INF1340 – Group 47 Final Project Results

GitHub Repository: github.com/Jerrryyy/final

Kai Xie, Nargiz Guliyeva, Jerry Li

Wednesday, December 12th, 2023

Course: Programming for Data Science
Graduate Unit: Faculty of Information
Professor: Dr. Maher Elshakankiri
Due Date: December 13th, 2023

Script (analyze_housing.py) Terminal Output

```
Jerry@Jerry-PC MINGW64 ~/GitHub/inf1340/final (main)
$ py analyze_housing.py data/housing.csv

Analyzing the housing market data of New York City,
Seattle, and Miami, United States, from 2017-2023.
Data source: Redfin (redfin.com/news/data-center/)...
-------------------------------------------------------------
Saving descriptive analytics to /results/...

A total of 2430 days of data entries were analyzed.
Please note that variance.csv should be the same for all cities.
-------------------------------------------------------------
Saving diagnostic analytics to /results/...
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Linear regression summary of New York City:

Out of 6 predictors...
Significant positive coefficients:
1. Total Homes Sold With Price Drops
2. Median New Listing Price

Significant negative coefficients:
1. Inventory
2. Age of Inventory
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
After fitting a generalized linear model
to the same data, the GLM had a higher
pseudo-R-squared value of 0.98.
-------------------------------------------------------------
Saving predictive analytics to /results/...
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Support vector regression summary of New York City:

Model Results:
Avg CV Score: 0.92
RMSE: 20494.31
MAE: 15911.25
R-squared: 0.91

SVR model saved to results/nyc/predictive/svr_nyc.pickle
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Support vector regression summary of all three cities:

Model Results:
Avg CV Score: 0.97
RMSE: 20582.95
MAE: 16069.54
R-squared: 0.98

Saving model fit performance plots...
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Predicting median sales prices for new entries
after August 2023 in New York City:

Prediction Accuracy:
Mean of Residuals: 963.52
Residual Standard Error: 5262.78
Correlation: 0.59
p-value: 0.05
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Lasso and Ridge regression of median sales price in Seattle:

After fitting both models, the Lasso
model had a higher test score of 0.93.
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Random forest regression of median sales price in Miami:

Random Forest Regression Performance:
RMSE: 11777.97
R-squared: 0.98
-------------------------------------------------------------
Saving prescriptive analytics to /results/...
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Support vector regression of New York City used the following:

18 features, of which only 9 were positively-weighted
in terms of predicting median sales price:
1. Total Homes Sold: 4.19
2. Total Homes Sold With Price Drops: 6.70
3. Average Sale to List Ratio: 10.16
4. Percent Homes Sold With Price Drops: 6.17
5. Median New Listing Price: 7.19
6. Median Days to Close: 11.49
7. Pending Sales: 0.38
8. Percent Homes Sold Above List: 11.01
9. Price Drops: 3.45
-------------------------------------------------------------
Analysis complete. Results saved to /results/.
Thank you for using our program!
- Jerry, Kai, Nargiz, 2023
```

All Cities

Files

- anova.csv
- heatmaps.svg
- post_hoc.csv
- svr.svg
- svr_all.pickle
- top10.png

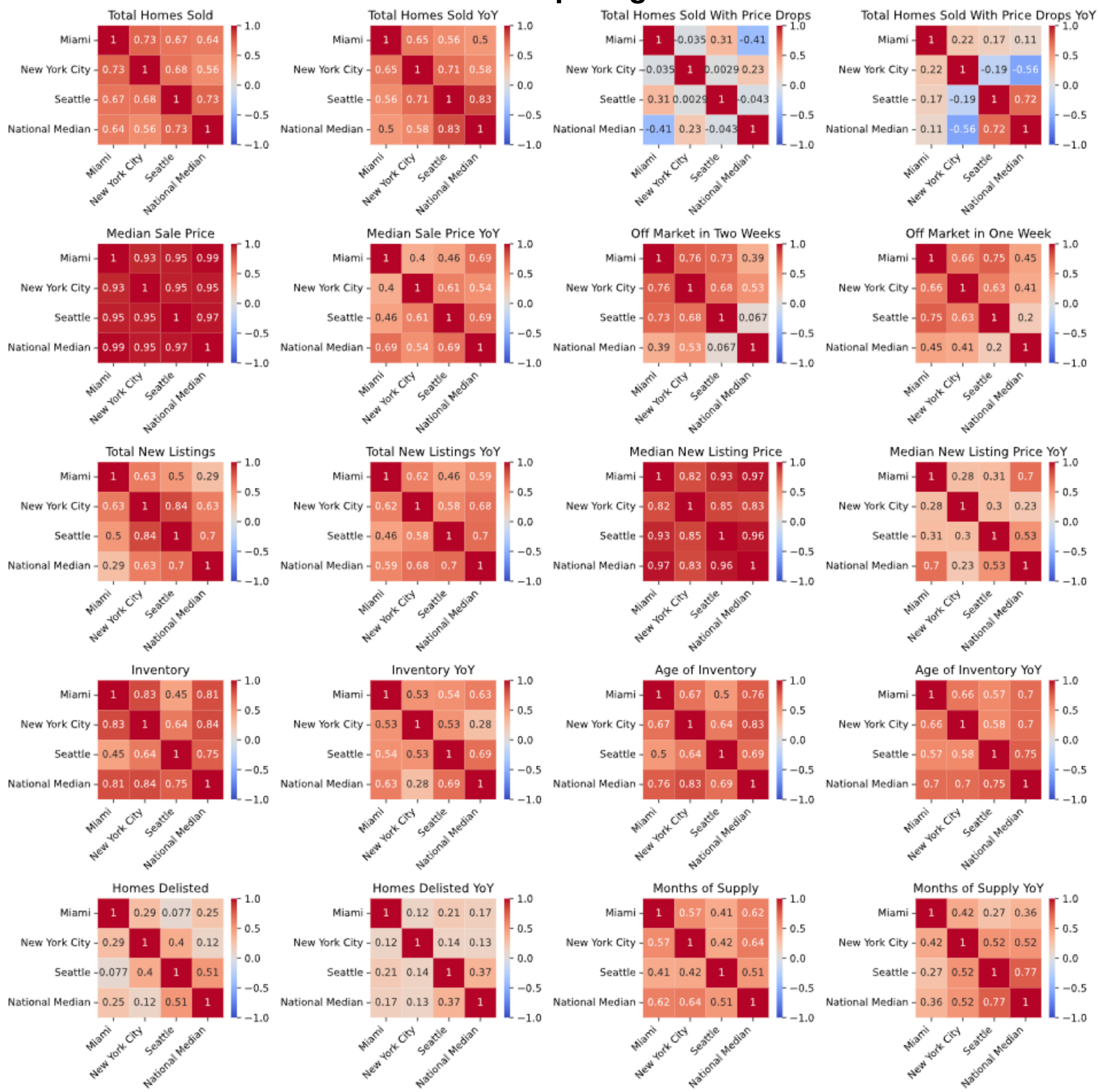
anova.csv

	Statistic	p-value	Result	Post-Hoc
median_sale_price	671.3732	1.63E-146	***	Required
total_homes_sold_with_price_drops	517.7933	3.65E-113	***	Required
median_new_listing_price	632.0382	5.68E-138	***	Required
inventory	927.1121	4.79E-202	***	Required
age_of_inventory	621.1468	1.32E-135	***	Required
months_of_supply	693.3413	2.77E-151	***	Required

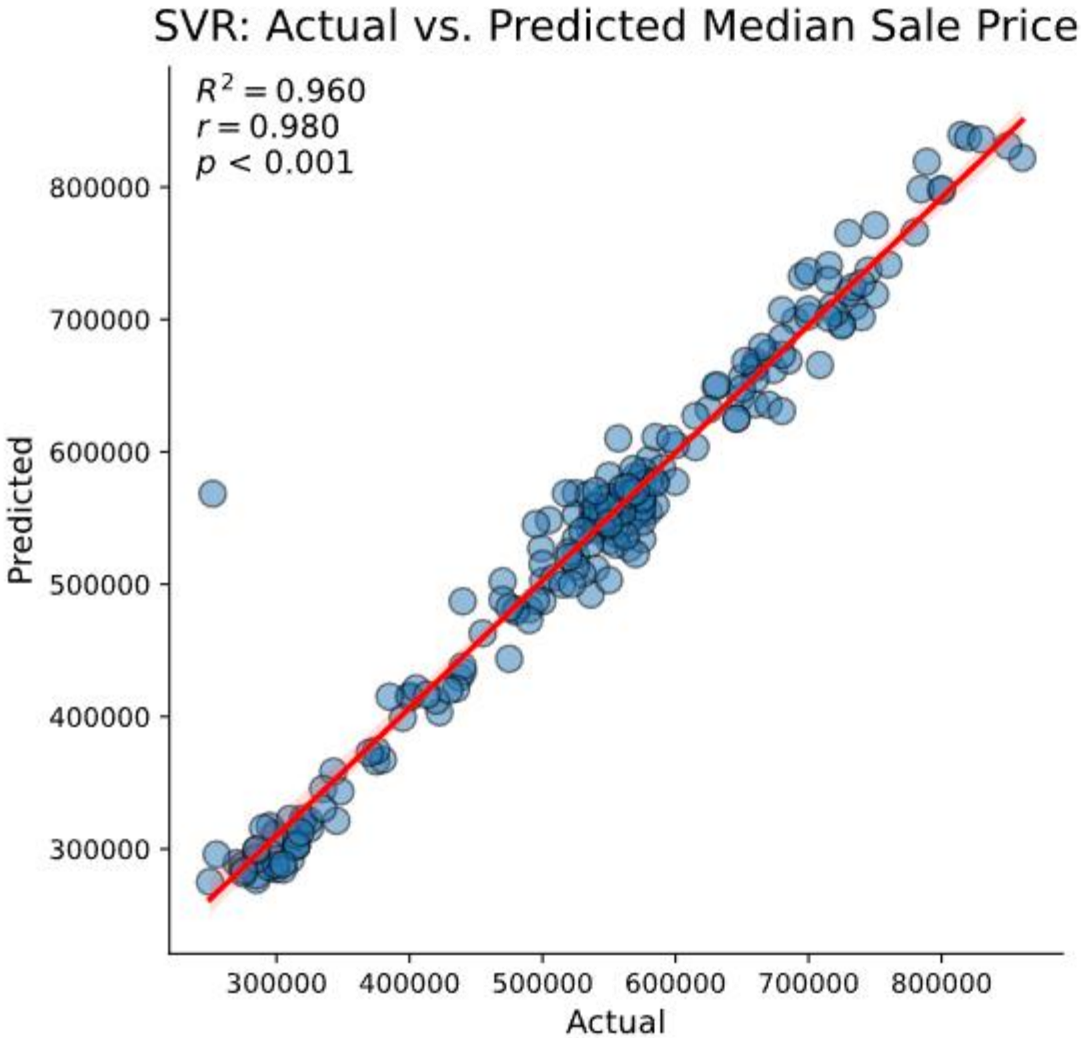
post_hoc.csv

median_sale_price	NYC	Seattle	Miami
NYC	1	0.000457	9.09E-112
Seattle	0.000457	1	6.76E-109
Miami	9.09E-112	6.76E-109	1
total_homes_sold_with_price_drops	NYC	Seattle	Miami
NYC	1	1.09E-84	2.90E-82
Seattle	1.09E-84	1	5.43E-09
Miami	2.90E-82	5.43E-09	1
median_new_listing_price	NYC	Seattle	Miami
NYC	1	0.002485	1.77E-112
Seattle	0.002485	1	4.56E-95
Miami	1.77E-112	4.56E-95	1
inventory	NYC	Seattle	Miami
NYC	1	2.27E-115	2.27E-115
Seattle	2.27E-115	1	2.27E-115
Miami	2.27E-115	2.27E-115	1
age_of_inventory	NYC	Seattle	Miami
NYC	1	2.07E-101	0.000319
Seattle	2.07E-101	1	1.56E-102
Miami	0.000319	1.56E-102	1
months_of_supply	NYC	Seattle	Miami
NYC	1	8.17E-115	0.029943
Seattle	8.17E-115	1	5.63E-114
Miami	0.029943	5.63E-114	1

heatmaps.svg



svr.svg



top10.png

	Region	Date	Median Sale Price ($)
3005375	Chase County, KS	2020-08-03	111111111.0
4041131	Nantucket County, MA	2023-06-26	38127000.0
4040283	Teton County, WY	2023-01-23	29450000.0
3204697	Fillmore County, NE	2023-08-28	26449214.2
2979303	Lake County, MT	2023-01-16	25000000.0
546239	Jefferson County, MT	2018-01-08	24500000.0
3509098	Treasure County, MT	2023-08-07	20000000.0
2164849	Meagher County, MT	2022-10-03	20000000.0
954029	Pitkin County, CO	2020-06-08	14650000.0
3125157	Rosebud County, MT	2022-02-14	13995000.0

New York City

Directories

- descriptive
- diagnostic
- predictive
- prescriptive

Descriptive Analytics

- histograms.svg
- monthly_means.csv
- normality.csv
- qqplots.svg
- summary.csv
- variance.csv
- violinplots.svg

monthly_means.csv

date	age_of_in	total_new	total_acti	inventory	total_hom	total_hom	average_s	percent_h	homes_de	median_n	median_d	median_d	median_s	months_o	pending_s	percent_c	percent_h	price_drops
1/31/2017	124.2	1566.6	34607.6	33338.8	1137.4	212.4	0.974382	0.187555	380	584253.2	110.8	39.8	499250	30.94002	373.6	0.16266	0.133746	771.4
2/28/2017	119.75	1904.25	35352.5	34026.75	1083.5	197.25	0.973517	0.181263	326.25	589475	114.5	38.25	497687.5	33.73491	475	0.19622	0.133334	747.5
3/31/2017	99.25	2112	36760.25	35345.75	1127.5	200	0.972735	0.178199	322.5	599250	114.125	35	497125	33.3288	621.5	0.201062	0.136122	783
4/30/2017	80.75	2219.25	37980.25	36503.75	1170.25	199.75	0.974378	0.17151	316.5	610500	105	33.625	506562.5	33.28818	721.75	0.167304	0.133256	695.75
5/31/2017	75.2	2285.2	39771.4	38075	1349.4	190.8	0.979627	0.141671	356.4	589199.8	96.4	29.4	523950	29.55953	832.6	0.148031	0.161582	813.2
6/30/2017	68.5	2696.5	40674.75	38086.75	1656	234.25	0.980686	0.142617	757.25	628500	90.75	28.625	556375	25.46337	995	0.217091	0.146724	970.75
7/31/2017	76.2	1820.4	39540.4	37653.6	1531.8	226.2	0.984404	0.147501	468.6	544789.8	82.6	38.1	551000	27.20767	825.2	0.126172	0.170296	906.4
8/31/2017	83.25	1561	37767.75	35890.5	1638	392.75	0.982878	0.23871	533.25	540500	81.75	38.75	545250	23.1757	731.75	0.141468	0.160751	791.25
9/30/2017	92.25	2024	37283.75	35566	1428	374.5	0.98151	0.261503	513.5	623850.3	84.75	40.75	521250	26.74214	626.25	0.14412	0.155031	911
10/31/2017	92.8	1744.2	37096	35399.2	1380.8	364.6	0.978921	0.26364	470.8	600900	84.4	44.8	512390	27.16417	653.8	0.138806	0.148142	781.4
11/30/2017	94	1263.5	35490.25	33865.75	1288.75	311	0.97894	0.240757	548.25	578000	87.25	45	528562.5	28.63935	559.75	0.126109	0.16593	696.25
12/31/2017	102	845	32834	31362.25	1421	328.5	0.977188	0.232843	466.25	523700	90.25	40.875	528537.4	23.68091	481	0.101173	0.172107	488
1/31/2018	112.5	1590.4	31619	30294.4	1112.2	244.8	0.976167	0.220027	430.4	614380	96	48.4	537610.2	29.93535	401.8	0.184113	0.170845	535
2/28/2018	106.5	2038.25	33172.75	31784.25	1040.75	226.5	0.974056	0.217538	372.25	624486	101	42.875	529999.9	32.39215	509.75	0.205162	0.150656	638
3/31/2018	84.5	2001.25	35240.75	33781.5	1110.25	252.5	0.975724	0.228265	348.5	621999.3	106.5	40.25	536767.3	32.18773	627.25	0.194299	0.168466	659
4/30/2018	72	2637.4	37897.6	36184.2	1196.8	248.4	0.978606	0.208754	428.4	636688.4	96	38.6	534606.2	31.98403	799.8	0.165873	0.190507	901
5/31/2018	69.75	2396.75	40847.25	39142.5	1269.75	228.75	0.980313	0.180527	389	614249.8	87.125	38	555676.4	32.33008	835	0.162984	0.184118	980.75
6/30/2018	69.25	2259.75	41332	39327.25	1559.75	285.5	0.980407	0.183232	532	598586.3	79.25	42.5	562812.5	27.34364	914.5	0.138704	0.202763	1208.75
7/31/2018	78.6	1887.8	40681.2	38870.2	1546.8	298.8	0.979726	0.191971	486.6	578980	77.7	40.4	562354	27.35113	759.4	0.116367	0.196893	1112.8
8/31/2018	88.25	1582.75	39678.25	37923.75	1533.75	296.75	0.979485	0.192781	509.75	566972	77	43	554181.3	26.05694	672	0.123753	0.194941	974.75
9/30/2018	92.25	2160	40007.25	38282.75	1361	284.75	0.977922	0.208143	534.75	643624.9	82.5	41.5	547500	29.7503	578.5	0.12856	0.17851	1208
10/31/2018	89.4	1932	40303.4	38559.4	1326.2	307.8	0.976858	0.231456	585.2	631299.9	85.6	43.6	534000	30.64955	596.2	0.123946	0.16344	1217.6
11/30/2018	89	1361	38976.25	37302.5	1183.25	292.5	0.975104	0.246035	650.5	580487.4	88	43.5	543737.5	33.95394	514.25	0.104727	0.156188	928.75
12/31/2018	101.8	997.6	36028.2	34528.6	1202.4	323.2	0.972233	0.270685	577.6	565580	92.7	42	533200	32.73777	415.2	0.091539	0.152291	600.6
1/31/2019	108.25	1810	35077.75	33405.5	1137.75	304.75	0.973793	0.267757	660.5	637874.9	97.25	47.125	531066.3	31.25214	424.5	0.158662	0.145667	799.75
2/28/2019	106.25	1989.5	35635.5	34100.75	1034	288	0.972377	0.278533	461	630725	108.5	41.75	528750	35.3236	511	0.17672	0.141412	772.75
3/31/2019	85.25	2310.75	37324.75	35657	1133.5	297	0.970049	0.261151	470.75	643658.9	107.875	39.5	523218.9	33.3566	633.5	0.175857	0.134395	830.25
4/30/2019	71.8	2603.2	40099.2	38378	1168.4	282.4	0.973926	0.242225	455.4	638999	103	40	553800	34.59966	769.2	0.150793	0.151748	938
5/31/2019	69.75	2358.5	42560.25	40788.25	1290.75	298.5	0.975932	0.23122	490.25	629737.5	93.5	43.25	554375	33.09941	776.75	0.135471	0.151769	1107
6/30/2019	75.25	2166.75	43440.25	41420.5	1603	342.5	0.977878	0.215512	562	607250	87.625	42.125	589212	28.03492	856	0.123296	0.150876	1227.25
7/31/2019	84	1769.2	41998.4	40038.6	1453.4	343.2	0.977713	0.235239	625	591177.6	79.9	45	556207.4	29.84981	733.2	0.099485	0.173255	1039.8
8/31/2019	94	1533.75	39978	38052	1506.5	373.25	0.976168	0.247303	632.5	579690	81.5	47	574375	26.78613	660	0.110006	0.163269	884.5
9/30/2019	93	2213.2	39391.2	37490.6	1334.2	367.4	0.974049	0.275151	765.6	669360	88.9	46.6	548954	29.90736	563.4	0.130181	0.149896	1123.4
10/31/2019	89	1669	38925	37209.25	1288.75	357.75	0.972752	0.278425	569.5	622322	90.75	50.5	549750	30.84505	590.75	0.12459	0.146225	979.25
11/30/2019	88.5	1281.75	37352.25	35724.25	1179.5	339	0.973659	0.288432	575.75	612737.5	94	46.25	555500	32.0674	544.75	0.12384	0.152445	741.5
12/31/2019	102.2	888	34214.6	32669.6	1235.4	362.4	0.972874	0.294756	639.4	594979.8	94.5	48.2	545250	29.32997	411.6	0.082741	0.151275	451.8

