Sticker Shock: Bay Area Housing Market

REDFIN

Predicting Home Prices with Linear Regression

tri le



Why Home Prices?

- Covid-19
- Inflation
- Economic Downturn
- Supply chain Shortages

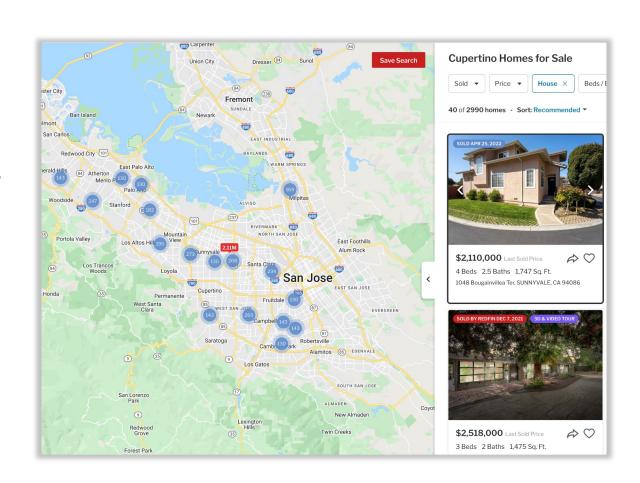




R

Why Redfin?

- Web Scraping-Friendly
- Extensive Real Estate Listings
- Feature-Rich Data and Filters



R

EDA



- Single Family Homes
- Sold within last 6 months
- 1600+ Data points
- 13 Features
- 7 Cities
- 18 Zip Codes



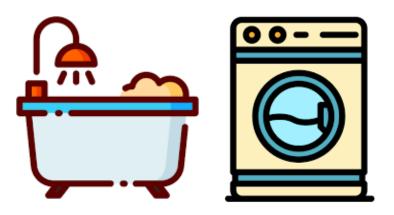
Data Point Example



Sold Price

- Beds
- Baths
- Floors
- Garage
- Lot Size
- Home Size
- Year Built

- School Score
- Laundry
- Heating
- A/C
- Pool
- City







Model Performance (Before)

Baseline Linear Regression Results

R-Squared: 0.783

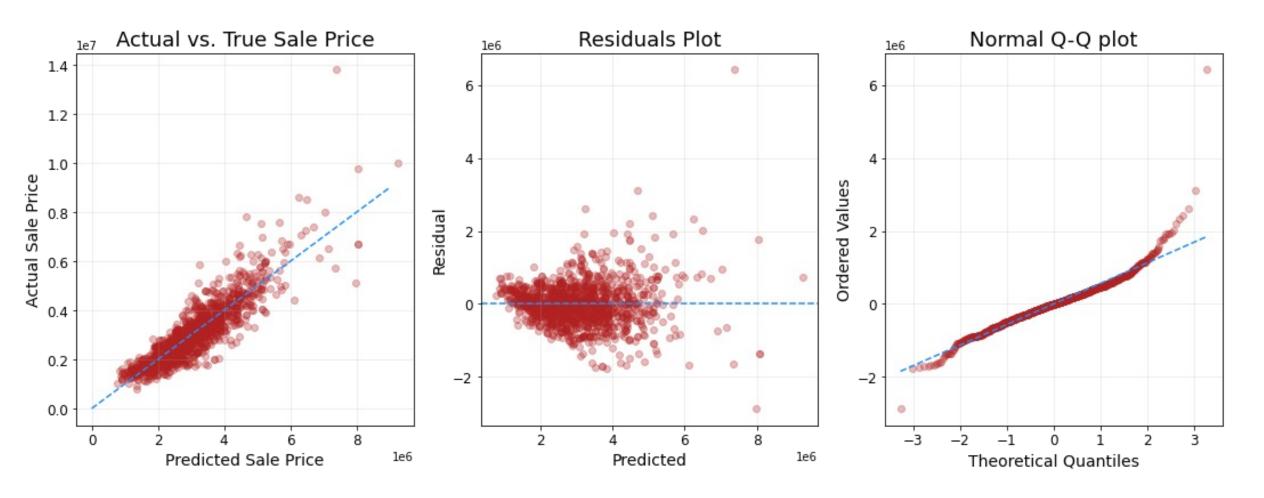
Adjusted R-Squared: 0.782

Mean Absolute Error: \$416,244

Root Mean Squared Error: \$586,488



Model Evaluation (Before)



