EDA

Out[1]: Click here to display all codes

Single Family

Out[5]:

	MLSNUM	STATUS	LISTPRICE	SOLDPRICE	LISTDATE	SOLDDATE	EXPIREDDATE	DOM	DTO
0	71876727	SLD	149000.0	149000.0	2015-07-21	2016-01-01	NaN	144	137.0
1	71882139	SLD	399903.0	397000.0	2015-07-31	2016-01-01	NaN	130	111.0
2	71892702	SLD	125000.0	125000.0	2015-08-20	2016-01-01	NaN	16	16.0
3	71900538	SLD	479900.0	450000.0	2015-09-08	2016-01-01	NaN	73	52.0
4	71906517	SLD	419900.0	405000.0	2015-09-18	2016-01-01	NaN	106	50.0

5 rows × 45 columns

Multi-family

Out[6]:

	MLSNUM	STATUS	LISTPRICE	SOLDPRICE	LISTDATE	SOLDDATE	EXPIREDDATE	DOM	DTO
0	71590427	SLD	80000	70000.0	2013-09-30	2016-01-29	NaN	269	242.0
1	71661835	SLD	199900	135000.0	2014-04-16	2016-01-20	NaN	760	443.0
2	71695436	SLD	508888	455000.0	2014-06-09	2016-01-22	NaN	469	468.0
3	71701476	SLD	394000	394000.0	2014-06-19	2016-01-06	NaN	596	338.0
4	71735686	SLD	449900	380000.0	2014-08-30	2016-01-28	NaN	514	514.0

5 rows × 45 columns

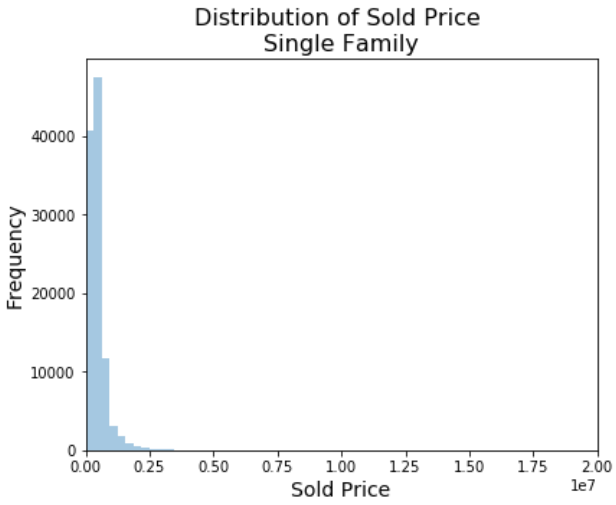
Condo

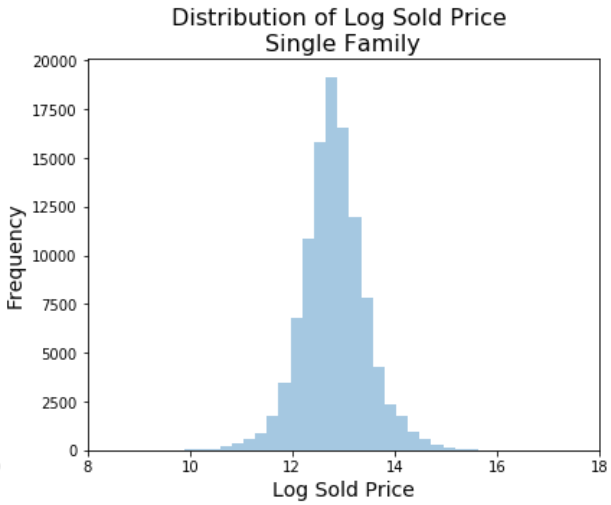
Out[7]:

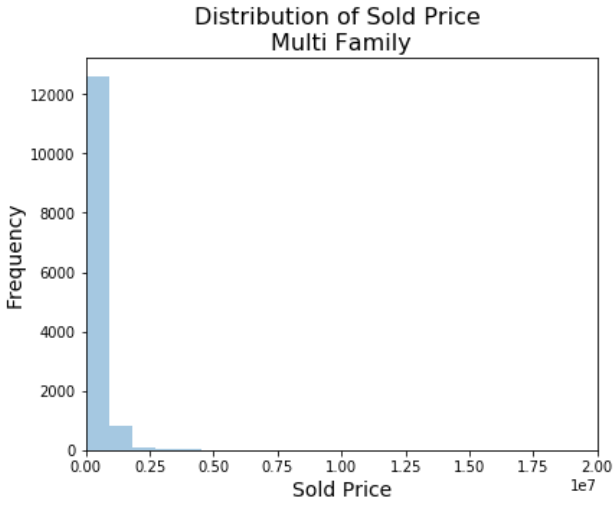
	MLSNUM	STATUS	LISTPRICE	SOLDPRICE	LISTDATE	SOLDDATE	EXPIREDDATE	DOM	DTO
0	71498924	SLD	169900.0	177500.0	2013-03-26	2016-01-06	NaN	709	618.0
1	71500500	SLD	279900.0	295850.0	2013-03-30	2016-01-29	NaN	897	870.0
2	71595485	SLD	449900.0	470519.0	2013-10-10	2016-01-14	NaN	658	658.0
3	71599223	SLD	389500.0	383500.0	2013-10-21	2016-01-22	NaN	804	759.0
4	71661544	SLD	237000.0	232000.0	2014-04-15	2016-01-07	NaN	467	467.0

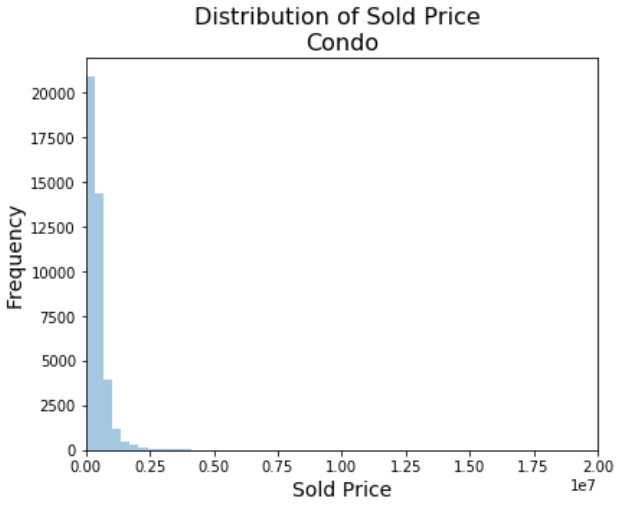
5 rows × 45 columns

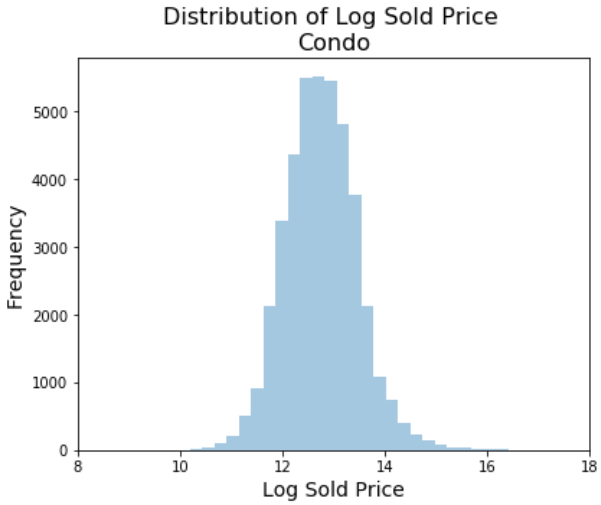
Distribution of Sold Price (Response)