summary.csv

	min	mean	median	max	range	stdev	sem
age_of_inventory	52	83.71983	81	125	73	17.352	0.930165
total_new_listings	429	1840.247	1849.5	3682	3253	561.3947	30.09391
total_active_listings	22405	33692.27	34193	43657	21252	5772.505	309.4388
inventory	21039	31983.61	32675	41630	20591	5683.03	304.6424
total_homes_sold	568	1338.066	1344.5	2329	1761	318.6045	17.07899
total_homes_sold_with_price_drops	137	365.5517	362	694	557	124.9609	6.698609
average_sale_to_list_ratio	0.967444	0.983717	0.981475	1.011693	0.044249	0.009944	0.000533
percent_homes_sold_with_price_drops	0.13144	0.273109	0.273141	0.444302	0.312862	0.066303	0.003554
homes_delisted	248	491.6437	466.5	2244	1996	179.0412	9.597613
median_new_listing_price	456000	644264.4	639200	799000	343000	57912.76	3104.45
median_days_on_market	39	79.31322	81.25	121.5	82.5	19.48674	1.044599
median_days_to_close	22	52.96408	50.5	80	58	12.69276	0.680403
median_sale_price	471000	596461.5	574500	740000	269000	66812.68	3581.535
months_of_supply	13.07548	26.57202	26.23219	56.08627	43.01079	7.556313	0.405061
pending_sales	242	743.2213	713.5	2033	1791	260.597	13.96947
percent_off_market_in_two_weeks	0.049618	0.176431	0.167883	0.450517	0.400899	0.062938	0.003374
percent_homes_sold_above_list	0.117068	0.202909	0.176494	0.391462	0.274394	0.064428	0.003454
price_drops	171	889.4971	883	1941	1770	275.9657	14.79331

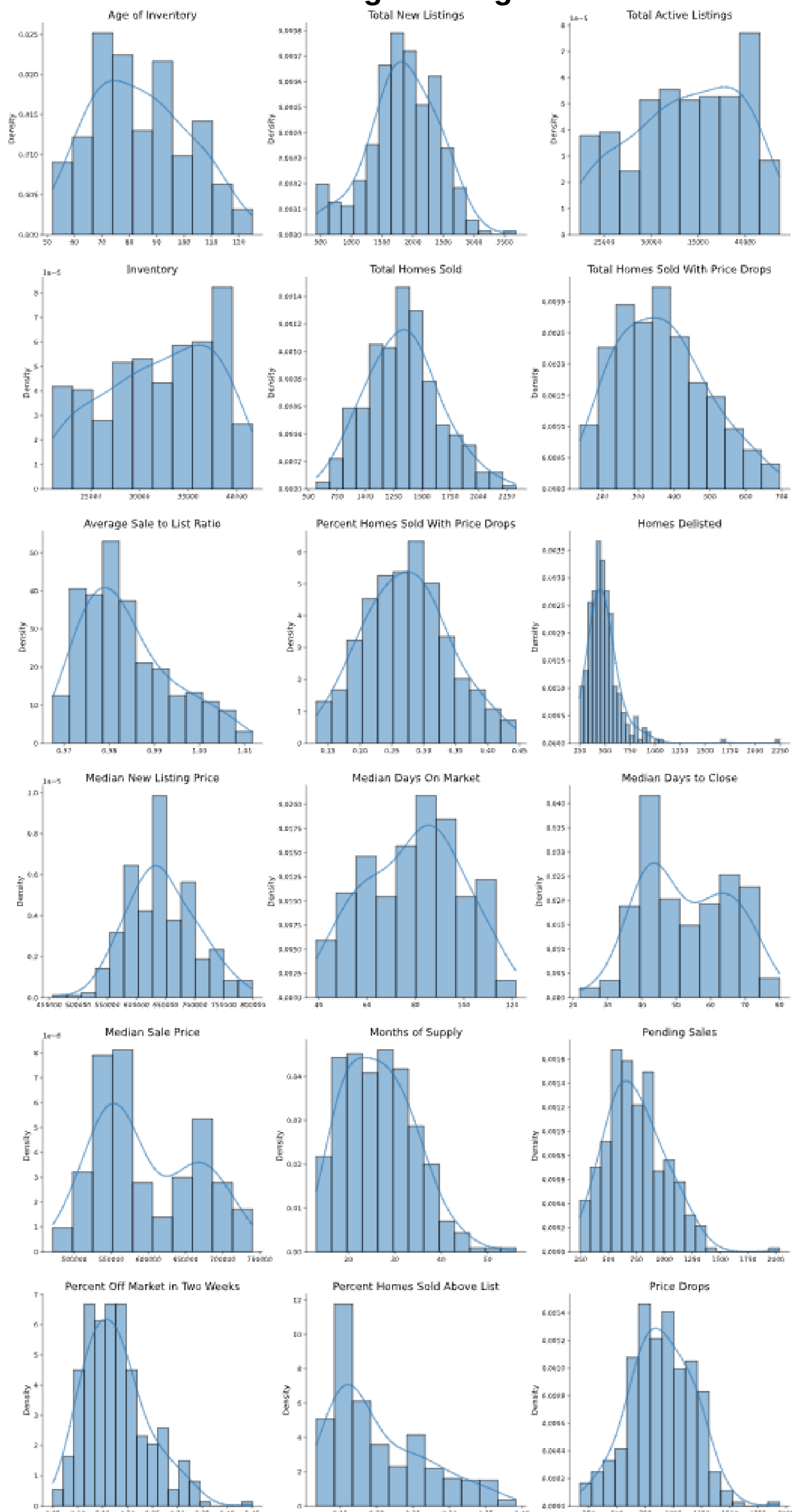
normality.csv

	Statistic	p-value	Skewness	Kurtosis	Result
age_of_inventory	0.972993	4.32E-06	0.277045	-0.7795	Non-normal
total_new_listings	0.984992	0.001108	-0.29463	0.144803	Non-normal
total_active_listings	0.952395	3.55E-09	-0.29165	-1.03249	Non-normal
inventory	0.94855	1.17E-09	-0.29943	-1.07347	Non-normal
total_homes_sold	0.991659	0.047357	0.300829	-0.04548	Non-normal
total_homes_sold_with_price_drops	0.973332	4.97E-06	0.430903	-0.48466	Non-normal
average_sale_to_list_ratio	0.936401	4.79E-11	0.772785	-0.19475	Non-normal
percent_homes_sold_with_price_drops	0.991565	0.044801	0.179026	-0.35838	Non-normal
homes_delisted	0.730406	2.35E-23	4.05096	31.13879	Non-normal
median_new_listing_price	0.989733	0.015301	0.189208	-0.12963	Non-normal
median_days_on_market	0.978565	4.76E-05	-0.09795	-0.84445	Non-normal
median_days_to_close	0.956162	1.11E-08	0.122611	-1.07424	Non-normal
median_sale_price	0.93309	2.15E-11	0.377158	-1.10382	Non-normal
months_of_supply	0.971033	1.97E-06	0.548045	0.064392	Non-normal
pending_sales	0.973431	5.17E-06	0.619411	0.926705	Non-normal
percent_off_market_in_two_weeks	0.959748	3.49E-08	0.79429	0.684637	Non-normal
percent_homes_sold_above_list	0.888352	3.01E-15	0.930449	-0.15566	Non-normal
price_drops	0.992997	0.104268	-0.03022	0.084372	Normal

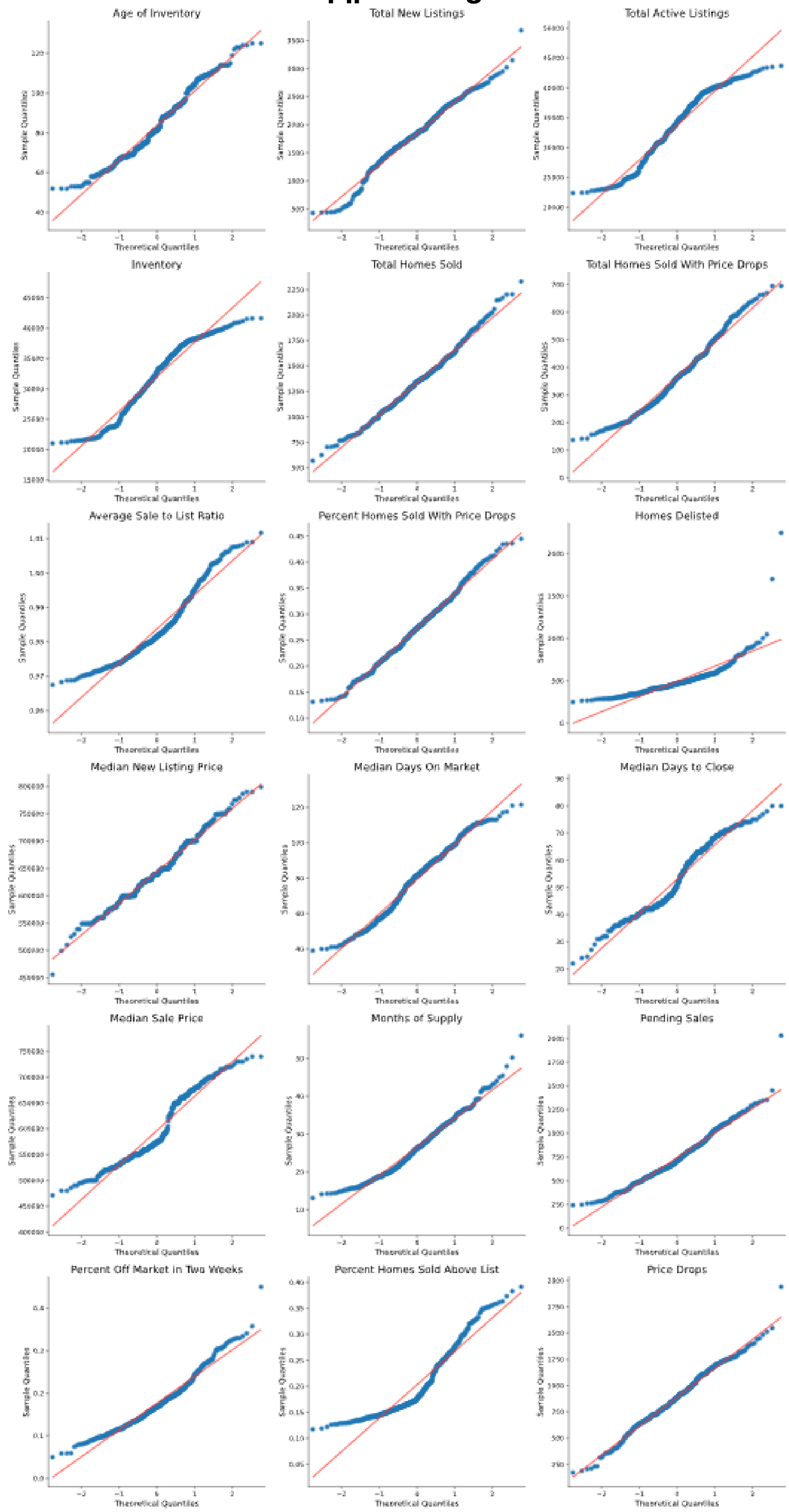
variance.csv

	Statistic	p-value	Result
age_of_inventory	1.461868	0.232278	Equal
total_new_listings	148.1618	2.38E-57	Unequal
total_active_listings	179.9023	7.83E-68	Unequal
inventory	181.5625	2.28E-68	Unequal
total_homes_sold	58.56987	7.86E-25	Unequal
total_homes_sold_with_price_drops	116.6218	1.99E-46	Unequal
average_sale_to_list_ratio	145.4323	2.00E-56	Unequal
percent_homes_sold_with_price_drops	24.96083	2.58E-11	Unequal
homes_delisted	57.70375	1.71E-24	Unequal
median_new_listing_price	51.66224	4.06E-22	Unequal
median_days_on_market	119.9248	1.35E-47	Unequal
median_days_to_close	399.2628	2.01E-129	Unequal
median_sale_price	41.25282	5.74E-18	Unequal
months_of_supply	77.38115	4.67E-32	Unequal
pending_sales	110.543	2.93E-44	Unequal
percent_off_market_in_two_weeks	186.4017	6.39E-70	Unequal
percent_homes_sold_above_list	193.774	2.89E-72	Unequal
price_drops	22.49501	2.73E-10	Unequal

