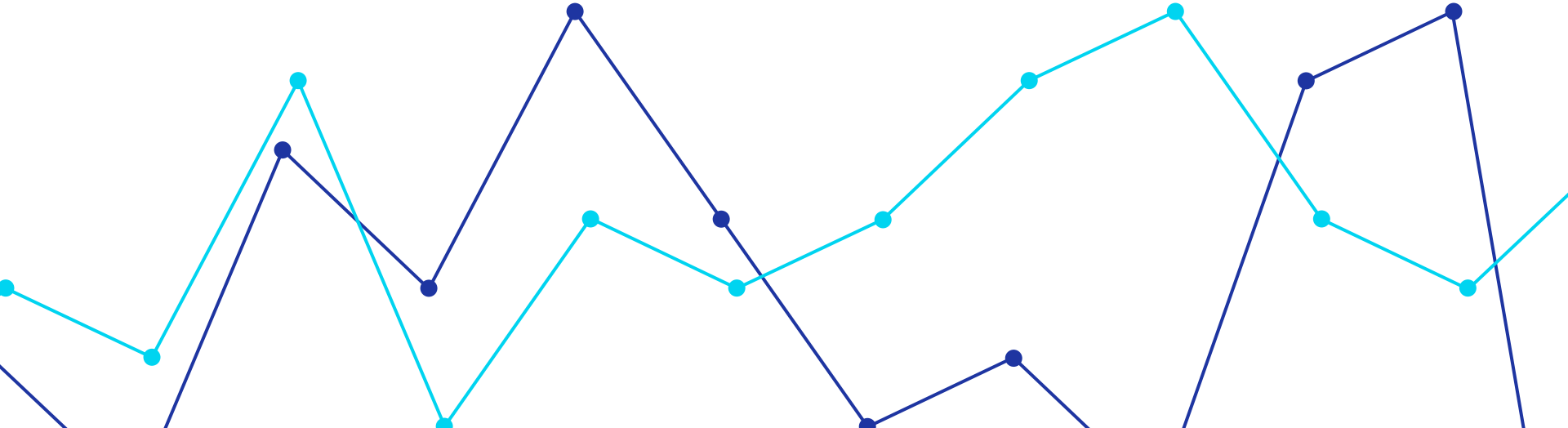
Trends in the New York Housing Market

Manav Bharath



Dataset

- New York Housing Market
    - Credit: https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market

Motivation

- Predict housing prices based on select features
- Analyze the distribution of house prices across the NYC area
- Understand the important brokers in the NYC market
- Study the effect of geography on house pricing

Data Exploration

Housing Data Table

	BROKERTITLE	TYPE	PRICE	BEDS	BATH	PROPERTYSQFT	ADDRESS	STATE	MAIN_ADDRESS	ADMINISTRATIVE_AREA_LEVEL_2
0	Brokered by Douglas Elliman -111 Fifth Ave	Condo for sale	315000	2	2.000000	1400.0	2 E 55th St Unit 803	New York, NY 10022	2 E 55th St Unit 803New York, NY 10022	New York County
1	Brokered by Serhant	Condo for sale	195000000	7	10.000000	17545.0	Central Park Tower Penthouse-217 W 57th New Yo...	New York, NY 10019	Central Park Tower Penthouse-217 W 57th New Yo...	United States
2	Brokered by Sowae Corp	House for sale	260000	4	2.000000	2015.0	620 Sinclair Ave	Staten Island, NY 10312	620 Sinclair AveStaten Island, NY 10312	United States
3	Brokered by COMPASS	Condo for sale	69000	3	1.000000	445.0	2 E 55th St Unit 908W33	Manhattan, NY 10022	2 E 55th St Unit 908W33Manhattan, NY 10022	United States
4	Brokered by Sotheby's International Realty - E...	Townhouse for sale	55000000	7	2.373861	14175.0	5 E 64th St	New York, NY 10065	5 E 64th StNew York, NY 10065	United States

