PYTHON WEB SCRAPING

• • •

By Sabbir Mohammed

A brief look at NYC Real Estate

About the project:

- Using scrapy on www.trulia.com

- Site offers information on 13K+ listings for recent sales over the last nine months (May 2018 - Jan 2019)

- What variables about the property factor into the final sale price and to what extent?

- Original plan to scrape site for the following variables/ features per listing:
  - Property Type
  - Property Age
  - Sale Price
  - Sale Date
  - Area (in sqft.)
  - Street Address
  - City
  - Postal Code
  - Borough

Screenshot: <https://www.trulia.com/sold/New_York,NY/>

A peek into the data:

	address	city	soldDate	soldPrice	sqft	zipCode	borough
0	245 E 93rd St #26C	New York	21-Dec-18	1399999.0	1056	10128	Manhattan
1	15 Broad St #2320	New York	21-Dec-18	2150000.0	1772	10005	Manhattan
2	510 W 110th St #12D	New York	21-Dec-18	660000.0	616	10025	Manhattan
3	405 W 50th St	New York	21-Dec-18	NaN	9775	10019	Manhattan
4	407 W 50th St	New York	21-Dec-18	NaN	9775	10019	Manhattan
5	155 W 70th St #12E	New York	21-Dec-18	950000.0	1302	10023	Manhattan
6	200 E 66th St #E2006	New York	21-Dec-18	NaN	955	10065	Manhattan
7	200 E 24th St #1809	New York	21-Dec-18	650000.0	158337	10010	Manhattan
8	181 E 90th St #26A	New York	21-Dec-18	3200000.0	1948	10128	Manhattan
9	58 W 129th St #2C	New York	21-Dec-18	494000.0	479	10027	Manhattan

13,267 listings in total

1,952 missing Sale Prices

Highest sale: **$44,580,000**
(home in Staten Island)

Largest Properties: **883,265 sq.ft.**
(2 listings: Brooklyn, NY 11201)

Overview:

	soldPrice	sqft	PricePerSQFT
count	11315.00	11315.00	11315.00
mean	880200.73	6056.46	573.92
std	858014.41	35225.27	765.58
min	10.00	224.00	0.00
25%	520000.00	1138.50	325.42
50%	705000.00	1524.00	451.13
75%	970000.00	2160.00	661.39
max	44580000.00	883265.00	68796.30

LOCATION

Summary Statistics by Borough:

	soldPrice			sqft			PricePerSQFT		
	count	mean	std	count	mean	std	count	mean	std
borough									
Bronx	1273	530538.121	265929.544	1273	2383.314	6412.720	1273	292.779	113.847
Brooklyn	3047	1064731.482	735927.796	3047	7136.969	42829.347	3047	618.849	361.348
Manhattan	1097	1716700.270	1616463.039	1097	11426.018	45640.966	1097	1333.850	678.809
Queens	3937	765146.343	332193.463	3937	7116.907	39106.759	3937	503.325	214.225
Staten Island	1961	583507.397	1020134.015	1961	1629.237	751.918	1961	403.258	1550.905