histograms.svg

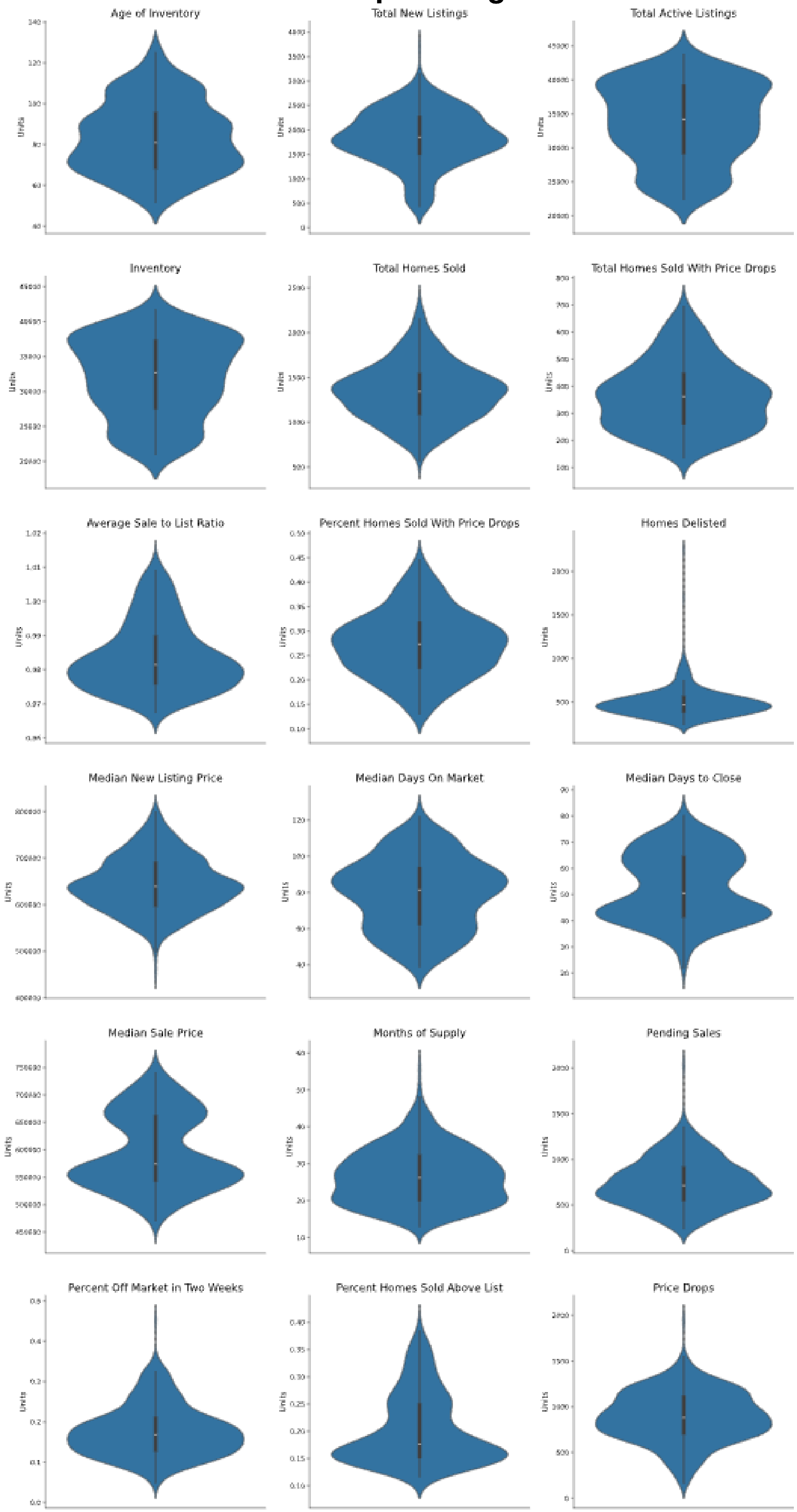


qqplots.svg



Age of Inventory
Total New Listings
Total Active Listings
Inventory
Total Homes Sold
Total Homes Sold With Price Drops
Average Sale to List Ratio
Percent Homes Sold With Price Drops
Homes Delisted
Median New Listing Price
Median Days On Market
Median Days to Close
Median Sale Price
Months of Supply
Pending Sales
Percent Off Market in Two Weeks
Percent Homes Sold Above List
Price Drops
Theoretical Quantiles
Sample Quantiles

violinplots.svg



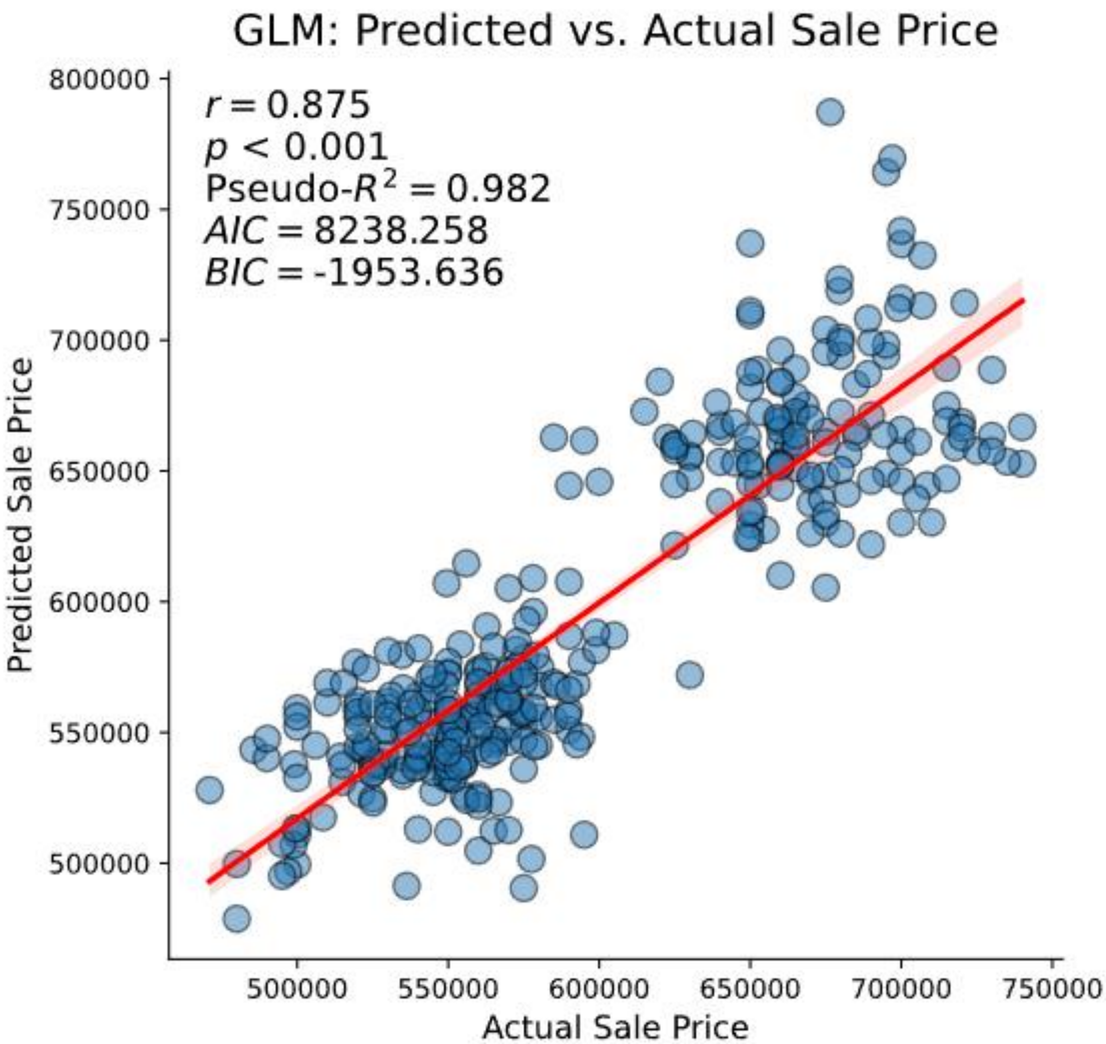
Diagnostic Analytics

glm.svg

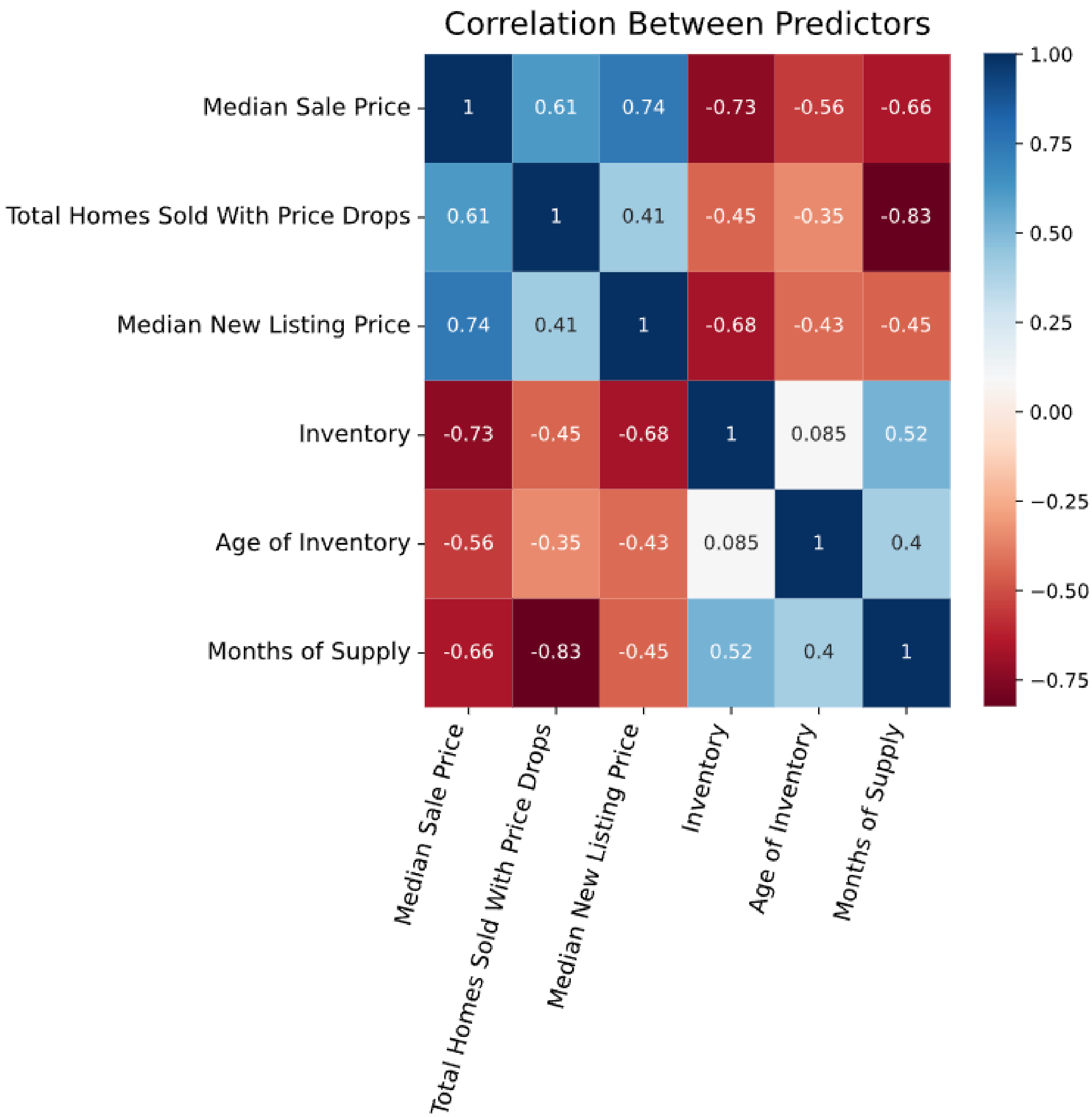
multicollinearity.svg

svr_nyc.pickle

glm.svg



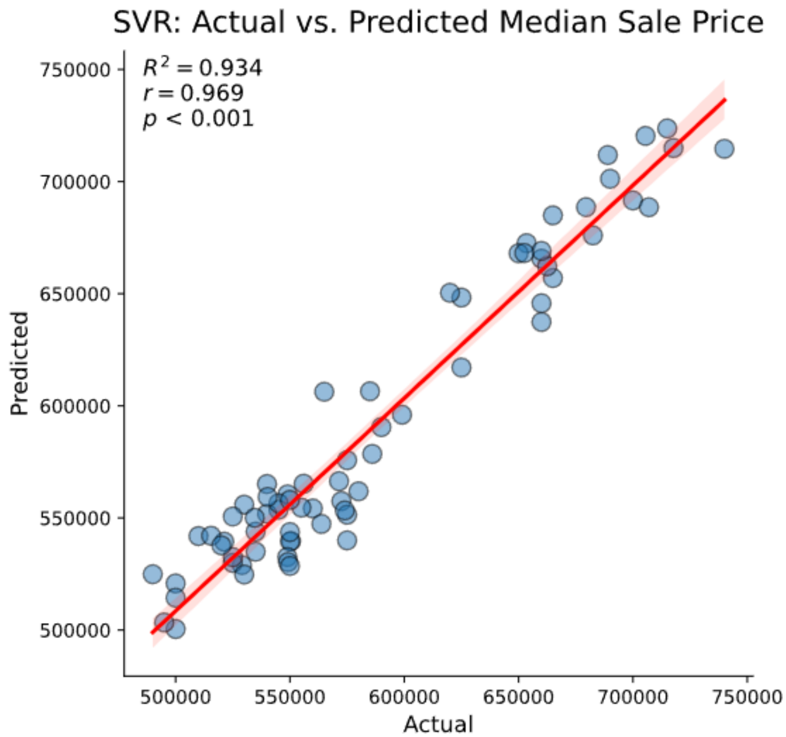
multicollinearity.svg



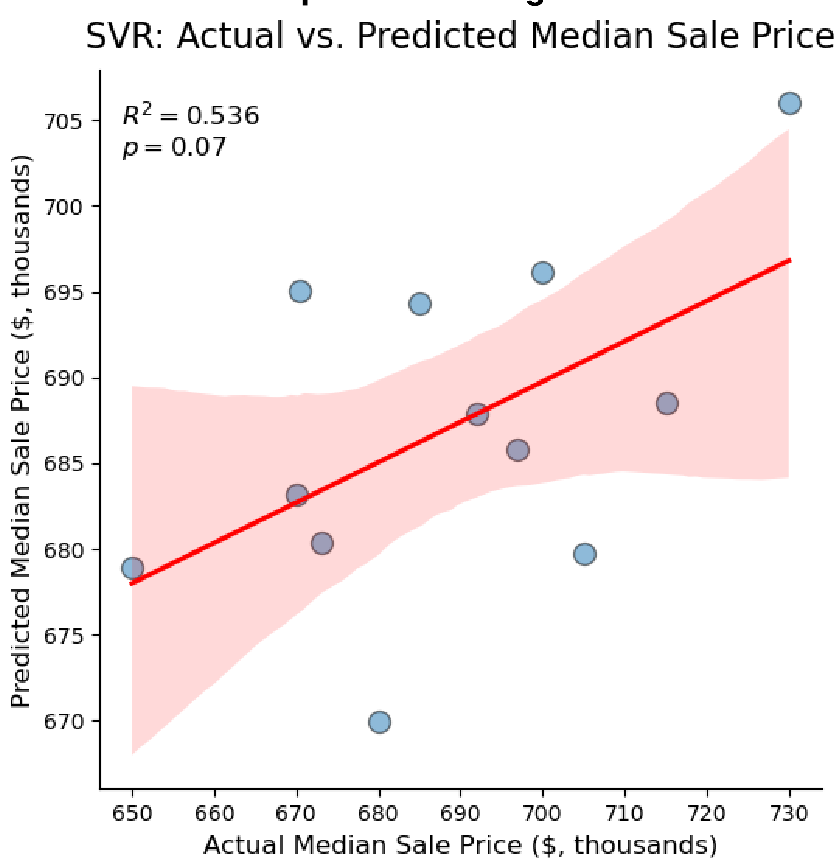
Predictive Analytics

- predictions.svg
- svr.svg

svr.svg



predictions.svg



Prescriptive Analytics

coefficients.svg

coefficients.png

	Coefficient
Date	0.868472
Age of Inventory	-10.498677
Total New Listings	-0.017170
Total Active Listings	-6.532139
Inventory	-6.392241
Total Homes Sold	4.863372
Total Homes Sold With Price Drops	6.748163
Average Sale to List Ratio	9.715982
Percent Homes Sold With Price Drops	5.356549
Homes Delisted	-4.830068
Median New Listing Price	9.576695
Median Days On Market	-12.120269
Median Days to Close	11.768876
Months of Supply	-8.259239
Pending Sales	3.028283
Percent Off Market in Two Weeks	0.948238
Percent Homes Sold Above List	11.679005
Price Drops	4.070729

Seattle

Directories

- descriptive
- diagnostic
- predictive
- prescriptive

Descriptive Analytics

- longitudinal.svg
- monthly_means.csv
- normality.csv
- summary.csv
- variance.csv

Formatting of .csv files match corresponding ones for New York City.