* Baseline Linear Regression - Untouched Dataset

Feature Engineering

Data	columns (total 14	columns):							
#	Column	Non-Null Count	Dtype						
0	sold price	1292 non-null	float64						
1	beds	1292 non-null	float64						
2	baths	1292 non-null	float64						
3	floors	1292 non-null	float64						
4	garage spaces	1292 non-null	int64						
5	lot size	1292 non-null	float64						
6	home size	1292 non-null	float64						
7	school score avg	1292 non-null	float64						
8	laundry	1292 non-null	bool						
9	heating	1292 non-null	bool						
10	air conditioning	1292 non-null	bool						
11	pool	1292 non-null	bool						
12	city	1292 non-null	object						
13	age of house	1292 non-null	float64						



Data	columns (total 32 c	olumns):	
#	•	Non-Null Count	Dtype
0	sold price	1292 non-null	float64
1	lot size	1292 non-null	float64
2	home size	1292 non-null	float64
3	school score avg	1292 non-null	float64
4	laundry	1292 non-null	int64
5	heating	1292 non-null	int64
6	air conditioning	1292 non-null	int64
7	pool	1292 non-null	int64
8	age of house	1292 non-null	float64
9	beds_3.0	1292 non-null	uint8
10	beds_4.0	1292 non-null	uint8
11	beds_5.0	1292 non-null	uint8
12	beds_6+	1292 non-null	uint8
13	baths_1.5	1292 non-null	uint8
14	baths_2.0	1292 non-null	uint8
15	baths_2.5	1292 non-null	uint8
	baths_3.0	1292 non-null	uint8
	baths_3.5	1292 non-null	uint8
	baths_4.0	1292 non-null	uint8
19	baths_4.5	1292 non-null	uint8
20	baths_5.0	1292 non-null	uint8
21	baths_6+	1292 non-null	uint8
22	floors_2.0	1292 non-null	uint8
23	floors_3.0	1292 non-null	uint8
24	garage spaces_1	1292 non-null	uint8
25	garage spaces_2	1292 non-null	uint8
26	garage spaces_3+	1292 non-null	uint8
27	city_LOS ALTOS	1292 non-null	uint8
28	city_MOUNTAIN VIEW		uint8
29	city_PALO ALTO	1292 non-null	uint8
30	city_SANTA CLARA		
31	city_SUNNYVALE	1292 non-null	uint8



Model Performance (After)

