Predicting Craigslist NYC Apartment Rental Prices with Linear Regression

Ami Tian

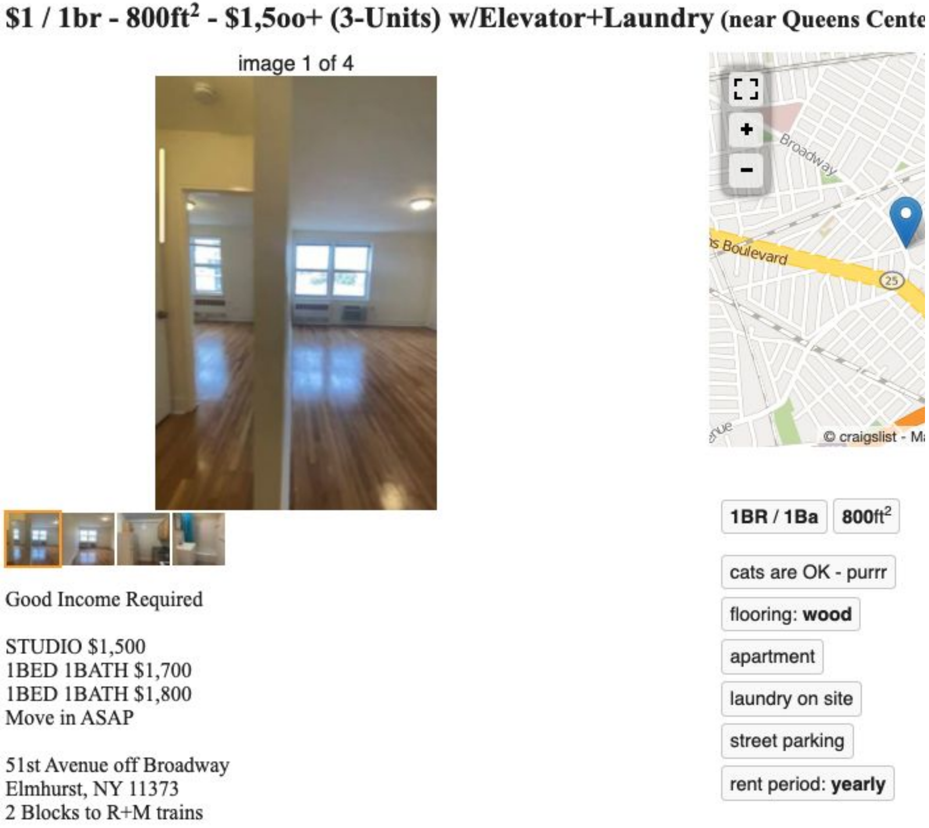
Project Goal: Do an apartment's features help predict its rental price?

The small counter. Cameron Knowlton

https://www.insider.com/worst-apartment-new-york-city-tiktok-reactions-2021-2

Web Scraping and Data Cleaning

- Selenium/BeautifulSoup for web scraping
- Features:
  - # of bedrooms (BR)
  - # of bathrooms (BA)
  - Location
  - Square footage (SQFT)
  - Pets allowed
  - Application Fee
  - Broker's Fee
  - Laundry
  - Parking
- Cleaning: Dealing with outliers, incorrect/misplaced information, removing duplicate posts