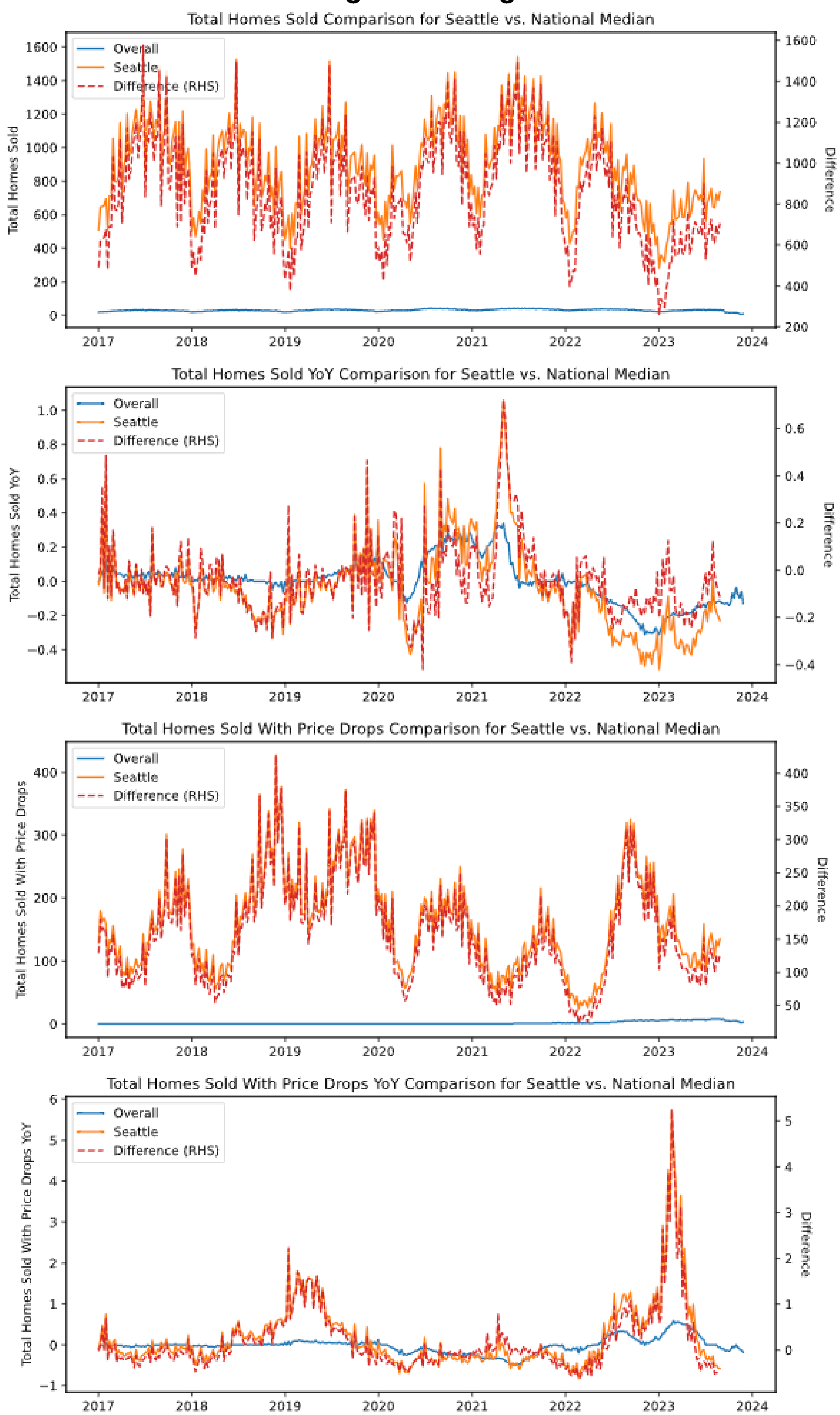
monthly_means.csv

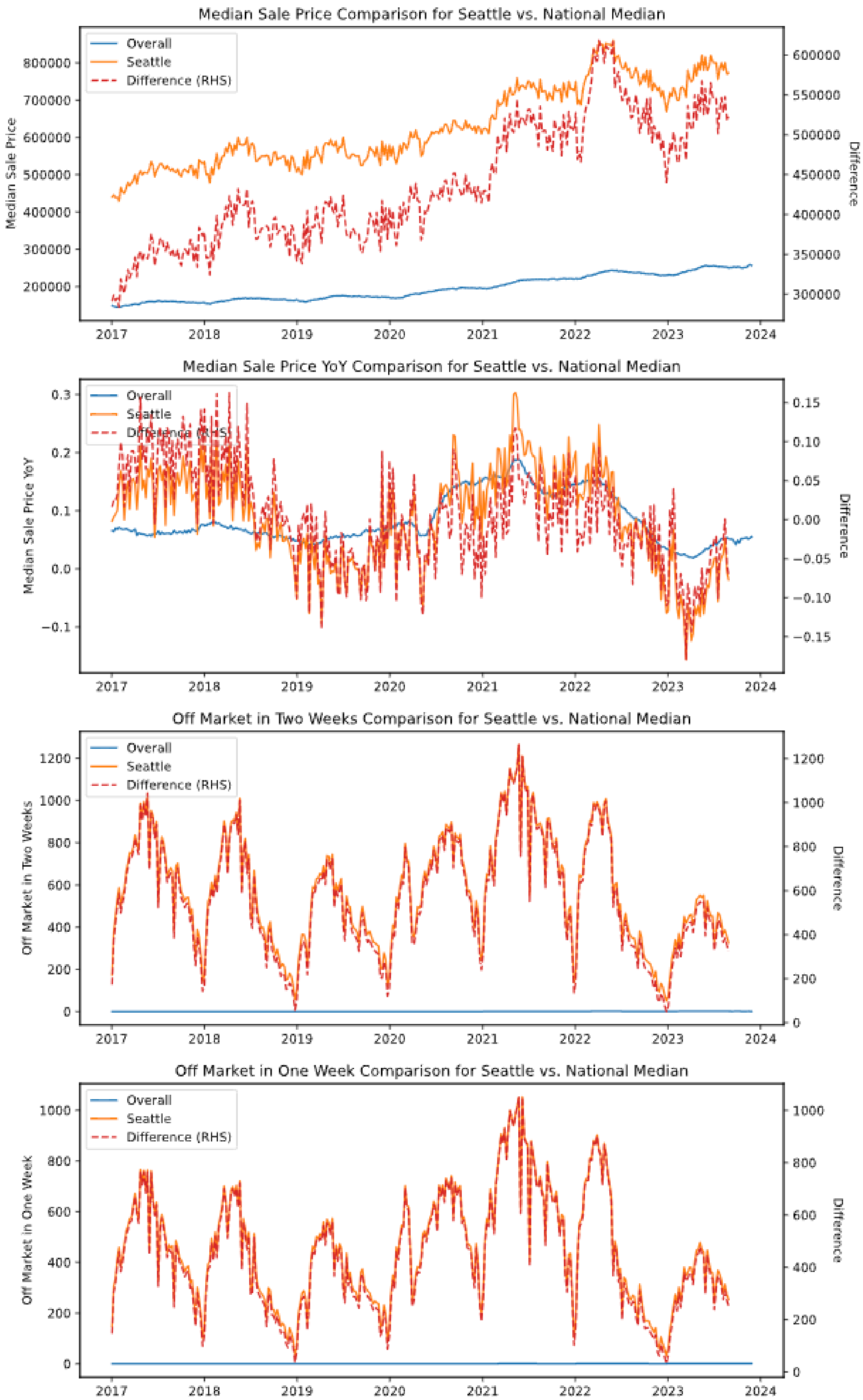
date	age_of_in	total_new	total_acti	inventory	total_hom	total_hom	average_s	percent_h	homes_de	median_n	median_d	median_d	median_s	months_o	pending_s	percent_c	percent_h	price_drops
1/31/2017	59	738	3980.4	3205.8	632.2	159.6	1.006414	0.25284	54.8	472990	21.4	39.8	438490	6.357703	604.8	0.677122	0.366482	128.2
2/28/2017	33.25	874	3855.25	2983.25	745.75	127.75	1.022735	0.17497	44.5	487195.8	9.5	33	458870.6	5.50638	741.25	0.771923	0.494626	115.5
3/31/2017	24.25	1070.25	4137.75	3103.5	892.5	114.5	1.033686	0.130418	49	498487.3	7.25	31.5	478246.9	4.735826	924.25	0.807003	0.539183	107
4/30/2017	22.25	1213.5	4463.75	3342	941.5	85.5	1.042906	0.091735	47.5	492600	6.75	31	496314	4.857897	1031.75	0.818952	0.605012	132.75
5/31/2017	18.4	1450	5423.6	4189.6	1085.6	87.8	1.046968	0.080924	51	507876	7	31.2	505915	5.022755	1153.4	0.791813	0.626019	204.6
6/30/2017	22.25	1380	6211.5	4883.75	1259.75	106.25	1.04457	0.083974	70.75	494223.8	7	31.5	516425	5.052849	1220.75	0.74138	0.617179	294.5
7/31/2017	25.4	1334	6517.6	5342.8	1125.6	134.6	1.03324	0.117884	75.8	498500	7.5	31.4	522990	5.85415	1060.4	0.676753	0.565241	335.8
8/31/2017	31.75	1125	6729	5542	1208.5	167.5	1.02517	0.138406	95.25	489450	8.25	31.75	517000	5.67861	1043.25	0.635054	0.509234	379.25
9/30/2017	32.625	1188	6756.5	5656	1111.5	204.25	1.018576	0.181651	95.5	528300	9.5	32	507243.8	6.322538	941.75	0.628847	0.47765	415
10/31/2017	37.6	922.2	6218.6	5102.2	1019.4	194.6	1.015026	0.190891	129.2	512894	10	31	507000	6.185683	911.4	0.59735	0.447581	369
11/30/2017	48.25	606.75	4750.25	3871.5	956.5	213.5	1.012208	0.224002	112.5	503551	10.75	31.75	515487.5	5.3139	687.75	0.561607	0.432233	197.75
12/31/2017	61	359.75	3426.5	2802	904.25	200.25	1.010226	0.221023	81.5	489856.3	13	33.5	523730	3.817317	475	0.598104	0.397878	73.75
1/31/2018	49.8	720	3192	2484.4	541.2	112.2	1.014361	0.207163	42	529914	12.2	33.2	506900	5.950237	589.4	0.724618	0.42004	75.8
2/28/2018	24.5	927.75	3467.25	2627.5	740.75	93.75	1.035531	0.128657	30.5	577802.5	7.25	29.25	526518.8	4.860151	747.75	0.828885	0.532575	94.75
3/31/2018	19.25	1123	4082.5	3016.75	871.5	84.75	1.039934	0.097016	32.5	569455	7	28	551718.8	4.762923	992	0.828461	0.589011	114
4/30/2018	18.6	1360.2	4843.4	3734.8	967.4	74	1.044842	0.076068	40.4	572375	6.2	28.6	570870	5.068984	1039.8	0.807089	0.620738	175.6
5/31/2018	19.5	1467.75	6058.25	4861.5	1086.75	84.75	1.040283	0.077747	57.75	576000	7	28.25	582224.4	5.595602	1110.75	0.757864	0.606667	295
6/30/2018	23.5	1471.25	7279.5	6076.75	1200.75	137.75	1.027107	0.112157	86	577750	7.25	29.5	578675	6.177931	1087.25	0.693367	0.542598	448
7/31/2018	30.8	1288.2	8099.4	7057.4	1036.2	171.2	1.012815	0.165045	111.4	542720	9	30	562790	7.902057	898	0.572418	0.449844	551.2
8/31/2018	36	1077.75	9022	8059.5	1002.75	212	1.002908	0.210473	134.5	532412.5	10.75	30	548533.4	9.114552	776.75	0.51034	0.361886	717.5
9/30/2018	41	1231.75	9790.25	8809.5	863	240.5	0.996303	0.274198	175.75	573237.5	14.75	30.25	540500	11.76863	734	0.474402	0.291203	863
10/31/2018	46.6	879.8	9782.8	8786	809.8	261.8	0.991407	0.321039	231.4	551734.5	19	29.6	550240	12.25776	665.8	0.398215	0.252093	894.2
11/30/2018	57	550.25	8427.25	7525.25	766.25	299.5	0.988066	0.387871	245	523461.9	26.25	31	541225	11.44708	535	0.345988	0.217844	544.75
12/31/2018	74.6	338.6	6528	5867.2	673.2	295.6	0.985724	0.440582	185.8	522834.5	38.4	30.8	526712.3	10.09549	352.2	0.351686	0.187556	273.4
1/31/2019	77.25	762.5	6006	5181	530.75	224.5	0.984248	0.424957	74.5	582087.8	44.625	32.25	518311	11.65142	555	0.573632	0.186487	326
2/28/2019	59	776.75	5672	4917.5	716.75	236.25	0.990784	0.339863	63.5	587731.3	33.875	29.75	536330.6	8.538677	550.25	0.617621	0.233903	230.75
3/31/2019	35	1217	6550.75	5485.25	777.5	200.5	0.997768	0.259511	60	599718.5	15.625	29	557243.8	8.789209	874.25	0.697338	0.290006	309.5
4/30/2019	30.8	1384.6	7371.8	6194.6	929.8	178.2	1.003657	0.191759	74	588919.8	8.6	29.4	565150	8.087031	1008.6	0.6651	0.361378	399.2
5/31/2019	31	1414.75	8477.5	7277.5	1078.75	192	1.003913	0.17829	88.5	586855.4	8	29.75	582932.1	7.918276	1047.25	0.640274	0.379665	548.5
6/30/2019	35.5	1341.75	9056.5	7833.5	1182.25	236.75	1.003325	0.197426	116.5	566225	9.5	30.5	574625	7.8438	1031	0.586323	0.366363	658.25
7/31/2019	43.4	1098.6	8882.6	7767.2	1013.2	255.4	0.999661	0.250339	130.6	539895	12.3	30.4	562063	8.944311	896.2	0.526648	0.322408	609.4
8/31/2019	47	929	8618.5	7582.75	1051.5	301.75	0.995886	0.287065	129.25	545675	16.25	30.75	552462.5	8.367845	793.75	0.497485	0.293559	581.25
9/30/2019	48.4	1041.4	8562	7525	880.4	268.2	0.994917	0.305059	139.6	575194	19.6	31.1	535500	9.887739	773.2	0.54141	0.267996	640.4
10/31/2019	50.75	784	7790	6754.25	908.5	271.25	0.994802	0.297517	175.5	562268.8	16.75	30.75	563872	8.651622	730	0.489777	0.283487	548.5
11/30/2019	60.75	515.5	6226.25	5321	834.75	274	0.993793	0.328123	168.5	542481.3	21.25	31.5	548958.3	7.510555	599.25	0.481795	0.260047	273.5
12/31/2019	71.8	375.8	4490.2	3854.4	776.2	273.2	0.992631	0.350858	116	535583.5	28.2	32	568415	5.946323	403.6	0.493898	0.245553	127

summary.csv

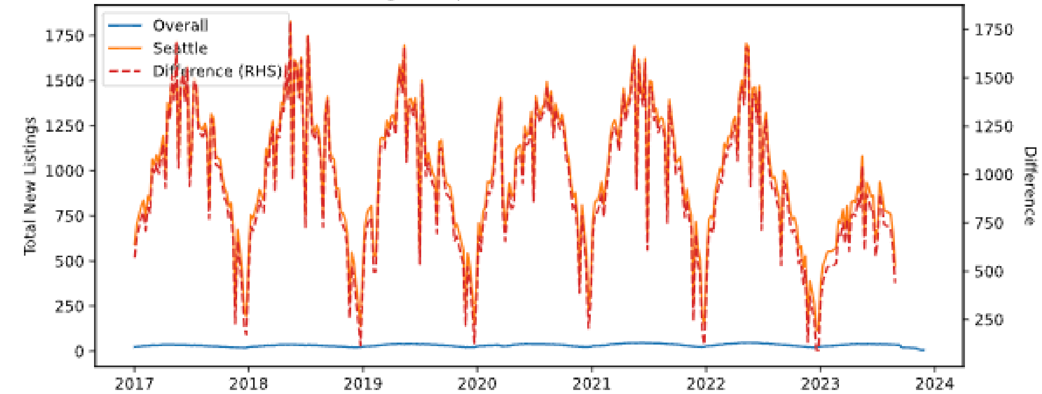
	min	mean	median	max	range	stdev	sem
age_of_inventory	9	36.01149	31.5	81	72	17.03323	0.913077
total_new_listings	114	981.6868	965	1826	1712	362.4431	19.42899
total_active_listings	1307	5447.408	5099.5	10043	8736	1895.395	101.6039
inventory	1045	4471.922	4158	9005	7960	1820.052	97.56501
total_homes_sold	282	894.8506	895.5	1609	1327	257.5806	13.80777
total_homes_sold_with_price_drops	23	159.3075	150	427	404	76.4504	4.098171
average_sale_to_list_ratio	0.976366	1.019797	1.013624	1.11999	0.143624	0.028763	0.001542
percent_homes_sold_with_price_drops	0.024691	0.193021	0.168513	0.469835	0.445144	0.106017	0.005683
homes_delisted	18	90.90805	73	362	344	54.90298	2.943109
median_new_listing_price	449950	631676.9	600875	849950	400000	102794.6	5510.372
median_days_on_market	5	11.92529	8	55	50	9.250836	0.495897
median_days_to_close	26	30.4842	30	42	16	2.000478	0.107237
median_sale_price	428950	630254.4	599972.5	860580	431630	107896.2	5783.842
months_of_supply	1.95952	6.454088	5.998038	15.13568	13.17616	2.569847	0.137758
pending_sales	165	821.1552	822	1442	1277	263.2177	14.10995
percent_off_market_in_two_weeks	0.308333	0.6685	0.682092	0.937355	0.629022	0.1489	0.007982
percent_homes_sold_above_list	0.146628	0.436405	0.442345	0.7481	0.601472	0.152216	0.00816
price_drops	12	302.9885	248	1031	1019	215.0918	11.53013

longitudinal.svg

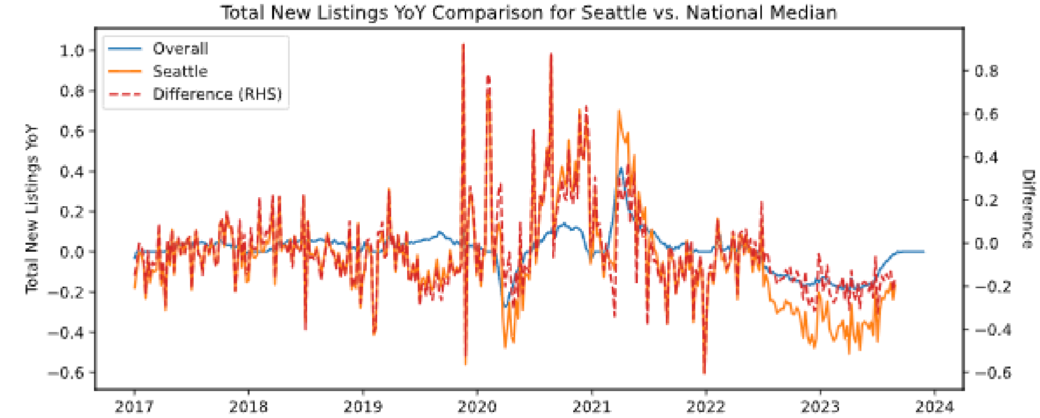



Median Sale Price Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Median Sale Price
Difference
Median Sale Price YoY Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Median Sale Price YoY
Difference
Off Market in Two Weeks Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Off Market in Two Weeks
Difference
Off Market in One Week Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Off Market in One Week
Difference