Housing Data Table

LOCALITY	SUBLOCALITY	STREET_NAME	LONG_NAME	FORMATTED_ADDRESS	LATITUDE	LONGITUDE
New York	Manhattan	East 55th Street	Regis Residence	Regis Residence, 2 E 55th St #803, New York, N...	40.761255	-73.974483
New York	New York County	New York	West 57th Street	217 W 57th St, New York, NY 10019, USA	40.766393	-73.980991
New York	Richmond County	Staten Island	Sinclair Avenue	620 Sinclair Ave, Staten Island, NY 10312, USA	40.541805	-74.196109
New York	New York County	New York	East 55th Street	2 E 55th St, New York, NY 10022, USA	40.761398	-73.974613
New York	New York County	New York	East 64th Street	5 E 64th St, New York, NY 10065, USA	40.767224	-73.969856

Adjust Columns to be Lowercase

brokertitle	type	price	beds	bath	propertysqft	address	state	main_address	administrative_area_level_2	locality

sublocality	street_name	long_name	formatted_address	latitude	longitude

Features

```
brokertitle                 object
type                        object
price                        int64
beds                         int64
bath                       float64
propertysqft               float64
address                     object
state                       object
main_address                object
administrative_area_level_2 object
locality                    object
sublocality                 object
street_name                 object
long_name                   object
formatted_address           object
latitude                   float64
longitude                  float64
```

Condensing Info

brokertitle
Brokered by Douglas Elliman -111 Fifth Ave
Brokered by Serhant
Brokered by Sowae Corp
Brokered by COMPASS
Brokered by Sotheby's International Realty - E...

brokertitle
Douglas Elliman
Serhant
Sowae Corp
COMPASS
Sotheby's International Realty

Faulty Data

- House apparently on sale for $2.1 billion!
- Listed price is actually $2.5 million
- Dropped this entry from the data table

brokertitle	Brokered by ANNE LOPA REAL ESTATE
type	House for sale
price	2147483647
beds	7
bath	6.0
propertysqft	10000.0
address	6659-6675 Amboy Rd
state	New York, NY 10309
main_address	6659-6675 Amboy RdNew York, NY 10309
administrative_area_level_2	United States
locality	New York
sublocality	Richmond County
street_name	Staten Island
long_name	Amboy Road
formatted_address	6659 Amboy Rd, Staten Island, NY 10309, USA
latitude	40.518484
longitude	-74.224418

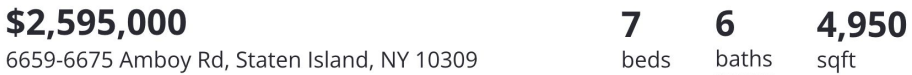


Image Credit: Zillow

Most Expensive Property

- Condo apparently on sale for $195 million!

brokertitle	Brokered by Serhant
type	Condo for sale
price	195000000
beds	7
bath	10.0
propertysqft	17545.0
address	Central Park Tower Penthouse-217 W 57th New Yo...
state	New York, NY 10019
main_address	Central Park Tower Penthouse-217 W 57th New Yo...
administrative_area_level_2	United States
locality	New York
sublocality	New York County
street_name	New York
long_name	West 57th Street
formatted_address	217 W 57th St, New York, NY 10019, USA
latitude	40.766393
longitude	-73.980991

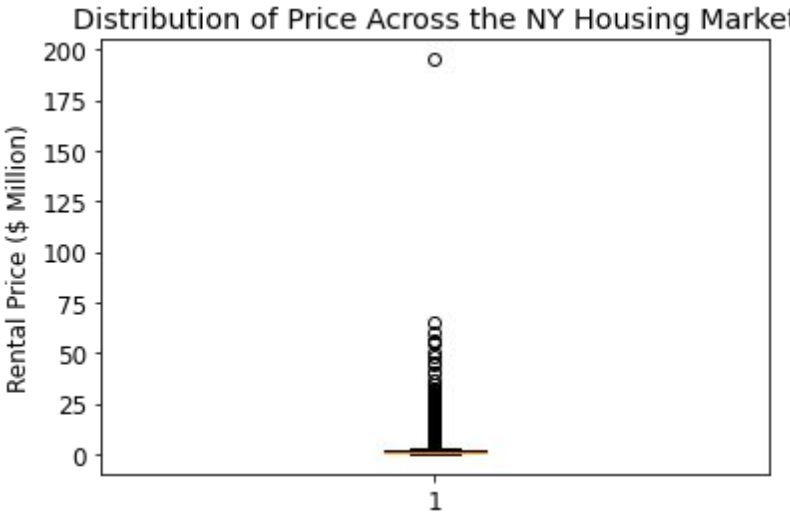
- Accurate data (as confirmed by Zillow)

