Project 2
Rental Home Prediction

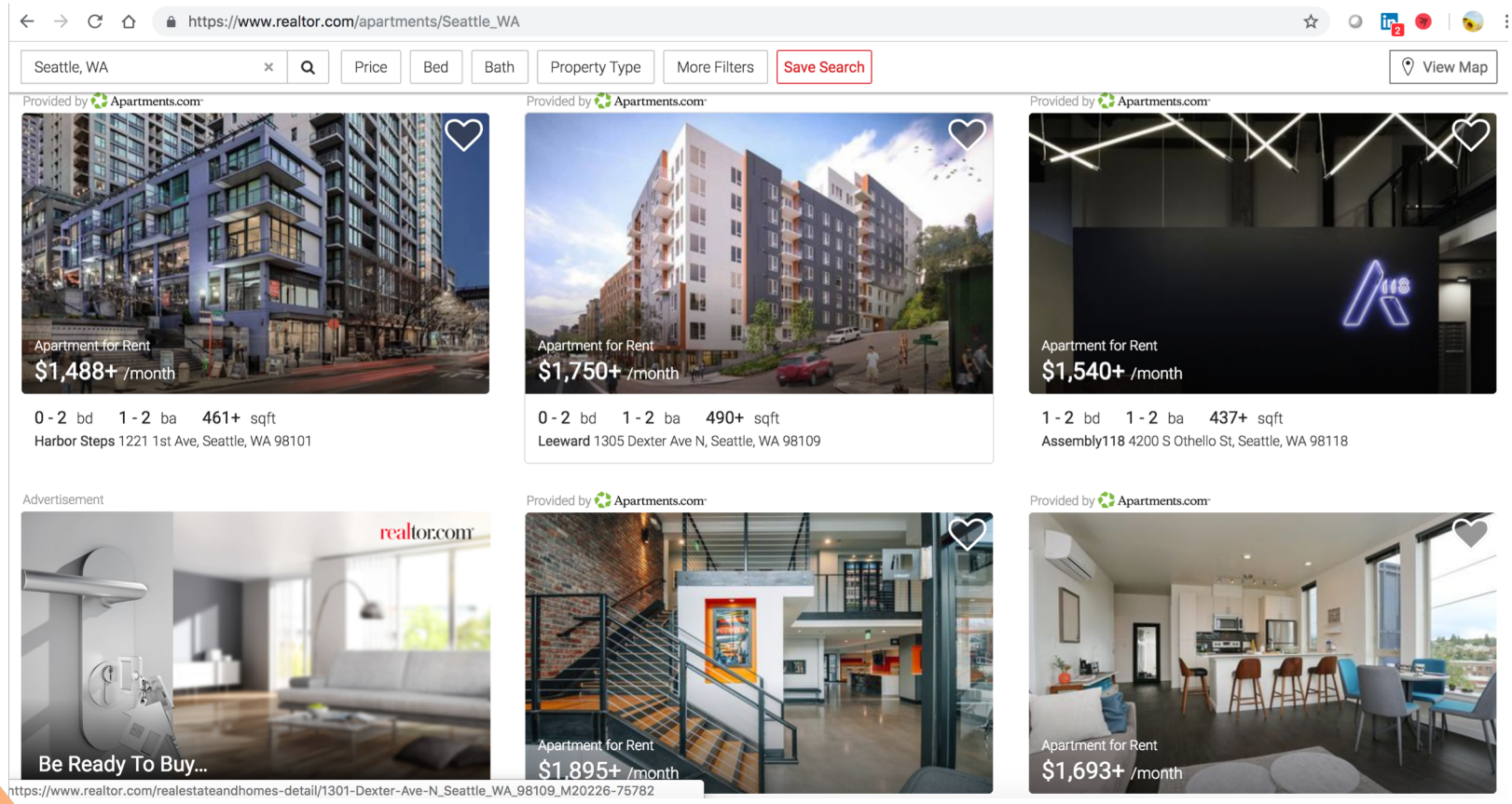
Leo Liu

Objective

- Evaluate a property

- Budget your rental

Data Source


https://www.realtor.com/apartments/Seattle_WA
Seattle, WA
Price
Bed
Bath
Property Type
More Filters
Save Search
View Map
Provided by Apartments.com
Apartment for Rent
$1,488+ /month
0 - 2 bd 1 - 2 ba 461+ sqft
Harbor Steps 1221 1st Ave, Seattle, WA 98101
Provided by Apartments.com
Apartment for Rent
$1,750+ /month
0 - 2 bd 1 - 2 ba 490+ sqft
Leeward 1305 Dexter Ave N, Seattle, WA 98109
Provided by Apartments.com
Apartment for Rent
$1,540+ /month
1 - 2 bd 1 - 2 ba 437+ sqft
Assembly118 4200 S Othello St, Seattle, WA 98118
Advertisement
realtor.com
Be Ready To Buy...