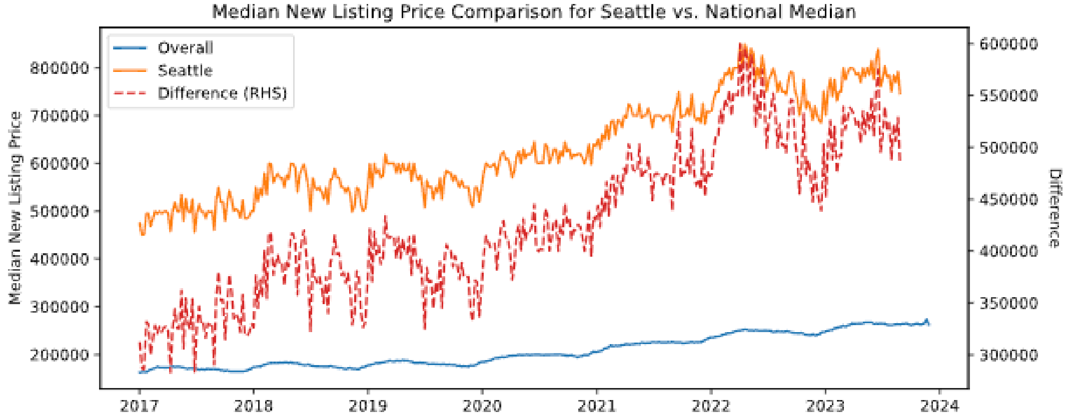
Total New Listings Comparison for Seattle vs. National Median



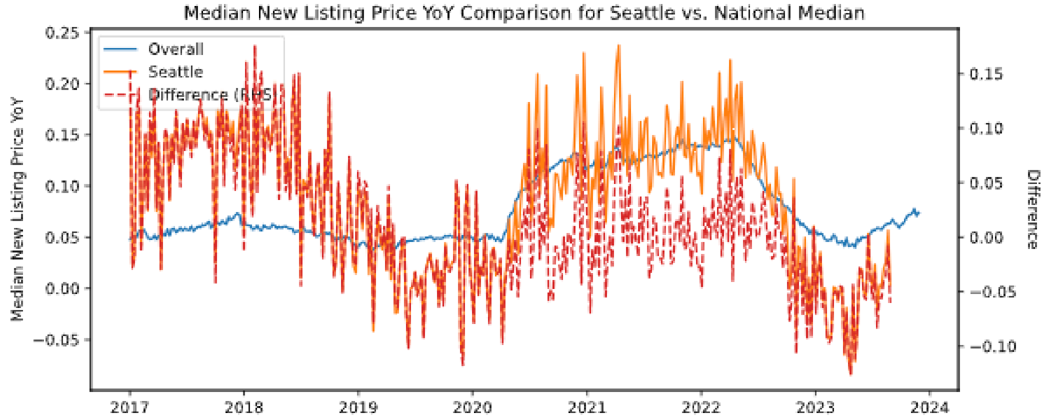
Total New Listings YoY Comparison for Seattle vs. National Median



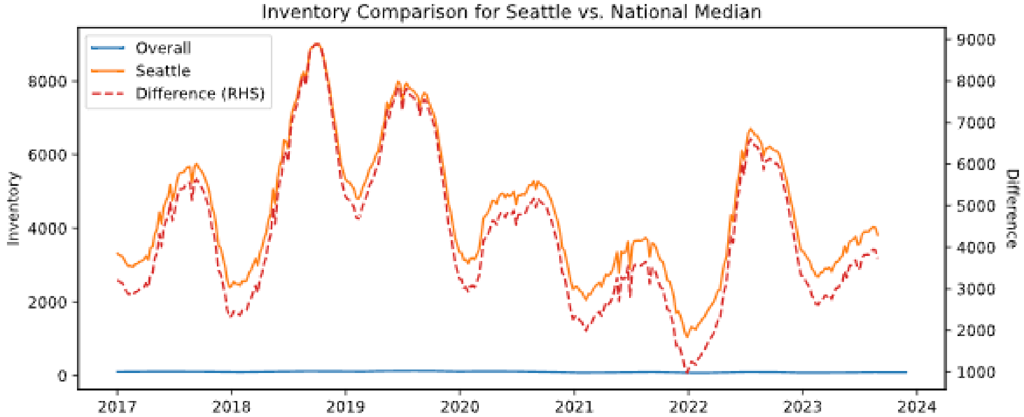
Median New Listing Price Comparison for Seattle vs. National Median



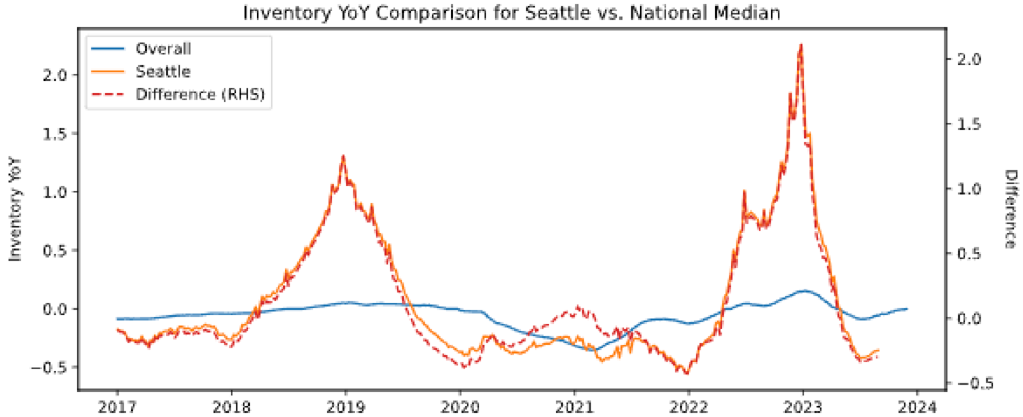
Median New Listing Price YoY Comparison for Seattle vs. National Median



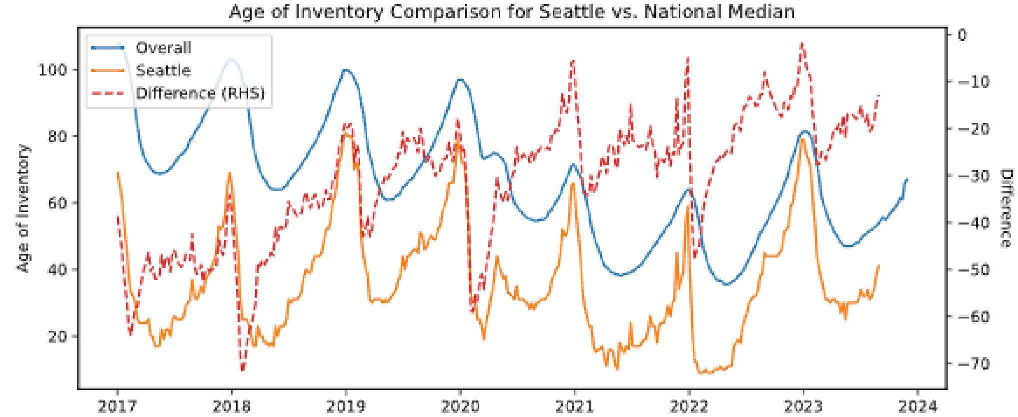
Inventory Comparison for Seattle vs. National Median



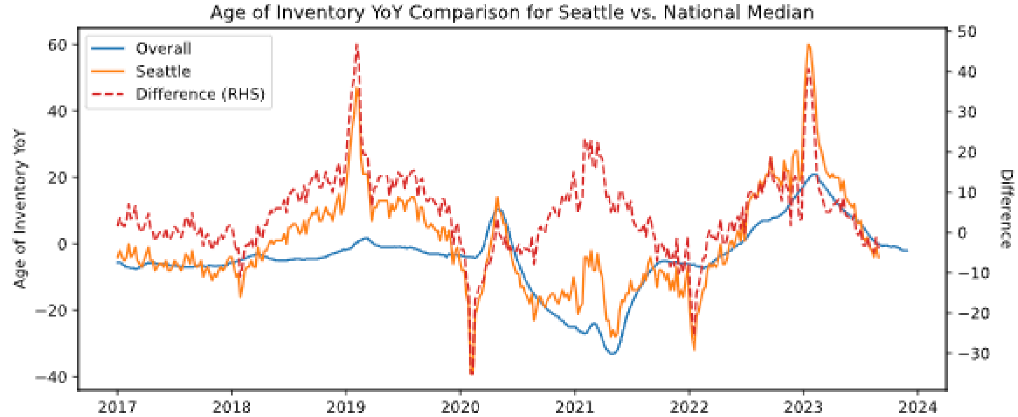
Inventory YoY Comparison for Seattle vs. National Median

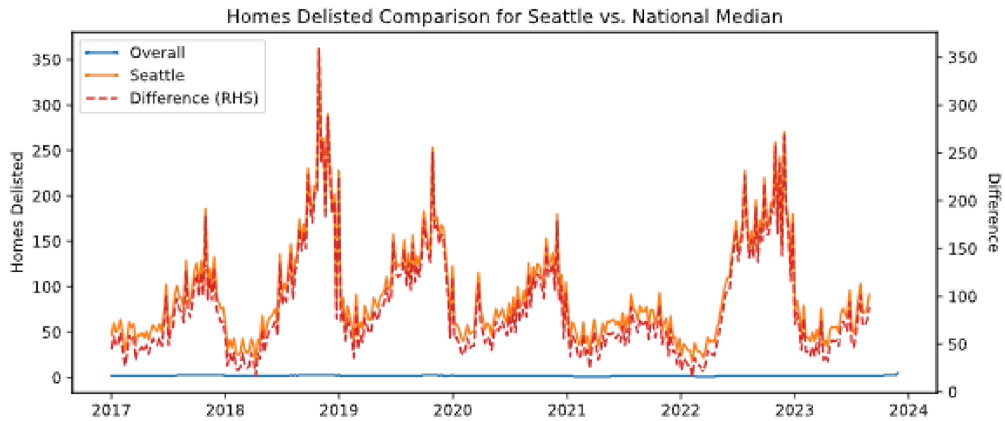


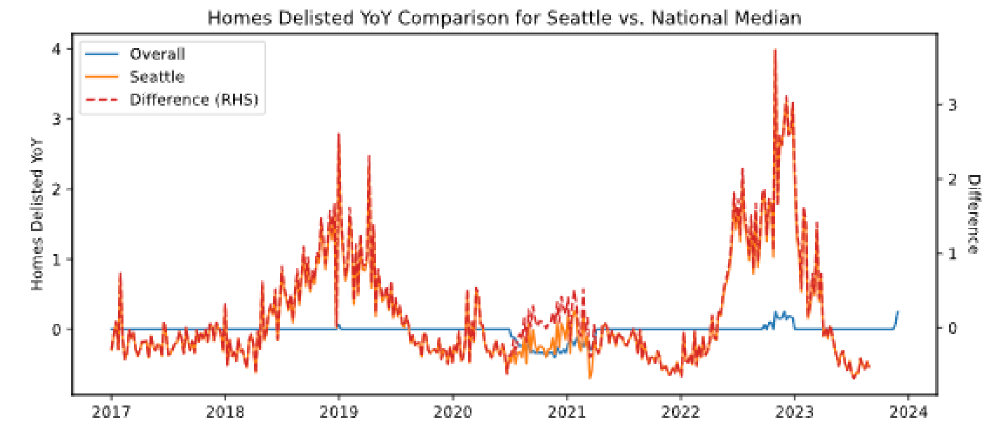
Age of Inventory Comparison for Seattle vs. National Median

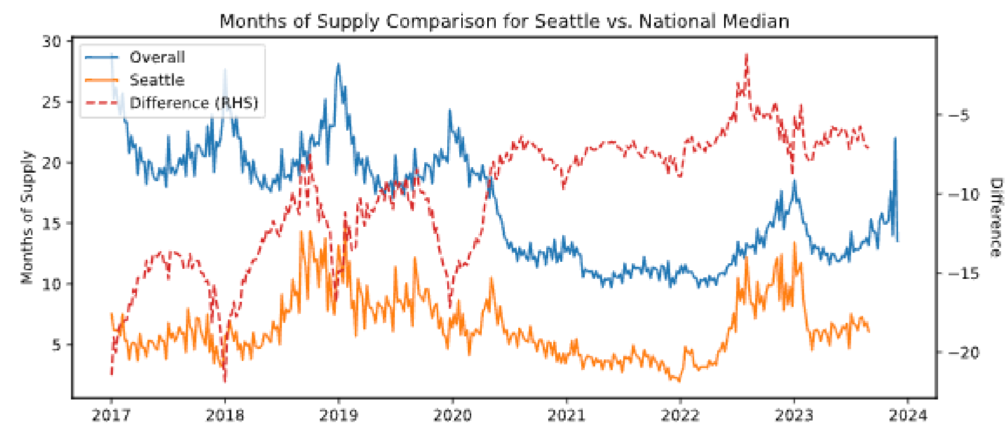


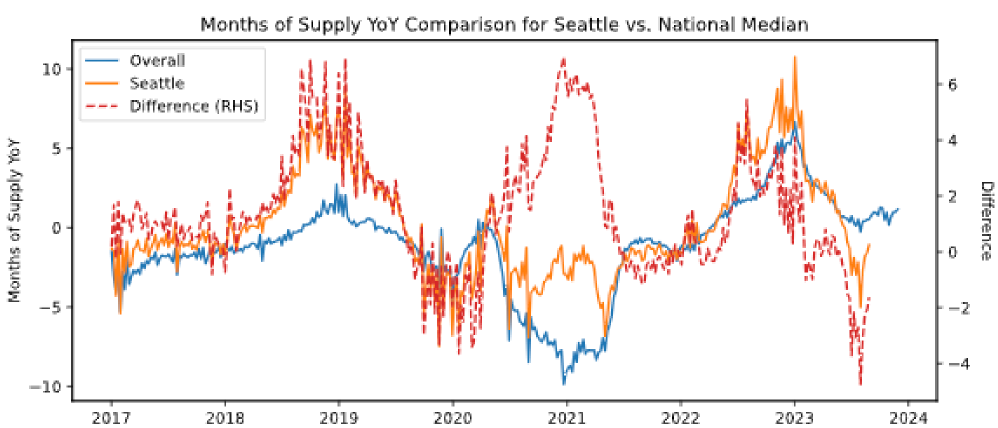
Age of Inventory YoY Comparison for Seattle vs. National Median




Homes Delisted Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Homes Delisted
Difference
0
50
100
150
200
250
300
350
2017
2018
2019
2020
2021
2022
2023
2024


Homes Delisted YoY Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Homes Delisted YoY
Difference
0
1
2
3
4
2017
2018
2019
2020
2021
2022
2023
2024

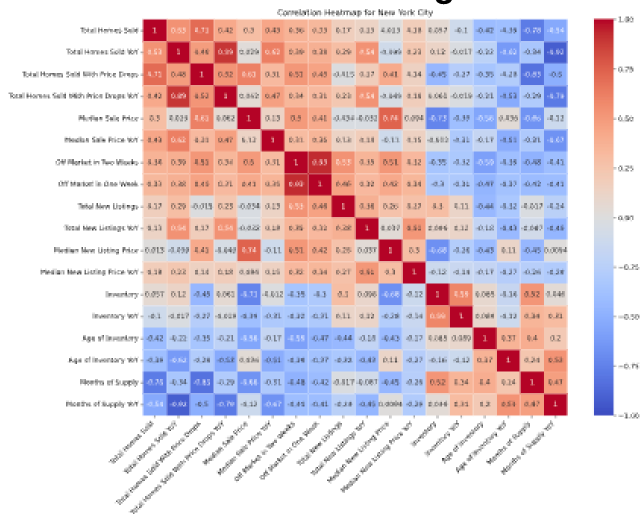

Months of Supply Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Months of Supply
Difference
5
10
15
20
25
30
−5
−10
−15
−20
2017
2018
2019
2020
2021
2022
2023
2024