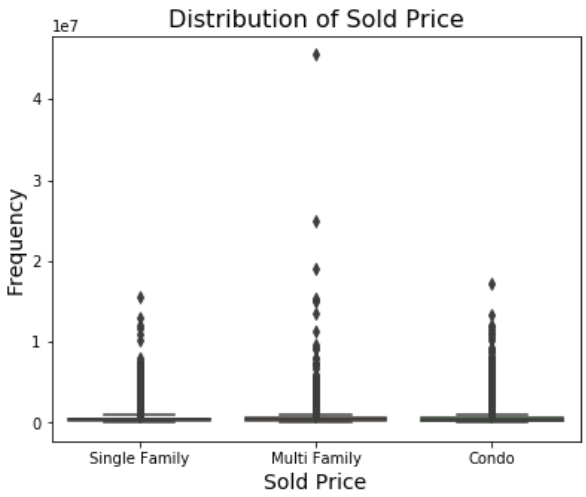


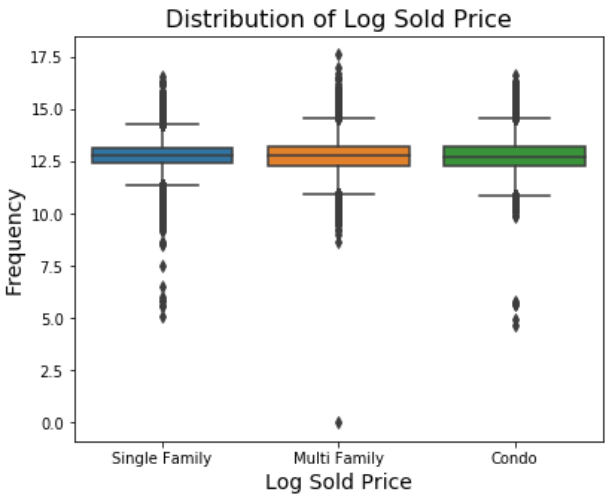






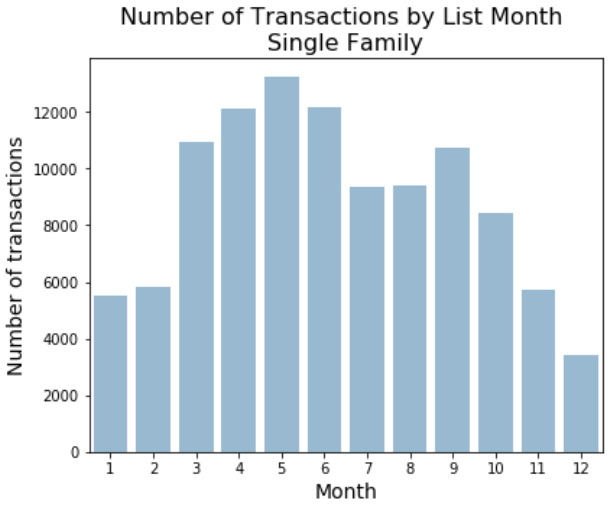


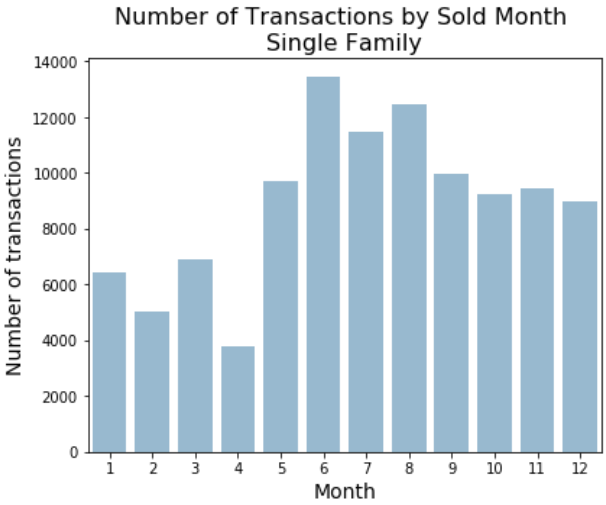


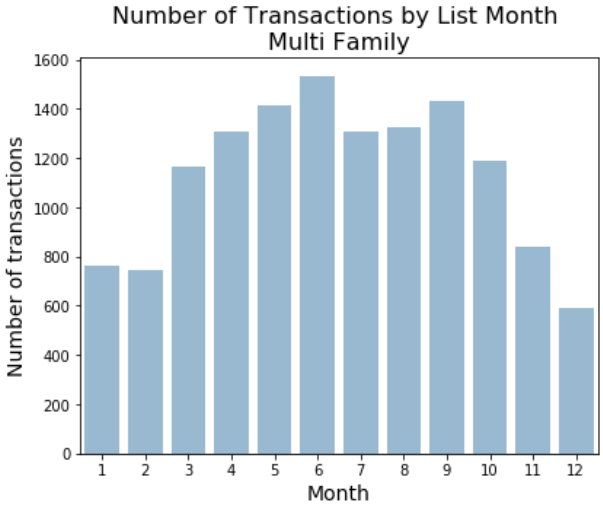


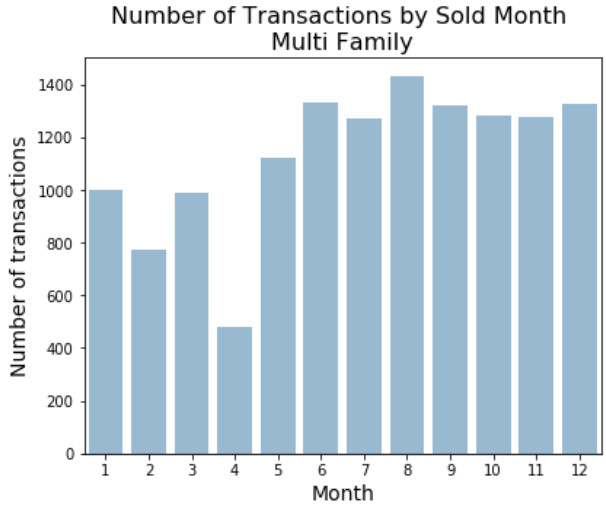
Distribution of sold price is very right skewed. Log transformation reduces the skewness, making the distribution more symmetrical. The sold price for each property type roughly follows log-normal distribution.

