# KC Housing Dataset
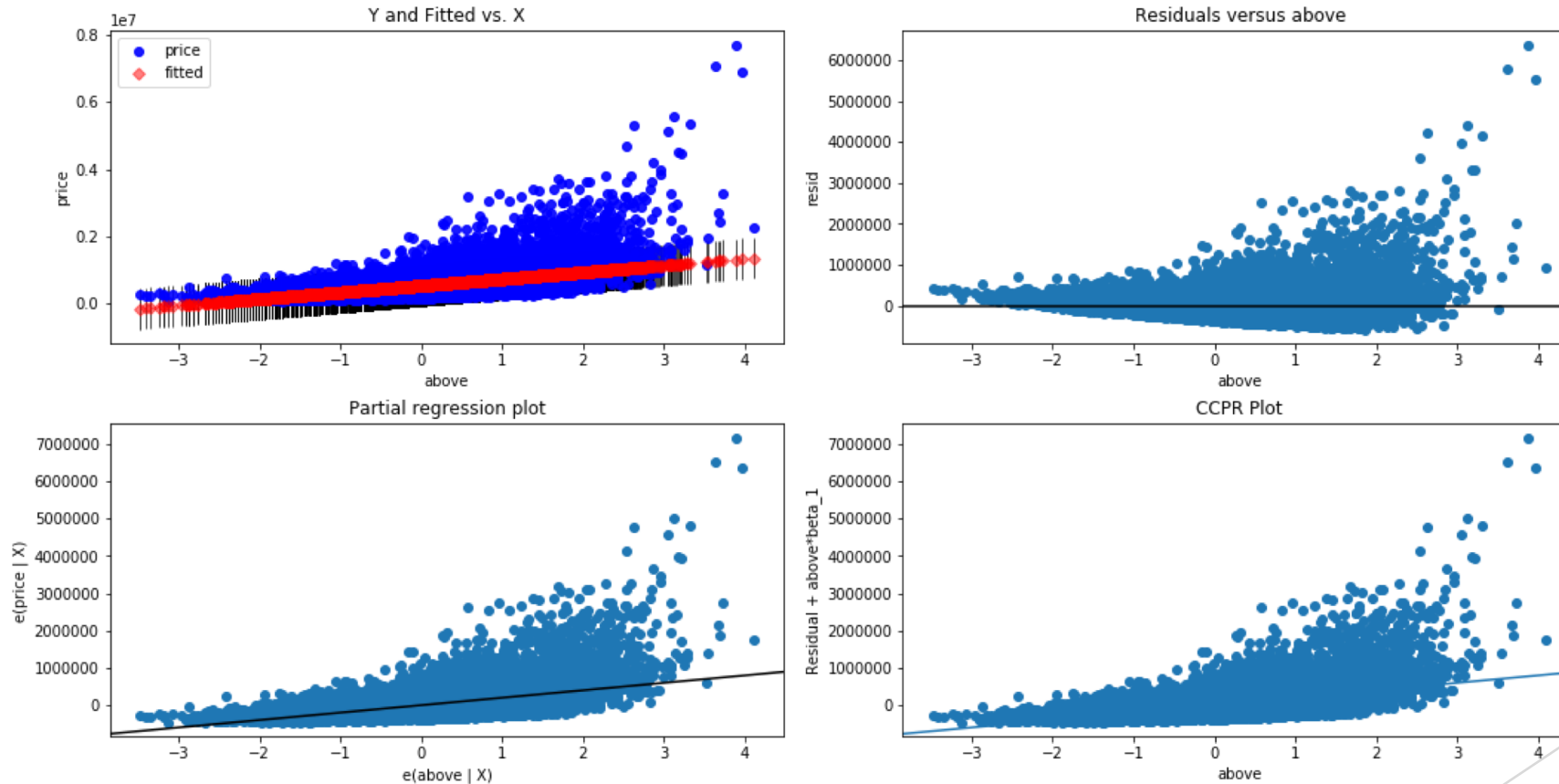
Regression Analysis of Kansas City Housing with Price as the Target
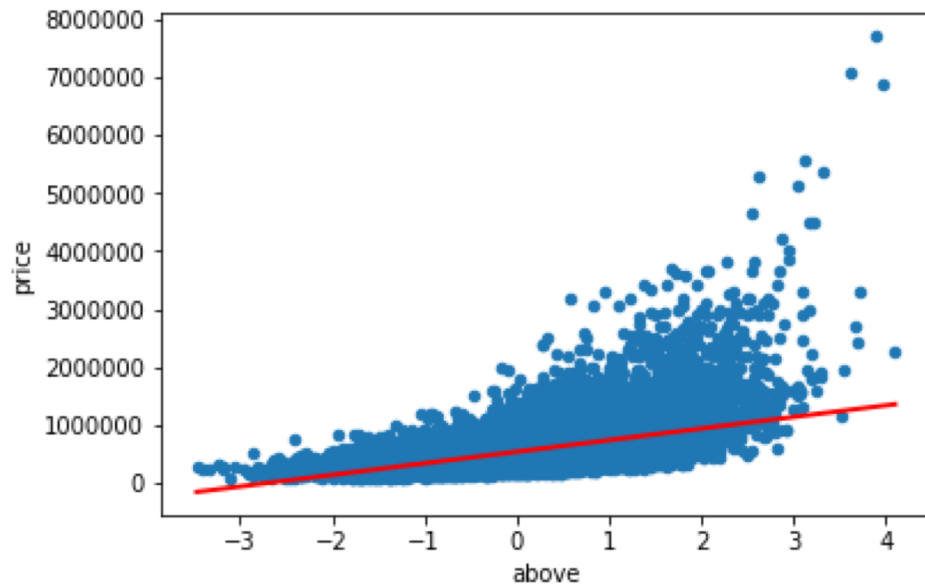
# Price vs Square Footage of House (apart from basement)