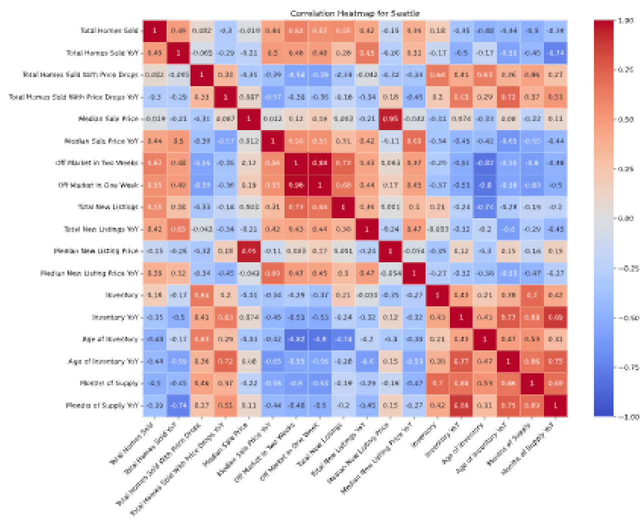

Months of Supply YoY Comparison for Seattle vs. National Median
Overall
Seattle
Difference (RHS)
Months of Supply YoY
Difference
−10
−5
0
5
10
−4
−2
0
2
4
6
2017
2018
2019
2020
2021
2022
2023
2024

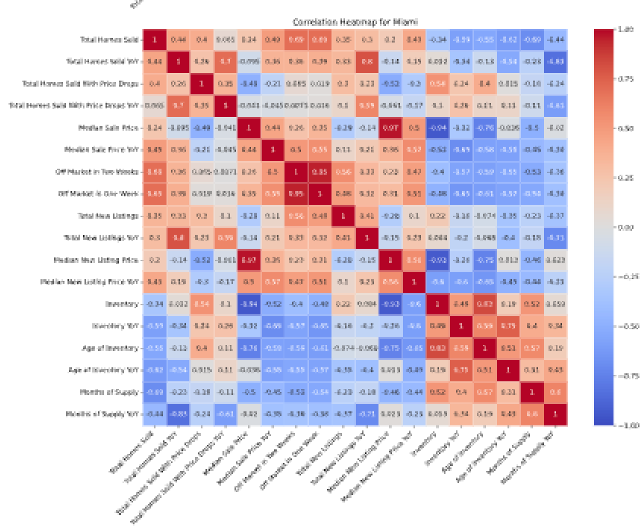
Diagnostic Analytics

correlations.svg

correlations.svg







Predictive Analytics

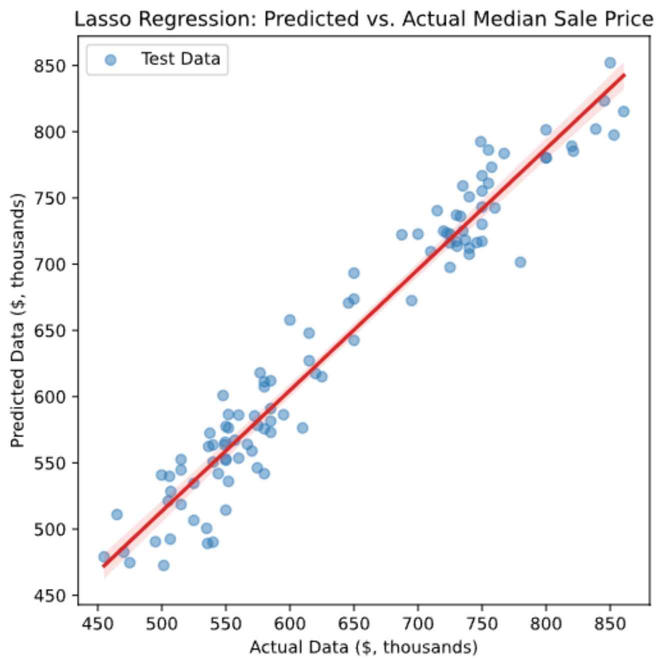
lasso.svg

ridge.svg

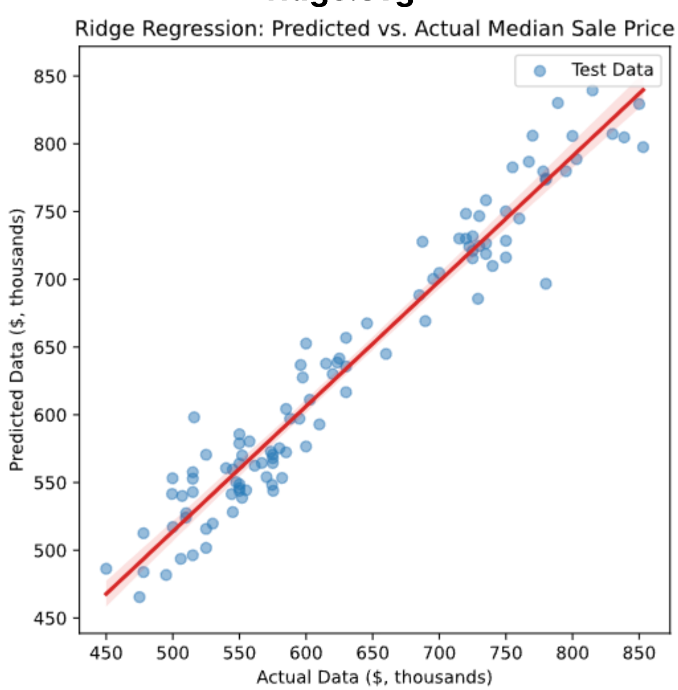
svr.svg

svr.svg matches the formatting of svr.svg for New York City.

lasso.svg



ridge.svg



Prescriptive Analytics

regression_results.png

regression_results.png + Analysis in README.md

Prescriptive Analytics

To better understand market dynamics in Seattle, one can infer useful information from the output of its regression table (shown below). We can see that the variables selected have an adjusted R^2 of 0.944.

OLS Regression Results

Dep. Variable:	median_sale_price	R-squared:	0.947
Model:	OLS	Adj. R-squared:	0.944
Method:	Least Squares	F-statistic:	373.9
Date:	Tue, 05 Dec 2023	Prob (F-statistic):	7.29e-204
Time:	22:01:27	Log-Likelihood:	-4100.3
No. Observations:	355	AIC:	8235.
Df Residuals:	338	BIC:	8300.
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.595e+04	2.76e+04	1.305	0.193	-1.83e+04	9.01e+04
total_homes_sold	109.8609	18.472	5.948	0.000	73.527	146.195
total_homes_sold_yoy	1.073e+04	1.48e+04	0.726	0.468	-1.83e+04	3.98e+04
total_homes_sold_with_price_drops	-227.8986	44.590	-5.111	0.000	-315.608	-140.189
total_homes_sold_with_price_drops_yoy	-1.033e+04	2951.080	-3.502	0.001	-1.61e+04	-4529.108
off_market_in_two_weeks	-90.1921	56.076	-1.608	0.109	-200.493	20.109
off_market_in_one_week	57.1180	60.156	0.950	0.343	-61.209	175.445
total_new_listings	-53.6013	8.778	-6.106	0.000	-70.868	-36.334
total_new_listings_yoy	-2955.5639	9409.175	-0.314	0.754	-2.15e+04	1.56e+04
median_new_listing_price	0.9855	0.023	43.223	0.000	0.941	1.030
median_new_listing_price_yoy	-1.066e+05	2.99e+04	-3.563	0.000	-1.65e+05	-4.78e+04
inventory	4.7114	3.223	1.462	0.145	-1.629	11.052
inventory_yoy	2.143e+04	8189.296	2.616	0.009	5317.751	3.75e+04
age_of_inventory	-619.9703	255.206	-2.429	0.016	-1121.962	-117.979
age_of_inventory_yoy	-17.6973	235.472	-0.075	0.940	-480.873	445.478
months_of_supply	-114.3352	2411.029	-0.047	0.962	-4856.846	4628.176
months_of_supply_yoy	-934.6215	1780.148	-0.525	0.600	-4436.186	2566.943

Omnibus:	3.269	Durbin-Watson:	1.653
Prob(Omnibus):	0.195	Jarque-Bera (JB):	3.766
Skew:	0.004	Prob(JB):	0.152
Kurtosis:	3.505	Cond. No.	1.48e+07

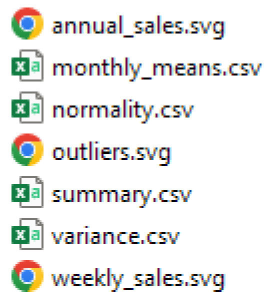
- From the results, we can see that transaction **volume** has a meaningful impact on the level of home prices: `total_homes_sold` has a significantly positive effect on Seattle's home price, while the year-over-year (`yoy`) change in total homes sold has a significantly negative effect. This suggests that, although home prices increase alongside total transaction volume, a "surge" in volume would actually lead to a decline in home prices, possibly due to the market being "swamped" by increased supply (i.e., more sellers than buyers).
- From the results, the level of **distress** in the market negatively impacts home prices. Specifically, `total_homes_sold_with_price_drops` has a significantly negative effect on home prices, and the effect of year-over-year change in total homes sold with price drops is also significantly negative. This suggests a "slippery slope" impact: price cuts coupled with an increase in price cuts, could serve as a "double whammy" for home prices.
- From the results, the most **popular** or **demanded** cohort of homes does not have a significant impact on home prices. Specifically, if one were to measure popular (i.e., more demand/desired) cohorts as homes that are sold or taken `off_market_in_one_week` or `two_weeks`, the coefficients on both are insignificant. This suggests that particular "hotness" of the market tends to be contained, and not materially "spreading" to impact overall home prices.
- Sellers' **asking prices** (`new_listing_price`) have a significantly positive correlation with median sales price, which is very intuitive. Interestingly, year-over-year changes in listing price exhibit a significantly negative correlation with sale prices. This is in line with the observation from the total homes sold metrics. As there are more sellers, and listing prices get overly increased, home prices could actually be negatively impacted.
- **Supply** in Seattle paints a mixed, but interesting, story. Here, we break down the notion of "housing supply" into its various forms or sources.
    - First, when viewed through the lens of `total_new_listings` (i.e., amount of new/fresh supply/interest in selling) in the market, an increase in new listings correlates with a statistically significant decline in home prices.
    - Second, when viewed through `total_inventory` (the total number of active listings on the market), the level itself actually does not have a statistically significant impact. However, the year-over-year change in inventory has a significantly positive correlation with home prices.
    - Third, `age_of_inventory` has a significantly negative correlation with home prices. Age of inventory is measured as the number of days active inventory has been on the market. As homes continue to stay on the market without being bought, this indicates demand is weaker than supply.
    - An interesting question arises: How can one reconcile the observations from the first and second point? The key takeaway is that housing prices benefit from stable availability of housing supply (which is measured by year-over-year change in total inventory). However, a sudden shock in new supply, which is measured by new listings during each time period (1-week in our analyzed dataset), causes a contemporaneous negative shock to home prices.
- **Policy recommendation**: Given the backdrop in this analysis, policy recommendations should be two-fold. Policy should be aimed at improving affordability and maintaining home price stability. Therefore, Seattle should continue to expand available housing, increasing housing supply at a gradual pace, while closely monitoring any abnormal activity in particular areas of the market to avoid potential shocks to the system (either excessive aging of inventory, excessive popularity of certain cohorts, or sudden supply shocks). Furthermore, the city should also construct more buyer-friendly assistance programs, such as initiatives that provide lower cost loans for first-time homebuyers and deeper collaborations with banks and non-bank lenders to better understand the city's evolving demographics. The initiatives should be designed with the purpose to help boost transaction volumes, and serve as a good matching mechanism to help different buyers with different needs find appropriate homes.

Miami

Intro

Miami housing market historically has been dynamic and affected by the factors such as population growth, job opportunities, international investments and others. The analysis below examines trends in the Miami housing market between 2017 and 2023.

Descriptive Analytics

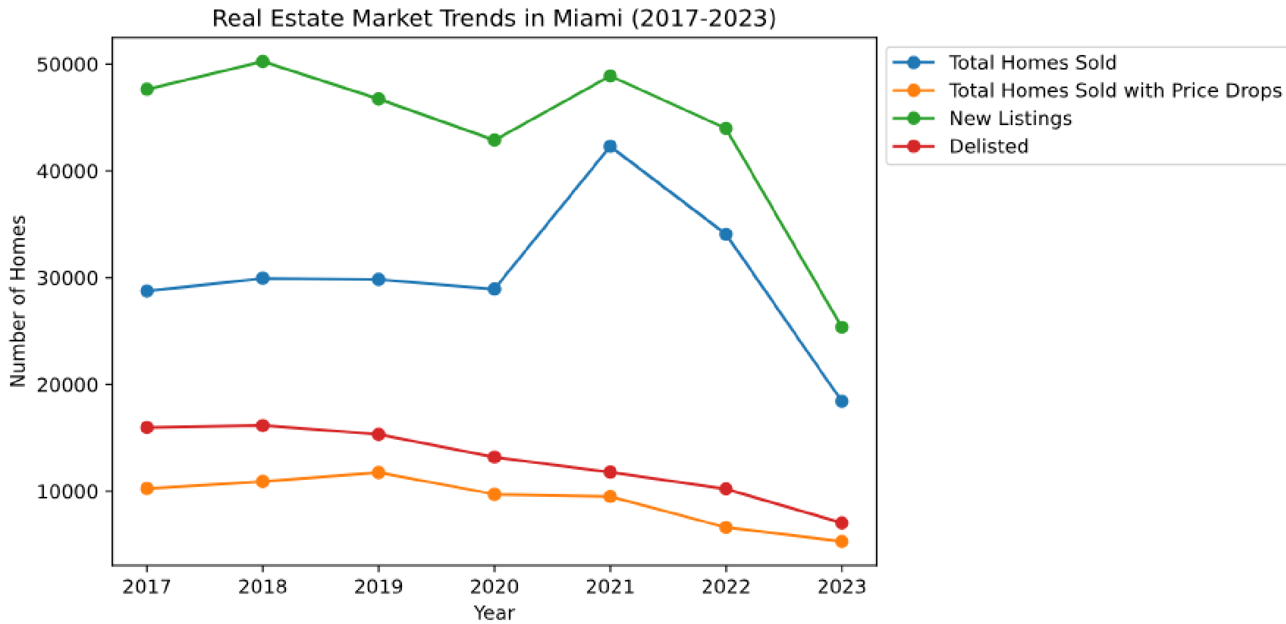


annual_sales.svg

The line graph illustrates the fluctuation in the volume of homes sold in Miami, delineating changes in different categories: total homes sold, total homes sold with a price drop, new listings, and delisted homes, i.e., unsold home listings that were removed from the market during a given period.

The graph shows that the number of homes with a price drop and delisted has been historically low, i.e., not exceeding 20,000 homes, and has been declining since 2017, reaching its lowest point in 2023. In contrast, total homes sold and new listing figures have been dynamically changing throughout the years.