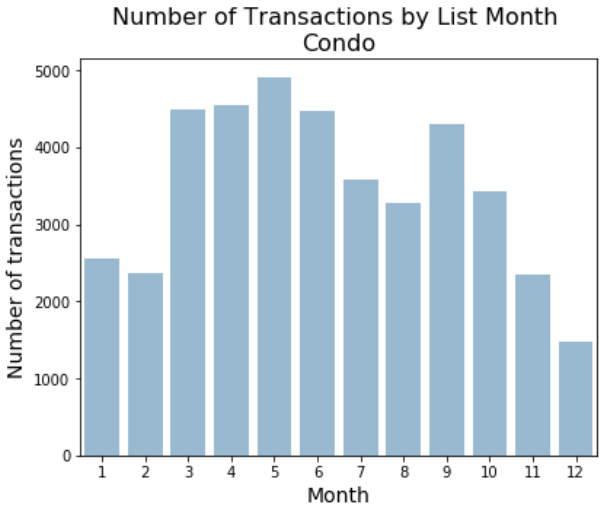
Time

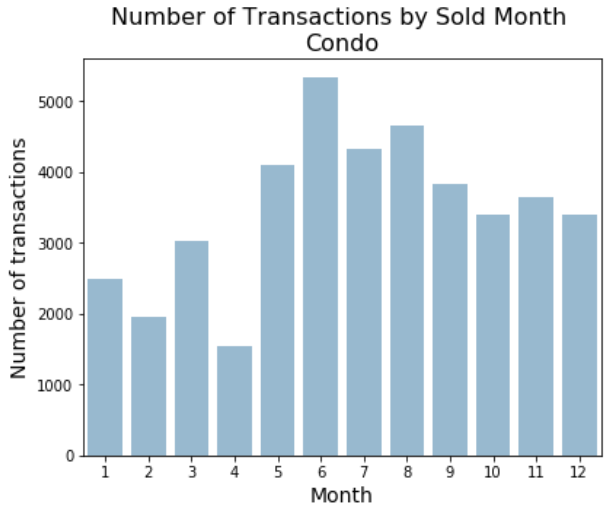

Number of Transactions by List Month
Single Family
Number of transactions
Month

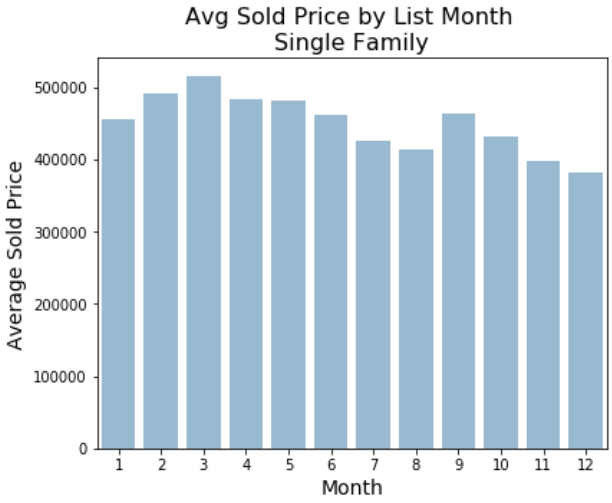

Number of Transactions by Sold Month
Single Family
Number of transactions
Month

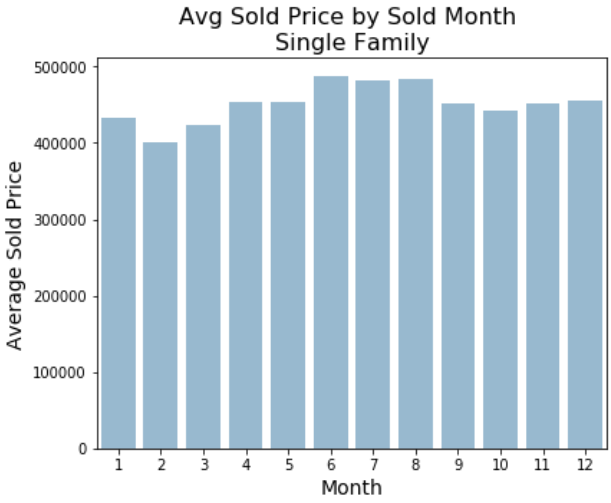

Number of Transactions by List Month
Multi Family
Number of transactions
Month

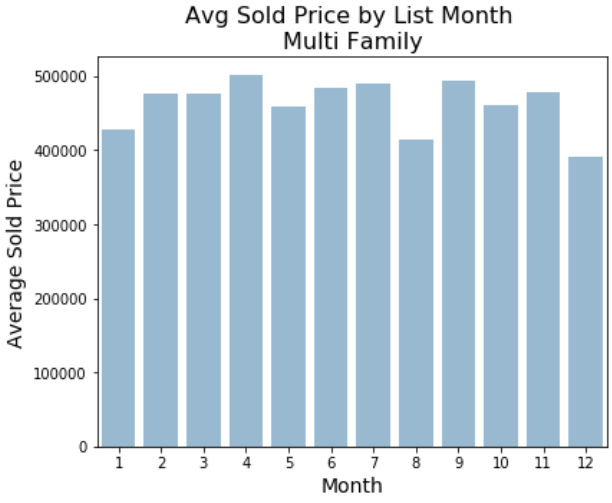

Number of Transactions by Sold Month
Multi Family
Number of transactions
Month

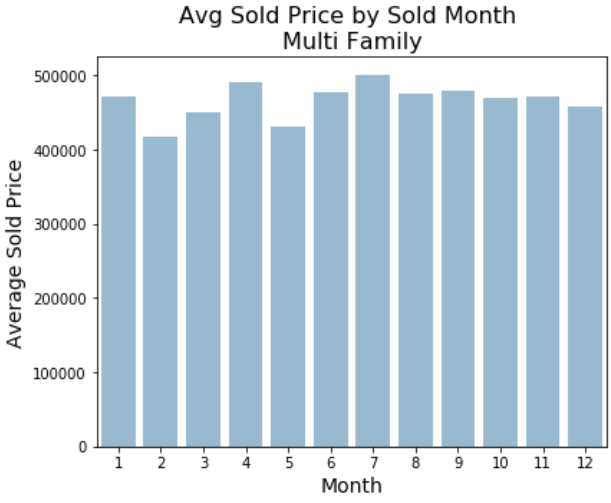

Number of Transactions by List Month
Condo
Number of transactions
Month

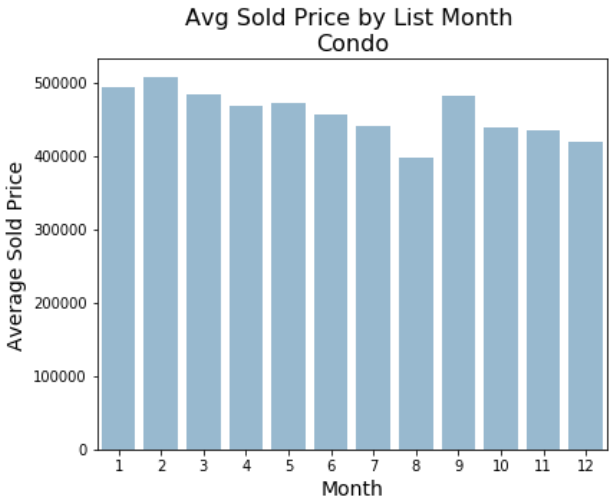

Number of Transactions by Sold Month
Condo
Number of transactions
Month

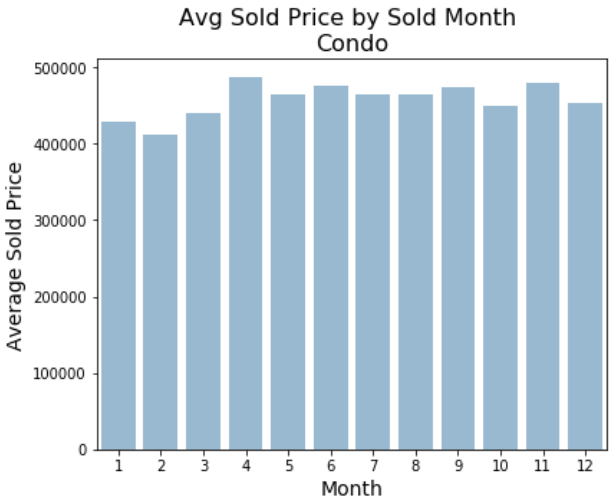

Avg Sold Price by List Month
Single Family
Average Sold Price
Month

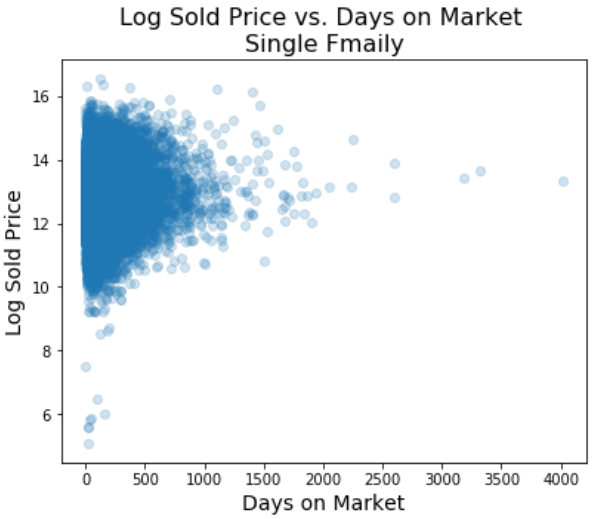

Avg Sold Price by Sold Month
Single Family
Average Sold Price
Month

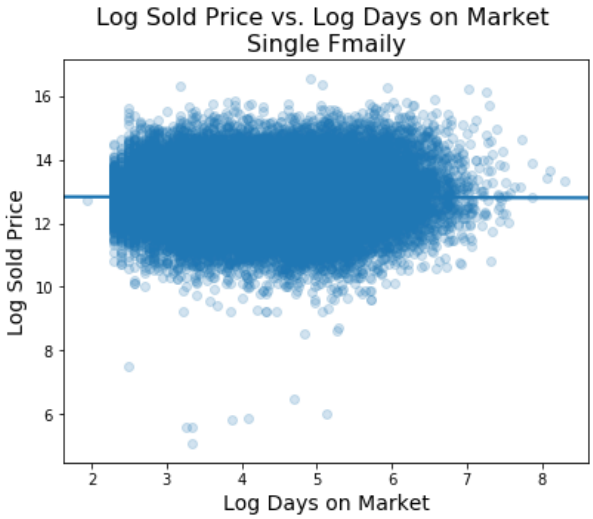

Avg Sold Price by List Month
Multi Family
Average Sold Price
Month

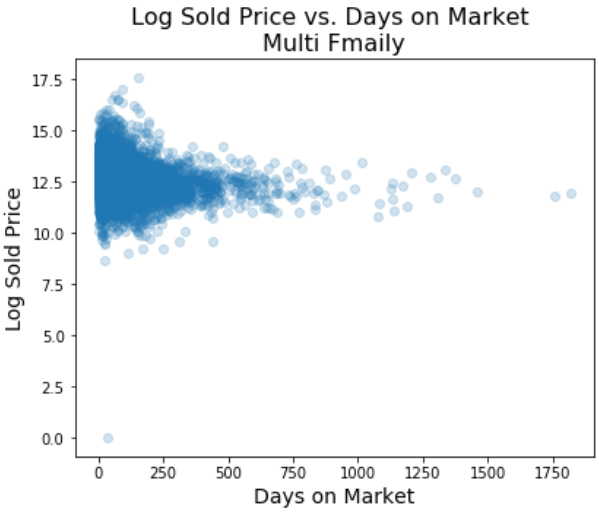

Avg Sold Price by Sold Month
Multi Family
Average Sold Price
Month

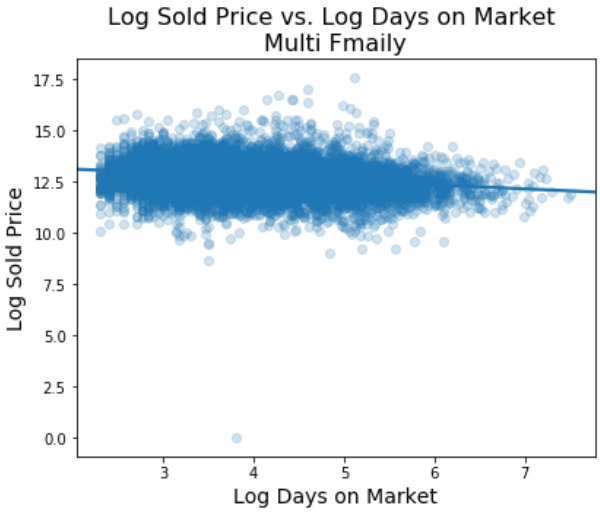

Avg Sold Price by List Month
Condo
Average Sold Price
Month

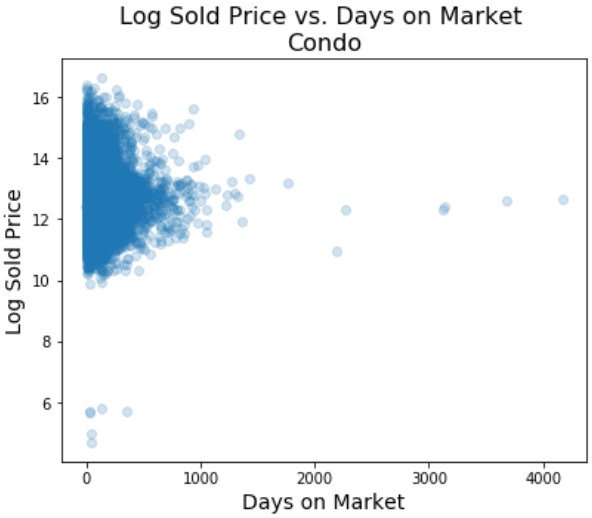

