

Izaak King

Professor Eleish

Data Science

September 19, 2025

### “Lab 1”

For this lab, I looked at the following Linear Models:

- |  |                 |
|--|-----------------|
| 1. Predictor: PropertySqFt,              | Response: Price |
| 2. Predictor: Bath + Beds                | Response: Price |
| 3. Predictor: PropertySqFt + Bath        | Response: Price |
| 4. Predictor: PropertySqFt + Bath + Beds | Response: Price |

I examined their fit for the data model, under various cleaning techniques. I will display the summary of each model for comparison, with some plots at the end.

#### A. No cleaning applied

- |                             |                 |
|-----------------------------|-----------------|
| 1. Predictor: PropertySqFt, | Response: Price |
|-----------------------------|-----------------|

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -837809.4   610850.8  -1.372    0.17
PROPERTYSQFT    1462.7     189.2    7.729 1.31e-14 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31170000 on 4799 degrees of freedom
Multiple R-squared:  0.0123,    Adjusted R-squared:  0.01209
F-statistic: 59.74 on 1 and 4799 DF,  p-value: 1.307e-14
```

## 2. Predictor: Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-421655	752696	-0.560	0.575
BATH	1575414	367735	4.284	1.87e-05 ***
BEDS	-286349	275127	-1.041	0.298

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31260000 on 4798 degrees of freedom

Multiple R-squared: 0.006524, Adjusted R-squared: 0.00611

F-statistic: 15.75 on 2 and 4798 DF, p-value: 1.516e-07

## 3. Predictor: PropertySqFt + Bath

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1653140.0	728920.4	-2.268	0.0234 *
PROPERTYSQFT	1248.4	216.2	5.775	8.18e-09 ***
BATH	540624.0	263928.7	2.048	0.0406 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31150000 on 4798 degrees of freedom

Multiple R-squared: 0.01316, Adjusted R-squared: 0.01275

F-statistic: 31.99 on 2 and 4798 DF, p-value: 1.58e-14

## 4. Predictor: PropertySqFt + Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-1301012.4	764839.2	-1.701	0.089 .
PROPERTYSQFT	1275.1	216.9	5.880	4.38e-09 ***
BATH	958232.7	381191.2	2.514	0.012 *
BEDS	-417607.9	275076.5	-1.518	0.129

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31150000 on 4797 degrees of freedom

Multiple R-squared: 0.01363, Adjusted R-squared: 0.01302

F-statistic: 22.1 on 3 and 4797 DF, p-value: 3.291e-14

## B. Removal of outliers from Price

### 1. Predictor: PropertySqFt,

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.030e+05	1.733e+04	23.26	<2e-16 ***
PROPERTYSQFT	2.681e+02	8.208e+00	32.66	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 544600 on 4240 degrees of freedom

Multiple R-squared: 0.201, Adjusted R-squared: 0.2008

