

# PYTHON WEB SCRAPING



By Sabbir Mohammed

# A brief look at NYC Real Estate



# About the project:

- Using scrapy on [www.trulia.com](http://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 was to scrape the site for the following features per listing (top 2 were not scraped):
  - 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/](https://www.trulia.com/sold/New_York,NY/)>

trulia Buy Rent Mortgage Local Info Additional Resources


New York, NY  Any Price ▾ All Beds ▾ All Home Types ▾ All Sold ▾ More ▾ [Save Search](#)

Sort: Newest ▾

Explore the Neighborhoods




**Upper East Side**  
435 Sold Homes  
Sold: \$295k - \$15.5m  
[See Local Highlights](#)




**Upper West Side**  
471 Sold Homes  
Sold: \$175k - \$18m  
[See Local Highlights](#)


Recently Sold Homes in New York, NY • 2996 sold homes on Trulia




**\$399,000**  
9,664 sqft  
249 E 7th St #1  
**SOLD on Feb 1, 2019**




**\$1,100,000**  
788 sqft  
392 Central Park W #4M  
**SOLD on Jan 31, 2019**




**\$1,525,000**  
836 sqft  
101 W 24th St #4H  
**SOLD on Jan 31, 2019**




**\$1,153,750**  
852 sqft  
555 W 23rd St #S7R  
**SOLD on Jan 31, 2019**



**\$527,500**  
368 sqft  
20 West St #44F  
**SOLD on Jan 31, 2019**



**\$1,150,000**  
842 sqft  
209-211 E 51st St #9F  
**SOLD on Jan 31, 2019**





# A peek into the source 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

# Overview of final dataset:

	<b>soldPrice</b>	<b>sqft</b>	<b>PricePerSQFT</b>
<b>count</b>	11289.00	11289.00	11289.00
<b>mean</b>	878191.73	5881.31	569.15
<b>std</b>	753059.93	33147.63	417.60
<b>min</b>	20000.00	224.00	0.43
<b>25%</b>	520000.00	1138.00	326.39
<b>50%</b>	705000.00	1520.00	451.61
<b>75%</b>	970000.00	2160.00	662.13
<b>max</b>	18000000.00	651914.00	7604.56

**LOCATION**

# The Five Boroughs: Number of Sales





# NYC Population per Borough (US Census Bureau, 2010):

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

# The Five Boroughs: Median Sale Price



# The Five Boroughs: Median Property Size (in sq.ft.)



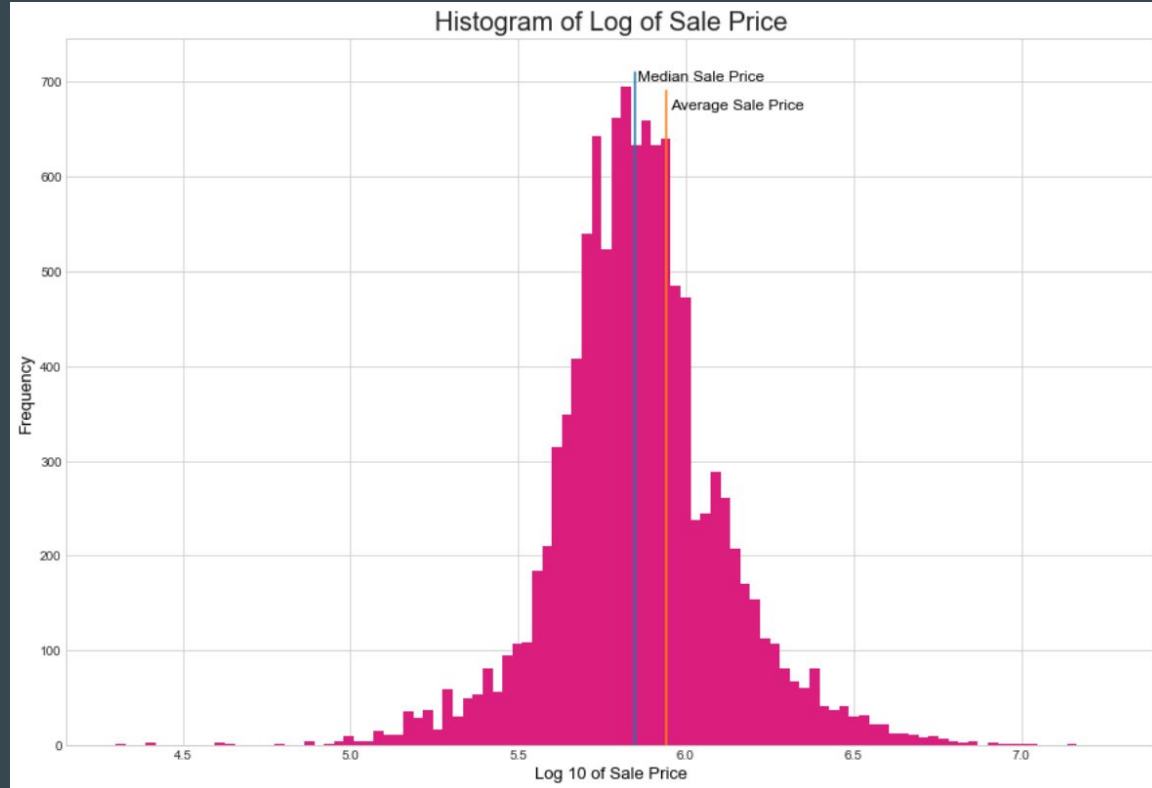
# The Five Boroughs: Average Price per sq.ft.