Avg Sold Price by Sold Month
Condo
Average Sold Price
Month

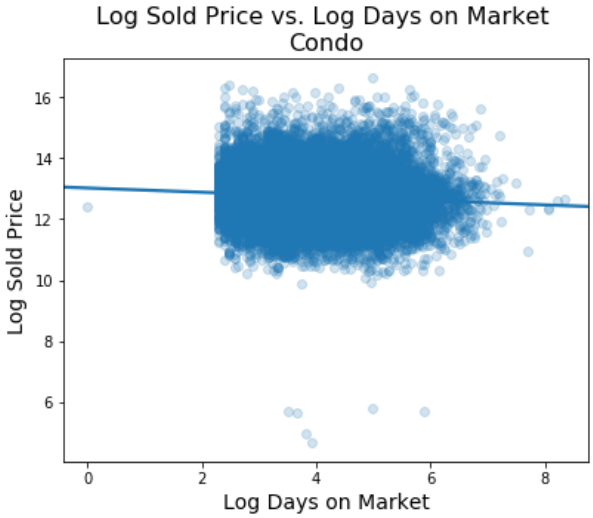






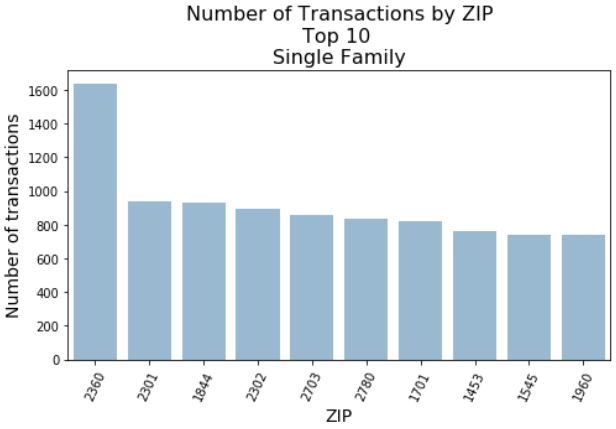


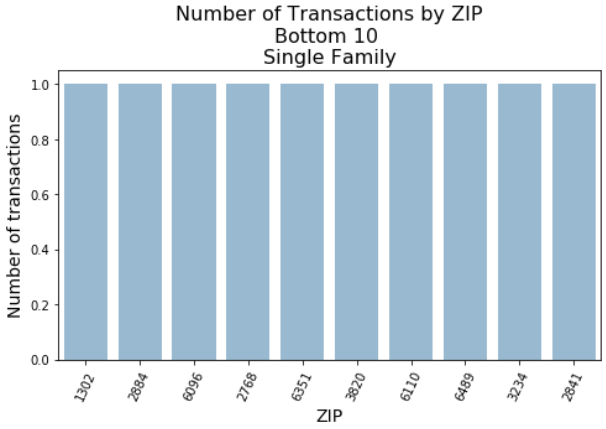


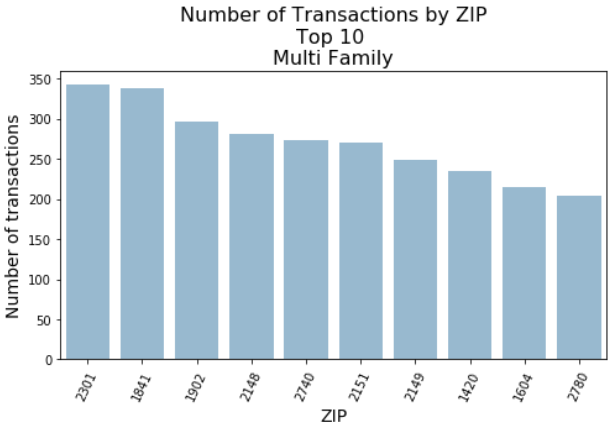


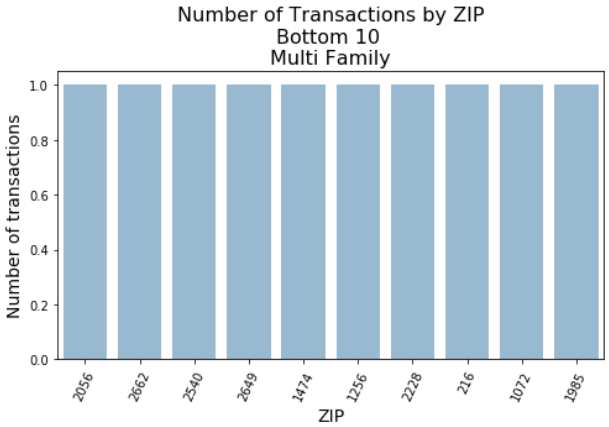
Most of the homeowners put their houses on the market in Spring, and people tend to buy houses in Summer. Houses listed in August and Winter tend to be sold for a lower price.

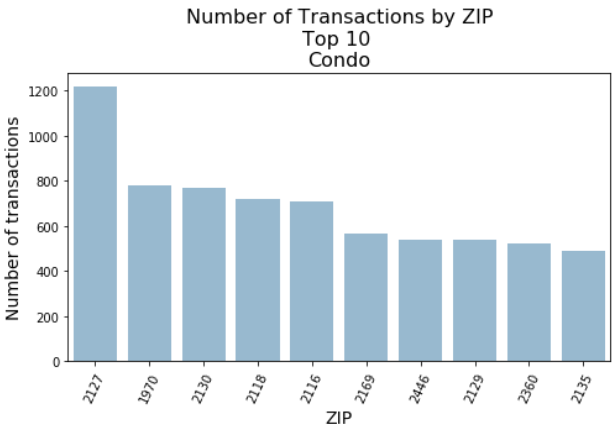
Location

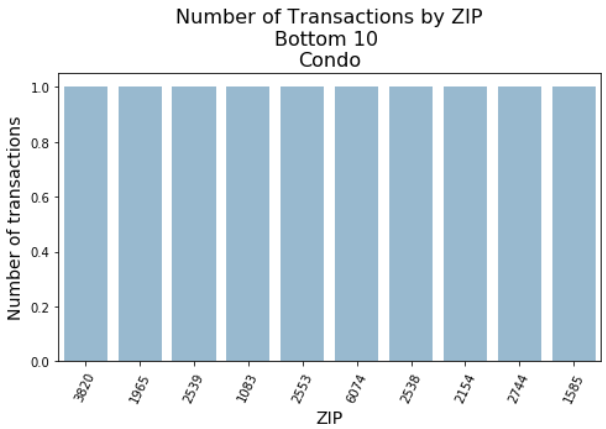

Number of Transactions by ZIP
Top 10
Single Family
Number of transactions
ZIP
2360
2301
1844
2302
2703
2780
1701
1453
1545
1960

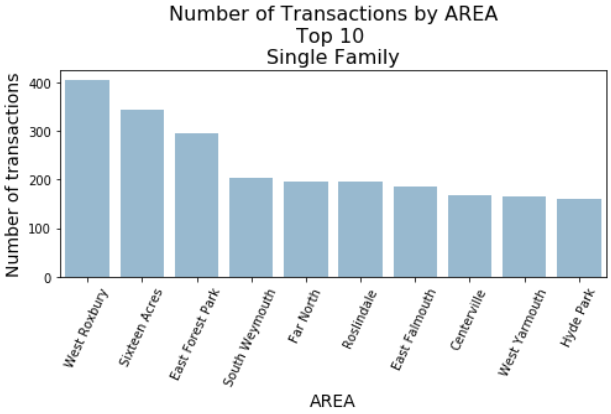

Number of Transactions by ZIP
Bottom 10
Single Family
Number of transactions
ZIP
1302
2884
6096
2768
6351
3820
6110
6489
3234
2841

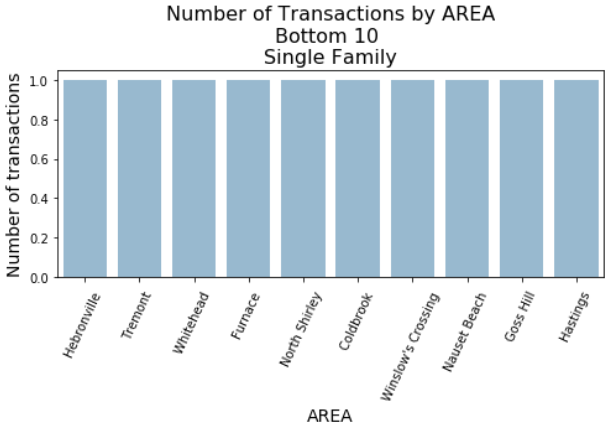

Number of Transactions by ZIP
Top 10
Multi Family
Number of transactions
ZIP
2301
1841
1902
2148
2740
2151
2149
1420
1604
2780

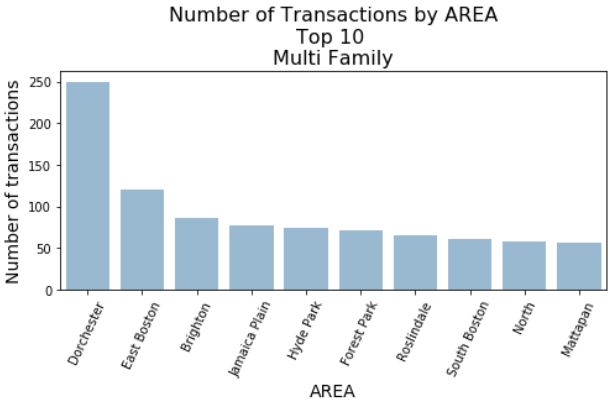

Number of Transactions by ZIP
Bottom 10
Multi Family
Number of transactions
ZIP
2056
2662
2540
2649
1474
1256
2228
216
1072
1985

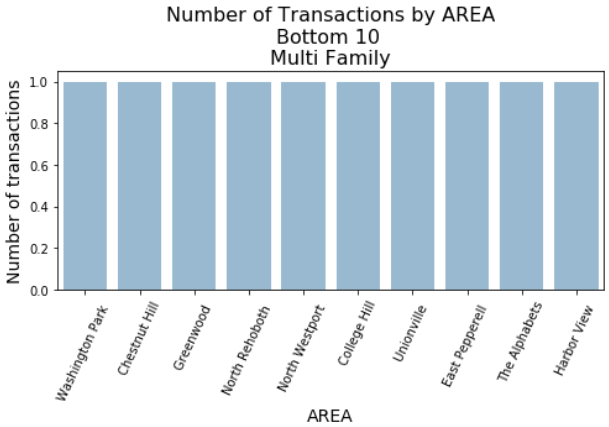

Number of Transactions by ZIP
Top 10
Condo
Number of transactions
ZIP
2127
1970
2130
2118
2116
2169
2446
2129
2360
2135