$1 / 1br - 800ft^2 - $1,5oo+ (3-Units) w/Elevator+Laundry (near Queens Cente

image 1 of 4

Good Income Required

STUDIO $1,500
1BED 1BATH $1,700
1BED 1BATH $1,800
Move in ASAP

51st Avenue off Broadway
Elmhurst, NY 11373
2 Blocks to R+M trains

1BR / 1Ba | 800ft^2

cats are OK - purrr

flooring: **wood**

apartment

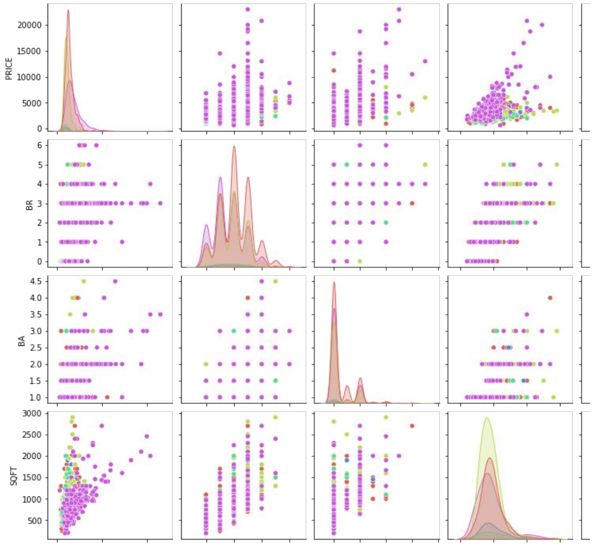
laundry on site

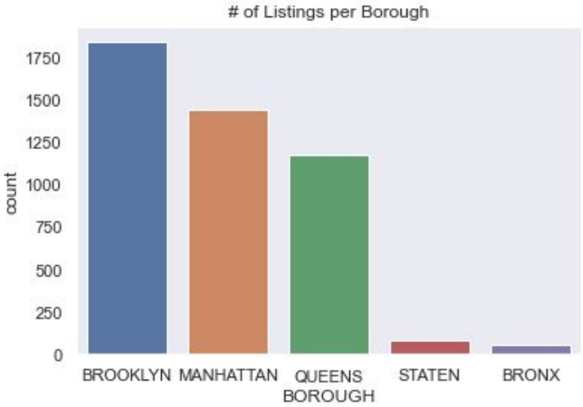
street parking

rent period: **yearly**

Data Prep and Feature Engineering

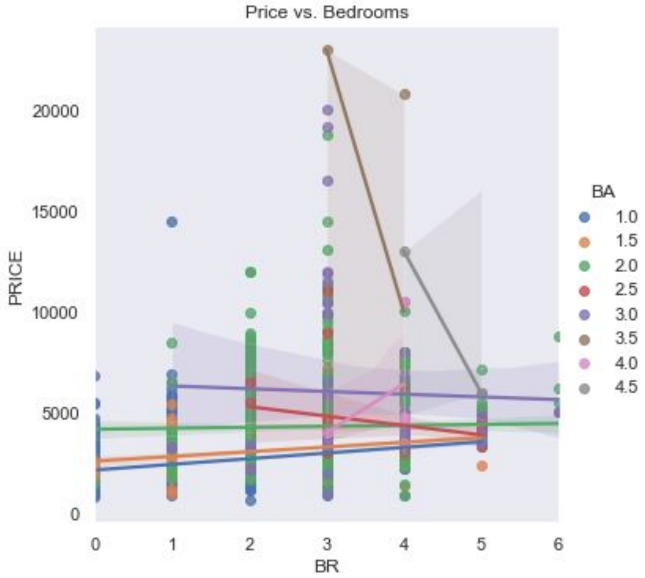
- Out of 4609 listings, only 969 listed square footage.
  - Imputed missing values based on median, borough, bedrooms.
- Created dummy variables for categoricals
  - Boolean: pets, application fee, broker's fee
  - Washer/Dryer: none, in-unit, in-building
  - Parking: none, off-street (includes carports and garages), valet
  - Borough: Brooklyn, Manhattan, Queens, the Bronx, Staten Island