# Price vs Sq. Ft. Above 2

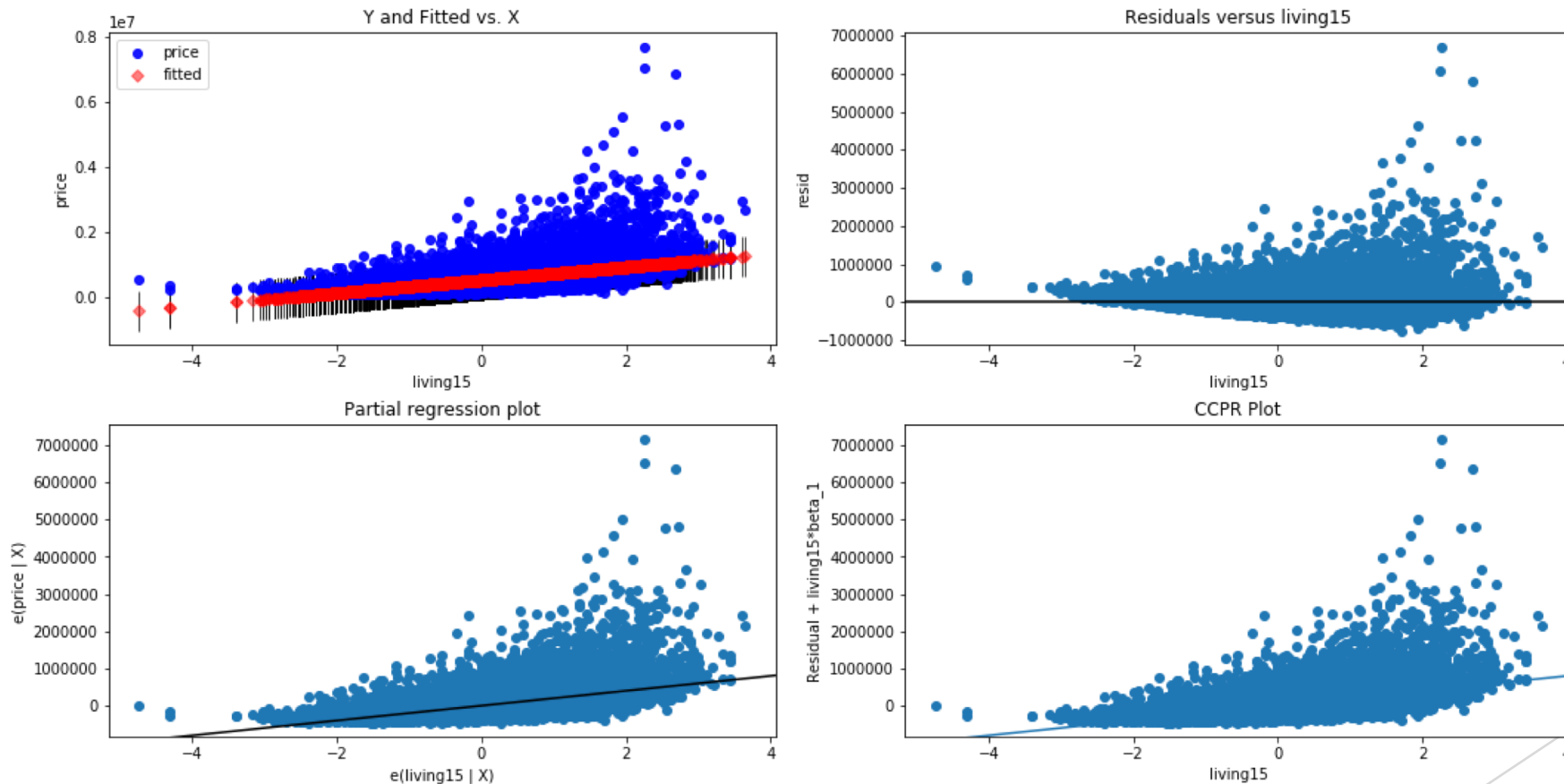|  | coef | std err | t | P>|t| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| above | 1.331e+05 | 2611.098 | 50.977 | 0.000 | 1.28e+05 | 1.38e+05 |