Image Credit: Zillow

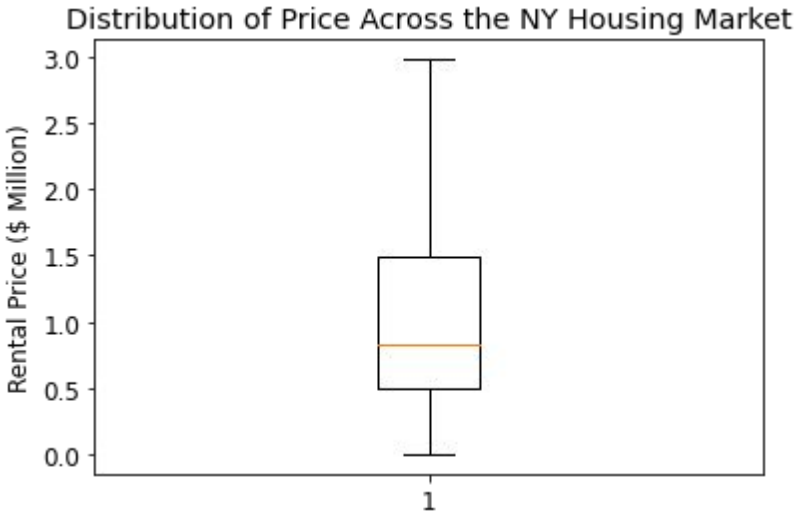
Data Visualization

Property sq. ft. and rental price trends

Price Distribution



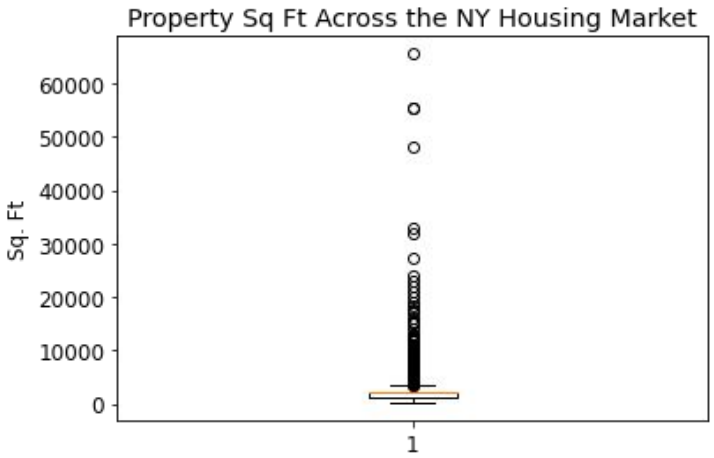
With outliers



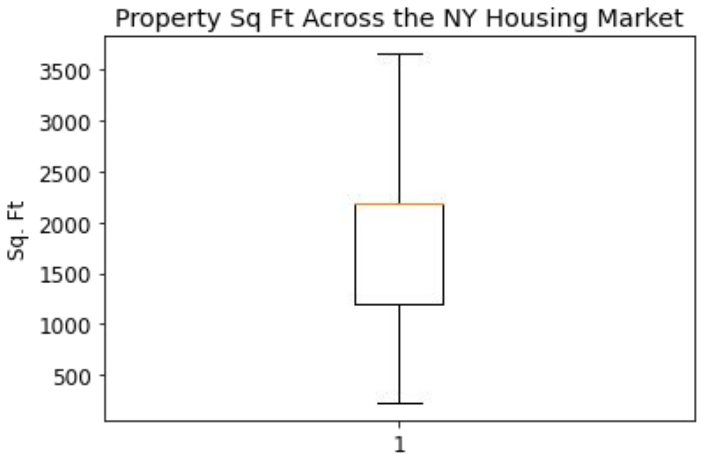
Without outliers

- The few most expensive properties (outliers) are vastly more expensive than the rest
- Most properties (75%) fall between $500,000 to $1.5 million