First, the notable disparity in the number of listed homes and total homes sold can be highlighted in 2017. This discrepancy suggests an influx of new listings, illustrated as a green line, outpacing the number of homes sold, illustrated as a blue line. This phenomenon might be attributed to many factors, including increased property development/construction and other factors. However, the demand and supply were getting narrower, peaking in 2021 and significantly dropping in 2023.



monthly_means.csv

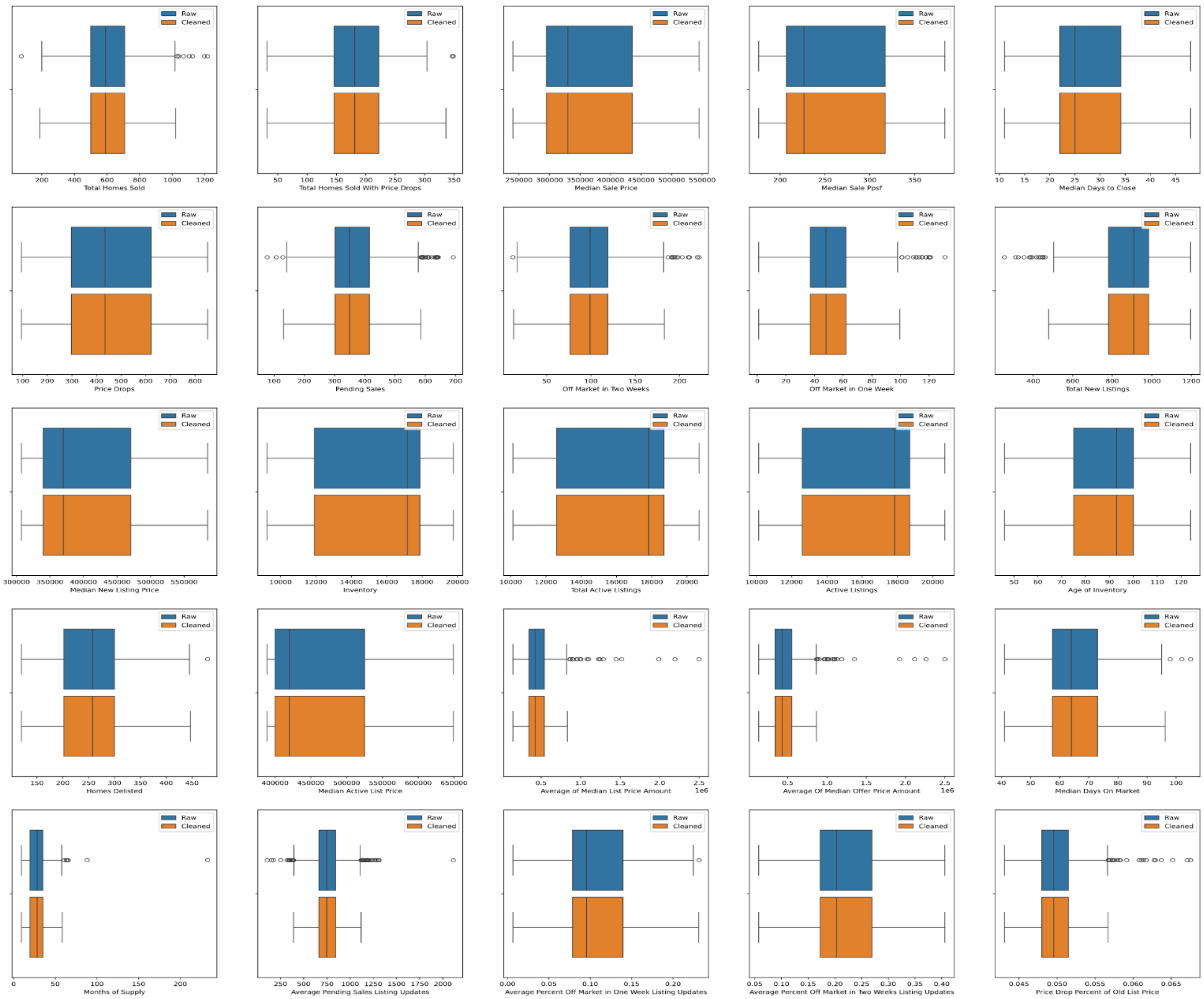
The table describes the monthly average value of 18 primary variables. When comparing the percent change between January 2021 and August 2023, significant changes can be observed in the averages of all reported variables. Notably, the median sales price has increased by almost 100%, the total homes sold increased by 20%, and the median new listing price has also increased by 65%. Downward trends can be observed in inventory dropped by about 40%, median days on the market decreased by about 30%, and the month of supply dropped by about 50%.

normality.csv

The normality test has been applied to assess whether the key variables follow a normal distribution. In this regard, the table computes the statistic (the degree to which the sample deviates from a normal distribution), p-value (probability of obtaining the observed results if the data were sampled from a truly normal distribution), skewness (measures the asymmetry of the distribution), kurtosis (tail of the distribution), and result. Based on the table, only 'total_homes_sold_with_price_drops' is normally distributed.

outliers.svg

The boxplots display the outliers before and after treatment, done with the cap method. This is an essential pre-step for building and running a predictive model.



summary.csv

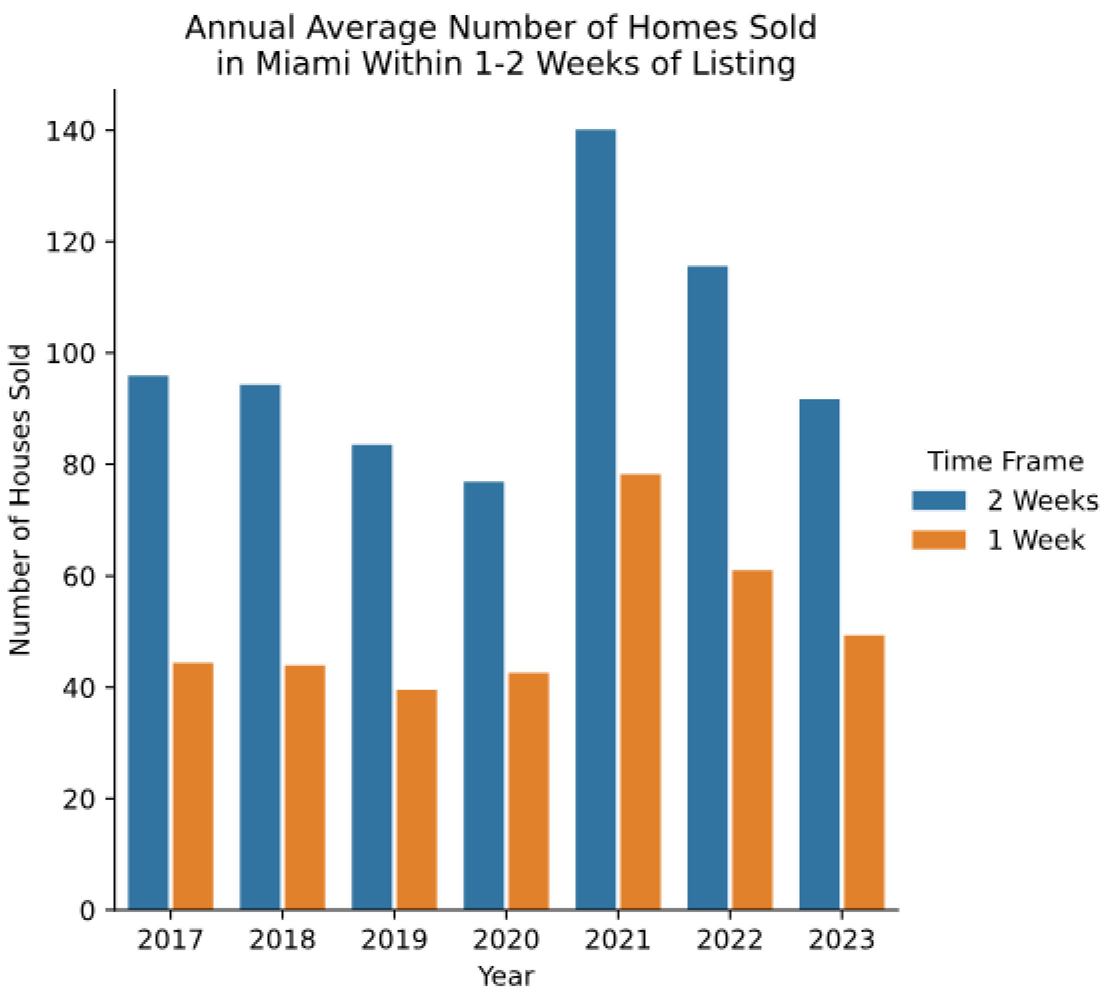
The table displays measurements such as mean, median, max, range, standard deviation, and standard error of the mean for the 19 key variables. The highest standard deviation is for 'median sale price' (305,000) and 'median new listing price' (277,505).

variance.csv

The table displays measurements (statistic, p-value, result) on the variance for 19 variables. According to the table, 'age_of_inventory' is the only variable with an equal variance.

weekly_sales.svg

The chart compares the annual difference between homes sold in Miami within the 1st or 2nd weeks of listing. Based on the chart, historically, homes are primarily sold within the 2nd week of listing, and sales within the 1st week are much lower. Overall, sales within both the 1st and 2nd weeks peaked in 2021, which is also validated by other findings in this analysis.

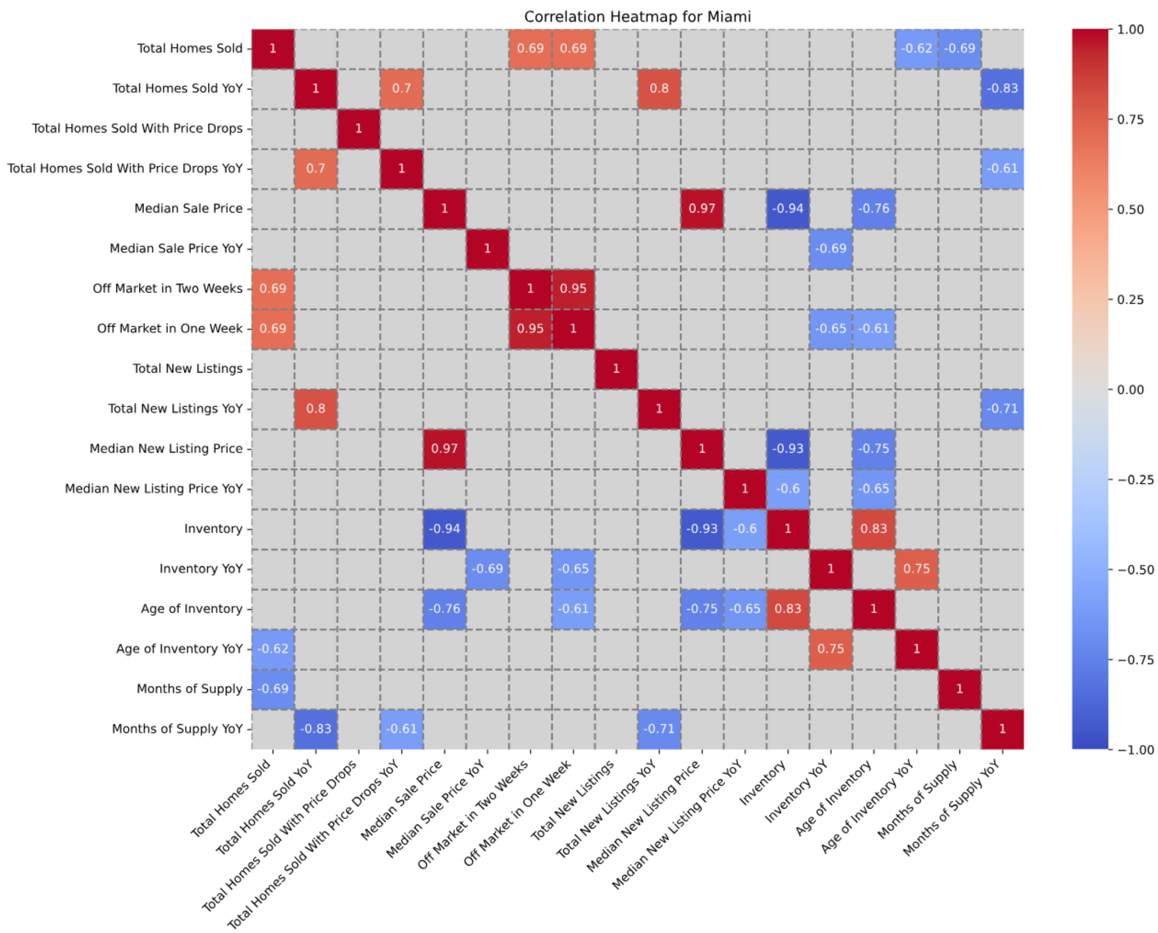


Diagnostic Analytics

- correlations_strong.svg
- distributions.svg
- homes_sold_months_supply.svg
- price_trends.svg
- sale_price_days_to_close.svg
- sale_price_median_days_to_close.svg

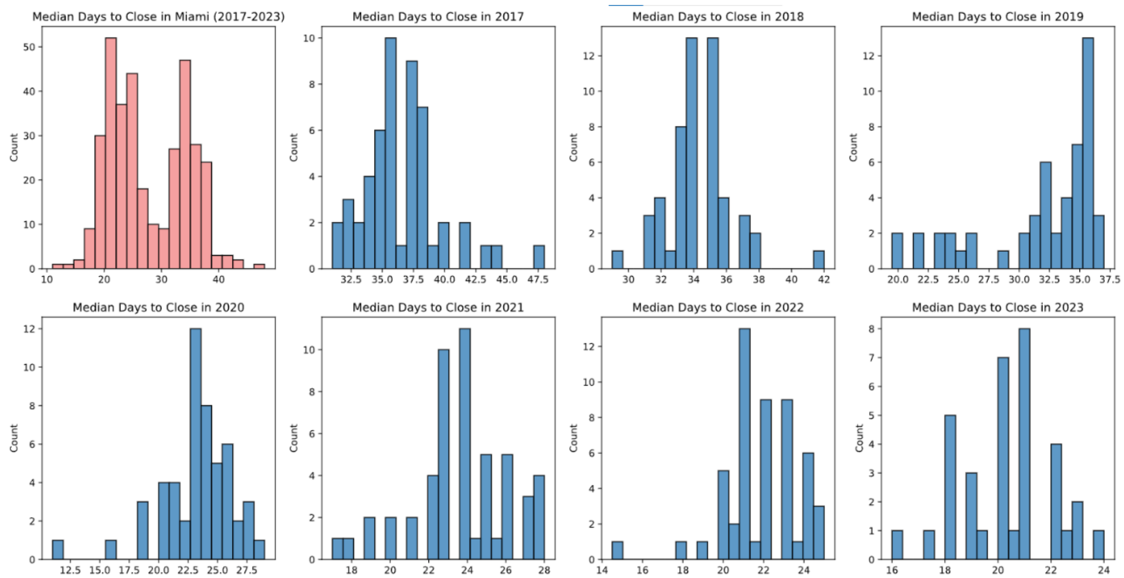
correlations_strong.svg

The correlation heatmap explores and illustrates the strong correlation (0.6 and above) between key variables in New York, Seattle, and Miami. The highest negative correlation for the dataframe on Miami has been observed between variables such as 'total homes sold,' 'total homes sold with a price drop,' 'age of inventory,' and 'monthly supply'; 'new listing prices,' 'median sale price,' and 'inventory.' The positive strong correlation is observed between 'total homes sold' and 'off-market in two weeks,' 'off-market in one week.'



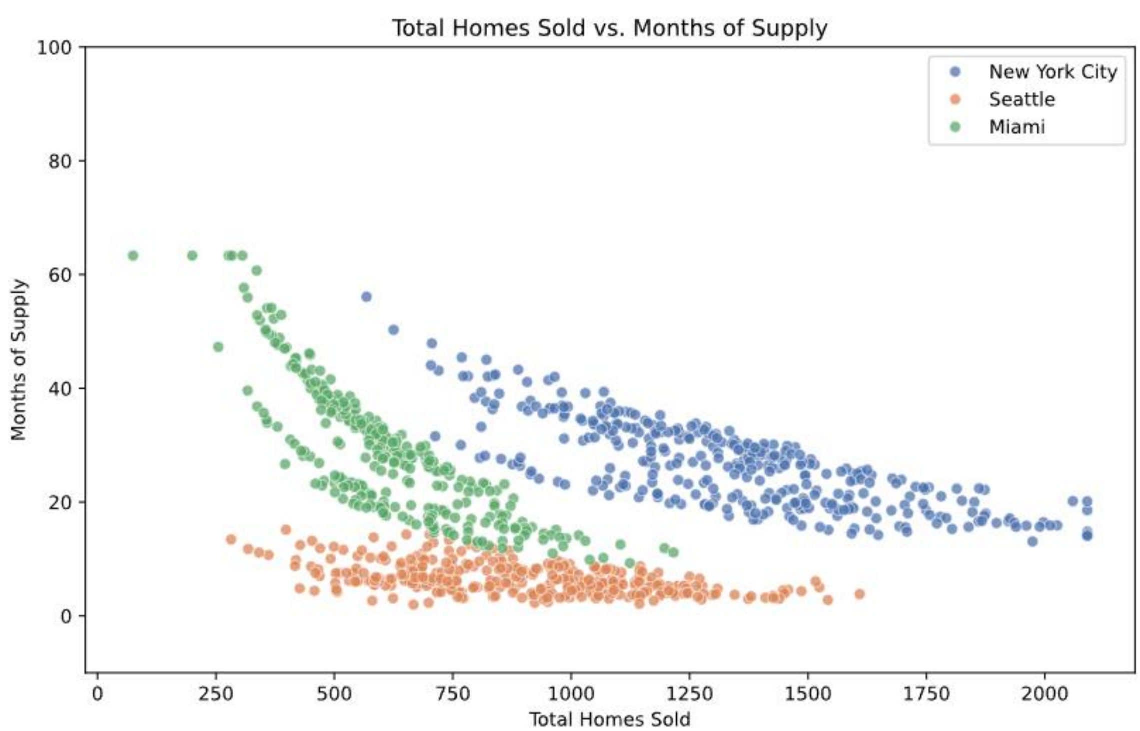
distributions.svg

The histogram illustrates median days to close in Miami in compiled (marked as pink) and then in annual (marked as blue) formats. The compiled chart for the median days to close in Miami has a Bernoulli distribution shape with peaks at around 20 days and about 35 days. Further examination of the data, as evidenced by the annual chart, reveals that the median days to close the transaction between 2017 and 2019 was about 35 days. It dropped to 23 days in 2020 and has continued decreasing in recent years.