The Five Boroughs: Number of Sales
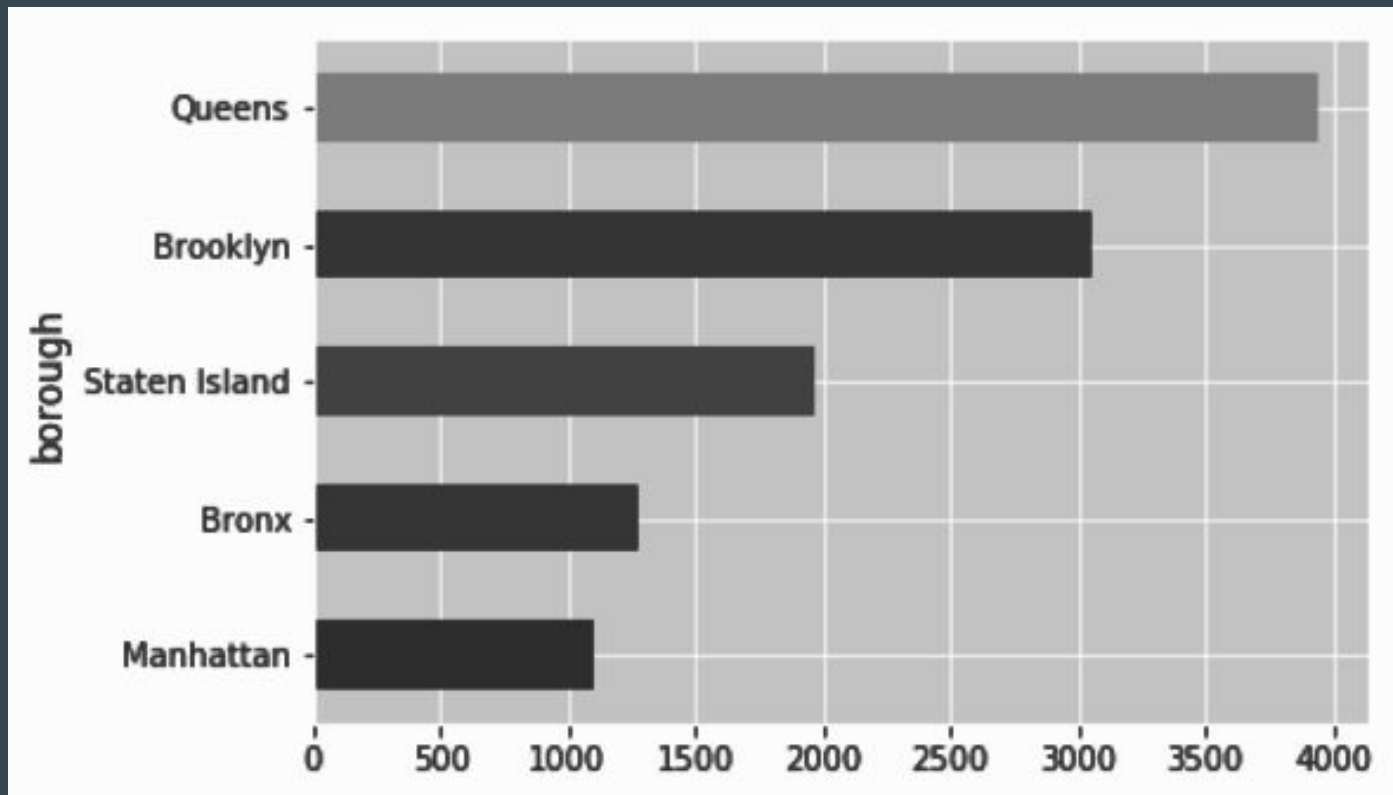
NYC Population per Borough (US Census Bureau, 2010):