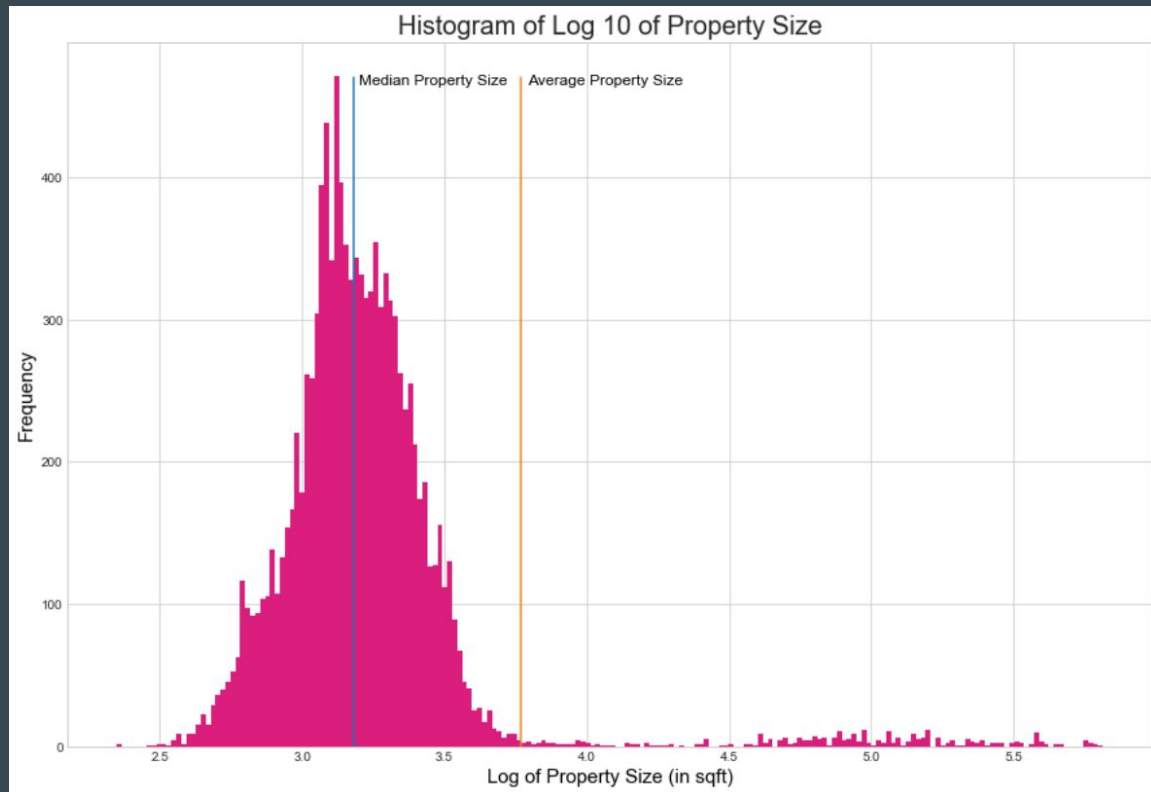
**Property Size**



# How are the listings spread? : Sale Price



# How are the listings spread? : Property Size



# Mapping Sale Price vs. Property Size



# Conclusions and Future Work :

- 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.