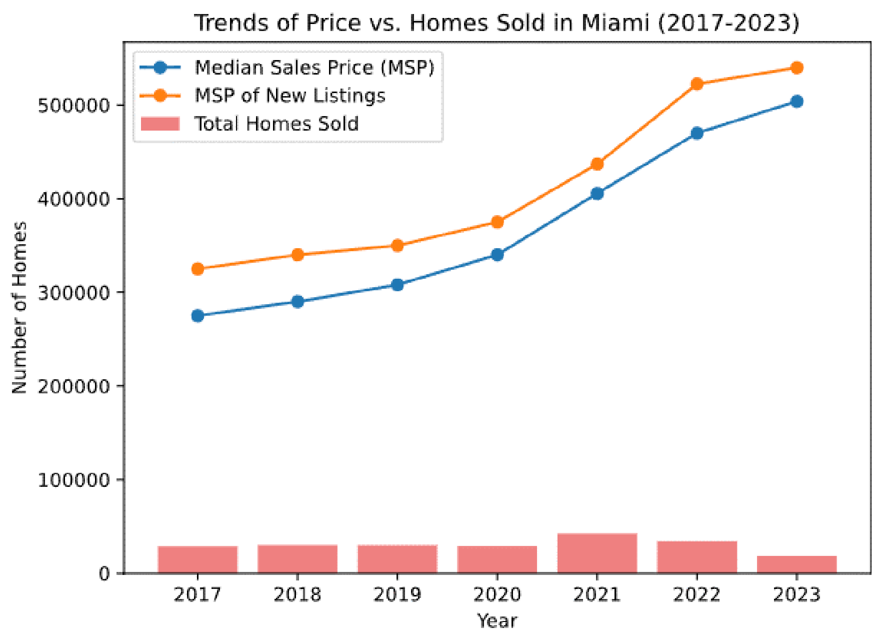
homes_sold_months_supply.svg

The scatterplot aims to compare the relationship between 'months of supply' and 'total homes sold' in three selected cities. Based on the plot, Seattle has the lowest value for the months of supply and total homes sold (this might be explained by the population size and overall demand); New York has more homes sold with the overall months having a higher minimum of 18 and a maximum of 60 months. Miami has a middle value with the lowest homes sold ranging from 15 up to 61 months of supply.



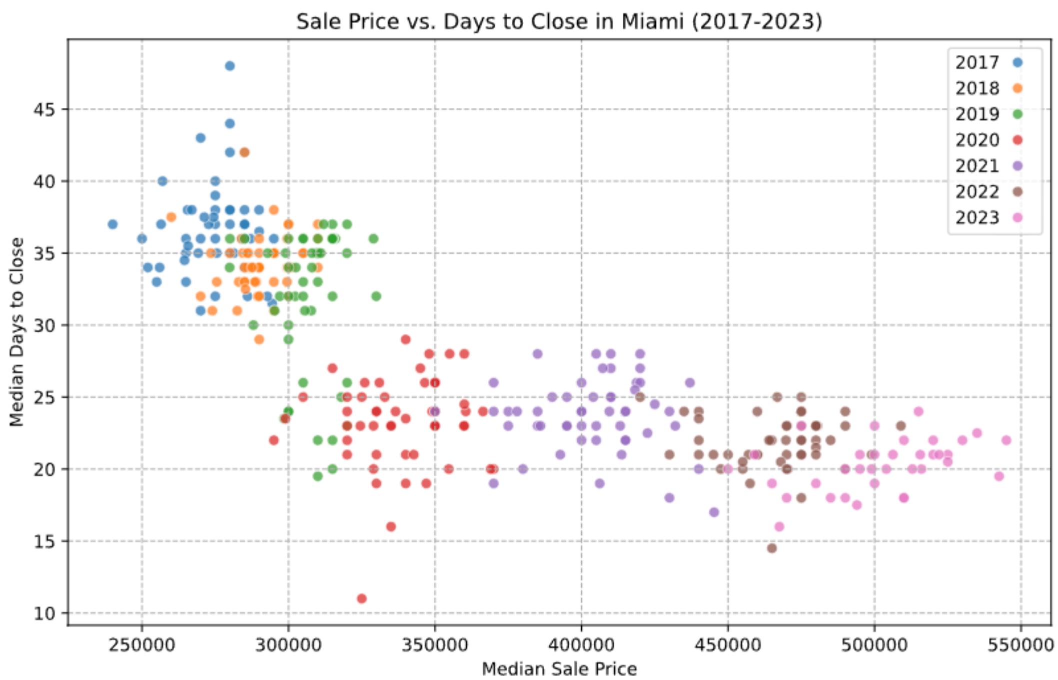
price_trends.svg

The line graphs and bar chart illustrate the relationship between 'median sales price,' 'median sales price of new listings,' and 'total homes sold in Miami.' The chart highlights several important trends in the housing market in Miami: (1) the new listing price is above the median price year after year; (2) housing prices grow steadily in the last years, with a sharp increase during the last 3 years; (3) the number of homes sold is negatively correlated with the price in 2023: with the highest price, the number of homes sold is the lowest. This can be explained that the data for 2023 is only until August.

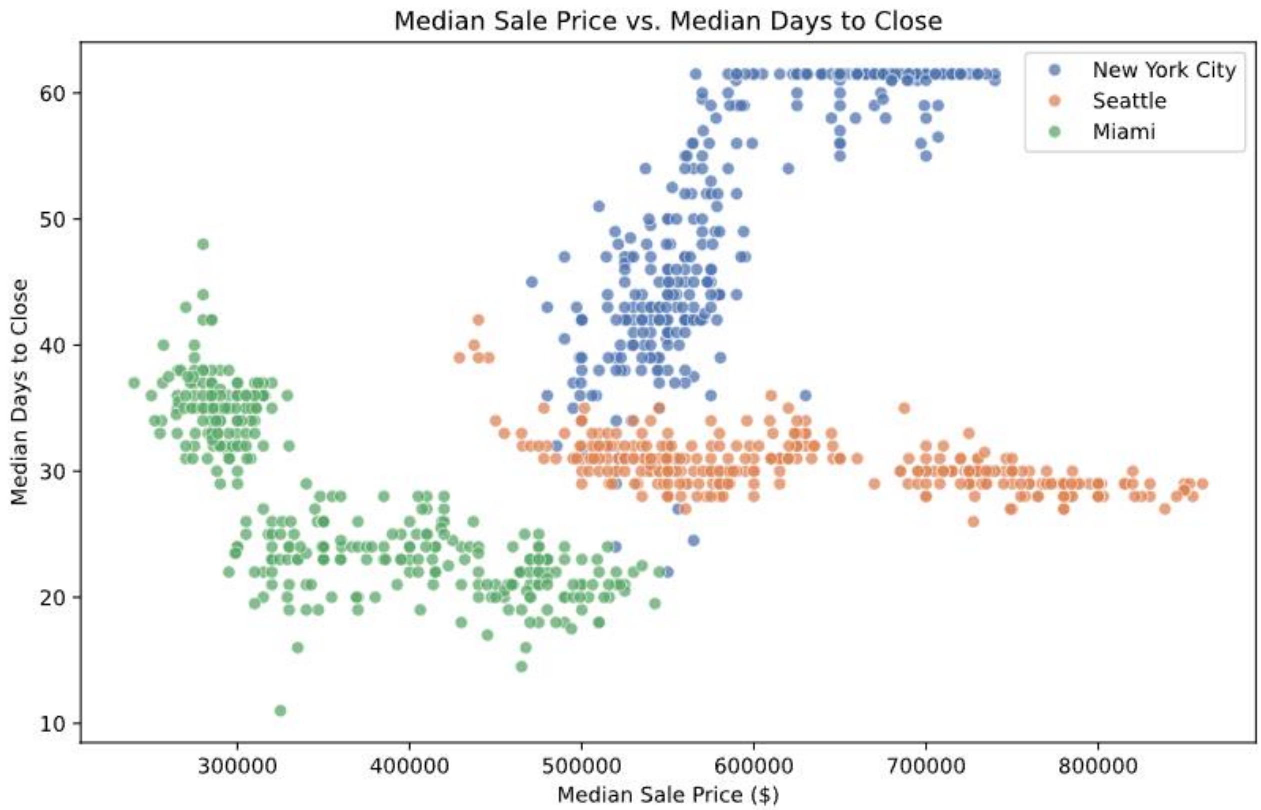


sale_price_days_to_close.svg

The scatter plot examines the median sales price and days to close the transaction in Miami in detail. The plot shows that the median price was fluctuating between USD 250,000 and USD 300,000, and the median days to close the transaction ranged between 30 and 45 days, with a drop in 2019. The number of days for the close transaction remained comparatively stable between 2019 and 2023, with the median prices gradually growing up to USD 550,000.



sale_price_median_days_to_close.svg
The scatterplot examines the relations between median sales price and median days to close, comparing three selected cities. Based on the plot, Miami has prices scattered between USD 250,000 and USD 550,000 and days to close transactions between 25 to 45 days, as also evidenced from the previous scatterplot. Notably, the housing price historically is lower in Miami than in Seattle and New York. Specifically, the feature of the housing market in Seattle is growing constantly and reaching the highest points in comparison with the other cities, remaining stable in terms of the median close transaction between 30 to 40 days. The housing market in New York City is highly skewed in comparison with Miami and Seattle: the price has increased from USD 550,000 to USD 750,000 with the highest median days to close the transaction above 60 days.



Predictive and Prescriptive Analytics

forest.svg

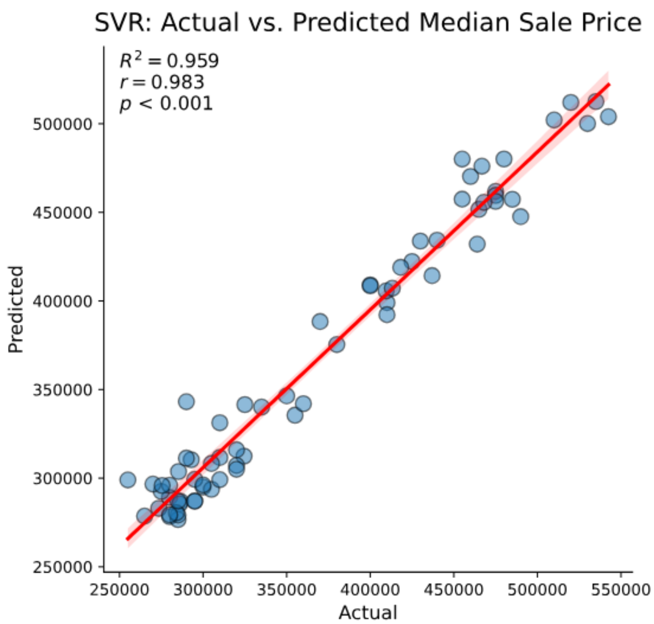
svr.svg

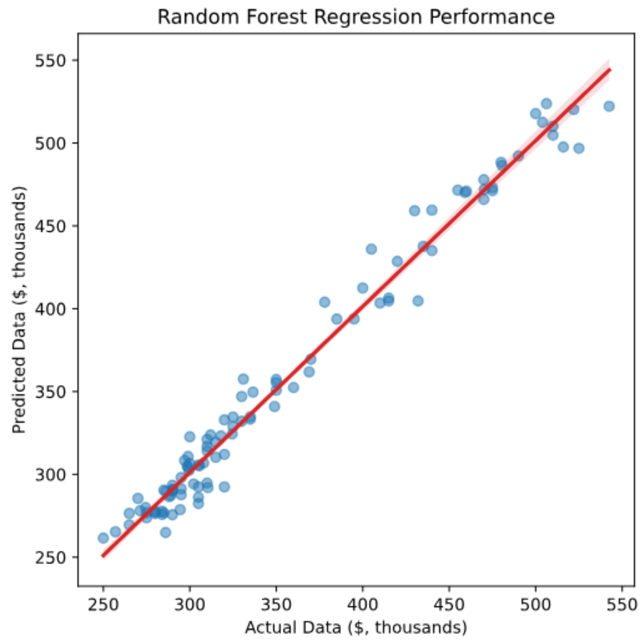
For the purposes of predictive and prescriptive analytics, Support Vector Regression (SVR) and Random Forest (RF) models were applied. The limitation of the dataframe and potential challenges in the models have been described in the README file.

In the SVR model, R-squared value of 0.959 indicates that approximately 95.9% of the variability in the median sales price can be explained by the model. The correlation coefficient (r) of 0.983 displays strong positive linear relationship between the predicted and actual median sales prices. The associated p-value, being less than 0.001, highlights the statistical significance of this relationship. These findings suggest that the SVR model demonstrates a high level of accuracy in predicting median sales prices in the Miami housing market.

In the Random Forest model, the R-squared score of 0.997 further underscores the model's exceptional accuracy, explaining approximately 99.7% of the variance in predicting median sales prices in the Miami housing market between

2017 and 2023. Additionally, Root Mean Squared Error (RMSE) of 4746.55 is also comparatively low. These suggest suggest that the Random Forest model is also highly effective in capturing and modeling the complex patterns in the housing market in Miami.





Insights on the Prescriptive Analytics

Overall, the housing market in Miami can be described as unaffordable. The analysis and predictive model presented above offer a glimpse into the dynamics of the market between January 2017 and August 2023. Market prices soared in 2019 and have continued to grow significantly since then. Furthermore, the initial high difference between supply (new listings) and demand (number of homes sold) has narrowed in recent years. These insights propel further investigation into the housing market in Miami to better understand the case in a larger context and provide recommendations on literature on the subject.

Wijburg (2021) compares the governance of affordable housing in Amsterdam and Miami, arguing that Miami relies heavily on market forces and private contributions [1] . However, turning points can be made with strong public commitments through an Affordable Housing Trust Fund. In a recent study, Weiss (2023) highlights that Miami is experiencing a housing affordability crisis, becoming increasingly unaffordable in the USA due to rising rates, record prices, insurance premiums, and unprecedented migration to Southeast Florida[2]. In this regard, the

[1] Retrieved from https://www.sciencedirect.com/science/article/pii/S0016718520302979
[2] Retrieved from : https://dspace.mit.edu/handle/1721.1/150284

policy recommendations on affordable housing in Seattle provided as part of this assignment are relevant for Miami as well.

Apart from affordability, one aspect highlighted in the research is the high vulnerability of Miami to climate change and its potential effects on the housing market [Rodziewicz et. al (2022)[3], Gosain et. al (2022)[4], Keys & Mulder (2020)[5]], which developers, policymakers, and others must consider when acting upon the housing market in Miami.

Forecasting the housing market has been widely explored both in academic literature and corporate non-academic sources. Some models presented below contrast SVM and RF models above. Møller et al. (2023) developed the Housing Search Index (HIS) with strong predictive power over subsequent changes using Miami as a case. [6] Treyz (2023) presents REMI, a comprehensive economic/demographic forecasting model, for 20 U.S metropolitan areas, including Miami, containing assumptions on housing price elasticities[7]. Hsieh & Lin (2021) present a Generative Adversarial Network (GAN), an unsupervised machine learning algorithm, and Long Short-Term Memory (LSTM), a neural network with long-term memory, for forecasting house prices. [8] In comparison with academic literature, some online commercial publications provide generalized insights into the Miami housing market, claiming that rising construction costs will significantly impact home pricing, low unemployment in Miami will affect the demand and purchasing power of buyers, and other predictions.

Given all of the above, additional recommendations include:

- Enhancing public-private dialogue in Miami to strengthen human-centered housing resilience towards disaster risk reduction. This will involve dialogue supported by policymakers to address potential housing crises induced by climate change.
- Strengthen public commitment towards affordable housing through existing structures such as the Affordable Housing Trust Fund, and encouragement of development to increase the margin for affordable housing.

---

[3] Retrieved from: https://link.springer.com/article/10.1007/s10669-022-09842-6

[4] https://www.sciencedirect.com/science/article/pii/S2212420922002801?casa_token=VMCkH2NzvCoAAAAA:-Dnsjwm8gY75wCijKK6wsHnltc-grb1q0Xh76J_DOF7jyXvfDw2FqxoeFo3oKe5SXmbXF5G7Wg

[5] Retrieved from: https://www.nber.org/papers/w27930

[6] Retrieved from: https://pubsonline.informs.org/doi/full/10.1287/mnsc.2023.4672?casa_token=i-wIAkPwWjsAAAAA%3AORiTxXd6iRc0zP2FFdg-7KeML2KOXyfHKhb-OfjNIaNqjKvQ6PbpFeTdEXW7QsMGCrmbB_Q_fA

[7] Retrieved from https://journals.sagepub.com/doi/full/10.1177/08912424221145186?casa_token=VTm2H-UGXuAAAAAA%3AeP-u62ZZ-qvphvPHBywNXIsr27GVjIrSVgKVVI2nS9Rv3lHHxSos2DbYGX-DXPxdDv_5GZ-Y_PWK

[8] Retrieved from: https://ieeexplore.ieee.org/abstract/document/9778012?casa_token=vRgq-YlDXkAAAAAA:yh3cYoyYz15LMPUlgfEwTuLJnJaxZ3gM-qKhozLESVUDiJKEOYmKhu6-6zDgpBJ1sWcXQQD2

Future research on the housing market in Miami can significantly benefit from merging Redfin data with data frames that describe demographic context and applying new models for forecasting housing prices.