- P score of 0 indicating linear relationship
- Highly predictive coefficient to follow trend
- R-squared value of 0.294306 so not the best fit
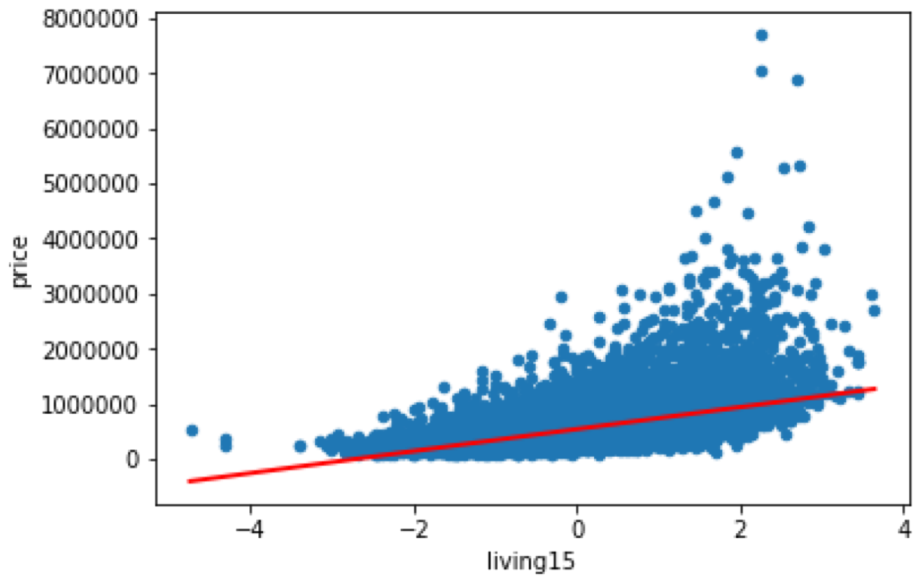- Some signs of heteroscedasticity