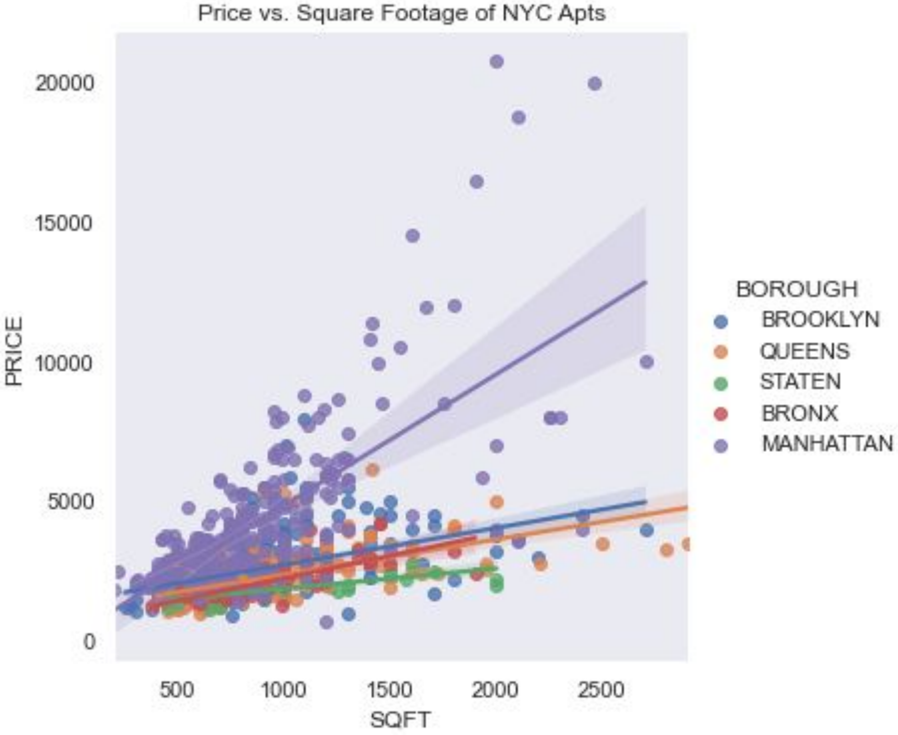




More Data Prep and Feature Engineering

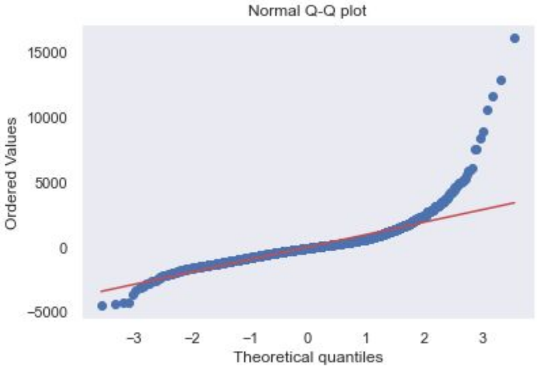
- Removed some features (application fee, pets)
- Added interaction terms
  - BR / BA
  - SQFT * MANHATTAN