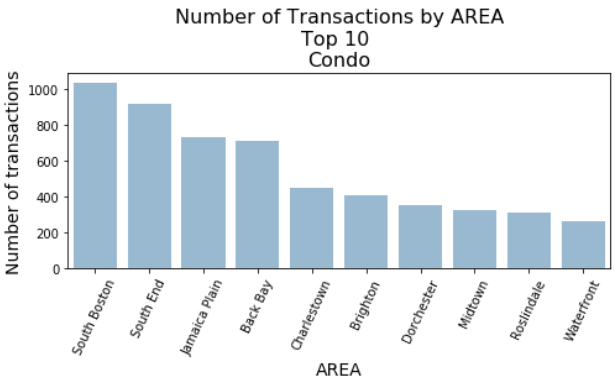

Number of Transactions by ZIP
Bottom 10
Condo
Number of transactions
ZIP
3820
1965
2539
1083
2553
6074
2538
2154
2744
1585

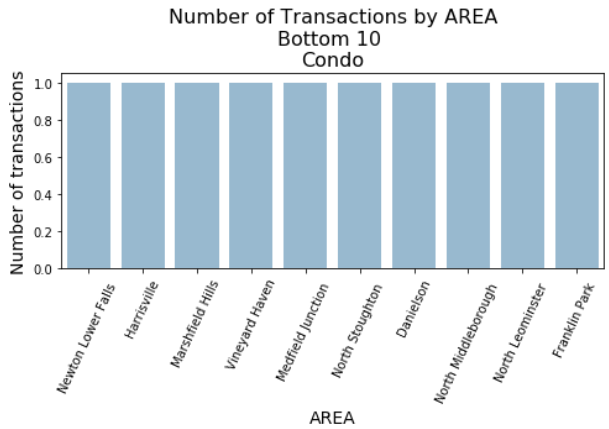

Number of Transactions by AREA
Top 10
Single Family
Number of transactions
AREA
West Roxbury
Sixteen Acres
East Forest Park
South Weymouth
Far North
Roslindale
East Falmouth
Centerville
West Yarmouth
Hyde Park

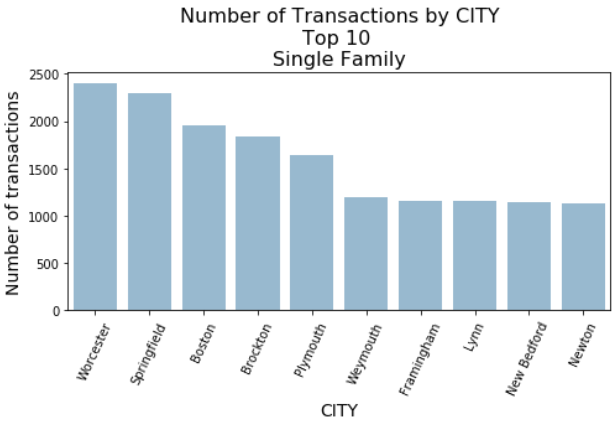

Number of Transactions by AREA
Bottom 10
Single Family
Number of transactions
AREA
Hebronville
Tremont
Whitehead
Furnace
North Shirley
Coldbrook
Winslow's Crossing
Nauset Beach
Goss Hill
Hastings

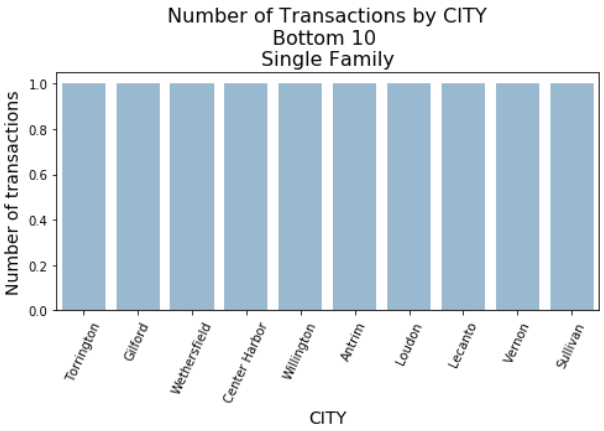

Number of Transactions by AREA
Top 10
Multi Family
Number of transactions
AREA
Dorchester
East Boston
Brighton
Jamaica Plain
Hyde Park
Forest Park
Roslindale
South Boston
North
Mattapan

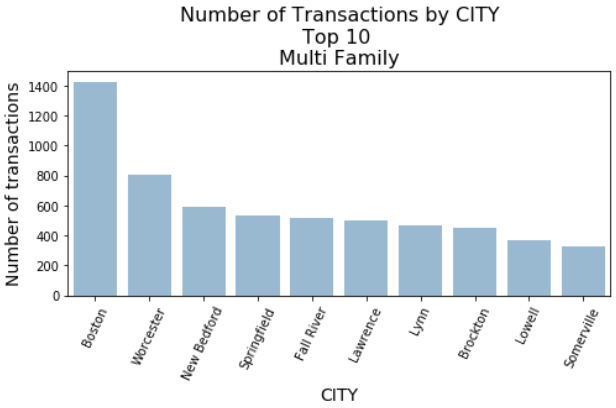

Number of Transactions by AREA
Bottom 10
Multi Family
Number of transactions
AREA
Washington Park
Chestnut Hill
Greenwood
North Rehoboth
North Westport
College Hill
Unionville
East Pepperell
The Alphabets
Harbor View

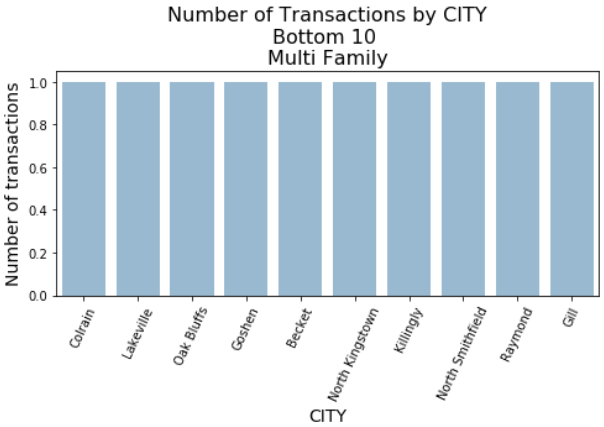

Number of Transactions by AREA
Top 10
Condo
Number of transactions
AREA
South Boston
South End
Jamaica Plain
Back Bay
Charlestown
Brighton
Dorchester
Midtown
Roslindale
Waterfront

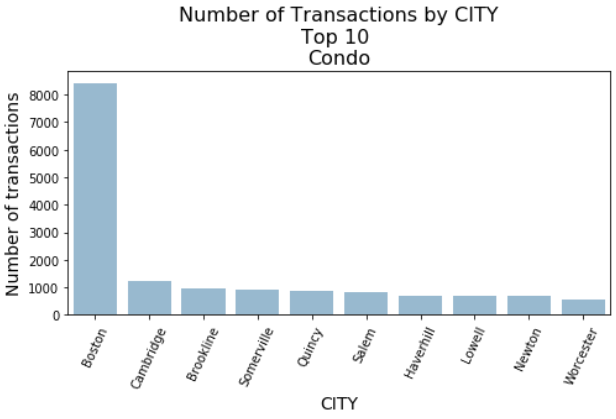

Number of Transactions by AREA
Bottom 10
Condo
Number of transactions
AREA
Newton Lower Falls
Harrisville
Marshfield Hills
Vineyard Haven
Medfield Junction
North Stoughton
Danielson
North Middleborough
North Leominster
Franklin Park

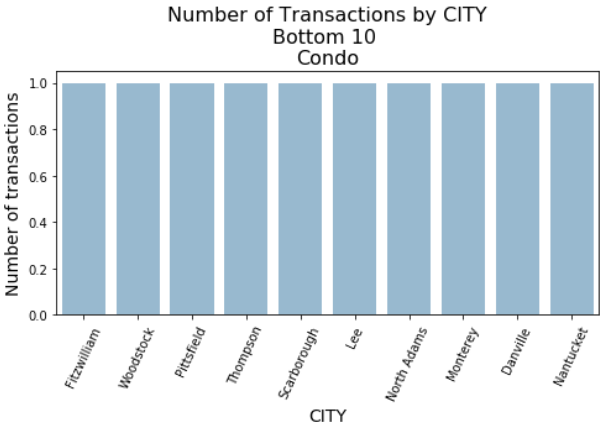

Number of Transactions by CITY
Top 10
Single Family
Number of transactions
2500
2000
1500
1000
500
0
Worcester
Springfield
Boston
Brockton
Plymouth
Weymouth
Framingham
Lynn
New Bedford
Newton
CITY