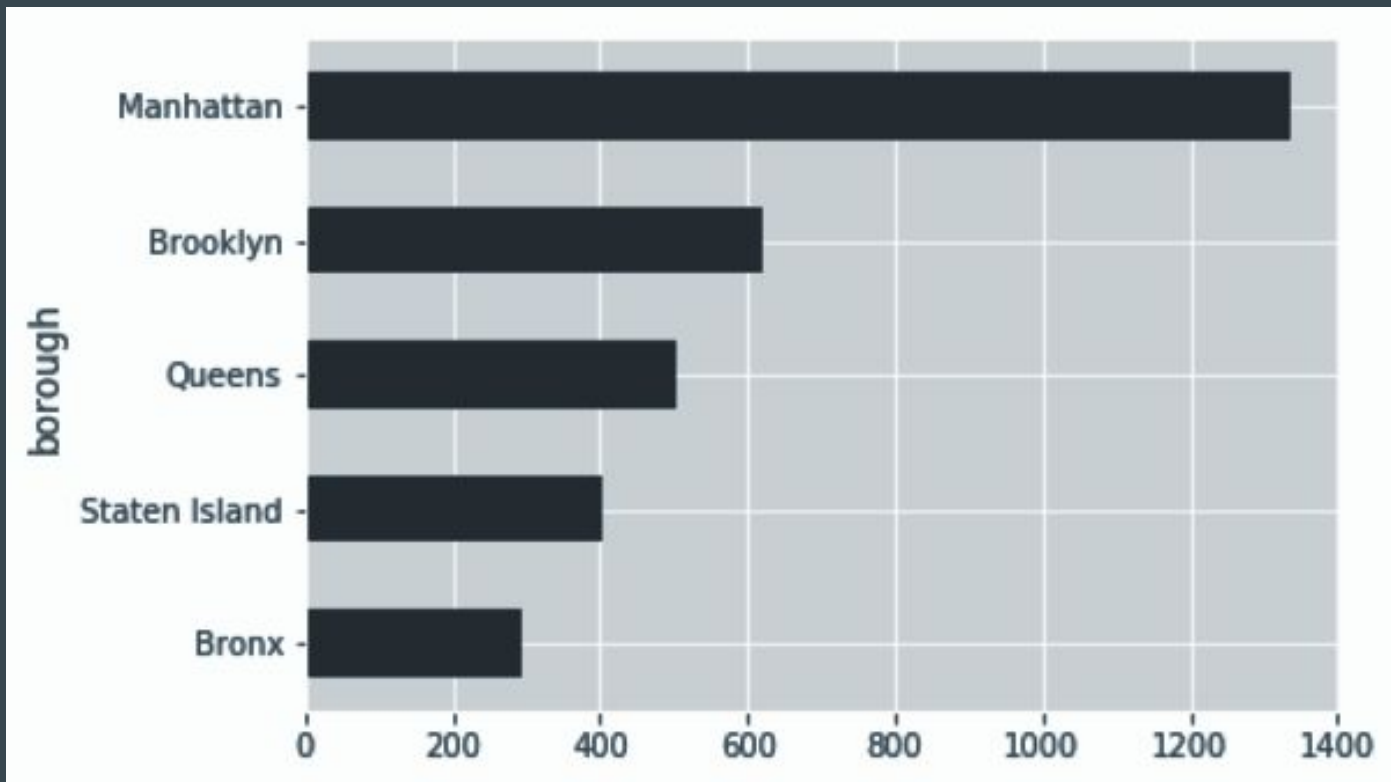| Borough	Population (2010)
Brooklyn	2,504,706
Queens	2,230,545
Manhattan	1,586,184
Bronx	1,384,794
Staten Island	468,730