Sq. Ft Distribution



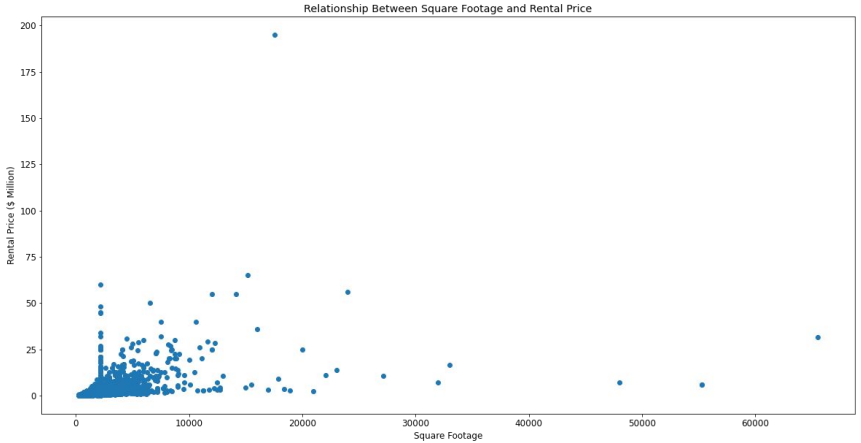
With outliers

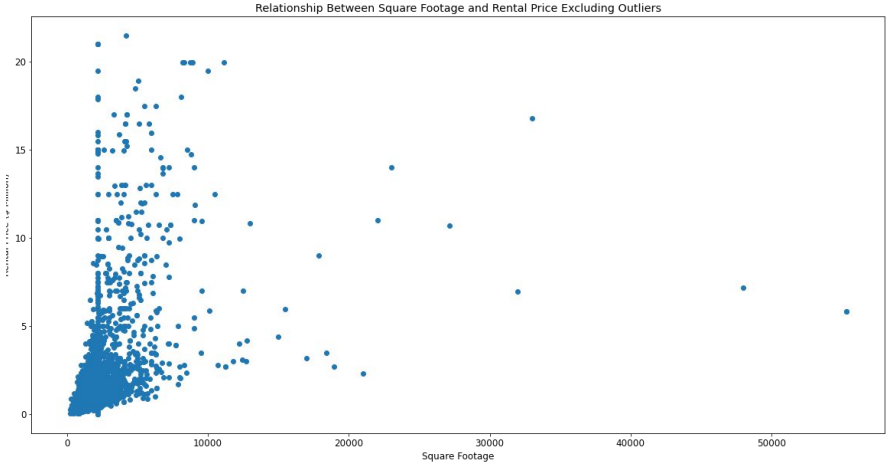


Without outliers

- The few most expensive properties (outliers) are vastly larger than the rest
- Most properties (75%) fall between 1250 to 2250 sq ft.

Sq Ft. vs Rental Price

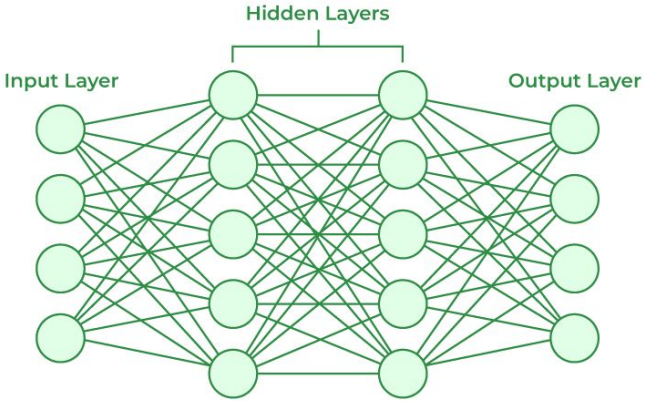




- Positive linear relationship between price and square footage for smaller properties
- More expensive properties don’t correspond to larger properties
  - Location has a larger impact than square footage