Provided by Apartments.com
Apartment for Rent
$1,895+ /month
Provided by Apartments.com
Apartment for Rent
$1,693+ /month
https://www.realtor.com/realestateandhomes-detail/1301-Dexter-Ave-N_Seattle_WA_98109_M20226-75782

Process

1 Web Scrapping

- Get URLs
- Iterate over pages
- Scrap each single page

2 Data Cleaning

- Account for only popular community and unit features

3 Modeling

- MVP
- Feature Screening
- Polynomial Features
- Variable Transformation

4 Training & Testing

- Check overfitting

Model

- Features:

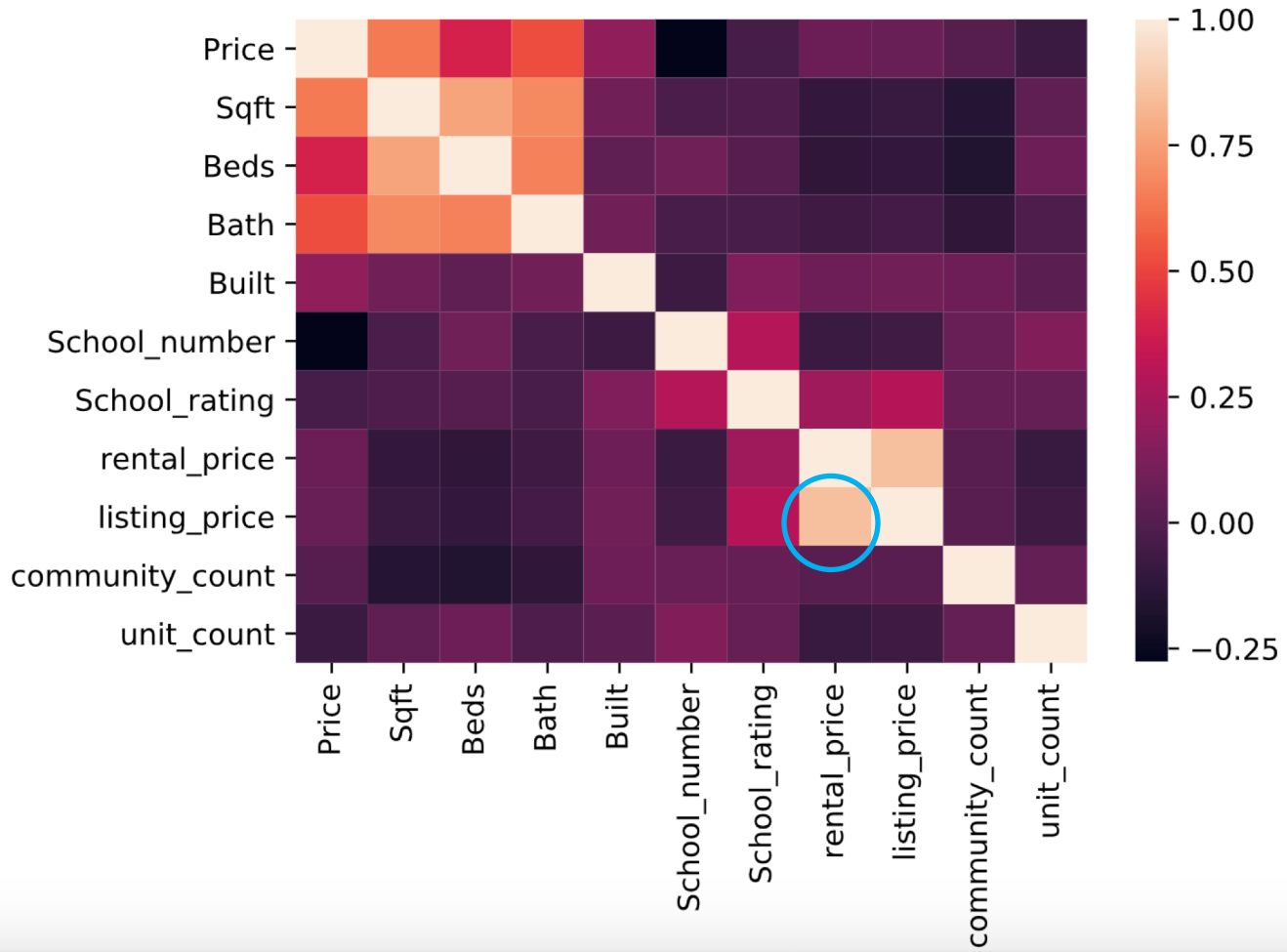
Bedrooms	Bathrooms
Sqft	YearBulit
Schools	Average School Rating
Community Features	Unit Features
Median Rental Price	Median Listing Price