The Five Boroughs: Average Sale Price
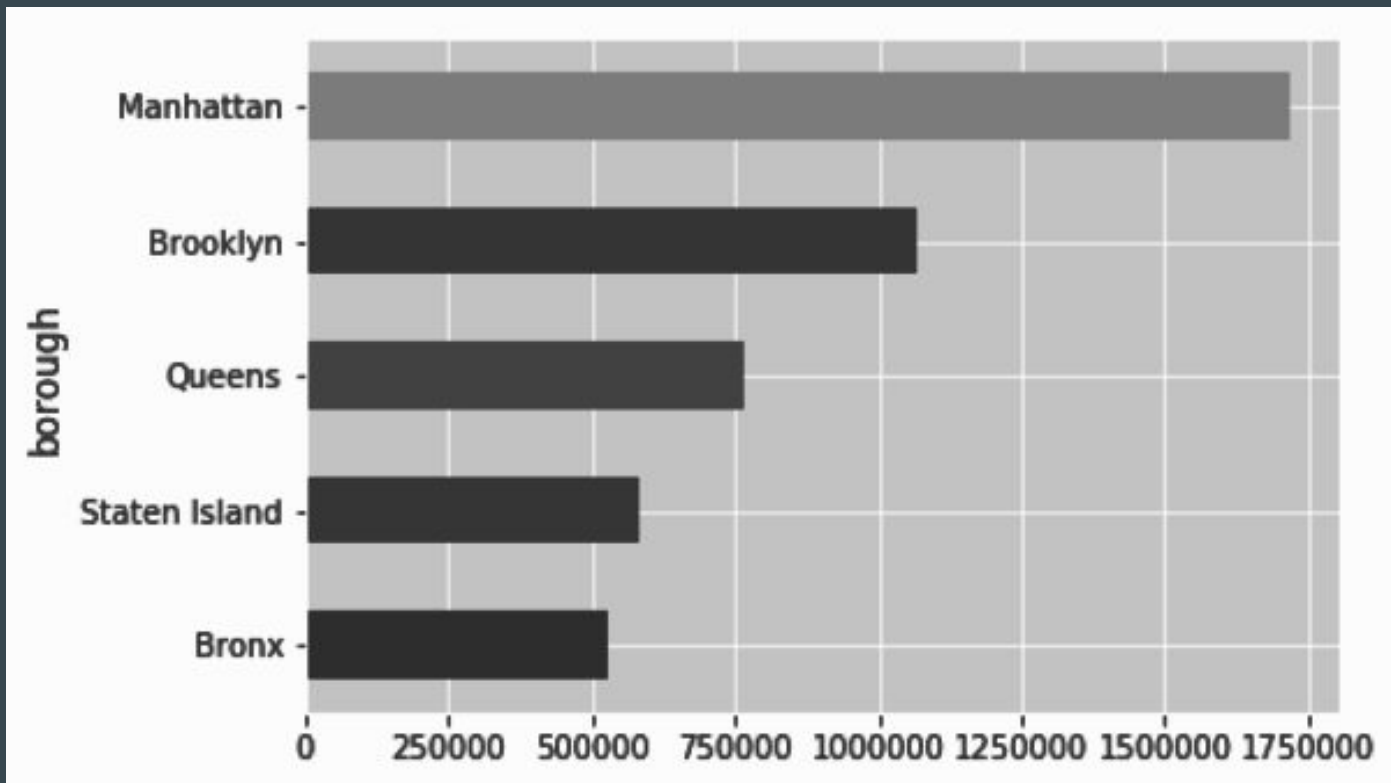
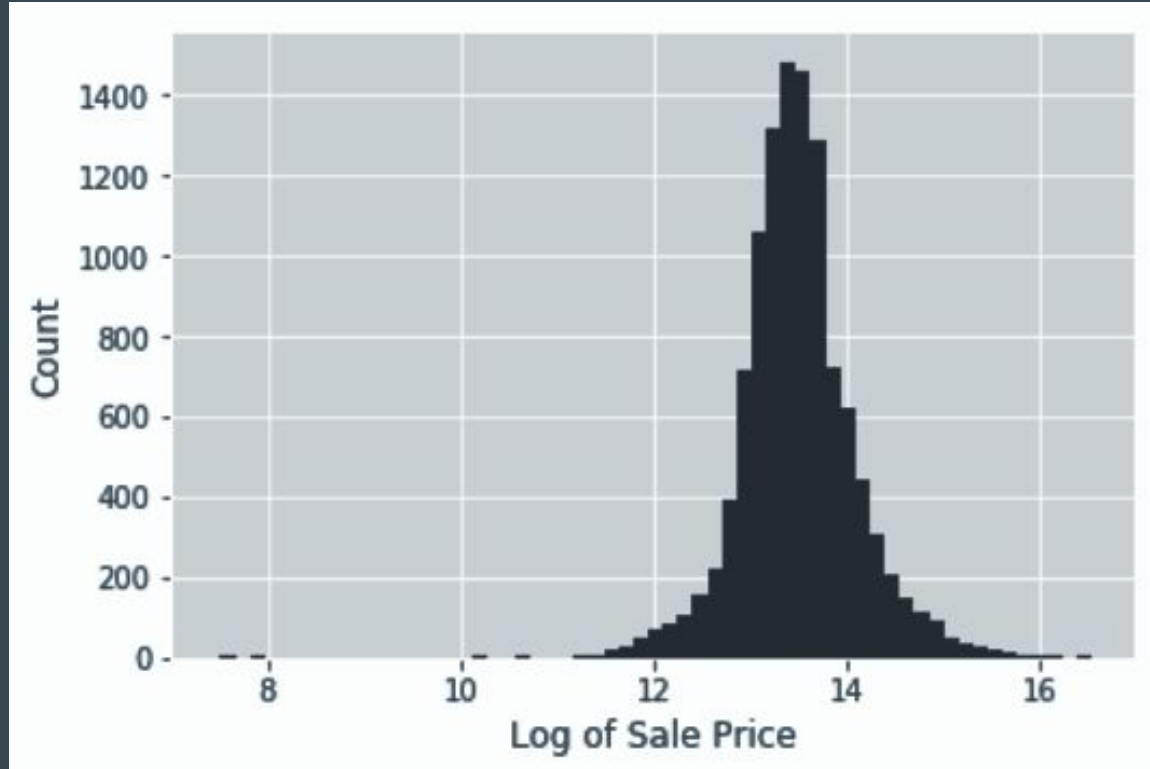# The