Predicting Property Price

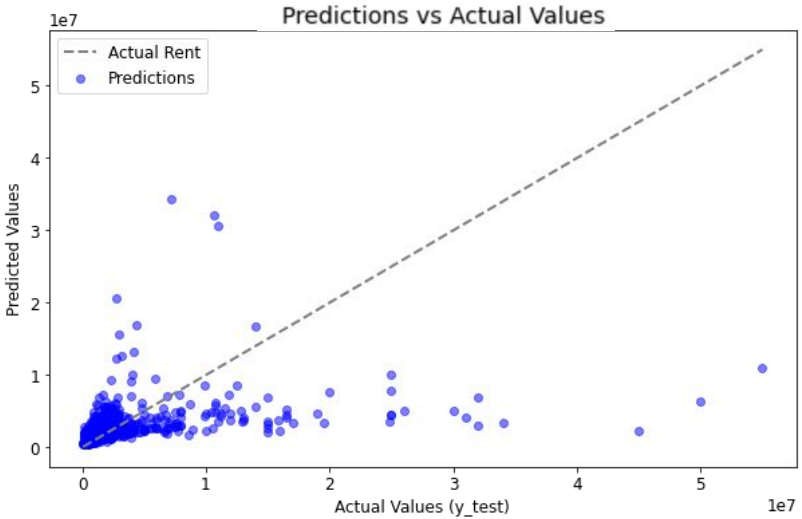
Neural Networks with PyTorch



PyTorch

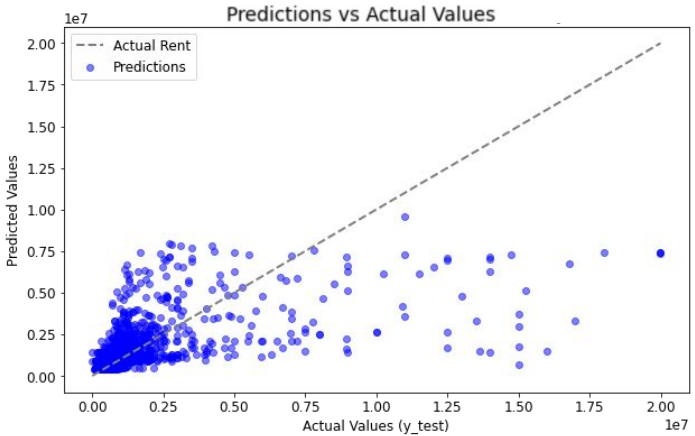
- Utilized PyTorch to easily generate a neural network to predict property price
- Input: “beds”, “bath”, “propertysqft”
- Adjust the weights of multiple nodes to fine-tune an output for a given set of inputs

Iteration 1: All Data



- Price of cheaper properties can be predicted reasonably well
- Overestimation of price: more expensive properties skew the predictions for cheaper properties
- Underestimation of price: location not factored

Iteration 2: Outliers Removed and Location Factored



- Input: “beds”, “bath”, “propertysqft”, “latitude”, “longitude”
- Price overestimation is reduced
- More accurate for a wider range of prices
- More factors might be needed to accurately determine price for more expensive properties

Neural Network Effectiveness

Iteration 1 MSE

```
Test MSE is 14032797958144.0
Test Root MSE is 3746037.634373686
```

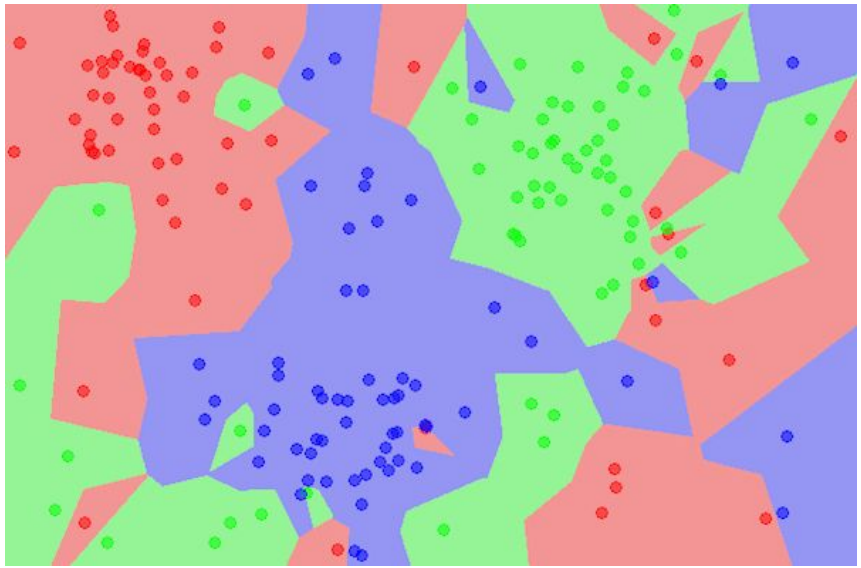
Iteration 2 MSE

```
Test MSE is 4232488419328.0
Test Root MSE is 2057301.2466160613
```

- Root MSE measures the average distance between the predicted price and the actual value
- MSE improved by ~$1.7 million across the 2 iterations (removing outliers and factoring in location)
- Still not an ideal model – more factors may need to be considered