- Responses:

  Rental Price

Model

Check multicollinearity



Model

Check p-values

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.389e+04	3093.447	-4.490	0.000	-2e+04	-7820.163
Sqft	2.4373	0.122	19.906	0.000	2.197	2.678
Beds	-356.2730	55.674	-6.399	0.000	-465.514	-247.032
Bath	596.9922	88.266	6.764	0.000	423.802	770.183
Built	5.8529	1.521	3.849	0.000	2.869	8.836
School_number	-345.3570	34.178	-10.105	0.000	-412.418	-278.296
School_rating	7.9883	20.666	0.387	0.699		48.538
rental_price	1.6230	0.491	3.304	0.001	0.659	2.587
listing_price	-0.0011	0.001	-0.790	0.429		0.002
community_count	8.0873	1.565	5.166	0.000	5.016	11.159
unit_count	-142.3902	57.050	-2.496	0.013	-254.330	-30.451

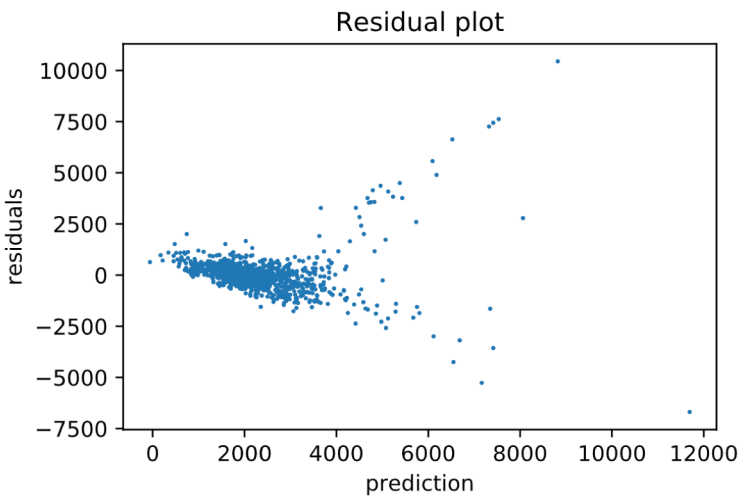
Model

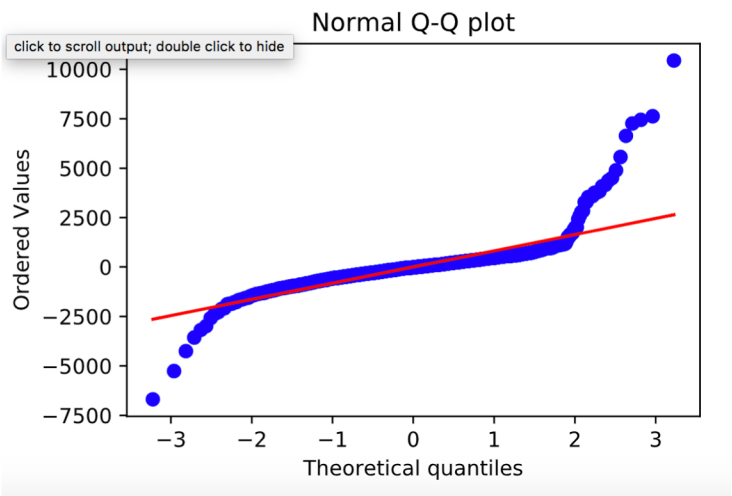
Feature Screening