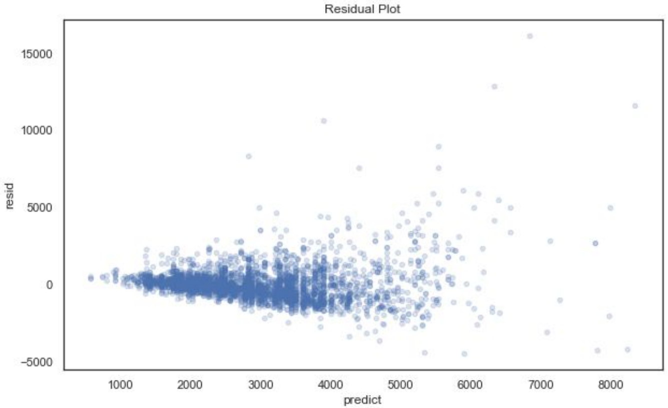


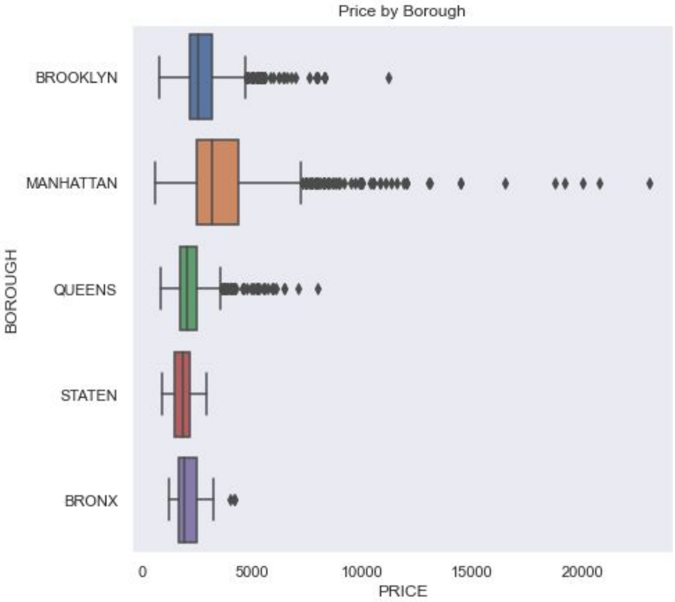


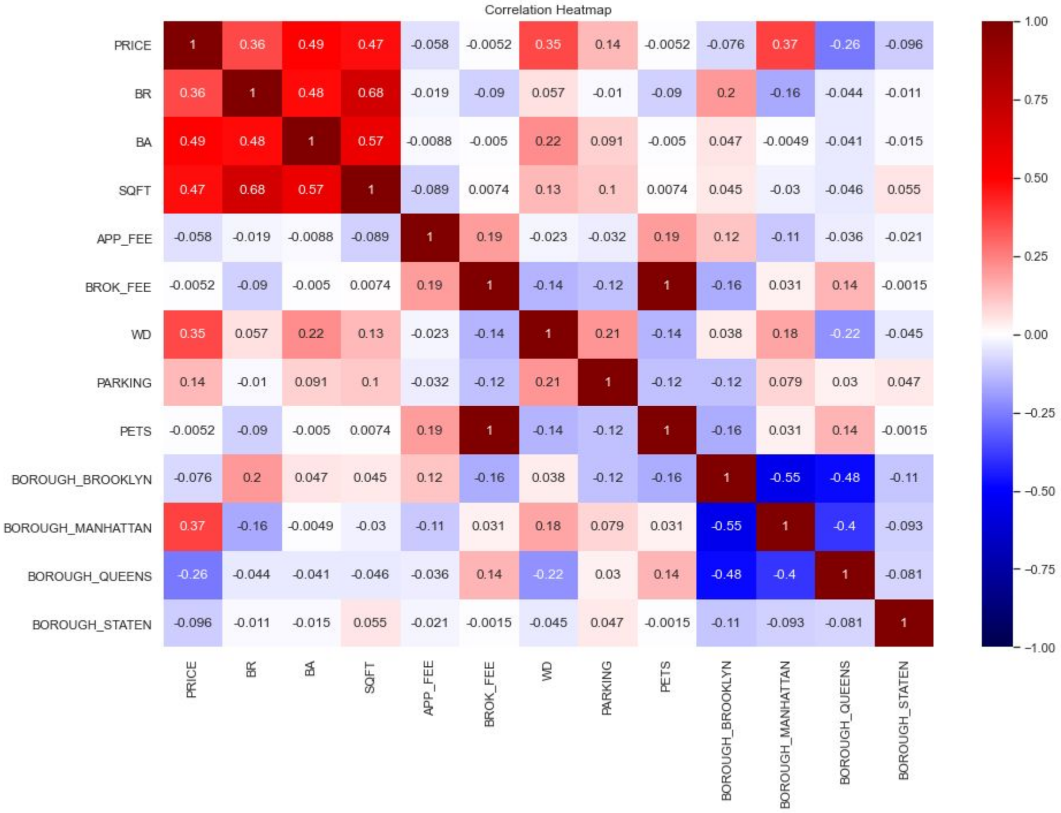
Issues

- Multicollinearity
- Heteroskedasticity

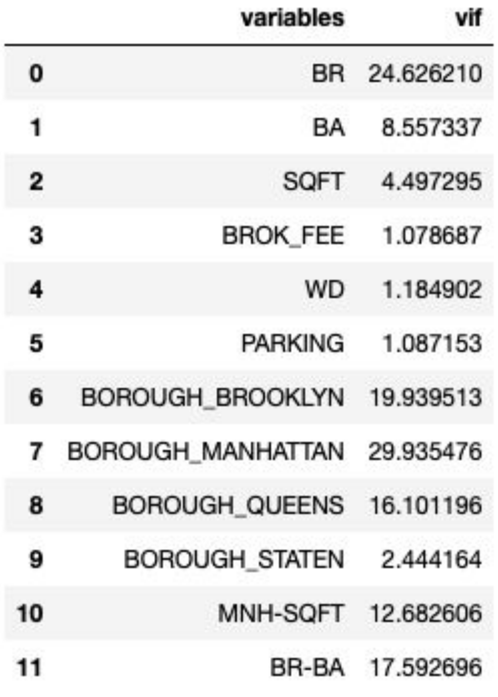








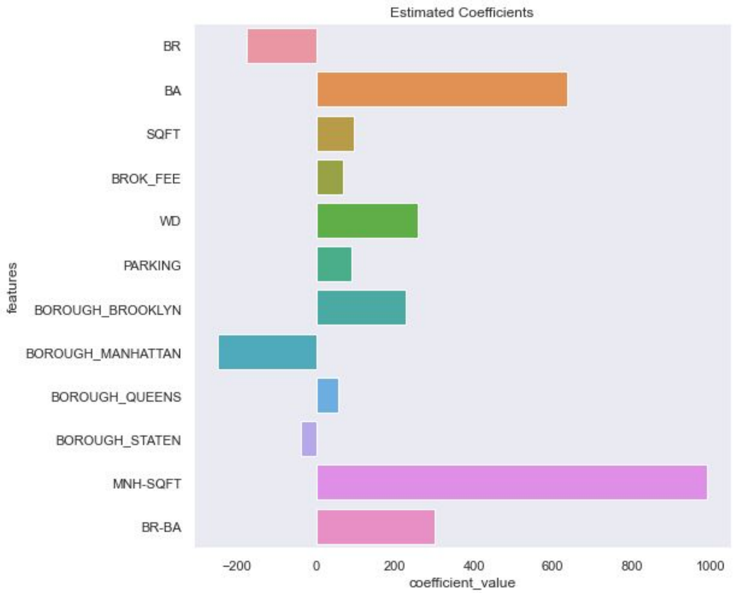
	PRICE	BR	BA	SQFT	APP_FEE	BROK_FEE	WD	PARKING	PETS	BOROUGH_BROOKLYN	BOROUGH_MANHATTAN	BOROUGH_QUEENS	BOROUGH_STATEN
PRICE	1	0.36	0.49	0.47	-0.058	-0.0052	0.35	0.14	-0.0052	-0.076	0.37	-0.26	-0.096
BR	0.36	1	0.48	0.68	-0.019	-0.09	0.057	-0.01	-0.09	0.2	-0.16	-0.044	-0.011
BA	0.49	0.48	1	0.57	-0.0088	-0.005	0.22	0.091	-0.005	0.047	-0.0049	-0.041	-0.015
SQFT	0.47	0.68	0.57	1	-0.089	0.0074	0.13	0.1	0.0074	0.045	-0.03	-0.046	0.055