Baseline Linear Regression Results

R-Squared: 0.801

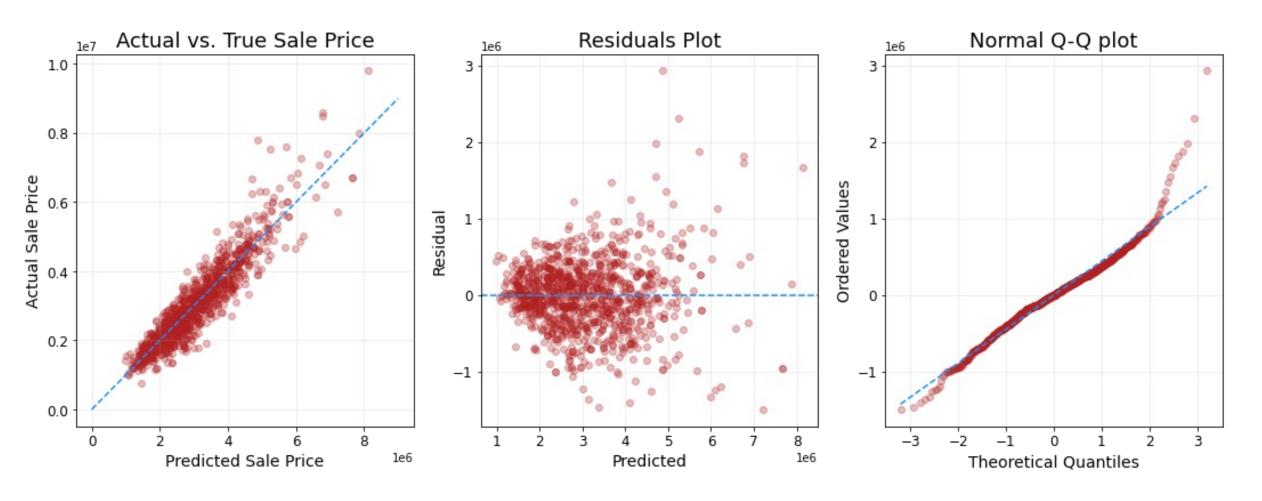
Adjusted R-Squared: 0.774

Mean Absolute Error: \$398,667

Root Mean Squared Error: \$543,227



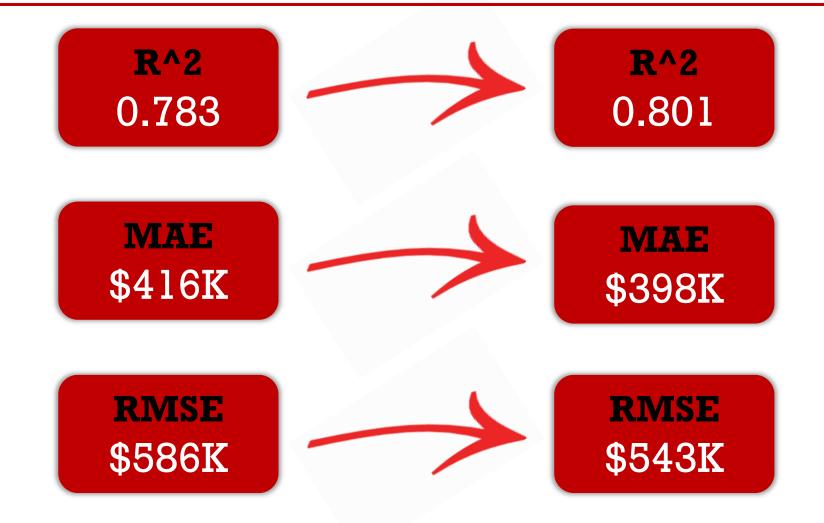
Model Evaluation (After)



* Elastic Net Regression – Test Dataset



Conclusion



^{*} Elastic Net Regression – Test Dataset



Conclusion

Interesting Feature Coefficients

home size: 773471.44

school score avg: 269845.16

lot size: 130766.26

*city - palo alto: 336679.56

*city – los altos: 205731.06

age of house: 0.00



Thank You

Questions?

Metis May 2022

Thank You

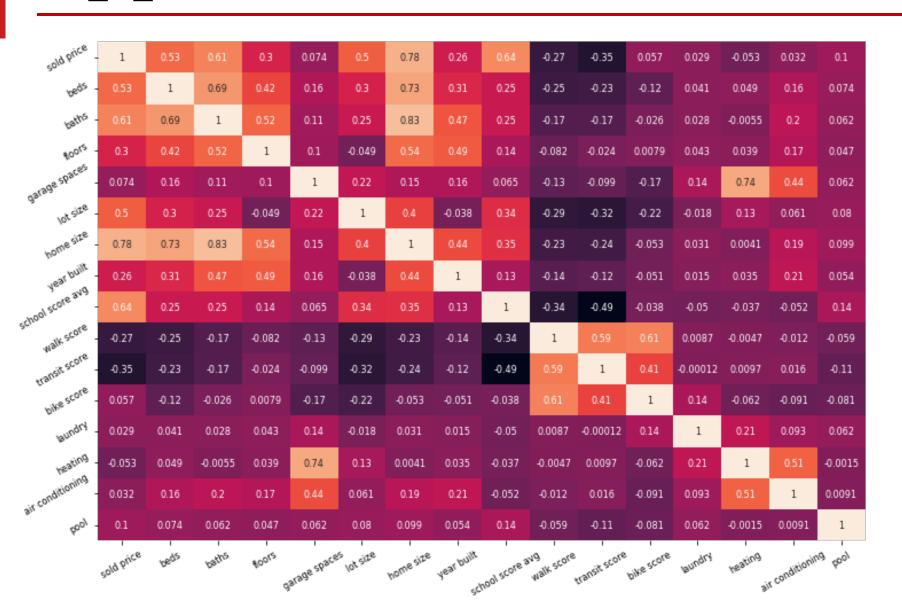
REDFIN

Questions?