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.388e+04	3039.351	-4.566	0.000	-1.98e+04	-7913.476
Sqft	2.4360	0.122	19.915	0.000	2.196	2.676
Beds	-354.4319	55.593	-6.375	0.000	-463.513	-245.351
Bath	594.7134	88.152	6.746	0.000	421.747	767.679
Built	5.8810	1.504	3.911	0.000	2.930	8.832
School_number	-341.6078	32.428	-10.534	0.000	-405.236	-277.980
rental_price	1.3132	0.258	5.099	0.000	0.808	1.819
community_count	8.1024	1.564	5.180	0.000	5.033	11.171
unit_count	-142.4432	56.978	-2.500	0.013	-254.241	-30.645

Model

Check heteroskedasticity

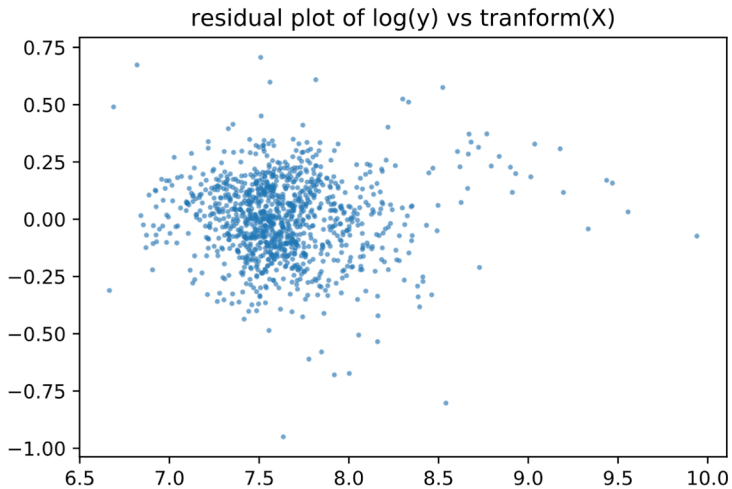


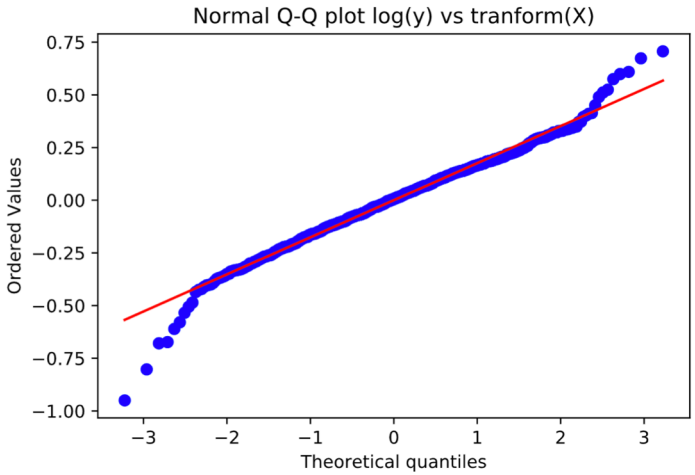


Residual plot and Normal Q-Q plot of 2-degree polynomial feature regression model

Model

Check heteroskedasticity





Residual plot and Normal Q-Q plot of 2-degree polynomial feature regression model with transformed variables