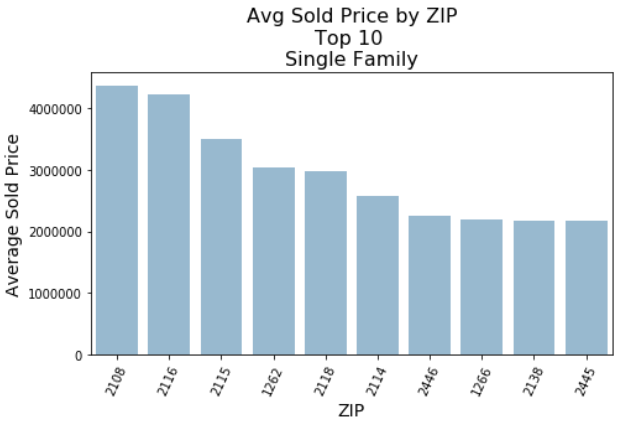

Number of Transactions by CITY
Bottom 10
Single Family
Number of transactions
1.0
0.8
0.6
0.4
0.2
0.0
Torrington
Gilford
Wethersfield
Center Harbor
Willington
Antrim
Loudon
Lecanto
Vernon
Sullivan
CITY

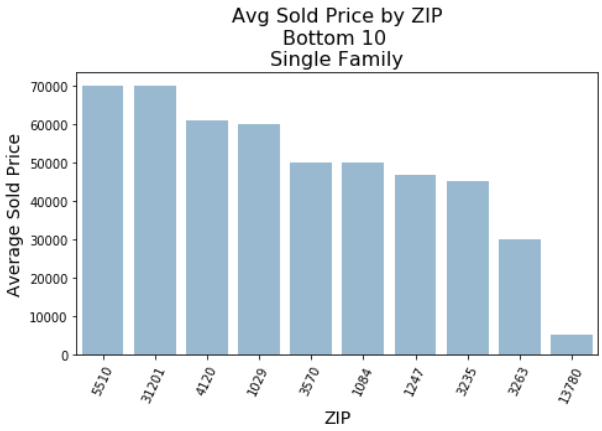

Number of Transactions by CITY
Top 10
Multi Family
Number of transactions
1400
1200
1000
800
600
400
200
0
Boston
Worcester
New Bedford
Springfield
Fall River
Lawrence
Lynn
Brockton
Lowell
Somerville
CITY

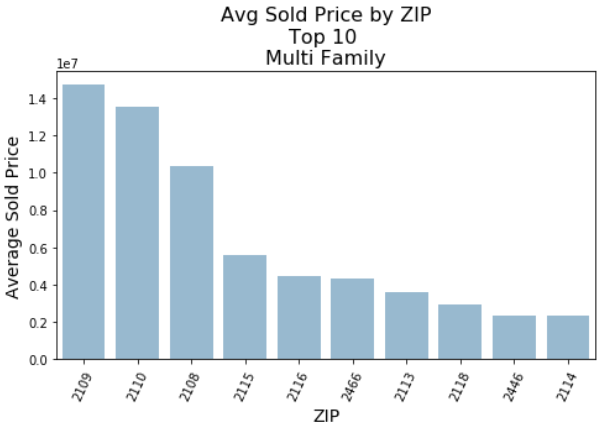

Number of Transactions by CITY
Bottom 10
Multi Family
Number of transactions
1.0
0.8
0.6
0.4
0.2
0.0
Colrain
Lakeville
Oak Bluffs
Goshen
Becket
North Kingstown
Killingly
North Smithfield
Raymond
Gill
CITY

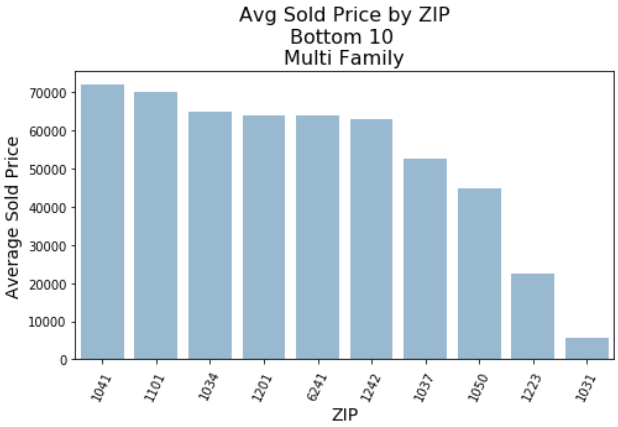

Number of Transactions by CITY
Top 10
Condo
Number of transactions
8000
7000
6000
5000
4000
3000
2000
1000
0
Boston
Cambridge
Brookline
Somerville
Quincy
Salem
Haverhill
Lowell
Newton
Worcester
CITY

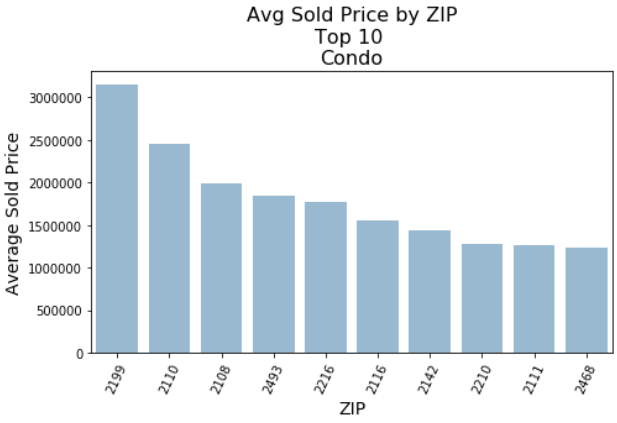

Number of Transactions by CITY
Bottom 10
Condo
Number of transactions
1.0
0.8
0.6
0.4
0.2
0.0
Fitzwilliam
Woodstock
Pittsfield
Thompson
Scarborough
Lee
North Adams
Monterey
Danville
Nantucket
CITY

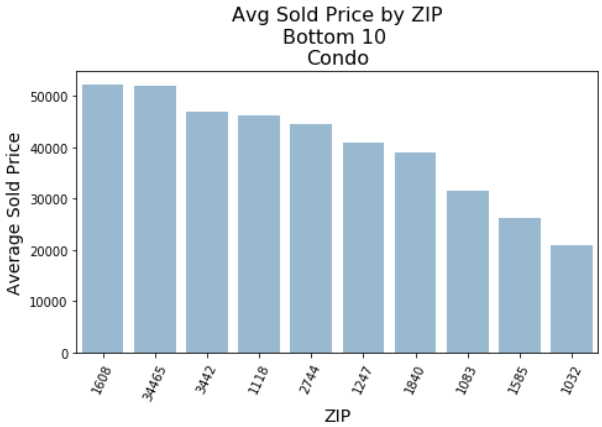

Avg Sold Price by ZIP
Top 10
Single Family
Average Sold Price
ZIP
2108
2116
2115
1262
2118
2114
2446
1266
2138
2445

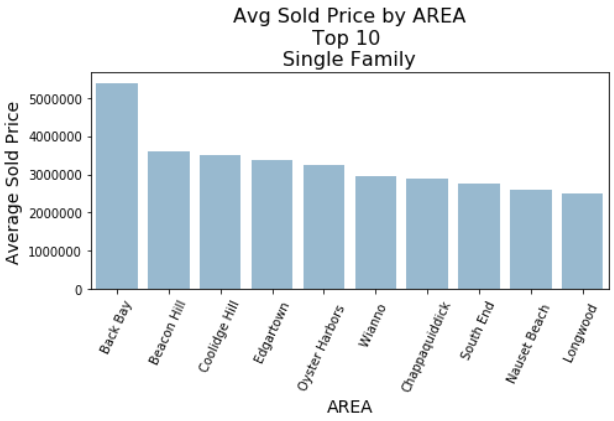

Avg Sold Price by ZIP
Bottom 10
Single Family
Average Sold Price
ZIP
5510
31201
4120
1029
3570
1084
1247
3235
3263
13780

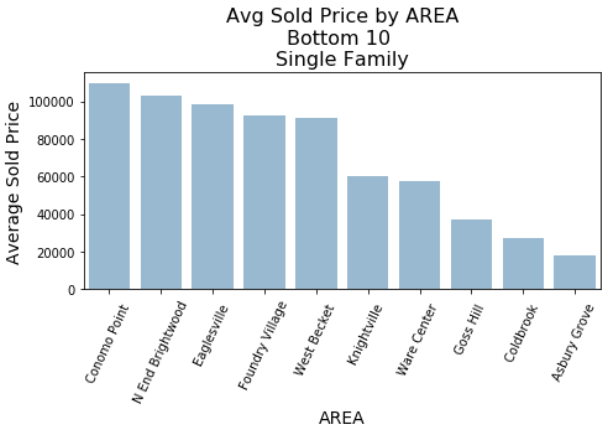

Avg Sold Price by ZIP
Top 10
Multi Family
1e7
Average Sold Price
ZIP
2109
2110
2108
2115
2116
2466
2113
2118
2446
2114

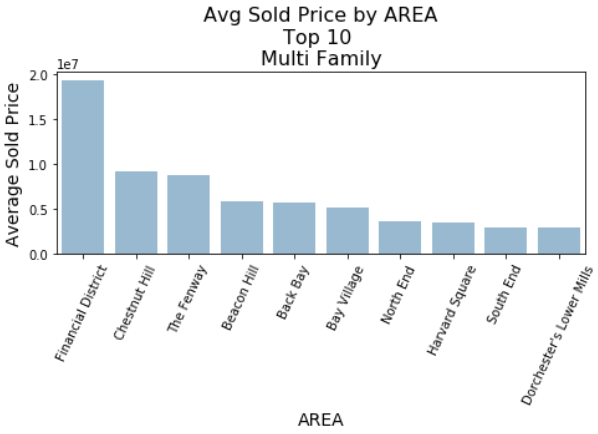

Avg Sold Price by ZIP
Bottom 10
Multi Family
Average Sold Price
ZIP
1041
1101
1034
1201
6241
1242
1037
1050
1223
1031