metis 2022 may

R

Appendix



- 1.0

- 0.8

- 0.6

- 0.4

- 0.2

- 0.0

- -0.2

- -0.4

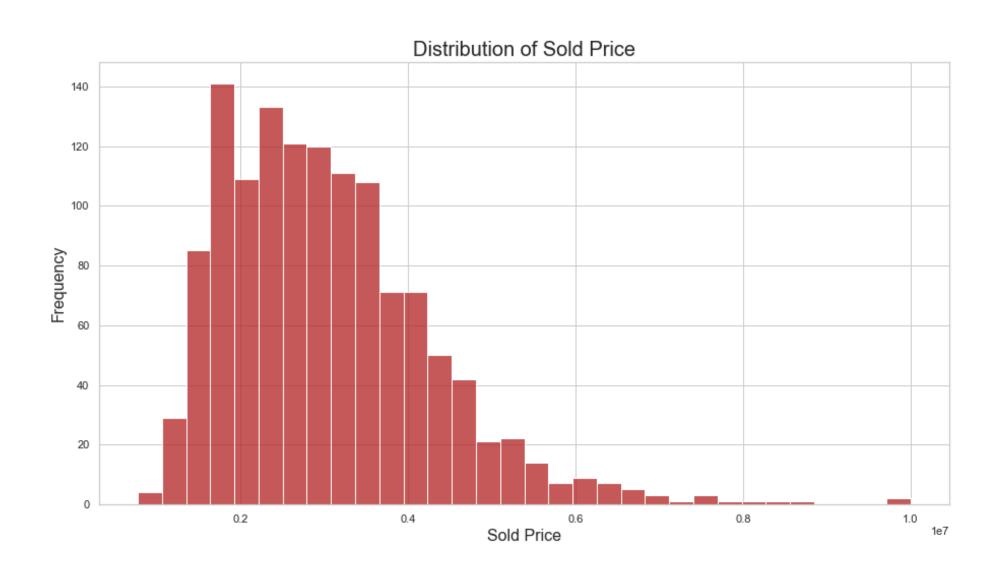


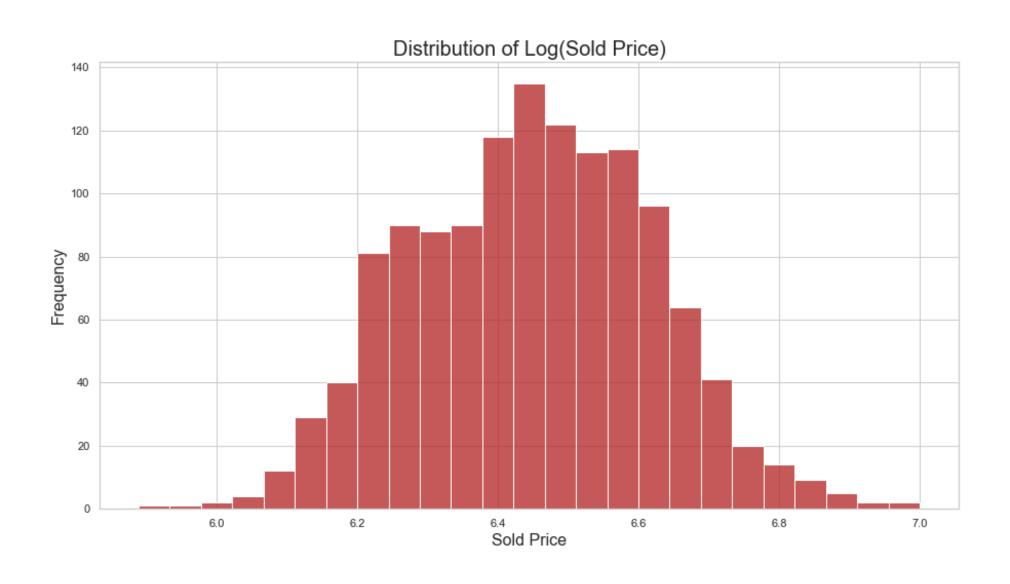
sold price	- 1	05	0.75	0.64	0.029	0.053	0.032	01	0.26	0.4	016	035	025	0.072	0.3	00075	015	012	019	025	022	031	028	0.095	0.11	0.037	013	042	0.016	043	0.45	-0.24
lot size	05	1	04	034	0.018		0.061				0.099	018					0.053		0.095		0.064		0.042							0033		-0.14
home size	0.78	04	1	035	0.031	00041		0.099	0.44	0.51				0.089	0.45	0.048													0.043			-0.16
school score avg	0.64	034	035	1	0.05		0.052						0.085	-0.099		0.049		0.061	0.061		0.094	0.06	015								-0.52	-0.26
laundry -	0.029	0.018	0031	0.05	1	021		0.062				0.033			0.0036				0.00037		0.049		0.035				0044	0.016		0.085	0.052	-0.11
heating .	0.053		00041		021	1						0.0086		0.00086					0.08												0.042	0.046
air conditioning	0.032	0061			0.093	051	1					0.096	0.046				0.062							0.089		03			0.0064			0.034
pool	01		0.099		0.062	0.0015	0.0091	1		0.055	0.048					0.036			0.0099					00043		006	0042	0.00062				0.11
age of house	0.26		0.44					0.054	1					0.083			0.086		0.0082	0.32	0.068		-0.42			0.093	-0.097					0.022
beds_3.0	-0.4		0.51						018	1	0.66	0.34				0.063					-0.09T		-0.33	-0.048		0.005			00056			012
beds_4.0	016	0.099	02					0.048	0.13	0.66	1	0.25		0.062	-0.14				0064	1.5e-05		0.053			0.09				0044	0.041	0.063	0.0084
beds_5.0	035		0.46			0,00086	0.096			0.34	0.25	1			0.31	0.051						0063			0.00068	0.032			0.086	0.063		0.1
beds_6+	025			0.083			0.046					0.057	1	0.019		0.056			0.069				0.099		0.0046			0.064				0.059
baths_1.5	0.072		-0.089	0.099		0.0086			0.085		-0.082	0.049	0.019	1		0.048	0.061					0.014										0.041
baths_2.0	0.3		0.45		0.0036									0.12	1	0.34	0.43				0.098		-0.43			0056			0.043	0.043	0.087	012
baths_2.5	00075		0.048	0.049					0.19	0.063			0.056	0.048	0.34	1			0.062	0.065	0.041	0.042		0.00056				0044			0.064	0.028
baths_3.0	015						0.062		0.086	0.27				0.061	0.43	0.15	1	-0.1		0.082					-0.048			0.085	0.028			0.05T