Train and Test

Cross Validation on Training Data

- **R-Squared = 0.730 +/- 0.033**
- **Mean Squared Error = 0.041 +/- 0.010**

Check Overfitting

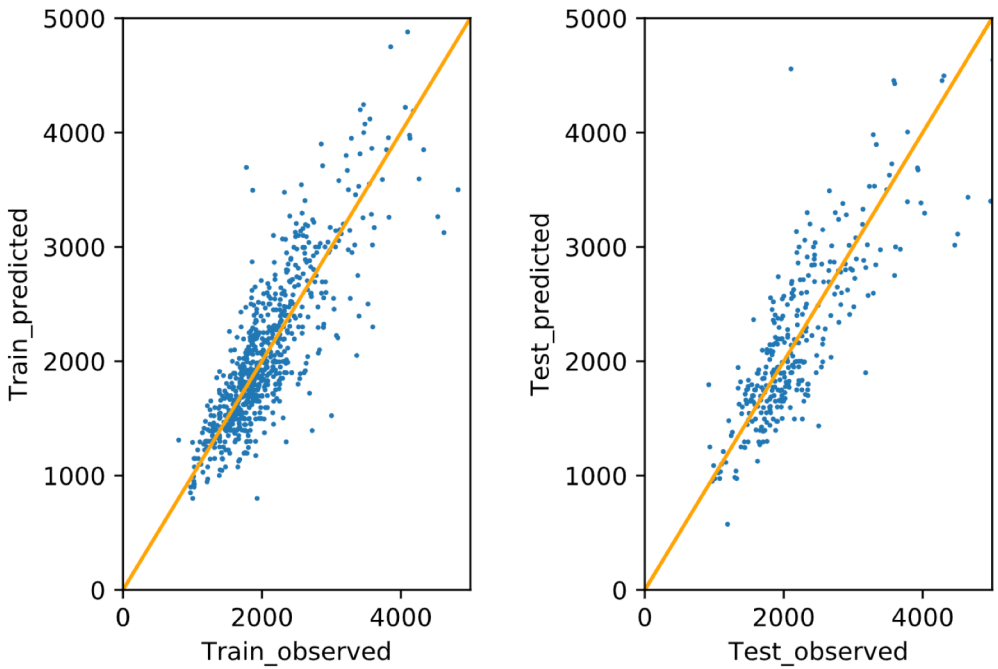
- **R-Squared(Train) = 0.791**
- **R-Squared(Test) = 0.780**
- **Mean Squared Error(Train) = 0.031**
- **Mean Squared Error(Test) = 0.037**

$$e^{\sqrt{0.037}} = 1.212$$

$$e^{\sqrt{0.031}} = 1.192$$

So the final deviation between train and test is about 1.212/1.192 - 1 ≈ 1.6%

Train and Test



Train & Test Prediction vs. Observation

Future Work

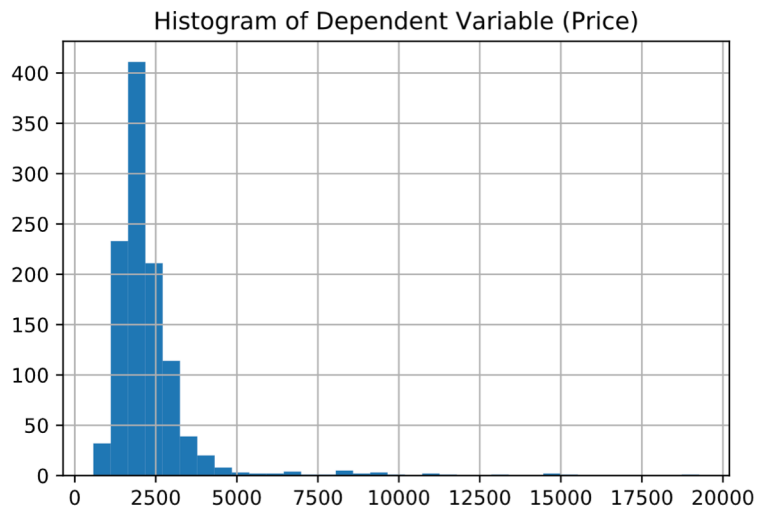
- Increase dataset
- Incorporate more features
- Create customized models for different property types