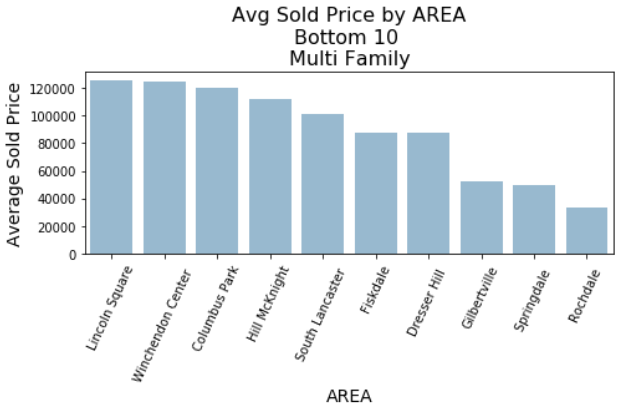

Avg Sold Price by ZIP
Top 10
Condo
Average Sold Price
ZIP
2199
2110
2108
2493
2216
2116
2142
2210
2111
2468


Avg Sold Price by ZIP
Bottom 10
Condo
Average Sold Price
ZIP
1608
34465
3442
1118
2744
1247
1840
1083
1585
1032

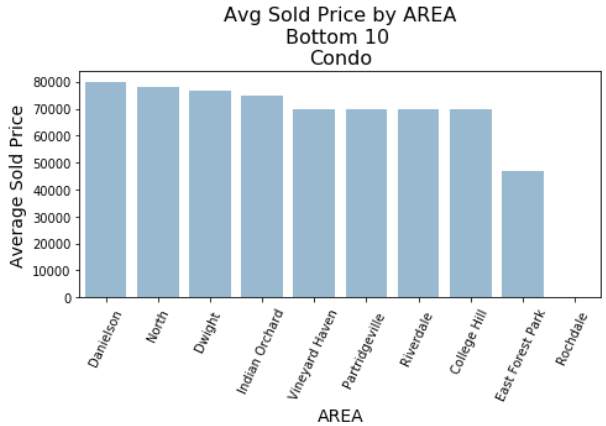

Avg Sold Price by AREA
Top 10
Single Family
Average Sold Price
5000000
4000000
3000000
2000000
1000000
0
Back Bay
Beacon Hill
Coolidge Hill
Edgartown
Oyster Harbors
Wianno
Chappaquiddick
South End
Nauset Beach
Longwood
AREA

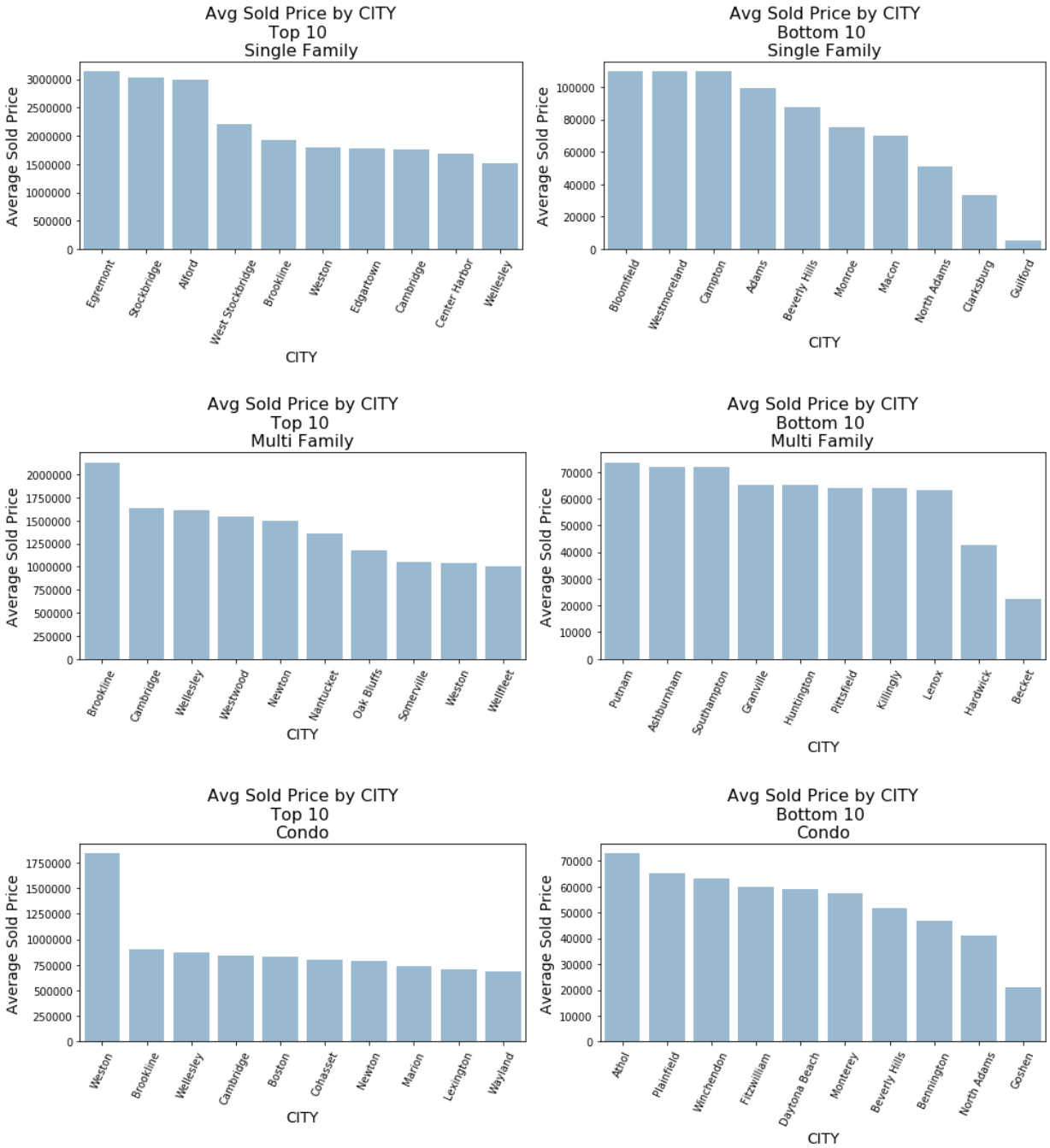

Avg Sold Price by AREA
Bottom 10
Single Family
Average Sold Price
100000
80000
60000
40000
20000
0
Conomo Point
N End Brightwood
Eaglesville
Foundry Village
West Becket
Knightville
Ware Center
Goss Hill
Coldbrook
Asbury Grove
AREA


Avg Sold Price by AREA
Top 10
Multi Family
1e7
Average Sold Price
2.0
1.5
1.0
0.5
0.0
Financial District
Chestnut Hill
The Fenway
Beacon Hill
Back Bay
Bay Village
North End
Harvard Square
South End
Dorchester's Lower Mills
AREA


Avg Sold Price by AREA
Bottom 10
Multi Family
Average Sold Price
120000
100000
80000
60000
40000
20000
0
Lincoln Square
Winchendon Center
Columbus Park
Hill McKnight
South Lancaster
Fiskdale
Dresser Hill
Gilbertville
Springdale
Rochdale
AREA

Avg Sold Price by AREA
Top 10
Condo
Average Sold Price
2000000
1500000
1000000
500000
0
Midtown
Back Bay
Beacon Hill
Agassiz
Waterfront
Seaport District
Wabian
South Hingham
South End
Harvard Square
AREA

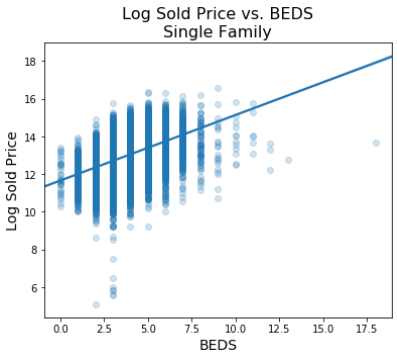

Avg Sold Price by AREA
Bottom 10
Condo
Average Sold Price
80000
70000
60000
50000
40000
30000
20000
10000
0
Danielson
North
Dwight
Indian Orchard
Vineyard Haven
Partridgeville
Riverdale
College Hill
East Forest Park
Rochdale
AREA

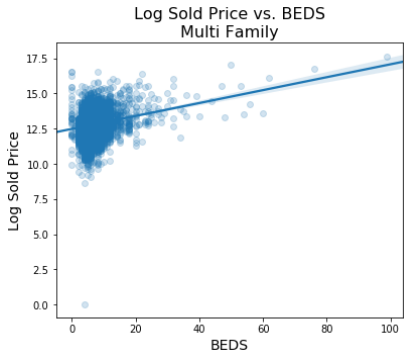


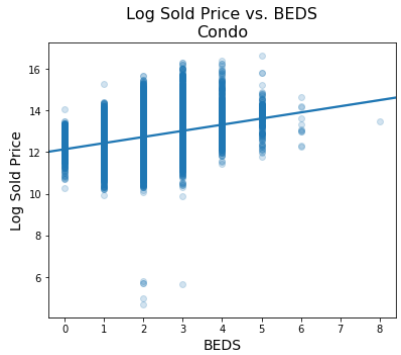
Number of transactions varies a lot for different areas. We summed up the number of transactions by zipcode, area and city. The number ranges from only 1 transaction to over 1k transactions. The type of properties in different area varies a lot. For example, it seems that most properties in Back Bay are condos while West Roxbury is dominated by single family houses.