baths_3.5	012															0.08	0.1	1	0.04T					0.099		0042	0.038				0.046	52e-05
baths_4.0	019	0.095			000037	0.08		0.0099	0.0062		0.064		0.069			0.062		-0.047	1	0.038			0.088		0.052					0.095	0.042	-0.061
baths_4.5	025								0.32		1.5e-05					0.065	0.062		0.036	1										0.096	0.056	0.068
baths_5.0	022	0.064		0.094	0.049		0.05		0.068						0.095	0.041	0.052			0.019	1								0.042	0041	0.056	0.069
bwths_6+	031											0.063		0.014							0.012	1	0.065						0.044		0.058	0.055
floors_2.0	028	0.042	0.47						0.42	0.33			0.099		0.43				0.088			0.065	1	0.086						0.099		0.063
floors_3.0	0.095		023				0.089	00043		0.048		00075				0.0086	0012	0.099				027	0.086	1	-0.019	0.046	0.043			0.038		0.071
garage spaces_1	0.11											0.00068	0.0046				0.048		0.052	0.081				0.019	1	0.52	0.069	0.091	0.0052			0.003
garage spaces_2	0.037					055			0.093	0.008								0042		0.0018				0046	0.52	1		0.0043	0.042			0.097
garage spaces_3+	013				0.044	0.073		0.042								0.0013		0.038						0043	0.069	-0.2	1	0.085		0.065		0.037
city_LOS ALTOS	0.42				0.016			0.00062					0.064			0044	0.085								-0.091	0.0043	0.085	1	0.12			-0.2
city_MOUNTAIN VIEW	0.016	-0.1	0.043	0.035			0.0064		0.078			0.086							0.052		0.042	0.044				0.042		-0.12	1	0.19		0.25
city_PALO ALTO	043	0055			0.085				0.009		0.041	0.065			0.043					0.096			0.099				0.065		-0.19	1	-0.25	0.3
city_SANTA CLARA	0.45	0.17	0.23	0.52	0052	0042					0.083	0.06T	0.078		0.087	0.064		0.046	0.042	0.056	0.056	0.058								0.25	1	-0.33
city_SUNNYVALE	0.24			0.26		0.046					0.0084		0.059	0041				52e-05	0.081	0.068	0.069		0.063								0.33	1
	8	8	8	5	-	- 0		-	9	0	0	0	+	1	-	1	0	1 10	- 0	1 10	-	+	- 0		· .	P1.	+	- 5	*	9	RA -	-
	E E	in si	9	8	þin	n se	ij	Ē.	E E	41,3.0	4	4	4	Ę.	2	27	ř,	ri,	4	4	E.	ė	ri,	ri,	80	80	, g	ALTO	VIEW	O ALTO	GAR	786
	8	=	Ę	ii ii		E	E E		ŏ	ă	ă	ă	ă	Ē	Ē	Ē	Ē	Ē	Ē	Ē	Ē	Ē	g g	g g	ě	ğ.	old.	Ry_LOS.	TANK.	98.	Ĕ	SUMP.
				8			1		E .																20	Ē	8	ALC:	MOUNTAIN	C By	SAMTA	20
				×			-																		В	8	8	~	, N	~	and the second	T
																													ū			