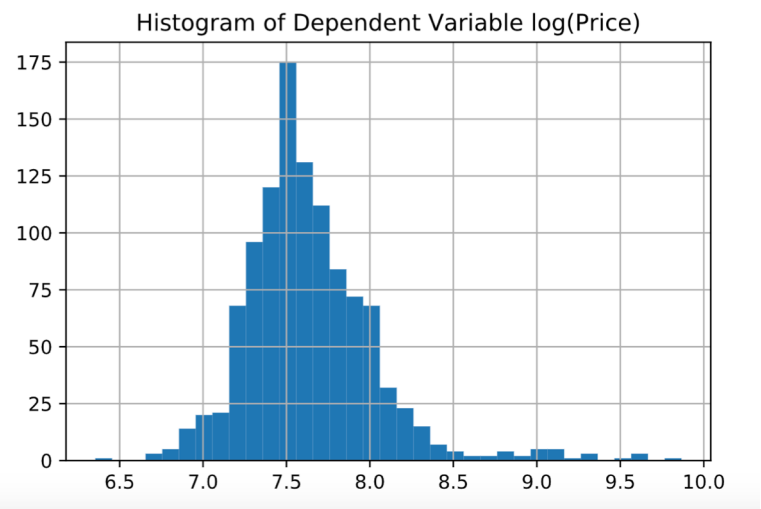
Thank You!

Appendix

Model

Check heteroskedasticity

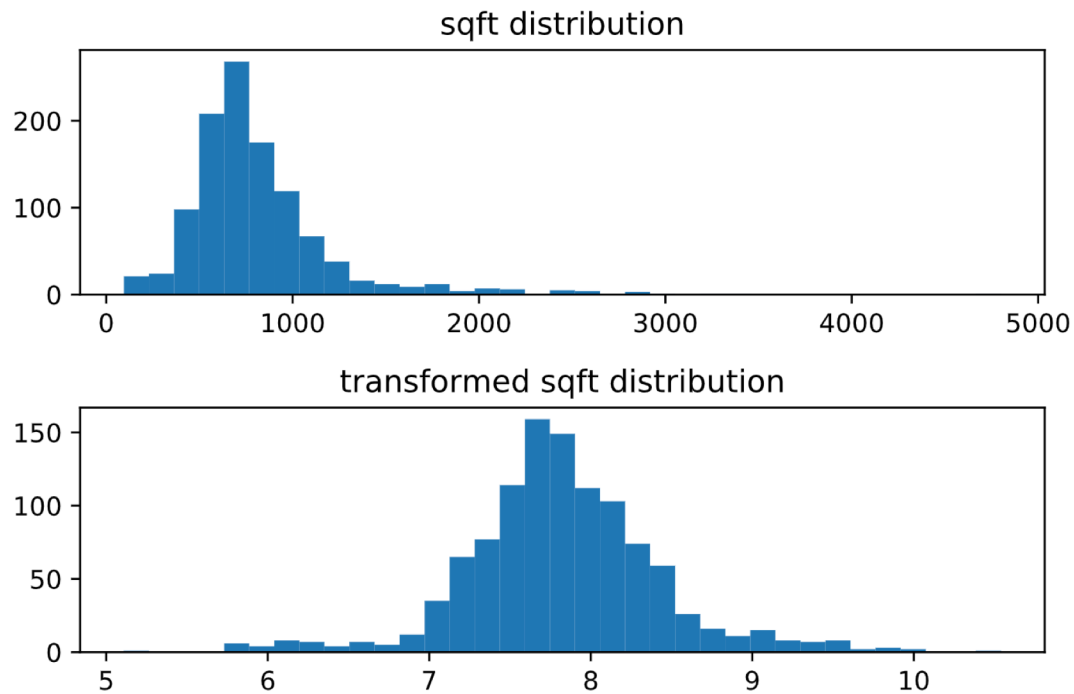




Distribution of y and log(y)

Model

Check heteroskedasticity



Distribution of Sqft and Boxcox transformed Sqft