# Price vs Living Space (amongst nearest 15 neighbors)

# Price vs Living Space 2

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| living15 | 8.416e+04 | 2308.537 | 36.457 | 0.000 | 7.96e+04 | 8.87e+04 |



- R-squared value of 0.295098 meaning not a good fit

- Still highly predictive of the trend in price

- P score of 0

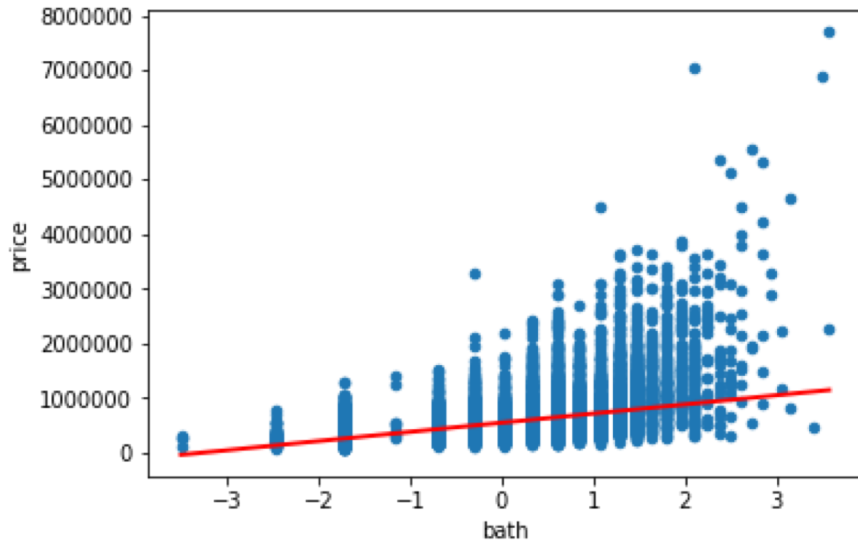- Some signs of heteroscedasticity and therefore non normal