Five Boroughs: Average Property Size (in sq.ft.)
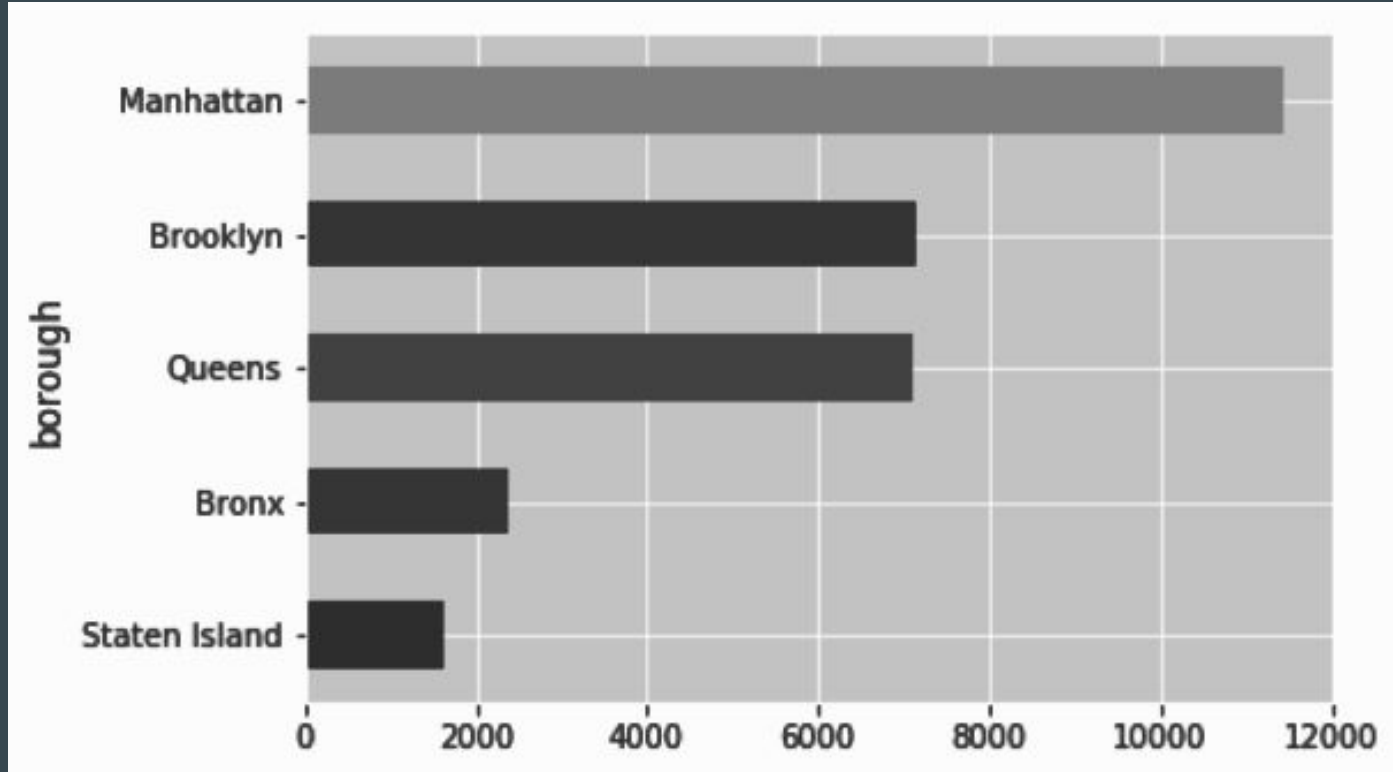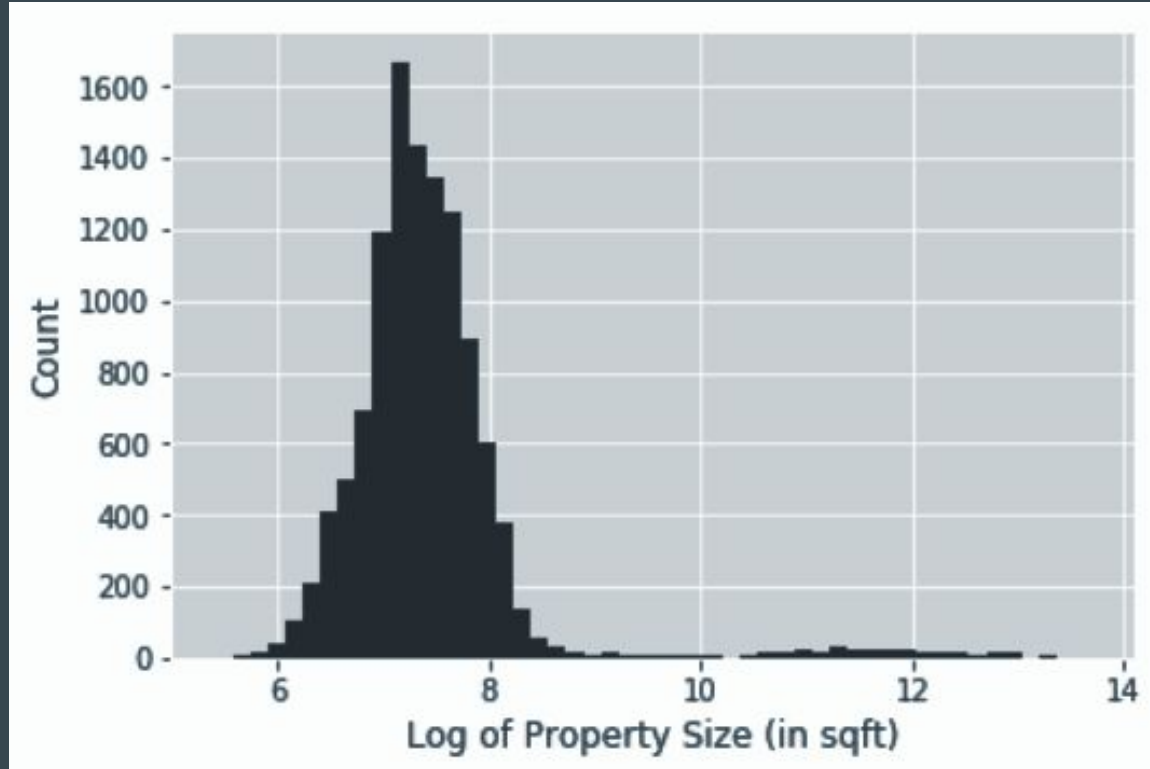
