	Features
0	3.0534	Total Sq. Ft.
1	1.3051	Lot Size Sq. Ft.
2	1.5105	Townhouse
3	1.0716	90024
4	1.0637	90026
5	1.0982	90041
6	1.3714	90049
7	1.1839	90056
8	1.0664	90221
9	1.0575	90301
10	1.2923	90403
11	1.0539	94601
12	1.1084	94602
13	1.1086	94605
14	1.0535	94606
15	1.0173	94607
16	1.0537	94609
17	1.1048	94610
18	1.2353	94611
19	1.1632	94618
20	1.1020	94619

SAMPLE OF DOWN-SELECTED FEATURES FROM LASSO REGRESSION

```
['Total Sq. Ft.',  
 'Year Built',  
 '90221',  
 '90232',  
 '94601',  
 '94603',  
 '94606',  
 '94618',  
 '94619*Total Sq. Ft.',  
 '94619*Lot Size Sq. Ft.',  
 '94619*Single Family Residential',  
 ...]
```

... 64 features ...

```
'90024*Total Sq. Ft.',  
'90024*Single Family Residential',  
'Los Angeles*Baths',  
'Los Angeles*Total Sq. Ft.',  
'Los Angeles*Single Family Residential',  
'Single Family Residential*Baths',  
'Single Family Residential*Total Sq. Ft.',  
'Single Family Residential*Lot Size Sq. Ft.']
```