F-statistic: 1066 on 1 and 4240 DF, p-value: < 2.2e-16

### 2. Predictor: Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	356137	15152	23.504	<2e-16 ***
BATH	245777	9198	26.721	<2e-16 ***
BEDS	8612	5761	1.495	0.135

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 508900 on 4239 degrees of freedom

Multiple R-squared: 0.3025, Adjusted R-squared: 0.3022

F-statistic: 919.2 on 2 and 4239 DF, p-value: < 2.2e-16

### 3. Predictor: PropertySqFt + Bath

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.631e+05	1.670e+04	15.75	<2e-16 ***
PROPERTYSQFT	1.121e+02	9.393e+00	11.94	<2e-16 ***
BATH	2.042e+05	7.323e+03	27.88	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 500700 on 4239 degrees of freedom

Multiple R-squared: 0.3248, Adjusted R-squared: 0.3245

F-statistic: 1020 on 2 and 4239 DF, p-value: < 2.2e-16

### 4. Predictor: PropertySqFt + Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	264540.961	16786.888	15.76	<2e-16 ***
PROPERTYSQFT	113.762	9.585	11.87	<2e-16 ***
BATH	209515.665	9551.450	21.93	<2e-16 ***
BEDS	-5029.322	5783.781	-0.87	0.385

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 500700 on 4238 degrees of freedom

Multiple R-squared: 0.3249, Adjusted R-squared: 0.3245

F-statistic: 680 on 3 and 4238 DF, p-value: < 2.2e-16

## C. Removal of outliers from Price, PropertySqFt, Beds, Bath

### 1. Predictor: PropertySqFt,

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	440501.79	23815.95	18.50	<2e-16 ***
PROPERTYSQFT	201.65	13.98	14.43	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 517100 on 3328 degrees of freedom  
Multiple R-squared: 0.05887, Adjusted R-squared: 0.05859  
F-statistic: 208.2 on 1 and 3328 DF, p-value: < 2.2e-16

### 2. Predictor: Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	90463	24210	3.737	0.00019 ***
BATH	443305	15788	28.079	< 2e-16 ***
BEDS	-22366	10173	-2.199	0.02798 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 463200 on 3327 degrees of freedom  
Multiple R-squared: 0.245, Adjusted R-squared: 0.2446  
F-statistic: 539.9 on 2 and 3327 DF, p-value: < 2.2e-16

### 3. Predictor: PropertySqFt + Bath

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-37216.07	26581.78	-1.40	0.162
PROPERTYSQFT	96.38	12.92	7.46	1.1e-13 ***
BATH	396084.86	13324.91	29.73	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 459700 on 3327 degrees of freedom  
Multiple R-squared: 0.2564, Adjusted R-squared: 0.2559  
F-statistic: 573.5 on 2 and 3327 DF, p-value: < 2.2e-16

### 4. Predictor: PropertySqFt + Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-14000.18	27347.99	-0.512	0.608736
PROPERTYSQFT	104.02	13.08	7.953	2.48e-15 ***
BATH	425974.71	15793.03	26.972	< 2e-16 ***
BEDS	-35888.98	10221.48	-3.511	0.000452 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 459000 on 3326 degrees of freedom  