K Nearest Neighbors

- Supervised learning classifier algorithm
- Makes predictions and classifications by forming clusters based on “distance” of data points
- Only requires tuning of one parameter (k) rather than tuning of multiple parameters (weights of nodes in a neural network)



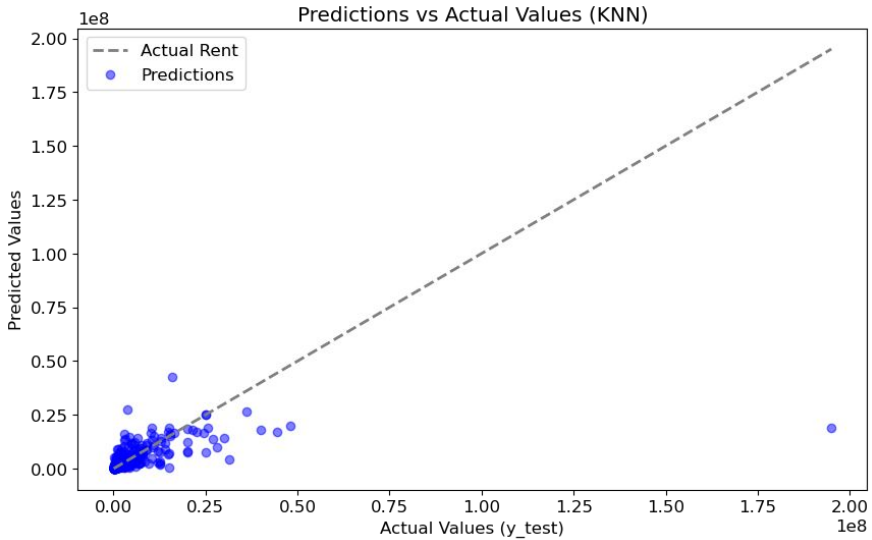
K Nearest Neighbors

- Adjustments: not using sq. footage as an input
    - Larger properties can be significantly cheaper than smaller properties
    - More likely to give faulty predictions if size is used

```
X = houses[['beds', 'bath', 'latitude','longitude']]
y = houses['price']
```

Iteration 1: KNN With Max Included

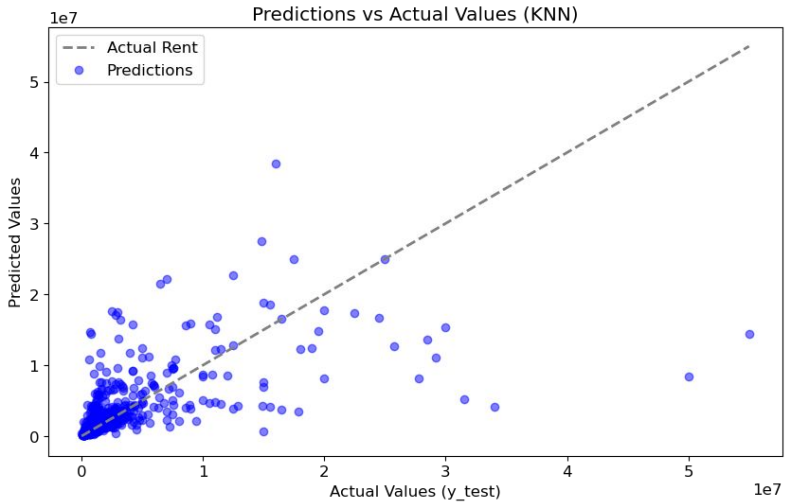
- $195 million property included
- Root MSE is greater than the neural network models
  - Less predictive model
- Upside: rest of data seems to follow the actual rent
- Adjustments: remove max data point



Root MSE: 5772615.0

Iteration 2: KNN With Max Removed

- Root MSE is smaller than the previous KNN
- Root MSE is larger than the 2nd neural network iteration
  - Less predictive model
  - Neural network is a better model for this data