APP_FEE	-0.058	-0.019	-0.0088	-0.089	1	0.19	-0.023	-0.032	0.19	0.12	-0.11	-0.036	-0.021
BROK_FEE	-0.0052	-0.09	-0.005	0.0074	0.19	1	-0.14	-0.12	1	-0.16	0.031	0.14	-0.0015
WD	0.35	0.057	0.22	0.13	-0.023	-0.14	1	0.21	-0.14	0.038	0.18	-0.22	-0.045
PARKING	0.14	-0.01	0.091	0.1	-0.032	-0.12	0.21	1	-0.12	-0.12	0.079	0.03	0.047
PETS	-0.0052	-0.09	-0.005	0.0074	0.19	1	-0.14	-0.12	1	-0.16	0.031	0.14	-0.0015
BOROUGH_BROOKLYN	-0.076	0.2	0.047	0.045	0.12	-0.16	0.038	-0.12	-0.16	1	-0.55	-0.48	-0.11
BOROUGH_MANHATTAN	0.37	-0.16	-0.0049	-0.03	-0.11	0.031	0.18	0.079	0.031	-0.55	1	-0.4	-0.093
BOROUGH_QUEENS	-0.26	-0.044	-0.041	-0.046	-0.036	0.14	-0.22	0.03	0.14	-0.48	-0.4	1	-0.081
BOROUGH_STATEN	-0.096	-0.011	-0.015	0.055	-0.021	-0.0015	-0.045	0.047	-0.0015	-0.11	-0.093	-0.081	1