The Five Boroughs: Average Price per sq.ft.
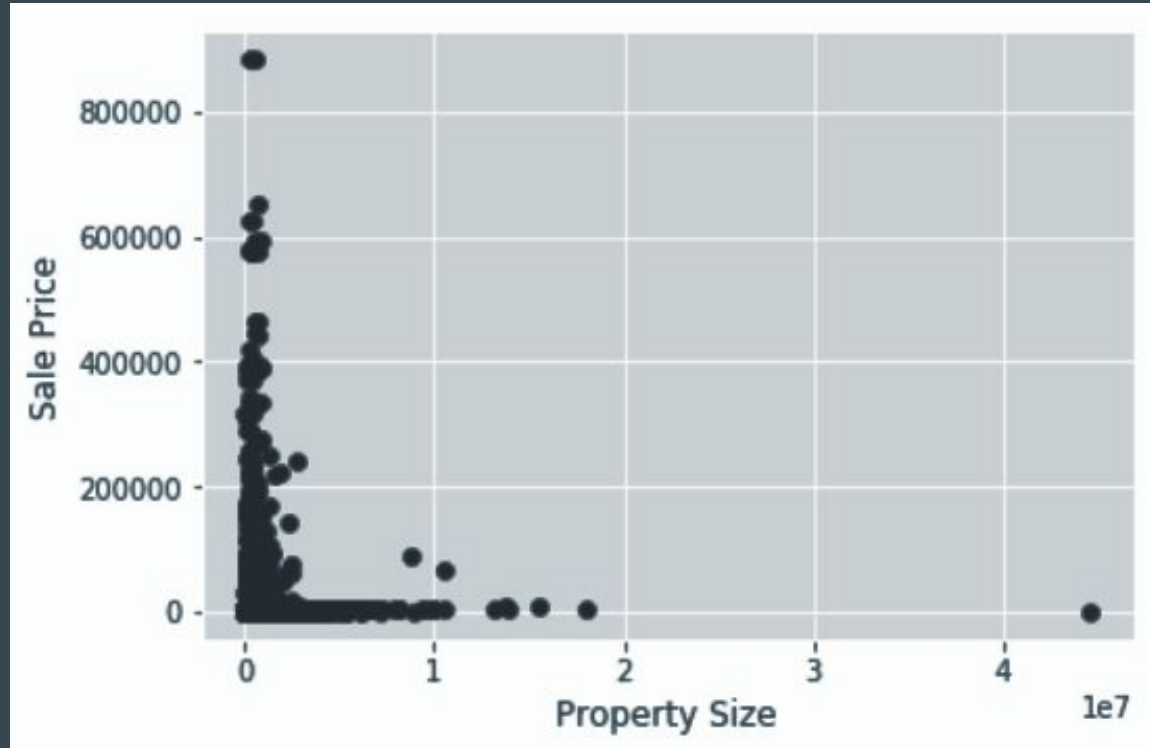
Property Size