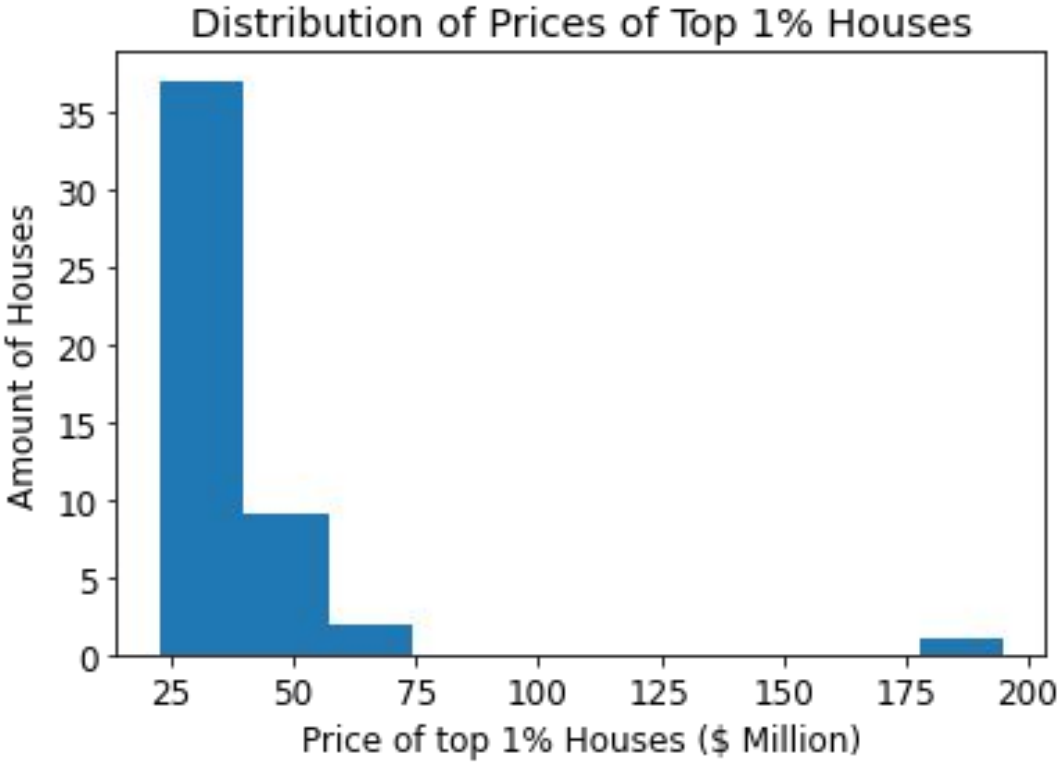
Neural Network with location

```
Test Root MSE is 2057301.2466160613
```

KNN

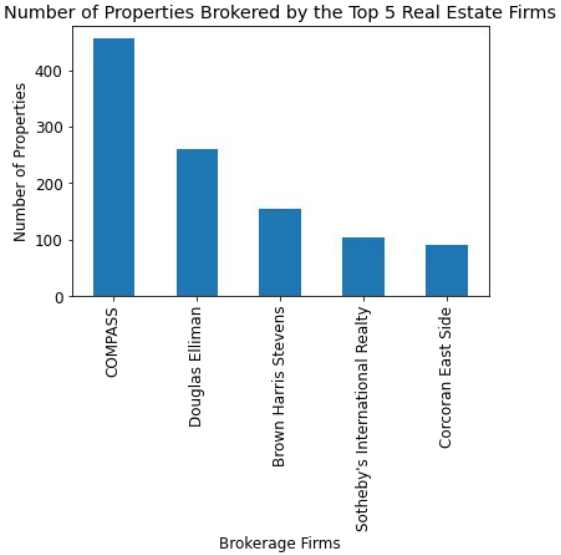
Root MSE: 3269728.8

Analysis of Most Expensive Properties

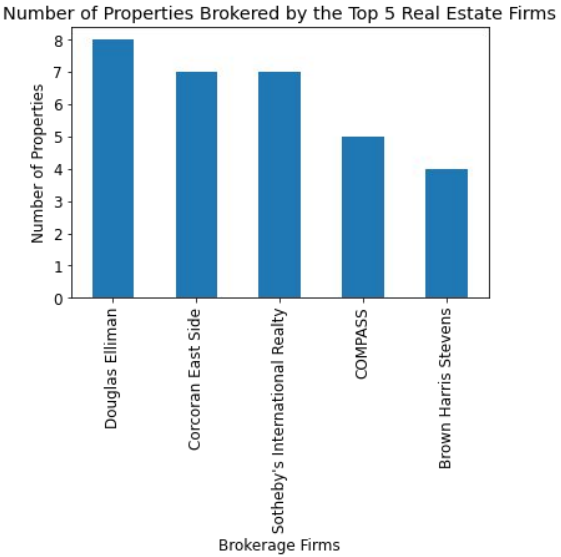

Distribution of Prices of Top 1% Houses
Amount of Houses
0
5
10
15
20
25
30
35
25
50
75
100
125
150
175
200
Price of top 1% Houses ($ Million)