Multiple R-squared: 0.2591, Adjusted R-squared: 0.2584  
F-statistic: 387.7 on 3 and 3326 DF, p-value: < 2.2e-16

## D. Filter price data by log, rather than with outliers

### 1. Predictor: PropertySqFt,

Response: Price

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.789e+00  7.876e-03  735.01  <2e-16 ***
PROPERTYSQFT 8.263e-05  2.440e-06   33.87  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.4018 on 4799 degrees of freedom  
Multiple R-squared: 0.1929, Adjusted R-squared: 0.1927  
F-statistic: 1147 on 1 and 4799 DF, p-value: < 2.2e-16

### 2. Predictor: Bath + Beds

Response: Price

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.789e+00  7.876e-03  735.01  <2e-16 ***
PROPERTYSQFT 8.263e-05  2.440e-06   33.87  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.4018 on 4799 degrees of freedom  
Multiple R-squared: 0.1929, Adjusted R-squared: 0.1927  
F-statistic: 1147 on 1 and 4799 DF, p-value: < 2.2e-16

### 3. Predictor: PropertySqFt + Bath

Response: Price

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.645e+00  8.596e-03  656.68  <2e-16 ***
PROPERTYSQFT 4.476e-05  2.549e-06   17.56  <2e-16 ***
BATH         9.554e-02  3.112e-03   30.70  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.3674 on 4798 degrees of freedom  
Multiple R-squared: 0.3254, Adjusted R-squared: 0.3251  
F-statistic: 1157 on 2 and 4798 DF, p-value: < 2.2e-16

### 4. Predictor: PropertySqFt + Bath + Beds

Response: Price

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.652e+00  9.015e-03  626.928  <2e-16 ***
PROPERTYSQFT 4.530e-05  2.556e-06   17.724  <2e-16 ***
BATH         1.040e-01  4.493e-03   23.153  <2e-16 ***
BEDS        -8.485e-03  3.242e-03   -2.617   0.0089 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.3672 on 4797 degrees of freedom  
Multiple R-squared: 0.3264, Adjusted R-squared: 0.3259  
F-statistic: 774.7 on 3 and 4797 DF, p-value: < 2.2e-16

## E. Filter Price data by log, withall outliers are removed

### 1. Predictor: PropertySqFt,

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.596e+00	1.337e-02	418.4	<2e-16 ***
PROPERTYSQFT	1.185e-04	7.849e-06	15.1	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2904 on 3328 degrees of freedom  
Multiple R-squared: 0.06411, Adjusted R-squared: 0.06383  
F-statistic: 228 on 1 and 3328 DF, p-value: < 2.2e-16

### 2. Predictor: Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.359020	0.013235	404.918	<2e-16 ***
BATH	0.256903	0.008630	29.767	<2e-16 ***
BEDS	0.002691	0.005561	0.484	0.629

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2532 on 3327 degrees of freedom  
Multiple R-squared: 0.2886, Adjusted R-squared: 0.2882  
F-statistic: 674.9 on 2 and 3327 DF, p-value: < 2.2e-16