How are the listings spread? : Sale Price
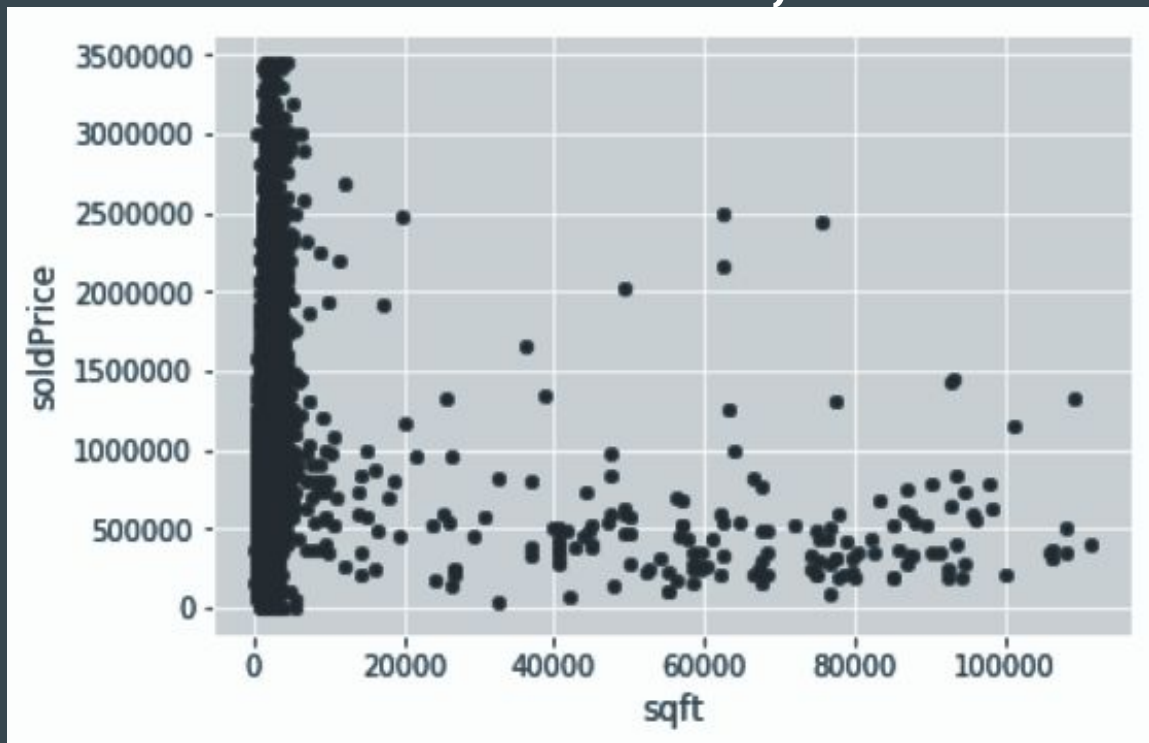
How are the listings spread? : Property Size

Mapping Sale Price vs. Property Size

Mapping Sale Price vs. Property Size
(ignoring values outside 3 standard deviations)

Pearson Correlation: **-0.045**

Conclusion (partial) :

- Property location impacts price significantly more than property size does.

- Considerable disparity between boroughs in terms of price

- Age of the Property and Property Type (condo, single or multi family house) would be interesting variables to analyze for further insight.