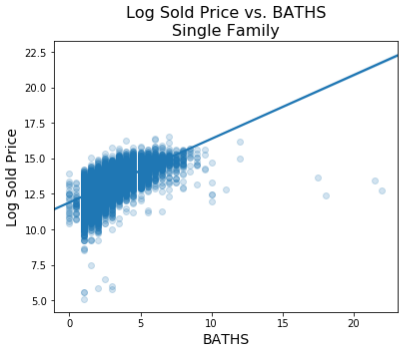
The price of houses is highly dependent on the area it's located in. For example, single family houses in Back Bay and Beacon Hill are hundreds times the price of those in Coldbrook and Asbury Grove. Multifamily houses in financial districts are two times more expensive than the neighbourhood that ranked the second expensive. Condos in Midtown, Back Bay and Beacon Hill are way more expensive than those in East Forest.

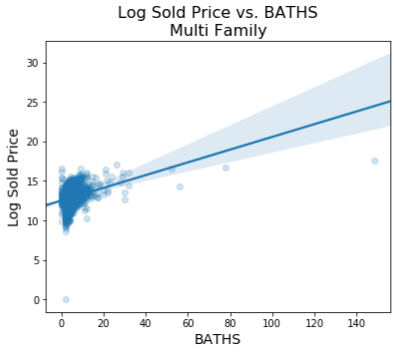
House features

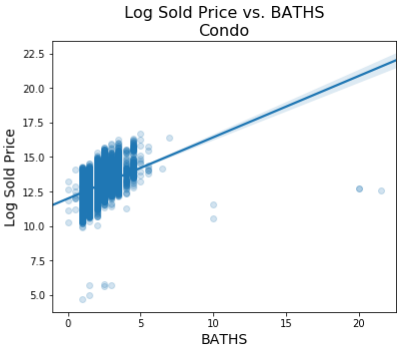

Log Sold Price vs. BEDS
Single Family
Log Sold Price
BEDS

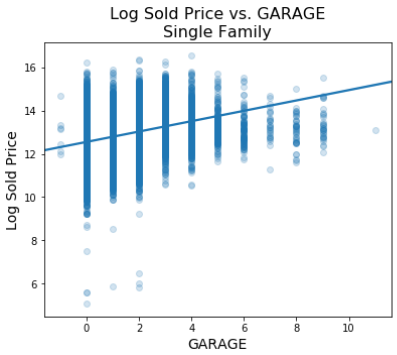

Log Sold Price vs. BEDS
Multi Family
Log Sold Price
BEDS

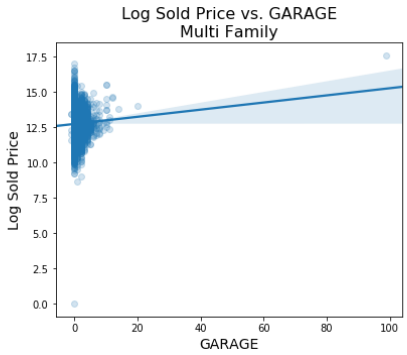

Log Sold Price vs. BEDS
Condo
Log Sold Price
BEDS

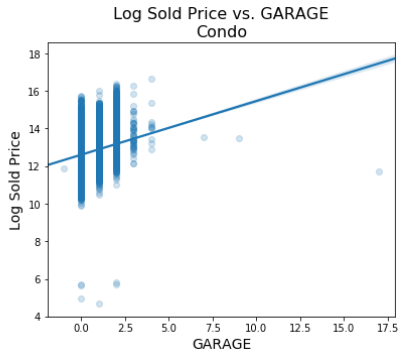

Log Sold Price vs. BATHS
Single Family
Log Sold Price
BATHS

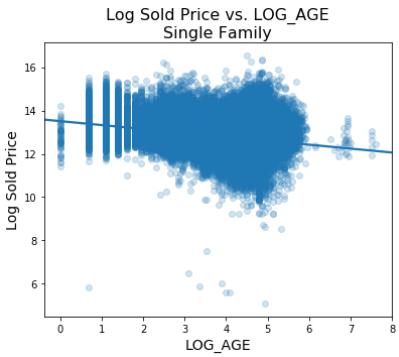

Log Sold Price vs. BATHS
Multi Family
Log Sold Price
BATHS


Log Sold Price vs. BATHS
Condo
Log Sold Price
BATHS

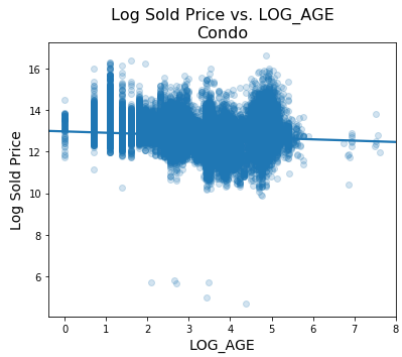

Log Sold Price vs. GARAGE
Single Family
Log Sold Price
GARAGE

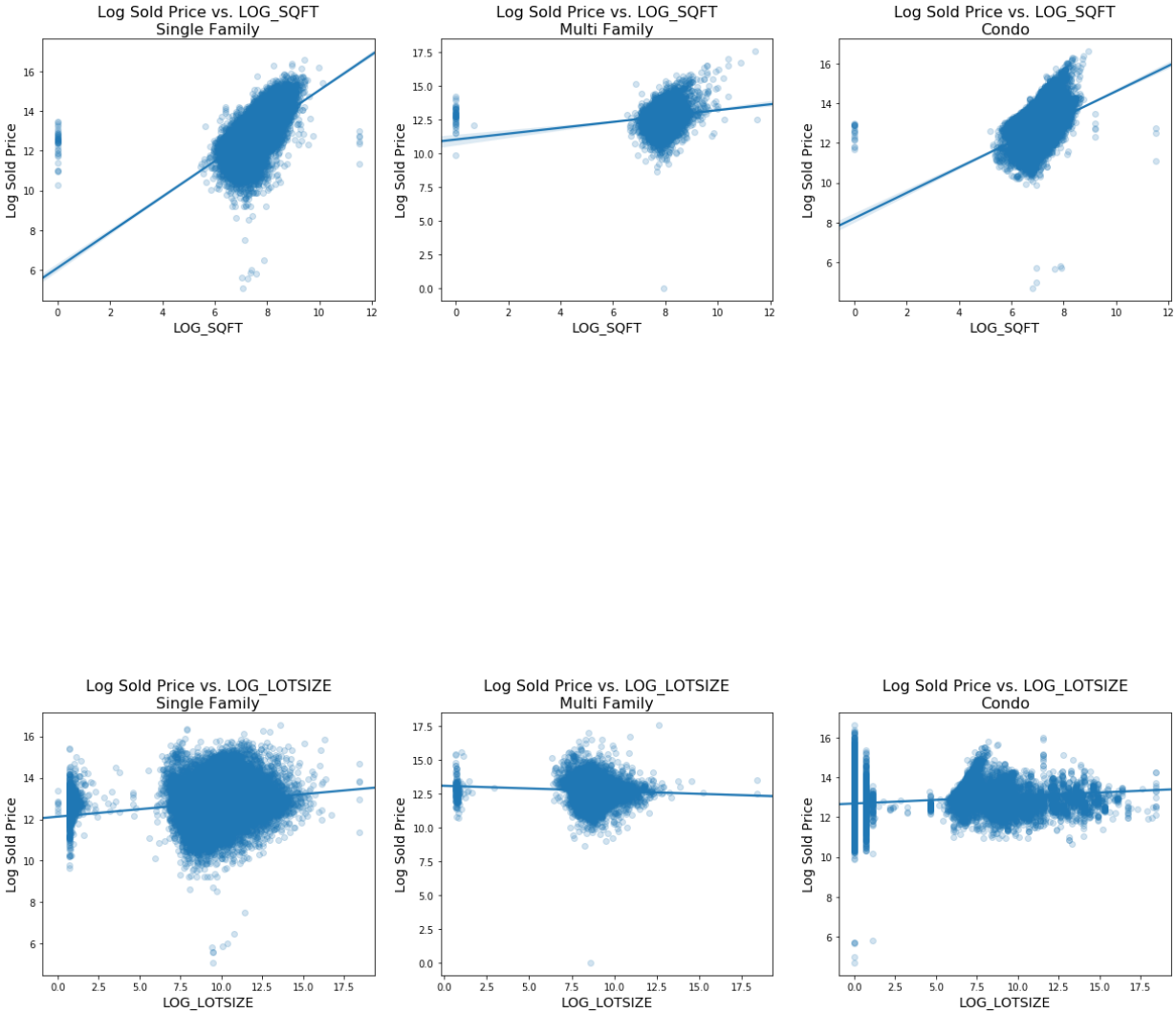

Log Sold Price vs. GARAGE
Multi Family
Log Sold Price
GARAGE


Log Sold Price vs. GARAGE
Condo
Log Sold Price
GARAGE


Log Sold Price vs. LOG_AGE
Single Family
Log Sold Price
LOG_AGE

Log Sold Price vs. LOG_AGE
Multi Family
Log Sold Price
LOG_AGE


Log Sold Price vs. LOG_AGE
Condo
Log Sold Price
LOG_AGE



From the plots above, we can observe that the log of the sold price is positively correlated with the number of bedrooms, bathrooms and garages. Newer houses tend to be sold for a higher price, even though the relationship between age and price is not very strong for multi-family houses and condos. Larger houses are more expensive, but the lot size does not affect the price as much as the size of the house itself.