	variables	vif
0	BR	24.626210
1	BA	8.557337
2	SQFT	4.497295
3	BROK_FEE	1.078687
4	WD	1.184902
5	PARKING	1.087153
6	BOROUGH_BROOKLYN	19.939513
7	BOROUGH_MANHATTAN	29.935476
8	BOROUGH_QUEENS	16.101196
9	BOROUGH_STATEN	2.444164
10	MNH-SQFT	12.682606
11	BR-BA	17.592696

Model Selection and Results

- Used kfold cross-validation on training set to get the R^2 score, mean absolute error (MAE) and root mean squared error (RMSE)
- Applied Ridge and Lasso regularization, compared scores
- Lasso regularization improved scores slightly



TRAINING DATA

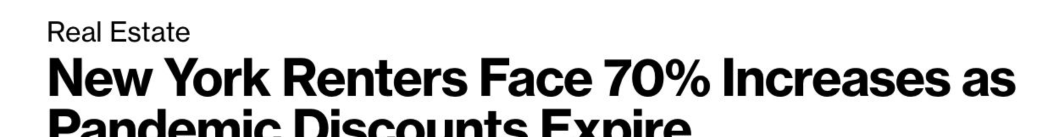
- R^2 value: 0.526
- Mean Absolute Error: $648.65
- Root Mean Squared Error: $1,027.45

TEST DATA

- R^2 value: 0.519
- Mean Absolute Error: $671.15
- Root Mean Squared Error: $1,127.72

Conclusions

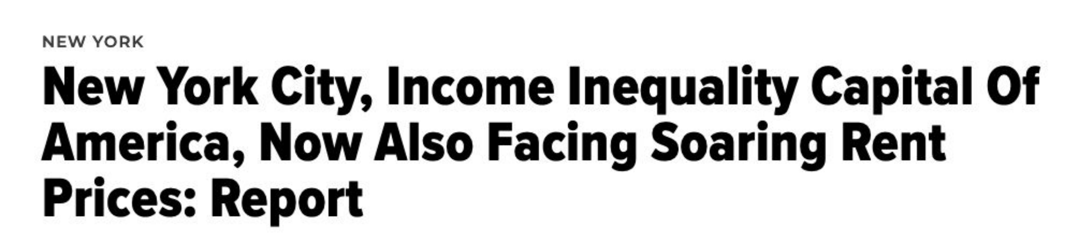
- The model can only explain about 50% of an apartment’s rental price
- Other possible factors: neighborhood, amenities, utilities
- External factors: housing policies, market forces



Real Estate

New York Renters Face 70% Increases as Pandemic Discounts Expire

The era of widespread Covid concessions for apartment hu



NEW YORK

New York City, Income Inequality Capital Of America, Now Also Facing Soaring Rent Prices: Report

Future Steps

- Get more data, use information from other sites (e.g. StreetEasy, apartments.com)
- Look into more features
- Figure out more accurate interaction terms
- Categorize by neighborhood instead of borough, get more information about location
- Try log transformation or use weighted least squares regression to deal with heteroskedasticity