# Price vs Number of Bathrooms
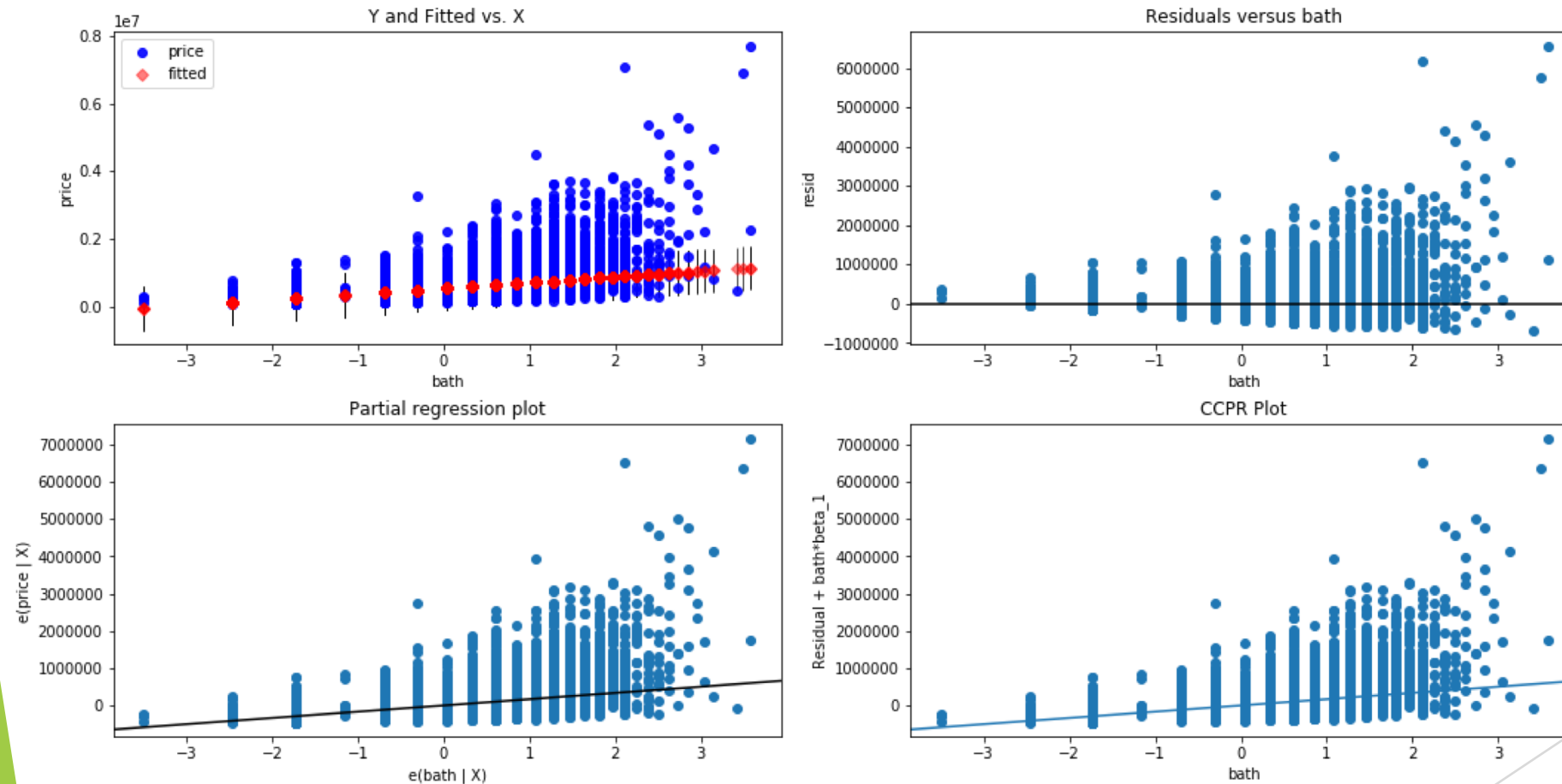
# Price vs Number of Bathrooms 2

| | coef | std err | t | P>|t| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| bath | 3.666e+04 | 2205.867 | 16.619 | 0.000 | 3.23e+04 | 4.1e+04 |



- R-squared score of 0.208114 again meaning not the best fit

- Yet still an excellent predictor of the direction of price

- P score of 0

- Clear linear and positive relationship

- Again signs of heteroscedasticity

# Model Performance

OLS Regression Results

| Dep. Variable: | price | R-squared: | 0.639 |
|---|---|---|---|
| Model: | OLS | Adj. R-squared: | 0.638 |
| Method: | Least Squares | F-statistic: | 517.3 |
| Date: | Thu, 22 Aug 2019 | Prob (F-statistic): | 0.00 |
| Time: | 15:45:53 | Log-Likelihood: | -2.9400e+05 |
| No. Observations: | 21420 | AIC: | 5.881e+05 |
| Df Residuals: | 21346 | BIC: | 5.887e+05 |
| Df Model: | 73 | | |
| Covariance Type: | nonrobust | | |

- ▶ R-squared score is low but much higher when considering all the parameters

- ▶ Coefficients include the previous 3 mentioned as well as categorical values for all the Zip Codes as well as Condition and Number of Bedrooms

# Test – Training Model



```
[price        4.402329e+10
dtype: float64, price        4.518984e+10
dtype: float64, price        4.873215e+10
dtype: float64, price        4.899870e+10
dtype: float64, price        4.705336e+10
dtype: float64]
[price        5.847043e+10
dtype: float64, price        5.389784e+10
dtype: float64, price        3.965527e+10
dtype: float64, price        3.858774e+10
dtype: float64, price        4.665075e+10
dtype: float64]
```

- ▶ Train Mean Squared Error: 45336513948.13085

- ▶ Test Mean Squared Error: 53244315653.16054

- ▶ Model is fairly predictive and not a huge difference in train/test splits

- ▶ Bottom image is residual results for training and testing K-fold cross-validation test