### 3. Predictor: PropertySqFt + Bath

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.301e+00	1.452e-02	365.19	< 2e-16 ***
PROPERTYSQFT	5.362e-05	7.056e-06	7.60	3.83e-14 ***
BATH	2.441e-01	7.277e-03	33.55	< 2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2511 on 3327 degrees of freedom  
Multiple R-squared: 0.3007, Adjusted R-squared: 0.3003  
F-statistic: 715.3 on 2 and 3327 DF, p-value: < 2.2e-16

### 4. Predictor: PropertySqFt + Bath + Beds

Response: Price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.304e+00	1.496e-02	354.529	< 2e-16 ***
PROPERTYSQFT	5.456e-05	7.156e-06	7.624	3.18e-14 ***
BATH	2.478e-01	8.640e-03	28.682	< 2e-16 ***
BEDS	-4.402e-03	5.592e-03	-0.787	0.431

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2511 on 3326 degrees of freedom  
Multiple R-squared: 0.3008, Adjusted R-squared: 0.3002  
F-statistic: 477 on 3 and 3326 DF, p-value: < 2.2e-16

Three particularly good models we found from different cleaning methods are:

Predictor: PropertySqFt + Bath

Response: Price

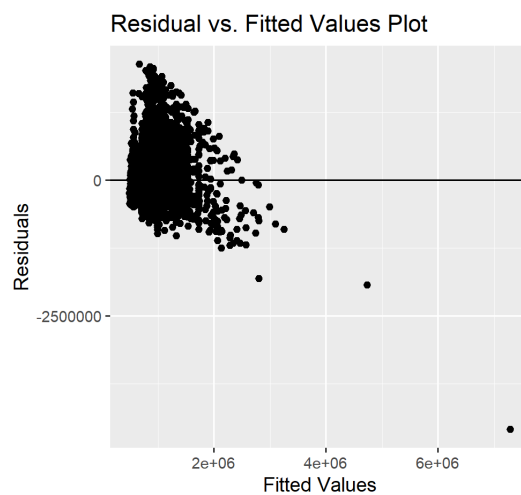
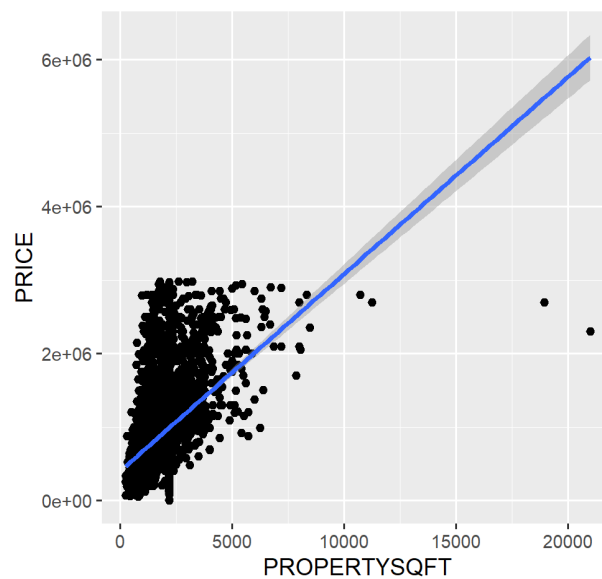
- Cleaning: Removal of outliers in price

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.631e+05	1.670e+04	15.75	<2e-16 ***
PROPERTYSQFT	1.121e+02	9.393e+00	11.94	<2e-16 ***
BATH	2.042e+05	7.323e+03	27.88	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 500700 on 4239 degrees of freedom  
Multiple R-squared: 0.3248, Adjusted R-squared: 0.3245  
F-statistic: 1020 on 2 and 4239 DF, p-value: < 2.2e-16



Predictor: PropertySqFt + Bath + Beds

Response: Price

- Filter price data by log, rather than with outliers

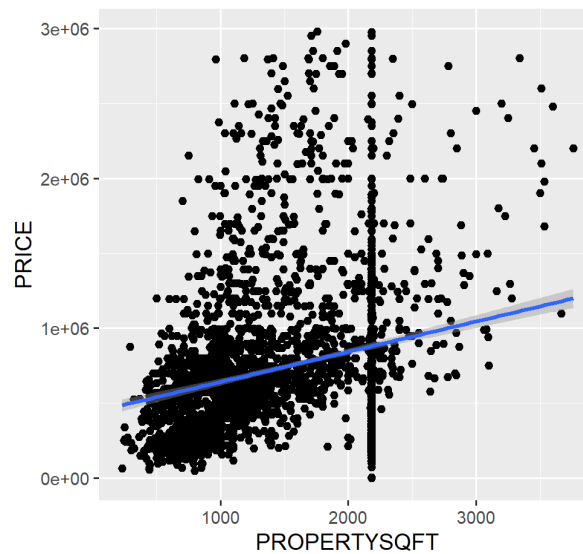
Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	5.652e+00	9.015e-03	626.928	<2e-16	***
PROPERTYSQFT	4.530e-05	2.556e-06	17.724	<2e-16	***
BATH	1.040e-01	4.493e-03	23.153	<2e-16	***
BEDS	-8.485e-03	3.242e-03	-2.617	0.0089	**

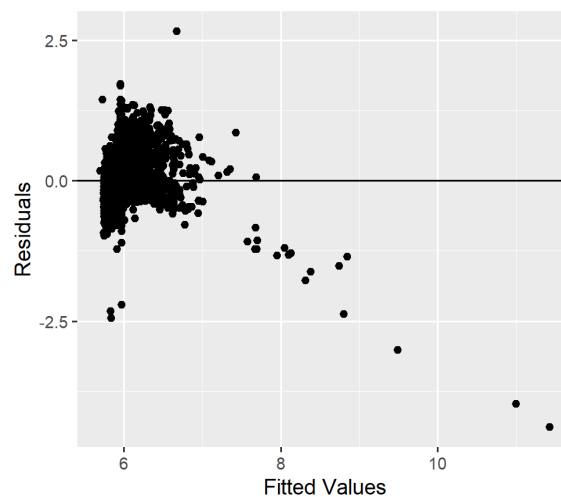
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3672 on 4797 degrees of freedom  
Multiple R-squared: 0.3264, Adjusted R-squared: 0.3259  
F-statistic: 774.7 on 3 and 4797 DF, p-value: < 2.2e-16



Residual vs. Fitted Values Plot





Predictor: PropertySqFt + Bath

Response: Price

- Filter Price data by log, once all outliers are removed

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5.301e+00	1.452e-02	365.19	< 2e-16 ***
PROPERTYSQFT	5.362e-05	7.056e-06	7.60	3.83e-14 ***
BATH	2.441e-01	7.277e-03	33.55	< 2e-16 ***

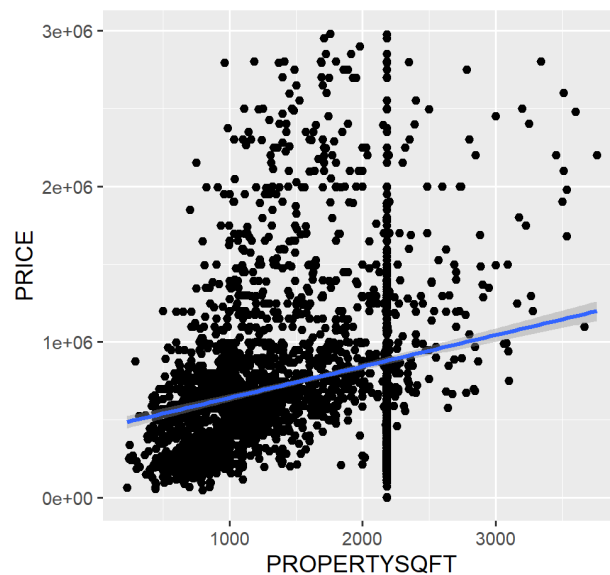
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2511 on 3327 degrees of freedom

Multiple R-squared: 0.3007, Adjusted R-squared: 0.3003

F-statistic: 715.3 on 2 and 3327 DF, p-value: < 2.2e-16



Residual vs. Fitted Values Plot