- 0.8

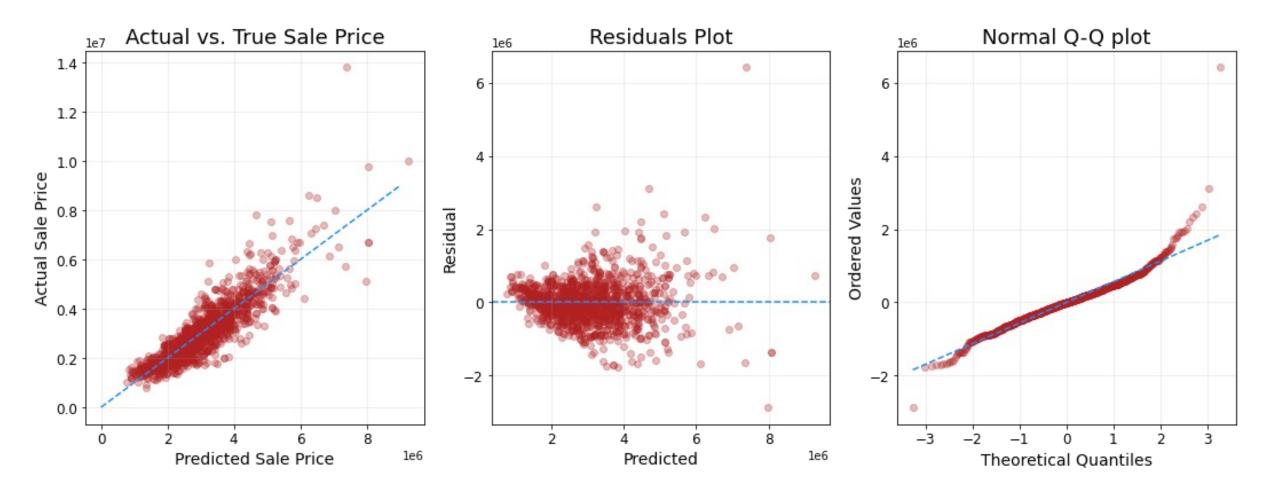
- 0.6

- 0.4



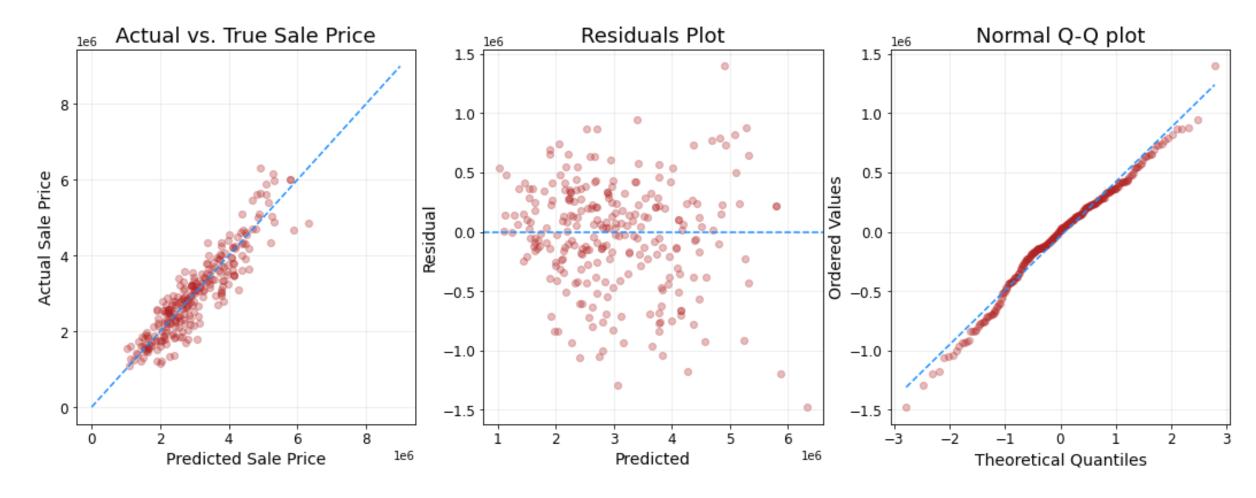






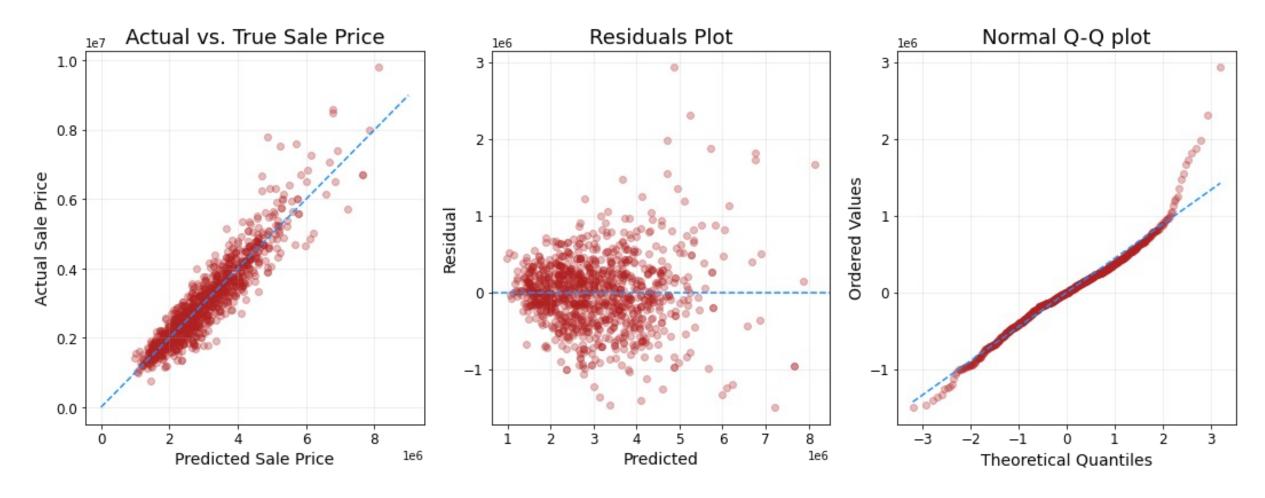
* Baseline Linear Regression – Untouched Dataset

R



* Elastic Net Regression – Validation Dataset





* Elastic Net Regression – Test Dataset

lot size : 130766.26 home size: 773471.44 school score avg: 269845.16 laundry : 3076.32 heating: -39490.85 air conditioning: -18532.17 pool: 15432.62 age of house: 0.00 beds 3.0 : 21291.99 beds 4.0 : 0.00 beds 5.0 : -31173.91 beds 6+: -71899.62 baths 1.5 : -0.00 baths 2.0 : 0.00 baths 2.5 : -28218.86 baths 3.0 : -0.00baths 3.5 : 0.00 baths 4.0 : -43796.74 baths 4.5 : 20622.92 baths 5.0 : 37970.59 baths 6+ : 52139.99 floors 2.0 : -36938.23 floors 3.0 : -96698.00 garage spaces 1 : 17165.13 garage spaces 2: 42698.15 garage spaces 3+: -37222.90 city LOS ALTOS : 205731.06 city MOUNTAIN VIEW: 110700.70 city PALO ALTO: 336679.56 city SANTA CLARA: -73019.63 city SUNNYVALE: 29654.19