Property Brokers Comparison

All Properties



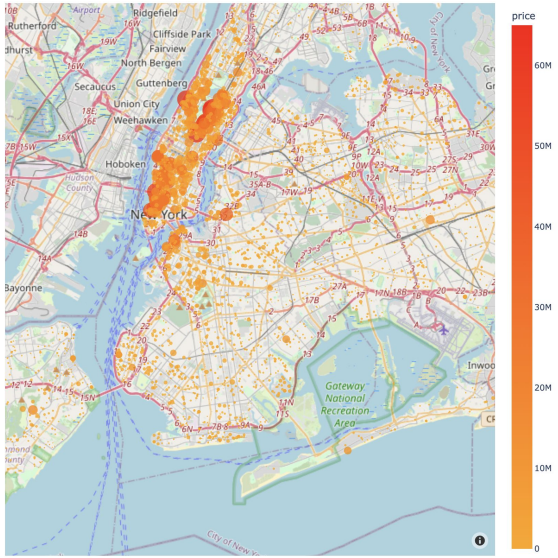
Top 1% Properties

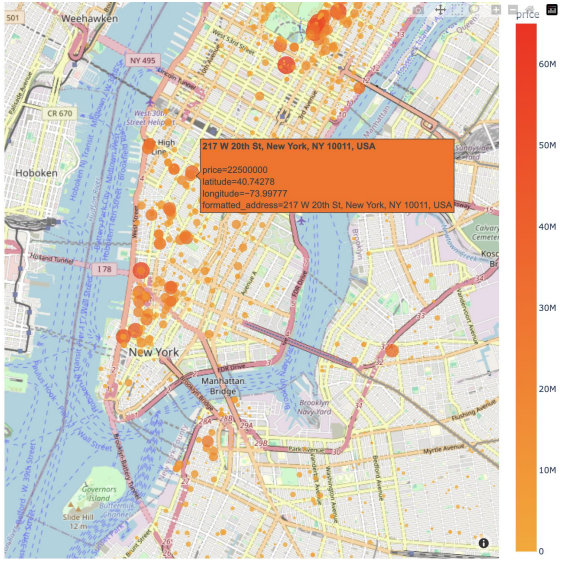


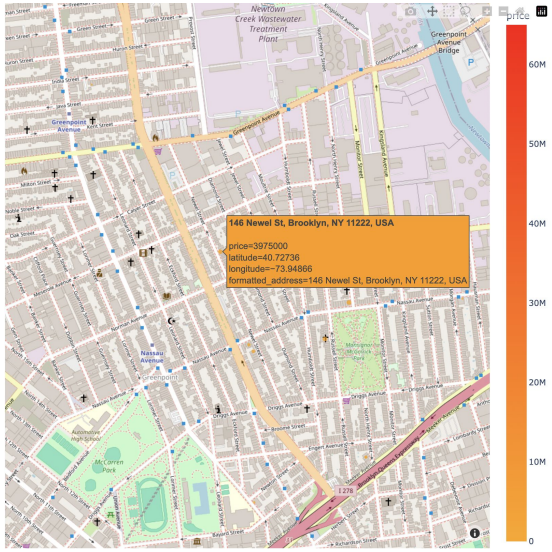
- While COMPASS is used for most properties, the most expensive properties often use Douglas Elliman and Corcoran
- Same group of brokerage firms for the most expensive properties and all properties (wide price range of properties sold)

Interactive Housing Price Exploration

Interactive Plot







- Used Plotly mapbox tools to generate a scrollable interactive map of housing in NYC colored by price

- Most expensive property ($195 million) not included to have a clearer color distribution

Additional Work

- I wanted to find other models that could generate more accurate predictions of the price or find additional ways to tweak the neural network and kNN models to be more accurate but I couldn’t.

- I wanted to find a way to host the interactive map element on a website so that other people could use it but I don’t have expertise in web development so I wasn’t able to accomplish this.